Interior-Point Algorithm: Theory and Analysis

Yinyu Ye

June 1996

To my family and parents
Daisun, Fei, Xienshu and Chunxuan

Yinyu Ye, Professor
Department of Management Sciences
The University of Iowa
Iowa City, IA 52242-1000
Phone: (319) 335-1947
E-mail: yyye@dollar.biz.uiowa.edu

Preface

On a sunny afternoon in 1984, one of my officemates told me that there would be a seminar given by N. Karmarkar, an AT&T scientist, on linear programming. At the time, my knowledge of linear programming was limited to one optimization course and one research project with Professor David Luenberger in the Department of Engineering-Economic Systems, Stanford University. As a second-year Ph. D. student, I was familiar with economic terms like cost, supply, demand, and price, rather than mathematical terms like vertex, polyhedron, inequality, and duality.

That afternoon the Terman auditorium was packed. I could not find a seat and had to sit on the floor for about one and a half hour, where I saw Professor George Dantzig and many other faculty members. I was not particular enthusiastic about the statement from the speaker that a new interior-point method would be 40 times faster than the simplex method but I was amazed by the richness and applicability of linear programming as a whole. That was how and when I determined to devote my Ph. D. study to mathematical programming.

I immediately took a series of courses from George. In those courses I was fortunate to learn from many distinguished researchers. I also went to Cornell to work under the guidance of Professor Michael Todd. Since then, my interest in linear programming has become stronger and my knowledge of interior-point algorithms has grown broader.

I decided to write a monograph about my understanding of interior-point algorithms and their complexities. I chose to highlight the underlying interior-point geometry, combinatorica, and potential theory for convex inequalities. I did not intend to cover the entire progress of linear programming and interior-point algorithms during the last decade in this write-up. For a complete survey, I refer the reader to several excellent articles or books by Goldfarb and Todd [116], Gonzaga [127], den Hertog [145], Nesterov and Nemirovskii [263], Terlaky [320], Todd [322, 328], Wright [366], etc.

Special thanks go to Erling Andersen for giving me detailed feedback from a course he taught in Odense University using an early version of the manuscript, and to Nadine Castellano for her proof read of the manuscript. I wish to thank my colleagues Kurt Anstreicher, Dingzhu Du, Panos Pardalos, Mike Todd, and Steve Vavasis for their support and advice on the process of writing the manuscript. I am grateful to the University of Iowa for the environment and resources which made this effort possible. This book owes a great deal to the students and other participants in the course

6K287 at the Department of Management Science. I also wish to thank the Institute of Applied Mathematics of Chinese Academy and Huazhong University of Science and Technology for their support and assistance during my many visits there.

My research efforts have been supported by the following grants and awards at various institutes: National Science Foundation Grants DDM-8922636, DDM-9207347 and DMI-9522507; the summer grants of the College of Business of Administration, the Obermann Fellowship, and the Center for Advanced Studies Interdisciplinary Research Grant at the University of Iowa; the Cornell Center for Applied Mathematics and the Advanced Computing Research Institute, a unit of the Cornell Theory Center, which receives major funding from the NSF and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute; NSF Coop. Agr. No. CCR-8809615 at the Center for Research in Parallel Computation of Rice University; the award of K. C. WONG Education Foundation at Hong Kong; and Dutch Organization for Scientific Research (NWO).

Finally, I thank my wife Daisun Zhou and daughter Fei for their love and support, and my parents Xienshu Ye and Chunxuan Wang for their inspiration.

Yinyu Ye
Iowa City, 1996.

Contents

List of Figures

Chapter 1

Introduction and Preliminaries

1.1 Introduction

Complexity theory is the foundation of computer algorithms. The goal of the theory is to develop criteria for measuring the effectiveness of various algorithms and the difficulty of various problems. The term "complexity" refers to the amount of resources required by a computation. In this book, running time or number of arithmetic operations is the major resource of interest.

Linear programming, hereafter LP, plays a very important role in complexity analysis. In one sense it is a continuous optimization problem in minimizing a linear objective function over a convex polyhedron; but it is also a combinatorial problem involving selecting an extreme point among a finite set of possible vertices. Businesses, large and small, use linear programming models to optimize communication systems, to schedule transportation networks, to control inventories, to plan investments, and to maximize productivity.

Linear inequalities define a polyhedron, properties of which have been studied by mathematicians for centuries. Ancient Chinese and Greeks studied calculating volumes of simple polyhedra in three-dimensional space. Fourier's fundamental research connecting optimization and inequalities dates back to the early 1800s. At the end of 19th century, Farkas and Minkowski began basic work on algebraic aspects of linear inequalities. In 1910 De La Vallée Poussin developed an algebraic technique for minimizing the infinity-norm of $b - Ax$ that can be viewed as a precursor of the simplex method. Beginning in the 1930s, such notable mathematicians as von Neumann, Kantorovich, and Koopmans studied mathematical economics based on linear inequalities. During World War II, it was observed that decisions involving the best movement of personnel and optimal allocation of resources could be posed and solved as linear programs. Linear programming began to assume its current popularity.

An optimal solution of a linear program always lies at a vertex of the feasible region, which itself is a polyhedron. Unfortunately, the number of vertices associated with a set of n inequalities in m variables can be exponential in the dimensions—in this case, up to $n!/m!(n-m)!$. Except for small values of m and n, this number is so large as to prevent examining all possible vertices for searching an optimal vertex.

The simplex method, invented in the mid-1940s by George Dantzig, is a procedure for examining optimal candidate vertices in an intelligent fashion. It constructs a sequence of adjacent vertices with improving values of the objective function. Thus, the method travels along edges of the polyhedron until it hits an optimal vertex. Improved in various way in the intervening four decades, the simplex method continues to be the workhorse algorithm for solving linear programming problems. On average, the number of vertices or iterations visited by the simplex method seems to be roughly linear in m and perhaps logarithmic n.

Although it performs well on average, the simplex method will indeed examine every vertex when applied to certain linear programs. Klee and Minty in 1972 gave such an example. These examples confirm that in the worst case, the simplex method needs an exponential number of iterations to find the optimal solution. As interest in complexity theory grew, many researchers believed that a good algorithm should be polynomial—i.e., broadly speaking, the running time required to compute the solution should be bounded above by a polynomial in the "size," or the total data length, of the problem. Thus, the simplex method is not a polynomial algorithm.

In 1979, a new approach to linear programming, Khachiyan's ellipsoid method, received dramatic and widespread coverage in the international press. Khachiyan proved that the ellipsoid method, developed during the 1970s by other mathematicians, is a polynomial algorithm for linear programming under a certain computational model. It constructs a sequence of shrinking ellipsoids with two properties: the current ellipsoid always contains the optimal solution set, and each member of the sequence undergoes a guaranteed reduction in volume, so that the solution set is squeezed more tightly at each iteration.

The ellipsoid method was studied intensively by practitioners as well as theoreticians. Based on the expectation that a polynomial linear programming algorithm would be faster than the simplex method, it was a great disappointment that the best implementations of the ellipsoid method were not even close to being competitive. In contrast to the simplex method, the number of steps required for the ellipsoid method to terminate was almost always close to the worst case bound—whose value, although defined by a polynomial, is typically very large. Thus, after the dust eventually settled, the prevalent view among linear programming researchers was that Khachiyan had proved a genuinely polynomial linear programming algorithm, but the simplex method remained the clear winner in practice.

This contradiction, the fact that an algorithm with the desirable theoretical property of polynomiality might nonetheless compare unfavorably with the (worst-case exponential) simplex method, set the stage for exciting new developments. It was no wonder, then, that the announcement by Karmarkar in 1984 of a new polynomial interior-point algorithm with the potential to dramatically improve the practical effectiveness of the simplex method made front-page news in major newspapers and magazines throughout the world.

Interior-point algorithms are continuous iterative algorithms. Computation experience with sophisticated procedures suggests that the number of iterations necessarily grows much more slowly than the dimension grows. Furthermore, they have an established worst-case polynomial iteration bound, providing the potential for dramatic improvement in computation effectiveness. The success of interior-point algorithms also brought much attention to complexity theory itself.

The goal of the book is to describe some of these recent developments and to suggest a few directions in which future progress might be made. The book is organized as follows. In Chapter 1, we discuss some necessary mathematical preliminaries. We also present models of computation and several basic optimization problems used throughout the text.

Chapter 2 is devoted to studying the geometry of inequalities and interior-point algorithms. At first glance, interior-point algorithms seem less geometric than the simplex or the ellipsoid methods. Actually, they also possess many rich geometric concepts. These concepts, such as "center," "volume," and "potential" of a polytope, are generally "non-combinatorial." These geometries are always helpful for teaching, learning and research.

In Chapter 3 we present some basic algorithms to compute a so-called analytic center, or, equivalently, to minimize a potential function for a polytope. They are key elements underlying interior-point algorithms. Then, we present several interior-point linear programming algorithms in Chapter 4. It is impossible to list all the literature in this field. Here, we select four algorithms: Karmarkar's projective algorithm, the path-following algorithm, the potential reduction algorithm, and the primal-dual algorithm including the predictor-corrector algorithm.

We analyze the worst-case complexity bound for interior-point algorithms in Chapter 5. The main issues are arithmetic operation, termination, and initialization techniques. We will use the real number computation model in our analysis because of the continuous nature of interior-point algorithms. We also compare the complexity theory with the convergence rate used in numerical analysis.

The worst-case complexity bound hardly serves as a practical criterion for judging the efficiency of algorithms. We will discuss a common phenomenon arising from using interior-point algorithms for solving optimization problems. It is often observed that effectiveness of an algorithm is dependent on the dimension or size of a problem instance as well as a parameter, called "condition number," inherited in the problem. This condition number represents the degree of difficulty of the problem instance. For two problems having the same dimension but different condition numbers, an algorithm may have drastically different performances. This classification will help us to understand algorithm efficiency and possibly improve the condition and, therefore, improve the complexity of the problem. We present condition-based complexity results for LP interior-point algorithms in Chapter 5.

While most of research has been focused on the worst-case performance of interior-point algorithms, many other complexity results were quietly established during the past several years. We try to cover these less-noticeable but significant results. In particular, we present some average and probabilistic complexity results in Chapter 6 and some asymptotic complexity (local convergence) results in Chapter 7. Average complexity bounds have been successfully established for the simplex method, and asymptotic or local convergence rates have been widely accepted by the numerical and continuous optimization community as major criteria in judging efficiency of iterative procedures.

Not only has the complexity of LP algorithms been significantly improved during the last decade, but also the problem domain solvable by interior-point algorithms has dramatically widened. We present complexity results for fractional programming, convex programming, positive semi-definite programming and non-polyhedron optimization in Chapter 8. We also discuss some approximation complexity results for solving nonconvex optimization problems in Chapter 9.

1.2 Mathematical Preliminaries

This section summarizes mathematical background material for linear algebra, linear programming, and nonlinear optimization.

1.2.1 Basic notations

By $\mathcal{R}$ we denote the set of real numbers. $\mathcal{R}_+$ denotes the set of nonnegative real numbers, and $\overset{\circ}{\mathcal{R}}_+$ denotes the set of positive numbers. For a natural number n, the symbol $\mathcal{R}^n$ ($\mathcal{R}^n_+$, $\overset{\circ}{\mathcal{R}}{}^n_+$) denotes the set of vectors with n components in $\mathcal{R}$ ($\mathcal{R}_+$, $\overset{\circ}{\mathcal{R}}_+$).

Addition of vectors and multiplication of vectors with scalars are standard. The vector inequality $x \geq y$ means $x_j \geq y_j$ for $j = 1, 2, ..., n$. 0 represents a vector whose entries are all zeros and e represents a vector whose entries are all ones, where their dimensions may vary according to other vectors in expressions. A vector is always considered as a column vector. The superscript "T" denotes transpose operation. The inner product in $\mathcal{R}^n$ is defined as follows:

$$x^T y := \sum_{j=1}^{n} x_j y_j \quad \text{for} \quad x, y \in \mathcal{R}^n.$$

The l_2 norm of a vector x is given by

$$\|x\| = \sqrt{x^T x},$$

and the l_∞ norm is

$$\|x\|_\infty = \max\{|x_1|, |x_2|, ..., |x_n|\}.$$

In general, the p norm is

$$\|x\|_p = \left(\sum_{1}^{n} |x_j|^p \right)^{1/p}, \quad p = 1, 2, ...$$

The dual of the p norm is the q norm where

$$\frac{1}{p} + \frac{1}{q} = 1.$$

For natural numbers m and n, $\mathcal{R}^{m \times n}$ denotes the set of real matrices with m rows and n columns. For $A \in \mathcal{R}^{m \times n}$, we assume that the row index set of A is $\{1, 2, ..., m\}$ and the column index set is $\{1, 2, ..., n\}$. The ith row of A is denoted by $a_{i.}$ and the jth column of A is denoted by $a_{.j}$; the i and jth component of A is denoted by a_{ij}. If I is a subset of the row index set and J is a subset of the column index set, then A_I denotes the submatrix of A whose rows belong to I, A_J denotes the submatrix of A whose columns belong to J, A_{IJ} denotes the submatrix of A induced by those components of A whose indices belong to I and J, respectively.

$\mathcal{M}^n$ denotes the space of symmetric matrices in $\mathcal{R}^{n\times n}$. The identity matrix is denoted by I. The null space of A is denoted $\mathcal{N}(A)$ and the range of A is $\mathcal{R}(A)$.

The determinant of an $n \times n$-matrix A is denoted by $\det(A)$. The trace of A, denoted by $\operatorname{tr}(A)$, is the sum of the diagonal entries in A. For a vector $x \in \mathcal{R}^n$, the upper case X represents a diagonal matrix in $\mathcal{R}^{n\times n}$ whose diagonal entries are the entries of x, i.e.,

$$X = \operatorname{diag}(x).$$

A matrix $Q \in \mathcal{R}^{n\times n}$ is said to be positive definite (PD) if

$$x^T Q x > 0, \quad \text{for all} \quad x \neq 0,$$

and positive semi-definite (PSD) if

$$x^T Q x \geq 0, \quad \text{for all} \quad x.$$

$\{x^k\}_0^\infty$ is an ordered sequence $x^1, x^2, ..., x^k, ...$ A sequence $\{x^k\}_0^\infty$ is convergent to $\bar{x}$, denoted $x^k \to \bar{x}$, if

$$\|x^k - \bar{x}\| \to 0.$$

A point x is a limit point of $\{x^k\}_0^\infty$ if there is a subsequence of $\{x^k\}$ convergent to x.

If $g(x) \geq 0$ is a real valued function of a real nonnegative variable, the notation $g(x) = O(x)$ means that $g(x) \leq \bar{c}x$ for some constant $\bar{c}$; the notation $g(x) = \Omega(x)$ means that $g(x) \geq \underline{c}x$ for some constant $\underline{c}$; the notation $g(x) = \theta(x)$ means that $\underline{c}x \leq g(x) \leq \bar{c}x$. Another notation is $g(x) = o(x)$, which means that $g(x)$ goes to zero faster than x does:

$$\lim_{x\to 0} \frac{g(x)}{x} = 0.$$

1.2.2 Convex sets

If x is a member of the set Ω, we write $x \in \Omega$; if y is not a member of Ω, we write $y \notin \Omega$. The union of two sets S and T is denoted $S \cup T$; the intersection of them is denoted $S \cap T$. A set can be specified in the form $\Omega = \{x : \ P(x)\}$ as the set of all elements satisfying property P.

For $y \in \mathcal{R}^n$ and $\epsilon > 0$, $B(y, \epsilon) = \{x : \ \|x - y\| \leq \epsilon\}$ is the ball or sphere of radius ϵ with center y. In addition, for a positive definite matrix Q of dimension n, $E(y, Q) = \{x : \ (x - y)^T Q(x - y) \leq 1\}$ is called an ellipsoid. The vector y is the center of E. When Q is diagonal, $E(y, Q)$ is called a regular ellipsoid (Figure 1.1).

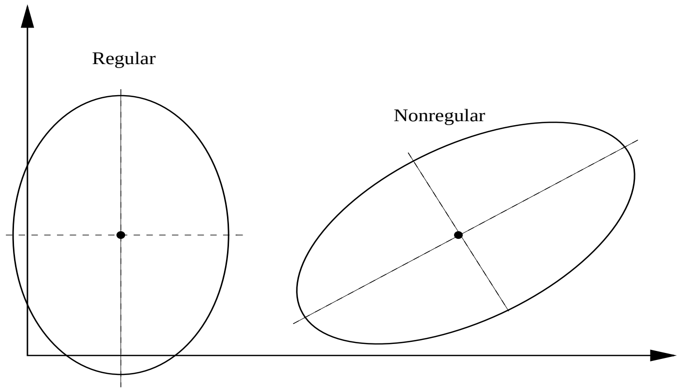

Figure 1.1: Regular and nonregular ellipsoids

A set Ω is closed if $x^k \to x$, where $x^k \in \Omega$, implies $x \in \Omega$. A set Ω is open if around every point $y \in \Omega$ there is a ball that is contained in Ω, i.e., there is an $\epsilon > 0$ such that $B(y, \epsilon) \subset \Omega$. A set is bounded if it is contained within a ball with finite radius. A set is compact if it is both closed and bounded. The (topological) interior of any set Ω, denoted $\mathring{\Omega}$, is the set of points in Ω which are the centers of some balls contained in Ω. The closure of Ω, denoted $\bar{\Omega}$, is the smallest closed set containing Ω. The boundary of Ω is the part of $\bar{\Omega}$ that is not in $\mathring{\Omega}$.

A set C is said to be convex if for every $x^1, x^2 \in C$ and every real number α, $0 < \alpha < 1$, the point $\alpha x^1 + (1 - \alpha)x^2 \in C$. The convex hull of a set Ω is the intersection of all convex sets containing Ω.

A set C is a cone if $x \in C$ implies $\alpha x \in C$ for all $\alpha > 0$. A cone that is also convex is a convex cone. For a cone $C \subset \Omega$, the dual of C is the cone

$$C^* := \{y : \langle x, y \rangle \geq 0 \quad \text{for all} \quad x \in C\},$$

where $\langle , \rangle$ is an inner product operation for space Ω. A cone C is polyhedral if

$$C = \{x : Ax \leq 0\}$$

for some matrix A (Figure 1.2).

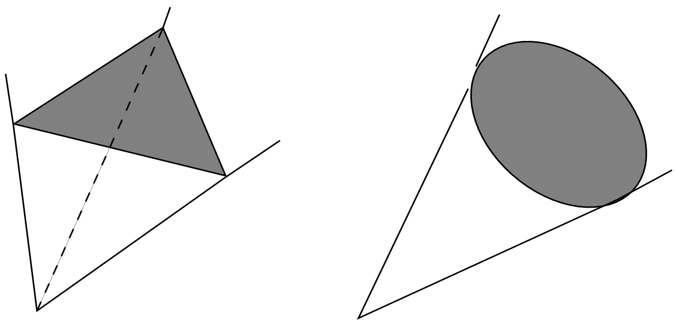

Figure 1.2: Polyhedral and nonpolyhedral cones

Theorem 1.1 *A convex cone C is polyhedral if and only if it is finitely generated, that is, the cone is generated by a finite number of vectors $b_1,\ldots,b_m$:*

$$C = cone(b_1, \ldots, b_m) := \left\{ \sum_{i=1}^{m} b_i y_i : \ y_i \geq 0, \ i = 1, \ldots, m \right\}.$$

Example 1.1 *The n-dimensional positive orthant, $\mathcal{R}^n_+ := \{x \in \mathcal{R}^n : \ x \geq 0\}$, is a convex cone. The set of all n-dimensional positive semi-definite matrices is a convex cone, called the positive semi-definite matrix cone. The set $\{(t, x \in \mathcal{R}^n) : \ t \geq \|x\|\}$ is a convex cone, called the second-order cone.*

Theorem 1.2 *(Carathéodory's theorem) Let convex polyhedral cone $C = cone(b_1, \ldots, b_m)$ and $x \in C$. Then, $x \in cone(b_{i_1}, \ldots, b_{i_d})$ for some linearly independent vectors $b_{i_1},\ldots,b_{i_d}$ chosen from $b_1,\ldots,b_m$.*

The most important type of convex set is a hyperplane. Hyperplanes dominate the entire theory of optimization. Let a be a nonzero n-dimensional vector, and let b be a real number. The set

$$H = \{x \in \mathcal{R}^n : a^T x = b\}$$

is a hyperplane in $\mathcal{R}^n$ (Figure 1.3). Relating to hyperplane, positive and negative closed half spaces are given by

$$H_+ = \{x : a^T x \geq b\}$$

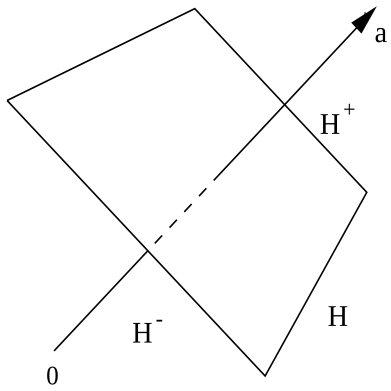

Figure 1.3: A hyperplane and half-spaces

$$H_{-} = \{x : a^T x \leq b\}.$$

A set which can be expressed as the intersection of a finite number of closed half spaces is said to be a convex polyhedron:

$$P = \{x : Ax \leq b\}.$$

A bounded polyhedron is called polytope.

Let P be a polyhedron in $\mathcal{R}^n$, F is a face of P if and only if there is a vector c for which F is the set of points attaining max $\{c^T x : x \in P\}$ provided the this maximum is finite. A polyhedron has only finite many faces; each face is a nonempty polyhedron.

The most important theorem about the convex set is the following separating theorem (Figure 1.4).

Theorem 1.3 *(Separating hyperplane theorem) Let $C \subset \mathcal{E}$ be a closed convex set and let y be a point exterior to C. Then there is a vector $a \in \mathcal{E}^*$ such that*

$$\langle a, y \rangle < \inf_{x \in C} \langle a, x \rangle.$$

Here $\mathcal{E}$ is a finite-dimensional real vector space and $\mathcal{E}^*$ is the dual space of $\mathcal{E}$. (In this book we often consider $\mathcal{E}$ being $\mathcal{R}^n$ or $\mathcal{M}^n$, so that $\mathcal{E}^* = \mathcal{E}$.)

1.2.3 Real functions

The real function $f(x)$ is said to be continuous at x if $x^k \to x$ implies $f(x^k) \to f(x)$. A continuous function f defined on a compact set Ω has

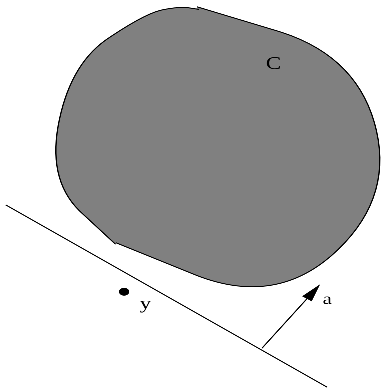

Figure 1.4: Illustration of the separating hyperplane theorem; an exterior point y is separated by a hyperplane from a convex set C.

a minimizer in Ω; that is, there is an $x^* \in \Omega$ such that for all $x \in \Omega$, $f(x) \geq f(x^*)$. This result is called the Weierstrass theorem.

A set of real-valued function $f_1, f_2, ..., f_m$ defined on $\mathcal{R}^n$ can be written as a single vector function $f = (f_1, f_2, ..., f_m)^T \in \mathcal{R}^m$. If f has continuous partial derivatives of order p, we say $f \in C^p$. The gradient vector of a real-valued function $f \in C^1$ is a row vector

$$\nabla f(x) = \{\partial f / \partial x_i\}, \quad \text{for} \quad i = 1, ..., n.$$

If $f \in C^2$, we define the Hessian of f to be the n-dimensional symmetric matrix

$$\nabla^2 f(x) = \left\{ \frac{\partial^2 f}{\partial x_i \partial x_j} \right\} \quad \text{for} \quad i = 1, ..., n; \; j = 1, ..., n.$$

If $f = (f_1, f_2, ..., f_m)^T \in \mathcal{R}^m$, then the Jacobian matrix of f is

$$\nabla f(x) = \begin{pmatrix} \nabla f_1(x) \\ \ldots \\ \nabla f_m(x) \end{pmatrix}.$$

f is a (continuous) convex function if and only if for $0 \leq \alpha \leq 1$,

$$f(\alpha x + (1 - \alpha) y) \leq \alpha f(x) + (1 - \alpha) f(y).$$

f is a (continuous) quasi-convex function if and only if for $0 \leq \alpha \leq 1$,

$$f(\alpha x + (1 - \alpha) y) \leq \max[f(x), f(y)].$$

Thus, a convex function is a quasi-convex function. The level set of f is given by

$$L_b = \{x : \quad f(x) \leq b\}.$$

f is a quasi-convex function implies that the level set of f is convex for any given b.

We have several propositions for real functions.

Theorem 1.4 *(Taylor expansion) Let $f \in C^2$ be in a region containing the line segment $[x, y]$. Then there is a $0 \leq \alpha \leq 1$ such that*

$$f(y) = f(x) + \nabla f(x)(y - x) + (1/2)(y - x)^T \nabla^2 f(\alpha x + (1 - \alpha) y)(y - x).$$

Proposition 1.5 *Let $f \in C^1$. Then f is convex over a convex set Ω if and only if*

$$f(y) \geq f(x) + \nabla f(x)(y - x)$$

for all $x,\ y \in \Omega$.

Proposition 1.6 *Let $f \in C^2$. Then f is convex over a convex set Ω if and only if the Hessian matrix of f is positive semi-definite throughout Ω.*

1.2.4 Inequalities

There are several important inequalities that are frequently used in algorithm design and complexity analysis.

Cauchy-Schwarz: given $x, y \in \mathcal{R}^n$ then

$$x^T y \leq \|x\| \|y\|.$$

Arithmetic-geometric mean: given $x \in \mathcal{R}^n_+$

$$\frac{\sum x_j}{n} \geq \left(\prod x_j\right)^{1/n}.$$

Harmonic: given $x \in \overset{\circ}{\mathcal{R}}{}^n_+$

$$\left(\sum x_j\right)\left(\sum 1/x_j\right) \geq n^2.$$

Hadamard: given $A \in \mathcal{R}^{m \times n}$ with columns $a_1, a_2, ..., a_n$, then

$$\sqrt{\det(A^T A)} \leq \prod \|a_j\|.$$

1.3 Decision and Optimization Problems

A decision or optimization problem has a certain form that is usually characterized by the number of decision variables and the number of constraints. A problem, $\mathcal{P}$, consists of two sets, data set $\mathcal{Z}_p$ and solution set $\mathcal{S}_p$. In general, $\mathcal{S}_p$ can be implicitly defined by the so-called optimality conditions. The solution set may be empty, i.e., problem $\mathcal{P}$ may have no solution.

In what follows, we list several decision and optimization problems that we will use later.

1.3.1 System of linear equations

Given $A \in \mathcal{R}^{m \times n}$ and $b \in \mathcal{R}^m$, the problem is to solve m linear equations for n unknowns:

$$Ax = b.$$

The data and solution sets are

$$\mathcal{Z}_p = \{A \in \mathcal{R}^{m \times n}, b \in \mathcal{R}^m\} \quad \text{and} \quad \mathcal{S}_p = \{x \in \mathcal{R}^n : Ax = b\}.$$

$\mathcal{S}_p$ is usually called an affine set. Given an x, one can easily check to see if x is in $\mathcal{S}_p$.

Theorem 1.7 *Each linear subspace of $\mathcal{R}^n$ is generated by finitely many vectors, and is also the intersection of finitely many linear hyperplanes; that is, for each linear subspace of L of $\mathcal{R}^n$ there are matrices A and C such that $L = \mathcal{N}(A) = \mathcal{R}(C)$.*

Theorem 1.8 *Let $A \in \mathcal{R}^{m \times n}$ and $b \in \mathcal{R}^m$. The system $\{x : \ Ax = b\}$ has a solution if and only if that $A^T y = 0$ implies $b^T y = 0$.*

A vector y, with $A^T y = 0$ and $b^T y \neq 0$, is called a infeasibility certificate for the system $\{x : \ Ax = b\}$.

1.3.2 Linear least-squares problem

Given $A \in \mathcal{R}^{m \times n}$ and $c \in \mathcal{R}^n$, the problem is to find an $s \in \mathcal{R}(A^T)$ such that $\|s - c\|$ is minimized. We can write the problem in the following format:

$$(LS) \quad \begin{array}{ll} \text{minimize} & \|s - c\|^2 \\ \text{subject to} & s \in \mathcal{R}(A^T). \end{array}$$

or

$$(LS) \quad \begin{array}{ll} \text{minimize} & \|A^T y - c\|^2 \\ \text{subject to} & y \in \mathcal{R}^m. \end{array}$$

In this format, the term $\|A^T y - c\|^2$ is called the objective function, y is called decision variables. Since y can be any point in $\mathcal{R}^m$, we say this (optimization) problem is unconstrained. The data and solution sets are

$$\mathcal{Z}_p = \{A \in \mathcal{R}^{m \times n}, c \in \mathcal{R}^n\}$$

and

$$\mathcal{S}_p = \{y \in \mathcal{R}^m : \|A^T y - c\|^2 \le \|A^T x - c\|^2 \quad \text{for every} \quad x \in \mathcal{R}^m\}.$$

Obviously, given a y, to see if $y \in \mathcal{S}_p$ is as the same as the original problem. However, from a projection theorem in linear algebra, the solution set can be characterized as

$$\mathcal{S}_p = \{y \in \mathcal{R}^m : AA^T y = Ac\},$$

which becomes a system of linear equations. The vector $s = A^T y = A^T(AA^T)^+ Ac$ is the projection of c onto the range of A^T, where AA^T is called normal matrix and $(AA^T)^+$ is called pseudo-inverse.

The vector $c - A^T y = (I - A^T(AA^T)^+ A)c$ is the projection of c onto the null space of A. It is the solution of the following least-squares problem:

$$\begin{array}{lll} (LS) & \text{minimize} & \|x - c\|^2 \\ & \text{subject to} & x \in \mathcal{N}(A). \end{array}$$

1.3.3 System of nonlinear equations

Given $f(x) : \mathcal{R}^n \to \mathcal{R}^m$, the problem is to solve m equations for n unknowns:

$$f(x) = 0.$$

The "data" and solution sets are

$$\mathcal{Z}_p = \{f\} \quad \text{and} \quad \mathcal{S}_p = \{x \in \mathcal{R}^n : f(x) = 0\}.$$

Here we call $\mathcal{Z}_p$ an oracle. For any input x, it returns the value and other numerical information of the function. Again, given an x one can easily check to see if x is in $\mathcal{S}_p$.

1.3.4 System of linear inequalities

Given $A \in \mathcal{R}^{m \times n}$ and $b \in \mathcal{R}^m$, the problem is to find a solution $x \in \mathcal{R}^n$ satisfying $Ax \le b$ or prove that the solution set is empty. The inequality problem may possess other form, for example, find x such that $Ax = b$ and

$x \geq 0$, which is a combination of linear equations and inequalities. The data and solution sets are

$$\mathcal{Z}_p = \{A \in \mathcal{R}^{m \times n}, b \in \mathcal{R}^m\} \quad \text{and} \quad \mathcal{S}_p = \{x \in \mathcal{R}^n : Ax = b,\ x \geq 0\}.$$

Traditionally, a point in $\mathcal{S}_p$ is called a feasible solution, and a strictly positive point in $\mathcal{S}_p$ is called a strictly or interior feasible solution.

Theorem 1.9 *(Farkas' lemma) Let $A \in \mathcal{R}^{m \times n}$ and $b \in \mathcal{R}^m$. Then, the system $\{x:\ Ax = b,\ x \geq 0\}$ has a solution x if and only if that $A^T y \leq 0$ implies $b^T y \leq 0$.*

A vector y, with $A^T y \leq 0$ and $b^T y > 0$, is called a (primal) infeasibility certificate for the system $\{x:\ Ax = b,\ x \geq 0\}$.

Theorem 1.10 *(Farkas' lemma variant) Let $A \in \mathcal{R}^{m \times n}$ and $c \in \mathcal{R}^n$. Then, the system $\{y:\ A^T y \leq c\}$ has a solution y if and only if that $Ax = 0$ and $x \geq 0$ imply $c^T x \geq 0$.*

Again, a vector $x \geq 0$, with $Ax = 0$ and $c^T x < 0$, is called a (dual) infeasibility certificate for the system $\{y:\ A^T y \leq c\}$.

Theorem 1.11 *(Approximate Farkas' lemma) Let $A \in \mathcal{R}^{m \times n}$, $b \in \mathcal{R}^m$, and $c \in \mathcal{R}^n$. Let*

$$\alpha_x := \min\{\|x\|:\ Ax = b,\ x \geq 0\},$$

$$\alpha_y := \min\{\|y\|:\ A^T y \leq c\},$$

$$\beta_u := \min\{\|u\|^*:\ A^T y \leq u,\ b^T y = 1\}$$

and

$$\beta_v := \min\{\|v\|^*:\ Ax = v,\ c^T x = 1,\ x \geq 0\}.$$

Then

$$\alpha_x \beta_u = \alpha_y \beta_v = 1.$$

Here, $\|.\|$ is an arbitrary norm and $\|.\|^*$ is the corresponding dual norm.

1.3.5 System of nonlinear inequalities

Given $f(x) : \mathcal{R}^n \rightarrow \mathcal{R}^m$, the problem is to find a solution $x \in \mathcal{R}^n$ satisfying $f(x) \leq 0$ or prove the solution set is empty. The oracle and solution sets are

$$\mathcal{Z}_p = \{f\} \quad \text{and} \quad \mathcal{S}_p = \{x \in \mathcal{R}^n : f(x) \leq 0\}.$$

Given an x, one can easily check to see if x is in $\mathcal{S}_p$.

$f(x) : \mathcal{R}^n \rightarrow \mathcal{R}^n$ is a monotone function over a open convex set Ω if and only if

$$(y-x)^T(f(y)-f(x)) \geq 0$$

for all $x,\ y \in \Omega$. Consider the following set

$$\Omega = \{x:\ f(x) \geq 0,\ x \geq 0\},$$

where $f(x)$ is continuous and monotone in $\mathcal{R}^n_+$. We have the following theorem.

Theorem 1.12 *Ω has a feasible solution if and only if for any sequence $(x^k \geq 0, \tau^k > 0)$ such that*

$$\lim\ \tau^k f(x^k/\tau^k) \geq 0,$$

then we have

$$\lim\ (x^k)^T f(x^k/\tau^k) \geq 0.$$

1.3.6 Linear programming

Given $A \in \mathcal{R}^{m\times n}$, $b \in \mathcal{R}^m$ and $c \in \mathcal{R}^n$, the linear programming (LP) problem is the following optimization problem.

$$\begin{array}{lll} (LP) & \text{minimize} & c^Tx \\ & \text{subject to} & Ax = b,\ x \geq 0. \end{array}$$

The linear function c^Tx is called the objective function, and x is called the decision variables. In this problem, $Ax = b$ and $x \geq 0$ enforce constraints on the selection of x. The set $\mathcal{F}_p = \{x : Ax = b, x \geq 0\}$ is called feasible set or feasible region. A point $x \in \mathcal{F}_p$ is called a feasible point, and a feasible point x^* is called an optimal solution if $c^Tx^* \leq c^Tx$ for all feasible points x. If there is a sequence $\{x^k\}$ such that x^k is feasible and $c^Tx^k \rightarrow -\infty$, then (LP) is said unbounded.

The data and solution sets for (LP), respectively, are

$$\mathcal{Z}_p = \{A \in \mathcal{R}^{m\times n}, b \in \mathcal{R}^m, c \in \mathcal{R}^n\}$$

and

$$\mathcal{S}_p = \{x \in \mathcal{F}_p : c^Tx \leq c^Ty \quad \text{for every} \quad y \in \mathcal{F}_p\}.$$

Again, given an x, to see if $x \in \mathcal{S}_p$ is as difficult as the original problem. However, due to the duality theorem, we can simplify the representation of the solution set significantly.

With every linear program, another linear program, called the dual (LD), is the following problem.

$$\begin{array}{lll} (LD) & \text{maximize} & b^T y \\ & \text{subject to} & A^T y + s = c, \; s \geq 0, \end{array}$$

where $y \in \mathcal{R}^m$ and $s \in \mathcal{R}^n$. The components of s are called dual slacks. Denote by $\mathcal{F}_d$ the sets of all (y, s) that are feasible for the dual.

Theorem 1.13 *(Weak duality theorem) Let $\mathcal{F}_p$ and $\mathcal{F}_d$ be non-empty. Then,*

$$c^T x \geq b^T y \quad \textit{where} \quad x \in \mathcal{F}_p, \; (y, s) \in \mathcal{F}_d.$$

Corollary 1.14 *(Strong duality theorem) Let $\mathcal{F}_p$ and $\mathcal{F}_d$ be non-empty. Then, x^* is optimal for (LP) if and only if the following conditions hold:*

i) $x^* \in \mathcal{F}_p$;

ii) *there is* $(y^*, s^*) \in \mathcal{F}_d$;

iii) $c^T x^* = b^T y^*$ *or* $(x^*)^T s^* = 0$.

From the optimality condition, the solution set for (LP) and (LD) is

$$\mathcal{S}_p = \{x \in \mathcal{F}_p, \; (y, s) \in \mathcal{F}_d : c^T x = b^T y\},$$

which is a system of linear inequalities and equations.

Theorem 1.15 *(LP duality theorem) If (LP) and (LD) both have feasible solutions then both problems have optimal solutions and the optimal objective values of the objective functions are equal.*

If one of (LP) or (LD) has no feasible solution, then the other is either unbounded or has no feasible solution. If one of (LP) or (LD) is unbounded then the other has no feasible solution.

For feasible x and (y, s), $c^T x - b^T y$ is usually called the duality gap and $x^T s$ the complementarity gap. If $x^T s = 0$, then we say x and s are complementary to each other.

Theorem 1.16 *(Strict complementarity theorem) If (LP) and (LD) both have feasible solutions then both problems have a pair of strictly complementary solutions x^* and s^* such that*

$$x^* + s^* > 0.$$

Moreover, the supports

$$P^* = \{j : \; x_j^* > 0\} \quad \textit{and} \quad Z^* = \{j : \; s_j^* > 0\}$$

are invariant for all pairs of strictly complementary solutions.

Given (LP) or (LD), the pair of P^* and Z^* is called the (strict) complementarity partition. $\{x : A_{P^*}x_{P^*} = b,\ x_{P^*} \geq 0,\ x_{Z^*} = 0\}$ is called the primal optimal face, and $\{y : c_{Z^*} - A^T_{Z^*}y \geq 0,\ c_{P^*} - A^T_{P^*}y = 0\}$ is called the dual optimal face.

Select m linearly independent columns, denoted by the index set B, from A. Then matrix A_B is nonsingular and we may uniquely solve

$$A_B x_b = b$$

for the m-vector x_B. By setting the variables of x corresponding to the remaining columns of A equal to zero, we obtain a solution x such that

$$Ax = b.$$

Then, x is said to be a (primal) basic solution to (LP) with respect to the basis A_B. The components of x_B are called basic variables. A dual vector y satisfying

$$A^T_B y = c_B$$

is said to be the corresponding dual basic solution. If a basic solution $x \geq 0$, then x is called a basic feasible solution. If the dual solution is also feasible, that is

$$s = c - A^T y \geq 0,$$

then x is called an optimal basic solution and A_B an optimal basis. A basic feasible solution is a vertex on the boundary of the feasible region. An optimal basic solution is an optimal vertex of the feasible region.

If one or more components in x_B has value zero, that basic solution x is said to be (primal) degenerate. Note that in a nondegenerate basic solution the basic variables and the basis can be immediately identified from the nonzero components of the basic solution. If all components in the corresponding dual slack vector s, except for s_B, are non-zero, then y is said to be (dual) nondegenerate. If both primal and dual basic solutions are nondegenerate, A_B is called a nondegenerate basis.

Theorem 1.17 *(LP fundamental theorem) Given (LP) and (LD) where A has full row rank m,*

i) *if there is a feasible solution, there is a basic feasible solution;*

ii) *if there is an optimal solution, there is an optimal basic solution.*

1.3.7 Quadratic programming

Given $Q \in \mathcal{R}^{n\times n}$, $A \in \mathcal{R}^{m\times n}$, $b \in \mathcal{R}^m$ and $c \in \mathcal{R}^n$, the quadratic programming (QP) problem is the following optimization problem.

$$\begin{array}{lll} (QP) & \text{minimize} & q(x) = (1/2)x^T Qx + c^T x \\ & \text{subject to} & Ax = b,\ x \geq 0, \end{array}$$

We may denote the feasible set by $\mathcal{F}_p$. The data and solution sets for (QP) are

$$\mathcal{Z}_p = \{Q \in \mathcal{R}^{n\times n}, A \in \mathcal{R}^{m\times n}, b \in \mathcal{R}^m, c \in \mathcal{R}^n\}$$

and

$$\mathcal{S}_p = \{x \in \mathcal{F}_p : q(x) \leq q(y) \quad \text{for every} \quad y \in \mathcal{F}_p\}.$$

A feasible point x^* is called a KKT point, where KKT stands for Karush-Kuhn-Tucker, if the following KKT conditions hold: there exists $(y^* \in \mathcal{R}^m, s^* \in \mathcal{R}^n)$ such that (x^*, y^*, s^*) is feasible for the following dual problem

$$\begin{array}{lll} (QD) & \text{maximize} & b^T y - (1/2)x^T Qx \\ & \text{subject to} & A^T y + s = c + Qx,\ x,\ s \geq 0, \end{array}$$

and

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

If Q is positive semi-definite, then x^* is an optimal solution for (QP) if and only if x^* is a KKT point for (QP). In this case, the solution set for (QP) is characterized by a system of linear inequalities and equations. One can see (LP) is a special case of (QP).

1.3.8 Complementarity problem

Given $M \in \mathcal{R}^{n\times n}$ and $q \in \mathcal{R}^n$, the linear complementarity problem (LCP) is to find a pair $x, s \in \mathcal{R}^n$ such that

$$s = Mx + q, \quad (x, s) \geq 0 \quad \text{and} \quad x_j s_j = 0, \quad j = 1, 2, ..., n.$$

A pair $(x, s) \geq 0$ satisfying $s = Mx + q$ is called a feasible pair. The data and solution sets for LCP are

$$\mathcal{Z}_p = \{M \in \mathcal{R}^{n\times n}, q \in \mathcal{R}^n\}$$

$$\mathcal{S}_p = \{(x, s) : s = Mx + q, \quad (x, s) \geq 0 \quad \text{and} \quad x^T s = 0\}.$$

One can verify that the solution set of (LP) and the KKT set of (QP) can be formulated as the solution set of an LCP. Hence, the LCP is a fundamental decision and optimization problem. The LCP also arises from economic equilibrium problems, noncooperative games, traffic assignment problems, and optimization problems. If M is positive semi-definite, i.e.,

$$x^T M x \geq 0, \quad \text{for all} \quad x,$$

then, the LCP is called a monotone LCP.

Let $f(x) : \mathcal{R}^n_+ \to \mathcal{R}^n$ be a real function. Then, the nonlinear complementarity problem (NCP) is to find a pair $x, s \in \mathcal{R}^n$ such that

$$s = f(x), \quad (x, s) \geq 0 \quad \text{and} \quad x_j s_j = 0, \quad j = 1, 2, ..., n.$$

If $f(x)$ is a monotone function over $\mathcal{R}^n_+$, meaning

$$(x^1 - x^2)^T (f(x^1) - f(x^2)) \geq 0 \quad \text{for all} \quad x^1, x^2 \in \mathcal{R}^n_+,$$

then the problem is called the monotone complementarity problem.

Theorem 1.18 *(Monotone complementarity theorem) If a monotone complementarity problem has a complementary solution, then it has a maximal complementary solution pair x^* and s^* meaning that the number of the positive components in vector $x^* + s^*$ is maximized. Moreover, the supports*

$$P^* = \{j : \; x^*_j > 0\} \quad \textit{and} \quad Z^* = \{j : \; s^*_j > 0\}$$

are invariant for all pairs of maximal complementary solutions.

1.3.9 Positive semi-definite programming

Given $C \in \mathcal{M}^n$, $A_i \in \mathcal{M}^n$, $i = 1, 2, ..., m$, and $b \in \mathcal{R}^m$, the positive semi-definite programming problem is to find a matrix $X \in \mathcal{M}^n$ for the optimization problem:

$$\begin{array}{lll} (PSP) & \inf & C \bullet X \\ & \text{subject to} & A_i \bullet X = b_i, i = 1, 2, ..., m, \; X \succeq 0, \end{array}$$

where the $\bullet$ operation is the generalization of the inner product to matrices

$$A \bullet B := \sum_{i,j} A_{i,j} B_{i,j} = \text{tr} A^T B,$$

and the notation $X \succeq 0$ means that X is positive semi-definite matrix. This problem has many applications in combinatorial optimization and control engineering.

The dual problem can be written as

$$\begin{array}{lll} (PSD) & \sup & b^T y \\ & \text{subject to} & \sum_i^m y_i A_i + S = C, \; S \succeq 0, \end{array}$$

which is analogous to the dual of LP. We also have analogous theorems.

Theorem 1.19 *(Farkas' lemma in PSP) Let $A_i \in \mathcal{M}^n$ for $i = 1, ..., m$ and $b \in \mathcal{R}^m$. Then, there exists a symmetric matrix $X \succ 0$ with*

$$A_i \bullet X = b_i, \quad i = 1, ..., m,$$

if and only if $\sum_i^m y_i A_i \preceq 0$ and $\sum_i^m y_i A_i \neq 0$ implies $b^T y < 0$.

Theorem 1.20 *(Weak duality theorem in PSP) Let $\mathcal{F}_p$ and $\mathcal{F}_d$, the feasible sets for the primal and dual, be non-empty. Then,*

$$C \bullet X \geq b^T y \quad \textit{where} \quad X \in \mathcal{F}_p, \; (y, S) \in \mathcal{F}_d.$$

Corollary 1.21 *(Strong duality theorem in PSP) Let $\mathcal{F}_p$ and $\mathcal{F}_d$ be non-empty and have an interior. Then, X is optimal for (PS) if and only if the following conditions hold:*

i) *$X \in \mathcal{F}_p$;*

ii) *there is $(y, S) \in \mathcal{F}_d$;*

iii) *$C \bullet X = b^T y$ or $X \bullet S = 0$.*

Two positive semi-definite matrices are complementary to each other if $X \bullet S = 0$. From the optimality conditions, the solution pair set for (PSP) and (PSD) is

$$\mathcal{S}_p = \{X \in \mathcal{F}_p, \; (y, S) \in \mathcal{F}_d : C \bullet X = b^T y\},$$

which is a system of linear matrix inequalities and equations.

Theorem 1.22 *(PSP duality theorem) If one of (PSP) or (PSD) has a strictly or interior feasible solution and its optimal value is finite, then the other is feasible and has the same optimal value. If one of (PSP) or (PSD) is unbounded then the other has no feasible solution.*

1.3.10 Nonlinear programming

Given $f : \mathcal{R}^n \rightarrow \mathcal{R}$, $h : \mathcal{R}^n \rightarrow \mathcal{R}^m$, $g : \mathcal{R}^n \rightarrow \mathcal{R}^d$, the nonlinear programming problem is the following optimization problem.

$$\begin{array}{lll} (NP) & \text{minimize} & f(x) \\ & \text{subject to} & h(x) = b, \ g(x) \geq 0. \end{array}$$

We may denote the feasible set by $\mathcal{F}$. The oracle and solution sets for (NP) are

$$\mathcal{Z}_p = \{f, h, g\} \quad \text{and} \quad \mathcal{S}_p = \{x \in \mathcal{F} : f(x) \leq f(y) \quad \text{for every} \quad y \in \mathcal{F}\}.$$

Note that the data set is a functional set.

A feasible point x^* is called a KKT point if the following KKT conditions hold: there exists $(y^* \in \mathcal{R}^m, s^* \in \mathcal{R}^d)$ such that (x^*, y^*, s^*) is satisfying

$$\nabla^T f(x^*) - \nabla h(x^*)y^* - \nabla g(x^*)s^* = 0, \ s^* \geq 0,$$

and

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

Theorem 1.23 *(Karush-Kuhn-Tucker theorem) Let x^* be a relative (local) minimum solution for (NP) and suppose x^* is a regular point for the equality and active inequality constraints, i.e., the Jacobian matrix of these constraints has full row rank. Then, x^* is an KKT point.*

If f is convex, h is affine, and g is convex, then x^* is optimal if and only if x^* is a KKT point for (NP).

1.4 Algorithms and Computation Models

An algorithm is a list of instructions to solve a problem. For every instance of problem $\mathcal{P}$, i.e., for every given data $Z \in \mathcal{Z}_p$, an algorithm for solving $\mathcal{P}$ either determines that $\mathcal{S}_p$ is empty or generates an output x such that $x \in \mathcal{S}_p$ or x is close to $\mathcal{S}_p$ in certain measure. The latter x is called an (approximate) solution.

Let us use $\mathcal{A}_p$ to denote the collection of all possible algorithm for solving every instance in $\mathcal{P}$. Then, the (operation) complexity of an algorithm $A \in \mathcal{A}_p$ for solving an instance $Z \in \mathcal{Z}_p$ is defined as the total arithmetic operations: $+$, $-$, $*$, $/$, and comparison on real numbers. Denote it by $c_o(A, Z)$. Sometimes it is convenient to define the iteration complexity, denoted by $c_i(A, Z)$, where we assume that each iteration costs a polynomial number (in m and n) of arithmetic operations. In most iterative algorithms, each iteration can be performed efficiently both sequentially and parallelly, such as solving a system of linear equations, rank-one updating the inversion of a matrix, pivoting operation of a matrix, multiplying a matrix by a vector, etc.

1.4.1 Worst-case complexity

The worst-case complexity of algorithm A for problem $\mathcal{P}$ is defined as

$$c(A) := \sup_{Z \in \mathcal{Z}_p} c(A, Z).$$

It is better to distinguish the worst-case complexity of an algorithm, A, from that of a problem $\mathcal{P}$. The worst-case complexity of the problem is

$$c^p := \inf_{A \in \mathcal{A}_p} c(A).$$

Analyzing the worst-case complexity of a problem is challenging since $\mathcal{A}_p$ is an unknown domain, and the analysis of the complexity of the algorithm is equally difficult since $\mathcal{P}$ is also immense. However, the complexity theory does not directly attack the algorithm complexity for every instance. Instead, it classifies $\mathcal{P}$ using its data bit-size L, where the data are assumed rational. This is the Turing Machine Model for computation. We may call this type of complexity size-based. Then, we express an upper bound $f_A(m,n,L)$, in terms of these parameters m, n, and L, for the size-based complexity of algorithm A as

$$c(A, L) := \sup_{Z \in \mathcal{Z}_p,\ \mathrm{size}(Z) \le L} c(A, Z) \le f_A(m, n, L).$$

Then, the size-based complexity of problem $\mathcal{P}$ has a relation

$$c^p(L) := \inf_{A \in \mathcal{A}_p} c(A, L) \le f_A(m, n, L).$$

We see that the complexity of algorithms is an upper bound for the complexity of the problem. Another active pursuit in computer science is the analysis of a lower bound for the problem's complexity, which is outside of the scope of this monograph.

If $f_A(m,n,L)$ is a polynomial in m, n, and L, then we say algorithm A is a polynomial-time or polynomial algorithm and problem $\mathcal{P}$ is polynomially solvable. If $f_A(m,n,L)$ is independent of L and polynomial in m and n, then we say algorithm A is a strongly polynomial algorithm.

In the real number model, the use of L is not suitable. We may use ϵ, the error for an approximate solution as a parameter. Let $c(A, Z, \epsilon)$ be the total number of operations of algorithm A for generating an ϵ-approximate solution, with a well-defined measure, to problem $\mathcal{P}$. Then,

$$c(A, \epsilon) := \sup_{Z \in \mathcal{Z}_p} c(A, Z, \epsilon) \leq f_A(m, n, \epsilon) \quad \text{for any} \quad \epsilon > 0.$$

We call this complexity model error-based.

If $f_A(m, n, \epsilon)$ is a polynomial in m, n, and $\log(1/\epsilon)$, then algorithm A is a polynomial algorithm and problem $\mathcal{P}$ is polynomially solvable. Again, if $f_A(m, n, \epsilon)$ is independent of ϵ and polynomial in m and n, then we say algorithm A is a strongly polynomial algorithm. If $f_A(m, n, \epsilon)$ is a polynomial in m, n, and $(1/\epsilon)$, then algorithm A is a polynomial approximation scheme or pseudo-polynomial algorithm . For some optimization problems, the complexity theory can be applied to prove not only that they cannot be solved in polynomial-time, but also that they do not have polynomial approximation schemes. In practice, approximation algorithms are widely used and accepted in practice.

We have to admit that the criterion of polynomiality is somewhat controversial. Many algorithms may not be polynomial but work fine in practice. This is because complexity theory is built upon the worst-case analysis. However, this criterion generally provides a qualitative statement: if a problem is polynomial solvable, then the problem is indeed relatively easy to solve regardless of the algorithm used. Furthermore, it is ideal to develop an algorithm with both polynomiality and practical efficiency.

1.4.2 Condition-based complexity

As we discussed before, in the Turing Machine Model the parameters are selected as the number of variables, the number of constraints, and the bit-size of the data of an instance. In fact, expressing the algorithm complexity in terms of the size of the problem does not really measure the difficulty of an instance of the problem. Two instances with the same size may result in drastically different performances by the same algorithm. Consider the steepest descent method for solving

$$\text{minimize} \quad (1/2)x^T Q x + c^T x$$

where $Q \in \mathcal{R}^{n \times n}$ is positive definite. It is well known that the algorithm generates a sequence of $\{x^k\}$ such that

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq \frac{e_n - e_1}{e_n + e_1},$$

where $e_1 \leq e_2, ... \leq e_n$ are n eigenvalues of Q with ascending order. Thus, two matrices with the same size but different eigenvalue-structure will possess quite different convergence speed. This phenomena is surprisingly common in optimization, due to mathematical bases upon which algorithms are designed. Thus, the upper bound for the complexity of an algorithm may be expressed as $f_A(m, n, \eta(Z))$ or $f_A(m, n, \epsilon, \eta(Z))$ in both the rational-number and the real number models, where $\eta(Z)$ can be viewed as a condition number for the instance Z. The better the condition number, the less difficult the instance. It is our goal to study this phenomena and to improve the condition number and, thereby, the performance of an algorithm.

1.4.3 Average complexity

Let $\mathcal{Z}_p$ be a random sample space, then one can define the average or expected complexity of the algorithm for the problem as

$$c^a(A) = E_{Z \in \mathcal{Z}_p}(c(A, Z)).$$

If we know the condition-based complexity of an algorithm for $\mathcal{P}$, then the average complexity of the algorithm is

$$c^a(A) \leq E_{Z \in \mathcal{Z}}(f_A(m, n, \eta(Z))).$$

In many cases, $f_A(m, n, \eta(Z))$ can be expressed as

$$f_A(m, n, \eta(Z)) = f^1(m, n) f^2(\eta(Z)).$$

Thus,

$$c^a(A) \leq f^1(m, n) E_{Z \in \mathcal{Z}}(f^2(\eta(Z))),$$

which will simplify analysis a great deal.

Another probabilistic model is called high probability analysis. We say that a problem $\mathcal{P}$ can be solved by algorithm A in $f_A(m, n)$ time with high probability if

$$\Pr\{c(A, Z) \leq f_A(m, n)\} \to 1$$

as $m, n \to \infty$. Again, if we have a condition-based complexity and if we have

$$\Pr\{f^2(\eta(Z)) \leq f^3(m, n)\} \to 1$$

as $m, n \to \infty$, then the algorithm solves P in

$$f_A(m, n) = f^1(m, n) f^3(m, n)$$

operations with high probability.

1.4.4 Asymptotic complexity

Most algorithms are iterative in nature. They generate a sequence of ever-improving points $x^0, x^1, ..., x^k, ...$ approaching the solution set. For many optimization problems and/or algorithms, the sequence will never exactly reach the solution set. One theory of iterative algorithms, referred to as local or asymptotic convergence analysis, is concerned with the rate at which the optimality error of the generated sequence converges to zero.

Obviously, if each iteration of competing algorithms requires the same amount of work, the speed of the convergence of the error reflects the speed of the algorithm. This convergence rate, although it holds locally or asymptotically, provides evaluation and comparison of different algorithms. It has been widely used by the nonlinear optimization and numerical analysis community as an efficiency criterion. In many cases, this criterion does explain practical behavior of iterative algorithms.

Consider a sequence of real numbers $\{r^k\}$ converging to zero. One can define several notions related to the speed of convergence of such a sequence.

Definition 1.1 . *Let the sequence $\{r^k\}$ converge to zero. The order of convergence of $\{r^k\}$ is defined as the supermum of the nonnegative numbers p satisfying*

$$0 \leq \limsup_{k\to\infty} \frac{|r^{k+1}|}{|r^k|^p} < \infty.$$

Definition 1.2 . *Let the sequence $\{r^k\}$ converge to zero such that*

$$\limsup_{k\to\infty} \frac{|r^{k+1}|}{|r^k|^2} < \infty.$$

Then, the sequence is said to converge quadratically to zero.

It should be noted that the order of convergence is determined only by the properties of the sequence that holds as $k \to \infty$. In this sense we might say that the order of convergence is a measure of how good the tail of $\{r^k\}$ is. Large values of p imply the faster convergence of the tail.

Definition 1.3 . *Let the sequence $\{r^k\}$ converge to zero such that*

$$\limsup_{k\to\infty} \frac{|r^{k+1}|}{|r^k|} = \beta < 1.$$

Then, the sequence is said to converge linearly to zero with convergence ratio β.

Linear convergence is the most important type of convergence behavior. A linearly convergence sequence, with convergence ratio β, can be said to have a tail that converges to zero at least as fast as the geometric sequence $c\beta^k$ for a fixed number c. Thus, we also call linear convergence geometric convergence.

As a rule, when comparing the relative effectiveness of two competing algorithms both of which produce linearly convergent sequences, the comparison is based on their corresponding convergence ratio—the smaller the ratio, the faster the algorithm. The ultimate case where $\beta = 0$ is referred to as superlinear convergence.

1.5 Basic Numerical Procedures

There are several basic numerical problems frequently solved by interior-point algorithms.

1.5.1 Gaussian elimination method

Probably the best-known algorithm for solving a system of linear equations is the Gaussian elimination method. Suppose we want to solve

$$Ax = b.$$

We may assume $a_{11} \neq 0$, where a_{ij} is the component of A in row i and column j. Then we can subtract appropriate multiples of the first equation from the other equations so as to have an equivalent system:

$$\begin{pmatrix} a_{11} & A_{1.} \\ 0 & A' \end{pmatrix} \begin{pmatrix} x_1 \\ x' \end{pmatrix} = \begin{pmatrix} b_1 \\ b' \end{pmatrix}.$$

This is a pivot step, where a_{11} is called pivot. Now, recursively, we solve the system of the last $m-1$ equations. Substituting the solution x' found into the first equation yields a value for x_1.

In matrix form, the Gaussian elimination method transforms A into the form

$$\begin{pmatrix} U & C \\ 0 & 0 \end{pmatrix}$$

where U is a nonsingular, upper-triangular matrix,

$$A = L \begin{pmatrix} U & C \\ 0 & 0 \end{pmatrix},$$

and L is a nonsingular, lower-triangular matrix. This is called the LU-decomposition.

Sometimes, the matrix is transformed further to a form

$$\begin{pmatrix} D & C \\ 0 & 0 \end{pmatrix}$$

where D is a nonsingular, diagonal matrix. This whole procedure uses about nm^2 arithmetic operations. Thus, it is a strong polynomial-time algorithm.

1.5.2 Choleski decomposition method

Another useful method is to solve the least squares problem:

$$(LS) \quad \text{minimize} \quad \|A^T y - c\|.$$

The theory says that y^* minimizes $\|A^T y - c\|$ if and only if

$$AA^T y^* = Ac.$$

So the problem is reduced to solving a system of linear equations with a symmetric semi-positive definite matrix. One method is Choleski's decomposition. In matrix form, the method transforms AA^T into the form

$$AA^T = U^T D U,$$

where U is an upper-triangular matrix and D is a diagonal matrix.

1.5.3 The Newton method

The Newton method is used to solve a system of nonlinear equations: given $f(x) : \mathcal{R}^n \to \mathcal{R}^n$, the problem is to solve n equations for n unknowns such that

$$f(x) = 0.$$

The Newton method is defined by the following iterative formula:

$$x^{k+1} = x^k - \alpha(\nabla f(x^k))^{-1} f(x^k),$$

where scalar $\alpha \geq 0$ is called step-size.

A modified or quasi Newton method is defined by

$$x^{k+1} = x^k - \alpha M^k f(x^k),$$

where M^k is an $n \times n$ positive definite matrix. In particular, if $M^k = I$, the method is called the steepest descent method.

1.5.4 Solve ball-constrained linear problem

The ball constrained linear problem has the following form

$$
\begin{array}{lll}
(BP) & \text{minimize} & c^T x \\
 & \text{subject to} & Ax = 0, \ \|x\|^2 \le 1,
\end{array}
$$

or

$$
\begin{array}{lll}
(BD) & \text{minimize} & b^T y \\
 & \text{subject to} & \|A^T y\|^2 \le 1.
\end{array}
$$

x^* minimizes (BP) if and only if there is y such that they satisfy

$$AA^T y = Ac,$$

and if $c - A^T y \neq 0$ then

$$x^* = -(c - A^T y)/\|c - A^T y\|;$$

otherwise any feasible x is a solution. The solution y^* for (BD) is given as follows: solve

$$AA^T \bar{y} = b,$$

and if $\bar{y} \neq 0$ then set

$$y^* = -\bar{y}/\|A^T \bar{y}\|;$$

otherwise any feasible y is a solution. So these two problems can be reduced to solving a system of linear equations.

1.5.5 Solve ball-constrained quadratic problem

The ball constrained quadratic problem has the following form

$$
\begin{array}{lll}
(BP) & \text{minimize} & (1/2)x^T Q x + c^T x \\
 & \text{subject to} & Ax = 0, \ \|x\|^2 \le 1,
\end{array}
$$

or simply

$$
\begin{array}{lll}
(BD) & \text{minimize} & (1/2)y^T Q y + b^T y \\
 & \text{subject to} & \|y\|^2 \le 1.
\end{array}
$$

This problem is a classical trust region problem used in nonlinear optimization. The optimality conditions for the minimizer y^* of (BD) are

$$(Q + \mu^* I)y^* = -b, \quad \mu^* \ge 0, \quad \|y^*\|^2 \le 1, \quad \mu^* \|y^*\|^2 = 0,$$

and

$$(Q + \mu^* I) \succeq 0.$$

These conditions are necessary and sufficient. This problem can be solved in polynomial time $\log(1/\epsilon)$ and $\log(\log(1/\epsilon))$ by the bisection method or a hybrid of the bisection and Newton methods, respectively. In practice, several trust region procedures have been very effective in solving this problem.

1.6 Notes

The term "complexity" was introduced by Hartmanis and Stearns [143]. Also see Garey and Johnson [101] and Papadimitriou and Steiglitz [272]. The NP theory was due to Cook [69] and Karp [174]. The importance of P was observed by Edmonds [84].

Linear programming and the simplex method were introduced by Dantzig [73]. Other inequality problems and convexity theories can be seen in Gritzmann and Klee [132], Grötschel, Lovász and Schrijver [133], Grünbaum [134], Rockafellar [291], and Schrijver [298]. Various complementarity problems can be found found in Cottle, Pang and Stone [72]. The positive semi-definite programming, an optimization problem in nonpolyhedral cones, and its applications can be seen in Nesterov and Nemirovskii [263], Alizadeh [8], and Boyd, Ghaoui, Feron and Balakrishnan [62]. Recently, Goemans and Williamson [108] obtained several breakthrough results on approximation algorithms using positive semi-definite programming. The KKT condition for nonlinear programming was given by Karush, Kuhn and Tucker [193].

It was shown by Klee and Minty [178] that the simplex method is not a polynomial-time algorithm. The ellipsoid method, the first polynomial-time algorithm for linear programming with rational data, was proven by Khachiyan [176], also see Bland, Goldfarb and Todd [53]. The method was originally proposed by Shor [303], and discussed in Nemirovskii and Yudin [258]. The interior-point method, another polynomial-time algorithm for linear programming, was developed by Karmarkar. It is related to the classical barrier-function method studied by Frisch [99] and Fiacco and McCormick [92], see Gill, Murray, Saunders, Tomlin and Wright [106], and Anstreicher [18]. For a brief LP history, see the excellent article by Wright [366].

The real computation model was developed by Blum, Shub and Smale [56] and Nemirovskii and Yudin [258]. The average setting can be seen in Traub, Wasilkowski and Wozniakowski [335]. The asymptotic convergence rate and ratio can be seen in Luenberger [198], Ortega and Rheinboldt [269], and Traub [334]. Other complexity issues in numerical optimization were discussed in Vavasis [359].

Many basic numerical procedures listed in this chapter can be found in Golub and Van Loan [119]. The ball-constrained quadratic problem and its solution methods can be seen in Moré [248], Sorenson [309], and Dennis and Schnable [76]. The complexity result of the ball-constrained quadratic problem was proved by Vavasis [359] and Ye [376, 380].

1.7 Exercises

1.1 *Prove Theorem 1.3.*

1.2 *Function f being a quasi-convex implies that the level set of f is convex.*

1.3 *Prove Propositions 1.5 and 1.6.*

1.4 *Prove Harmonic inequality described in Section 1.2.4.*

1.5 *Prove Theorem 1.8.*

1.6 *Prove Theorems 1.9 and 1.10.*

1.7 *Prove Theorem 1.11.*

1.8 *Prove Theorem 1.12.*

1.9 *Prove Theorem 1.17.*

1.10 *If (LP) and (LD) have a nondegenerate optimal basis A_B, then the strict complementarity partition in Theorem 1.16*

$$P^* = B.$$

1.11 *If Q is positive semi-definite, then x^* is an optimal solution for (QP) if and only if x^* is a KKT point for (QP).*

1.12 *Prove Theorem 1.18.*

1.13 *Prove Theorem 1.19.*

1.14 *The optimality conditions for the minimizer y^* of (BD) in Section 1.5.5:*

$$(Q + \mu^* I)y^* = -b, \quad \mu^* \geq 0, \quad \|y^*\| \leq 1, \quad \mu^*\|y^*\| = 0,$$

and

$$(Q + \mu^* I) \succeq 0,$$

are necessary and sufficient.

Chapter 2

Geometry of Convex Inequalities

Most optimization algorithms are iterative in nature, that is, they generate a sequence of improved points. Algorithm design is closely related to how the improvement is measured. Most optimization algorithms use a merit or descent function to measure the progress. Some merit or descent functions are based on the objective function. For example, if we know a lower bound z of the optimal objective value, $f(x) - z$ is a measure of how far x is from the solution set. Another example measures the residual or error of the optimality conditions represented by a system of equations and inequalities involving the derivatives of the objective and constraint functions as discussed in the preceding chapter.

One particular merit or descent function measures the "size" of the containing set—a set that contains a solution. A typical example is the bisection method for finding a root of a continuous function within an interval. The method measures the length of the containing interval. In each step, the middle point of the containing interval is tested, and, subsequently, a new containing interval is selected and its length is a half of the previous one. Thus, these containing intervals shrink at a constant rate $1/2$.

A generic center-section algorithm for multiple-variable problems can be described as follows. Given x^k, a "good" interior point in a containing set, we check to see if x^k is desirable. If not, we generate a separating hyperplane and place it through x^k and cut the containing set into two parts. The separating hyperplane can be generated from an oracle. If we can assure that the solution set lies in one of the two parts, then the other part can be deleted. This leads to a new containing set that is smaller than the previous one. A new "good" interior point in the new containing set can be tested and the process continues. Obviously, this method can be applied to any convex problems.

The question that arises is how to select the test point and where to place the cut. Ideally, we would like to select a "center" point that divides the containing set into two approximately "equal" parts with respect to certain measures of the containing set. Then, we will have a shrinking rate of about $1/2$ for the sequence of the containing sets.

Various centers were considered as test points. In this chapter, we review these centers and their associated measures. We show that, similar to these center-section algorithms, interior-point algorithms use a new measure of the containing set represented by linear inequalities. This measure is "analytic", and is relatively easy to compute. Its associated center is called the analytic center.

2.1 Convex Bodies and Ellipsoids

A natural choice of the measure would be the volume of the convex body. Interest in measure of convex bodies dates back as far as the ancient Greeks and Chinese who computed centers, areas, perimeters and curvatures of circles, triangle, and polygons. Unlike the length of a line segment, the computation of volumes, even in two and three dimensional spaces, is not an easy task for a slightly complex shaped body. In order to measure cultivated lands, Chinese farmers weighted the amount of sand contained in a down-scaled body whose shape is identical to the land, then compared the weight to the amount of sand in an equally down-scaled unit-square box.

Associated with the volume of a convex body, the center of gravity will be the choice as the test point (Figure 2.1). We have the following theorem.

Theorem 2.1 *Let Ω be a compact convex body in $\mathcal{R}^m$ with center of gravity y^g, and let Ω^+ and Ω^- be the bodies in which a hyperplane H passing through y^g divides Ω. Then the volumes $V(\Omega^+)$ and $V(\Omega^-)$ satisfy the inequality*

$$V(\Omega^*) \leq \left(1-(1-\frac{1}{n+1})^n\right)V(\Omega), \quad \textit{where} \quad * = + \textit{ or } - .$$

This result shows that by successively cutting through the center of gravity, these convex bodies shrink at a constant rate of at most $(1 - 1/\exp(1))$, where $\exp(1)$ is the natural number 2.718.... This rate is just slightly worse than $1/2$ in the bisection method.

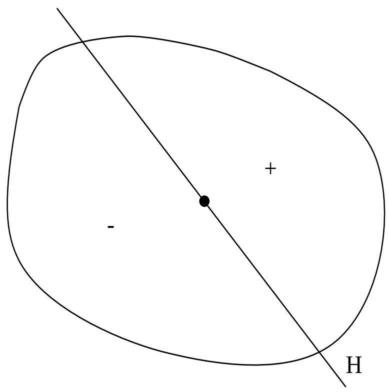

Figure 2.1: A hyperplane H cuts through the center of gravity of a convex body.

Let the solution set contain a ball with radius r and the initial containing set be contained in a ball with radius R. Then we know that the volumes of the containing sets are bounded from below by $\pi_m r^m$ and bounded from above by $\pi_m R^m$, where π_m is the volume of the unit ball in $\mathcal{R}^m$. Thus, a solution point must be found in $O(m \log(R/r))$ center-section steps, because the volumes of the containing sets eventually become too small to contain the solution set.

The difficulty with the gravity-center section method lies in computing the center and volume of a convex body. It is well-known that computing the volume of a convex polytope, either given as a list of facets or vertices, is as difficult as computing the permanent of a matrix, which is itself $\#P$-Hard. Since the computation of the center of gravity is closely related to the volume computation, it seems reasonable to conclude that no efficient algorithm can compute the center of gravity of Ω.

Although computing the center of gravity is difficult for general convex bodies, it is relatively easy for some simple convex bodies like a cube, a simplex, or an ellipsoid. This leads researchers to use some simple convex bodies to estimate Ω.

It is known that every convex body contains a unique ellipsoid of maximal volume and is contained in a unique ellipsoid of minimal volume (Figure 2.2). We have the following general theorem.

Theorem 2.2 *For every full dimensional convex body $\Omega \subset \mathcal{R}^m$ there exists a unique ellipsoid $\underline{E}(\Omega)$ of maximal volume contained in Ω. Moreover, Ω is contained in the ellipsoid obtained from $\underline{E}(\Omega)$ by enlarging it from its* *center by a factor of* m.

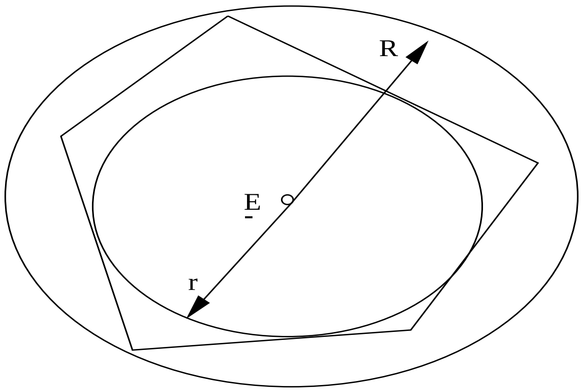

Figure 2.2: The max-volume ellipsoid inscribing a polytope and the co-centered ellipsoid circumscribing the polytope ($R/r \leq m$)

Let $\underline{y}^e$ be the center of the max-volume ellipsoid inscribing Ω. Through the center we place a hyperplane H and divide Ω into two bodies Ω^+ and Ω^-. Then, we have a center-section theorem:

Theorem 2.3 *The Volumes of the new ellipsoids satisfy the inequality*

$$V(\underline{E}(\Omega^*)) \leq 0.843 V(\underline{E}(\Omega)), \quad \textit{where} \quad * = + \textit{ or } - .$$

Thus, one can use the max-volume inscribing ellipsoid as an estimate of Ω. These ellipsoids will shrink at a constant rate in the center-section method. The volume of $\underline{E}(\Omega)$ is also bounded from below by $\pi_m r^m$ and bounded from above by $\pi_m R^m$. Thus, a solution point must be found in $O(m \log(R/r))$ center-section steps. Apparently, to compute $\underline{y}^e$ one needs to use the structure of the convex body. If the convex body is represented by linear inequalities, there is a polynomial complexity bound for computing an approximate point of $\underline{y}^e$.

Another approach is the original ellipsoid method, which monitors the volume of an ellipsoid that contains the solution set. This is based on a similar theorem:

Theorem 2.4 *For every convex body $\Omega \subset \mathcal{R}^m$ there exists a unique ellipsoid $\overline{E}(\Omega)$ of minimal volume containing Ω. Moreover, Ω contains the ellipsoid obtained from $\overline{E}(\Omega)$ by shrinking it from its center by a factor of* m.

Moreover, let $\overline{y}^e$ be the center of an ellipsoid $E \subset \mathcal{R}^m$. Through its centers we place a hyperplane H and divide E into two bodies (half ellipsoids) E^+ and E^-. Let $\overline{E}(E^+)$ and $\overline{E}(E^-)$ be the new min-volume ellipsoids containing E^+ and E^-, respectively (Figure 2.3). Then, we have the following center-section theorem:

Theorem 2.5 *The Volumes of the new ellipsoids satisfy the inequality*

$$V(\overline{E}(E^*)) \leq \exp(-.5/m)V(E), \quad \textit{where} \quad * = + \textit{ or } - .$$

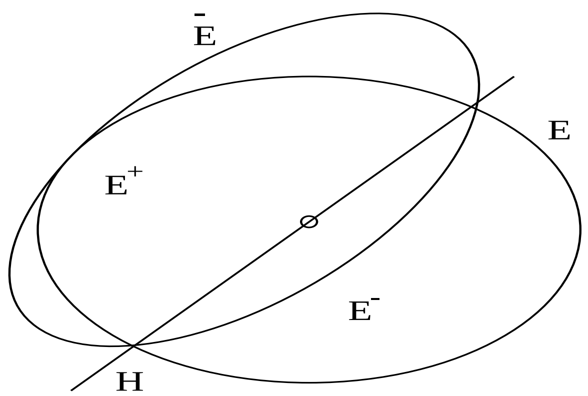

Figure 2.3: Illustration of the min-volume ellipsoid containing a half ellipsoid

Furthermore, the new containing ellipsoid can be easily constructed and its center can be computed in $O(m^2)$ arithmetic operations. Since the volumes of the ellipsoids are bounded from below by $\pi_m r^m$ and the initial one is bounded from above by $\pi_m R^m$, a solution point must be found in $O(m^2 \log(R/r))$ center-section steps. We see that the ellipsoid method does not keep the "knowledge" of the cutting plane after the new containing ellipsoid is updated.

2.2 Max-Potential and Analytic Center

The centers discussed in the preceding section are "universal," meaning that they are invariant of the representation of a convex body. A drawback of these centers is that they generally cannot be computed cost-effectively. For the ellipsoid method, its advantage in not keeping knowledge of the cutting planes is also a disadvantage to practical efficiency for solving certain problems, such as linear programs. Thus, another type of center, called the analytic center for a convex polyhedron given by linear inequalities, was introduced.

Let Ω be a bounded polytope in $\mathcal{R}^m$ defined by n $(> m)$ linear inequalities, i.e.,

$$\Omega = \{y \in \mathcal{R}^m : c - A^T y \geq 0\},$$

where $A \in \mathcal{R}^{m\times n}$ and $c \in \mathcal{R}^n$ are given and A has rank m. Denote the interior of Ω by

$$\overset{\circ}{\Omega} = \{y \in \mathcal{R}^m : c - A^T y > 0\}.$$

Given a point in Ω, define the distance function of y to the boundary of Ω which satisfies

1. $d(y, \Omega) = 0$ if y is on the boundary of Ω;
2. $d(y, \Omega) > 0$ if y is in $\overset{\circ}{\Omega}$;
3. If $c' \leq c$ and let $\Omega' = \{y \in \mathcal{R}^m : c' - A^T y \geq 0\}$ (thereby, $\Omega' \subset \Omega$), then $d(y, \Omega') \leq d(y, \Omega)$.

This definition is very similar to Huard's generic distance function, with one exception, where property (3) was stated as "If $\Omega' \subset \Omega$, then $d(y, \Omega') \leq d(y, \Omega)$." The reason for the difference is that the distance function may return different values even if we have the same polytope but two different representations. In other words, the distance function is dependent on the representation of Ω. We may have $d(y, \Omega') \leq d(y, \Omega)$ even when $\Omega \subset \Omega'$ geometrically. Thus, the distance function is really a function of point y, and the data A and c as well.

One choice of the distance functions is

$$d(y, \Omega) = \prod_{j=1}^{n} (c_j - a_j^T y), \quad y \in \Omega,$$

where a_j is the jth column of A. Traditionally, we let $s := c - A^T y$ and call it a slack vector. Thus, the distance function is the product of all slack variables. Its logarithm is called the barrier or (dual) potential function,

$$\mathcal{B}(y, \Omega) = \log d(y, \Omega) = \sum_{j=1}^{n} \log(c_j - a_j^T y) = \sum_{j=1}^{n} \log s_j.$$

The interior point, denoted by y^a, in Ω that maximizes the function is called the analytic center of Ω, i.e.,

$$\mathcal{B}(\Omega) := \mathcal{B}(y^a, \Omega) = \max_{y\in\Omega} \log d(y, \Omega).$$

y^a is uniquely defined, since the barrier function is strictly concave in $\overset{\circ}{\Omega}$. It satisfies the following optimality conditions. There exists an x^a such that

$$X^a s^a = e, \quad Ax^a = 0, \quad \text{and} \quad s^a = c - A^T y^a > 0. \tag{2.1}$$

Note that the scale of x^a is immaterial in these conditions. Also, adding or deleting a redundant inequality changes the location of the analytic center.

Example 2.1 *Consider* $\Omega = \{y \in R : y \geq 0, \ -y \geq -1\}$ *which is interval* $[0, 1]$. *The analytic center is* $y^a = 1/2$ *with* $x^a = (2, 2)^T$.

Example 2.2 *Consider* $\Omega = \{y \in R : y \geq 0, \ ..., \ y \geq 0, \ -y \geq -1\}$ *which is again interval* $[0, 1]$ *but "*$y \geq 0$*" is copied* n *times. The analytic center for this system is* $y^a = n/(n+1)$ *with* $x^a = ((n+1)/n, \ ..., \ (n+1)/n, \ (n+1))^T$.

The analytic center can be defined when the interior is empty or equalities are presented, such as

$$\Omega = \{y \in \mathcal{R}^m : c - A^T y \geq 0, \ By = b\}$$

Then the analytic center is chosen on the hyperplane $\{y : By = b\}$ to maximize the product of the slack variables $s = c - A^T y$. Thus, the interior of Ω is not used in the sense that the topological interior for a set is used. Rather, it refers to the interior of the positive orthant of slack variables: $\mathcal{R}^n_+ := \{s : \ s \geq 0\}$. When say Ω as an interior, we mean that

$$\overset{\circ}{\mathcal{R}}{}^n_+ \cap \{s : \ s = c - A^T y \text{ for some } y \text{ where } By = b\} \neq \emptyset.$$

Again $\overset{\circ}{\mathcal{R}}{}^n_+ := \{s \in \mathcal{R}^n_+ : \ s > 0\}$, i.e., the interior of the orthant $\mathcal{R}^n_+$. Thus, if Ω has only a single point y with $s = c - A^T y > 0$, we still say $\overset{\circ}{\Omega}$ is not empty.

Example 2.3 *Consider the system* $\Omega = \{x : \ Ax = 0, \ e^T x = n, \ x \geq 0\}$, *which is called Karmarkar's canonical set. If* $x = e$ *is in* Ω, *then* e *is the analytic center of* Ω *(Figure 2.4).*

2.2.1 Dual potential function

We may represent $\Omega = \{y \in \mathcal{R}^m : \ c - A^T y \geq 0\}$ using the slack variable s only:

$$\mathcal{S}_\Omega := \{s \in \mathcal{R}^n : \ A^T y + s = c, \ s \geq 0\},$$

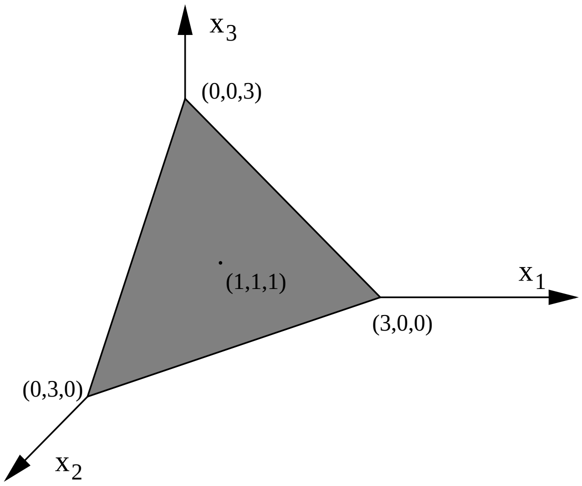

Figure 2.4: Illustration of the Karmarkar (simplex) polytope and its analytic center

or

$$\mathcal{S}_\Omega := \{s \in \mathcal{R}^n : \ s - c \in \mathcal{R}(A^T), \ s \geq 0\},$$

which is the intersection of the affine set

$$\mathcal{A}_\Omega = \{s \in \mathcal{R}^n : \ s - c \in \mathcal{R}(A^T)\}$$

and the positive cone (orthant) $\mathcal{R}^n_+$. The interior of $\mathcal{S}_\Omega$ is denoted by

$$\overset{\circ}{\mathcal{S}}_\Omega := \mathcal{A}_\Omega \cap \overset{\circ}{\mathcal{R}}{}^n_+.$$

Let s be an interior point in $\mathcal{S}_\Omega$. Then consider the ellipsoid

$$E_s = \{t \in \mathcal{R}^n : \ \|S^{-1}(t-s)\| \leq 1\}.$$

This is a regular ellipsoid centered at s and inscribing the positive orthant $\mathcal{R}^n_+$. The volume of the regular ellipsoid is

$$V(E_s) = \pi_n \prod_{j=1}^{n} s_j.$$

Moreover, we have

$$(E_s \cap \mathcal{A}_\Omega) \subset \mathcal{S}_\Omega,$$

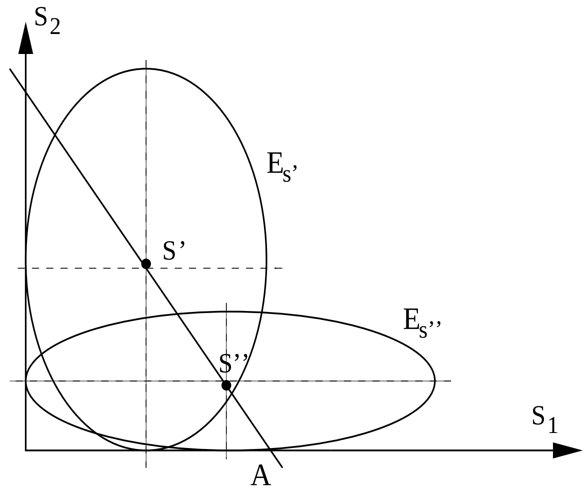

Figure 2.5: Regular (dual) ellipsoids centered at points s's on the intersection of an affine set A and the positive orthant; they are also contained by the positive orthant.

that is, the intersection of E_s and $\mathcal{A}_\Omega$ is contained in $\mathcal{S}_\Omega$ (Figure 2.5).

Thus, the potential function at $s \in \overset{\circ}{\mathcal{S}}_\Omega$ plus $\log \pi_n$ is the logarithmic volume of the regular ellipsoid centered at s and inscribing $\mathcal{R}^n_+$. Therefore, the inscribing regular ellipsoid centered at the analytic center of $\mathcal{S}_\Omega$, among all of these inscribing regular ellipsoids, has the maximal volume. We denote this max-barrier or max-potential of $\mathcal{S}_\Omega$ by $\mathcal{B}(\Omega)$.

We now argue that the exponential of $\mathcal{B}(\Omega)$ is an "analytic" measure of $\mathcal{S}_\Omega$ or Ω.

1. $\exp(\mathcal{B}(\Omega)) = 0$ if $\overset{\circ}{\Omega} = \emptyset$;

2. $\exp(\mathcal{B}(\Omega)) > 0$ if $\overset{\circ}{\Omega} \neq \emptyset$, and if Ω contains a full dimensional ball with radius r or $\{y : \ A^T y \leq c - re\} \neq \emptyset$ (here we assume that $\|a_j\| = 1$ for $j = 1, 2, ..., n$), then $\mathcal{B}(\Omega) \geq n \log r$;

3. If $c' \leq c$ and let $\Omega' = \{y \in \mathcal{R}^m : c' - A^T y \geq 0\}$ (thereby, $\Omega' \subset \Omega$), then $\mathcal{B}(\Omega') \leq \mathcal{B}(\Omega)$.

Note that the max-barrier or max-potential $\mathcal{B}(\Omega)$ is a function of data A and c.

Let s^a (or y^a) be the analytic center of $\mathcal{S}_\Omega$ (or Ω). We now consider the ellipsoid

$$nE_{s^a} = \{t \in \mathcal{R}^n : \ \|(S^a)^{-1}(t - s^a)\| \le n\},$$

which is enlarged from E_{s^a} by a factor of n. The question is whether or not this enlarged regular ellipsoid $nE_{s^a} \cap \mathcal{A}_\Omega$ contains $\mathcal{S}_\Omega$. The answer is "yes" according to the following theorem.

Theorem 2.6 *The analytic center $s^a \in \mathcal{S}_\Omega$ is the center of the unique maximal-volume regular ellipsoid E_{s^a} inscribing the orthant $\mathcal{R}^n_+$. Its intersection with $\mathcal{A}_\Omega$ is contained by polytope $\mathcal{S}_\Omega$. Moreover, polytope $\mathcal{S}_\Omega$ itself is contained by the intersection of $\mathcal{A}_\Omega$ and the ellipsoid obtained from E_{s^a} by enlarging it from its center by a factor n.*

Proof. The uniqueness of the analytic center is resulted from the fact that the potential function is strictly concave in the interior of the polytope and A has a full row-rank. Let y^a be the analytic center and $s^a = c - A^T y^a$, then there is x^a such that $Ax^a = 0$ and $X^a s^a = e$. Thus, we have $x^a > 0$ and $c^T x^a = n$. For all $s = c - A^T y \ge 0$ we have

$$\|(S^a)^{-1}(s-s^a)\|^2 = \|X^a s - e\|^2 = \|X^a s\|^2 - n \le ((x^a)^T s)^2 - n = n^2 - n < n^2.$$

This completes the proof.

□

2.2.2 Analytic center section

We now develop two center-section inequalities for the analytic center. They resemble the results for the previously discussed centers. First, we study how translating a hyperplane in $\Omega = \{y : \ c - A^T y \ge 0\}$ will affect the max-potential value. More specifically, we have the following problem: If one inequality in Ω, say the first one, of $c - A^T y \ge 0$ needs to be translated, change $c_1 - a_1^T y \ge 0$ to $a_1^T y^a - a_1^T y \ge 0$; i.e., the first inequality is parallelly translated, and it cuts through the center y^a and divides Ω into two bodies (Figure 2.6). Let

$$\Omega^+ := \{y : \ a_1^T y^a - a_1^T y \ge 0, \ c_j - a_j^T y \ge 0, \ j = 2, ...n\}$$

and let $\bar{y}^a$ be the analytic center of Ω^+. Then, the max-potential for the new convex polytope Ω^+ is

$$\exp(\mathcal{B}(\Omega^+)) = (a_1^T y^a - a_1^T \bar{y}^a) \prod_{j=2}^n (c_j - a_j^T \bar{y}^a).$$

Regarding $\mathcal{B}(\Omega)$ and $\mathcal{B}(\Omega^+)$, we prove the following theorem.

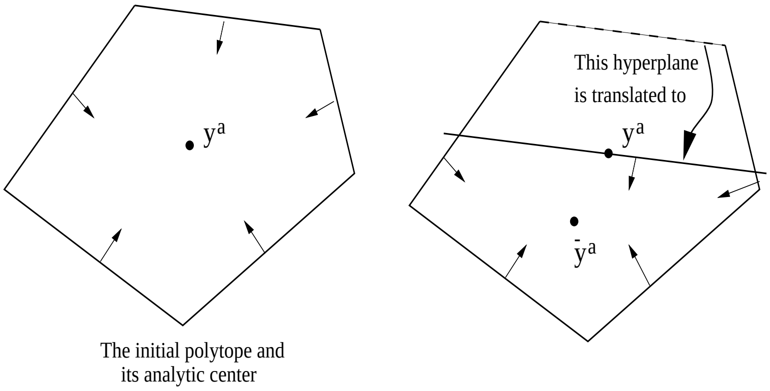

Figure 2.6: Translation of a hyperplane

Theorem 2.7 *Let Ω and Ω^+ be defined as the above. Then*

$$\mathcal{B}(\Omega^+) \leq \mathcal{B}(\Omega) - 1,$$

or

$$\exp(\mathcal{B}(\Omega^+)) \leq \exp(-1)\exp(\mathcal{B}(\Omega)),$$

where $\exp(-1) = 1/\exp(1)$.

Proof. Since y^a is the analytic center of Ω, there exists $x^a > 0$ such that

$$X^a(c - A^T y^a) = e \quad \text{and} \quad Ax^a = 0. \tag{2.2}$$

Recall that e is the vector of all ones and X^a designates the diagonal matrix of x^a. Thus, we have

$$s^a = (c - A^T y^a) = (X^a)^{-1}e \quad \text{and} \quad c^T x^a = n.$$

Let $\bar{c}_j = c_j$ for $j = 2, ..., n$ and $\bar{c}_1 = a_1^T y^a$, and let $\bar{s}^a = \bar{c} - A^T \bar{y}^a$. Then, we have

$$\begin{aligned} e^T X^a \bar{s}^a &= e^T X^a(\bar{c} - A^T \bar{y}^a) = e^T X^a \bar{c} \\ &= c^T x^a - x_1^a(c_1 - a_1^T y^a) = n - 1. \end{aligned}$$

Thus,

$$\frac{\exp(\mathcal{B}(\Omega^+))}{\exp(\mathcal{B}(\Omega))} = \prod_{j=1}^{n} \frac{\bar{s}_j^a}{s_j^a}$$

$$
\begin{aligned}
&= \prod_{j=1}^{n} x_j^a \bar{s}_j^a \\
&\leq \left(\frac{1}{n}\sum_{j=1}^{n} x_j^a \bar{s}_j^a\right)^n \\
&= \left(\frac{n-1}{n}\right)^n \leq \exp(-1).
\end{aligned}
$$

□

Now suppose we translate the first hyperplane by a β, $0 \leq \beta \leq 1$, the fractional distance to the analytic center, i.e.,

$$\Omega^+ := \{y: \ (1-\beta)c_1 + \beta a_1^T y^a - a_1^T y \geq 0, \ c_j - a_j^T y \geq 0, \ j = 2, ...n\}.$$

If $\beta = 0$, then there is no translation; if $\beta = 1$, then the hyperplane is translated through the analytic center as in the above theorem. Regarding $\mathcal{B}(\Omega)$ and $\mathcal{B}(\Omega^+)$, we have the following inequality.

Corollary 2.8

$$\mathcal{B}(\Omega^+) \leq \mathcal{B}(\Omega) - \beta.$$

Now suppose we translate $k(< n)$ hyperplanes, say $1, 2, ..., k$, cutting through Ω; i.e., use multiple cuts passing through Ω, and let

$$\Omega^+ := \{y : (1-\beta_j)c_j + \beta_j a_j^T y^a - a_j^T y \geq 0, \ j = 1, ..., k,$$

$$c_j - a_j^T y \geq 0, \ j = k+1, ..., n, \}$$

where $0 \leq \beta_j \leq 1$, $j = 1, 2, ..., k$. Then, we have the following corollary.

Corollary 2.9

$$\mathcal{B}(\Omega^+) \leq \mathcal{B}(\Omega) - \sum_{j=1}^{k} \beta_j.$$

This corollary will play an important role in establishing the current best complexity result for linear inequality and linear programming problems.

These corollaries show the shrinking nature of the regular ellipsoids after a cut is translated. They enable us to develop an algorithm that resembles the center-section method. Again, if a lower bound on the max-potential of the solution set is $n \log r$ and the max-potential of the initial containing set is $n \log R$, then a solution must be found in $O(n \log(R/r))$ center-section

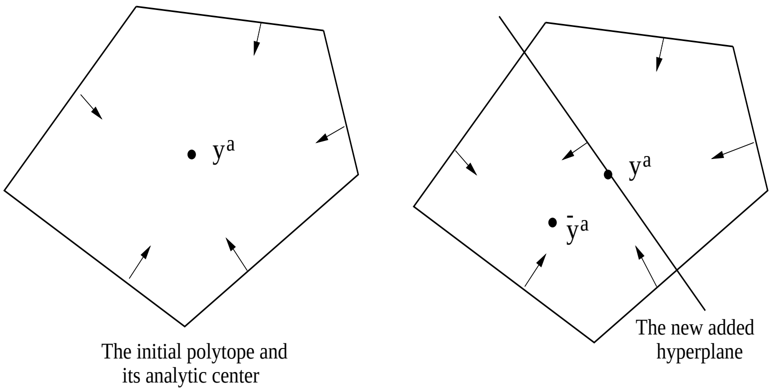

Figure 2.7: Addition of a new hyperplane

steps. Moreover, if we can translate multiple inequalities and the max-potential is reduced by $\theta(\sqrt{n})$ at each step, a solution must be found in $O(\sqrt{n}\log(R/r))$ center-section steps.

In the following, we study the problem when placing additional hyperplanes through the analytic center of Ω (Figure 2.7). When a hyperplane is added, the new convex body is represented by

$$\Omega^+ = \{y : c - A^T y \geq 0, \quad a_{n+1}^T y^a - a_{n+1}^T y \geq 0\}.$$

Again, the question is how the max-potential of the new polytope changes compared to that of Ω. Let

$$r(\Omega)_1 = \sqrt{a_{n+1}^T (A(X^a)^2 A^T)^{-1} a_{n+1}} = \sqrt{a_{n+1}^T (A(S^a)^{-2} A^T)^{-1} a_{n+1}}.$$

Then, we have an inequality as follows.

Theorem 2.10

$$\mathcal{B}(\Omega^+)) \leq \mathcal{B}(\Omega) + \log(r(\Omega)_1) + 2\log 2 - 1.5.$$

Proof. Again, x^a and (y^a, s^a) satisfy condition (2.2). Let

$$\bar{s}^a = c - A^T \bar{y}^a \quad \text{and} \quad \bar{s}_{n+1}^a = c_{n+1} - a_{n+1}^T \bar{y}^a. \tag{2.3}$$

Then,

$$\begin{aligned} \bar{s}_{n+1}^a &= a_{n+1}^T (y^a - \bar{y}^a) \\ &= a_{n+1}^T (A(X^a)^2 A^T)^{-1} (A(X^a)^2 A^T)(y^a - \bar{y}^a) \end{aligned}$$

$$
\begin{aligned}
&= a_{n+1}^T(A(X^a)^2A^T)^{-1}A(X^a)^2(A^Ty^a - A^T\bar{y}^a) \\
&= a_{n+1}^T(A(X^a)^2A^T)^{-1}A(X^a)^2(-c + A^Ty^a + c - A^T\bar{y}^a) \\
&= a_{n+1}^T(A(X^a)^2A^T)^{-1}A(X^a)^2(\bar{s}^a - (X^a)^{-1}e) \\
&= a_{n+1}^T(A(X^a)^2A^T)^{-1}AX^a(X^a\bar{s}^a - e). \qquad (2.4)
\end{aligned}
$$

Note that we have

$$
e^TX^a\bar{s}^a = e^TX^a(c - A^T\bar{y}^a) = e^TX^ac = n. \qquad (2.5)
$$

Thus, from (2.4)

$$
\begin{aligned}
\frac{\exp(\mathcal{B}(\Omega^+))}{\exp(\mathcal{B}(\Omega))r(\Omega)_1} &= \frac{\bar{s}_{n+1}^a}{r(\Omega)_1}\prod_{j=1}^n \frac{\bar{s}_j^a}{s_j^a} \\
&= \frac{a_{n+1}^T(A(X^a)^2A^T)^{-1}AX^a(X^a\bar{s}^a - e)}{r(\Omega)_1}\prod_{j=1}^n (x_j^a\bar{s}_j^a) \\
&\le \frac{\|a_{n+1}^T(A(X^a)^2A^T)^{-1}AX^a\|\|X^a\bar{s}^a - e\|}{r(\Omega)_1}\prod_{j=1}^n (x_j^a\bar{s}_j^a) \\
&= \|X^a\bar{s}^a - e\|\prod_{j=1}^n (x_j^a\bar{s}_j^a). \qquad (2.6)
\end{aligned}
$$

Let $\alpha = X^a\bar{s}^a$. Then, to evaluate (2.6) together with (2.5) we face a maximum problem

$$
\begin{array}{ll}
\text{maximize} & f(\alpha) = \|\alpha - e\|\prod_{j=1}^n \alpha_j \\
\text{subject to} & e^T\alpha = n, \alpha > 0.
\end{array}
$$

This maximum is achieved, without loss of generality, at $\alpha_1 = \beta > 1$ and $\alpha_2 = \ldots = \alpha_n = (n-\beta)/(n-1) > 0$. Hence,

$$
\begin{aligned}
f(\alpha) &\le (\beta - 1)\sqrt{\frac{n}{n-1}}\beta(\frac{n-\beta}{n-1})^{n-1} \\
&\le 4\sqrt{\frac{n}{n-1}}\frac{\beta-1}{2}\frac{\beta}{2}(\frac{n-\beta}{n-1})^{n-1} \\
&\le 4\sqrt{\frac{n}{n-1}}(\frac{n-.5}{n+1})^{n+1} \\
&\le \frac{4}{\mathrm{e}^{1.5}}.
\end{aligned}
$$

This completes the proof.

□

Note that $r(\Omega)_1 = \sqrt{a_{n+1}^T (A(S^a)^{-2}A^T)^{-1} a_{n+1}}$ is the maximal value of the problem:

$$\max \ a_{n+1}^T (y - y^a) \quad s.t. \quad y \in \{y : \|(S^a)^{-1}A^T(y - y^a)\| \le 1\} \subset \Omega.$$

Thus, if the initial ellipsoid is contained in a unit ball, then $r(\Omega)_1 \le 1$.

In general, if k additional hyperplanes cut through the analytic center, i.e.,

$$\Omega^+ = \{y : c - A^T y \ge 0, \ a_{n+1}^T y^a - a_{n+1}^T y \ge 0, ..., a_{n+k}^T y^a - a_{n+k}^T y \ge 0\},$$

we have

Corollary 2.11

$$\mathcal{B}(\Omega^+) \le \mathcal{B}(\Omega) + \sum_{i=1}^{k} \log(r(\Omega)_i) + (k+1)\log(k+1) - (k + \frac{k}{k+1}),$$

where

$$r(\Omega)_i = \sqrt{a_{n+i}^T (A(S^a)^{-2}A^T)^{-1} a_{n+i}}, \quad i = 1, ..., k.$$

2.3 Primal and Primal-Dual Potential Functions

From the duality theorem, we have a (homogeneous) linear programming problem related to $\Omega = \{y : \ c - A^T y \ge 0\}$ as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = 0, x \ge 0, \end{array}$$

which we called the primal problem. If Ω is nonempty, then the minimal value of the problem is 0; if Ω is bounded and has an interior, then the interior of $\mathcal{X}_\Omega := \{x \in \mathcal{R}^n : \ Ax = 0, \ x \ge 0\}$ is nonempty and $x = 0$ is the unique (primal) minimal solution.

2.3.1 Primal potential function

One can define a potential function for $\mathcal{X}_\Omega$ as

$$\mathcal{P}(x, \Omega) = n \log(c^T x) - \sum_{j=1}^{n} \log x_j, \quad x \in \overset{\circ}{\mathcal{X}}_\Omega \ .$$

This is the so-called Karmarkar potential function. We now show that this quantity represents the logarithmic volume of a regular ellipsoid whose intersection with $\mathcal{A}_\Omega$ contains $\mathcal{S}_\Omega = \mathcal{R}^n_+ \cap \mathcal{A}_\Omega$.

Recall that we are interested in finding a point $s \in \mathcal{S}_\Omega$. Let $x \in \overset{\circ}{\mathcal{X}}_\Omega$. Then, for all $s \in \mathcal{S}_\Omega$ we must have

$$\|Xs\|^2 \le (x^T s)^2 = (c^T x)^2. \tag{2.7}$$

Let $\bar{y} = (AX^2A^T)^{-1}AX^2c$ and $\bar{s} = c - A^T\bar{y}$. Then, we see for any $s \in \mathcal{S}_\Omega$, we have

$$\|Xs\|^2 = \|X(s-\bar{s}) + X\bar{s}\|^2 = \|X(s-\bar{s})\|^2 + \|X\bar{s}\|^2,$$

or

$$\|X(s-\bar{s})\|^2 = \|Xs\|^2 - \|X\bar{s}\|^2 \le \|Xs\|^2 \le (c^T x)^2.$$

Thus, let E_x be the regular ellipsoid $E_x = \{s \in \mathcal{R}^n : \ \|X(s-\bar{s})\| \le c^T x\}$ that is centered at $\bar{s}$ (Figure 2.8). Then, we must have

$$\mathcal{S}_\Omega \subset (E_x \cap \mathcal{A}_\Omega).$$

Furthermore, the volume of E_x is

$$V(E_x) = \frac{\pi_n (c^T x)^n}{\det(X)} = \frac{\pi_n (c^T x)^n}{\prod_{j=1}^n x_j}, \tag{2.8}$$

where π_n is the volume of the unit ball in $\mathcal{R}^n$. Thus,

$$\mathcal{P}(x, \Omega) = \log V(E_x) - \log \pi_n.$$

In the next chapter, we will show that Karmarkar's algorithm actually generates sequences $\{0 < x^k \in X\}$ such that

$$\mathcal{P}(x^{k+1}, \Omega) \le \mathcal{P}(x^k, \Omega) - .3$$

for $k = 0, 1, 2, ...$ In other words

$$\frac{V(E_{x^{k+1}})}{V(E_{x^k})} \le \exp(-.3).$$

That is, the volume of the containing ellipsoids shrinks at a constant rate. Note that E_{x^k} contains the solution set $\mathcal{S}_\Omega$. Therefore, Karmarkar's algorithm conceptually resembles the ellipsoid method.

Since the primal potential function represents the volume of a regular ellipsoid containing $\mathcal{S}_\Omega$, let us minimize it over all $x \in \overset{\circ}{\mathcal{X}}_\Omega$. Note that the

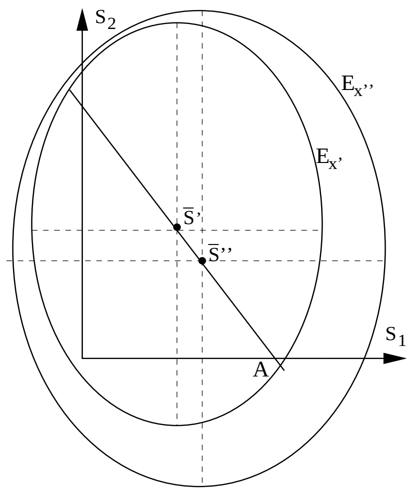

Figure 2.8: Regular (primal) ellipsoids centered at points $\bar{s}$'s on an affine set A; they also contain the intersection of A and the positive orthant.

primal potential function is homogeneous degree 0 in x, so that we can fix $c^T x = n$. Then the problem minimizes a strictly convex function and the optimality conditions become

$$x \in \mathcal{X}_\Omega \text{ and } X(c - A^T y) = e \text{ for some } y \in \Omega. \tag{2.9}$$

One can see that y in (2.9) is in fact the analytic center y^a of Ω, since this condition is identical to condition (2.1) for defining the analytic center of Ω. Let x^a, called the primal analytic center, satisfy these conditions. Recall that $s^a = c - A^T y^a$ is the analytic center of $\mathcal{S}_\Omega$. Then, we have

Theorem 2.12 *There is a unique minimal-volume regular ellipsoid E_{x^a}, where $x^a \in \overset{\circ}{\mathcal{X}}_\Omega$, whose intersection with $\mathcal{A}_\Omega$ contains polytope $\mathcal{S}_\Omega$. Moreover, polytope $\mathcal{S}_\Omega$ contains the intersection of $\mathcal{A}_\Omega$ and the ellipsoid obtained from E_{x^a} by shrinking it from its center by a factor of n. In fact, the two ellipsoids E_{x^a} and E_{s^a} (in Theorem 2.6) are co-centered, and they can be obtained from each other by enlarging or shrinking with a factor n (Figures 2.5 and 2.8).*

Proof. To prove they are co-centered, we only need to prove that the center of E_{x^a} is the analytic center of $\mathcal{S}_\Omega$. Recall that $\bar{s}$ is the center of E_{x^a} with $\bar{s} = c - A^T \bar{y}$, where

$$\bar{y} = (A(X^a)^2 A^T)^{-1} A(X^a)^2 c.$$

On the other hand, from (2.9) we have

$$X^a(c - A^T y^a) = e \quad \text{or} \quad X^a c = e + X^a A^T y^a,$$

where y^a is the analytic center of Ω. Thus

$$X^a(c - A^T\bar{y}) = e + X^a A^T y^a - X^a A^T (A(X^a)^2 A^T)^{-1} A X^a (e + X^a A^T y^a)$$

$$= e - X^a A^T (A(X^a)^2 A^T)^{-1} A X^a e = e.$$

Thus, $\bar{y} = y^a$ and $\bar{s} = s^a$ since $\bar{y}$ also satisfies the optimality condition of (2.1).

□

2.3.2 Primal-dual potential function

For $x \in \overset{\circ}{\mathcal{X}}_\Omega$ and $s \in \overset{\circ}{\mathcal{S}}_\Omega$ consider a primal-dual potential function which has the form

$$\psi_n(x, s) := n \log(x^T s) - \sum \log(x_j s_j)$$

$$= n \log(c^T x) - \sum \log x_j - \sum \log s_j = \mathcal{P}(x, \Omega) - \mathcal{B}(y, \Omega).$$

This is the logarithmic ratio of the volume of E_x over the volume of E_s. We also have, from the arithmetic-geometric mean inequality,

$$\psi_n(x, s) = n \log(x^T s) - \sum \log(x_j s_j) \geq n \log n,$$

and

$$\psi_n(x^a, s^a) = n \log n.$$

This is the precise logarithmic ratio of the volumes of two co-centered ellipsoids whose radii are differentiated by a factor n (Figures 2.5 and 2.8).

2.4 Potential Functions for Linear Programming

We now consider a linear program in the standard form (LP) and (LD) with the optimal value z^*. Denote the feasible sets of (LP) and (LD) by $\mathcal{F}_p$ and $\mathcal{F}_d$, respectively. Denote by $\mathcal{F} = \mathcal{F}_p \times \mathcal{F}_d$.

2.4.1 Primal potential function

Consider the level set

$$\Omega(z) = \{y : \; c - A^T y \geq 0, -z + b^T y \geq 0\}$$

where $z \leq z^*$. Let (x', x'_0) satisfy

$$Ax' - bx'_0 = 0, \; (x', x'_0) > 0,$$

and let $x := x'/x'_0 \in \overset{\circ}{\mathcal{F}}_p$, i.e.,

$$Ax = b, \; x > 0.$$

Then, the primal potential function for $\Omega(z)$ (Figure 2.9), as described in the preceding section, is

$$\mathcal{P}(x', \Omega(z)) = (n+1)\log(c^T x' - z x'_0) - \sum_{j=0}^{n} \log x'_j$$

$$= (n+1)\log(c^T x - z) - \sum_{j=1}^{n} \log x_j =: \mathcal{P}_{n+1}(x, z).$$

The later, $\mathcal{P}(x, z)_{n+1}$, is the Karmarkar potential function in the standard LP form with a lower bound z for z^*. We see that it represents the volume of a regular ellipsoid whose intersection with $\mathcal{A}_{\Omega(z)}$ contains $\mathcal{S}_{\Omega(z)}$, as we discussed earlier.

As we illustrated before, one can represent $\Omega(z)$ differently:

$$\Omega(z) = \{y : \; c - A^T y \geq 0, -z + b^T y \geq 0, ..., -z + b^T y \geq 0\}, \tag{2.10}$$

where "$-z + b^T y \geq 0$" is copied ρ times. Geometrically, this representation does not change $\Omega(z)$, but it changes the location of its analytic center. Let (x', x'_0) satisfy

$$Ax' - b(\rho x'_0) = 0, \; (x', x'_0) > 0,$$

and let $x = x'/(\rho x'_0) \in \overset{\circ}{\mathcal{F}}_p$. Then, the primal potential function for the new $\Omega(z)$ given by (2.10) is

$$\mathcal{P}(x, \Omega(z)) = (n+\rho)\log(c^T x' - z\rho x'_0) - \sum_{j=1}^{n} \log x'_j - \rho \log x'_0$$

$$= (n+\rho)\log(c^T x - z) - \sum_{j=1}^{n} \log x_j + \rho \log \rho$$

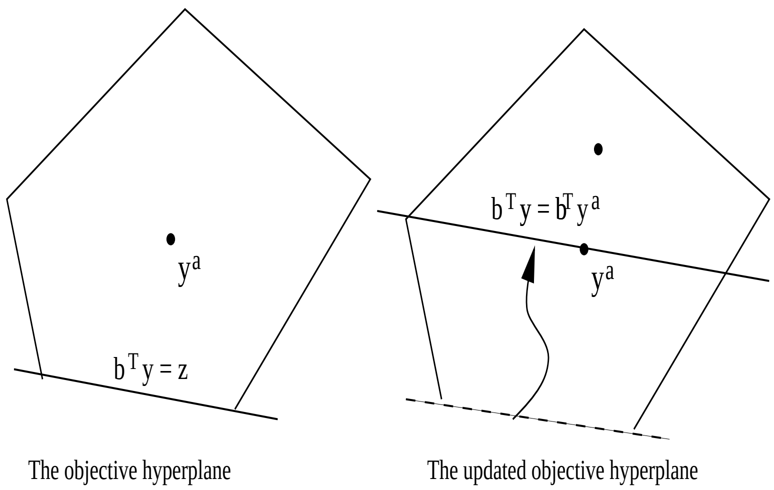

Figure 2.9: Intersections of a dual feasible region and the objective hyperplane; $b^Ty \geq z$ on the left and $b^Ty \geq b^Ty^a$ on the right.

$$:= \mathcal{P}_{n+\rho}(x,z) + \rho\log\rho.$$

$\mathcal{P}_{n+\rho}(x,z)$ is an extension of the Karmarkar potential function in the standard LP form with a lower bound z for z^*. It represents the volume of a regular ellipsoid whose intersection with $\mathcal{A}_{\Omega(z)}$ contains $\mathcal{S}_{\Omega(z)}$.

2.4.2 Primal-dual potential function

Finally, for $x \in \overset{\circ}{\mathcal{F}}_p$ and $(y,s) \in \overset{\circ}{\mathcal{F}}_d$ consider the primal-dual potential function

$$\psi_{n+\rho}(x,s) := (n+\rho)\log(x^Ts) - \sum_{j=1}^n \log(x_js_j)$$

$$= (n+\rho)\log(c^Tx - b^Ty) - \sum_{j=1}^n \log x_j - \sum_{j=1}^n \log s_j,$$

$$= \mathcal{P}_{n+\rho}(x, b^Ty) - \sum_{j=1}^n \log s_j,$$

where $\rho \geq 0$. Since

$$\psi_{n+\rho}(x,s) = \rho\log(x^Ts) + \psi_n(x,s) \geq \rho\log(x^Ts) + n\log n,$$

then, for $\rho > 0$, $\psi_{n+\rho}(x,s) \to -\infty$ implies that $x^T s \to 0$. More precisely, we have

$$x^T s \leq \exp(\frac{\psi_{n+\rho}(x,s) - n\log n}{\rho}).$$

We have the following theorem.

Theorem 2.13 *Define the level set*

$$\Psi(\delta) := \{(x,y,s) \in \overset{\circ}{\mathcal{F}}:\ \psi_{n+\rho}(x,s) \leq \delta\}.$$

i)

$$\Psi(\delta^1) \subset \Psi(\delta^2) \quad \textit{if} \quad \delta^1 \leq \delta^2.$$

ii)

$$\overset{\circ}{\Psi}(\delta) = \{(x,y,s) \in \mathcal{F}:\ \psi_{n+\rho}(x,s) < \delta\}.$$

iii) *For every δ, $\Psi(\delta)$ is bounded and its closure $\hat{\Psi}(\delta)$ has non-empty intersection with the solution set.*

Later we will show that a potential reduction algorithm generates sequences $\{x^k, y^k, s^k\} \in \overset{\circ}{\mathcal{F}}$ such that

$$\psi_{n+\sqrt{n}}(x^{k+1}, y^{k+1}, s^{k+1}) \leq \psi_{n+\sqrt{n}}(x^k, y^k, s^k) - .05$$

for $k = 0, 1, 2, ...$ This indicates that the level sets shrink at least a constant rate independently of m or n.

2.5 Central Path of Linear Programming

Again we consider a linear program in the standard form (LP) and (LD). Assume that $\overset{\circ}{\mathcal{F}} \neq \emptyset$, i.e., both $\overset{\circ}{\mathcal{F}}_p \neq \emptyset$ and $\overset{\circ}{\mathcal{F}}_d \neq \emptyset$, and denote z^* the optimal objective value.

A central path can be expressed as

$$\mathcal{C} = \left\{(x,y,s) \in \overset{\circ}{\mathcal{F}}:\ Xs = \frac{x^T s}{n} e\right\}$$

in the primal-dual form. We also see

$$\mathcal{C} = \left\{(x,y,s) \in \overset{\circ}{\mathcal{F}}:\ \psi_n(x,s) = n\log n\right\}.$$

The central path theory is one of the most important foundations for the development of interior-point algorithms.

For any $\mu > 0$ one can derive the central path simply by minimizing the primal LP with a logarithmic barrier function:

$$\begin{array}{lll} (P) & \text{minimize} & c^T x - \mu \sum_{j=1}^n \log x_j \\ & \text{subject to} & Ax = b, \ x \geq 0. \end{array}$$

Let $x(\mu)$ be the unique minimizer of (P). Then, it satisfies the optimality conditions

$$x \in \overset{\circ}{\mathcal{F}}_p, \quad Xs = \mu e, \quad \text{and} \quad s = c - A^T y.$$

Consider minimizing the dual LP with the barrier function:

$$\begin{array}{lll} (D) & \text{maximize} & b^T y + \mu \sum_{j=1}^n \log s_j \\ & \text{subject to} & A^T y + s = c, \ s \geq 0. \end{array}$$

Let $(y(\mu), s(\mu))$ be the unique minimizer of (D). Then, it satisfies the optimality conditions

$$(y, s) \in \overset{\circ}{\mathcal{F}}_d, \quad Xs = \mu e, \quad \text{and} \quad Ax = b.$$

Comparing these two conditions, we see that they are identical. Thus, both minimizers $(x(\mu), y(\mu), s(\mu))$ are on the central path with $x(\mu)^T s(\mu) = n\mu$.

Another way to derive the central path is to consider the dual level set

$$\Omega(z) = \{y : \ c - A^T y \geq 0, -z + b^T y \geq 0\}$$

for any $z < z^*$ (Figure 2.10). Then, the analytic center $(y(z), s(z))$ of $\Omega(z)$ and a point $(x'(z), x_0'(z))$ satisfies

$$Ax'(z) - bx_0'(z) = 0, \ X'(z)s = e, \ s = c - A^T y, \ \text{and} \ x_0'(z)(b^T y - z) = 1.$$

Let $x(z) = x'(z)/x_0'(z)$, then we have

$$Ax(z) = b, \ X(z)s(z) = e/x_0'(z) = (b^T y(z) - z)e.$$

Thus, the center pair $(x(z), y(z), s(z))$ is on the central path with $\mu = b^T y(z) - z$ and $c^T x(z) - b^T y(z) = x(z)^T s(z) = n(b^T y(z) - z) = n\mu$.

Theorem 2.14 *Let $(x(\mu), y(\mu), s(\mu))$ be on the central path.*

i) *The central pair $(x(\mu), s(\mu))$ is bounded where $0 < \mu \leq \mu^0$ for any given $\mu^0 > 0$.*

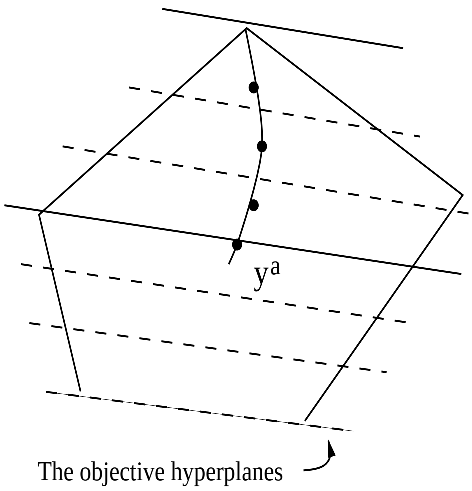

Figure 2.10: The central path of $y(z)$ in a dual feasible region

ii) *For* $0 < \mu' < \mu$,

$$c^T x(\mu') < c^T x(\mu) \quad \textit{and} \quad b^T y(\mu') > b^T y(\mu).$$

iii) $(x(\mu), s(\mu))$ *converges to an optimal solution pair for (LP) and (LD). Moreover, the limit point* $x(0)_{P^*}$ *is the analytic center on the primal optimal face, and the limit point* $s(0)_{Z^*}$ *is the analytic center on the dual optimal face, where* (P^*, Z^*) *is the strict complementarity partition of the index set* $\{1, 2, ..., n\}$.

Proof. Note that

$$(x(\mu^0) - x(\mu))^T (s(\mu^0) - s(\mu)) = 0$$

since $(x(\mu^0) - x(\mu)) \in \mathcal{N}(A)$ and $(s(\mu^0) - s(\mu)) \in \mathcal{R}(A^T)$. This can be rewritten as

$$\sum_j^n \left(s(\mu^0)_j x(\mu)_j + x(\mu^0)_j s(\mu)_j \right) = n(\mu^0 + \mu) \leq 2n\mu^0,$$

or

$$\sum_j^n \left(\frac{x(\mu)_j}{x(\mu^0)_j} + \frac{s(\mu)_j}{s(\mu^0)_j} \right) \leq 2n.$$

Thus, $x(\mu)$ and $s(\mu)$ are bounded, which proves (i).

We leave the proof of (ii) as an exercise.

Let $x^*_{P^*}$ ($x^*_{Z^*} = 0$) and $s^*_{Z^*}$ ($s^*_{P^*} = 0$), respectively, be the analytic centers on the primal and dual optimal faces: $\{x_{P^*} : \ A_{P^*}x_{P^*} = b, \ x_{P^*} \geq 0\}$ and $\{s_{Z^*} : \ s_{Z^*} = c_{Z^*} - A^T_{Z^*}y \geq 0, \ c_{P^*} - A^T_{P^*}y = 0\}$. Again, we have

$$\sum_j^n \left(s^*_j x(\mu)_j + x^*_j s(\mu)_j\right) = n\mu,$$

or

$$\sum_{j\in P^*}\left(\frac{x^*_j}{x(\mu)_j}\right) + \sum_{j\in Z^*}\left(\frac{s^*_j}{s(\mu)_j}\right) = n.$$

Thus, we have

$$x(\mu)_j \geq x^*_j/n > 0, \ j \in P^*$$

and

$$s(\mu)_j \geq s^*_j/n > 0, \ j \in Z^*.$$

This implies that

$$x(\mu)_j \to 0, \ j \in Z^*$$

and

$$s(\mu)_j \to 0, \ j \in P^*.$$

Furthermore,

$$\left(\prod_{j\in P^*}\frac{x^*_j}{x(\mu)_j}\right)\left(\prod_{j\in Z^*}\frac{s^*_j}{s(\mu)_j}\right) \leq 1$$

or

$$\left(\prod_{j\in P^*} x^*_j\right)\left(\prod_{j\in Z^*} s^*_j\right) \leq \left(\prod_{j\in P^*} x(\mu)_j\right)\left(\prod_{j\in Z^*} s(\mu)_j\right).$$

However, $(\prod_{j\in P^*} x^*_j)(\prod_{j\in Z^*} s^*_j)$ is the maximal value of the barrier function over all interior point pairs on the optimal face, and $x(0)_{P^*}$ and $s(0)_{Z^*}$ is one interior point pair on the optimal face. Thus, we must have

$$\left(\prod_{j\in P^*} x^*_j\right)\left(\prod_{j\in Z^*} s^*_j\right) = \left(\prod_{j\in P^*} x(0)_j\right)\left(\prod_{j\in Z^*} s(0)_j\right).$$

Therefore,

$$x(0)_{P^*} = x^*_{P^*} \quad \text{and} \quad s(0)_{Z^*} = s^*_{Z^*},$$

since $x^*_{P^*}$ and $s^*_{Z^*}$ are the unique maximizer pair of the barrier function.

$\square$

We usually define a neighborhood of the central path as

$$\mathcal{N}(\beta) = \left\{ (x, y, s) \in \overset{\circ}{\mathcal{F}}\colon \ \|Xs - \mu e\| \leq \beta\mu \quad \text{and} \quad \mu = \frac{x^T s}{n} \right\},$$

where $\|.\|$ can be any norm, or even a one-sided "norm" as

$$\|x\|_{-\infty} = |\min(0, \min(x))|.$$

We have the following theorem

Theorem 2.15 *Let* $(x, y, s) \in \mathcal{N}(\beta)$.

i) *The* $\mathcal{N}(\beta)$ *is bounded where* $0 < \mu \leq \mu^0$ *for any given* $\mu^0 > 0$.

ii) *Any limit point of* $\mathcal{N}(\beta)$ *is an optimal solution pair for (LP) and (LD). Moreover, for any* $j \in P^*$

$$x_j \geq \frac{(1-\beta)x_j^*}{n}$$

where x^* *is any optimal primal solution; for any* $j \in Z^*$

$$s_j \geq \frac{(1-\beta)s_j^*}{n},$$

where s^* *is any optimal dual solution.*

2.6 Notes

General convex problems, such as membership, separation, validity, and optimization, can be solved by the center-section method, see Grötschel, Lovász and Schrijver [133].

Levin [197] and Newman [255] considered the center of gravity of a convex body, Elzinga and Moore [87] considered the center of the max-volume sphere contained in a convex body, a number of Russian mathematicians (for example, Tarasov, Khachiyan and Érlikh [319]) considered the center of the max-volume ellipsoid inscribing the body, Huard and Liêu [152, 153] considered a generic center in the body that maximizes a distance function, and Vaidya [349] considered the volumetric center, the maximizer of the determinant of the Hessian matrix of the logarithmic barrier function. See Kaiser, Morin and Trafalis [167] for a complete survey.

Grünbaum [135] first proved Theorem 2.1, with a more geometric proof given by Mityagin [227]. Dyer and Frieze [82] proved that computing the volume of a convex polytope, either given as a list of facets or vertices, is itself $\#P$-Hard. Furthermore, Elekes [86] has shown that no polynomial time algorithm can compute the volume of a convex body with less than exponential relative error. Bárány and Fürendi [34] further showed that for $\Omega \in \mathcal{R}^m$, any polynomial time algorithm that gives an upper and lower bound on the volume of Ω, represented as $\overline{V}(\Omega)$ and $\underline{V}(\Omega)$, respectively, necessarily has an exponential gap between them. They showed

$$\overline{V}(\Omega)/\underline{V}(\Omega) \geq (cm/\log m)^m,$$

where c is a constant independent of m. In other words, there is no polynomial time algorithm that would compute $\overline{V}(\Omega)$ and $\underline{V}(\Omega)$ such that

$$\overline{V}(\Omega)/\underline{V}(\Omega) < (cm/\log m)^m.$$

Recently, Dyer, Frieze and Kannan [83] developed a random polynomial time algorithm that can, with high probability, find a good estimate for the volume of Ω.

Apparently, the result that every convex body contains a unique ellipsoid of maximal volume and is contained in a unique ellipsoid of minimal volume, was discovered independently by several mathematicians—see for example, Danzer, Grünbaum and Klee [74]. These authors attributed the first proof to K. Löwner. John [165] later proved Theorem 2.2.

Tarasov, Khachiyan, and Érlikh [319] proved the center-section Theorem 2.3. Khachiyan and Todd [177] established a polynomial complexity bound for computing an approximate point of the center of the maximal inscribing ellipsoid if the convex body is represented by linear inequalities, Theorem 2.5 was proved by Shor [303] and Nemirovskii and Yudin [258].

The "analytic center" for a convex polyhedron given by linear inequalities was introduced by Huard [152], and later by Sonnevend [306]. The logarithmic distance function, called the barrier function, was introduced by Frisch [99]. Theorem 2.6 was first proved by Sonnevend [306], also see Karmarkar [173] for a canonical form.

Todd [321] and Ye [372] proved that Karmarkar's potential function represents the logarithmic volume of a regular ellipsoid who contains the feasible region. The Karmarkar potential function in the standard form (LP) with a lower bound z for z^* was seen in Todd and Burrell [329], Anstreicher [19], Gay [103], and Ye and Kojima [384]. The primal potential function with $\rho > 1$ was proposed by Gonzaga [124], Freund [96], and Ye [373, 375]. The primal-dual potential function was proposed by Tanabe [316], and Todd and Ye [330]. Noma [268] proved Theorem 2.13.

McLinden [212] earlier, then Bayer and Lagarias [37, 38], Megiddo [216], and Sonnevend [306] analyzed the central path for linear programming and convex optimization. Megiddo [216] derived the central path simply minimizing the primal with a logarithmic barrier function as in Fiacco and McCormick [92]. McLinden [212] proved Theorem 2.14.

2.7 Exercises

2.1 *Find the min-volume ellipsoid containing a half of the unit ball.*

2.2 *Verify Examples 2.1 and 2.2.*

2.3 *Prove Example 2.3.*

2.4 *Let* $\overset{\circ}{\Omega} = \{y \in \mathcal{R}^m : c - A^T y > 0\} \neq \emptyset$, $\overset{\circ}{\Omega}{}' = \{y \in \mathcal{R}^m : c' - A^T y > 0\} \neq \emptyset$, *and* $c' \leq c$. *Then* $\mathcal{B}(\Omega') \leq \mathcal{B}(\Omega)$.

2.5 *Prove Corollary 2.8.*

2.6 *Prove Corollary 2.9.*

2.7 *Prove Corollary 2.11.*

2.8 *Consider the problem*

$$\begin{array}{ll} \textit{maximize} & f(x) = \|x - e\| \prod_{j=1}^n x_j \\ \textit{subject to} & e^T x = n, \ x \geq 0 \in \mathcal{R}^n. \end{array}$$

Prove that $x_1 = \beta > 1$ *and* $x_2 = \ldots = x_n = (n - \beta)/(n - 1) > 0$ *is a maximizer.*

2.9 *If* $\Omega = \{y : \ c - A^T y \geq 0\}$ *is nonempty, then the minimal value of the primal problem described at the beginning of Section s2:3 is* 0*; if* Ω *is bounded and has an interior, then the interior of* $\mathcal{X}_\Omega := \{x \in \mathcal{R}^n : \ Ax = 0, \ x \geq 0\}$ *is nonempty and* $x = 0$ *is the unique primal solution.*

2.10 *Prove (ii) of Theorem 2.13.*

2.11 *Prove (ii) of Theorem 2.14.*

2.12 *Prove Theorem 2.15.*

Chapter 3

Computation of Analytic Center

As we mentioned in the preceding chapter, a favorable property of the analytic center is that it is relatively easy to compute. In this chapter, we discuss how to compute the analytic center using the dual, primal, and primal-dual algorithms in three situations: 1) from an approximate analytic center, 2) from an interior-point, and 3) from an exterior point.

3.1 Proximity to Analytic Center

Before we introduce numerical procedures to compute it, we need to discuss how to measure proximity to the analytic center. Recall that Ω is a bounded polytope in $\mathcal{R}^m$ defined by n $(> m)$ linear inequalities, i.e.,

$$\Omega = \{y \in \mathcal{R}^m : \ c - A^T y \geq 0\}.$$

For a point $y \in \overset{\circ}{\Omega}$, denote the barrier or potential function of Ω

$$\mathcal{B}(y, \Omega) = \sum_{j=1}^{n} \log s_j, \quad s = c - A^T y,$$

simply by $\mathcal{B}(y)$ in this section. Ideally, a measure of closeness of $y \in \Omega$ to the analytic center y^a would be

$$\mathcal{B}(y^a) - \mathcal{B}(y) = \left(\max_{y \in \Omega} \mathcal{B}(y)\right) - \mathcal{B}(y).$$

The problem is that we have no knowledge of y^a or $\mathcal{B}(y^a)$.

Since y^a is the maximizer of $\mathcal{B}(y)$, one measure would be the residual of the optimality condition at y. Note that

$$\nabla \mathcal{B}(y) = -AS^{-1}e \quad \text{and} \quad \nabla^2 \mathcal{B}(y) = -AS^{-2}A^T, \quad \text{where} \quad s = c - A^T y > 0.$$

Thus, the optimality condition is

$$\nabla \mathcal{B}(y) = -AS^{-1}e = 0.$$

Consider a normalized gradient vector of $\mathcal{B}$ at y and $s = c - A^T y$:

$$\begin{array}{rcl} p(s) & := & S^{-1}A^T(AS^{-2}A^T)^{-1}\nabla \mathcal{B}(y) \\ & = & -S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1}e, \end{array} \tag{3.1}$$

and

$$\begin{array}{rcl} \delta(s)^2 & := & \|p(s)\|^2 \\ & = & e^T S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1}e. \end{array} \tag{3.2}$$

Note that $\delta(s) = 0$ implies $\nabla \mathcal{B}(y) = 0$. Let

$$x(s) = S^{-1}(I - S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1})e. \tag{3.3}$$

Then, we have

$$Ax(s) = 0,$$

$$p(s) = Sx(s) - e \quad \text{and} \quad \delta(s) = \|Sx(s) - e\|.$$

Thus, if $\delta(s) = 0$, then $y = y^a$ and $s = s^a$, and also $x(s^a) = x^a$ which is the minimizer of

$$\begin{array}{ll} \text{minimize} & \mathcal{P}(x, \Omega) \\ \text{subject to} & x \in \overset{\circ}{\mathcal{X}}_\Omega = \{x : \ Ax = 0, \ x > 0\}, \end{array}$$

where the primal (homogeneous) potential function

$$\mathcal{P}(x, \Omega) = n \log(c^T x) - \sum_{j=1}^{n} \log x_j.$$

As we discussed earlier, this quantity represents the logarithmic volume of a regular ellipsoid that contains $\mathcal{S}_\Omega$, and the minimization of the potential function results in the analytic center of Ω. In other words, x and y are the analytic center pair for Ω. Denote $\mathcal{P}(x, \Omega)$ simply by $\mathcal{P}(x)$ in this section.

Another measure is to use the Newton direction of the primal potential function. Note that $\mathcal{P}(x)$ is homogeneous so that we can fix $c^T x = n$. Then

$$\nabla \mathcal{P}(x) = \frac{n}{c^T x} c - X^{-1}e = c - X^{-1}e.$$

Consider a scaled projection of $c - X^{-1}e$:

$$\begin{aligned} p(x) &:= (I - XA^T(AX^2A^T)^{-1}AX)X(c - X^{-1}e) \\ &= (I - XA^T(AX^2A^T)^{-1}AX)(Xc - e), \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} \delta(x)^2 &:= \|p(x)\|^2 \\ &= (Xc - e)^T(I - XA^T(AX^2A^T)^{-1}AX)(Xc - e). \end{aligned} \tag{3.5}$$

Let

$$y(x) = (AX^2A^T)^{-1}AX(Xc - e) \quad \text{and} \quad s(x) = c - A^Ty(x). \tag{3.6}$$

Then, we have

$$p(x) = Xs(x) - e \quad \text{and} \quad \delta(x) = \|Xs(x) - e\|.$$

Again, if $\delta(x) = 0$, $x = x^a$ and $y(x^a) = y^a$, which are the analytic center pair for Ω.

The third measure is to use both the primal and dual. For an $x \in \mathcal{X}_\Omega = \{x : \ Ax = 0, \ x \geq 0\}$ and a $y \in \Omega$ or $s = c - A^Ty \in \mathcal{S}_\Omega = \{s : \ s = c - A^Ty, \ s \geq 0\}$, the measure would be defined as

$$\eta(x, s) := \|X(c - A^Ty) - e\| = \|Xs - e\|. \tag{3.7}$$

Obviously, if $\eta(x, s) = 0$, $x = x^a$, $y = y^a$, and $\psi_n(x, s) = n \log n$, where the primal-dual potential function is

$$\psi_n(x, s) = n \log(x^Ts) - \sum_{j=1}^n \log(x_js_j).$$

To find relations among these measures, we first present a lemma whose proof is omitted.

Lemma 3.1 *If $d \in \mathcal{R}^n$ such that $\|d\|_\infty < 1$ then*

$$e^Td \geq \sum_{i=1}^n \log(1 + d_i) \geq e^Td - \frac{\|d\|^2}{2(1 - \|d\|_\infty)}.$$

We have the following theorem to equalize these measures.

Theorem 3.2 *Let (y, s) be an interior point and (y^a, s^a) be the analytic center of Ω or $\mathcal{S}_\Omega$, and let x be an interior point of $\mathcal{X}_\Omega$ and x^a be the primal potential minimizer with $c^Tx^a = n$.*

i) *If* $\eta(x,s) < 1$ *then*

$$\delta(s) \le \eta(x,s)$$

and

$$\delta(x) \le \eta(x,s).$$

Conversely, if $\delta(s) < 1$ *then there is an* $x(s) \in \overset{\circ}{\mathcal{X}}_\Omega$ *such that*

$$\eta(x(s),s) \le \delta(s),$$

and there is a $s(x) \in \overset{\circ}{\mathcal{S}}_\Omega$ *such that*

$$\eta(x,s(x)) \le \delta(x).$$

ii) *If* $\eta(x,s) < 1$*, then there is a* $\hat{x} \ge 0$ *with* $A\hat{x} = 0$ *and* $c^T\hat{x} = s^T\hat{x} = n$ *such that*

$$\eta(\hat{x},s) \le \frac{\eta(x,s)}{\sqrt{1-\eta(x,s)^2/n}}.$$

iii) *If* $\eta(x,s) < 1$ *with* $c^Tx = s^Tx = n$*, then*

$$\psi_n(x,s) - \psi_n(x^a,s^a) \le \frac{\eta(x,s)^2}{2(1-\eta(x,s))},$$

$$\mathcal{P}(x) - \mathcal{P}(x^a) \le \frac{\eta(x,s)^2}{2(1-\eta(x,s))},$$

and

$$\mathcal{B}(y^a) - \mathcal{B}(y) \le \frac{\eta(x,s)^2}{2(1-\eta(x,s))}.$$

iv) *Let the dual Newton procedure be:*

$$d_y = -(AS^{-2}A^T)^{-1}AS^{-1}e, \quad y^+ = y + d_y, \quad \textit{and} \quad s^+ = c - A^Ty^+. \tag{3.8}$$

Then, if $\delta(s) < 1$*,*

$$s^+ > 0 \quad \textit{and} \quad \delta(s^+) \le \delta(s)^2.$$

Let the primal Newton procedure be:

$$d_x = X(I - XA^T(AX^2A^T)^{-1}AX)(Xc - e) \quad \textit{and} \quad x^+ = x + d_x. \tag{3.9}$$

Then, if $\delta(x) < 1$*,*

$$x^+ > 0, \quad Ax^+ = 0, \quad \textit{and} \quad \delta(x^+) \le \delta(x)^2.$$

Let the primal-dual Newton procedure be: Solve for d_x, d_y and d_s from the system of linear equations

$$\begin{array}{rcl} Sd_x + Xd_s & = & e - Xs, \\ Ad_x & = & 0, \\ A^T d_y + d_s & = & 0, \end{array} \tag{3.10}$$

and assign

$$x^+ = x + d_x, \quad y^+ = y + d_y, \quad s^+ = s + d_s.$$

Then, if $\eta(x,s) < 1$,

$$x^+ > 0, \quad Ax^+ = 0, \quad s^+ > 0$$

and

$$\eta(x^+, s^+) \le \frac{\sqrt{2}\eta(x,s)^2}{4(1-\eta(x,s))}.$$

v) *If $\eta(x,s) < 1$, then*

$$\|S^{-1}s^a - e\| \le \frac{\eta(x,s)}{1-\eta(x,s)}$$

and

$$\|X^{-1}x^a - e\| \le \frac{\eta(x,s)}{1-\eta(x,s)}.$$

Proof. i) Given $s > 0$ we can verify that $\delta(s)$ is the minimal value and $x(s)$ is the minimizer of the least squares problem

$$\begin{array}{ll} \text{minimize} & \|Sx - e\| \\ \text{subject to} & Ax = 0. \end{array}$$

Since x in $\eta(x,s)$ is a feasible point for this problem, we must have

$$\delta(s) = \eta(x(s), s) \le \eta(x,s).$$

Conversely, $x = x(s)$ will do it.

Similarly, given $x > 0$ we can verify that $\delta(x)$ is the minimal value and $(y(x), s(x))$ is the minimizer of the least squares problem

$$\begin{array}{ll} \text{minimize} & \|Xs - e\| \\ \text{subject to} & s = c - A^T y. \end{array}$$

Since (y, s) in $\eta(x, s)$ is any point for this problem, we must have

$$\delta(x) = \eta(x, s(x)) \leq \eta(x, s).$$

Conversely, $y = y(x)$ and $s = s(x)$ will do it.

ii) Let $\hat{x} = (n/x(s)^T s)x(s)$ and $\eta(\hat{x}, s) = \|S\hat{x} - e\|$. Then

$$A\hat{x} = 0 \quad \text{and} \quad c^T\hat{x} = s^T\hat{x} = n.$$

Furthermore,

$$\begin{aligned}
\delta(s)^2 &= \|Sx(s) - e\|^2 \\
&= \|Sx(s) - (x(s)^T s/n)e + (x(s)^T s/n)e - e\|^2 \\
&= \|Sx(s) - (x(s)^T s/n)e\|^2 + \|(x(s)^T s/n)e - e\|^2 \\
&= \|S\hat{x} - e\|^2 (x(s)^T s/n)^2 + (1 - x(s)^T s/n)^2 n \\
&= \eta(\hat{x}, s)^2 (x(s)^T s/n)^2 + (1 - x(s)^T s/n)^2 n.
\end{aligned}$$

Thus, we have

$$(x(s)^T s/n)^2(\eta(\hat{x}, s)^2 + n) - 2(x(s)^T s/n)n + n - \delta(s)^2 = 0.$$

Consider this relation as a quadratic equation with variable $x(s)^T s/n$. Since it has a real root, we have

$$4n^2 - 4(\eta(\hat{x}, s)^2 + n)(n - \delta(s)^2) \geq 0$$

or

$$\eta(\hat{x}, s)^2 \leq \frac{n\delta(s)^2}{n - \delta(s)^2},$$

which gives the desired result.

iii) Let $\eta = \eta(x, s) < 1$. From Lemma 3.1 and $c^T x = s^T x = n$,

$$\sum_{j=1}^{n} \ln(x_j s_j) \geq e^T X s - n - \frac{\eta^2}{2(1-\eta)} = -\frac{\eta^2}{2(1-\eta)}.$$

Denote by x^a and y^a $(s^a = c - A^T y^a)$ the center pair of Ω. Noting that $X^a s^a = e$, we have

$$\psi_n(x, s) - \psi_n(x^a, s^a) = \sum_{j=1}^{n} \ln(x_j^a s_j^a) - \sum_{j=1}^{n} \ln(x_j s_j) \leq \frac{\eta^2}{2(1-\eta)}. \tag{3.11}$$

The left-hand side of (3.11) can be written as

$$\sum_{j=1}^{n} \ln(x_j^a) - \sum_{j=1}^{n} \ln(x_j) + \sum_{j=1}^{n} \ln(s_j^a) - \sum_{j=1}^{n} \ln(s_j) = \mathcal{P}(x) - \mathcal{P}(x^a) + \mathcal{B}(y^a) - \mathcal{B}(y).$$

Since y^a maximizes $\mathcal{B}(y)$ over the interior of Ω and x^a minimizes $\mathcal{P}(x)$ over the interior of $\mathcal{X}$, we have

$$\mathcal{B}(y^a) - \mathcal{B}(y) \geq 0$$

and

$$\mathcal{P}(x) - \mathcal{P}(x^a) \geq 0.$$

Thus, we have the desired result.

iv) Note that from the proof of (i)

$$\delta(s^+) = \|S^+ x(s^+) - e\| \leq \|S^+ x(s) - e\|.$$

But

$$\begin{aligned} \|S^+ x(s) - e\|^2 &= \|(2S - S^2 x(s))x(s) - e\|^2 \\ &= \sum_{j=1}^{n} (s_j x(s)_j - 1)^4 \\ &\leq (\sum_{j=1}^{n} (s_j x(s)_j - 1)^2)^2 \\ &= \|Sx(s) - e\|^4 \\ &= \delta(s)^4. \end{aligned}$$

Similarly, we can prove the primal procedure result.

To prove the primal-dual procedure result we first see that

$$\|X^+ s^+ - e\| = \|D_x d_s\|.$$

Multiplying the both sides of the first equation of (3.10) by $(XS)^{-1/2}$, we see

$$Dd_x + D^{-1} d_s = r := (XS)^{-1/2}(e - Xs),$$

where $D = S^{1/2} X^{-1/2}$. Let $p = Dd_x$ and $q = D^{-1} d_s$. Note that $p^T q = d_x^T d_s \geq 0$. Then,

$$\begin{aligned} \|D_x d_s\|^2 &= \|Pq\|^2 \\ &= \sum_{j=1}^{n} (p_j q_j)^2 \\ &\leq \left(\sum_{p_j q_j > 0}^{n} p_j q_j \right)^2 + \left(\sum_{p_j q_j < 0} p_j q_j \right)^2 \end{aligned}$$

$$
\begin{aligned}
&\le 2\left(\sum_{p_jq_j>0}^{n} p_jq_j\right)^2 \\
&\le 2\left(\sum_{p_jq_j>0}^{n} (p_j+q_j)^2/4\right)^2 \\
&\le 2\left(\|r\|^2/4\right)^2 .
\end{aligned}
$$

Furthermore,

$$
\|r\|^2 \le \|(XS)^{-1/2}\|^2\|e-Xs\|^2 \le \frac{\eta^2}{(1-\eta)},
$$

which gives the desired result.

v) From (i), $\delta(s) \le \eta(x,s)$. Let $y^0 = y$, $s^0 = s$, and the sequence $\{y^k, s^k\}$ be generated by the Newton method. Then, we show by induction that

$$
\|(S^0)^{-1}s^k - e\| \le \sum_{j=1}^{2^k-1} \delta(s)^j .
$$

Obviously, this relation is true for $k=1$ by the definition of $\delta(s^0) = \delta(s)$. Now assuming it is true for k. Since we have

$$
s^{k+1} = s^k - S^k(S^kx(s^k) - e),
$$

$$
\begin{aligned}
\|(S^0)^{-1}(s^{k+1}-s^k)\| &= \|(S^0)^{-1}S^k(S^kx(s^k)-e)\| \\
&\le \|(S^0)^{-1}S^k\|\|(S^kx(s^k)-e)\| \\
&\le \left(1+\sum_{j=1}^{2^k-1}\delta(s)^j\right)\|(S^kx(s^k)-e)\| \\
&\le \left(1+\sum_{j=1}^{2^k-1}\delta(s)^j\right)\delta(s)^{2^k} \\
&= \left(\sum_{j=0}^{2^k-1}\delta(s)^j\right)\delta(s)^{2^k} .
\end{aligned}
$$

Thus,

$$
\|(S^0)^{-1}s^{k+1}-e\| \quad\le\quad \|(S^0)^{-1}s^k-e\| + \|(S^0)^{-1}(s^{k+1}-s^k)\|
$$

$$
\begin{aligned}
&\le \left(\sum_{j=1}^{2^k-1}\delta(s)^j\right) + \left(\sum_{j=0}^{2^k-1}\delta(s)^j\right)\delta(s)^{2^k} \\
&= \sum_{j=1}^{2^{k+1}-1}\delta(s)^j.
\end{aligned}
$$

Similarly, we can prove the primal result. This proves (v).

□

3.2 Dual Algorithms

3.2.1 Dual Newton procedure

Given $y \in \Omega$, we call it an (η-)approximate (analytic) center if $\delta(s) \le \eta < 1$. The dual Newton procedure in (iv) of Theorem 3.2, once a y satisfies $\delta(s) \le \eta < 1$, will generate a sequence $\{y^k\}$ that converges to y^a quadratically.

The next question is how to generate an (η-)approximate center with $\eta < 1$.

3.2.2 Dual potential algorithm

Let y be any interior point in Ω. Let us apply the Newton method with a controlled step-size, which is equivalent to solving a ball-constrained linear problem with radius $\alpha < 1$:

$$
\begin{array}{ll}
\text{maximize} & \nabla\mathcal{B}(y)^T d_y \\
\text{subject to} & d_y^T(-\nabla^2\mathcal{B}(y))d_y^T \le \alpha^2.
\end{array}
$$

or

$$
\begin{array}{ll}
\text{maximize} & -e^T S^{-1}A^T d_y \\
\text{subject to} & d_y^T AS^{-2}A^T d_y^T \le \alpha^2.
\end{array}
$$

Note that

$$
s^+ = c - A^T y^+ = s - A^T d_y = S(e - S^{-1}A^T d_y),
$$

thus we must have $s^+ > 0$, i.e., $y^+ = y + d_y$ remains in the interior of Ω. Note that

$$
d_y = -\alpha\frac{(AS^{-2}A^T)^{-1}AS^{-1}e}{\sqrt{e^T S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1}e}}.
$$

Recall from (3.1) and (3.3)

$$
p(s) = -S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1}e = Sx(s) - e.
$$

Then,

$$d_y = -\alpha \frac{(AS^{-2}A^T)^{-1}AS^{-1}e}{\|p(s)\|},$$

$$d_s = -A^T d_y = -\alpha Sp(s)/\|p(s)\|,$$

and

$$e^T S^{-1} d_s = \alpha \|p(s)\|.$$

Furthermore, $s^+ = s + d_s = s - \alpha Sp(s)/\|p(s)\| = S(e - \alpha p(s)/\|p(s)\|)$, i.e., $\|S^{-1}s^+ - e\| = \alpha$. Thus, from Lemma 3.1 we have

$$\begin{aligned} \mathcal{B}(y^+) - \mathcal{B}(y) &\geq e^T S^{-1}(s^+ - s) - \frac{\|S^{-1}(s^+ - s)\|^2}{2(1 - \|S^{-1}(s^+ - s)\|_\infty)} \\ &\geq \alpha \|p(s)\| - \frac{\alpha^2}{2(1-\alpha)}. \end{aligned} \tag{3.12}$$

Thus, if $\|p(s)\| \geq \alpha$, we have

$$\mathcal{B}(y^+) - \mathcal{B}(y) \geq \alpha^2 - \frac{\alpha^2}{2(1-\alpha)}.$$

Hence as long as $\alpha < 1/2$ we have

$$\mathcal{B}(y^+) - \mathcal{B}(y) \geq \delta,$$

where constant

$$\delta = \alpha^2 - \frac{\alpha^2}{2(1-\alpha)} > 0.$$

In other words, the potential function is increased by a constant. Note that the potential function is bounded above by the max-potential $\mathcal{B}(y^a)$. Thus, in a finite time we must have $\delta(s) = \|p(s)\| < \alpha$, which implies that the quadratic-convergence condition is satisfied or y is an approximate center. The total number of iterations to reach this condition must be bounded by $O(\mathcal{B}(y^a) - \mathcal{B}(y))$.

3.2.3 Center-section algorithm

The next question is how to compute the analytic center if an interior point y is not known. This can be done with a center-section method. Consider the following set

$$\Omega(\hat{c}) = \{y \in \mathcal{R}^m : A^T y \leq \hat{c}\}.$$

Obviously, $\Omega(c) = \Omega$, whose interior is assumed nonempty and bounded. Then, for any given $\hat{c} \geq c$, $\overset{\circ}{\Omega}(\hat{c})$ is also nonempty and bounded. Moreover,

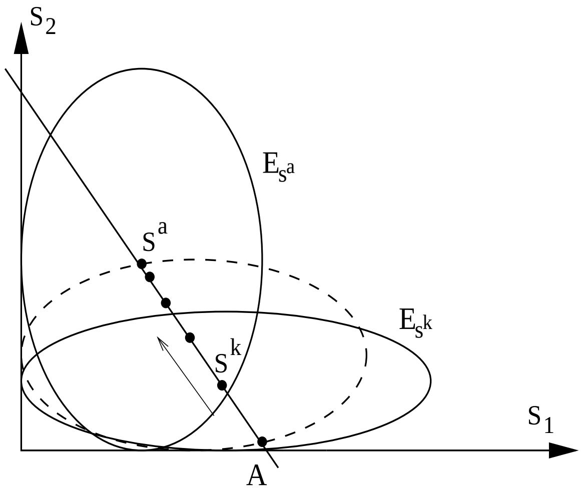

Figure 3.1: Illustration of the dual potential algorithm; it generates a sequence of contained regular ellipsoids whose volumes increase.

$y = 0$ is an interior point in $\Omega(\hat{c})$ if $\hat{c} > 0$. Let us choose c^0 such that $c^0 \geq e$ and $c^0 \geq c$. Then from the result in the preceding section we will generate an approximate center for $\Omega^0 := \Omega(c^0)$ in $O(\mathcal{B}(\Omega^0) - \mathcal{B}(0, \Omega^0))$ iterations. Let y^0 be an approximate center for Ω^0. From now on, we will generate a sequence of $\{c^k\}$ and $\{y^k\}$, where $c \leq c^{k+1} \leq c^k$ and y^k is an approximate center for $\Omega^k := \Omega(c^k)$. Moreover,

$$\mathcal{B}(\Omega) \leq \mathcal{B}(\Omega^{k+1}) \leq \mathcal{B}(\Omega^k) - \delta,$$

where δ is a positive constant, until $c^{k+1} = c$. This process terminates with y^{k+1} as an approximate center for $\Omega = \Omega(c)$.

We first describe a conceptual center-section algorithm.

Algorithm 3.1 *(***Conceptual Algorithm***) Let (y^0, s^0) be the analytic center of $\Omega^0 = \Omega(c^0)$ and let β be a constant in $(0, 1)$. Set $k := 0$.*

While $c^k \neq c$ **do***:*

1. *Translating Inequality: Find i such that $c_i^k > c_i$ and update*

$$c_i^{k+1} = \max\{c_i, \beta(c_i^k - a_i^T y^k) + a_i^T y^k\} = \max\{c_i, \beta s_i^k + a_i^T y^k\},$$

$$c_j^{k+1} = c_j^k \quad \textit{for} \quad j \neq i.$$

Then, from the center-section theorem in the preceding chapter, we have either

$$\mathcal{B}(\Omega^{k+1}) \leq \mathcal{B}(\Omega^k) - (1-\beta) \quad \text{if} \quad c_i^{k+1} > c_i; \tag{3.13}$$

or

$$\mathcal{B}(\Omega^{k+1}) < \mathcal{B}(\Omega^k) \quad \text{if} \quad c_i^{k+1} = c_i. \tag{3.14}$$

Note that the later case can happen only n times.

2. *Updating Center: Compute the center y^{k+1} of Ω^{k+1}, using Newton's method from y^k which is an approximate center of Ω^{k+1}.*

3. *Set $k := k+1$ and return to Step 1.*

Clearly, the center-section algorithm will stop after $O(\mathcal{B}(\Omega^0) - \mathcal{B}(\Omega)) + n$ iterations. If

$$\mathcal{B}(\Omega^0) \leq n \log R \quad \text{and} \quad \mathcal{B}(\Omega) \geq n \log r,$$

then, $O(n\log(R/r)) + n$ iterations suffice.

Numerically, we will never be able to compute the exact analytic center. We must use approximate centers instead of perfect centers in the center-section algorithm. We discuss this issue now.

Algorithm 3.2 *(***Using Approximate Centers***) Let (y^0, s^0) be an approximate analytic center of $\Omega^0 = \Omega(c^0)$, with $\delta(s^0) \leq \eta < 1$, and let β be a constant in $(0,1)$ such that $\eta + (1-\beta)(1+\eta) < 1$. Set $k := 0$.*

While $c^k \neq c$ **do***:*

1. *Translating Inequality: Find i such that $c_i^k > c_i$ and update*

$$c_i^{k+1} = \max\{c_i, \beta s_i^k + a_i^T y^k\},$$

$$c_j^{k+1} = c_j^k \quad \text{for} \quad j \neq i.$$

2. *Updating Approximate Center: Compute an approximate analytic center y^{k+1} of Ω^{k+1} so that $\delta(s^{k+1}) \leq \eta$, using one of the Newton procedures in Theorem 3.2 starting from y^k, which is an approximate center of Ω^{k+1}.*

3. *Set $k := k+1$ and return to Step 1.*

To show that y^k is an approximate center of Ω^{k+1}, we prove the following Lemma.

Lemma 3.3 *There exists a point $x^+ > 0$ such that*

$$Ax^+ = 0 \quad and \quad \|X^+ s^+ - e\| \le \eta + (1-\beta)(1+\eta) < 1,$$

where $s^+ = c^{k+1} - A^T y^k$.

Proof. Let $x^+ = x(s^k) > 0$ with $s^k = c^k - A^T y^k$. Then,

$$Ax^+ = Ax(s^k) = 0$$

and

$$\|x^+ s^k - e\| = \delta(s^k) \le \eta.$$

Note that $s_j^+ = s_j^k$ for $j \ne i$ and $s_i^k \ge s_i^+ \ge \beta s_i^k$. Thus,

$$\begin{aligned}
\|X^+ s^+ - e\| &= \|X^+ s^k - e + X^+ s^+ - X^+ s^k\| \\
&\le \|X^+ s^k - e\| + \|X^+ s^+ - X^+ s^k\| \\
&\le \|X^+ s^k - e\| + |x_i^+ s_i^k (\beta - 1)| \\
&\le \eta + (1-\beta)(1+\eta).
\end{aligned}$$

□

Lemma 3.3 shows that, after an inequality is translated, y^k is still in the "quadratic convergence" region of the center of Ω^{k+1}, because we choose $\eta + (1-\beta)(1+\eta) < 1$. Thus, a closer approximate center, y^{k+1} with $\delta(s^{k+1}) \le \eta$ for Ω^{k+1}, can be updated from y^k in a constant number of Newton's steps. We now verify that the potential function is still reduced by a constant for a small η after a translation.

Lemma 3.4 *Let (y^k, s^k) be an approximate center for Ω^k with $\delta(s^k) \le \eta$ and let Ω^{k+1} be defined above. Then, if $c_i^{k+1} > c_i$ in the update*

$$\mathcal{B}(\Omega^{k+1}) \le \mathcal{B}(\Omega^k) - \delta \quad \text{for a constant} \quad \delta > 0;$$

otherwise $c_i^{k+1} = c_i$ and

$$\mathcal{B}(\Omega^{k+1}) \le \mathcal{B}(\Omega^k).$$

The later case can happen only n times.

Proof. The proof of the first case is similar to Theorem 2.7 of Chapter 2. Let (y^a, s^a) and (y_+^a, s_+^a) be the centers of Ω^k and Ω^{k+1}, respectively. Note that $A(S^a)^{-1}e = 0$, $(c^k)^T (S^a)^{-1} e = (s^a)^T (S^a)^{-1} e = n$, and

$$s_+^a = c^{k+1} - A^T y_+^a$$

where

$$c_j^{k+1} = c_j^k,\ j \neq i \quad \text{and} \quad c_i^{k+1} = c_i^k - (1-\beta)s_i^k.$$

Note that we still have

$$e^T(S^a)^{-1}s_+^a = e^T(S^a)^{-1}c^{k+1} = n - (1-\beta)(s_i^k/s_i^a) \leq n - (1-\beta)(1-\eta).$$

The last inequality is due to (v) of Theorem 3.2. Therefore,

$$\begin{aligned}
\frac{\exp(\mathcal{B}(\Omega^{k+1}))}{\exp(\mathcal{B}(\Omega^k))} &= \prod_{j=1}^{n} \frac{(s_+^a)_j}{s_j^a} \\
&\leq (\frac{n-(1-\beta)(1-\eta)}{n})^n \\
&\leq \exp(-(1-\beta)(1-\eta)).
\end{aligned}$$

The proof of the later case is straightforward.

□

From the lemma, we can conclude that

Theorem 3.5 *In $O(\mathcal{B}(\Omega^0) - \mathcal{B}(\Omega)) + n$ center-section steps, the algorithm will generate an approximate analytic center for Ω.*

3.3 Primal Algorithms

3.3.1 Primal Newton procedure

Given $x \in \mathcal{X}_\Omega$, we call it an (η-)approximate (analytic) center if $\delta(x) \leq \eta < 1$. The primal Newton procedure in (iv) of Theorem 3.2, once a x satisfies $\delta(x) \leq \eta < 1$, will generate a sequence $\{x^k\}$ that converges to x^a quadratically.

The next question is how to generate an (η-)approximate center with $\eta < 1$.

3.3.2 Primal potential algorithm

Consider the primal potential function

$$\mathcal{P}(x) = \mathcal{P}(x, \Omega) := n \log(c^T x) - \sum_{j=1}^{n} \log x_j,$$

where $x \in \overset{\circ}{\mathcal{X}}_\Omega = \{x : \ Ax = 0,\ x > 0\}$. Again, this quantity represents the logarithmic volume of a regular ellipsoid that contains $\mathcal{S}_\Omega$. We now establish a simple fact about $\mathcal{P}$. Given $x \in \overset{\circ}{\mathcal{R}}{}^n_+$ and $d_x \in \mathcal{R}^n$, let $x^+ = x + d_x$ and $\|X^{-1}d_x\| < 1$. Then, from the concavity of log function we have

$$n\log(c^Tx^+) - n\log(c^Tx) \le \frac{n}{c^Tx}c^T(x^+ - x) = \frac{n}{c^Tx}c^Td_x,$$

and from Lemma 3.1 we have

$$-\sum \log x_j^+ + \sum \log x_j \quad \le \quad -e^TX^{-1}d_x + \frac{\|X^{-1}d_x\|^2}{2(1-\|X^{-1}d_x\|_\infty)}.$$

Thus,

$$\begin{aligned} \mathcal{P}(x^+) - \mathcal{P}(x) &\le \frac{n}{c^Tx}c^Td_x - e^TX^{-1}d_x + \frac{\|X^{-1}d_x\|^2}{2(1-\|X^{-1}d_x\|_\infty)} \\ &= \nabla\mathcal{P}(x)^Td_x + \frac{\|X^{-1}d_x\|^2}{2(1-\|X^{-1}d_x\|_\infty)}. \end{aligned} \tag{3.15}$$

Karmarkar's algorithm

We now describe Karmarkar's algorithm to reduce the potential function. Since the primal potential function is degree-0 homogeneous, we can normalize $e^Tx = n$ and work in the region

$$\mathcal{K}_p = \{x : \ Ax = 0, \ e^Tx = n, \ x \ge 0\}.$$

This is the so-called Karmarkar canonical form. Its related LP canonical problem is given as

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & x \in \mathcal{K}_p. \end{array}$$

Starting from any x^0 in $\overset{\circ}{\mathcal{K}}_x$ we generate a sequence $\{x^k\}$ such that

$$\mathcal{P}(x^{k+1}) \le \mathcal{P}(x^k) - 1/8$$

for $k = 0, 1, 2, ...$ until an approximate center of Ω is generated.

One observation regarding $\mathcal{K}_p$ is that if $Ae = 0$, e is the analytic center of $\mathcal{K}_p$. Unfortunately, in general we may not have $Ae = 0$. However, with a given $x^k > 0$ and $Ax^k = 0$, we may transform the LP problem into

$$\begin{array}{lll} (LP') & \text{minimize} & (c^k)^Tx' \\ & \text{subject to} & x' \in \mathcal{K}_{p'} := \{x' : \ A^kx' = 0, \ e^Tx' = n, \ x' \ge 0\}. \end{array}$$

where

$$c^k = X^kc \quad \text{and} \quad A^k = AX^k.$$

Note that if a pure affine scaling transformation,

$$x' = (X^k)^{-1}x,$$

had been used, the last inequality constraint would become $e^T X^k x' = n$. But as we discussed before, the potential function is homogeneous, so that we can use a projective transformation,

$$x' = T(x) = \frac{n(X^k)^{-1}x}{e^T(X^k)^{-1}x} \quad \text{for} \quad x \in \mathcal{K}_p,$$

whose inverse transformation is

$$x = T^{-1}(x') = \frac{nX^k x'}{e^T X^k x'} \quad \text{for} \quad x' \in \mathcal{K}_{p'}.$$

Obviously, $T(x^k) = e$ is the analytic center for $\mathcal{K}_{p'}$. In other words, Karmarkar transforms x^k into the analytic center of $\mathcal{K}_{p'}$ in (LP').

Note that the potential function for (LP') is

$$\mathcal{P}'(x') = n\log((c^k)^T x') - \sum_{j=1}^{n} \log x'_j.$$

The difference of the potential function values at two points of $\overset{\circ}{\mathcal{K}}_p$ is invariant under this projective transformation, i.e.,

$$\mathcal{P}'(T(x^2)) - \mathcal{P}'(T(x^1)) = \mathcal{P}(x^2) - \mathcal{P}(x^1).$$

To reduce $\mathcal{P}'(x')$, one may reduce its linearized function

$$\nabla\mathcal{P}'(e)^T x' = (\frac{n}{(c^k)^T e}c^k - e)^T x' = (\frac{n}{c^T x^k}c^k - e)^T x' = \frac{n}{c^T x^k}(c^k)^T x' - n.$$

Since $c^T x^k$ and n are fixed, we simply solve the following ball-constrained problem

$$\begin{array}{ll} \text{minimize} & (c^k)^T(x' - e) \\ \text{subject to} & A^k(x' - e) = 0,\ e^T(x' - e) = 0,\ \|x' - e\| \le \alpha. \end{array}$$

According to the discussion in Chapter 1, the solution of the problem is

$$x' - e = -\alpha\frac{p^k}{\|p^k\|},$$

where

$$
\begin{aligned}
p^k &= \left(I - \begin{pmatrix} A^k \\ e \end{pmatrix}^T \begin{pmatrix} A^k(A^k)^T & 0 \\ 0 & n \end{pmatrix}^{-1} \begin{pmatrix} A^k \\ e \end{pmatrix} \right) c^k \\
&= c^k - (A^k)^T y^k - \lambda^k e,
\end{aligned}
$$

where $\lambda^k = (c^k)^T e/n = c^T x^k/n$. Thus

$$p^k = X^k(c - A^T y^k) - (c^T x^k/n)e.$$

Using relation (3.15) we consider the difference of the potential values

$$\mathcal{P}'(x') - \mathcal{P}'(e) \leq \nabla \mathcal{P}'(e)(x' - e) + \frac{\alpha^2}{2(1-\alpha)} = -\alpha \frac{n\|p^k\|}{c^T x^k} + \frac{\alpha^2}{2(1-\alpha)}.$$

Thus, as long as $\|p^k\| \geq \eta \frac{c^T x^k}{n} > 0$, we may choose an appropriate α such that

$$\mathcal{P}'(x') - \mathcal{P}'(e) \leq -\zeta,$$

for a positive constant $\zeta = \alpha\eta - \alpha^2/2(1-\alpha)$. Note we have $\zeta = 1/8$ if $\alpha = 1/2$ and $\eta = 3/4$. Let

$$x^{k+1} = T^{-1}(x').$$

Then,

$$\mathcal{P}(x^{k+1}) - \mathcal{P}(x^k) \leq -\zeta.$$

Thus, in $O(\mathcal{P}(x^0) - \mathcal{P}(x^a))$ iterations, we shall generate a pair (x^k, y^k) such that

$$\|X^k(c - A^T y^k) - (c^T x^k/n)e\| \leq \eta \frac{c^T x^k}{n},$$

or

$$\|\frac{n}{c^T x^k} X^k (c - A^T y^k) - e\| \leq \eta < 1,$$

which indicates that y^k is an approximate analytic center for Ω.

Affine potential algorithm

In this section, we show that we can use the simple affine scaling transformation to achieve the same reduction for $\mathcal{P}$. Given $x^k > 0$ such that $Ax^k = 0$, the gradient vector of the potential function at x^k is

$$\nabla \mathcal{P}(x^k) = \frac{n}{c^T x^k} c - (X^k)^{-1} e.$$

Thus, we directly solve the problem

$$\begin{array}{ll} \text{minimize} & \nabla \mathcal{P}(x^k)^T(x - x^k) \\ \text{subject to} & A(x - x^k) = 0, \ \|(X^k)^{-1}(x - x^k)\| \le \alpha. \end{array}$$

Let the solution be x^{k+1}. Then

$$x^{k+1} - x^k = -\alpha \frac{X^k p^k}{\|p^k\|},$$

and

$$\mathcal{P}(x^{k+1}) - \mathcal{P}(x^k) \le -\alpha \|p^k\| + \frac{\alpha^2}{2(1-\alpha)}$$

where

$$p^k = (I - X^k A^T (A (X^k)^2 A^T)^{-1} A X^k) X^k \nabla \mathcal{P}(x^k) = X^k \nabla \mathcal{P}(x^k) - X^k A^T y^k$$

or

$$p^k = X^k (\frac{n}{c^T x^k} c - A^T y^k) - e,$$

and

$$y^k = (A(X^k)^2 A^T)^{-1} A (X^k)^2 \nabla \mathcal{P}(x^k).$$

If x^k is normalized such that $c^T x^k = n$, then we have

$$p^k = X^k (c - A^T y^k) - e.$$

Thus, as long as $\|p^k\| \ge \eta > 0$, we may choose an appropriate α and use (3.15) to guarantee

$$\mathcal{P}(x^{k+1}) - \mathcal{P}(x^k) \le -\zeta$$

for some positive constant ζ.

Thus, in $O(\mathcal{P}(x^0) - \mathcal{P}(x^a))$ iterations, we shall generate a pair (x^k, y^k) such that

$$\|X^k(c - A^T y^k) - e\| \le \eta,$$

which again indicates that y^k is an approximate analytic center for Ω.

3.4 Primal-Dual (Symmetric) Algorithms

3.4.1 Primal-dual Newton procedure

Clearly, the primal-dual Newton procedure in (iv) of Theorem 3.2, once an approximate center pair pair $x \in \mathcal{X}_\Omega$ and $s \in \mathcal{S}_\Omega$ satisfies $\eta(x,s) \le \eta < 1$, will generate a sequence $\{x^k, y^k, s^k\}$ that converges to (x^a, y^a, s^a) quadratically.

The next question is how to generate such a pair with

$$\eta(x,s) = \|Xs - e\| \le \eta < 1, \tag{3.16}$$

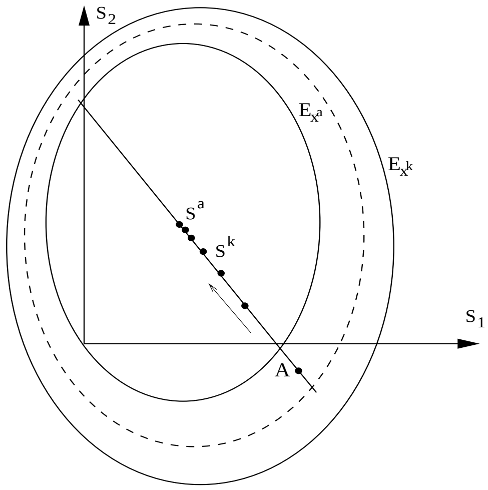

Figure 3.2: Illustration of the primal potential algorithm; it generates a sequence of containing regular ellipsoids whose volumes decrease.

3.4.2 Primal-dual potential algorithm

Given any $x \in \overset{\circ}{\mathcal{X}}_\Omega$ and $s \in \overset{\circ}{\mathcal{S}}_\Omega$ with $s^T x = n$, we show how to use the primal-dual algorithm to generate an approximate analytic center pair using the primal-dual potential function.

Lemma 3.6 *Let the directions (d_x, d_y, d_s) be generated by equation (3.10), and let*

$$\theta = \frac{\alpha\sqrt{\min(Xs)}}{\|(XS)^{-1/2}(e - Xs)\|}, \tag{3.17}$$

where α is a positive constant less than 1 and $\min(v \in \mathcal{R}^n) = \min_j\{v_j|\ j = 1, ..., n\}$. *Then, we have*

$$\psi_n(x + \theta d_x, s + \theta d_s) - \psi_n(x, s)$$

$$\leq -\alpha\sqrt{\min(Xs)}\|(XS)^{-1/2}(e - Xs)\| + \frac{\alpha^2}{2(1-\alpha)}.$$

Proof. It can be verified from the proof of the primal-dual procedure ((iv) of Theorem 3.2) that

$$\tau := \max(\|\theta S^{-1}d_s\|_\infty, \|\theta X^{-1}d_x\|_\infty) < 1.$$

This implies that

$$x^+ := x + \theta d_x > 0 \quad \text{and} \quad s^+ := s + \theta d_s > 0.$$

Then, from Lemma 3.1 and (3.15) we derive

$$\begin{aligned}
&\psi_n(x^+, s^+) - \psi_n(x, s) \\
&\quad \le \quad \theta e^T (X d_s + S d_x) - \theta e^T (S^{-1} d_s + X^{-1} d_x) \\
&\qquad\quad + \frac{\|\theta S^{-1} d_s\|^2 + \|\theta X^{-1} d_x\|^2}{2(1-\tau)}.
\end{aligned}$$

The choice of θ in (3.17) implies that

$$\|\theta S^{-1} d_s\|^2 + \|\theta X^{-1} d_x\|^2 \le \alpha^2.$$

Hence, we have $\tau \le \alpha$ and

$$\frac{\|\theta S^{-1} d_s\|^2 + \|\theta X^{-1} d_x\|^2}{2(1-\tau)} \le \frac{\alpha^2}{2(1-\alpha)}. \tag{3.18}$$

Moreover,

$$\begin{aligned}
&\theta e^T (X d_s + S d_x) - \theta e^T (S^{-1} d_s + X^{-1} d_x) \\
&\quad = \quad \theta \left(e^T (X d_s + S d_x) - e^T (S^{-1} d_s + X^{-1} d_x) \right) \\
&\quad = \quad \theta \left(e^T (X d_s + S d_x) - e^T (XS)^{-1} (X d_s + S d_x) \right) \\
&\quad = \quad \theta (e - (XS)^{-1} e)^T (X d_s + S d_x) \\
&\quad = \quad \theta (e - (XS)^{-1} e)^T (e - Xs) \quad (\text{from } (3.10)) \\
&\quad = \quad -\theta (e - XS)^T (XS)^{-1} (e - XS) \\
&\quad = \quad -\theta \|(XS)^{-1/2} (e - Xs)\|^2 \\
&\quad = \quad -\alpha \sqrt{\min(Xs)} \|(XS)^{-1/2} (e - Xs)\|.
\end{aligned} \tag{3.19}$$

Therefore, we have the desired result combining (3.18) and (3.19).

□

Theorem 3.7 *Let x^+ and s^+ be defined in Lemma 3.6. Then, if $\eta(x, s) \ge \eta$ for a positive constant $\eta < 1$, we can choose α such that*

$$\psi_n(x + \theta d_x, s + \theta d_s) - \psi_n(x, s) \le -\zeta$$

for a positive constant ζ.

Proof. Consider

$$\alpha\sqrt{\min(Xs)}\|(XS)^{-1/2}(e-Xs)\| = \alpha\sqrt{\min(Xs)}\|(XS)^{-1/2}e-(XS)^{1/2}e\|.$$

Let $z=Xs$ and $\min(z)=z_1\le 1$, since $e^Tz=n$. If $z_1\le 1/2$, then

$$\alpha\sqrt{\min(z)}\|Z^{-1/2}e-Z^{1/2}e\|\ge\alpha|1-z_1|\ge\alpha/2.$$

Thus, we can select an α such that

$$\psi_n(x^+,s^+)-\psi_n(x,s)\le -\frac{\alpha}{2}+\frac{\alpha^2}{2(1-\alpha)}=-\zeta$$

for a constant ζ.

Now let $\min(z)\ge 1/2$, that is,

$$\alpha\sqrt{\min(z)}\|Z^{-1/2}e-Z^{1/2}e\|\ge(\alpha/\sqrt{2})\|Z^{-1/2}e-Z^{1/2}e\|.$$

Let $z_n=\max(z)\ge 1$. If $z_n\ge 2$, then

$$(\alpha/\sqrt{2})\|Z^{-1/2}e-Z^{1/2}e\|\ge(\alpha/\sqrt{2})|\sqrt{z_n}-1/\sqrt{z_n}|\ge\alpha/2.$$

Again, we have

$$\psi_n(x^+,s^+)-\psi_n(x,s)\le -\frac{\alpha}{2}+\frac{\alpha^2}{2(1-\alpha)}=-\zeta$$

for a constant ζ.

Now let $\min(z)\ge 1/2$ and $\max(z)\le 2$. Then

$$\alpha\sqrt{\min(z)}\|Z^{-1/2}e-Z^{1/2}e\|=\alpha\sqrt{\min(z)}\|Z^{-1/2}(e-z)\|\ge(\alpha/2)\|e-z\|.$$

Thus, if $\|e-z\|=\|e-Xs\|\ge\eta$, we will have

$$\psi_n(x^+,s^+)-\psi_n(x,s)\le-\zeta$$

for a constant ζ.

□

Thus, until $\|Xs-e\|<\eta<1$, the primal-dual potential function at (x^+,s^+) will be reduced by a constant for some α. Therefore, in $O(\psi_n(x^0,s^0)-n\log n)$ iterations, we shall generate a pair (x,y) such that

$$\eta(x,s)=\|Xs-e\|=\|X(c-A^Ty)-e\|<\eta,$$

which indicates that x and y are an approximate analytic center pair for Ω. Note that this complexity bound depends only on the initial point (x^0,s^0).

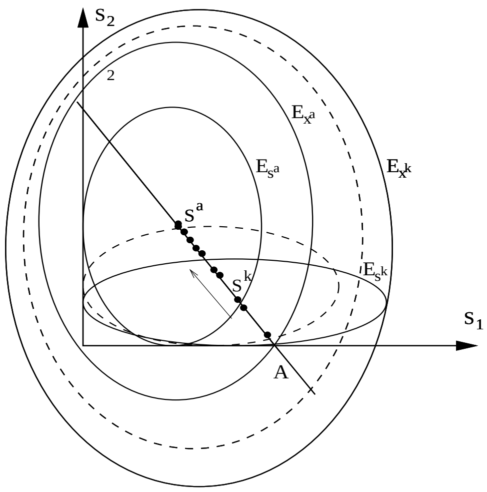

Figure 3.3: Illustration of the primal-dual potential algorithm; it generates a sequence of containing and contained regular ellipsoids whose logarithmic volume-ratio reduces to $n \log n$.

3.5 Notes

The proof of Lemma 3.1 is due to Karmarkar [173] and that of Theorem 3.2 is mostly due to Gonzaga [125, 126], Renegar and Shub [286, 288], Roos and Vial [293], Tseng [337], Vaidya [347], and Goffin et al. [110].

The dual algorithm with a starting interior point, described in this chapter, is similar to the one of Vaidya [347]. The primal or dual affine potential algorithm was proposed by Gonzaga [124], Freund [96], and Ye [373, 375]. The primal-dual algorithm described in this chapter is adapted from the one in Kojima, Mizuno and Yoshise [184, 183, 185] and Monteiro and Adler [239, 240]. The primal-dual procedure result in (iv) of Theorem 3.2 was proved by Mizuno [229].

Finally, we remark the relation among potential reduction algorithms. We provide a simple argument that the primal-dual potential function is also reduced in either the primal or the dual potential reduction algorithm described earlier.

Given x and (y, s) in the interior of $\mathcal{X}_\Omega$ and $\mathcal{S}_\Omega$, respectively. We have shown in the preceding chapter that

$$\psi_n(x, s) = n \log(s^T x) - \sum_{j=1}^{n} \log(x_j s_j) = \mathcal{P}(x) - \mathcal{B}(y).$$

Thus, if we update the dual (y, s) to (y^+, s^+) such that

$$\mathcal{B}(y^+) \geq \mathcal{B}(y) + \zeta,$$

then we must also have

$$\psi_n(x, s^+) \leq \psi_n(x, s) - \zeta;$$

or if we update the primal x to x^+ such that

$$\mathcal{P}(x^+) \leq \mathcal{P}(x) - \zeta,$$

then we must also have

$$\psi_n(x^+, s) \leq \psi_n(x, s) - \zeta.$$

Thus, to make either the primal or dual potential reduction leads to the same reduction in the primal-dual potential function. Therefore, all these algorithms must stop in $O(\psi_n(x^0, s^0) - n \log n)$ iterations. Again, this complexity bound depends only on the initial point (x^0, s^0). Moreover, the primal algorithm does not need knowledge of s^0 and the dual algorithm does not need knowledge of x^0, while the primal-dual algorithm uses both x^0 and s^0.

3.6 Exercises

3.1 *Prove Lemma 3.1.*

3.2 *Given $s > 0$ verify that $x(s)$ is the minimizer of the least squares problem*

$$\begin{array}{ll} \textit{minimize} & \|Sx - e\| \\ \textit{subject to} & Ax = 0. \end{array}$$

Given $x > 0$ verify that $y(x)$ is the minimizer of the least squares problem

$$\begin{array}{ll} \textit{minimize} & \|Xs - e\| \\ \textit{subject to} & s = c - A^T y. \end{array}$$

3.3 *Prove the primal procedure result in (iv) of Theorem 3.2.*

3.4 *Prove the primal inequality in (v) of Theorem 3.2.*

3.5 *Let $e \in \mathcal{K}_p$. Then, prove that*

$$\mathcal{K}_p \subset \{x : \|x - e\| \leq \sqrt{n(n-1)}\}.$$

3.6 *Consider the projective transformation and Karmarkar's potential function. Prove*

$$\mathcal{P}'(T(x^2)) - \mathcal{P}'(T(x^1)) = \mathcal{P}(x^2) - \mathcal{P}(x^1).$$

Chapter 4

Linear Programming Algorithms

In the preceding chapter we have used several interior algorithms to compute an approximate analytic center of a polytope specified by inequalities. The goal of this chapter is to extend these algorithms to solving the standard linear programming problem described in Section 1.3.6.

We assume that both $\overset{\circ}{\mathcal{F}}_p$ and $\overset{\circ}{\mathcal{F}}_d$ are nonempty. Thus, the optimal faces for both (LP) and (LD) are bounded. Furthermore, we assume $b \neq 0$, since otherwise (LD) reduces to finding a feasible point in $\mathcal{F}_d$.

Let z^* denote the optimal value and $\mathcal{F} = \mathcal{F}_p \times \mathcal{F}_d$. In this chapter, we are interested in finding an ϵ approximate solution for the LP problem:

$$c^T x - z^* \leq \epsilon \quad \text{and} \quad z^* - b^T y \leq \epsilon.$$

For simplicity, we assume that a central path pair (x^0, y^0, s^0) with $\mu^0 = (x^0)^T s^0 / n$ is known. We will use it as our initial point throughout this chapter. We first prove the following proposition.

Proposition 4.1 *Consider the dual level set:*

$$\Omega(z) = \{y : \; c - A^T y \geq 0, \; -z + b^T y \geq 0\},$$

where $z^0 := b^T y^0 - \mu^0 \leq z < z^*$*. There exists an* $x(z) \in \mathcal{F}_p$ *such that the max-potential of* $\Omega(z)$

$$\mathcal{B}(\Omega(z)) \geq \mathcal{B}(\Omega(z^0)) + (n+1) \log \frac{c^T x(z) - z}{c^T x^0 - z^0} - n \log 2, \qquad (4.1)$$

and for all $x \in \overset{\circ}{\mathcal{F}}_p$ *the primal potential*

$$\mathcal{P}_{n+1}(x,z) \geq \mathcal{P}_{n+1}(x^0,z^0) + (n+1)\log\frac{c^T x(z) - z}{c^T x^0 - z^0} - n\log 2. \qquad (4.2)$$

Proof. First, recall that the primal potential function associated with polytope $\Omega(z)$ is

$$\mathcal{P}_{n+1}(x,z) = \mathcal{P}(x,\Omega(z)) = (n+1)\log(c^T x - z) - \sum_{j=1}^{n}\log x_j, \quad x \in \overset{\circ}{\mathcal{F}}_p,$$

and the max-potential of $\Omega(z)$ is

$$\mathcal{B}(\Omega(z)) = \max_y \left\{ \sum_{j=1}^{n} \log(c_j - a_j^T y) + \log(b^T y - z) \right\},$$

and we have

$$\mathcal{P}_{n+1}(x,z) - \mathcal{B}(\Omega(z)) \geq (n+1)\log(n+1), \quad \forall\, x \in \overset{\circ}{\mathcal{F}}_p \quad \forall\, z < z^*. \qquad (4.3)$$

$\mathcal{P}_{n+1}(x,z)$ represents the logarithmic volume of a regular ellipsoid containing $\Omega(z)$, and $\mathcal{B}(\Omega(z))$ represents the logarithmic volume of the max-volume regular ellipsoid contained in $\Omega(z)$; see discussion in Chapter 2.

Let $y(z)$ be the analytic center of $\Omega(z)$ and let $s(z) = c - A^T y(z))$. Then, from the central path theory in Section 2.5, there is an $x' \in \mathcal{R}^{n+1}$ such that

$$Ax'[n] - bx'_{n+1} = 0 \text{ and } X'[n]s(z) = e,\ x'_{n+1}(b^T y(z) - z) = 1$$

or $x(z) = x'[n]/x'_{n+1} \in \mathcal{R}^n$ such that

$$Ax(z) = b \text{ and } X(z)s(z) = \mu(z)e,$$

where

$$\mu(z) = \frac{x(z)^T s(z)}{n} = \frac{c^T x(z) - b^T y(z)}{n} = b^T y(z) - z = 1/x'_{n+1}.$$

Here $x'[n]$ denotes the vector of the first n components of $x' \in \mathcal{R}^{n+1}$. Thus,

$$\mathcal{B}(\Omega(z)) \quad = \quad -\sum_{j=1}^{n+1}\log x'_j$$

$$\begin{aligned}
&= (n+1)\log\mu(z) - \sum_{j=1}^{n}\log x(z)_j \\
&= (n+1)\log(n+1)\mu(z) - \sum_{j=1}^{n}\log x(z)_j - (n+1)\log(n+1) \\
&= (n+1)\log(c^Tx(z)-z) - \sum_{j=1}^{n}\log x(z)_j - (n+1)\log(n+1) \\
&= (n+1)\log\frac{c^Tx(z)-z}{c^Tx^0-z^0} - \sum_{j=1}^{n}\log\frac{x(z)_j}{x_j^0} \\
&\quad +\mathcal{P}_{n+1}(x^0,z^0) - (n+1)\log(n+1). \qquad (4.4)
\end{aligned}$$

Note that $x(z)$ is on the central path with $\mu(z)$ and $\mu(z) \le \mu^0$ (see Exercise). Furthermore, we have

$$(x(z)-x^0)^T(s(z)-s^0) = 0 \quad \text{or} \quad x(z)^Ts^0 + s(z)^Tx^0 = n(\mu^0+\mu(z)),$$

which implies that

$$\sum_{j=1}^{n}\left(\frac{x(z)_j}{x_j^0}\mu^0 + \frac{x_j^0}{x(z)_j}\mu(z)\right) = n(\mu^0+\mu(z)),$$

and, therefore,

$$\sum_{j=1}^{n}\left(\frac{x(z)_j}{x_j^0}\right) \le n\left(1+\frac{\mu(z)}{\mu^0}\right) \le 2n. \qquad (4.5)$$

Combining inequalities (4.3), (4.4) and (4.5), we have

$$\begin{aligned}
\mathcal{B}(\Omega(z)) &\ge \mathcal{P}_{n+1}(x^0,z^0) - (n+1)\log(n+1) \\
&\quad +(n+1)\log\frac{c^Tx(z)-z}{c^Tx^0-z^0} - n\log 2 \\
&\ge \mathcal{B}(\Omega(z^0)) + (n+1)\log\frac{c^Tx(z)-z}{c^Tx^0-z^0} - n\log 2.
\end{aligned}$$

Also, for all $x \in \overset{\circ}{\mathcal{F}}_p$ we have from (4.3)

$$\begin{aligned}
\mathcal{P}_{n+1}(x,z) &\ge \mathcal{B}(\Omega(z)) + (n+1)\log(n+1) \\
&\ge \mathcal{P}_{n+1}(x^0,z^0) + (n+1)\log\frac{c^Tx(z)-z}{c^Tx^0-z^0} - n\log 2.
\end{aligned}$$

These inequalities lead to the desired result.

□

The proposition indicates that if the net reduction of max-potential or the primal potential of $\Omega(z)$, where $z = b^T y$ for some $y \in \mathcal{F}_d$, is greater than $(n+1)\log(c^T x^0 - z^0)/\epsilon + n\log 2$, then we have

$$(n+1)\log\frac{c^T x(z) - z}{c^T x^0 - z^0} - n\log 2 \le (n+1)\log\frac{\epsilon}{c^T x^0 - z^0} - n\log 2,$$

which implies that

$$z^* - b^T y = z^* - z \le c^T x(z) - z \le \epsilon$$

and

$$c^T x(z) - z^* \le c^T x(z) - z \le \epsilon,$$

i.e., $x(z)$ and y are ϵ approximate solutions for (LP) and (LD).

In the proposition, $x(z)$ is chosen as the minimizer of the primal potential for $\Omega(z)$. This need not be the case. The proposition holds for any $x(z) \in \overset{\circ}{\mathcal{F}}_p$ that is an approximate minimizer, that is, it satisfies

$$\left\| \begin{pmatrix} X(z)(c - A^T y) \\ b^T y - z \end{pmatrix} - \frac{c^T x(z) - z}{n+1} e \right\| \le \beta \frac{c^T x(z) - z}{n+1} \tag{4.6}$$

for a constant $0 \le \beta < 1$. More specifically, we have

Corollary 4.2 *Consider the dual level set where $z^0 < z < z^*$:*

$$\Omega(z) = \{y : \ c - A^T y \ge 0, \ -z + b^T y \ge 0\}.$$

Let $x(z) \in \overset{\circ}{\mathcal{F}}_p$ satisfy condition (4.6). Then

$$\mathcal{B}(\Omega(z)) \ge \mathcal{B}(\Omega(z^0)) + (n+1)\log\frac{c^T x(z) - z}{c^T x^0 - z^0} - O(n),$$

and for all $x \in \overset{\circ}{\mathcal{F}}_p$ the primal potential

$$\mathcal{P}_{n+1}(x, z) \ge \mathcal{P}_{n+1}(x^0, z^0) + (n+1)\log\frac{c^T x(z) - z}{c^T x^0 - z^0} - O(n).$$

Several algorithms presented in this chapter will generate a sequence of $x^k \in \overset{\circ}{\mathcal{F}}_p$ and $z^k \le b^T y^k$, where $y^k \in \mathcal{F}_d$, and either the primal potential $\mathcal{P}_{n+1}(x^k, z^k)$ or the max-potential of $\Omega(z^k)$ tends to $-\infty$, so that $z^* - z^k \to 0$. Moreover, there is a subsequence of $\{(x^k, y^k)\}$ which satisfies condition (4.6) with $z = z^k$. Thus, along this subsequence $c^T x^k - b^T y^k \to 0$ as well.

The basic idea is the following. Given z^{k-1}, we reduce $\mathcal{P}_{n+1}(x, z^{k-1})$ until an approximate center pair (x^k, y^k, s^k) of $\Omega(z^k)$ is generated, which is guaranteed by the discussions of the preceding chapter. Then, we must have $b^T y^k > z^{k-1}$. Thus, we update $z^k = b^T y^k > z^{k-1}$, that is, translate the hyperplane $b^T y \geq z^{k-1}$ to $b^T y \geq z^k$ through y^k. This results in reduction of both the primal potential and max-potential of the containing set $\Omega(z^k)$.

4.1 Karmarkar's Algorithm

We first show how Karmarkar's algorithm may be adapted to solve (LP) and (LD). Starting from (x^0, y^0, s^0), the algorithm, which will be described next, generates sequences $\{x^k \in \overset{\circ}{\mathcal{F}}_p\}$, $\{y^k \in \mathcal{F}_d\}$, and $\{z^k \leq b^T y^k\}$ such that

$$\mathcal{P}_{n+1}(x^{k+1}, z^{k+1}) \leq \mathcal{P}_{n+1}(x^k, z^k) - \delta \qquad \text{for} \qquad k = 0, 1, \ldots.$$

where constant $\delta \geq .2$. In particular, if

$$\mathcal{P}_{n+1}(x^k, z^k) - \mathcal{P}_{n+1}(x^0, z^0) \leq (n+1) \log \frac{\epsilon}{c^T x^0 - z^0} - n \log 2,$$

then from Proposition 4.1, we must have

$$z^* - b^T y^k \leq z^* - z^k \leq \epsilon.$$

That is, we have an ϵ approximation solution for (LD). Moreover, we generate a subsequence of $\{(x^k, y^k)\}$ which satisfies condition (4.6) with $z = z^k$. Thus, along this subsequence $c^T x^k - b^T y^k \to 0$.

Here is how we do it. Given $x^k \in \overset{\circ}{\mathcal{F}}_p$ and $z^k \leq b^T y^k$ where $y^k \in \mathcal{F}_d$, we again transform the (LP) problem into Karmarkar's canonical form:

$$\begin{array}{lll} (LP') & \text{minimize} & (c^k)^T x' \\ & \text{subject to} & x' \in \mathcal{K}_{p'}. \end{array}$$

where

$$\mathcal{K}_{p'} := \{x' \in \mathcal{R}^{n+1} : \ A^k x' = 0, \ e^T x' = n+1, \ x' \geq 0\},$$

$$c^k = \begin{pmatrix} X^k c \\ -z^k \end{pmatrix}, \quad \text{and} \quad A^k = \begin{pmatrix} AX^k, & -b \end{pmatrix}.$$

This is accomplished via an extended Karmarkar's projective transformation

$$x' = T(x) = \begin{pmatrix} \frac{(n+1)(X^k)^{-1} x}{1 + e^T (X^k)^{-1} x} \\ \frac{(n+1)}{1 + e^T (X^k)^{-1} x} \end{pmatrix} \quad \text{from} \quad x \in \mathcal{F}_p.$$

Obviously, $T(x^k) = e$ is the analytic center for $\mathcal{K}_{p'}$. In other words, the projective transformation maps x^k to the analytic center of $\mathcal{K}_{p'}$ in (LP'). Each feasible point $x \in \mathcal{F}_p$ is also mapped to a feasible point $x' \in \mathcal{K}_{p'}$. Conversely, each feasible point $x' \in \mathcal{K}_{p'}$ can be transformed back to an $x \in \mathcal{F}_p$ via the inverse transformation, T^{-1}, given by

$$x = T^{-1}(x') = \frac{X^k x'[n]}{x'_{n+1}} \quad \text{from} \quad x' \in \mathcal{K}_{p'},$$

where $x'[n]$ denotes the vector of the first n components of $x' \in \mathcal{R}^{n+1}$.

The projective transformation T also induces the potential function $\mathcal{P}'_{n+1}$ associated with (LP'):

$$\mathcal{P}'_{n+1}(x', z^k) = (n+1)\log((c^k)^T x') - \sum_{i=1}^{n+1} \log(x'_i), \quad x' \in \overset{\circ}{\mathcal{K}}_{p'} .$$

Again, the difference of the potential values at two points $x^1,\ x^2 \in \overset{\circ}{\mathcal{F}}_p$ is invariant under the projective transformation, i.e.,

$$\mathcal{P}'_{n+1}(T(x^2), z^k) - \mathcal{P}'_{n+1}(T(x^1), z^k) = \mathcal{P}_{n+1}(x^2, z^k) - \mathcal{P}_{n+1}(x^1, z^k).$$

Therefore, a reduction of the potential $\mathcal{P}'_{n+1}$ for the transformed problem induces the same amount of reduction of the potential $\mathcal{P}_{n+1}$ for the original problem.

To reduce $\mathcal{P}'_{n+1}(x', z^k)$, we again solve the following ball-constrained problem

$$\begin{array}{ll} \text{minimize} & (c^k)^T(x' - e) \\ \text{subject to} & A^k(x' - e) = 0,\ e^T(x' - e) = 0,\ \|x' - e\| \le \alpha. \end{array}$$

According to the discussion in Chapter 3, the solution of the problem is

$$x' - e = -\alpha \frac{p^k}{\|p^k\|},$$

where

$$\begin{aligned} p^k &= \left(I - \begin{pmatrix} A^k \\ e \end{pmatrix}^T \begin{pmatrix} A^k(A^k)^T & 0 \\ 0 & n \end{pmatrix}^{-1} \begin{pmatrix} A^k \\ e \end{pmatrix} \right) c^k \\ &= c^k - (A^k)^T y^k - \frac{c^T x^k - z^k}{n+1} e, \end{aligned}$$

and

$$
\begin{array}{rcl}
y^k & = & y(z^k) := y_2 + z^k y_1, \\
y_1 & = & (A^k (A^k)^T)^{-1} A X^k b, \\
y_2 & = & (A^k (A^k)^T)^{-1} A (X^k)^2 c.
\end{array} \tag{4.7}
$$

Consider the difference of the potential values

$$
\mathcal{P}'_{n+1}(x', z^k) - \mathcal{P}'_{n+1}(e, z^k) \le -\alpha \frac{(n+1)\|p^k\|}{c^T x^k - z^k} + \frac{\alpha^2}{2(1-\alpha)}.
$$

Thus, as long as $\|p^k\| \ge \eta \frac{c^T x^k - z^k}{n+1}$ for a constant $0 < \eta < 1$ we can choose an appropriate α such that

$$
\mathcal{P}'_{n+1}(x', z^k) - \mathcal{P}'_{n+1}(e, z^k) \le -\delta,
$$

for a positive constant $\delta = \alpha\eta - \alpha^2/2(1-\alpha)$. Let

$$
x^{k+1} = T^{-1}(x').
$$

Then,

$$
\mathcal{P}_{n+1}(x^{k+1}, z^k) - \mathcal{P}_{n+1}(x^k, z^k) \le -\delta.
$$

In this case, we simply let $z^{k+1} = z^k$.

What is happening when

$$
\|p^k\| < \eta \frac{c^T x^k - z^k}{n+1}?
$$

Note that p^k can be decomposed as

$$
\begin{array}{rcl}
p^k = p(z^k) & := & \begin{pmatrix} X^k c \\ -z^k \end{pmatrix} - \begin{pmatrix} X^k A^T \\ -b^T \end{pmatrix} y(z^k) - \dfrac{c^T x^k - z^k}{n+1} e \\
 & = & \begin{pmatrix} X^k (c - A^T y(z^k)) \\ b^T y(z^k) - z^k \end{pmatrix} - \dfrac{c^T x^k - z^k}{n+1} e.
\end{array} \tag{4.8}
$$

Thus, $\|p^k\| < \eta \frac{c^T x^k - z^k}{n+1}$ implies that $(x^k, y(z^k), z^k)$ satisfies condition (4.6), which means that $(x^k, y(z^k))$ is an approximate center pair for $\Omega(z^k)$. Consequently,

$$
\begin{pmatrix} X^k (c - A^T y(z^k)) \\ b^T y(z^k) - z^k \end{pmatrix} > 0,
$$

that is

$$
A^T y(z^k) < c \quad \text{and} \quad z^k < b^T y(z^k).
$$

By the duality theorem, $z^{k+1} = b^T y(z^k) > z^k$ is a new lower bound for z^*. We may update $\Omega(z^k)$ to $\Omega(z^{k+1})$, that is, we translate the inequality $b^T y \ge z^k$ through $y(z^k)$ to cut $\Omega(z^k)$.

However, we can place a deeper cut. We can do a simple ratio test to obtain

$$\bar{z} = \arg\max_{z}\{b^T y(z) : \; A^T y(z) \le c\}. \tag{4.9}$$

This leads us to a new dual feasible point and a new lower bound $b^T y(\bar{z}) \ge b^T y(z^k) > z^k$, and

$$\|p(b^T y(\bar{z}))\| \ge \frac{c^T x^k - z^k}{n+1}.$$

We describe the algorithm as follows.

Algorithm 4.1 *Given* $x^0 \in \overset{\circ}{\mathcal{F}}_p$ *and* $z^0 = b^T y^0$ *where* $y^0 \in \overset{\circ}{\mathcal{F}}_d$. *Set* $k := 0$.
While $(c^T x^k - z^k) > \epsilon$ **do**:

1. *Compute* y_1 *and* y_2 *from (4.7).*
2. *Compute* $\bar{z}$ *from (4.9) and let* $z^{k+1} = b^T y(\bar{z})$ *and* $y^{k+1} = y_2 + \bar{z} y_1$.
3. *Let*
$$x' = e - \frac{\alpha}{\|p(z^{k+1})\|} p(z^{k+1})$$
and
$$x^{k+1} = T^{-1}(x').$$
4. *Set* $k := k+1$ *and return to Step 1.*

We have

Lemma 4.3 *For* $k = 0, 1, 2, \ldots$

$$\mathcal{P}_{n+1}(x^{k+1}, z^{k+1}) \le \mathcal{P}_{n+1}(x^k, z^k) - \delta,$$

where $\delta \ge .2$.

This lemma leads to the following theorem:

Theorem 4.4 *In at most* $O(n \log(c^T x^0 - z^0)/\epsilon + n)$ *iterations, Algorithm 4.1 will generate* $x^k \in \overset{\circ}{\mathcal{F}}_p$ *and* $y^k \in \mathcal{F}_d$ *with* $z^k \le b^T y^k < z^*$ *such that*

$$z^* - z^k \le \epsilon.$$

Moreover, there is a subsequence $\{x^k, y^k\}$ *which is a sequence of approximate center pairs for* $\Omega(z^k)$ *with*

$$c^T x^k - b^T y^k \le \epsilon.$$

4.2 Path-Following Algorithm

While Karmarkar's algorithm reduces the primal potential function, which represents the logarithmic volume of a regular ellipsoid containing $\Omega(z^k)$, the path-following algorithm directly down-sizes $\Omega(z^k)$. Let (y^k, s^k) be an approximate analytic center of $\Omega(z^k)$ with $\delta(s^k) \le \eta < 1$, where

$$s^k = \begin{pmatrix} c - A^T y^k \\ b^T y^k - z^k \end{pmatrix} > 0.$$

Let β be a constant in $(0,1)$. Then, similar to what has been done in the center-section algorithm of Chapter 3, we update z^{k+1} from z^k at the kth iteration

$$z^{k+1} = b^T y^k - \beta s^k_{n+1}.$$

Accordingly,

$$\mathcal{B}(\Omega(z^{k+1})) \le \mathcal{B}(\Omega(z^k)) - \delta,$$

where δ is a positive constant. This process stops with y^k as an approximate center for $\Omega(z^k)$, where $z^k \ge z^* - \epsilon$. The total number of iterations is bounded by $O(n \log(c^T x^0 - z^0)/\epsilon + n \log 2)$ from Proposition 4.1, which is the same bound as in Karmarkar's algorithm.

In an additional effort, Renegar developed a new method to improve the iteration bound by a factor $\sqrt{n}$. A similar algorithm can be described as the following. Consider

$$\Omega(z) = \{y : \ c - A^T y \ge 0, -z + b^T y \ge 0, ..., -z + b^T y \ge 0\},$$

where "$-z + b^T y \ge 0$" is copied n times. Note that the slack vector $s \in \mathcal{R}^{2n}$ and

$$s = \begin{pmatrix} c - A^T y \\ b^T y - z \\ \cdots \\ b^T y - z \end{pmatrix} > 0.$$

Thus, $s_{n+1} = ... = s_{2n}$. The primal potential function for this $\Omega(z)$ is

$$\mathcal{P}_{2n}(x, z) = 2n \log(c^T x - z) - \sum_{j=1}^{n} \log x_j, \quad x \in \overset{\circ}{\mathcal{F}}_p \ .$$

Following the proof of Proposition 4.1, we have

$$\mathcal{B}(\Omega(z)) - \mathcal{B}(\Omega(z^0)) \ge 2n \log \frac{c^T x(z) - z}{c^T x^0 - z^0} - O(n).$$

Algorithm 4.2 *Given an approximate analytic center y^0 of $\Omega(z^0)$ with $\delta(s^0) \le \eta < 1$, set $k := 0$.*
While $(c^T x(z^k) - b^T y^k) > \epsilon$ **do**:

1. *Update $z^{k+1} = b^T y^k - \beta s^k_{2n}$.*

2. *Via the Newton procedure compute an approximate analytic center y^{k+1} with $\delta(s^{k+1}) \le \eta$ for $\Omega(z^{k+1})$. Let $x' = x(s^{k+1})$ be given by (3.3) for $\Omega(z^{k+1})$ (note here that the matrix A is augmented to $(A, -b, ..., -b)$) and $x(z^{k+1}) = x'[n]/(nx'_{2n})$.*

3. *Set $k := k+1$ and return to Step 1.*

We now prove the following Lemma.

Lemma 4.5 *Let $\beta = 1 - \eta/\sqrt{n}$. Then, there exists a point $x^+ > 0$ such that*

$$(A, -b, ..., -b)x^+ = 0 \quad \textit{and} \quad \|X^+ s^+ - e\| \le \delta < 1,$$

where

$$s^+ = \begin{pmatrix} c - A^T y^k \\ b^T y^k - z^{k+1} \\ \cdots \\ b^T y^k - z^{k+1} \end{pmatrix} > 0.$$

Proof. Let $x^+ = x(s^k) > 0$ of (3.3) for $\Omega(z^k)$, where matrix A is augmented to $(A, -b, ..., -b)$. Then, $x^+_{n+1} = ..., = x^+_{2n}$,

$$(A, -b, ..., -b)x^+ = Ax(s^k) = 0,$$

and

$$\|X^+ s^k - e\| = \delta(s^k) \le \eta.$$

Note that $s^+_j = s^k_j$ for $j \le n$ and $s^+_j = \beta s^k_j$ for $j \ge n+1$. Thus,

$$\begin{aligned} \|X^+ s^+ - e\| &= \|X^+ s^k - e + X^+ s^+ - X^+ s^k\| \\ &\le \|X^+ s^k - e\| + \|X^+ s^+ - X^+ s^k\| \\ &\le \|X^+ s^k - e\| + \sqrt{n}|x^+_{2n} s^k_{2n}(\beta - 1)| \\ &\le \eta + \sqrt{n}(1-\beta)(1+\eta) \\ &\le \eta + \eta(1+\eta). \end{aligned}$$

□

Lemma 4.5 shows that, even though it is not perfectly centered, y^k is in the "quadratic convergence" region of the center of $\Omega(z^{k+1})$, if we choose $\eta + \eta(1+\eta) < 1$. Thus, an approximate center y^{k+1} with $\delta(s^{k+1}) \le \eta$ for $\Omega(z^{k+1})$ can be updated from y^k in a constant number of Newton's steps. We now verify that the max-potential of $\Omega(z^{k+1})$ is reduced by $\sqrt{n}\delta$ for a constant δ.

Lemma 4.6 *Let (y^k, s^k) be an approximate center for Ω^k with $\delta(s^k) \le \eta$ and let $\Omega(z^{k+1})$ be given as in Step 1 of Algorithm 4.2. Then,*

$$\mathcal{B}(\Omega(z^{k+1})) \le \mathcal{B}(\Omega(z^k)) - \sqrt{n}\delta \quad \textit{for a constant} \quad \delta > 0.$$

Proof. The proof is very similar to Theorem 2.7 in Chapter 2. Let (y^a, s^a) and (y^a_+, s^a_+) be the centers of $\Omega(z^k)$ and $\Omega(z^{k+1})$, respectively. Note we have

$$s^a_{n+1} = ... = s^a_{2n} = b^T y^a - z^k.$$

Also

$$(A, -b, ..., -b)(S^a)^{-1}e = 0,$$
$$(c, -z^k, ..., -z^k)^T (S^a)^{-1} e = (s^a)^T (S^a)^{-1} e = 2n,$$

and

$$s^a_+ = \begin{pmatrix} c - A^T y^a_+ \\ b^T y^a_+ - z^{k+1} \\ \cdots \\ b^T y^a_+ - z^{k+1} \end{pmatrix} > 0,$$

where

$$z^{k+1}_{n+1} = ... = z^{k+1}_{2n} = z^k + (1-\beta) s^k_{2n}.$$

Then we have

$$\begin{array}{rcl} e^T (S^a)^{-1} s^a_+ & = & e^T (S^a)^{-1} (c; -z^{k+1}; ...; -z^{k+1}) \\ & = & e^T (S^a)^{-1} (c; -z^k; ...; -z^k) - n(1-\beta)(s^k_{2n}/s^a_{2n}) \\ & = & 2n - n(1-\beta)(s^k_{2n}/s^a_{2n}) \\ & \le & 2n - n(1-\beta)(1-\eta). \end{array}$$

The last inequality is due to (v) of Theorem 3.2. Therefore,

$$\begin{array}{rcl} \dfrac{\exp \mathcal{B}(\Omega(z^{k+1}))}{\exp \mathcal{B}(\Omega(z^k))} & = & \displaystyle\prod_{j=1}^{2n} \frac{(s^a_+)_j}{s^a_j} \\ & \le & (\dfrac{2n - n(1-\beta)(1-\eta)}{2n})^{2n} \\ & \le & \exp(-n(1-\beta)(1-\eta)) \\ & = & \exp(-\sqrt{n}\eta(1-\eta)). \end{array}$$

This completes the proof.

□

From the lemma, we can conclude that

Theorem 4.7 *In at most $O(\sqrt{n}\log(c^Tx^0 - z^0)/\epsilon + \sqrt{n})$ iterations, Algorithm 4.2 will generate a $(x(z^k), y^k) \in \overset{\circ}{\mathcal{F}}$ such that it is an approximate center for $\Omega(z^k)$ where*

$$c^Tx(z^k) - b^Ty^k \leq c^Tx(z^k) - z^k \leq \epsilon.$$

4.3 Potential Reduction Algorithm

At this point, we can see the difference between Karmarkar's and the path-following algorithms. The former, called potential reduction algorithms, are equipped with a primal potential functions, which are solely used to measure the solution's progress. There is no restriction on either step-size or z-update during the iterative process; the greater the reduction of the potential function, the faster the convergence of the algorithm. The path-following algorithms are equipped with the max-potential, so that the z-update needs to be carefully chosen and each step needs to stay close to the analytic center. Thus, from a practical point of view, Karmarkar's algorithm has an advantage.

The next question is whether or not we can improve the complexity bound by the same factor for potential reduction algorithms. Let $x \in \overset{\circ}{\mathcal{F}}_p$ and $(y, s) \in \overset{\circ}{\mathcal{F}}_d$. Then consider the primal-dual potential function

$$\psi_{n+\rho}(x, s) = (n + \rho)\log(x^Ts) - \sum_{j=1}^{n}\log(x_js_j),$$

where $\rho = \sqrt{n}$. Let $z = b^Ty$, then $s^Tx = c^Tx - z$ and we have

$$\psi_{n+\rho}(x, s) = \mathcal{P}_{n+\rho}(x, z) - \sum_{j=1}^{n}\log s_j,$$

where the primal potential function

$$\mathcal{P}_{n+\rho}(x, z) = (n + \rho)\log(c^Tx - z) - \sum_{j=1}^{n}\log x_j.$$

Consider a pair of $(x^k, y^k, s^k) \in \overset{\circ}{\mathcal{F}}$. Fix $z^k = b^T y^k$, then the gradient vector of the primal potential function at x^k is

$$\nabla \mathcal{P}_{n+\rho}(x^k, z^k) = \frac{(n+\rho)}{(s^k)^T x^k} c - (X^k)^{-1} e.$$

Now we directly solve the problem

$$\begin{array}{ll} \text{minimize} & \nabla \mathcal{P}_{n+\rho}(x^k, z^k)^T (x - x^k) \\ \text{subject to} & A(x - x^k) = 0, \ \|(X^k)^{-1}(x - x^k)\| \le \alpha. \end{array}$$

Let the minimizer be x^{k+1}. Then

$$x^{k+1} - x^k = -\alpha \frac{X^k p^k}{\|p^k\|}, \tag{4.10}$$

and, in view of Section 3.3.2,

$$\mathcal{P}_{n+\rho}(x^{k+1}, z^k) - \mathcal{P}_{n+\rho}(x^k, z^k) \le -\alpha \|p^k\| + \frac{\alpha^2}{2(1-\alpha)},$$

where

$$\begin{aligned} p^k &= (I - X^k A^T (A (X^k)^2 A^T)^{-1} A X^k) X^k \nabla \mathcal{P}_{n+\rho}(x^k, z^k) \\ &= X^k \nabla \mathcal{P}_{n+\rho}(x^k, z^k) - X^k A^T y^k \end{aligned}$$

or

$$p^k = X^k \left(\frac{(n+\rho)}{(s^k)^T x^k} c - A^T y^k\right) - e,$$

and

$$y^k = (A(X^k)^2) A^T)^{-1} A (X^k)^2 \nabla \mathcal{P}_{n+\rho}(x^k, z^k).$$

Thus, as long as $\|p^k\| \ge \eta > 0$, we may choose an appropriate α such that

$$\mathcal{P}_{n+\rho}(x^{k+1}, z^k) - \mathcal{P}_{n+\rho}(x^k, z^k) \le -\delta$$

for some positive constant δ.

Now, we focus on the expression of p^k, which can be rewritten as

$$\begin{aligned} p^k = p(z^k) &= (I - X^k A^T (A (X^k)^2 A^T)^{-1} A X^k) \left(\frac{(n+\rho)}{(s^k)^T x^k} X^k c - e\right) \\ &= \frac{(n+\rho)}{(s^k)^T x^k} X^k s(z^k) - e, \end{aligned} \tag{4.11}$$

with

$$s(z^k) = c - A^T y(z^k) \tag{4.12}$$

and

$$y(z^k) = y_2 - \frac{(s^k)^T x^k}{(n+\rho)} y_1, \tag{4.13}$$

where y_1 and y_2 are given in equation (4.7).

Regarding $\|p^k\| = \|p(z^k)\|$, we have the following lemma.

Lemma 4.8 *Let*

$$\mu^k = \frac{(x^k)^T s^k}{n} = \frac{c^T x^k - z^k}{n} \quad \textit{and} \quad \mu = \frac{(x^k)^T s(z^k)}{n}.$$

If

$$\|p(z^k)\| < \min(\eta\sqrt{\frac{n}{n+\eta^2}}, 1-\eta), \tag{4.14}$$

then the following three inequalities hold:

$$s(z^k) > 0, \quad \|X^k s(z^k) - \mu e\| < \eta\mu, \quad \textit{and} \quad \mu < (1 - .5\eta/\sqrt{n})\mu^k. \tag{4.15}$$

Proof. The proof is by contradiction.

i) If the first inequality of (4.15) is not true, then $\exists\, j$ such that $s_j(z^k) \le 0$ and

$$\|p(z^k)\| \ge 1 - \frac{(n+\rho)}{n\mu^k} x_j s_j(z^k) \ge 1.$$

ii) If the second inequality of (4.15) does not hold, then

$$\begin{aligned}
\|p(z^k)\|^2 &= \|\frac{(n+\rho)}{n\mu^k} X^k s(z^k) - \frac{(n+\rho)\mu}{n\mu^k} e + \frac{(n+\rho)\mu}{n\mu^k} e - e\|^2 \\
&= (\frac{(n+\rho)}{n\mu^k})^2 \|X^k s(z^k) - \mu e\|^2 + \|\frac{(n+\rho)\mu}{n\mu^k} e - e\|^2 \\
&\ge (\frac{(n+\rho)\mu}{n\mu^k})^2 \eta^2 + (\frac{(n+\rho)\mu}{n\mu^k} - 1)^2 n \qquad (4.16) \\
&\ge \eta^2 \frac{n}{n+\eta^2},
\end{aligned}$$

where the last relation prevails since the quadratic term yields the minimum at

$$\frac{(n+\rho)\mu}{n\mu^k} = \frac{n}{n+\eta^2}.$$

iii) If the third inequality of (4.15) is violated, then

$$\frac{(n+\rho)\mu}{n\mu^k} \ge (1 + \frac{1}{\sqrt{n}})(1 - \frac{.5\eta}{\sqrt{n}}) \ge 1,$$

which in view of (4.16) leads to

$$\begin{aligned}
\|p(z^k)\|^2 &\geq (\frac{(n+\rho)\mu}{n\mu^k} - 1)^2 n \\
&\geq ((1+\frac{1}{\sqrt{n}})(1-\frac{.5\eta}{\sqrt{n}}) - 1)^2 n \\
&\geq (1-\frac{\eta}{2}-\frac{\eta}{2\sqrt{n}})^2 \\
&\geq (1-\eta)^2.
\end{aligned}$$

$\square$

Based on this lemma, we have the following potential reduction theorem.

Theorem 4.9 *Given* $x^k \in \overset{\circ}{\mathcal{F}}_p$ *and* $(y^k, s^k) \in \overset{\circ}{\mathcal{F}}_d$. *Let* $\rho = \sqrt{n}$, $z^k = b^T y^k$, x^{k+1} *be given by (4.10), and* $y^{k+1} = y(z^k)$ *in (4.13) and* $s^{k+1} = s(z^k)$ *in (4.12). Then, either*

$$\psi_{n+\rho}(x^{k+1}, s^k) \leq \psi_{n+\rho}(x^k, s^k) - \delta$$

or

$$\psi_{n+\rho}(x^k, s^{k+1}) \leq \psi_{n+\rho}(x^k, s^k) - \delta$$

where $\delta > 1/20$.

Proof. If (4.14) does not hold, i.e.,

$$\|p(\underline{z}^k)\| \geq \min(\eta\sqrt{\frac{n}{n+\eta^2}}, 1-\eta),$$

then

$$\mathcal{P}_{n+\rho}(x^{k+1}, z^k) - \mathcal{P}_{n+\rho}(x^k, z^k) \leq -\alpha \min(\eta\sqrt{\frac{n}{n+\eta^2}}, 1-\eta) + \frac{\alpha^2}{2(1-\alpha)},$$

hence from the relation between $\mathcal{P}_{n+\rho}$ and $\psi_{n+\rho}$,

$$\psi_{n+\rho}(x^{k+1}, s^k) - \psi_{n+\rho}(x^k, s^k) \leq -\alpha \min(\eta\sqrt{\frac{n}{n+\eta^2}}, 1-\eta) + \frac{\alpha^2}{2(1-\alpha)}.$$

Otherwise, from Lemma 4.8 the inequalities of (4.15) hold:

i) The first of (4.15) indicates that y^{k+1} and s^{k+1} are in $\overset{\circ}{\mathcal{F}}_d$.

ii) Using the second of (4.15) and applying Lemma 3.1 to vector $X^k s^{k+1}/\mu$, we have

$$\begin{aligned}
& n\ln(x^k)^T s^{k+1} - \sum_{j=1}^n \ln(x_j^k s_j^{k+1}) \\
= \quad & n\ln n - \sum_{j=1}^n \ln(x_j^k s_j^{k+1}/\mu) \\
\leq \quad & n\ln n + \frac{\|X^k s^{k+1}/\mu - e\|^2}{2(1-\|X^k s^{k+1}/\mu - e\|_\infty)} \\
\leq \quad & n\ln n + \frac{\eta^2}{2(1-\eta)} \\
\leq \quad & n\ln(x^k)^T s^k - \sum_{j=1}^n \ln(x_j^k s_j^k) + \frac{\eta^2}{2(1-\eta)}.
\end{aligned}$$

iii) According to the third of (4.15), we have

$$\sqrt{n}(\ln(x^k)^T s^{k+1} - \ln(x^k)^T s^k) = \sqrt{n}\ln\frac{\mu}{\mu^k} \leq -\frac{\eta}{2}.$$

Adding the two inequalities in ii) and iii), we have

$$\psi_{n+\rho}(x^k, s^{k+1}) \leq \psi_{n+\rho}(x^k, s^k) - \frac{\eta}{2} + \frac{\eta^2}{2(1-\eta)}.$$

Thus, by choosing $\eta = .43$ and $\alpha = .3$ we have the desired result.

$\square$

Theorem 4.9 establishes an important fact: the *primal-dual* potential function can be reduced by a constant no matter where x^k and y^k are. In practice, one can perform the line search to minimize the primal-dual potential function. This results in the following primal-dual potential reduction algorithm.

Algorithm 4.3 *Given* $x^0 \in \overset{\circ}{\mathcal{F}}_p$ *and* $(y^0, s^0) \in \overset{\circ}{\mathcal{F}}_d$. *Let* $z^0 = b^T y^0$. *Set* $k := 0$.

While $(s^k)^T x^k \geq \epsilon$ **do**

1. *Compute* y_1 *and* y_2 *from (4.7).*
2. *Set* $y^{k+1} = y(\bar{z})$, $s^{k+1} = s(\bar{z})$, $z^{k+1} = b^T y^{k+1}$ *with*

$$\bar{z} = \arg\min_{z \geq z^k} \psi_{n+\rho}(x^k, s(z)).$$

3. *Let* $x^{k+1} = x^k - \bar{\alpha} X^k p(z^{k+1})$ *with*

$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi_{n+\rho}(x^k - \alpha X^k p(z^{k+1}), s^{k+1}).$$

4. *Set* $k := k + 1$ *and return to Step 1.*

The performance of the algorithm results from the following corollary.

Corollary 4.10 *Let* $\rho = \sqrt{n}$. *Then, Algorithm 4.3 terminates in at most* $O(\sqrt{n}\log(c^T x^0 - b^T y^0)/\epsilon)$ *iterations with*

$$c^T x^k - b^T y^k \leq \epsilon.$$

Proof. In $O(\sqrt{n}\log((x^0)^T s^0/\epsilon))$ iterations

$$\begin{array}{rcl} -\sqrt{n}\log((x^0)^T s^0/\epsilon) & = & \psi_{n+\rho}(x^k, s^k) - \psi_{n+\rho}(x^0, s^0) \\ & \geq & \sqrt{n}\log(x^k)^T s^k + n\log n - \psi_{n+\rho}(x^0, s^0) \\ & = & \sqrt{n}\log((x^k)^T s^k/(x^0)^T s^0). \end{array}$$

Thus,

$$\sqrt{n}\log(c^T x^k - b^T y^k) = \sqrt{n}\log(x^k)^T s^k \leq \sqrt{n}\log\epsilon,$$

i.e.,

$$c^T x^k - b^T y^k = (x^k)^T s^k \leq \epsilon.$$

□

4.4 Primal-Dual (Symmetric) Algorithm

Once we have a pair $(x, y, s) \in \overset{\circ}{\mathcal{F}}$ with $\mu = x^T s/n$, we can apply the primal-dual Newton method to generate a new iterate x^+ and (y^+, s^+) as follows: Solve for d_x, d_y and d_s from the system of linear equations:

$$\begin{array}{rcl} Sd_x + Xd_s & = & \gamma\mu e - Xs, \\ Ad_x & = & 0, \\ A^T d_y + d_s & = & 0. \end{array} \tag{4.17}$$

It is well-known that this is the Newton step starting from (x, s) which helps to find the point on the central path with duality gap $n\gamma\mu$. Note that $d_x^T d_s = -d_x^T A^T d_y = 0$. We present the following lemma whose proof is very similar to Lemma 3.6 and will be omitted.

Lemma 4.11 *Let the direction d_x, d_y and d_s be generated by equation (4.17) with $\gamma = n/(n+\rho)$, and let*

$$\theta = \frac{\alpha\sqrt{\min(Xs)}}{\|(XS)^{-1/2}(\frac{x^T s}{(n+\rho)}e - Xs)\|}, \tag{4.18}$$

where α is a positive constant less than 1. Let

$$x^+ = x + \theta d_x, \quad y^+ = y + \theta d_y, \quad \text{and} \quad s^+ = s + \theta d_s.$$

Then, we have $(x^+, y^+, s^+) \in \overset{\circ}{\mathcal{F}}$ and

$$\psi_{n+\rho}(x^+, s^+) - \psi_{n+\rho}(x, s)$$

$$\leq -\alpha\sqrt{\min(Xs)}\|(XS)^{-1/2}(e - \frac{(n+\rho)}{x^T s}Xs)\| + \frac{\alpha^2}{2(1-\alpha)}.$$

Let $v = Xs$. Then consider

$$\frac{(n+\rho)\sqrt{\min(v)}}{e^T v}\|V^{-1/2}(\frac{e^T v}{(n+\rho)}e - v)\|.$$

We now prove the following lemma

Lemma 4.12 *Let v be a n-dimensional positive vector and $\rho \geq \sqrt{n}$. Then,*

$$\sqrt{\min(v)}\|V^{-1/2}(e - \frac{(n+\rho)}{e^T v}v)\| \geq \sqrt{3/4}.$$

From these two lemmas we have

$$\psi_{n+\rho}(x^+, s^+) - \psi_{n+\rho}(x, s)$$

$$\leq -\alpha\sqrt{3/4} + \frac{\alpha^2}{2(1-\alpha)} = -\delta$$

for a constant δ. This leads to

Algorithm 4.4 *Given $(x^0, y^0, s^0) \in \overset{\circ}{\mathcal{F}}$. Set $\rho = \sqrt{n}$ and $k := 0$.*
While $(s^k)^T x^k \geq \epsilon$ **do**

1. *Set $(x, s) = (x^k, s^k)$ and $\gamma = n/(n+\rho)$ and compute (d_x, d_y, d_s) from (4.17).*

2. *Let $x^{k+1} = x^k + \bar{\alpha} d_x$, $y^{k+1} = y^k + \bar{\alpha} d_y$, and $s^{k+1} = s^k + \bar{\alpha} d_x$ where*

$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi_{n+\rho}(x^k + \alpha d_x, s^k + \alpha d_s).$$

3. Set $k := k + 1$ and return to Step 1.

Theorem 4.13 *Let $\rho = \sqrt{n}$. Then, Algorithm 4.4 terminates in at most $O(\sqrt{n}\log((x^0)^T s^0/\epsilon))$ iterations with*

$$c^T x^k - b^T y^k \leq \epsilon.$$

4.5 Adaptive Path-Following Algorithms

Here we describe and analyze several additional primal-dual interior-point algorithms for linear programming. In some sense these methods follow the central path

$$\mathcal{C} = \left\{ (x, s) \in \overset{\circ}{\mathcal{F}} \colon Xs = \mu e \quad \text{where} \quad \mu = \frac{x^T s}{n} \right\}$$

in primal-dual form, but certain algorithms allow a very loose approximation to the path.

Suppose we have a pair $(x, s) \in \mathcal{N}$, a neighborhood of $\mathcal{C}$, where $\mathcal{C} \subset \mathcal{N} \subset \overset{\circ}{\mathcal{F}}$. Consider the neighborhood

$$\mathcal{N}_2(\eta) = \left\{ (x, s) \in \overset{\circ}{\mathcal{F}} \colon \|Xs - \mu e\| \leq \eta\mu \quad \text{where} \quad \mu = \frac{x^T s}{n} \right\}$$

for some $\eta \in (0, 1)$. We will first analyze an adaptive-step path-following algorithm that generates a sequence of iterates in $\mathcal{N}_2(1/4)$. Actually, the algorithm has a predictor-corrector form, so that it also generates intermediate iterates in $\mathcal{N}_2(1/2)$.

Next we consider adaptive-step algorithms generating sequences of iterates in either

$$\mathcal{N}_\infty(\eta) = \left\{ (x, s) \in \overset{\circ}{\mathcal{F}} \colon \|Xs - \mu e\|_\infty \leq \eta\mu \quad \text{where} \quad \mu = \frac{x^T s}{n} \right\}$$

or

$$\mathcal{N}_\infty^-(\eta) = \left\{ (x, s) \in \overset{\circ}{\mathcal{F}} \colon \|Xs - \mu e\|_\infty^- \leq \eta\mu \quad \text{where} \quad \mu = \frac{x^T s}{n} \right\},$$

for any $\eta \in (0, 1)$. Here, for any $z \in R^n$,

$$\|z\|_\infty^- := \|z^-\|_\infty$$

and

$$\|z\|_\infty^+ := \|z^+\|_\infty,$$

where $(z^-)_j := \min\{z_j, 0\}$ and $(z^+)_j := \max\{z_j, 0\}$ and $\|\cdot\|_\infty$ is the usual ℓ_∞ norm. Note that $\|z\|_\infty = \max\{\|z\|_\infty^+, \|z\|_\infty^-\}$ and that neither $\|\cdot\|_\infty^-$ nor $\|\cdot\|_\infty^+$ is a norm, although they obey the triangle inequality.

We easily see that

$$\mathcal{C} \subset \mathcal{N}_2(\eta) \subset \mathcal{N}_\infty(\eta) \subset \mathcal{N}_\infty^-(\eta) \subset \overset{\circ}{\mathcal{F}} \text{ for each } \eta \in (0,1).$$

Our results indicate that when we use a wider neighborhood of the central path, the worst-case number of iterations grows, while the practical behavior might be expected to improve.

Given $(x, s) \in \mathcal{N}$, we again generate a search direction $d = (d_x, d_s)$ using the primal-dual scaling method by solving (4.17). To show the dependence of d on the current pair and the parameter γ, we write $d = d(x, s, \gamma)$.

Having obtained the search direction d, we let

$$\begin{array}{rcl} x(\theta) & := & x + \theta d_x, \\ y(\theta) & := & y + \theta d_y, \\ s(\theta) & := & s + \theta d_s. \end{array} \tag{4.19}$$

We will frequently let the next iterate be $(x^+, s^+) = (x(\bar\theta), s(\bar\theta))$, where $\bar\theta$ is as large as possible so that $(x(\theta), s(\theta))$ remains in the neighborhood $\mathcal{N}$ for $\theta \in [0, \bar\theta]$.

Let $\mu(\theta) = x(\theta)^T s(\theta)/n$ and $X(\theta) = \operatorname{diag}(x(\theta))$. In order to get bounds on $\bar\theta$, we first note that

$$\mu(\theta) = (1-\theta)\mu + \theta\gamma\mu, \tag{4.20}$$

$$X(\theta)s(\theta) - \mu(\theta)e = (1-\theta)(Xs - \mu e) + \theta^2 D_x d_s, \tag{4.21}$$

where $D_x = \operatorname{diag}(d_x)$. Thus $D_x d_s$ is the second-order term in Newton's method to compute a new point of $\mathcal{C}$. Hence we can usually choose a larger $\bar\theta$ (and get a larger decrease in the duality gap) if $D_x d_s$ is smaller. In this section we obtain several bounds on the size of $D_x d_s$.

First, it is helpful to re-express $D_x d_s$. Let

$$\begin{array}{rcl} p & := & X^{-.5} S^{.5} d_x, \\ q & := & X^{.5} S^{-.5} d_s, \\ r & := & (XS)^{-.5}(\gamma\mu e - Xs), \end{array} \tag{4.22}$$

Lemma 4.14 *With the notations above,*

i)

$$\|Pq\| \le \frac{\sqrt{2}}{4}\|r\|^2;$$

ii)

$$-\frac{\|r\|^2}{4} \leq p_j q_j \leq \frac{r_j^2}{4} \quad \textit{for each} \quad j;$$

iii)

$$\begin{aligned}
\|Pq\|_\infty^- &\leq \frac{\|r\|^2}{4} \leq \frac{n\|r\|_\infty^2}{4}, \\
\|Pq\|_\infty^+ &\leq \frac{\|r\|_\infty^2}{4}, \quad \textit{and} \\
\|Pq\|_\infty &\leq \frac{\|r\|^2}{4} \leq \frac{n\|r\|_\infty^2}{4}.
\end{aligned}$$

The bounds in lemma 4.14 cannot improved by much in the worst case: consider the case where

$$\begin{aligned}
r &= e = (1, 1, \cdots, 1)^T, \\
p &= (1/2, 1/2, \cdots, 1/2, (1+\sqrt{n})/2)^T, \quad \text{and} \\
q &= (1/2, 1/2, \cdots, 1/2, (1-\sqrt{n})/2)^T.
\end{aligned}$$

To use lemma 4.14 we also need to bound r. The following result is useful:

Lemma 4.15 *Let r be as above.*

i) *If $\gamma = 0$, then $\|r\|^2 = n\mu$.*

ii) *If $\eta \in (0,1)$, $\gamma = 1$ and $(x,s) \in \mathcal{N}_2(\eta)$, then $\|r\|^2 \leq \eta^2\mu/(1-\eta)$.*

iii) *If $\eta \in (0,1)$, $\gamma \in (0,1)$, $\gamma \leq 2(1-\eta)$ and $(x,s) \in \mathcal{N}_\infty^-(\eta)$, then $\|r\|^2 \leq n\mu$. Moreover, if $(x,s) \in \mathcal{N}_\infty(\eta)$ then*

$$\sqrt{1-\eta} \geq r_j/\sqrt{\mu} \geq -\sqrt{1+\eta},$$

so $\|r\|_\infty^2 \leq (1+\eta)\mu$.

Proof. i) If $\gamma = 0$, $r = -(XS)^{-.5}Xs$, so $\|r\|^2 = x^T s = n\mu$.

ii) Now $r = (XS)^{-.5}(\mu e - Xs)$, so $\|r\| \leq \frac{1}{\sqrt{(1-\eta)\mu}}\eta\mu$, which yields the desired result.

iii) In this case

$$\|r\|^2 = \sum_{j=1}^n \frac{(\gamma\mu - x_j s_j)^2}{x_j s_j}$$

$$
\begin{aligned}
&= \sum_{j=1}^{n} \left(\frac{(\gamma\mu)^2}{x_j s_j} - 2\gamma\mu + x_j s_j \right) \\
&\leq \frac{n(\gamma\mu)^2}{(1-\eta)\mu} - 2n\gamma\mu + n\mu \quad (\text{since } \ x_j s_j \geq (1-\eta)\mu) \\
&\leq n\mu \quad (\text{since } \ \gamma \leq 2(1-\eta)).
\end{aligned}
$$

Now suppose $(x,s) \in \mathcal{N}_\infty(\eta)$, so that $x_j s_j \in [(1-\eta)\mu, (1+\eta)\mu]$ for each j. Thus, for each j,

$$
\frac{\gamma\mu}{\sqrt{1-\eta}\sqrt{\mu}} - \sqrt{1-\eta}\sqrt{\mu} \geq r_j = \frac{\gamma\mu}{\sqrt{x_j s_j}} - \sqrt{x_j s_j} \geq \frac{\gamma\mu}{\sqrt{1+\eta}\sqrt{\mu}} - \sqrt{1+\eta}\sqrt{\mu},
$$

which yields the final result since $0 \leq \gamma \leq 2(1-\eta)$.

□

4.5.1 Predictor-corrector algorithm

In this section we describe and analyze an algorithm that takes a single "corrector" step to the central path after each "predictor" step to decrease μ. Although it is possible to use more general values of η, we will work with nearly-centered pairs in $\mathcal{N}_2(\eta)$ with $\eta = 1/4$ (iterates after the corrector step), and intermediate pairs in $\mathcal{N}_2(\alpha)$ with $\alpha = 2\eta$ (iterates after a predictor step).

Algorithm 4.5 *Given* $(x^0, s^0) \in \mathcal{N}_2(\eta)$ *with* $\eta = 1/4$. *Set* $k := 0$.
While $(x^k)^T s^k > \epsilon$ **do**:

1. *Predictor step: set* $(x,s) = (x^k, s^k)$ *and compute* $d = d(x,s,0)$ *from (4.17); compute the largest* $\bar{\theta}$ *so that*

$$
(x(\theta), s(\theta)) \in \mathcal{N}_2(2\eta) \quad \textit{for} \quad \theta \in [0, \bar{\theta}].
$$

2. *Corrector step: set* $(x', s') = (x(\bar{\theta}), s(\bar{\theta}))$ *and compute* $d' = d(x', s', 1)$ *from (4.17); set* $(x^{k+1}, s^{k+1}) = (x' + d'_x, s' + d'_s)$.

3. *Set* $k := k+1$ *and return to Step 1.*

To analyze this method, we start by showing

Lemma 4.16 *For each* k, $(x^k, s^k) \in \mathcal{N}_2(\eta)$.

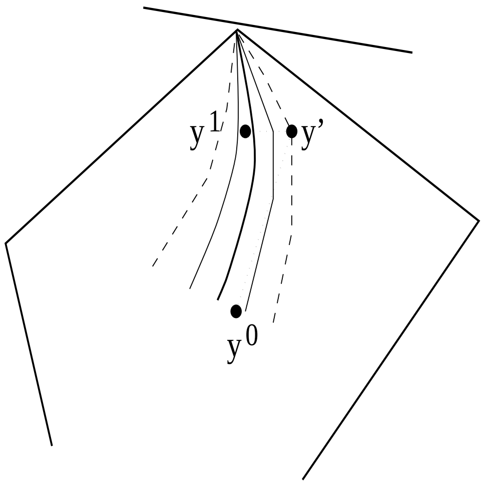

Figure 4.1: Illustration of the predictor-corrector algorithm; the predictor step moves y^0 in a narrower neighborhood of the central path to y' on the boundary of a wider neighborhood and the corrector step then moves y' to y^1 back in the narrower neighborhood.

Proof. The claim holds for $k = 0$ by hypothesis. For $k > 0$, let (x', s') be the result of the predictor step at the kth iteration and let $d' = d(x', s', 1)$, as in the description of the algorithm. Let $x'(\theta)$ and $s'(\theta)$ be defined as in (4.19) and p', q' and r' as in (4.22) using x', s' and d'. Let $\mu'(\theta) := x'(\theta)^T s'(\theta)/n$ for all $\theta \in [0, 1]$ with $\mu' := \mu'(0) = (x')^T s'/n$ and $\mu^{k+1} := \mu'(1) = (x^{k+1})^T s^{k+1}/n$.

From (4.20),

$$\mu'(\theta) = \mu' \quad \text{for all} \quad \theta, \tag{4.23}$$

and in particular $\mu^{k+1} = \mu'$. From (4.21),

$$\begin{aligned} X'(\theta)s'(\theta) - \mu'(\theta)e &= (1-\theta)(X's' - \mu'e) + \theta^2 D'_x d'_s \\ &= (1-\theta)(X's' - \mu'e) + \theta^2 P'q', \end{aligned} \tag{4.24}$$

where $X'(\theta) = \operatorname{diag}(x'(\theta))$, etc. But by lemma 4.14(i), lemma 4.15(ii) and $(x', s') \in \mathcal{N}(2\eta)$ with $\eta = 1/4$,

$$\|P'q'\| \le \frac{\sqrt{2}}{4}\|r'\|^2 \le \frac{\sqrt{2}}{4}\frac{(2\eta)^2}{1-2\eta}\mu' < \frac{1}{4}\mu'.$$

It follows that

$$\|X'(\theta)s'(\theta) - \mu'e\| \le (1-\theta)\frac{\mu'}{2} + \theta^2\frac{\mu'}{4} \le \frac{1}{2}\mu'. \tag{4.25}$$

Thus $X'(\theta)s'(\theta) \geq \frac{\mu'}{2}e > 0$ for all $\theta \in [0,1]$, and this implies that $x'(\theta) > 0$, $s'(\theta) > 0$ for all such θ by continuity. In particular, $x^{k+1} > 0$, $s^{k+1} > 0$, and (4.25) gives $(x^{k+1}, s^{k+1}) \in \mathcal{N}_2(1/4)$ as desired when we set $\theta = 1$.

$\square$

Now let $(x, s) = (x^k, s^k)$, $d = d(x, s, 0)$, $\mu = \mu^k = x^T s/n$, and p, q and r be as in (4.22); these quantities all refer to the predictor step at iteration k. By (4.20),

$$\begin{aligned} \mu' &= (1-\bar{\theta})\mu, \quad \text{or} \\ \mu^{k+1} &= (1-\bar{\theta})\mu^k. \end{aligned} \tag{4.26}$$

Hence the improvement in the duality gap at the kth iteration depends on the size of $\bar{\theta}$.

Lemma 4.17 *With the notation above, the step-size in the predictor step satisfies*

$$\bar{\theta} \geq \frac{2}{1 + \sqrt{1 + 4\|Pq/\mu\|/\eta}}.$$

Proof. By (4.21) applied to the predictor step,

$$\begin{aligned} \|X(\theta)s(\theta) - \mu(\theta)e\| &= \|(1-\theta)(Xs - \mu e) + \theta^2 Pq\| \\ &\leq (1-\theta)\|Xs - \mu e\| + \theta^2\|Pq\| \\ &\leq (1-\theta)\eta\mu + \theta^2\|Pq\|, \end{aligned}$$

after using lemma 4.16. We see that for

$$0 \leq \theta \leq \frac{2}{1 + \sqrt{1 + 4\|Pq/\mu\|/\eta}}$$

$$\begin{aligned} \|X(\theta)s(\theta) - \mu(\theta)e\|/\mu &\leq (1-\theta)\eta + \theta^2\|Pq/\mu\| \\ &\leq 2\eta(1-\theta). \end{aligned}$$

This is because the quadratic term in θ:

$$\|Pq/\mu\|\theta^2 + \eta\theta - \eta \leq 0$$

for θ between zero and the root

$$\frac{-\eta + \sqrt{\eta^2 + 4\|Pq/\mu\|\eta}}{2\|Pq/\mu\|} = \frac{2}{1 + \sqrt{1 + 4\|Pq/\mu\|/\eta}}.$$

Thus,

$$\|X(\theta)s(\theta) - \mu(\theta)e\| \le 2\eta(1-\theta)\mu = 2\eta\mu(\theta)$$

or $(x(\theta), s(\theta)) \in \mathcal{N}_2(2\eta)$ for

$$0 \le \theta \le \frac{2}{1 + \sqrt{1 + 4\|Pq/\mu\|/\eta}}.$$

□

We can now show

Theorem 4.18 *Let $\eta = 1/4$. Then Algorithm 4.5 will terminate in at most $O(\sqrt{n}\log((x^0)^T s^0/\epsilon))$ iterations with*

$$c^T x^k - b^T y^k \le \epsilon.$$

Proof. Using lemma 4.14(i) and lemma 4.15(i), we have

$$\|Pq\| \le \frac{\sqrt{2}}{4}\|r\|^2 = \frac{\sqrt{2}}{4} n\mu,$$

so that

$$\bar{\theta} \ge \frac{2}{1 + \sqrt{1 + \sqrt{2}n/\eta}} = \frac{2}{1 + \sqrt{1 + 4\sqrt{2}n}}$$

at each iteration. Then (4.26) and lemma 4.17 imply that

$$\mu^{k+1} \le \left(1 - \frac{2}{1 + \sqrt{1 + 4\sqrt{2}n}}\right)\mu^k$$

for each k. This yields the desired result.

□

4.5.2 Wide-neighborhood algorithm

In this section we consider algorithms of the following form based on $\gamma \in (0,1)$ and $\mathcal{N}$, where $\mathcal{N}$ is a wide neighborhood of either $\mathcal{N}_\infty$ or $\mathcal{N}_\infty^-$.

Algorithm 4.6 *Let $\eta \in (0,1)$ and $\gamma \in (0,1)$ with $\gamma \le 2(1-\eta)$. Given $(x^0, s^0) \in \mathcal{N}(\eta)$. Set $k := 0$.*
While $(x^k)^T s^k > \epsilon$ **do**:

1. *Set $(x, s) = (x^k, s^k)$ and compute $d = d(x, s, \gamma)$ from (4.17).*

2. Compute the largest $\bar\theta$ *so that*

$$(x(\theta), s(\theta)) \in \mathcal{N} \quad \textit{for} \quad \theta \in [0, \bar\theta];$$

set $(x^{k+1}, s^{k+1}) = (x(\bar\theta), s(\bar\theta))$.

3. Set $k := k+1$ *and return to Step 1.*

Again the selection of $\bar\theta$ makes this an adaptive-step method. We will analyze this algorithm for $\mathcal{N} = \mathcal{N}_\infty(\eta)$ and $\mathcal{N}_\infty^-(\eta)$, where $\eta \in (0,1)$. In either case, computing $\bar\theta$ involves the solution of at most $2n$ single-variable quadratic equations.

Note that, if $\mu^k := (x^k)^T s^k/n$, (4.20) implies

$$\mu^{k+1} = (1 - \bar\theta(1-\gamma))\mu^k, \tag{4.27}$$

so we wish to bound $\bar\theta$ from below.

Lemma 4.19 *Let* $\eta \in (0,1)$, $\gamma \in (0,1)$, *and* $\mathcal{N} = \mathcal{N}_\infty(\eta)$ *or* $\mathcal{N}_\infty^-(\eta)$. *Let* x, s, d *and* $\bar\theta$ *be as in the kth iteration of Algorithm 4.6, and define* p, q *and* r *by (4.22). Then*

$$\bar\theta \geq \theta_2 := \min\left\{1, \frac{\eta\gamma\mu^k}{\|Pq\|_\infty}\right\} \quad \textit{if } \mathcal{N} = \mathcal{N}_\infty(\eta),$$

$$\bar\theta \geq \theta_2^- := \min\left\{1, \frac{\eta\gamma\mu^k}{\|Pq\|_\infty^-}\right\} \quad \textit{if } \mathcal{N} = \mathcal{N}_\infty^-(\eta).$$

Proof. Suppose first $\mathcal{N} = \mathcal{N}_\infty^-(\eta)$. Then, for each $\theta \in [0, \theta_2^-]$, (4.20) and (4.21) imply

$$\begin{aligned} & X(\theta)s(\theta) - \mu(\theta)e \\ & \quad = \quad (1-\theta)(Xs - \mu e) + \theta^2 Pq, \\ & \quad \geq \quad -\left((1-\theta)\|Xs - \mu e\|_\infty^- + \theta^2\|Pq\|_\infty^-\right)e \\ & \quad \geq \quad -\left((1-\theta)\eta\mu^k + \theta\eta\gamma\mu^k\right)e \\ & \quad = \quad -\eta\mu(\theta)e. \end{aligned}$$

Hence, as in the proof of lemma 4.16, $(x(\theta), s(\theta)) \in \mathcal{N}_\infty^-(\eta)$ for $\theta \in [0, \theta_2^-]$, when $\bar\theta \geq \theta_2^-$. If $\mathcal{N} = \mathcal{N}_\infty(\eta)$, a similar proof gives

$$\eta\mu(\theta)e \geq X(\theta)s(\theta) - \mu(\theta)e \geq -\eta\mu(\theta)e$$

for $\theta \in [0, \theta_2]$, which again implies $\bar\theta \geq \theta_2$.

□

We can now prove

Theorem 4.20 *Let $\eta \in (0,1)$ and $\gamma \in (0,1)$ be constants with $\gamma \leq 2(1-\eta)$. Then Algorithm 4.6, with $\mathcal{N} = \mathcal{N}_\infty(\eta)$ or $\mathcal{N}_\infty^-(\eta)$, will terminate in $O(n \log((x^0)^T s^0/\epsilon))$ iterations.*

Proof. In either case, each iterate lies in $\mathcal{N}_\infty^-(\eta)$, whence

$$\|Pq\|_\infty^- \leq \|Pq\|_\infty \leq \|r\|^2/4 \leq n\mu^k/4,$$

using lemma 4.14(iii) and lemma 4.15(iii). Hence

$$\theta_2^- \geq \theta_2 \geq 4\eta\gamma/n.$$

Then lemma 4.19 and (4.27) give

$$\mu^{k+1} \leq \left(1 - \frac{4\eta\gamma(1-\gamma)}{n}\right)\mu^k, \tag{4.28}$$

which yields the result.

□

The algorithms for the neighborhoods $\mathcal{N}_\infty(\eta)$ and $\mathcal{N}_\infty^-(\eta)$ generate sequences of points lying in the boundaries of these sets. Since the results hold for arbitrary $\eta \in (0,1)$, the algorithms can generate sequences of points in a wide area of the feasible region. In particular,

$$\mathcal{N}_\infty^-(1) = \overset{\circ}{\mathcal{F}},$$

so when η is close to 1, the neighborhood $\mathcal{N}_\infty^-(\eta)$ spreads over almost all of the feasible region $\mathcal{F}$, and the points generated by the algorithm based on $\mathcal{N}_\infty^-(\eta)$ are close to the boundary rather than the central path.

4.6 Notes

A similar result to Proposition 4.1 has been proved by Todd [328]. This proposition plays an important role in analyzing several interior-point algorithms.

The Karmarkar projective algorithm in the LP standard form with a lower bound z for z^* was first developed and analyzed by Todd and Burrell [329], Anstreicher [19], Gay [103], and Ye and Kojima [384]. de Ghellinck and Vial [104] developed a projective algorithm that has a unique feature: it does not need to start from a feasible interior point. All of these algorithms have an iteration complexity $O(nL)$. Other extensions and analyses of Karmarkar's algorithm can be found in Akgül [7], Anstreicher [17], Asic, Kovacevic-Vujcic and Radosavljevic-Nikolic [29], Betke and Gritzmann [47], Blair [52], Blum [54], Dennis, Morshedi and Turner [75], Diao [77], Gonzaga [123], Kalantari [168], Kovacevic–Vujcic [191], Lagarias [195], McDiarmid [211], Nemirovskii [256], Nesterov [260], Padberg [270], Sherali [301], Shub [304], Tseng [336], Wei [365], Wu and Wu [368], Zimmermann [395], etc.

The path-following algorithm, described in Section 4.2, is a variant of Renegar [286]. The difference is the analysis used in proving the complexity bound. Renegar measured the duality-gap, while we used the max-potential of the level set. A primal path-following algorithm is independently analyzed by Gonzaga [122]. Both Gonzaga [122] and Vaidya [348] developed a rank-one updating technique in solving the Newton equation of each iteration, and proved that each iteration uses $O(n^{2.5})$ arithmetic operations on average. Kojima, Mizuno and Yoshise [183] and Monteiro and Adler [239] developed a symmetric primal-dual path-following algorithm with the same iteration and arithmetic operation bounds. The algorithm was proposed earlier by Tanabe [316]. Other variants of path-following or homotopy algorithms can be found in Blum [55], Boggs, Domich, Donaldson and Witzgall [57], Nazareth [253, 254], etc.

Recently, Vaidya [349] developed a new center, the volumetric center, for linear inequalities and a path-following algorithm for convex programming. The arithmetic operations complexity bound is identical to that of the ellipsoid method, but its iteration complexity bound is less than that of the ellipsoid method. Also see Anstreicher [21].

The primal potential function with $\rho > 1$ and the affine potential reduction algorithm were developed by Gonzaga [124]. His algorithm has iteration complexity $O(nL)$. The primal-dual potential function and algorithm were analyzed by Anstreicher and Bosch [23], Freund [96], Gonzaga [124], Todd and Ye [330], and Ye [373]. These algorithms possess $O(\sqrt{n}L)$ iteration complexity. Using this function, Ye [375] further developed a projective algorithm with $O(\sqrt{n}L)$ iteration complexity, also see Goldfarb and Xiao [117].

The primal-dual potential reduction algorithm described in Section 4.3 is in the primal form. One can develop a potential reduction algorithm in dual form, where z, an upper bound for the optimal objective value z^*, is updated down in each iteration, see Ye [373]. The symmetric primal-dual potential algorithm of Section 4.4 was proposed by Kojima, Mizuno and Yoshise [185]. Other potential reduction algorithms are by Gonzaga and Todd [130], Huang and Kortanek [151], and Tuncel [345]. Todd [324] proposed an extremely simple and elegant $O(\sqrt{n}L)$ algorithm.

The adaptive primal-dual algorithms were developed by Mizuno, Todd and Ye [233], also see Barnes, Chopra and Jensen [36]. A more practical predictor-corrector algorithm was proposed by Mehrotra [221], based on the power series algorithm of Bayer and Lagarias [39] and the primal-dual version of Monteiro, Adler and Resende [242], also see Carpenter, Lustig, Mulvey and Shanno [63] and Zhang and Zhang [393]. His technique has been used in almost all of the LP interior-point implementations. Furthermore, Hung [155] developed a $O(n^{\frac{n+1}{2n}}L)$-iteration variant that uses wider neighborhoods. As n becomes large, this bound approaches the best bound for linear programming algorithms that use the small neighborhood (which are not practical). Other polynomial wide-neighborhood algorithms can be found in Jansen [158] and Sturm and Zhang [311].

There was another polynomial interior-point algorithm, a multiplicative barrier function method, which was developed by Iri and Imai [156], also see Sturm and Zhang [310].

Another popular interior-point algorithm, called the affine scaling algorithm, was developed by Dikin [78, 79] and was rediscovered by Barnes [35], Cavalier and Soyster [64], Kortanek and Shi [188], Sherali, Skarpness and Kim [302], Vanderbei and Lagarias [357], Vanderbei, Meketon and Freedman [358], Andersen [15], and Monteiro, Adler and Resende [242]. The algorithm also has three forms as the potential algorithm has; the difference is that ρ is chosen as ∞ in the direction of the affine scaling algorithm. The primal or dual algorithm has no polynomial complexity bound yet, but has been proved convergent under very weak conditions, see Tsuchiya [339, 340], Tsuchiya and Muramatsu [342], Monteiro, Tsuchiya and Wang [246], Saigal [295], Sun [312], and Tseng and Luo [338]. Mascarenhas [210] provided a divergence example for the affine scaling algorithm. Polynomial affine-scaling-type algorithms can be found in Monteiro, Adler and Resende [242] and Jansen, Roos and Terlaky [159].

A modified (shifted) barrier function theory and methods were developed by Polyak [277], also see Pan [271], and Polak, Higgins and Mayne [276].

Interior-point algorithm computational results can be found in Adler, Karmarkar, Resende and Veiga [4], Altman and Gondzio [12], Bixby, Gregory, Lustig, Marsten and Shanno [50], Choi, Monma and Shanno [66, 67], Christiansen and Kortanek [68], Domich, Boggs, Rogers and Witzgall [81], Fourer and Mehrotra [93], Gondzio [120, 121], Lustig, Marsten and Shanno [203, 206, 204, 205], McShane, Monma and Shanno [214], Mehrotra [221], Monma [235], Ponnambalam, Vannelli and Woo [279], Vanderbei [355], and Xu, Hung and Ye [369].

There are several comprehensive books which cover interior-point linear programming algorithms. They are Bazaraa, Jarvis and Sherali [40], Fang and Puthenpura [88], den Hertog [145], Saigal [296], Murty [250], etc.

Many researchers have applied interior-point algorithms to solving convex QP and monotone LCP problems. The algorithms can be divided into three groups: the primal scaling algorithm, see Anstreicher, den Hertog, Roos and Terlaky [24], Ben–Daya and Shetty [42], Goldfarb and Liu [114], Ponceleon [278], and Ye [387, 374]; the dual scaling algorithm, see Jarre [160], Mehrotra and Sun [222], Nesterov and Nemirovskii [262], and Renegar and Shub [288]; and the primal-dual scaling algorithm, see Kojima, Mizuno and Yoshise [184, 185, 183], Mizuno [229], Monteiro and Adler [240], Monteiro, Adler and Resende [242], and Vanderbei [354].

Relations among these algorithms can be seen in den Hertog and Roos [147]. Given an interior point (x, y, s), the following is a summary of directions generated by the three potential algorithms. They all satisfy

$$Ad_x = 0, \quad d_s = -A^T d_y \quad \text{for LP},$$

$$Ad_x = 0, \quad d_s = Qd_x - A^T d_y \quad \text{for QP},$$

and

$$d_s = Md_x \quad \text{for LCP}.$$

Furthermore, they satisfy, respectively,

Primal:

$$d_s + \frac{x^T s}{(n+\rho)} X^{-2} d_x = -s + \frac{x^T s}{(n+\rho)} X^{-1} e,$$

Dual:

$$d_x + \frac{x^T s}{(n+\rho)} S^{-2} d_s = -x + \frac{x^T s}{(n+\rho)} S^{-1} e,$$

and

Primal-dual:

$$Xd_s + Sd_x = -Xs + \frac{x^T s}{(n+\rho)} e,$$

where $\rho \geq \sqrt{n}$. These algorithms will reduce the primal-dual potential function by a constant, leading to $O(\rho L)$ iteration complexity.

4.7 Exercises

4.1 *Prove inequality (4.3).*

4.2 *Let* $x(z) \in \mathcal{F}_p$ *be on the central path associated with* $\Omega(z)$ *in Proposition 4.1. Then,* $z^* > z \geq z^0$ *implies* $\mu(z) \leq \mu(z^0)$.

4.3 *Prove Corollary 4.2.*

4.4 *Let* $\bar{z}$ *be chosen by (4.9) in Karmarkar's algorithm. Then show*

$$\|p(b^T y(\bar{z}))\| \geq \frac{c^T x^k - z^k}{n+1}.$$

4.5 *Develop a potential reduction algorithm in dual form, with* z *as a upper bound for the optimal objective value* z^*.

4.6 *Prove Lemma 4.12.*

4.7 *Prove Lemma 4.14.*

4.8 *Describe the primal affine scaling algorithm mentioned at the end of Section 4.6. Starting from* $x = e$, *use it to complete the first three iterations for solving*

$$\begin{array}{ll} \textit{minimize} & x_1 + 3x_2 \\ \textit{subject to} & x_1 + x_2 + x_3 = 3, \\ & x_1, x_2, x_3 \geq 0. \end{array}$$

Chapter 5

Worst-Case Analysis

There are several remaining key issues concerning interior-point algorithms for LP. The first is the arithmetic operation complexity. In the previous chapters, we have analyzed the total number of iterations needed to solve a LP problem approximately. Since it solves a system of linear equations with dimension $m+n$ and $n \geq m$, each iteration of all interior-point algorithms uses $O(n^3)$ arithmetic operations. Thus, their best operation complexity bound is $O(n^{3.5}\log(R/\epsilon))$, when the initial gap $(x^0)^T s^0 \leq R$. (We assume $(x^0)^T s^0 \leq R$ throughout this section.) The question is whether or not the arithmetic operations can be reduced in each iteration.

The second issue involves termination. Unlike the simplex method for linear programming which terminates with an exact solution, interior-point algorithms are continuous optimization algorithms that generate an infinite solution sequence converging to the optimal solution set. If the data of an LP instance are integral or rational, an argument is made that, after the worst-case time bound, an exact solution can be "rounded" from the latest approximate solution. Thus, several questions arise. First, under the real number computation model (i.e., the LP data consists of real numbers) how do we argue termination to an exact solution? Second, regardless of the data's status, can we utilize a practical test, one which can be computed cost-effectively during the iterative process, to identify an exact solution so that the algorithm can be terminated before the worse-case time bound? Here, the exact solution means a mathematical solution form using exact arithmetic, such as the solution of a system of linear equations or the solution of a least-squares problem, which can be computed in a number of arithmetic operations bounded by a polynomial in n.

The third issue involves initialization. Almost all interior-point algorithms solve the LP problem under the regularity assumption that $\overset{\circ}{\mathcal{F}} \neq \emptyset$. A related issue is that interior-point algorithms have to start at a strictly feasible point. Since no prior knowledge is usually available, one way is to explicitly bound the feasible region of the LP problem by a big M number. If the LP problem has integral data, this number can be set to 2^L in the worst case, where L is the length of the LP data in binary form. This setting is impossible in solving large problems. Moreover, if the LP problem has real data, no computable bound is known to set up the big M.

5.1 Arithmetic operation

The primary computation cost of each iteration of interior-point algorithms is to inverse a normal matrix AX^2A^T in the primal form, $AS^{-2}A^T$ in the dual form, or $AXS^{-1}A^T$ in the primal-dual form. In this section, we show how to use a rank-one technique to update the inverse of the matrix during the iterative progress. This can be described as follows.

Consider the primal potential reduction algorithm in Section 4.3. Replacing X^k in the normal matrix by a positive diagonal matrix D such that

$$\frac{1}{1.1} \leq \frac{d_j}{x_j^k} \leq 1.1 \quad \text{for} \quad j = 1, \ldots, n,$$

we now have

$$x^{k+1} - x^k = -\alpha \frac{D\hat{p}(z^k)}{\|\hat{p}(z^k)\|},$$

where

$$\hat{p}(z^k) = (I - DA^T(AD^2A^T)^{-1}AD)D\nabla\mathcal{P}_{n+\rho}(x^k, z^k).$$

Then

$$\nabla\mathcal{P}_{n+\rho}(x^k, z^k)^T(x^{k+1} - x^k) = -\beta\|\hat{p}(z^k)\|,$$

and the reduction of the potential function is

$$\mathcal{P}_{n+\rho}(x^{k+1}, z^k) - \mathcal{P}_{n+\rho}(x^k, z^k) \leq -\alpha\|\hat{p}(z^k)\| + \frac{(1.1\alpha)^2}{2(1-1.1\alpha)},$$

since

$$\begin{aligned} \|(X^k)^{-1}(x^{k+1} - x^k)\| &= \|(X^k)^{-1}DD^{-1}(x^{k+1} - x^k)\| \\ &\leq \|(X^k)^{-1}D\|\|D^{-1}(x^{k+1} - x^k)\| \\ &\leq 1.1\|D^{-1}(x^{k+1} - x^k)\| = 1.1\alpha. \end{aligned}$$

Note that $\hat{p}(z^k)$ can be written as

$$\hat{p}(z^k) = \frac{n+\rho}{c^T x^k - z^k} D s(z^k) - D(X^k)^{-1} e = D(X^k)^{-1} p(z^k), \tag{5.1}$$

where the expressions of $p(z^k)$ and $s(z^k)$ are again given by (4.11) and (4.12) with

$$y(z^k) = (AD^2A^T)^{-1} AD(Dc - \frac{(s^k)^T x^k}{n+\rho} D(X^k)^{-1} e). \tag{5.2}$$

Thus, we have

$$\|\hat{p}(z^k)\| = \|D(X^k)^{-1} p(z^k)\| \geq \|p(z^k)\| / \|D^{-1} X^k\| \geq \|p(z^k)\| / 1.1.$$

Noting that Lemma 4.8 still holds for $p(z^k)$, we only need to modify the potential reduction inequality in the proof of Theorem 4.9 by

$$\mathcal{P}_{n+\rho}(x^{k+1}, z^k) - \mathcal{P}_{n+\rho}(x^k, z^k) \leq \frac{\alpha}{1.1} \min(\beta \sqrt{\frac{n}{n+\beta^2}}, 1-\beta) + \frac{(1.1\alpha)^2}{2(1-1.1\alpha)}.$$

Therefore, upon choosing $\beta = 0.43$ and $\alpha = 0.25$, Theorem 4.9 is still valid for $\delta > 0.04$. As a result, the following modified primal algorithm can be developed.

Algorithm 5.1 *Given $x^0 \in \overset{\circ}{\mathcal{F}}_p$ and $(y^0, s^0) \in \overset{\circ}{\mathcal{F}}_d$. Let $z^0 = b^T y^0$ and $D^0 = X^0$. Set $\alpha = 0.25$ and $k := 0$;*
while $(s^k)^T x^k \geq \epsilon$ **do**

1. *For $j = 1, \ldots, n$, if $d_j^k / x_j^k \notin [1/1.1, 1.1]$ then $d_j^k = x_j^k$. Let $D := D^k$. Then, compute $y(z^k)$ of (5.2), $s(z^k)$ of (4.12), $p(z^k)$ of (4.11) and $\hat{p}(z^k)$ of (5.1).*
2. *Set $y^{k+1} = y(\bar{z})$, $s^{k+1} = s(\bar{z})$, $z^{k+1} = b^T y^{k+1}$ where*
$$\bar{z} = \arg \min_{z \geq z^k} \psi_{n+\rho}(x^k, s(z)).$$
3. *Let $x^{k+1} = x^k - \alpha D \hat{p}(z^{k+1}) / \|\hat{p}(z^{k+1})\|$.*
4. *Set $D^{k+1} = D^k$ and $k := k+1$, and return to Step 1.*

The inverse of the normal matrix AD^2A^T in (5.2) can be calculated using a rank-one updating technique whenever d_j is changed, and each update uses $O(n^2)$ arithmetic operations. Now we estimate the total number of updates needed before the algorithm terminates. Let

$$I^t = \left\{ i : \; \frac{d_j^t}{x_j^t} \notin [1/1.1, 1.1] \right\}.$$

Then, the computation work in the tth iteration is $O(n^2|I^t|)$ operations, where $|I|$ denotes the number of elements in set I. Thus, the total operations up to the kth iteration is $O(n^3 + n^2 \sum_{t=1}^{k} |I^t|)$, where n^3 in the estimate is the amount of work at the initial iteration $t = 0$. We have the following lemma to bound this estimate.

Lemma 5.1 *In Algorithm 5.1,*

$$\|(D^t)^{-1}(x^t - d^t)\|_\infty \le 0.1 \quad \textit{for any} \quad t = 0, 1, ...$$

and

$$\sum_{t=1}^{k} |I^t| \le 11 \sum_{t=1}^{k} \|(D^{t-1})^{-1}(x^t - x^{t-1})\|_1.$$

Proof. The proof of the first inequality is straightforward. Let

$$\sigma^t = \|(D^t)^{-1}(x^t - d^t)\|_1 \quad \text{for} \quad t = 0, 1, ...$$

Then, for $t = 1, 2, ...$

$$\begin{aligned}
\sigma^t &= \sum_{i \in I^t} \frac{|x_i^t - d_i^t|}{d_i^t} + \sum_{i \notin I^t} \frac{|x_i^t - d_i^t|}{d_i^t} \\
&= \sum_{i \notin I^t} \frac{|x_i^t - d_i^{t-1}|}{d_i^{t-1}} \\
&\le \sum_{i \in I^t} \left(\frac{|x_i^t - d_i^{t-1}|}{d_i^{t-1}} - \frac{1}{11}\right) + \sum_{i \notin I^t} \frac{|x_i^t - d_i^{t-1}|}{d_i^{t-1}} \\
&= \|(D^{t-1})^{-1}(x^t - d^{t-1})\|_1 - |I^t|/11 \\
&= \|(D^{t-1})^{-1}(x^t - x^{t-1} + x^{t-1} - d^{t-1})\|_1 - |I^t|/11 \\
&\le \|(D^{t-1})^{-1}(x^t - x^{t-1})\| + \|(D^{t-1})^{-1}(x^{t-1} - d^{t-1})\|_1 - |I^t|/11 \\
&= \|(D^{t-1})^{-1}(x^t - x^{t-1})\|_1 + \sigma^{t-1} - |I^t|/11.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\sum_{t=1}^{k} |I^t|/11 &\le \sum_{t=1}^{k} \left(\|(D^{t-1})^{-1}(x^t - x^{t-1})\|_1 + \sigma^{t-1} - \sigma^t\right) \\
&= \sigma^0 - \sigma^k + \sum_{t=1}^{k} \|(D^{t-1})^{-1}(x^t - x^{t-1})\|_1.
\end{aligned}$$

Since $\sigma^0 = 0$ and $\sigma^k \ge 0$, we have the desired result.

$\square$

From the update in Step 3 of Algorithm 5.1 we have

$$\|(D^{t-1})^{-1}(x^t - x^{t-1})\| \le \alpha,$$

for $t = 1, 2, ...$, which implies that

$$\|(D^{t-1})^{-1}(x^t - x^{t-1})\|_1 \le \alpha\sqrt{n}.$$

Hence

$$\sum_{t=1}^{k} |I^t| \le 11k\alpha\sqrt{n}.$$

If k is about $O(\sqrt{n}\log(R/\epsilon))$, then the total number of rank-one updates is bounded by $O(n\log(R/\epsilon))$ and the total number of operations by $O(n^3\log(R/\epsilon))$. To summarize, we have

Theorem 5.2 *Let $\rho = \sqrt{n}$ and $\psi_{n+\rho}(x^0, s^0) \le O(\sqrt{n}\log R)$. Then, Algorithm 5.1 terminates in $O(\sqrt{n}\log(R/\epsilon))$ iterations and uses $O(n^3\log(R/\epsilon))$ total arithmetic operations.*

5.2 Termination

We now turn our attention to the termination of interior-point algorithms, the object of a great deal of research efforts. These efforts resulted in four basic approaches.

- The first is the standard purification procedure to find a feasible vertex whose objective value is at least as good as the current interior point. This approach can be done in strongly polynomial time when using the simplex method, and it works for LP with real number data. One difficulty which arises with this method is that many non-optimal vertices may be close to the optimal face, and the simplex method still requires many pivot steps for some "bad" problems. Moreover, the (simplex) purification procedure is *sequential* in nature.

- The second is a theoretical effort to identify an optimal basis. A test procedure was also developed to show that, if the LP problem is nondegenerate, the unique optimal basis can be identified before the worst-case time bound. The procedure seemed to work fine for some LP problems but it has difficulty with degenerate LP problems. Unfortunately, most real LP problems are degenerate. The difficulty arises simply because any degenerate LP problem has multiple optimal bases.

- The third approach is to slightly randomly perturb the data such that the new LP problem is nondegenerate and its optimal basis remains one of the optimal bases of the original LP problem. Questions remain on how and when to randomize the data during the iterative process, decisions which significantly affect the success of the effort.
- The fourth approach is to guess the optimal face and to find a feasible solution on the face. It consists of two phases: the first phase uses interior point algorithms to identify the complementarity partition (P^*, Z^*), and the second phase solves two linear feasibility problems to find the optimal primal and dual solutions. One can use the solution resulting from the first phase as a starting point for the second phase.

In this section we develop a termination procedure to obtain an exact solution on the interior of the optimal face. We shall see that (i) the termination criterion is guaranteed to work in finite time, (ii) the projection procedure (solving a least-squares problem) is strongly polynomial and can be efficiently performed in *parallel*, and (iii) the approach identifies the optimal face, which is useful in sensitivity analysis.

It has been noticed in practice that many interior-point algorithms generate a sequence of solutions converging to a strictly complementary solution for linear programming. It was subsequently proved that numerous interior-point algorithms for linear programming indeed generate solution sequences that converge to strictly complementary solutions, or interior solutions on the optimal face. Recall that the primal optimal face is

$$\Omega_p = \{x_{P^*} : A_{P^*}x_{P^*} = b,\ x_{P^*} \geq 0\},$$

and the one for the dual is

$$\Omega_d = \{(y, s_{Z^*}) : A_{P^*}^T y = c_{P^*},\quad s_{Z^*} = c_{Z^*} - A_{Z^*}^T y \geq 0\},$$

where (P^*, Z^*) is the strict complementarity partition of the LP problem. Note that these faces have strictly feasible solutions. Define

$$\begin{array}{rcl} \xi_p(A,b,c) & := & \min_{j\in P^*}\{\max_{x_{P^*}\in\Omega_p} x_j\} > 0, \\ \xi_d(A,b,c) & := & \min_{j\in Z^*}\{\max_{(y,s_{Z^*})\in\Omega_d} s_j\} > 0, \\ \xi(A,b,c) & := & \min\{\xi_p(A,b,c), \xi_d(A,b,c)\} > 0. \end{array} \tag{5.3}$$

5.2.1 Project an interior point onto the optimal face

To measure the magnitude of positivity of a point $x \in \mathcal{R}^n_+$, we let $\sigma(x)$ be the support, i.e., index set of positive components in x, that is,

$$\sigma(x) \ = \ \{i : x_i > 0\}\,.$$

We first prove the following theorem.

Theorem 5.3 *Given an interior solution x^k and s^k in the solution sequence generated by any of interior-point algorithms possessing property (5.7) below, define*

$$\sigma^k = \{j : x_j^k \geq s_j^k\}.$$

Then, for $K = O(\sqrt{n}(\log(R/\xi^2(A,b,c)) + \log n))$ we have

$$\sigma^k = P^* \quad \textit{for all} \quad k \geq K.$$

Proof. For simplicity, we use $\xi = \xi(A,b,c)$. For a given $j \in P^*$, let (x^*, s^*) be a complementarity solution such that x_j^* is maximized on the primal optimal face Ω_p, i.e, $x_j^* \geq \xi_p(A,b,c) \geq \xi$. Since

$$(x^k - x^*)^T(s^k - s^*) = 0,$$

$$\sum_{i \in P^*} x_i^* s_i^k + \sum_{i \in Z^*} s_i^* x_i^k = (x^k)^T s^k. \tag{5.4}$$

Thus, if $(x^k)^T s^k < \epsilon$, then

$$s_j^k \leq \frac{(x^k)^T s^k}{x_j^*} < \epsilon/\xi. \tag{5.5}$$

On the other hand, inequality (5.4) can be written as

$$\sum_{i \in P^*} \frac{x_i^*}{x_i^k}(x_i^k s_i^k) + \sum_{i \in Z^*} \frac{s_i^*}{s_i^k}(x_i^k s_i^k) = (x^k)^T s^k. \tag{5.6}$$

Almost all interior-point algorithms generate a solution sequence (or subsequence) (x^k, s^k) such that

$$\frac{\min(X^k s^k)}{(x^k)^T s^k} > n^{-\alpha}, \tag{5.7}$$

where α is a positive constant. Thus, from (5.6) and (5.7) we have

$$\frac{x_j^*}{x_j^k}(x_j^k s_j^k) < (x^k)^T s^k,$$

or

$$\frac{x_j^k}{x_j^*} > \frac{x_j^k s_j^k}{(x^k)^T s^k} \geq n^{-\alpha},$$

or

$$x_j^k > n^{-\alpha} x_j^* \geq n^{-\alpha}\xi. \tag{5.8}$$

Thus, if $\epsilon \leq n^{-\alpha}\xi^2$, recalling (5.5) we must have

$$s_j^k < \epsilon/\xi \leq n^{-\alpha}\xi < x_j^k.$$

Similarly, we can prove this result for each $j \in P^*$. Moreover, for each $j \in Z^*$,

$$x_j^k < \epsilon/\xi \leq n^{-\alpha}\xi < s_j^k.$$

Due to the polynomial bound of these interior-point algorithms, in $O(\sqrt{n}(\log(R/\xi^2) + \log n))$ iterations we shall have

$$(x^k)^T s^k \leq \epsilon = n^{-\alpha}\xi^2.$$

This concludes the proof of the theorem.

$\square$

In practice, we don't wait to see if $\sigma^k = P^*$. In the following we develop a projection procedure to test if $\sigma^k = P^*$ and if an exactly optimal solution can be reached. For simplicity, let $P = \sigma^k$ and the rest be Z. Then we solve

$$\begin{array}{lll} (PP) & \text{minimize} & \|(X_P^k)^{-1}(x_P - x_P^k)\| \\ & \text{subject to} & A_P x_P = b, \end{array}$$

and

$$\begin{array}{lll} (DP) & \text{minimize} & \|(S_Z^k)^{-1}A_Z^T(y - y^k)\| \\ & \text{subject to} & A_P^T y = c_P. \end{array}$$

Without loss of generality we assume that A_P has full row rank. These two problems can be solved as two least-squares problems. The amount of work is equivalent to the computation in one iteration of interior-point algorithms. Furthermore, if the resulting solutions x_P^* and y^* satisfy

$$x_P^* > 0 \quad \text{and} \quad s_Z^* = c_Z - A_Z^T y^* > 0,$$

then obviously $x^* = (x_P^*, 0)$ and y^* are (exact) optimal solutions for the original LP problems, and $\sigma^k = P^*$ (Figure 5.1).

Let $d_x = (X_P^k)^{-1}(x_P - x_P^k)$ and $d_y = y - y^k$ and $d_s = (S_Z^k)^{-1}A_Z^T(y - y^k)$. Then, the two problems can be rewritten as

$$\begin{array}{lll} (PP) & \text{minimize} & \|d_x\| \\ & \text{subject to} & A_P X_P^k d_x = b - A_P x_P^k = A_Z x_Z^k, \end{array}$$

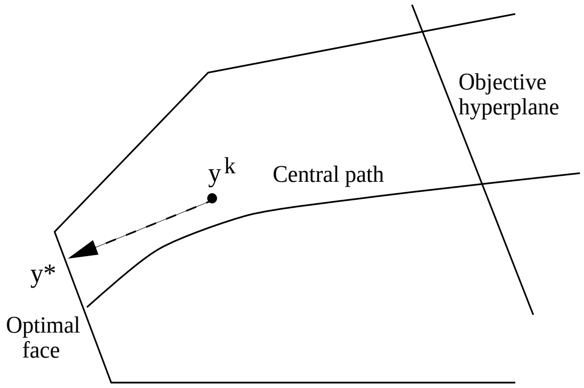

Figure 5.1: Illustration of the projection of y^k onto the dual optimal face

and

$$(DP) \quad \begin{array}{ll} \text{minimize} & \|(S_Z^k)^{-1} A_Z^T d_y\| \\ \text{subject to} & A_P^T d_y = c_P - A_P^T y^k = s_P^k. \end{array}$$

Thus, if both $\|d_x\|_\infty$ and $\|d_s\|_\infty$ are less than 1, we must have $x_P > 0$ and $s_Z = c_Z - A_Z^T y > 0$. Since $\sigma^k \to P^*$ from the proof of Theorem 5.3 and the right-hand sides of the constraints of both problems converge to zero as $(x^k)^T s^k \to 0$, $\|d_x\|$ and $\|d_s\|$ must converge to zero. After $\sigma^k = P = P^*$, both (PP) and (PD) are feasible. Note that the solutions of (PP) and (PD) are

$$d_x^* = X_P^k A_P^T (A_P (X_P^k)^2 A_P^T)^{-1} A_Z x_Z^k \text{ and } d_s^* = (S_Z^k)^{-1} A_Z^T (A_P A_P^T)^{-1} A_P s_P^k.$$

Thus,

$$\begin{array}{rl} \|d_x^*\|_\infty & \\ = & \|X_P^k A_P^T (A_P (X_P^k)^2 A_P^T)^{-1} A_P X_P^k (X_P^k)^{-1} A_P^T (A_P A_P^T)^{-1} A_Z x_Z^k\|_\infty \\ \le & \|X_P^k A_P^T (A_P (X_P^k)^2 A_P^T)^{-1} A_P X_P^k\| \|(X_P^k)^{-1} A_P^T (A_P A_P^T)^{-1} A_Z x_Z^k\|_\infty \\ \le & \|(X_P^k)^{-1} A_P^T (A_P A_P^T)^{-1} A_Z x_Z^k\|_\infty \\ \le & \|(X_P^k)^{-1}\| \|A_P^T (A_P A_P^T)^{-1} A_Z\| \|x_Z^k\|_\infty. \end{array}$$

Let

$$\zeta(A, b, c) = \max(1, \|A_{P^*}^T (A_{P^*} A_{P^*}^T)^{-1} A_{Z^*}\|).$$

Then, if $\sigma^k = P = P^*$ and $\min(x_P^k) > \zeta(A, b, c) \max(x_Z^k)$,

$$\|d_x^*\|_\infty < 1 \quad \text{which implies that} \quad x_P^* > 0.$$

Similarly, when $\min(s_Z^k) > \zeta(A,b,c)\max(s_P^k)$ then

$$\|d_s^*\|_\infty \le 1 \quad \text{which implies that} \quad s_P^* > 0.$$

Recall (5.5) that each component of x_Z^k and s_P^k is less than $\epsilon/\xi(A,b,c)$ and (5.8) that each component of x_P^k and s_Z^k is greater than $\xi(A,b,c)n^{-\alpha}$. This essentially shows that in

$$O\left(\sqrt{n}\left(\log(\frac{R\zeta(A,b,c)}{\xi^2(A,b,c)}) + \log n\right)\right)$$

iterations the above projection procedure will succeed in generating the complementarity partition and an exact optimal solution pair. To summarize, we have a condition-based complexity bound.

Theorem 5.4 *All $O(\sqrt{n}\log(R/\epsilon))$-iteration polynomial-time interior-point algorithms discussed earlier, coupled with the termination procedure, will generate an optimal solution in $O(\sqrt{n}(\log(R\zeta(A,b,c)/\xi^2(A,b,c)) + \log n))$ iterations and $O(n^3(\log(R\zeta(A,b,c)/\xi^2(A,b,c)) + \log n))$ arithmetic operations. If the LP problem has integral or rational data, then*

$$R \le 2^L, \quad \zeta(A,b,c) \le 2^L, \quad \text{and} \quad \xi(A,b,c) \ge 2^{-L},$$

where L is the size of the LP data. Thus, $\sigma^k = P^*$ and an exact solution will be generated in $O(\sqrt{n}L)$ iterations and $O(n^3L)$ operations.

When an interior solution x_P^*, $P = P^*$, on the primal optimal face is obtained, it can be cornered to a basic solution in no more than $n - m$ pivot operations. For example, if $|\sigma(x_P^*)| \le m$, then x^* is a basic solution; otherwise, we do the following. Find any nonzero direction d_P in the null space of P, i.e.,

$$A_P d_P = 0.$$

(If the secondary objective exists, we may set d_P as the projection to the null space of A_P from the secondary objective vector.) Note that the null space was already available when the test procedure was performed. Assume $\max(d_P) > 0$ (otherwise, let $d_P = -d_P$). Then assign

$$x_P^* := x_P^* - \alpha d_P,$$

where α is chosen such that $x_P^* \ge 0$ but at least one of its components becomes zero. Delete the corresponding column from A_P. This process can be continued until $|\sigma(x_P^*)| \le m$, i.e., a basic solution is obtained. Also note that the next null space can be updated from the previous one in $O(m^2)$ arithmetic operations. The total number of required pivots in the process is at most $|\sigma(x_{P^*}^*)| - m \le n - m$.

5.3 Initialization

Most interior-point algorithms have to start at a strictly feasible point. The complexity of obtaining such an initial point is the same as that of solving the LP problem itself. More importantly, a complete LP algorithm should accomplish two tasks: 1) affirmatively detect the infeasibility or unboundedness status of the LP problem, then 2) generate an optimal solution if the problem is neither infeasible nor unbounded.

Several approaches have been proposed to resolve these issues:

- Combining the primal and dual into a single linear feasibility problem. Theoretically, this approach can retain the currently best complexity result. Practically, the disadvantage of this approach is the doubled dimension of the system of equations which must be solved at each iteration.

- The big M method, i.e., add one or more artificial column(s) and/or row(s) and a huge penalty parameter M to force solutions to become feasible during the algorithm. Theoretically, this approach holds the best complexity. The major disadvantage of this approach is the numerical problems caused by the addition of coefficients of magnitude. It also makes the algorithms slow to converge. This disadvantage also occurs in the primal-dual "exterior" or "infeasible" algorithm. A polynomial complexity can be established for this approach if the LP problem possesses an optimal solution and if the initial point is set to Me. Thus, the big M difficulty even remains in these polynomial infeasible interior-point algorithms.

- Phase I-then-Phase II method, i.e., first try to find a feasible point (and possibly one for the dual problem), and then start to look for an optimal solution if the problem is feasible and bounded. Theoretically, this approach can maintain the polynomial complexity result. The major disadvantage of this approach is that the two (or three) related LP problems are solved sequentially.

- Combined Phase I-Phase II method, i.e., approach feasibility and optimality simultaneously. To our knowledge, the currently "best" complexity of this approach is $O(n \log(R/\epsilon))$. Other disadvantages of the method include the assumption of non-empty interior and/or the use of the big M lower bound. Also, the method works exclusively in either the primal or the dual form.

In this section, we present a homogeneous and self-dual (HSD) LP algorithm to overcome the difficulties mentioned above. The algorithm possesses the following features:

- It solves the linear programming problem without any regularity assumption concerning the existence of optimal, feasible, or interior feasible solutions, while it retains the currently best complexity result

- It can start at any positive primal-dual pair, feasible or infeasible, near the central ray of the positive orthant (cone), and it does not use any big M penalty parameter or lower bound.

- Each iteration solves a system of linear equations whose dimension is almost the same as that solved in the standard (primal-dual) interior-point algorithms.

- If the LP problem has a solution, the algorithm generates a sequence that approaches feasibility and optimality simultaneously; if the problem is infeasible or unbounded, the algorithm will produce an infeasibility certificate for at least one of the primal and dual problems.

5.3.1 A HSD linear program

Our algorithm is based on the construction of a homogeneous and self-dual linear program related to (LP) and (LD). We now briefly explain the two major concepts, homogeneity and self-duality, used in our construction.

In the context of interior-point algorithms, the idea of attacking a standard-form LP by solving a related homogeneous artificial linear program can be traced to many earlier works. (By a homogeneous linear program, we do not mean that all constraints must be homogeneous, or equivalently all right-hand sides zero. We allow a single inhomogeneous constraint, often called a normalizing constraint.) Karmarkar's original canonical form is a homogeneous linear program. One advantage of working in the homogeneous form is that we don't need to be concerned about the magnitude of solutions, since a solution is represented by a ray whose quality is scale-invariant. A disadvantage is that these related homogeneous problems, especially if they do not use any big M parameters, usually involve combining the primal and dual constraints and thus usually lead to algorithms requiring the solution of linear systems roughly twice as large as other methods.

Self-dual linear programs, meaning that the dual of the problem is equivalent to the primal, were introduced many years ago. We state the form of such problems, with inequality constraints, and their properties in the following result, whose proof is omitted.

Proposition 5.5 *Let $\tilde{A} \in R^{p\times p}$ be skew-symmetric, and let $\tilde{b} = -\tilde{c} \in R^p$. Then the problem*

$$\begin{array}{lll} (SDP) & \textit{minimize} & \tilde{c}^T\tilde{u} \\ & \textit{subject to} & \tilde{A}\tilde{u} \geq \tilde{b},\ \tilde{u} \geq 0, \end{array}$$

is equivalent to its dual. Suppose that (SDP) has a feasible solution $\tilde{u}$. Then $\tilde{u}$ is also feasible in the dual problem, and the two objective values sum to zero. Moreover, in this case (SDP) has an optimal solution, and its optimal value is zero.

The advantage of self-duality is that we can apply a primal-dual interior-point algorithm to solve the self-dual problem *without* doubling the dimension of the linear system solved at each iteration.

We now present a homogeneous and self-dual (artificial) linear program (HSDP) relating (LP) and (LD). Given any $x^0 > 0$, $s^0 > 0$, and y^0, we let $n^0 = (x^0)^T s^0 + 1$ and formulate

$$\begin{array}{llllllll} (HSDP) & \min & & & & n^0\theta & \\ & \text{s.t.} & & Ax & -b\tau & +\bar{b}\theta & =0, \\ & & -A^Ty & & +c\tau & -\bar{c}\theta & \geq 0, \\ & & b^Ty & -c^Tx & & +\bar{z}\theta & \geq 0, \\ & & -\bar{b}^Ty & +\bar{c}^Tx & -\bar{z}\tau & & =-n^0, \\ & & y \text{ free}, & x\geq 0, & \tau \geq 0, & \theta \text{ free}, & \end{array}$$

where

$$\bar{b} = b - Ax^0, \quad \bar{c} = c - A^Ty^0 - s^0, \quad \bar{z} = c^Tx^0 + 1 - b^Ty^0. \tag{5.9}$$

Here $\bar{b}$, $\bar{c}$, and $\bar{z}$ represent the "infeasibility" of the initial primal point, dual point, and primal-dual "gap," respectively.

Note that the top three constraints in (HSDP), with $\tau = 1$ and $\theta = 0$, represent primal and dual feasibility (with $x \geq 0$) and reversed weak duality, so that they define primal and dual optimal solutions. Making τ a homogenizing variable adds the required dual variable to the third constraint. Then, to achieve feasibility for $x = x^0, (y, s) = (y^0, s^0)$, we add the artificial variable θ with appropriate coefficients, and then the fourth constraint is added to achieve self-duality.

Denote by s the slack vector for the second constraint and by κ the slack scalar for the third constraint. Denote by $\mathcal{F}_h$ the set of all points $(y, x, \tau, \theta, s, \kappa)$ that are feasible for (HSDP). Denote by $\mathcal{F}_h^0$ the set of strictly feasible points with $(x, \tau, s, \kappa) > 0$ in $\mathcal{F}_h$. Note that by combining the constraints, we can write the last (equality) constraint as

$$(s^0)^T x + (x^0)^T s + \tau + \kappa - n^0 \theta = n^0, \tag{5.10}$$

which serves as a normalizing constraint for (HSDP). Also note that the constraints of (HSDP) form a skew-symmetric system, which is basically why it is a self-dual linear program.

With regard to the selection of (x^0, y^0, s^0), note that if x^0 (respectively, (y^0, s^0)) is feasible in (LP) ((LD)), then $\bar{b}$ ($\bar{c}$) is zero, and then every feasible solution to (HSDP) with $\tau > 0$ has x/τ feasible in (LP) ($(y, s)/\tau$ feasible in (LD)). Conversely, if $\bar{z} < 0$, then every feasible solution to (HSDP) with $\theta > 0$ and $\tau > 0$ has $c^T x - b^T y \le \bar{z}\theta < 0$, so either x/τ or $(y, s)/\tau$ must be infeasible.

Now let us denote by (HSDD) the dual of (HSDP). Denote by y' the dual multiplier vector for the first constraint, by x' the dual multiplier vector for the second constraint, by τ' the dual multiplier scalar for the third constraint, and by θ' the dual multiplier scalar for the fourth constraint. Then, we have the following result.

Theorem 5.6 *Consider problems (HSDP) and (HSDD).*

i) *(HSDD) has the same form as (HSDP), i.e., (HSDD) is simply (HSDP) with (y, x, τ, θ) being replaced by (y', x', τ', θ').*

ii) *(HSDP) has a strictly feasible point*

$$y = y^0, \quad x = x^0 > 0, \quad \tau = 1, \quad \theta = 1, \quad s = s^0 > 0, \quad \kappa = 1.$$

iii) *(HSDP) has an optimal solution and its optimal solution set is bounded.*

iv) *The optimal value of (HSDP) is zero, and*

$$(y, x, \tau, \theta, s, \kappa) \in \mathcal{F}_h \quad \textit{implies that} \quad n^0 \theta = x^T s + \tau \kappa.$$

v) *There is an optimal solution $(y^*, x^*, \tau^*, \theta^* = 0, s^*, \kappa^*) \in \mathcal{F}_h$ such that*

$$\begin{pmatrix} x^* + s^* \\ \tau^* + \kappa^* \end{pmatrix} > 0,$$

which we call a strictly self complementary solution. (Similarly, we sometimes call an optimal solution to (HSDP) a self-complementary solution; the strict inequalities above need not hold.)

Proof. In what follows, denote the slack vector and scalar in (HSDD) by s' and κ', respectively. The proof of (i) is based on the skew-symmetry of the linear constraint system of (HSDP). We omit the details. Result (ii) can be easily verified. Then (iii) is due to the self-dual property: (HSDD) is also feasible and it has non-empty interior. The proof of (iv) can be constructed as follows. Let $(y, x, \tau, \theta, s, \kappa)$ and $(y', x', \tau', \theta', s', \kappa')$ be feasible points for (HSDP) and (HSDD), respectively. Then the primal-dual gap is

$$n^0(\theta + \theta') = x^T s' + s^T x' + \tau\kappa' + \kappa\tau'.$$

Let $(y', x', \tau', \theta', s', \kappa') = (y, x, \tau, \theta, s, \kappa)$, which is possible since any feasible point $(y', x', \tau', \theta', s', \kappa')$ of (HSDD) is a feasible point of (HSDP) and vice versa. Thus, we have (iv). Note that (HSDP) and (HSDD) possess a strictly complementary solution pair: the primal solution is the solution for (HSDP) in which the number of positive components is maximized, and the dual solution is the solution for (HSDD) in which the number of positive components is maximized. Since the supporting set of positive components of a strictly complementary solution is invariant and since (HSDP) and (HSDD) are identical, the strictly complementary solution $(y^*, x^*, \tau^*, \theta^* = 0, s^*, \kappa^*)$ for (HSDP) is also a strictly complementary solution for (HSDD) and vice versa. Thus, we establish (v).

$\square$

Henceforth, we simply choose

$$y^0 = 0, \quad x^0 = e, \quad \text{and} \quad s^0 = e. \tag{5.11}$$

Then, $n^0 = n + 1$ and (HSDP) becomes

$$\begin{array}{llrrrrl}
(HSDP) & \min & & & & (n+1)\theta & \\
& \text{s.t.} & & Ax & -b\tau & +\bar{b}\theta & = 0, \\
& & -A^T y & & +c\tau & -\bar{c}\theta & \geq 0, \\
& & b^T y & -c^T x & & +\bar{z}\theta & \geq 0, \\
& & -\bar{b}^T y & +\bar{c}^T x & -\bar{z}\tau & & = -(n+1), \\
& & y \text{ free}, & x \geq 0, & \tau \geq 0, & \theta \text{ free}, &
\end{array}$$

where

$$\bar{b} = b - Ae, \quad \bar{c} = c - e, \quad \text{and} \quad \bar{z} = c^T e + 1. \tag{5.12}$$

Again, combining the constraints we can write the last (equality) constraint as

$$e^T x + e^T s + \tau + \kappa - (n+1)\theta = n + 1. \tag{5.13}$$

Since $\theta^* = 0$ at every optimal solution for (HSDP), we can see the normalizing effect of equation (5.13) for (HSDP).

We now relate optimal solutions to (HSDP) to those for (LP) and (LD).

Theorem 5.7 *Let $(y^*, x^*, \tau^*, \theta^* = 0, s^*, \kappa^*)$ be a strictly self complementary solution for (HSDP).*

i) *(LP) has a solution (feasible and bounded) if and only if $\tau^* > 0$. In this case, x^*/τ^* is an optimal solution for (LP) and $(y^*/\tau^*, s^*/\tau^*)$ is an optimal solution for (LD).*

ii) *(LP) has no solution if and only if $\kappa^* > 0$. In this case, x^*/κ^* or s^*/κ^* or both are certificates for proving infeasibility: if $c^T x^* < 0$ then (LD) is infeasible; if $-b^T y^* < 0$ then (LP) is infeasible; and if both $c^T x^* < 0$ and $-b^T y^* < 0$ then both (LP) and (LD) are infeasible.*

Proof. If (LP) and (LD) are both feasible, then they have a complementary solution pair $\bar{x}$ and $(\bar{y}, \bar{s})$ for (LP) and (LD), such that

$$(\bar{x})^T \bar{s} = 0.$$

Let

$$\alpha = \frac{n+1}{e^T \bar{x} + e^T \bar{s} + 1} > 0.$$

Then one can verify (see (5.13)) that

$$\tilde{y}^* = \alpha \bar{y}, \quad \tilde{x}^* = \alpha \bar{x}, \quad \tilde{\tau}^* = \alpha, \quad \tilde{\theta}^* = 0, \quad \tilde{s}^* = \alpha \bar{s}, \quad \tilde{\kappa}^* = 0$$

is a self-complementary solution for (HSDP). Since the supporting set of a strictly complementary solution for (HSDP) is unique, $\tau^* > 0$ at any strictly complementary solution for (HSDP).

Conversely, if $\tau^* > 0$, then $\kappa^* = 0$, which implies that

$$Ax^* = b\tau^*, \quad A^T y^* + s^* = c\tau^*, \quad \text{and} \quad (x^*)^T s^* = 0.$$

Thus, x^*/τ^* is an optimal solution for (LP) and $(y^*/\tau^*, s^*/\tau^*)$ is an optimal solution for (LD). This concludes the proof of the first statement in the theorem.

Now we prove the second statement. If one of (LP) and (LD) is in feasible, say (LD) is infeasible, then we have a certificate $\bar{x} \geq 0$ such that $A\bar{x} = 0$ and $c^T \bar{x} = -1$. Let $(\bar{y} = 0, \bar{s} = 0)$ and

$$\alpha = \frac{n+1}{e^T \bar{x} + e^T \bar{s} + 1} > 0.$$

Then one can verify (see (5.13)) that

$$\tilde{y}^* = \alpha\bar{y}, \quad \tilde{x}^* = \alpha\bar{x}, \quad \tilde{\tau}^* = 0, \quad \tilde{\theta}^* = 0, \quad \tilde{s}^* = \alpha\bar{s}, \quad \tilde{\kappa}^* = \alpha$$

is a self-complementary solution for (HSDP). Since the supporting set of a strictly complementary solution for (HSDP) is unique, $\kappa^* > 0$ at any strictly complementary solution for (HSDP).

Conversely, if $\tau^* = 0$, then $\kappa^* > 0$, which implies that $c^T x^* - b^T y^* < 0$, i.e., at least one of $c^T x^*$ and $-b^T y^*$ is strictly less than zero. Let us say $c^T x^* < 0$. In addition, we have

$$Ax^* = 0, \quad A^T y^* + s^* = 0, \quad (x^*)^T s^* = 0 \quad \text{and} \quad x^* + s^* > 0.$$

From Farkas' lemma, x^*/κ^* is a certificate for proving dual infeasibility. The other cases hold similarly.

□

From the proof of the theorem, we deduce the following

Corollary 5.8 *Let $(\bar{y}, \bar{x}, \bar{\tau}, \bar{\theta} = 0, \bar{s}, \bar{\kappa})$ be any optimal solution for (HSDP). Then if $\bar{\kappa} > 0$, either (LP) or (LD) is infeasible.*

5.3.2 Solving (HSD)

The following theorem resembles the central path analyzed for (LP) and (LD).

Theorem 5.9 *Consider problem (HSDP).*

i) *For any $\mu > 0$, there is a unique $(y, x, \tau, \theta, s, \kappa)$ in $\mathcal{F}_h^0$ such that*

$$\begin{pmatrix} Xs \\ \tau\kappa \end{pmatrix} = \mu e.$$

ii) *Let $(d_y, d_x, d_\tau, d_\theta, d_s, d_\kappa)$ be in the null space of the constraint matrix of (HSDP) after adding surplus variables s and κ, i.e.,*

$$\begin{array}{rrrrrrcl} & Ad_x & -bd_\tau & +\bar{b}d_\theta & & & = & 0, \\ -A^T d_y & & +cd_\tau & -\bar{c}d_\theta & -d_s & & = & 0, \\ b^T d_y & -c^T d_x & & +\bar{z}d_\theta & & -d_\kappa & = & 0, \\ -\bar{b}^T d_y & +\bar{c}^T d_x & -\bar{z}d_\tau & & & & = & 0. \end{array} \tag{5.14}$$

Then

$$(d_x)^T d_s + d_\tau d_\kappa = 0.$$

Proof. For any $\mu > 0$, there is a unique feasible point $(y, x, \tau, \theta, s, \kappa)$ for (HSDP) and a unique feasible point $(y', x', \tau', \theta', s', \kappa')$ for (HSDD) such that

$$Xs' = \mu e, \quad Sx' = \mu e, \quad \tau\kappa' = \mu, \quad \kappa\tau' = \mu.$$

However, if we switch the positions of $(y, x, \tau, \theta, s, \kappa)$ and $(y', x', \tau', \theta', s', \kappa')$ we satisfy the same equations. Thus, we must have

$$(y', x', \tau', \theta', s', \kappa') = (y, x, \tau, \theta, s, \kappa),$$

since (HSDP) and (HSDD) have the identical form. This concludes the proof of (i) in the theorem.

The proof of (ii) is simply due to the skew-symmetry of the constraint matrix. Multiply the first set of equations by d_y^T, the second set by d_x^T, the third equation by d_τ and the last by d_θ and add. This leads to the desired result.

$\square$

We see that Theorem 5.9 defines an endogenous path within (HSDP):

$$\mathcal{C} = \left\{ (y, x, \tau, \theta, s, \kappa) \in \mathcal{F}_h^0 : \begin{pmatrix} Xs \\ \tau\kappa \end{pmatrix} = \frac{x^T s + \tau\kappa}{n+1} e, \right\}$$

which we may call the (self-)central path for (HSDP). Obviously, if $X^0 s^0 = e$, then the initial interior feasible point proposed in Theorem 5.6 is on the path with $\mu = 1$. Our choice (5.11) for x^0 and s^0 satisfies this requirement. We can define a neighborhood of the path as

$$\mathcal{N}(\beta) = \left\{ (y, x, \tau, \theta, s, \kappa) \in \mathcal{F}_h^0 : \| \begin{pmatrix} Xs \\ \tau\kappa \end{pmatrix} - \mu e \| \le \beta\mu, \right.$$
$$\left. \text{where} \quad \mu = \frac{x^T s + \tau\kappa}{n+1} \right\}$$

for some $\beta \in (0, 1)$. Note that from statement (iv) of Theorem 5.6 we have $\theta = \mu$ for any feasible point in $\mathcal{F}_h$.

Since the (HSDP) model constructed and analyzed does not rely on any particular algorithm for solving it, we may use any interior-point algorithm, as long as it generates a strictly complementary solution. Given an interior feasible point $(y^k, x^k, \tau^k, \theta^k, s^k, \kappa^k) \in \mathcal{F}_h^0$, consider solving the following system of linear equations for $(d_y, d_x, d_\tau, d_\theta, d_s, d_\kappa)$:

$$\begin{array}{rcl} (d_y, d_x, d_\tau, d_\theta, d_s, d_\kappa) & & \text{satisfies (5.14)}, \\ \begin{pmatrix} X^k d_s + S^k d_x \\ \tau^k d_\kappa + \kappa^k d_\tau \end{pmatrix} & = & \gamma\mu^k e - \begin{pmatrix} X^k s^k \\ \tau^k \kappa^k \end{pmatrix}. \end{array} \tag{5.15}$$

In what follows we apply the predictor-corrector algorithm in Chapter 4 to solving (HSDP):

Predictor Step. For any even number k, we have $(y^k, x^k, \tau^k, \theta^k, s^k, \kappa^k) \in \mathcal{N}(\beta)$ with $\beta = 1/4$. We solve the linear system (5.14,5.15) with $\gamma = 0$. Then let

$$\begin{array}{rr} y(\alpha) := y^k + \alpha d_y, & x(\alpha) := x^k + \alpha d_x, \\ \tau(\alpha) := \tau^k + \alpha d_\tau, & \theta(\alpha) := \theta^k + \alpha d_\theta, \\ s(\alpha) := s^k + \alpha d_s, & \kappa(\alpha) := \kappa^k + \alpha d_\kappa. \end{array}$$

We determine the step size using

$$\bar{\alpha} := \max\left\{\alpha : (y(\alpha), x(\alpha), \tau(\alpha), \theta(\alpha), s(\alpha), \kappa(\alpha)) \in \mathcal{N}(2\beta)\right\}. \tag{5.16}$$

Then compute the next points by $y^{k+1} = y(\bar{\alpha})$, $x^{k+1} = x(\bar{\alpha})$, $\tau^{k+1} = \tau(\bar{\alpha})$, $\theta^{k+1} = \theta(\bar{\alpha})$, $s^{k+1} = s(\bar{\alpha})$, and $\kappa^{k+1} = \kappa(\bar{\alpha})$.

Corrector Step. For any odd number k, we solve the linear system (5.14,5.15) with $\gamma = 1$. Then let $y^{k+1} = y^k + d_y$, $x^{k+1} = x^k + d_x$, $\tau^{k+1} = \tau^k + d_\tau$, $\theta^{k+1} = \theta^k + d_\theta$, $s^{k+1} = s^k + d_s$, and $\kappa^{k+1} = \kappa^k + d_\kappa$. We have

$$(y^{k+1}, x^{k+1}, \tau^{k+1}, \theta^{k+1}, s^{k+1}, \kappa^{k+1}) \in \mathcal{N}(\beta).$$

Termination. We use the termination technique described earlier to terminate the algorithm. Define σ^k be the index set $\{j : x_j^k \geq s_j^k, \quad j = 1, 2, ..., n\}$, and let $P = \sigma^k$ and the rest be Z. Then, we again use a least-squares projection to create an optimal solution that is strictly self-complementary for (HSDP).

Case 1: If $\tau^k \geq \kappa^k$, we solve for y, x_P, and τ from

$$\begin{array}{llll} \min & \|(S_Z^k)^{-1} A_Z^T (y^k - y)\|^2 & +\|(X_P^k)^{-1}(x_P^k - x_P)\|^2 & \\ \text{s.t.} & & A_P x_P & = b\tau^k, \\ & -A_P^T y & & = -c_P \tau^k, \end{array}$$

otherwise,

Case 2: $\tau^k < \kappa^k$, and we solve for y and x_P from

$$\begin{array}{llll} \min & \|(S_Z^k)^{-1} A_Z^T (y^k - y)\|^2 & +\|(X_P^k)^{-1}(x_P^k - x_P)\|^2 & \\ \text{s.t.} & & A_P x_P & = 0, \\ & -A_P^T y & & = 0, \\ & b^T y & -c_P^T x_P & = \kappa^k. \end{array}$$

This projection guarantees that the resulting x_P^* and s_Z^* ($s_Z^* = c_Z \tau^k - A_Z^T y^*$ in Case 1 or $s_Z^* = -A_Z^T y^*$ in Case 2) are positive, as long as $(x^k)^T s^k + \tau^k \kappa^k$ is reduced to be sufficiently small by the algorithm according to our discussion in the preceding section on termination.

Theorem 5.10 *The $O(\sqrt{n}\log(R/\epsilon))$ interior-point algorithm, coupled with the termination technique described above, generates a strictly self complementary solution for (HSDP) in $O(\sqrt{n}(\log(c(A,b,c))+\log n))$ iterations and $O(n^3(\log(c(A,b,c))+\log n))$ operations, where $c(A,b,c)$ is a positive number depending on the data (A,b,c). If (LP) and (LD) have integer data with bit length L, then by the construction, the data of (HSDP) remains integral and its length is $O(L)$. Moreover, $c(A,b,c)\le 2^L$. Thus, the algorithm terminates in $O(\sqrt{n}L)$ iterations and $O(n^3L)$ operations.*

Now using Theorem 5.7 we obtain

Corollary 5.11 *Within $O(\sqrt{n}(\log(c(A,b,c))+\log n))$ iterations and $O(n^3(\log(c(A,b,c))+\log n))$ operations, where $c(A,b,c)$ is a positive number depending on the data (A,b,c), the $O(\sqrt{n}\log(R/\epsilon))$ interior-point algorithm, coupled with the termination technique described above, generates either optimal solutions to (LP) and (LD) or a certificate that (LP) or (LD) is infeasible. If (LP) and (LD) have integer data with bit length L, then $c(A,b,c)\le 2^L$.*

Again, $c(A,b,c)$ plays the condition number for the data set (A,b,c).

Note that the algorithm may not detect the infeasibility status of both (LP) and (LD). Consider the example where

$$A=\begin{pmatrix} -1 & 0 & 0 \end{pmatrix},\quad b=1,\qquad \text{and}\quad c=\begin{pmatrix} 0 & 1 & -1 \end{pmatrix}.$$

Then,

$$y^*=2,\quad x^*=(0,2,1)^T,\quad \tau^*=0,\quad \theta^*=0,\quad s^*=(2,0,0)^T,\quad \kappa^*=1$$

could be a strictly self-complementary solution generated for (HSDP) with

$$c^Tx^*=1>0,\quad b^Ty^*=2>0.$$

Thus (y^*,s^*) demonstrates the infeasibility of (LP), but x^* doesn't show the infeasibility of (LD). Of course, if the algorithm generates instead $x^*=(0,1,2)^T$, then we get demonstrated infeasibility of both.

5.3.3 Further analysis

In practice, we may wish to stop the algorithm at an approximate solution. Thus, we wish to analyze the asymptotic behavior of τ^k vs. θ^k.

Theorem 5.12 . *If (LP) possesses an optimal solution, then*

$$\tau^k\ge\frac{1-2\beta}{(e^T\bar{x}+e^T\bar{s}+1)}\quad \textit{for all}\quad k$$

where $\bar{x}$ *and* $(\bar{y}, \bar{s})$ *are any optimal solution pair for (LP) and (LD); otherwise,*

$$\kappa^k \geq \frac{1-2\beta}{(e^T\bar{x}+e^T\bar{s}+1)} \quad \textit{for all} \quad k$$

where $\bar{x}$ *and* $(\bar{y}, \bar{s})$ *are any certificate for proving the infeasibility of (LP) or (LD), and moreover,*

$$\frac{1-2\beta}{2(n+1)} \leq \frac{1-2\beta}{\kappa^k} \leq \frac{\tau^k}{\theta^k} \leq \frac{1+2\beta}{\kappa^k} \leq \frac{(e^T\bar{x}+e^T\bar{s}+1)(1+2\beta)}{(1-2\beta)} \quad \textit{for all} \quad k$$

where ϵ *is a fixed positive number independent of* k.

Proof. Note that the sequence generated by the predictor-corrector algorithm is in $\mathcal{N}(2\beta)$. Note that

$$y^* = \alpha\bar{y}, \quad x^* = \alpha\bar{x}, \quad \tau^* = \alpha, \quad \theta^* = 0, \quad s^* = \alpha\bar{s}, \quad \kappa^* = 0,$$

where

$$\alpha = \frac{n+1}{e^T\bar{x}+e^T\bar{s}+1} > 0,$$

is a self-complementary solution for (HSDP). Now we use

$$(x^k - x^*)^T(s^k - s^*) + (\tau^k - \tau^*)(\kappa^k - \kappa^*) = 0,$$

which follows by subtracting the constraints of (HSDP) for $(y^*, \ldots, \kappa^*)$ from those for $(y^k, \ldots, \kappa^k)$ and then multiplying by $((y^k - y^*)^T, \ldots, \kappa^k - \kappa^k)$. This can be rewritten as

$$(x^k)^T s^* + (s^k)^T x^* + \kappa^k \tau^* = (n+1)\mu^k = (n+1)\theta^k.$$

Thus,

$$\tau^k \geq \frac{\tau^k\kappa^k}{(n+1)\mu^k}\tau^* \geq \frac{1-2\beta}{n+1}\tau^* = \frac{1-2\beta}{(e^T\bar{x}+e^T\bar{s}+1)}.$$

The second statement follows from a similar argument. We know that there is an optimal solution for (HSDP) with

$$\kappa^* \geq \frac{n+1}{(e^T\bar{x}+e^T\bar{s}+1)} > 0.$$

Thus

$$\kappa^k \geq \frac{1-2\beta}{(e^T\bar{x}+e^T\bar{s}+1)}$$

for all k. In addition, from relation (5.13) we have $\kappa^k \leq (n+1)+(n+1)\theta^k \leq 2(n+1)$ for all k.

$\square$

Theorem 5.12 indicates that either τ^k stays bounded away from zero for all k, which implies that (LP) has an optimal solution, or τ^k and θ^k converge to zero at the same rate, which implies that (LP) does not have an optimal solution. In practice, for example, we can adopt the following two convergence criteria:

- $(x^k/\tau^k)^T(s^k/\tau^k) \le \epsilon_1$, and $(\theta^k/\tau^k)\|(\bar{b}, \bar{c})\| \le \epsilon_2$,
- $\tau^k \le \epsilon_3$.

Here ϵ_1, ϵ_2, and ϵ_3 are small positive constants. Since both $(x^k)^T s^k + \tau^k \kappa^k$ and θ^k decrease by at least $(1 - 1/\sqrt{n+1})$ in every two iterations, one of the above convergence criteria holds in $O(\sqrt{n}t)$ iterations for $t = \max\{\ln((x^0)^T s^0/(\epsilon_1 \epsilon_3^2)), \ln(\|\bar{b}, \bar{c}\|/(\epsilon_2 \epsilon_3))\}$. If the algorithm terminates by the first criterion then we get approximate optimal solutions of (LP) and (LD); otherwise we detect that (LP) and (LD) have no optimal solutions such that $\|(\bar{x}, \bar{s})\|_1 \le (1 - 2\beta)/\epsilon_3 - 1$ from Theorem 5.12.

5.3.4 Implementation

From the implementation point of view, each iteration of our algorithm solves the linear system (5.14,5.15). It can be shown that

$$d_\theta = \gamma - 1.$$

Then eliminating d_s and d_κ, we face the KKT system of linear equations:

$$\begin{pmatrix} X^k S^k & -X^k A^T & X^k c \tau^k \\ A X^k & 0 & -\tau^k b \\ -\tau^k c^T X^k & \tau^k b^T & \tau^k \kappa^k \end{pmatrix} \begin{pmatrix} (X^k)^{-1} d_x \\ d_y \\ (\tau^k)^{-1} d_\tau \end{pmatrix}$$

$$= \begin{pmatrix} \gamma \mu^k e - X^k s^k \\ 0 \\ \gamma \mu^k - \tau^k \kappa^k \\ 0 \end{pmatrix} + (1 - \gamma) \begin{pmatrix} -X^k \bar{c} \\ \bar{b} \\ \tau^k \bar{z} \end{pmatrix}.$$

Thus, the dimension of the system is increased only by 1 over the case when strictly feasible points for both (LP) and (LD) are known and used for starting primal-dual interior-point algorithms. It seems that the benefit of knowing a starting interior point is not great.

5.4 Notes

Using the rank-one updating technique to improve the arithmetic operation complexity of interior-point algorithms by a factor of $\sqrt{n}$ was first due to Karmarkar [173]. Gonzaga [122] and Vaidya [348] used this technique to obtain the current-best LP arithmetic complexity bound $O(n^3 L)$, see also Mizuno [228] for a general treatment. Parallel worst-case complexity results on interior-point algorithms can be seen in Goldberg, Plotkin, Shmoys and Tardos [112, 113] and and Nesterov and Nemirovskii [262].

The convergence behavior of various interior-point trajectories was studied by Adler and Monteiro [5], Güler [136], Lagarias [194], McLinden [212], Megiddo and Shub [219], Monteiro [236] [237], and Monteiro and Tsuchiya [245]. The analysis of identifying the optimal partition of variables at a strictly complementary solution was due to Güler and Ye [140]. Adler and Monteiro [6], Jansen, Roos and Terlaky [158], Greenberg [131], and Monteiro and Mehrotra [243] provided a post-optimality analysis based on the optimal partition of variables.

The termination procedure described here was developed by Mehrotra and Ye [224]. They also reported effective computational results for solving Netlib problems. A more recent termination or cross-over procedure for obtaining a basic optimal solution is developed by Andersen and Ye [13], Bixby and Saltzman [51], Kortanek and Zhu [189], and Megiddo [217]. Andersen and Ye proved a polynomial bound and showed its practical effectiveness. For a comprehensive survey on identifying an optimal basis and the optimal partition, see El–Bakry, Tapia and Zhang [85].

The homogeneous and self-dual algorithm is due to Mizuno, Todd and Ye [386], which is based on the homogeneous model of Goldman and Tucker [118, 344]. The algorithm is simplified and implemented by Xu, Hung and Ye [369] (also see Tutuncu [346]). A combined phase I and phase II algorithm was proposed by Anstreicher [20], also see Freund [97]. Other infeasible-starting algorithms, which are very popular and effective, were developed and analyzed by Lustig [202], Kojima, Megiddo and Mizuno [180], Mizuno [230], Mizuno, Kojima and Todd [232], Potra [280], Tanabe [317], Wright [367], and Zhang [389]. These algorithms start from some $x^0 > 0$, $s^0 > 0$ and y^0, then each iteration generates directions from

$$\begin{array}{rcl} Ad_x \qquad\qquad & = & b - Ax^k, \\ -A^T d_y \qquad\qquad -d_s & = & s^k - c + A^T y^k, \end{array}$$

and

$$X^k d_s + S^k d_x \;=\; \gamma \mu^k e - X^k s^k.$$

Under certain conditions for choosing the initial point, these algorithms have polynomial iteration bounds (e.g. Zhang [389] and Mizuno [230]). A surface theory of all infeasible interior-point algorithms can be seen in Mizuno, Todd and Ye [234].

There have also been efforts to look for *lower* bounds on the number of iterations required; see Anstreicher [22], Bertsimas and Luo [46], Ji and Ye [164], Powell [283], Sonnevend, Stoer and Zhao [307, 308], and Zhao and Stoer [394]. One important recent result is due to Todd [327], who obtains a bound of at least $n^{1/3}$ iterations to achieve a constant factor decrease in the duality gap. The algorithm he studies is the primal-dual affine-scaling algorithm, which is close to methods used in practical implementations. He allows almost any reasonable step size rule, such as going 99.5% of the way to the boundary of the feasible region, again as used in practical codes; such step size rules definitely do not lead to iterates lying close to the central path. The weakness of the primal-dual affine-scaling algorithm is that no polynomiality or even global convergence has been established for it, except for the case of very small step sizes, and practical experiments indicate that the algorithm alone may not perform well.

Todd also showed that his lower bound extends to other polynomial primal-dual interior-point methods that use directions, including some centering component if the iterates are restricted to a certain neighborhood of the central path. Todd and Ye [331] further extended his result to long-step primal-dual variants that restrict the iterates to a wider neighborhood. This neighborhood seems the least restrictive while also guaranteeing polynomiality for primal-dual path-following methods, and the variants are even closer to what is implemented in practice.

Recently, Atkinson and Vaidya [350] used a combined logarithmic and volumetric potential function to derive an algorithm for LP in $O(n^{1/4}m^{1/4}L)$ iterations. Their algorithm is simplified and improved by Anstreicher [21] and Ramaswamy and Mitchell [284].

Condition-based complexity analyses could be found in Renegar [287], who developed a general condition number and ill-posedness theory for the generalized linear programming. See also Filipowski [91], Freund and Vera [98], and Todd and Ye [332] for a related discussion. More recently, Vavasis and Ye [362] proposed a primal-dual "layered-step" interior point (LIP) algorithm for linear programming. This algorithm follows the central path, either with short steps or with a new type of step called a "layered least squares" (LLS) step. The algorithm returns an exact optimum after a finite number of steps; in particular, after $O(n^{3.5}c(A))$ iterations, where $c(A)$ is a function of the coefficient matrix, which is independent of b and c. One consequence of the new method is a new characterization of the central path: we show that it composed of at most n^2 alternating straight and curved segments. If the LIP algorithm is applied to integer data, we get as another corollary a new proof of a well-known theorem of Tardos that linear programming can be solved in strongly polynomial time provided that A contains small-integer entries. Megiddo, Mizuno and Tsuchiya [218] further proposed an enhanced version of the LIP algorithm.

5.5 Exercises

5.1 *Verify inequality*

$$\mathcal{P}_{n+\rho}(x^{k+1}, z^k) - \mathcal{P}_{n+\rho}(x^k, z^k) \leq \frac{\alpha}{1.1}\min(\beta\sqrt{\frac{n}{n+\beta^2}}, 1-\beta) + \frac{(1.1\alpha)^2}{2(1-1.1\alpha)}$$

in Section 5.1

5.2 *In the termination section, prove if both* $\|d_x\|_\infty$ *and* $\|d_s\|_\infty$ *are less than* 1*, then one must have* $x_P > 0$ *and* $s_Z = c_Z - A_Z^T y > 0$*, which imply that* $\sigma^k = P^*$*.*

5.3 *Analyze the complexity bound if* A_P *has no full row rank in the termination procedure.*

5.4 *Prove that if the LP problem has integral data, then*

$$\zeta(A,b,c) \leq 2^L \quad \textit{and} \quad \xi(A,b,c) \geq 2^{-L},$$

where L *is the size of the binary LP data.*

5.5 *Prove that the total number of required pivots in the process described at the end of Section 5.2.1 is at most* $|\sigma(x^*_{P^*})| - m \leq n - m$*.*

5.6 *Prove Proposition 5.5.*

5.7 *Prove Theorem 5.10 for the predictor-corrector algorithm described in Section 5.3.2.*

5.8 *Similar to* $\xi(A,b,c)$ *and* $\zeta(A,b,c)$*, derive an expression for the condition number* $c(A,b,c)$ *in Theorem 5.10. Prove that if the LP problem has integral data, then*

$$c(A,b,c) \leq 2^L,$$

where L *is the size of the binary LP data.*

5.9 *If every feasible solution of an LP problem is large, i.e.,* $\|x\|$ *is large, then the problem is near infeasible. Prove this statement using Theorem 1.11.*

Chapter 6

Average-Case Analysis

The idea of average-case analysis is to obtain rigorous probabilistic bounds on the number of iterations required by an iterative algorithm to reach some termination criterion. Although many interior point algorithms devised in the last several years are polynomial time methods, in practice they generally perform much better than worst-case bounds would indicate. A "gap" between theory and practice exists that average-case analysis might (at least partially) close.

There are two main viewpoints in the probabilistic analysis of algorithms. First one can develop randomized algorithms, and show that, on a worst-case instance of a problem, the average running time of the algorithm has a certain bound, or the running time satisfies a certain bound with high probability, or the running time always satisfies a certain bound and the algorithm gives a correct answer with high probability, meaning converging to 1 as the dimension of the problem goes to ∞.

Second one can consider the expected running time of a deterministic algorithm when applied to problem instances generated according to some probability distribution (or class of such distributions). For linear programming, researchers have provided some theoretical justification for the observed practical efficiency of the simplex method, despite its exponential worst-case bound. Of course, this viewpoint might be less compelling, since one can always argue that the distribution chosen for problem instances is inappropriate.

Another minor viewpoint is the so called "one-step analysis:" at an iteration we make an nonrigorous but plausible assumption concerning the current data generated by the algorithm, and then address the expected behavior or behavior which occurs with high probability at that iteration. The anticipated number of iterations is then defined to be the number of iterations required if this behavior actually occurs at every iteration (or at least once every, say, ten iterations). This analysis is distinct from the two just described. As the reader will see, the assumptions we make at each iteration can be inconsistent with one another. Nevertheless, such an approach might add insight in the case where a more rigorous analysis seems intractable.

In this chapter, we first develop a one-step analysis for several adaptive interior-point algorithms described in Section 4.5, which all have complexities of $O(n^{1/2}\log(1/\epsilon))$ or $O(n\log(1/\epsilon))$ iterations to attain precision ϵ. (Here we assume, without loss of generality, that $(x^0)^T s^0 = R = 1$.) Based on the one-step analysis, we anticipate that these algorithms would only require $O(n^{1/4}\log(1/\epsilon))$ or $O((\log n)\log(1/\epsilon))$ iterations, where n is the number of variables in (LP).

We then develop a rigorous analysis, based on the second main viewpoint of probabilistic analysis, of interior-point algorithms coupled with the termination procedure described in Chapter 5. We will first show that a random linear feasibility problem can be solved in $O(\sqrt{n}\log n)$ iterations with high probability. Using the homogeneous and self-dual algorithm described in Chapter 5, we then show that the expected number of iterations required to solve a random LP problem is bounded above by $O(\sqrt{n}\log n)$.

Let us formally define high probability: an event in n-dimensional space is true with probability approaching one as $n \to \infty$. Such an event is called a high probability event. Note that a result based on high probability may be stronger than the one based on the standard expected or average analysis. We first derive some **Observations**:

1) Let events E_1 and E_2 be true with high probability. Then the event $E_1 \cap E_2$ is also true with high probability.

2) Let the event E_1 be true with high probability, and let E_1 imply E_2. Then the event E_2 is also true with high probability.

Observation (1) can not be induced to n events. However, we have

Lemma 6.1 *Let n events be E_1, E_2,..., E_n and their complements be $\bar{E}_1$, $\bar{E}_2$,..., $\bar{E}_n$, respectively. Then, if the probability*

$$\lim_{n\to\infty}\sum_{j=1}^{n} P(\bar{E}_j) = 0,$$

then $E_1 \cap E_2 \cap \ldots \cap E_n$ is true with high probability.

Proof. The proof simply follows

$$P(E_1 \cap E_2 \cap ... \cap E_n) \geq 1 - \sum_{j=1}^{n} P(\bar{E}_j).$$

□

6.1 One-Step Analysis

Consider two adaptive interior-point algorithms in Section 4.5: the predictor-corrector and wide-neighborhood algorithms, with worst-case complexities $O(n^{1/2}\log(1/\epsilon))$ and $O(n\log(1/\epsilon))$ iterations to attain precision ϵ, respectively. The progress will be far greater in the predictor-corrector algorithm if $\|Pq\|$ are typically much smaller than the bound given by Lemma 4.14. From (4.23) and (4.24), the corrector step will be much better centered than is guaranteed by Lemma 4.16, and the predictor step will be much larger than $O(n^{-1/2})$ by Lemma 4.17.

On the other hand, Lemma 4.14 shows that, for the wide-neighborhood algorithm, $\|Pq\|_\infty$ and $\|Pq\|_\infty^-$ can only be bounded by a multiple of $\|r\|^2$, not $\|r\|_\infty^2$, unless an extra factor of n is introduced. But $\|r\|$ may be large compared to $\|r\|_\infty$, which is related to β with $(x, s) \in \mathcal{N}_\infty(\beta)$. Again, if $\|Pq\|_\infty$ and $\|Pq\|_\infty^-$ are typically much smaller than the bound given by Lemma 4.14, then the duality gap reduction will be far greater.

For now, we note

Corollary 6.2 *Consider the predictor-corrector and wide-neighborhood algorithms in Section 4.5.*

i) *If at a particular iteration we have $\|Pq\| \leq n^{1/2}\mu$ in the predictor step of the predictor-corrector algorithm, then the duality gap at that iteration will decrease at least by a factor of $(1 - \frac{2}{1+\sqrt{1+8\sqrt{n}}})$.*

ii) *Let β and γ be as in Theorem 4.20. If at a particular iteration of the wide-neighborhood algorithm we have $\|Pq\|_\infty \leq \mu^k \log n$ for $\mathcal{N} = \mathcal{N}_\infty(\beta)$ and $\|Pq\|_\infty^- \leq \mu^k \log n$ for $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$, then the duality gap at that iteration will decrease at least by a factor $(1 - \frac{\beta\gamma(1-\gamma)}{\log n})$ with either $\mathcal{N}_\infty(\beta)$ or $\mathcal{N}_\infty^-(\beta)$.*

Proof. These follow immediately from Lemmas 4.17 and 4.19, and inequalities (4.26) and (4.27).

□

6.1.1 High probability behavior

In this section we provide heuristic arguments as to why we might expect $\|Pq\|$, $\|Pq\|_\infty$, and $\|Pq\|_\infty^-$ to be of the sizes stated in the above corollary. Recall that p and q are the projections of $r \in \mathcal{R}^n$ onto the subspaces U and $U^\perp$ respectively. In this section we suppose r is fixed, but assume that

Assumption 6.1 *U is a random subspace of $\mathcal{R}^n$ of dimension $d := n-m$, drawn from the unique distribution on such subspaces that is invariant under orthogonal transformations.*

Given that U is the null space of $AX^{1/2}S^{-1/2} =: \tilde{A}$, this assumption would hold, for example, if each entry of the matrix $\tilde{A}$ were independently drawn from a standard normal distribution. Note that such assumptions, made at different iterations and hence values of X and S, are not consistent with one another. Further, for several interior-point algorithms the asymptotic behavior of (x^k, s^k) is known, and this behavior is also inconsistent with our assumption, see the next chapter. We will comment further on our approach at the end of the section. For now, we examine the consequences on Pq of our assumption. Note that to compensate for the deficiencies of our assumption, the results we obtain hold with probability approaching one as $n \to \infty$.

We now establish the following theorem.

Theorem 6.3 *Let $\rho = \|r\|_\infty / \|r\|$. Then, with the assumption above,*

i)

$$\Pr\left(\|Pq\| \le \frac{\|r\|^2}{4}\left(2\rho^2 + \frac{6.5}{n}\right)^{1/2} \right) \to 1 \quad \text{as} \quad n \to \infty;$$

ii)

$$\Pr\left(\|Pq\|_\infty^- \le (\log(n)/n)\|r\|^2 \right) \to 1 \quad \text{as} \quad n \to \infty.$$

Before we show how these results are proved, we indicate how they relate to the bounds on $\|Pq\|$ that form the hypotheses of Corollary 6.2. In Corollary 6.2(i), we are analyzing the predictor step, so $r = -(XS)^{1/2}e$ and $(x,s) \in \mathcal{N}_2(1/4)$. Hence $\|r\|^2 = x^T s = n\mu$ and $\|r\|_\infty^2 = \|Xs\|_\infty \le \mu + \|Xs - \mu e\| \le 5\mu/4$. Thus $\rho^2 \le 5/(4n)$ and by Theorem 6.3(i), with probability approaching 1

$$\|Pq\| \le \frac{\|r\|^2}{4}\left(2\frac{5}{4n} + \frac{6.5}{n}\right)^{1/2} = \frac{3\|r\|^2}{4n^{1/2}} < n^{1/2}\mu,$$

which is the hypothesis of Corollary 6.2(i).

For Corollary 6.2(ii), we consider first the case where $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$. Then by Theorem 6.3(ii) and Lemma 4.15(iii), with probability approaching 1

$$\|Pq\|_\infty^- \le (\log(n)/n)\|r\|^2 \le \log(n)\mu^k,$$

which gives the hypothesis of Corollary 6.2(ii) in this case. Now suppose $\mathcal{N} = \mathcal{N}_\infty(\beta)$. Then with high probability

$$\|Pq\|_\infty^- \le \log(n)\mu^k$$

as above. Also, by Lemma 4.14(iii) and Lemma 4.15(iii),

$$\|Pq\|_\infty^+ \le \frac{\|r\|_\infty^2}{4} \le \frac{(1+\beta)\mu^k}{4} \le \frac{\mu^k}{2}.$$

Hence $\|Pq\|_\infty \le \log(n)\mu^k$ with probability approaching 1, which gives the hypothesis for Corollary 6.2(ii) with $\mathcal{N} = \mathcal{N}_\infty(\beta)$.

6.1.2 Proof of the theorem

Now we indicate the proof of Theorem 6.3. The proof of (i) is long and technical and hence we omit it here. However, we will prove a slightly weaker version of (i) at the end of this section.

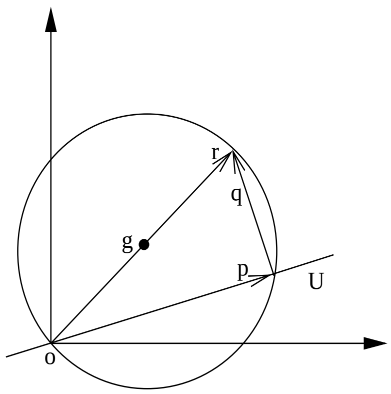

Figure 6.1: Illustration of the projection of r onto a random subspace U

Because p and q are homogeneous of degree 1 in $\|r\|$, we assume henceforth without loss of generality that r is scaled so that

$$g = r/2 \ \text{ satisfies } \ \|g\| = 1.$$

Let $F = (g, H)$ be an orthogonal $n \times n$ matrix. If we express the vector p in terms of the basis consisting of the columns of F, we get

Lemma 6.4 *We can write*

$$p = (1 + \zeta)g + \eta H v, \tag{6.1}$$

where $\frac{1+\zeta}{2}$ *has a beta distribution with parameters* $\frac{d}{2}$ *and* $\frac{m}{2}$*;* $\eta = \sqrt{1-\zeta^2}$*; and* v *is uniformly distributed on the unit sphere in* $\mathcal{R}^{n-1}$*.*

Proof. Since p and q are orthogonal with $p + q = r$, p lies on the sphere of center $r/2 = g$ and radius $\|g\| = 1$. Thus p can be written in the form (6.1), with $\eta = \sqrt{1-\zeta^2}$ and $\|v\| = 1$. We need to establish that ζ and v have the given distributions.

Note that $\|p\|^2 = (1+\zeta)^2 + \eta^2 = 2(1+\zeta)$. However, we can obtain the distribution of $\|p\|^2$ directly. The invariance under orthogonal transformations implies that we can alternatively take U as a fixed d-subspace, say $\{x \in \mathcal{R}^n : x_{d+1} = \cdots = x_n = 0\}$, and r uniformly distributed on a sphere of radius 2. Then r can be generated as

$$\left(\frac{2\lambda_1}{\|\lambda\|}, \frac{2\lambda_2}{\|\lambda\|}, \cdots, \frac{2\lambda_n}{\|\lambda\|}\right)^T,$$

where $\lambda \sim N(0, I)$ in $\mathcal{R}^n$ (i.e., the components of λ are independent normal random variables with mean 0 and variance 1, denoted by $N(0, 1)$). But then

$$p = \left(\frac{2\lambda_1}{\|\lambda\|}, \frac{2\lambda_2}{\|\lambda\|}, \cdots, \frac{2\lambda_d}{\|\lambda\|}, 0, \cdots, 0\right)^T,$$

and $\|p\|^2 = 4(\lambda_1^2 + \cdots + \lambda_d^2)/(\lambda_1^2 + \cdots + \lambda_n^2)$. This has the distribution of four times a beta random variable with parameter $\frac{d}{2}$ and $\frac{m}{2}$, which confirms the distribution of ζ.

Now let W be an orthogonal matrix with $Wg = g$. W can be thought of as rotating the sphere with center g around its diameter from 0 to $2g = r$. We can view the random d-subspace U as the null space of an $m \times n$ random matrix $\bar{A}$ with independent standard normal entries. The fact that p is the projection of r onto U is then equivalent to $\bar{A}p = 0$, $r - p = \bar{A}^T v$ for some v. But then $(\bar{A}W^T)Wp = 0$ and $r - Wp = Wr - Wp = (\bar{A}W^T)^T v$, so that Wp is the projection of r onto $U' = \{x : (\bar{A}W^T)x = 0\}$. If $\bar{A}$ has independent standard normal entries, so does $\bar{A}W^T$, so U' is also a random d-subspace. Thus Wp has the same distribution as p. But writing W as $HW'H^T + gg^T$, where W' is an arbitrary orthogonal matrix of order $n-1$, we see that v has the same distribution as $W'v$. Since $\|v\| = 1$, v is uniformly distributed on the unit sphere $\mathcal{R}^{n-1}$.

□

Since $p + q = r = 2g$, relation (6.1) implies

$$
\begin{aligned}
q &= (1-\zeta)g - \eta Hv, \quad \text{so that} \\
Pq &= \eta^2 g^2 - 2\zeta\eta GHv - \eta^2 (Hv)^2 \qquad (6.2)\\
&= -(Hv)^2 + (\eta g - \zeta Hv)^2 \\
&\geq -\|Hv\|_\infty^2 e, \qquad (6.3)
\end{aligned}
$$

where $G := \mathrm{diag}(g)$, and g^2, $(Hv)^2$, and $(\eta g - \zeta Hv)^2$ denote the vectors whose components are the squares of those of g, Hv, and $\eta g - \zeta Hv$ respectively.

The proof of Theorem 6.3(i) proceeds by using (6.2) to evaluate $\|Pq\|^2$, and then analyzing all the terms in the resulting expression. The proof of Theorem 6.3(ii) follows from (6.3) (which gives $\|Pq\|_\infty^- \leq \|Hv\|_\infty^2$) and the following result:

Lemma 6.5 *Let $F = [g, H]$ be an orthogonal matrix. If v is uniformly distributed on the unit sphere in $\mathcal{R}^{n-1}$,*

$$\Pr\left(\|Hv\|_\infty \leq \sqrt{3\frac{\log n}{n}}\right) \to 1 \quad as \quad n \to \infty.$$

Proof. Since v is uniformly distributed on the unit sphere in $\mathcal{R}^{n-1}$, it can be generated as follows: $v = \lambda/\|\lambda\|$, where $\lambda \sim N(0, I)$ (the standard normal distribution in $\mathcal{R}^{n-1}$). Hence we wish to obtain an upper bound on $\|H\lambda\|_\infty$ and a lower bound on $\|\lambda\|$, both of which hold with high probability. Now $\|\lambda\|^2$ is a χ^2 random variable with $n-1$ degrees freedom, so

$$
\begin{aligned}
\mathrm{E}(\|\lambda\|^2) &= n - 1, \\
\mathrm{Var}(\|\lambda\|^2) &= 2(n-1).
\end{aligned}
$$

From Chebychev's inequality, we have

$$\Pr(\|\lambda\| \geq (1-\epsilon)\sqrt{n-1}) \to 1 \quad \text{as} \quad n \to \infty \qquad (6.4)$$

for any $\epsilon > 0$.

Let λ_0 be a standard normal variable, and let $\lambda' = (\lambda_0, \lambda)$, also $N(0, I)$ but in $\mathcal{R}^n$. Then $\|\lambda'\|_\infty = \max\{\nu_j : j = 0, 1, 2, \cdots, n-1\}$ where $\nu_j = |\lambda_j|$ has the positive normal distribution. Then $1 - N_+(x) = 2(1 - N(x))$ where N_+ is the distribution function of ν, and N is the normal distribution function. It now follows from results in extreme value theory [1] that

$$\Pr\left(\|\lambda'\|_\infty \leq \sqrt{2\log(2n)}\right) \to 1 \text{ as } n \to \infty.$$

[1] S. I. Resnick, *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag (1987), pp. 42 and 71.

Since $F\lambda'$ is also $N(0, I)$,

$$\Pr\left(\|F\lambda'\|_\infty \leq \sqrt{2\log(2n)}\right) \to 1 \text{ as } n \to \infty.$$

Now we have

$$\|H\lambda\|_\infty \leq \|F\lambda'\|_\infty + \|\lambda_0 g\|_\infty.$$

Since $\|g\| = 1$,

$$\Pr\left(\|\lambda_0 g\|_\infty \leq \epsilon\sqrt{\log n}\right) \to 1 \text{ as } n \to \infty$$

for any $\epsilon > 0$. From the above relations and (6.4), we get the result of the lemma.

□

We conclude this section by showing how (6.2) and Lemma 6.5 imply a slightly weaker form of Theorem 6.3(i). Indeed, (6.2) yields

$$\begin{aligned}\|Pq\| &\leq \eta^2\|g^2\| + 2|\zeta\eta|\|g\|_\infty\|Hv\| + \eta^2\|(Hv)^2\| \\ &\leq \|g\|_\infty\|g\| + 2\|g\|_\infty + \|Hv\|_\infty\|Hv\| \\ &= 3\rho + \|Hv\|_\infty.\end{aligned}$$

By lemma 6.5, this is at most $3\rho + \sqrt{3\log(n)/n}$ with probability approaching 1 as $n \to \infty$. This bound would lead one to hope that $\|Pq\|$ would be at most $(n\log(n))^{1/2}\mu$ at a typical predictor step. The predictor-corrector algorithm, with the worst-case bound $O(\sqrt{n}\log(1/\epsilon))$, would require at most $O((n\log(n))^{1/4}\log(1/\epsilon))$ iterations of this type, while the wide-neighborhood algorithm, with the worst-case bound $O(n\log(1/\epsilon))$, would require at most $O((\log n)\log(1/\epsilon))$ iterations of this type.

6.2 Random-Problem Analysis I

We now develop a rigorous analysis, based on the second main viewpoint introduced at the beginning of the chapter, of interior-point algorithms coupled with the termination procedure described in Chapter 5. We use a simple problem, the homogeneous linear feasibility problem, to illustrate the analysis and show that a random such problem can be solved in $O(\sqrt{n}\log n)$ iterations with high probability.

Consider finding a feasible point for the homogeneous linear system

$$\mathcal{F} = \{x : Ax = 0, \quad x \geq 0, \quad x \neq 0\}. \tag{6.5}$$

We assume that $A \in \mathcal{R}^{m\times n}$ has full row-rank. Let us reformulate the problem as a Phase I LP problem

$$\begin{array}{ll} \text{minimize} & z \\ \text{subject to} & Ax + (-Ae)z = 0, \quad e^T x = 1, \ (x,z) \ge 0, \end{array} \tag{6.6}$$

and its dual

$$\begin{array}{ll} \text{maximize} & \lambda \\ \text{subject to} & s = -A^T y - e\lambda \ge 0, \quad s_z = 1 + e^T A^T y \ge 0. \end{array} \tag{6.7}$$

Obviously, LP problem (6.6) has nonempty interior, its optimal solution set is bounded, and its optimal objective value is 0 if and only if $\mathcal{F}$ is nonempty. In fact, we can select an initial feasible point as $x^0 = e/n$, $z^0 = 1/n$ for the primal, and $y^0 = 0$, $\lambda^0 = -1$, $s^0 = e$ and $s_z^0 = 1$ for the dual. Thus, (x^0, z^0) and (y^0, λ^0) are "centered," and the initial primal-dual gap is $(1 + 1/n)$.

We now specialize the termination procedure proposed in Chapter 5 for solving problem (6.6). Suppose $A = (A_P, A_Z)$, where

$$P = \{j : \ x_j^k \ge s_j^k\} \quad \text{and} \quad Z = \{j : \ x_j^k < s_j^k\}.$$

We solve the least-squares problem

$$\begin{array}{lll} (PP) & \text{minimize} & \|(X_P^k)^{-1}(x_P - x_P^k)\| \\ & \text{subject to} & A_P(x_P - x_P^k) = A_Z x_Z^k + (-Ae)z^k. \end{array}$$

and

$$\begin{array}{lll} (DD) & \text{minimize} & \|(S_Z^k)^{-1}A_Z^T(y - y^k)\| \\ & \text{subject to} & A_P^T(y - y^k) = s_P^k. \end{array}$$

Here, we have ignored variable z and the last (normalization) equality constraint $e^T x = 1$ in the problem when we apply the termination projection.

In particular, if $A = A_P$, then the minimizer $x^* = x_P^*$ of (PP) satisfies

$$(X^k)^{-1}(x^* - x^k) = X^k A^T (A(X^k)^2 A^T)(-Ae)z^k.$$

Thus,

$$\begin{array}{ll} \multicolumn{2}{l}{\|(X^k)^{-1}(x^* - x^k)\|} \\ = & \|X^k A^T (A(X^k)^2 A^T)(-Ae)z^k\| \\ = & \|X^k A^T (A(X^k)^2 A^T) A X^k (X^k)^{-1} A^T (AA^T)^{-1}(-Ae)z^k\| \\ \le & \|X^k A^T (A(X^k)^2 A^T) A X^k\| \|(X^k)^{-1} A^T (AA^T)^{-1}(-Ae)z^k\| \\ \le & \|(X^k)^{-1} A^T (AA^T)^{-1} Ae(-z^k)\| \\ \le & \|(X^k)^{-1}\| \|A^T (AA^T)^{-1} Ae\| |z^k| \\ \le & \sqrt{n} z^k \|(X^k)^{-1}\|. \end{array}$$

(Note that $\zeta(A,b,c)$ defined in Section 5.2.1 is less than or equal to $\sqrt{n}$ here.) This implies that if $\min(x^k) > \sqrt{n}z^k$, then the projection x^* satisfies

$$\|(X^k)^{-1}x^* - e\| \le \sqrt{n}z^k\|(X^k)^{-1}\| < 1, \tag{6.8}$$

and x^* must be a point in $\overset{\circ}{\mathcal{F}}$.

Let the optimal partition of problem (6.6) be (P^*, Z^*). If system (6.5) has an interior feasible point, then $A = A_{P^*}$ and $z^* = 0$. Using Theorem 5.4 with $\alpha = 1$ in (5.7), we have, when the duality gap $z^k - \lambda^k \le \xi^2/n^2$,

$$s_j^k < \xi/n^2 < \xi/n < x_j^k,\ j \in P^* \quad \text{and} \quad s_j^k > \xi/n > \xi/n^2 > x_j^k,\ j \in Z^*,$$

or

$$ns_j^k < x_j^k,\ j \in P^* \quad \text{and} \quad s_j^k > nx_j^k,\ j \in Z^*,$$

where recall from (5.3) that for the standard LP problem

$$\xi := \xi(A,b,c) = \min\{\xi_p, \xi_d\}.$$

Thus, in $O(\sqrt{n}(|\log\xi| + \log n))$ iterations we have $A_P = A_P^* = A$ and (6.8), and therefore we generate an interior-point in $\overset{\circ}{\mathcal{F}}$.

Consider the case that system (6.5) is empty, then the optimal value $z^* = \lambda^*$ of problem (6.6) is positive and we can choose $y = y^k$ in (DD) and have

$$s = -A^Ty^k = s^k + e\lambda^k.$$

Thus, if $\lambda^k \ge 0$, then we must have $s > 0$ which proves that $\mathcal{F}$ is empty from the Farkas lemma. Note that in $O(\sqrt{n}(|\log z^*| + \log n))$ iterations we have the duality gap $z^k - \lambda^k \le z^*$ or $\lambda^k \ge z^k - z^* \ge 0$.

Let us estimate ξ for problem (6.6) if system (6.5) has an interior feasible point $\bar{x}$ such that

$$p(1/n) \le \bar{x}_j \le p(n) \quad \text{for} \quad j = 1,2,...,n+1, \tag{6.9}$$

where $p(\alpha)$ is a polynomial α^d for a constant $d \ge 1$. Then, for problem (6.6) we must have

$$\xi_p \ge p(1/n)/(np(n)) \quad \text{and} \quad \xi_d = 1,$$

since $(\bar{x}/e^T\bar{x}, 0)$ is a primal optimal solution and $\bar{y} = 0$ is a dual optimal solution with $\bar{s} = (0,\ 1)^T$. Thus, $\xi \ge p(1/n)/(np(n))$.

On the other hand, if system (6.5) is empty then $\{s :\ s = -A^Ty \ge 0\}$ has an interior feasible point $(\bar{y}, \bar{s})$. Let $(\bar{y}, \bar{s})$ satisfy

$$p(1/n) \le \bar{s}_j \le p(n) \quad \text{for} \quad j = 1,2,...,n+1. \tag{6.10}$$

Then, the dual LP problem (6.7) has a feasible point $y = \bar{y}/e^T\bar{s}$, $s = \bar{s}/e^T\bar{s}$, $\lambda = \min(s)$, and $s_z = 0$. Thus, $z^* \geq \lambda \geq p(1/n)/(np(n))$.

To summarize, we have

Theorem 6.6 *Let $p(\alpha) = \alpha^d$ for a constant $d \geq 1$ and let the homogeneous system (6.5) either have a feasible point $\bar{x}$ satisfying (6.9) or be empty with an $\bar{s} = -A^T\bar{y}$ satisfying (6.10). Then, finding a feasible point for system (6.5) or proving it empty can be completed in $O(\sqrt{n}\log n)$ iterations by an $O(\sqrt{n}\log(1/\epsilon))$ interior-point algorithm, where each iteration solves a system of linear equations.*

We emphasize that ξ or z^* is a non-combinatorial measure of the feasible region $\mathcal{F}$ or its dual. For an LP problem (as opposed to a feasibility problem), ξ or z^* is determined by the geometry of the optimal face.

6.2.1 High probability behavior

From Lemma 6.1 we can derive several propositions.

Proposition 6.7 *Let $\hat{x}_j$, $j = 1, 2, ..., n$, have the identical standard Gauss distribution $N(0, 1)$ and condition on the event that $\hat{x}_j \geq 0$ for $j = 1, ..., n$. Then, with high probability*

$$p(1/n) \leq \hat{x}_j \leq p(n) \quad \textit{for} \quad j = 1, 2, ..., n.$$

Proposition 6.8 *Let $\hat{x}_j$, $j = 1, 2, ..., n$, have the identical Cauchy distribution, i.e., the quotient of two independent $N(0, 1)$ random variables, and condition on the event that $\hat{x}_j \geq 0$ for $j = 1, ..., n$. Then, with high probability*

$$p(1/n) \leq \hat{x}_j \leq p(n) \quad \textit{for} \quad j = 1, 2, ..., n.$$

Proposition 6.9 *Let λ_0, λ_1,..., λ_m be independent $N(0, 1)$ random variables and condition on the event that $\lambda_i \leq |\lambda_0|/\sqrt{d}$ for $i = 1, \ldots, m$ ($d = n - m \geq 1$). Then, the non-negative random variables, $\hat{x}_i := 1 - \sqrt{d}\lambda_i/|\lambda_0|$ for $i = 1, \ldots, m$, satisfy*

$$p(1/n) \leq \hat{x}_i \leq p(n) \quad \textit{for} \quad i = 1, \ldots, m,$$

with high probability, where $p(n) = n^d$ for some constant d.

The first two propositions are relatively easy to prove. To prove the third, we first prove a similar proposition:

Proposition 6.10 *Let λ_0, λ_1,..., λ_m be independent $N(0,1)$ random variables and condition on the event that $\lambda_i \leq |\lambda_0|/\sqrt{d}$ for $i = 1, \ldots, m$ ($d = n - m \geq 1$). Then, the non-negative random variables, $x_i := |\lambda_0|/\sqrt{d} - \lambda_i$ for $i = 1, \ldots, m$, satisfy*

$$p(1/n) \leq x_i \leq p(n) \quad \text{for} \quad i = 1, \ldots, m,$$

with high probability.

Proof. In proving Proposition 6.10, we fix $p(n) = n^4$. Let $f(\lambda)$ be the probability density function of $N(0,1)$,

$$P(m) := P(x_1 \geq 0, x_2 \geq 0, \ldots, x_m \geq 0)$$

and

$$P(m-1) := P(x_2 \geq 0, \ldots, x_m \geq 0).$$

Also note that $|N(0,1)|$ has the probability density function $2f(\lambda)$ in $[0,\infty)$. Then, we have

$$\begin{aligned}
&P(m)\\
&= P(x_1 \geq 0, x_2 \geq 0, \ldots, x_m \geq 0)\\
&= \int_0^\infty 2f(\lambda_0) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i\\
&\geq \int_0^\infty 2f(\lambda_0) \int_{-\infty}^{0} f(\lambda_1) \int_{-\infty}^{0} f(\lambda_2) \cdots \int_{-\infty}^{0} f(\lambda_m) \prod_{i=0}^m d\lambda_i\\
&= (1/2)^m.
\end{aligned}$$

We also have

$$\begin{aligned}
&P(m)\\
&= P(x_1 \geq 0, x_2 \geq 0, \ldots, x_m \geq 0)\\
&= \int_0^\infty 2f(\lambda_0) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i\\
&\geq \int_0^\infty 2f(\lambda_0) \int_{-\infty}^{0} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i\\
&= (1/2) \int_0^\infty 2f(\lambda_0) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0, i\neq 1}^m d\lambda_i\\
&= P(m-1)/2.
\end{aligned}$$

Consider the probability

$$P_1^- := P(x_1 \le p(1/n) | x_1 \ge 0, ..., x_m \ge 0).$$

We have

$$\begin{aligned}
P_1^- & \\
= \;& \frac{1}{P(m)} \int_0^\infty 2f(\lambda_0) \int_{\frac{\lambda_0}{\sqrt{d}}-n^{-4}}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i \\
\le \;& \frac{1}{P(m)} \int_0^\infty 2f(\lambda_0) \int_{\frac{\lambda_0}{\sqrt{d}}-n^{-4}}^{\frac{\lambda_0}{\sqrt{d}}} (\frac{1}{\sqrt{2\pi}}) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i \\
= \;& \frac{n^{-4}}{\sqrt{2\pi}} \frac{P(m-1)}{P(m)} \\
\le \;& \frac{2n^{-4}}{\sqrt{2\pi}} = O(n^{-2}).
\end{aligned}$$

Now consider the probability

$$P_1^+ := P(x_1 \ge p(n) | x_1 \ge 0, ..., x_m \ge 0).$$

We have

$$\begin{aligned}
P_1^+ & \\
= \;& \frac{1}{P(m)} \int_0^\infty 2f(\lambda_0) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}-n^4} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i \\
= \;& \frac{1}{P(m)} \int_0^{n^2} 2f(\lambda_0) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}-n^4} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i \\
& + \frac{1}{P(m)} \int_{n^2}^\infty 2f(\lambda_0) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}-n^4} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i \\
:= \;& P_1'^+ + P_1''^+.
\end{aligned}$$

For $P_1''^+$, we have

$$\begin{aligned}
P_1''^+ & \\
= \;& \frac{1}{P(m)} \int_{n^2}^\infty 2f(\lambda_0) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}-n^4} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^m d\lambda_i \\
\le \;& \frac{1}{P(m)} \int_{n^2}^\infty 2f(\lambda_0) d\lambda_0
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{P(m)} \int_{n^2}^{\infty} \frac{2}{\sqrt{2\pi}} \exp(-x^2/2) dx \\
&= \frac{1}{P(m)} \int_{n^2}^{\infty} \frac{2n^2}{\sqrt{2\pi}} [\exp(-x^2/2n^4)]^{n^4} d(x/n^2) \\
&\leq \frac{1}{P(m)} \int_{n^2}^{\infty} \frac{2n^2}{\sqrt{2\pi}} \exp(-(n^4-1)/2) \exp(-x^2/2n^4) d(x/n^2) \\
&= \frac{n^2 \exp(-(n^4-1)/2)}{P(m)} \int_{1}^{\infty} \frac{2}{\sqrt{2\pi}} \exp(-x^2/2) d(x) \\
&\leq 2^m n^2 \exp(-(n^4-1)/2) \\
&= O(n^{-2})
\end{aligned}
$$

for n large enough. For $P_1'^+$, we have

$$
\begin{aligned}
&P_1'^+ \\
&= \frac{1}{P(m)} \int_0^{n^2} 2f(\lambda_0) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}} - n^4} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^{m} d\lambda_i \\
&\leq \frac{1}{P(m)} \int_0^{n^2} 2f(\lambda_0) \int_{-\infty}^{n^2/\sqrt{d} - n^4} f(\lambda_1) \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_2) \cdots \int_{-\infty}^{\frac{\lambda_0}{\sqrt{d}}} f(\lambda_m) \prod_{i=0}^{m} d\lambda_i \\
&\leq \frac{P(m-1)}{P(m)} \int_{-\infty}^{n^2/\sqrt{d} - n^4} f(\lambda_1) d\lambda_1 \\
&\leq 2 \int_{-\infty}^{n^2/\sqrt{d} - n^4} f(\lambda_1) d\lambda_1 \\
&= 2 \int_{n^4 - n^2/\sqrt{d}}^{\infty} f(\lambda_1) d\lambda_1 \\
&= O(n^{-2}).
\end{aligned}
$$

Thus, the probability, P_1, that either $0 < x_1 \leq p(1/n)$ or $x_1 \geq p(n)$ satisfies

$$P_1 := P_1^- + P_1^+ = P_1^- + P_1'^+ + P_1''^+ \leq O(n^{-2}).$$

The same result holds for P_i, $i = 2, ..., m$, the probability that either $0 < x_i \leq p(1/n)$ or $x_i \geq p(n)$. Thus, we shall have

$$\sum_{i=1}^{m} P_i \leq O(n^{-1}),$$

which approaches zero as $n \to \infty$. Using Lemma 6.1 we prove Proposition 6.10.

□

In a similar manner, we can prove that $P(n^{-2} \leq |\lambda_0| \leq n^2 | x_1 \geq 0, ..., x_m \geq 0)$ approaches 1 as $n \to \infty$. This leads to the final proof of Proposition 6.9 for $d = 7$.

In the next section, we prove that the conditions in Theorem 6.6 are satisfied with high probability for some random LP problems.

6.2.2 Random linear problems

Let $A \in \mathcal{R}^{m \times n}$ of the homogeneous linear system (6.5) be standard Gaussian: Each entry of A is independently distributed as a standard Gauss random variable.

Since each column, a_j, of A is also standard Gaussian, $a_j/\|a_j\|$ is uniformly distributed on the unit sphere in $\mathcal{R}^m$. This is a special case of a rotationally-symmetric distribution. Denote by $d = n - m$. Todd has shown that the null space of A is an $d \times n$ standard Gaussian matrix.

Corollary 6.11 *The probability that system (6.5) has a feasible point is*

$$p_{nd} = \frac{\binom{n-1}{0} + ... + \binom{n-1}{d-1}}{2^{n-1}}.$$

Corollary 6.12 *The probability that system (6.5) has an interior feasible point is*

$$p_{nd} = \frac{\binom{n-1}{0} + ... + \binom{n-1}{d-1}}{2^{n-1}}.$$

Proof. From the strictly complementary property of linear systems, this probability is the probability that the dual

$$s = A^T y \geq 0, \qquad s \neq 0 \tag{6.11}$$

is infeasible. However, the latter probability is exact

$$1 - p_{nm} = p_{nd}.$$

□

Theorem 6.13 *With probability one, exactly one of systems (6.5) and (6.11) is empty and the other has an interior feasible point.*

Proof. The probability that (6.11) has an interior feasible solution is p_{nm}. Note that these two events are exclusive, and $p_{nd} + p_{nm} = 1$.

□

We now prove another lemma.

Lemma 6.14 *System (6.5) is feasible if and only if there is a partition* $A = (A_B,\ A_N)$*, where* A_B *is* m *by* m*, such that*

$$A_B x_B + (A_N e) x_{m+1} = 0, \quad x \neq 0 \quad \textit{and} \quad x \geq 0 \tag{6.12}$$

is feasible.

Proof. It is obvious that system (6.12) being feasible implies that (6.5) is feasible. Conversely, if (6.5) is feasible, then it implies that 0 belongs to the polytope P defined as the convex hull of the columns of A. Let $(a_1, a_2, ..., a_d)$ be a minimal affinely independent set of columns of A such that 0 belongs to the convex hull of $(a_1, a_2, ..., a_d)$. By Carathéodory's theorem $d \leq m+1$, if $d < m+1$, take as columns of A_B as $(a_1, a_2, ..., a_d)$ plus $m-d$ any other columns of A. If $d = m+1$, then there is an $(m+1)$-vector $\bar{u}$ such that

$$(a_1, a_2, ..., a_{m+1})\bar{u} = 0 \quad \text{and} \quad \bar{u} > 0.$$

Let b be the sum of the rest of the columns in A, and let $(\bar{v},\ 1)$ be a vector in the null space of the matrix $(a_1, a_2, ..., a_{m+1}, b)$ (since b can be expressed as a linear combination of $(a_1, a_2, ..., a_{m+1})$). Then, for scalars α and β,

$$x(\alpha, \beta) = \alpha \begin{pmatrix} \bar{u} \\ 0 \end{pmatrix} + \beta \begin{pmatrix} \bar{v} \\ 1 \end{pmatrix}$$

is also in the null space of $(a_1, a_2, ..., a_{m+1}, b)$. Let

$$k = \arg\min\{\frac{\bar{v}_j}{\bar{u}_j} : \ j = 1, 2, ..., (m+1)\}.$$

If $\bar{v}_k = 1$, then select $\alpha^* = 0$ and $\beta^* = 1$; if $\bar{v}_k < 1$, then select $\alpha^* = 1$ and $\beta^* = \bar{u}_k/(1-\bar{v}_k)$; else select $\alpha^* = -1$ and $\beta^* = \bar{u}_k/(\bar{v}_k - 1)$. In all of the three cases, $x^* = x(\alpha^*, \beta^*) > 0$ and $x^*_k = x^*_{m+2}$. In other words, the lemma is true by selecting $A_B = (a_1, ..., a_{k-1}, a_{k+1}, ..., a_{m+1})$.

□

Let us call the partition satisfying the condition in Lemma 6.14 a *basic feasible* partition. We now analyze a feasible solution of (6.5) or (6.11). We develop the following result.

Theorem 6.15 *With high probability, a feasible point of system (6.5) or (6.11) satisfies condition (6.9) or (6.10), respectively.*

Proof. Let $(A_B,\ A_N)$ be any partition of A and $b = A_N e/\sqrt{n-m}$. Consider the system

$$(A_B,\ b)x = 0, \quad x \neq 0 \quad \text{and} \quad x \geq 0. \tag{6.13}$$

Since $(A_B,\ b)$ is standard Gaussian, the vector $\hat{x}$ in the null space of $(A_B,\ b)$ is the line generated by a standard Gauss random vector $(\lambda_1, \lambda_2, ..., \lambda_{m+1})$, that is,

$$\hat{x}_i = \delta\lambda_i \quad \text{for} \quad i = 1, 2, ..., m+1,$$

where δ is a scalar. Without loss of generality, we can let $\delta > 0$. Hence, $(A_B,\ A_N)$ is a *basic feasible* partition or $\hat{x}$ is feasible for system (6.13) if and only if $\hat{x}_i = \lambda_i \geq 0$ for $i = 1, 2, ..., m+1$. Thus, each component of a feasible solution of (6.13) has the identical distribution $|N(0,1)|$. Thus, due to Proposition 6.7, with high probability

$$p(1/n) \leq \hat{x}_i \leq p(n) \quad \text{for} \quad i = 1, 2, ..., m+1.$$

Note that $\hat{x}$ induces a feasible solution for system (6.5) by assigning

$$x_B = (\hat{x}_1, ..., \hat{x}_m)^T \quad \text{and} \quad x_N = (\hat{x}_{m+1}/\sqrt{n-m})e.$$

This completes the proof for (6.5).

The same result applies to the dual slack vector s of system (6.11) when it is feasible, where m is replaced by $n - m$.

□

Based on Theorems 6.13 and 6.15, we have the final result

Theorem 6.16 *With high probability the homogeneous random linear feasibility problem (6.5) can be solved in $O(\sqrt{n}\log n)$ iterations.*

Proof. From Theorem 6.13, with probability one either $A = A_P^*$ or $z^* > 0$ for problem (6.6) associated with (6.5). Then from Theorem 6.15, with high probability, there exists positive primal variables or positive dual slacks (not both) satisfying condition (6.9) or (6.10). Thus, the theorem follows from Theorem 6.6.

□

Note that the non-homogeneous linear system

$$Ax = b, \quad x \geq 0, \tag{6.14}$$

where $(A,\ b)$ is standard Gaussian, can be solved by solving the system (6.5) with $A := (A,\ b)$, which remains standard Gaussian. Note that system (6.14) is feasible if and only if $b \in P^*$ where (P^*, Z^*) is the optimal partition of problem (6.6). Thus,

Corollary 6.17 *With high probability the random linear feasibility problem (6.14) can be solved in $O(\sqrt{n}\log n)$ iterations.*

6.3 Random-Problem Analysis II

This section analyzes the average complexity of interior-point algorithms for solving the probabilistic LP model of Todd. This model allows for degenerate optimal solutions, and does not provide a feasible starting point. We refer to this model as "Todd's degenerate model." The lack of an initial solution in the degenerate model is problematic for many interior point algorithms, which require an interior solution to start. We obtain a bound of $O(\sqrt{n}\ln n)$ iterations for the expected number of iterations before termination with an exact optimal solution, using the homogeneous and self–dual algorithm of Chapter 5 as applied to this model.

Denote by $\mathcal{F}_h$ the set of all points that are feasible for (HSDP). Denote by $\mathcal{F}_h^0$ the set of strictly feasible (interior) points in $\mathcal{F}_h$, with $(x, \tau, s, \kappa) > 0$. It is easily seen that (HSDP) has a strictly feasible point: $y = 0,\ x = e > 0,\ \tau = 1,\ \theta = 1,\ s = e > 0,\ \kappa = 1$.

From Theorem 5.7, it is clear that the key to solving a LP problem, or alternatively detecting its infeasibility or unboundedness, is to find a strictly self-complementary solution to (HSDP). Many interior point algorithms might be used to solve (HSDP), as long as they generate a sequence or subsequence of feasible pairs which converge to a strictly complementary solution of the problem being solved, such as the predictor-corrector or wide-neighborhood algorithm described in Chapter 4. By using the analysis employed in Section 5.3, with $\beta = 1/4$, we generate a sequence $(y^k, x^k, \tau^k, \theta^k, s^k, \kappa^k) \in \mathcal{N}(\beta)$, and

$$\frac{\theta^{k+1}}{\theta^k} = \frac{\mu^{k+1}}{\mu^k} \leq \frac{2}{1 + \sqrt{1 + 4\sqrt{2}(n+1)}}. \tag{6.15}$$

6.3.1 Termination scheme

In this section we consider the problem of generating an exact optimal solution to (HSDP). For simplicity, we denote

$$u = \begin{pmatrix} x \\ \tau \end{pmatrix} \in \mathcal{R}^{n+1}, \quad v = \begin{pmatrix} s \\ \kappa \end{pmatrix} \in \mathcal{R}^{n+1}.$$

To begin, let $(u^*, y^*, v^*, \theta^* = 0)$ be any strictly self-complementary solution, i.e., $u^* + v^* > 0$. Note that

$$e^T u^* + e^T v^* = e^T x^* + \tau^* + e^T s^* + \kappa^* = n + 1.$$

Define

$$\sigma_h^* = \{i : 0 \le i \le n+1, \ \ u_i^* > 0\}, \ \text{ and } \ \xi_h^* = \min_i (u_i^* + v_i^*).$$

We refer to σ_h^* as the self-complementary partition of (HSDP), and clearly $0 < \xi_h^* \le 1$. Our goal here is to use the iterates (u^k, v^k) of the algorithm to eventually identify the complementary partition, and to generate an exact optimal solution of (HSDP). Using the techniques developed in Chapter 5, we can prove the following result.

Lemma 6.18 *Let $\zeta = (1-\beta)\xi_h^*/(n+1)$. Then in order to obtain $v_j^k < \zeta \le u_j^k$, $j \in \sigma_h^*$, and $u_j^k < \zeta \le v_j^k$, $j \notin \sigma_h^*$, it suffices to have*

$$\theta^k < \frac{1-\beta}{(n+1)^2}(\xi_h^*)^2. \tag{6.16}$$

Given an iterate $(x^k, \tau^k, y^k, s^k, \kappa^k)$, let A_P denote the columns of A having $x_j^k > s_j^k$, and let x_P denote the corresponding components of x. Similarly let A_Z and s_Z denote the remaining columns of A, and the corresponding components of s. Note that (6.15) implies that $\theta^k \to 0$, so Lemma 6.18 implies that eventually we always have $P = \sigma_h^* \setminus \{n+1\}$. In what follows, we assume that k is in fact large enough so that $P = \sigma_h^* \setminus \{n+1\}$. We employ the following projections to generate an exact optimal solution. We distinguish two cases:

Case 1. $(\tau^k > \kappa^k)$. Find the solution $\hat{x}_P^k$ of

$$\begin{array}{lll} (PP1) & \min & \|x_P - x_P^k\| \\ & \text{s.t.} & A_P x_P = b\tau^k. \end{array}$$

If $\hat{x}_P^k \geq 0$, then compute the solution $\hat{y}^k$ of

$$\begin{array}{lll}(DP1) & \min & \left\|s_Z - s_Z^k\right\| \\ & \text{s.t.} & A_P^T y = c_P \tau^k, \; c_Z \tau^k - A_Z^T y = s_Z,\end{array}$$

and set

$$\hat{s}_Z^k = c_Z \tau^k - A_Z^T \hat{y}^k = s_Z^k - A_Z^T(\hat{y}^k - y^k).$$

Case 2. $(\tau^k \leq \kappa^k)$. Find the solution $\hat{x}_P^k$ of

$$\begin{array}{lll}(PP2) & \min & \left\|x_P - x_P^k\right\| \\ & \text{s.t.} & A_P x_P = 0.\end{array}$$

If $\hat{x}_P^k \geq 0$, then compute the solution $\hat{y}^k$ of

$$\begin{array}{lll}(DP2) & \min & \left\|s_Z - S_Z^k\right\| \\ & \text{s.t.} & A_P^T y = 0, \; -A_Z^T y = s_Z,\end{array}$$

and set $\hat{s}_Z^k = -A_Z^T \hat{y}^k$, and $\hat{\kappa}^k = b^T \hat{y}^k - c_P^T \hat{x}_P^k$.

According to Lemma 6.18, exactly one of the above cases occurs for all sufficiently large k. Also, Case 1 eventually occurs exactly when (LP) has an optimal solution, in which case (PP1) and (DP1) are both clearly feasible. In what follows, we consider Case 1 only, since our random model always has a solution.

It is easily seen from the definition of (HSDP) that:
(PP1) is equivalent to

$$\begin{array}{ll}\min & \left\|x_P - x_P^k\right\| \\ \text{s.t.} & A_P(x_P - x_P^k) = A_Z x_Z^k + \bar{b}\theta^k;\end{array} \tag{6.17}$$

(DP1) is equivalent to

$$\begin{array}{ll}\min & \left\|\bar{c}_Z \theta^k - A_Z^T(y - y^k)\right\| \\ \text{s.t.} & A_P^T(y - y^k) = \bar{c}_P \theta^k + s_P^k,\end{array} \tag{6.18}$$

because

$$s_Z^k - s_Z^k = A_Z^T(y^k - y^k) + \bar{c}_Z \theta^k. \tag{6.19}$$

From (5.4) and $\theta^k \to 0$, we conclude that $(x_Z^k, \; s_P^k) \to 0$ as $k \to \infty$, and also $\kappa^k \to 0$ if $\{n+1\} \in \sigma_h^*$. Using these facts and (6.17)–(6.18) we can easily deduce that

$$(\hat{x}_P^k - x_P^k) \to 0 \text{ and } (\hat{s}_Z^k - s_Z^k) \to 0 \quad \text{as} \quad k \to \infty.$$

From this relation and Lemma 6.18 it follows that $(\hat{x}_P^k, \tau^k > 0, \hat{s}_Z^k) > 0$ (if k is large enough) is a strictly complementary solution to (HSDP).

The above discussion shows that our projection scheme works provided k is large enough. Below we give a more precise characterization of this fact. Again, for our probabilistic model, to be described below, only Case 1 occurs provided k is large enough. Therefore in what follows we assume that k is large enough and that we are always in Case 1.

A matrix A_P satisfies the Haar condition if every square submatrix of A_P is invertible. It is well known that the standard Gaussian matrix A_P is a Haar matrix with probability one. Thus, for the purposes of studying probabilistic behavior, we only have to deal with matrices that satisfy the Haar condition. Let $\hat{A}_B$ denote any square submatrix of A_P with its full row or column dimension. Also, if A_P has more rows than columns, let $\hat{A}$, $\hat{A}_Z$, and $\hat{b}$ denote the $\hat{A}_B$-corresponding rows of A, A_Z, and b, respectively; Otherwise, $\hat{A} = A$, $\hat{A}_Z = A_Z$, and $\hat{b} = b$. Then we have

Lemma 6.19 *Let $(x_P^* > 0, \tau^* > 0, y^*, s_Z^* > 0)$ be any strictly (self) complementary solution for (HSDP). Then Case 1 occurs and (PP1) generates $\hat{x}_P^k > 0$ and $\hat{s}_Z^k > 0$ whenever*

$$\theta^k \le \frac{(1-\beta)(\xi_h^*)^2}{(n+1)^2(1+\sqrt{n}\left\|\hat{A}_B^{-1}\hat{A}_Z\right\|)}. \tag{6.20}$$

Proof. Assume that (6.20) holds. Since (6.20) implies (6.16), we have $\tau^k > \kappa^k$ and P must be the self-complementary partition $\sigma_h^* \setminus \{n+1\}$, by Lemma 6.18. From (6.17), the constraint in (PP1) is clearly equivalent to

$$A_P(x_P - x_P^k) = A_Z x_Z^k + (b - Ae)\theta^k,$$

and it is consistent. Note that $b = A_P x_P^*$. We have

$$A_P(x_P - x_P^k + e\theta^k - x_P^*\theta^k) = A_Z(x_Z^k - e\theta^k). \tag{6.21}$$

One solution to (6.21) is

$$x_P - x_P^k + e\theta^k - x_P^*\theta^k = \hat{A}_B^{-1}\hat{A}_Z(x_Z^k - e\theta^k)$$

if A_P has more rows than columns, or

$$x_P - x_P^k + e\theta^k - x_P^*\theta^k = \begin{pmatrix} \hat{A}_B^{-1}A_Z(x_Z^k - e\theta^k) \\ 0 \end{pmatrix}$$

otherwise. Thus, the solution $\hat{x}_P^k$ of (PP1) must satisfy

$$\left\|\hat{x}_P^k - x_P^k\right\| \le \left\|\hat{A}_B^{-1}\hat{A}_Z(x_Z^k - e\theta^k)\right\| + \theta^k\left\|e - x_P^*\right\|. \tag{6.22}$$

For the first term in (6.22), we have

$$\begin{aligned}
\left\|\hat{A}_B^{-1}\hat{A}_Z(x_Z^k - e\theta^k)\right\| &\leq \left\|\hat{A}_B^{-1}\hat{A}_Z\right\| \|x_Z^k - e\theta^k\| \\
&\leq \max\{\max(x_Z^k), \theta^k\} \|e\| \left\|\hat{A}_B^{-1}\hat{A}_Z\right\| \\
&\leq \frac{(n+1)\theta^k}{\xi_h^*}\sqrt{|Z|}\left\|\hat{A}_B^{-1}\hat{A}_Z\right\|,
\end{aligned}$$

where the last inequality is from (5.5). For the second term of (6.22), we have

$$\begin{aligned}
\theta^k \|e - x_P^*\| &\leq \theta^k\sqrt{|P| - 2e^T x_P^* + \|x_P^*\|^2} \\
&\leq \theta^k\sqrt{|P| - 2e^T x_P^* + (e^T x_P^*)^2} \\
&\leq \theta^k(n+1),
\end{aligned}$$

since $e^T x_P^* \leq n+1$. Substituting the above two inequalities into (6.22) results in

$$\begin{aligned}
\|\hat{x}_P^k - x_P^k\| &\leq \frac{(n+1)\theta^k}{\xi_h^*}\sqrt{|Z|}\left\|\hat{A}_B^{-1}\hat{A}_Z\right\| + (n+1)\theta^k \\
&\leq (1+\sqrt{n}\left\|\hat{A}_B^{-1}\hat{A}_Z\right\|)(n+1)\frac{\theta^k}{\xi_h^*}, \qquad (6.23)
\end{aligned}$$

so (6.20), (6.23) and (5.7) imply that $\hat{x}_P^k > 0$.

Now consider (DP1). From (6.18), the constraint in (DP1) is clearly equivalent to

$$A_P^T(y - y^k) = (c_P - e)\theta^k + s_P^k;$$

and it is consistent. Note that $c_P = P^T y^*$. We have

$$A_P^T(y - y^k - y^*\theta^k) = s_P^k - e\theta^k. \qquad (6.24)$$

Similarly, the solution $\hat{y}_P^k$ of (DP1) must satisfy

$$\|\hat{s}_Z^k - s_Z^k\| \leq \left\|\hat{A}_Z^T(\hat{A}_B^T)^{-1}(s_P^k - e\theta^k)\right\| + \theta^k\|e - s_Z^*\|. \qquad (6.25)$$

For the first term in (6.25), we have

$$\begin{aligned}
\left\|\hat{A}_Z^T(\hat{A}_B^T)^{-1}(s_P^k - e\theta^k)\right\| &\leq \left\|\hat{A}_B^{-1}\hat{A}_Z\right\| \|s_P^k - e\theta^k\| \\
&\leq \max\{\max(s_P^k), \theta^k\} \|e\| \left\|\hat{A}_B^{-1}\hat{A}_Z\right\| \\
&\leq \frac{(n+1)\theta^k}{\xi_h^*}\sqrt{|P|}\left\|\hat{A}_B^{-1}\hat{A}_Z\right\|,
\end{aligned}$$

where the last inequality is from (5.5). For the second term of (6.25), we have

$$\begin{aligned}\theta^k \|e - s_Z^*\| &\leq \theta^k \sqrt{|Z| - 2e^T x_P^* + \|s_Z^*\|^2} \\ &\leq \theta^k \sqrt{|Z| - 2e^T x_P^* + (e^T s_Z^*)^2} \\ &\leq \theta^k (n+1).\end{aligned}$$

Substituting the above two inequalities into (6.25) results in

$$\begin{aligned}\left\|\hat{s}_Z^k - s_Z^k\right\| &\leq \frac{(n+1)\theta^k}{\xi_h^*}\sqrt{|P|}\left\|\hat{A}_B^{-1}\hat{A}_Z\right\| + (n+1)\theta^k \\ &\leq (1+\sqrt{n}\left\|\hat{A}_B^{-1}\hat{A}_Z\right\|)(n+1)\frac{\theta^k}{\xi_h^*}, \qquad (6.26)\end{aligned}$$

so (6.20), (6.26) and (5.7) imply that $\hat{s}_Z^k > 0$.

□

6.3.2 Random model and analysis

In this section we describe a random LP model proposed by Todd, and perform a probabilistic analysis of the behavior of the homogeneous and self-dual algorithm, using the finite termination scheme described above. We will refer to the model under consideration as "Todd's degenerate model."

Todd's Degenerate Model. Let $A = (A_1, A_2)$, where A_i is $m \times n_i$, $n_i \geq 1$, $n_1 + n_2 = n$, and each component of A is i.i.d. from the $N(0,1)$ distribution. Let

$$\hat{x} = \begin{pmatrix} \hat{x}_1 \\ 0 \end{pmatrix}, \quad \hat{s} = \begin{pmatrix} 0 \\ \hat{s}_2 \end{pmatrix},$$

where the components of $\hat{x}_1$ and $\hat{s}_2$ are i.i.d. from the $|N(0,1)|$ distribution. Finally let $b = A\hat{x}, c = \hat{s} + A^T\hat{\pi}$. We assume that either $\hat{\pi} = 0$, or the components of $\hat{\pi}$ are i.i.d. from any distribution with $O(1)$ mean and variance.

Clearly this model allows for degenerate solutions, and produces instances of (LP) having no easy feasible starting point. This presents an obstacle for most interior point methods, which require interior feasible points for initialization. Since an instance of Todd's degenerate model always has an optimal solution, it follows from Theorem 5.7 that $n+1 \in \sigma_h^*$. Therefore, if the homogeneous and self–dual algorithm described in Chapter

5 is applied to an instance of Todd's degenerate model, we are eventually always in Case 1.

Now, we begin a probabilistic analysis of the self–dual algorithm equipped with the termination scheme described in the preceding section. Since our finite termination criterion in Lemma 6.19 depends on ξ_h^*, from a strictly complementary solution (x^*, τ^*, s^*) to (HSDP), we must first infer a valid value of ξ_h^* from the given strictly complementary solution $(\hat{x}, \hat{s})$ for (LP) and (LD).

Let

$$\hat{\xi} = \min(\hat{x} + \hat{s}) = \min \begin{pmatrix} \hat{x}_1 \\ \hat{s}_2 \end{pmatrix}, \quad \hat{\rho} = 1 + e^T\hat{x} + e^T\hat{s}. \tag{6.27}$$

Note that $x^* = (n+1)\hat{x}/\hat{\rho}$, $\tau^* = (n+1)/\hat{\rho}$, $y^* = (n+1)\hat{\pi}/\hat{\rho}$, $\kappa^* = 0$, and $s^* = (n+1)\hat{s}/\hat{\rho}$ is a strictly self-complementary solution to (HSDP). Thus, we have the following proposition.

Proposition 6.20 *Consider Todd's degenerate model with optimal solution $(\hat{x}, \hat{s})$. Then there is a strictly self-complementary solution $(x^*, \tau^*, y^*, s^*, \kappa^*)$ to (HSDP) such that $\xi_h^* \geq \hat{\xi}/\hat{\rho}$.*

This proposition and Lemma 6.19 lead to

Lemma 6.21 *Consider an instance of Todd's degenerate model, and let $\hat{\xi}$ and $\hat{\rho}$ be as in (6.27). Suppose that k is large enough so that the following inequality is satisfied:*

$$\theta^k \leq \frac{(1-\beta)\hat{\xi}^2}{(n+1)^2\hat{\rho}^2(1+\sqrt{n}\|\hat{A}_B^{-1}\hat{A}_Z\|)}, \tag{6.28}$$

where

$$A_P = A_1 \quad \text{and} \quad A_Z = A_2.$$

Then (PP1) and (DP1) generate solutions $\hat{x}_P^k > 0$ and $\hat{y}^k, \hat{s}_Z^k > 0$, so that $\hat{x} = (\hat{x}_P, 0)$ and $\hat{y}, \hat{s} = (0, \hat{s}_Z)$ solve (LP) and (LD), where $\hat{x}_P = \hat{x}_P^k/\tau^k$, $\hat{y} = \hat{y}^k/\tau^k$, and $\hat{s}_Z = \hat{s}_Z^k/\tau^k$.

Using the criterion in the previous lemma, we can terminate the algorithm once (6.28) holds. From $\theta^0 = 1$, (6.15), and (6.28), this definitely happens if

$$\theta^k \leq \left(1 - \frac{2}{1+\sqrt{1+4\sqrt{2}(n+1)}}\right)^k \leq \frac{(1-\beta)\hat{\xi}^2}{(n+1)^2\hat{\rho}^2(1+\sqrt{n}\|\hat{A}_B^{-1}\hat{A}_Z\|)},$$

which requires

$$k = O(\sqrt{n}) \left(\ln n + \ln \hat{\rho} + \ln(1 + \sqrt{n} \| \hat{A}_B^{-1} \hat{A}_Z \|) - \ln \hat{\xi} \right).$$

We now introduce a lemma which is frequently used later and whose straightforward proof is omitted.

Lemma 6.22 *Let ζ and η be two continuous random variables, with sample space $(0, \infty)$. Define the new variables $\xi = \min(\zeta, \eta)$ and $\rho = \max(\zeta, \eta)$. Then, for any $x \geq 0$,*

$$f_\xi(x) \leq f_\zeta(x) + f_\eta(x) \quad \textit{and} \quad f_\rho(x) \leq f_\zeta(x) + f_\eta(x)$$

where $f_\chi(\cdot)$ is the probability density function (p.d.f.) of a random variable χ.

Let λ have distribution $|N(0,1)|$ with the p.d.f.

$$f_\lambda(x) = \sqrt{2/\pi} \exp(-x^2/2).$$

Then,

$$\begin{aligned}
E(\ln \hat{\xi}) &= \int_0^{1/n} \ln x f_{\hat{\xi}}(x) dx + \int_{1/n}^{\infty} \ln x f_{\hat{\xi}}(x) dx \\
&\geq -\ln n + \int_0^{1/n} \ln x f_{\hat{\xi}}(x) dx \\
&\geq -\ln n - \int_0^{1/n} |\ln x| f_{\hat{\xi}}(x) dx.
\end{aligned}$$

Using Lemma 6.22, we have

$$\begin{aligned}
\int_0^{1/n} |\ln x| f_{\hat{\xi}}(x) dx &\leq n \int_0^{1/n} |\ln x| f_\lambda(x) dx \\
&\leq n\sqrt{2/\pi} \int_0^{1/n} |\ln x| \exp(-x^2/2) dx \\
&\leq n\sqrt{2/\pi} \int_0^{1/n} |\ln x| dx \\
&= -n\sqrt{2/\pi} \int_0^{1/n} \ln x dx \\
&= n\sqrt{2/\pi}(1 + \ln n)/n \\
&< 1 + \ln n.
\end{aligned}$$

Also, we have

$$E[\ln \hat{\rho}] = E[\ln(1 + e^T \hat{x}_1 + e^T \hat{s}_2)] \leq \ln(1 + E[e^T \hat{x}_1 + e^T s_2]) = O(\ln n).$$

Moreover, consider

$$\begin{aligned} E[\ln(1+\sqrt{n}\|\hat{A}_B^{-1}\hat{A}_Z\|)] &\leq E\left[\ln\left(1+\sqrt{n}\left(\sum_{j\in Z}\|\hat{A}_B^{-1}\hat{a}_j\|^2\right)^{1/2}\right)\right] \\ &\leq E\left[\ln\left((1+\sqrt{n})\left(|Z|+\sum_{j\in Z}\|\hat{A}_B^{-1}\hat{a}_j\|^2\right)^{1/2}\right)\right] \\ &= \ln(1+\sqrt{n}) + (1/2)E\left[\ln\sum_{j\in Z}(1+\|\hat{A}_B^{-1}\hat{a}_j\|^2)\right], \end{aligned}$$

where $\hat{a}_j$ is the jth column of $\hat{A}$. Note that $(\hat{A}_B, -\hat{a}_j)$ is a Gaussian matrix, $\hat{A}_B^{-1}\hat{a}_j$ has the distribution of the Cauchy random variables λ_i/λ_0 where λ_i, $i = 0, 1, ..., |\hat{A}_B|$, are independent $N(0,1)$ random variables. Note that $|\hat{A}_B|$, the dimension of $\hat{A}_B$, is less than or equal to m. Without losing generality, we assume $|\hat{A}_B| = m$. Hence

$$1 + \|\hat{A}_B^{-1}\hat{a}_j\|^2 \sim \frac{\lambda_0^2 + \lambda_1^2 + ... + \lambda_m^2}{\lambda_0^2} \sim \frac{\eta_j^2}{\nu_j^2},$$

where η_j^2 has a chi-square distribution with $m+1$ degrees of freedom, $\chi^2(m+1)$, and ν_j is a $|N(0,1)|$ random variable. Thus,

$$\begin{aligned} E[\ln\sum_{j\in Z}(1+\|\hat{A}_B^{-1}\hat{a}_j\|^2)] &\leq E[\ln\frac{\max_{j\in Z}\{\eta_j^2\}}{\min_{j\in Z}\{\nu_j^2\}}] \\ &= E[\ln\max_{j\in Z}\{\eta_j^2\}] - E[\ln(\min_{j\in Z}\{\nu_j\})^2] \\ &:= E[\ln\hat{\eta}^2] - E[\ln(\hat{\nu})^2]. \end{aligned}$$

Using Lemma 6.22 again, we have

$$\begin{aligned} E[\ln\hat{\eta}^2] &= \int_0^\infty \ln x f_{\hat{\eta}^2}(x)dx \\ &\leq \ln(\int_0^\infty x f_{\hat{\eta}^2}(x)dx) \end{aligned}$$

$$
\begin{aligned}
&\leq \ln\left(\int_0^\infty x|Z|f_{\eta_j^2}(x)dx\right) \\
&\leq \ln\left(|Z|\int_0^\infty x f_{\eta^2}(x)dx\right) \\
&= \ln(|Z|(m+1)),
\end{aligned}
$$

where η^2 is a $\chi^2(m+1)$ random variable, whose expected value is $m+1$.

Finally,

$$
\begin{aligned}
E[\ln(\hat{\nu})^2] &= 2E[\ln\hat{\nu}] \\
&= 2\int_0^\infty \ln x f_{\hat{\nu}}(x)dx \\
&= 2\int_0^{1/n} \ln x f_{\hat{\nu}}(x)dx + 2\int_{1/n}^\infty \ln x f_{\hat{\nu}}(x)dx \\
&\geq -2\ln n + 2\int_0^{1/n} \ln x f_{\hat{\nu}}(x)dx \\
&\geq -2\ln n - 2\int_0^{1/n} |\ln x| f_{\hat{\nu}}(x)dx
\end{aligned}
$$

and

$$
\begin{aligned}
\int_0^{1/n} |\ln x| f_{\hat{\nu}}(x)dx &\leq \int_0^{1/n} |\ln x||Z|f_\lambda(x)dx \\
&\leq |Z|\sqrt{2/\pi}\int_0^{1/n} |\ln x|\exp(-x^2/2)dx \\
&\leq |Z|\sqrt{2/\pi}\int_0^{1/n} |\ln x|dx \\
&\leq |Z|\sqrt{2/\pi}(1+\ln n)/n \\
&< 1+\ln n.
\end{aligned}
$$

Therefore, termination occurs on an iteration k, whose expected value is bounded as

$$E[k] \leq O(\sqrt{n}\ln n).$$

Thus we have proved the main result:

Theorem 6.23 *Assume that the homogeneous and self–dual algorithm, using the termination scheme described in the preceding section, is applied to an instance of Todd's degenerate model. Then the expected number of iterations before termination with an exact optimal solution of (LP) is bounded above by $O(\sqrt{n}\ln n)$.*

6.4 Notes

For examples in linear programming randomized algorithms, we cite the recent paper of Seidel [299], who gives a simple randomized algorithm whose expected running time for (LP) is $O(m!n)$, and the references therein.

For the expected running time of the simplex method, a deterministic algorithm, when applied to LP instances generated according to some probability distribution (or class of such distributions), see Adler, Karp and Shamir [2], Adler and Megiddo [3], Borgwardt [60], Megiddo [215], Smale [305], Todd [325], and the references cited there.

"One-step analysis" of a variant of Karmarkar's method can be seen in Nemirovsky [256]. Similar analysis of a primal-dual method can be seen in Gonzaga and Todd [130]. The analysis described in Section 6.1 is due to Mizuno et al. [233]. Let us describe a possible program to make one-step analysis rigorous. Suppose we assume that our original problem (LP) is generated probabilistically as follows: the entries of A are independent standard normal random variables, $b = Ae$ and $c = A^T y + e$ for some y. Then $(x, s) = (e, e)$ is an initial point on the central path $\mathcal{C}$. Moreover, for all of our algorithms, r is a multiple of e and U is a random subspace with the orthogonal transformation-invariant distribution. Hence our analysis holds at the initial iteration. We now apply an algorithm which requires that each iterate lies in $\mathcal{C}$ and hence $r = e$ at each iteration. However, the null space U of $AX^{1/2}S^{-1/2}$ will have a different induced distribution at later iterations. We could hope that before (x, s) gets too close to an optimal pair, this induced distribution is somewhat close to what we have assumed in Section 6.1.2, so that its Radon-Nikodym derivative with respect to our distribution is suitably bounded. In this case, the probability that $\|Pq\|$ exceeds $n^{1/2}\mu$, which is small under the distribution we have assumed, will also be small under the distribution induced by the initial probabilistic generation of (LP). Hence, for most iterations, the improvement in the duality gap would be as in Corollary 6.2. A great number of difficulties need to be resolved before such an approach could succeed. We would probably need bounds on how fast the probabilities in Theorem 6.3 approach 1, and clearly as (x, s) approaches the optimum the induced distribution differs drastically from what we have assumed.

In the meantime, we hope that the one-step nonrigorous analysis has lent some insight into the practical behavior of primal-dual algorithms. Our algorithms using $\mathcal{N} = \mathcal{N}_\infty^-(\beta)$ for β close to 1 are quite close to implemented primal-dual methods, and the result of our nonrigorous analysis, that $O((\log n)\log(1/\epsilon))$ iterations typically suffice, is borne out by several large-scale tests.

The properties of the Gaussian matrix in Section 6.2 and the rotationally-symmetric distribution LP model can be found in Borgwardt [60]. In particular, system (6.5) was discussed in Girko [107], Schmidt and Mattheiss [297], and Todd [326]. Our probabilistic analysis is essentially focused on the initialization and termination of interior-point algorithms. In other words, we have focused on the factor $\xi(A, b, c)$ and $\sigma(A)$ in the worst complexity result of Section 5.2. Essentially, we have proved that, for the above random problem, $\sigma(A) = 1$ with probability 1 and $\xi(A, b, c) \geq p(1/n)$ with high probability. Possible new topics for further research in this area include whether our analysis will hold for other probability distributions and the expected behavior.

Most of results in Section 6.3 are due to Anstreicher et al. [26, 25], where they proved that Theorem 6.23 holds for a more general degenerate model:

Todd's Degenerate Model. Let $A = (A_1, A_2, A_3)$, where A_i is $m \times n_i$, $n_i \geq 1$, $n_1 < m$, $n_1 + n_2 + n_3 = n$, and each component of A is i.i.d. from the $N(0, 1)$ distribution. Let

$$\hat{x} = \begin{pmatrix} \hat{x}_1 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{s} = \begin{pmatrix} 0 \\ 0 \\ \hat{s}_3 \end{pmatrix},$$

where the components of $\hat{x}_1$ and $\hat{s}_3$ are i.i.d. from the $|N(0, 1)|$ distribution. Finally let $b = A\hat{x}, c = \hat{s} + A^T\hat{\pi}$. We assume that either $\hat{\pi} = 0$, or the components of $\hat{\pi}$ are i.i.d. from any distribution with $O(1)$ mean and variance.

6.5 Exercises

6.1 *Prove Proposition 6.7.*

6.2 *Prove Proposition 6.8.*

6.3 *Prove Corollary 6.11.*

6.4 *Let λ_0, λ_1,..., λ_m be independent $N(0, 1)$ random variables and condition on the event that $\lambda_i \leq |\lambda_0|/\sqrt{d}$ for $i = 1, \ldots, m$ ($d = n - m \geq 1$). Prove that $P(n^{-2} \leq |\lambda_0| \leq n^2 | x_1 \geq 0, ..., x_m \geq 0)$ approaches 1 as $n \rightarrow \infty$, where $x_i := |\lambda_0|/\sqrt{d} - \lambda_i$ for $i = 1, \ldots, m$.*

6.5 *Prove Proposition 6.20.*

6.6 *Prove Lemma 6.22.*

Chapter 7

Asymptotic-Case Analysis

Interior-point algorithms generate a sequence of ever-improving points x^0, x^1, ..., x^k, ... approaching the solution set. For many optimization problems, the sequence never exactly reaches the solution set. One theory of iterative algorithms is referred to as local or asymptotic convergence analysis and is concerned with the rate at which the optimality error, $\{r^k\}$, of the generated sequence converges to zero. Obviously, if each iteration of competing algorithms requires the same amount of work, the speed of the convergence reflects the effectiveness of the algorithm. This convergence rate, although it holds locally or asymptotically, allows quantitative evaluation and comparison among different algorithms. It has been widely used in classical optimization and numerical analyses as an efficiency criterion. Generally, this criterion does explain the practical behavior of many iterative algorithms.

In this chapter we analyze the asymptotic convergence rate of iteration sequences generated by some interior-point algorithms. The asymptotic complexity presented in this chapter has several surprising but pleasing aspects. First, the theory is simple in nature. Second, it partially explains the excellent behavior of interior-point algorithms in practice. Third, it provides a tool to identify the strict complementarity partition for the termination method discussed in Chapter 5.

7.1 Rate of Convergence

The asymptotic convergence rate is a rich and yet elementary theory to predict the relative efficiency of a wide class of algorithms. It consists of two measures: the order and ratio of convergence.

7.1.1 Order of convergence

In Section 1.4.4 we have introduced, p, the order of convergence. To ensure that those definitions are applicable to any sequence, they are usually stated in terms of limit superior rather than just limit and $0/0$ is regarded as a finite number. In optimization, these technicalities are rarely necessary since $\{r^k \geq 0\}$ represents a measure towards optimality, and $r^k = 0$ implies that optimality is exactly reached.

We might say that the order of convergence is a measure of how good the tail of $\{r^k\}$ is in the worst case. Large values of p imply the faster convergence of the tail. The convergence of order equal two is called (sub) quadratic convergence. Indeed, if the sequence has order $p > 1$ and the limit

$$\lim_{k\to\infty} \frac{r^{k+1}}{(r^k)^p} = \beta < \infty$$

exists, then there exists a finite K, such that

$$\frac{r^{k+1}}{(r^k)^p} \leq 2\beta$$

or

$$(2\beta)^{1/(p-1)} r^{k+1} \leq [(2\beta)^{1/(p-1)} r^k]^p$$

and

$$(2\beta)^{1/(p-1)} r^K < 1$$

for all $k \geq K$. Thus, if we wish to reduce

$$(2\beta)^{1/(p-1)} r^k \leq \epsilon,$$

we need only

$$k - K = \frac{\log\log(1/\epsilon) + \log\log\left[(2\beta)^{1/(p-1)} r^K\right]^{-1}}{\log p}$$

iterations, since

$$(2\beta)^{1/(p-1)} r^k \leq \left[(2\beta)^{1/(p-1)} r^K\right]^{p^{k-K}}.$$

We also have the following proposition.

Proposition 7.1 *Let the positive sequence $\{r^k\}$ converge to zero. Then, the order of convergence equals*

$$\liminf_{k\to\infty} \frac{\log r^{k+1}}{\log r^k}.$$

Example 7.1 *The sequence with* $r^k = (\alpha)^k$ *where* $0 < \alpha < 1$ *converges to zero with order unity.*

Example 7.2 *The sequence with* $r^k = (\alpha)^{2^k}$ *where* $0 < \alpha < 1$ *converges to zero with order two, and therefore the sequence is quadratically convergent.*

7.1.2 Linear convergence

Most of iterative algorithms have an order of convergence equal to unity, i.e., $p = 1$. It is therefore appropriate to consider this class in greater detail and develop another measure of speed for this class: the ratio of linear convergence, which was introduced in Section 1.4.4.

Linear convergence is the most important type of convergence behavior. A linearly convergence sequence, with convergence ratio β, can be said to have a tail that converges to zero at least as fast as the geometric sequence $M(\beta)^k$ for some fixed positive number M independent of k. Thus, we also call linear convergence geometric convergence.

As a rule, when comparing the relative effectiveness of two competing algorithms both of which produce linearly convergent sequences, the comparison is based on their corresponding convergence ratio—the smaller the ratio the faster the convergence. The ultimate case where $\beta = 0$ is referred to as superlinear convergence. We note immediately that convergence of any order greater than unity is superlinear. It is possible for superlinear convergence to have unity convergence order.

Example 7.3 *The sequence with* $r^k = 1/k$ *converges to zero. The convergence is of order one but it is not linear, since* $\lim(r^{k+1}/r^k) = 1$, *that is,* β *is not strictly less than one.*

Example 7.4 *The sequence with* $r^k = (1/k)^k$ *is of order unity, and it is superlinearly convergent.*

7.1.3 Average order

In practical optimization, the convergence order at each iteration may not be the same during the iterative process. We now define the average order related to the speed of convergence of such a sequence.

Definition 7.1 *Let the positive sequence* $\{r^k\}$ *converge to zero. The average order of convergence of* $\{r^k\}$ *between* k *and* $k + K$ *is defined as*

$$\bar{p} = (\prod_{i=1}^{K} p_i)^{1/K},$$

where p_i is the convergence order from $k+i-1$ to $k+i$.

In other words, the average convergence order during this period is the geometric mean of the orders of each iteration. Using the average order, from k to $k+K$ we should have

$$r^{k+K} = (r^k)^{\bar{p}^K} = (r^k)^{\prod_{i=1}^K p_i}.$$

The right hand side is precisely the accumulated convergence orders between k and $K+k$.

Example 7.5 *The sequence with $r^0 = \alpha$, $0 < \alpha < 1$, $r^{k+1} = (r^k)^2$ if k is even and $r^{k+1} = r^k$ if k is odd. Then, the average converge order between k to $k+2$ is $\sqrt{2}$.*

7.1.4 Error function

In optimization, the decision variables form a vector in $\mathcal{R}^n$, and iterative algorithms generate a sequence $\{x^k\}$ in $\mathcal{R}^n$ space. Thus, if $\{x^k\}$ converges to the optimal solution set, the convergence properties of such a sequence are defined with respect to some particular error or residual function, $r(x)$, that converts the vector sequence into a real number sequence. Such an error function satisfies the property that $r(x) > 0$ for all non-optimal solutions and $r(x) = 0$ for every optimal solution. Hence, the convergence rate of $\{x^k\}$ is represented by the convergence rate of $\{r^k := r(x^k)\}$.

It is common to choose the error function by which to measure convergence as the same function that defines the objective function of the original optimization problem. This means that we measure convergence by how fast the objective converges to its optimum. Alternatively, we sometimes use the function $\min_{x \in X^*} \|x^k - x\|$ which represents the distance from x^k to the optimal solution set X^*.

In the analysis of interior-point algorithms the error function is chosen as the primal-dual or complementary gap $x^T s$, which should be zero at an optimal solution pair. For an optimization problem that possesses a strict complementarity solution, the above two error functions will have the same convergence rate.

7.2 Superlinear Convergence: LP

Consider the predictor-corrector algorithm described in Section 4.5. We will show that this $O(\sqrt{n}\log(R/\epsilon))$-iteration algorithm actually forces quadratic convergence of the duality gap $r^k := (x^k)^T s^k > 0$ to zero. In the context of the present work it is important to emphasize that the notions of convergence, superlinear convergence, or quadratic convergence of the duality gap sequence in no way require the convergence of the iteration sequence $\{(x^k, s^k)\}$.

We follow the same notations in Section 4.5. At the kth predictor step, let $\mu^k = (x^k)^T s^k/n$, $(d_x, d_s) := d(x^k, s^k, 0)$, and

$$\delta^k = \frac{D_x d_s}{\mu^k}. \tag{7.1}$$

If $\theta^k = \bar{\theta}$ is the largest step-size chosen in Algorithm 4.5, then from Lemma 4.17 (note $\delta^k = Pq/\mu^k$),

$$\begin{aligned} 1-\theta^k &\le 1 - \frac{2}{1+\sqrt{1+4\|\delta^k\|/\beta}} \\ &= \frac{\sqrt{1+4\|\delta^k\|/\beta}-1}{1+\sqrt{1+4\|\delta^k\|/\beta}} \\ &= \frac{4\|\delta^k\|/\beta}{(1+\sqrt{1+4\|\delta^k\|/\beta})^2} \\ &\le \|\delta^k\|/\beta \end{aligned} \tag{7.2}$$

and

$$(x^{k+1})^T s^{k+1} \le (1-\theta^k)(x^k)^T s^k \le \frac{\|\delta^k\|}{\beta}(x^k)^T s^k. \tag{7.3}$$

Our goal is to prove that $\|\delta^k\| = O((x^k)^T s^k)$. Then, inequality (7.3) guarantees quadratic convergence of $(x^k)^T s^k$ to zero. (In this section, the big "O" notation represents a positive quantity that may depend on n and/or the original problem data, but which is independent of the iteration k.).

7.2.1 Technical results

We first introduce several technical lemmas. For simplicity, we drop the index k and recall the linear system during the predictor step

$$\begin{aligned} Xd_s + Sd_x &= -Xs \\ Ad_x &= 0 \\ A^T d_y + d_s &= 0. \end{aligned} \tag{7.4}$$

Let $\mu = x^T s/n$ and $z = Xs$. Then from $(x, s) \in \mathcal{N}_2(\alpha)$ we must have

$$(1-\alpha)\mu \le z_j \le (1+\alpha)\mu \quad \text{for} \quad j = 1, 2, ..., n. \tag{7.5}$$

We shall estimate $\|d_x\|$ and $\|d_s\|$. Our present objective is to demonstrate that $\|d_x\| = O(\mu)$ and $\|d_s\| = O(\mu)$. We start by characterizing the solution to (7.4).

Lemma 7.2 *If d_x and d_s satisfy the equation*

$$Xd_s + Sd_x = -Xs$$

and the inequality

$$(d_x)^T d_s \geq 0,$$

then,

$$\|D^{-1}d_x\|^2 + \|Dd_s\|^2 \leq x^T s,$$

where $D = X^{1/2}S^{-1/2}$.

Proof. Multiplying the diagonal matrix $(XS)^{-1/2}$ on both sides of the equation, we have

$$D^{-1}d_x + Dd_s = -(XS)^{1/2}e.$$

Take the norms of both sides.

$$\|D^{-1}d_x + Dd_s\|^2 = x^T s$$

or

$$\|D^{-1}d_x\|^2 + \|Dd_s\|^2 + (d_x)^T d_s = x^T s.$$

Using the linear inequality, we have the lemma.

□

Let $(P, Z) = (P^*, Z^*)$ be the strict complementarity partition of the LP problem. For all k, Theorem 2.15 and relation (5.8) imply that

$$\begin{array}{lll} \xi \leq x_j^k \leq 1/\xi & \text{for} & j \in P \\ \xi \leq s_j^k \leq 1/\xi & \text{for} & j \in Z, \end{array} \tag{7.6}$$

where $\xi < 1$ is a fixed positive quantity independent of k.

Lemma 7.3 *If d_x and d_s are obtained from the linear system (7.4) and $\mu = x^T s/n$, then*

$$\|(d_x)_Z\| = O(\mu) \quad \textit{and} \quad \|(d_s)_P\| = O(\mu).$$

Proof. From Lemma 7.2 and relation (7.6), we obtain

$$\begin{aligned}\|(d_x)_Z\| &= \|D_Z D_Z^{-1}(d_x)_Z\| \\ &\leq \|D_Z\|\|D_Z^{-1}(d_x)_Z\| \\ &\leq \|D_Z\|O(\sqrt{\mu}) \\ &= \|(X_Z S_Z)^{1/2} S_Z^{-1}\|O(\sqrt{\mu}) \\ &= O(\sqrt{\mu})O(\sqrt{\mu}) = O(\mu).\end{aligned}$$

This proves that $\|(d_x)_Z\| = O(\mu)$. The proof that $\|(d_s)_P\| = O(\mu)$ is similar.

□

The proofs of $\|(d_x)_P\| = O(\mu)$ and $\|(d_s)_Z\| = O(\mu)$ are more involved. Towards this end, we first note

$$\begin{aligned} x + d_x &\in \mathcal{R}(D^2 A^T), \\ s + d_s &\in \mathcal{N}(AD^2). \end{aligned} \tag{7.7}$$

This is because from the first equation of (7.4) we have

$$\begin{aligned} S(x + d_x) &= -X d_s \\ X(s + d_s) &= -S d_x. \end{aligned}$$

Thus,

$$\begin{aligned} x + d_x &= -(XS^{-1})d_s = D^2 A^T d_y \\ s + d_s &= -(SX^{-1})d_x = -D^{-2} d_x, \end{aligned}$$

which gives relation (7.7).

Lemma 7.4 *If d_x and d_s are obtained from the linear system (7.4), then $(d_x)_P$ is the solution to the (weighted) least-squares problem*

$$\begin{array}{ll} \min_u & (1/2)\|D_P^{-1}u\|^2 \\ \textit{subject to} & A_P u \;=\; -A_Z(d_x)_Z \end{array}$$

and $(d_s)_Z = A_Z^T v$ and v is the solution to the (weighted) least-squares problem

$$\begin{array}{ll} \min_v & (1/2)||D_Z v||^2 \\ \textit{subject to} & A_P^T v \;=\; -(d_s)_P. \end{array}$$

Proof. From (7.7), we see that

$$x_P + (d_x)_P \in R(D_P^2 A_P^T). \tag{7.8}$$

Since $s_P^* = 0$ for all optimal s^*, we must have $c_P \in R(A_P^T)$. Thus,

$$s_P = c_P - A_P^T y \in R(A_P^T),$$

which implies that

$$x_P = D_P^2 s_P \in R(D_P^2 A_P^T). \tag{7.9}$$

From (7.8) and (7.9) we have

$$(d_x)_P \in R(D_P^2 A_P^T).$$

Moreover, $(d_x)_P$ satisfies the equation

$$A_P(d_x)_P = -A_Z(d_x)_Z.$$

Thus, $(d_x)_P$ satisfies the KKT conditions for the first least squares problem.

Since $AD^2(s + d_s) = -Ad_x = 0$ and $AD^2 s = Ax = b$, it follows that

$$-b = AD^2 d_s = A_P D_P^2 (d_s)_P + A_Z D_Z^2 (d_s)_Z. \tag{7.10}$$

Also, since $x_Z^* = 0$ for all optimal x^*, we have $A_P x_P^* = b$ implying $b \in \mathcal{R}(A_P)$. Therefore, relation (7.10) implies

$$A_Z D_Z^2 (d_s)_Z \in \mathcal{R}(A_P).$$

Moreover, d_y satisfies the equation

$$A_P^T d_y = -(d_s)_P.$$

Thus, d_y satisfies the KKT conditions for the second least squares problem.

$\square$

7.2.2 Quadratic convergence

Theorem 7.5 *If d_x and d_s are obtained from the linear system (7.4) and $\mu = x^T s/n$, then*

$$\|(d_x)_P\| = O(\mu) \quad \textit{and} \quad \|(d_s)_Z\| = O(\mu).$$

Proof. Since the first least-squares problem is always feasible, there must be a *feasible* $\bar{u}$ such that

$$\|\bar{u}\| = O(\|(d_x)_Z\|),$$

which together with Lemma 7.3 implies

$$\|\bar{u}\| = O(\mu).$$

Furthermore, from Lemma 7.4 and relations (7.5) and (7.6)

$$\begin{array}{rcl}
\|(d_x)_P\| & = & \|D_P D_P^{-1}(d_x)_P\| \\
& \le & \|D_P\|\|D_P^{-1}(d_x)_P\| \\
& \le & \|D_P\|\|D_P^{-1}\bar{u}\| \\
& \le & \|D_P\|\|D_P^{-1}\|\|\bar{u}\| \\
& = & \|(X_P S_P)^{-1/2}X_P\|\|(X_P S_P)^{1/2}X_P^{-1}\|\|\bar{u}\| \\
& \le & \|(X_P S_P)^{-1/2}\|\|X_P\|\|(X_P S_P)^{1/2}\|\|X_P^{-1}\|\|\bar{u}\| \\
& = & O(\|\bar{u}\|) = O(\mu).
\end{array}$$

Similarly, we can prove the second statement of the theorem.

□

Theorem 7.5 indicates that at the kth predictor step, d_x^k and d_s^k satisfy

$$\|(d_x^k)_P\| = O(\mu^k) \quad \text{and} \quad \|(d_s^k)_Z\| = O(\mu^k), \tag{7.11}$$

where $\mu^k = (x^k)^T s^k/n$. We are now in a position to state our main result.

Theorem 7.6 *Let $\{(x^k, s^k)\}$ be the sequence generated by Algorithm 4.5. Then, with constants $0 < \beta \le 1/4$ and $\alpha = 2\beta$:*

i) *the Algorithm has iteration complexity $O(\sqrt{n}\log(R/\epsilon))$;*

ii) $1 - \theta^k = O((x^k)^T s^k)$;

iii) $(x^k)^T s^k \to 0$ *quadratically.*

Proof. The proof of (i) is in Theorem 4.18, which also establishes

$$\lim_{k\to\infty} \mu^k = 0.$$

From relation (7.1), Lemma 7.3 and Theorem 7.5 we have

$$\|\delta^k\| = \|D_x d_s/\mu^k\| \le O((x^k)^T s^k),$$

which together with inequality (7.2) establishes (ii).

From inequality (7.3) we see that (ii) implies (iii). This proves the theorem.

□

7.3 Superlinear Convergence: LCP

Given $M \in \mathcal{R}^{n\times n}$ and $q \in \mathcal{R}^n$, recall that the LCP is to find a pair $x, s \in \mathcal{R}^n$ such that

$$s = Mx + q, \quad x^T s = 0, \quad \text{and} \quad x, s \geq 0.$$

In this section, we consider the monotone LCP extension of the predictor-corrector LP algorithm. We show that this $O(\sqrt{n}\log(R/\epsilon))$-iteration algorithm for the monotone LCP actually possesses quadratic convergence assuming

Assumption 7.1 *The monotone LCP possesses a strict complementarity solution.*

This assumption is restrictive since in general it does not hold for the monotone LCP. We will actually show by example, however, that assumption 7.1 appears to be necessary in order to achieve superlinear convergence for the algorithm.

Again, the LCP being monotone means that the iterate direction

$$d_s = Md_x \quad \text{implies} \quad d_x^T d_s \geq 0.$$

Note that for LP, we have $d_x^T d_s = 0$. This is the only difference between LP and LCP analyses. Almost all technical results on iterate directions developed for LP ($d_x^T d_s = 0$) hold for the monotone MCP ($d_x^T d_s \geq 0$).

7.3.1 Predictor-corrector algorithm for LCP

In this section, we briefly describe the predictor-corrector LCP algorithm. We let $\overset{\circ}{\mathcal{F}}$ denote the collection of all strictly feasible points (x, s). Consider the neighborhood

$$\mathcal{N}_2(\alpha) = \{(x, s) \in \overset{\circ}{\mathcal{F}} : \|Xs/\mu - e\| \leq \alpha\},$$

where $\mu = x^T s/n$ and α is a constant between 0 and 1.

To begin with choose $0 < \beta \leq 1/4$ (a typical choice would be $1/4$). All search directions d_x, d_s, and d_y will be defined as the solutions of the following system of linear equations

$$\begin{aligned} Xd_s + Sd_x &= \gamma\mu e - Xs \\ d_s &= Md_x, \end{aligned} \tag{7.12}$$

where $0 \leq \gamma \leq 1$. To show the dependence of $d = (d_x, d_s)$ on the pair (x, s) and parameter γ, we write $d = d(x, s, \gamma)$.

A typical iteration of the algorithm proceeds as follows. Given $(x^k, s^k) \in \mathcal{N}_2(\beta)$, we solve system (7.12) with $(x, s) = (x^k, s^k)$ and $\gamma = 0$, i.e., $(d_x, d_s) = d(x^k, s^k, 0)$. For some step length $\theta \geq 0$ let

$$x(\theta) = x^k + \theta d_x, \quad s(\theta) = s^k + \theta d_s,$$

and $\mu(\theta) = x(\theta)^T s(\theta)/n$. This is the predictor step.

Again, we can choose the *largest* step length $\theta = \bar{\theta} \leq 1$ such that $(x(\theta), s(\theta)) \in \mathcal{N}_2(\alpha)$ for $\alpha = \beta + \tau$ and some $0 < \tau \leq \beta$, and let

$$x' = x(\bar{\theta}) \quad \text{and} \quad s' = s(\bar{\theta}).$$

We can compute $\bar{\theta}$ by finding the roots of a quartic equation.

Next we solve system (7.12) with $(x, s) = (x', s') \in \mathcal{N}(\beta + \tau)$, $\mu' = (x')^T s'/n$, and $\gamma = 1$, i.e., $(d'_x, d'_s) = d(x', s', 1)$. Let $x^{k+1} = x' + d'_x$ and $s^{k+1} = s' + d'_s$. This is the corrector (or centering) step.

Similar to Lemma 4.16, for all k we can show that

$$(x^k, s^k) \in \mathcal{N}_2(\beta) \tag{7.13}$$

as long as $0 < \beta \leq 1/4$ and $0 < \tau \leq \beta$, and

$$\begin{array}{rcl} (x')^T s' & = & (1 - \theta^k)(x^k)^T s^k + (\theta^k)^2 (d_x)^T d_s \\ (x^{k+1})^T s^{k+1} & = & (x')^T s' + (d'_x)^T d'_s. \end{array} \tag{7.14}$$

One can also show that

$$\begin{array}{rcl} (d_x)^T d_s & \leq & (x^k)^T s^k/4 \\ (d'_x)^T d'_s & \leq & (x')^T s'/(8n). \end{array} \tag{7.15}$$

Let $\delta^k = D_x d_s/\mu^k$ in the predictor step. Then, we can show that

$$\|\delta^k\| \leq \sqrt{2}n/4, \tag{7.16}$$

and the following lemma, which resembles Lemma 4.17.

Lemma 7.7 *If $\theta^k := \bar{\theta}$ is the largest θ such that $(x(\theta), s(\theta)) \in \mathcal{N}_2(\alpha)$ with $\alpha = \beta + \tau$ and $0 < \tau \leq \beta$, then*

$$\theta^k \geq \frac{2}{1 + \sqrt{1 + 4\|\delta^k\|/\tau}}.$$

Clearly, this lemma together with (7.14), (7.15) and (7.16) implies that the iteration complexity of the algorithm is $O(\sqrt{n}\log(R/\epsilon))$ for a constant $0 < \tau \leq \beta$. Note again that

$$1 - \theta^k \leq \frac{\|\delta^k\|}{\tau}. \tag{7.17}$$

Relations (7.14), (7.15), (7.16), and (7.17), and Lemma 7.7, imply

$$\mu^{k+1} \le (1+1/8n)(\frac{\|\delta^k\|}{\tau}\mu^k + (d_x)^T d_s/n) = (1+1/8n)(\frac{\|D_x d_s\|}{\tau} + \frac{\|D_x d_s\|}{\sqrt{n}}). \tag{7.18}$$

From (7.18), we see that if

$$\|d_x\| = O(\mu^k) \quad \text{and} \quad \|d_s\| = O(\mu^k),$$

then the complementary gap converges to zero quadratically.

7.3.2 Technical results

For a LCP possessing a strict complementarity solution, a unique partition P and Z, where $P \cap Z = \{1, 2, ..., n\}$ and $P \cup Z = \emptyset$, exists such that $x^*_Z = 0$ and $s^*_P = 0$ in every complementarity solution and at least one complementarity solution has $x^*_P > 0$ and $s^*_Z > 0$. We can also prove that relation (7.6) holds for the sequence generated by the predictor-corrector MCP algorithm. Let $\mu = x^T s/n$ and $z = Xs$. We must also have relation (7.5) if $(x, s) \in \mathcal{N}_2(\alpha)$.

We now introduce several technical lemmas. For simplicity, we drop the index k and recall the linear system during the predictor step

$$\begin{array}{rcl} Xd_s + Sd_x &=& -Xs \\ d_s &=& Md_x. \end{array} \tag{7.19}$$

Define $D = X^{1/2}S^{-1/2}$. We now estimate $\|d_x\|$ and $\|d_s\|$. Since M is monotone, i.e., $(d_x)^T d_s \ge 0$, both Lemma 7.2 and the following lemma hold.

Lemma 7.8 *If d_x and d_s are obtained from the linear system (7.19), and $\mu = x^T s/n$, then*

$$\|(d_x)_Z\| = O(\mu) \quad \textit{and} \quad \|(d_s)_P\| = O(\mu).$$

The proofs of $\|(d_x)_P\| = O(\mu)$ and $\|(d_s)_Z\| = O(\mu)$ are again more involved. We first note

$$\begin{array}{rcl} S(x + d_x) &=& -Xd_s, \\ X(s + d_s) &=& -Sd_x, \end{array}$$

and therefore

$$\begin{array}{rcl} x + d_x &=& -(XS^{-1})d_s = -D^2 d_s \\ s + d_s &=& -(X^{-1}S)d_x = -D^{-2}d_x. \end{array} \tag{7.20}$$

Before proceeding, we need some results regarding (non-symmetric) positive semi-definite (PSD) matrices that may be of independent interest. In what follows, we will consider M to be partitioned (following a re-ordering of rows and columns) as

$$M = \begin{pmatrix} M_{PP} & M_{PZ} \\ M_{ZP} & M_{ZZ} \end{pmatrix}. \tag{7.21}$$

Lemma 7.9 *Let M be a PSD matrix, partitioned as in (7.21). Then $M_{PP}x_P = 0$ if and only if $M_{PP}^T x_P = 0$. Furthermore, $M_{PP}x_P = 0$ implies that $(M_{ZP} + M_{PZ}^T)x_P = 0$.*

Proof. Let $x = (x_P^T, 0^T)^T$. If either $M_{PP}x_P = 0$ or $M_{PP}^T x_P = 0$, then $x^T M x = 0$, so x is a global minimizer of the quadratic form $y^T M y$. Consequently $(M + M^T)x = 0$, which is exactly

$$\begin{aligned} (M_{PP} + M_{PP}^T)x_P &= 0 \\ (M_{ZP} + M_{PZ}^T)x_P &= 0. \end{aligned}$$

□

Lemma 7.10 *Let M be a PSD matrix, partitioned as in (7.21). Then*

$$\mathcal{R}\begin{pmatrix} M_{PP} & M_{PZ} \\ 0 & I \end{pmatrix} = \mathcal{R}\begin{pmatrix} M_{PP}^T & M_{ZP}^T \\ 0 & -I \end{pmatrix}.$$

Proof. From the fundamental theorem of linear algebra, it is equivalent to prove that

$$\mathcal{N}\begin{pmatrix} M_{PP}^T & 0 \\ M_{PZ}^T & I \end{pmatrix} = \mathcal{N}\begin{pmatrix} M_{PP} & 0 \\ M_{ZP} & -I \end{pmatrix},$$

where $\mathcal{N}(\cdot)$ denotes the null space of a matrix. To begin, assume that

$$\begin{pmatrix} M_{PP}^T & 0 \\ M_{PZ}^T & I \end{pmatrix}\begin{pmatrix} x_P \\ x_Z \end{pmatrix} = 0. \tag{7.22}$$

From Lemma 7.9, $M_{PP}x_P = 0$. Also $x_Z = -M_{PZ}^T x_P$, so showing that $M_{ZP}x_P - x_Z = 0$ is equivalent to showing that $(M_{ZP} + M_{PZ}^T)x_P = 0$, which also holds by Lemma 7.9. Thus

$$\begin{pmatrix} M_{PP} & 0 \\ M_{ZP} & -I \end{pmatrix}\begin{pmatrix} x_P \\ x_Z \end{pmatrix} = 0\ . \tag{7.23}$$

The argument that (7.23) implies (7.22) is similar.

□

7.3.3 Quadratic convergence

Now we can establish

Lemma 7.11 *If d_x and d_s are obtained from the linear system (7.19), and $\mu = x^T s/n$, then $u = (d_x)_P$ and $v = (d_s)_Z$ are the solutions to the (weighted) least-squares problem*

$$\begin{array}{ll} \min_{u,v} & (1/2)\|D_P^{-1}u\|^2 + (1/2)\|D_Z^1 v\|^2 \\ s.t. & M_{PP}u = -M_{PZ}(d_x)_Z + (d_s)_P \\ & M_{ZP}u - v = -M_{ZZ}(d_x)_Z. \end{array} \tag{7.24}$$

Proof. Note that from (7.19), $u = (d_x)_P$, $v = (d_s)_Z$ is certainly feasible in the problem (7.24). Next, from (7.19) and (7.20), we see that

$$\begin{array}{rcl} x_P + (d_x)_P & = & -D_P^2 M_{B\cdot} d_x \\ s_Z + (d_s)_Z & = & -D_Z^{-2}(d_x)_Z. \end{array} \tag{7.25}$$

Since $s_P^* = 0$ for all optimal s^*, with $x_Z^* = 0$, we must have $q_P = -M_{PP}x_P^* \in \mathcal{R}(M_{PP})$. Therefore,

$$D_P^{-2}x_P = s_P = M_{B\cdot}x + q_P = M_{PP}(x_P - x_P^*) + M_{PZ}x_Z.$$

Substituting this into the first equation of (7.25) obtains

$$D_P^{-2}(d_x)_P = -M_{PP}(x_P - x_P^* + (d_x)_P) - M_{PZ}(x_Z + (d_x)_Z). \tag{7.26}$$

Also $s_Z = D_Z^{-2}x_Z$, which substituted into the second equation of (7.25) yields

$$D_Z^2(d_s)_Z = -x_Z - (d_x)_Z. \tag{7.27}$$

Then (7.26) and (7.27) together imply that

$$\begin{pmatrix} D_P^{-2}(d_x)_P \\ D_Z^2(d_s)_Z \end{pmatrix} \in \mathcal{R}\begin{pmatrix} M_{PP} & M_{PZ} \\ 0 & I \end{pmatrix}.$$

Applying Lemma 7.10, we conclude that

$$\begin{pmatrix} D_P^{-2}(d_x)_P \\ D_Z^2(d_s)_Z \end{pmatrix} \in \mathcal{R}\begin{pmatrix} M_{PP}^T & M_{ZP}^T \\ 0 & -I \end{pmatrix},$$

which shows exactly that $u = (d_x)_P$, $v = (d_s)_Z$ satisfies the KKT conditions for optimality in the least squares problem (7.24).

$\square$

Theorem 7.12 *If d_x and d_s are obtained from the linear system (7.19), and $\mu = x^T s/n$, then $\|d_x\| = O(\mu)$ and $\|d_s\| = O(\mu)$.*

Proof. Due to Lemma 7.8, we only need to prove

$$\|(d_x)_P\| = O(\mu) \quad \text{and} \quad \|(d_s)_Z\| = O(\mu).$$

Since the least-squares problem (7.24) is always feasible, there must be *feasible* $\bar{u}$ and $\bar{v}$ such that

$$\|\bar{u}\| = O(\|(d_x)_Z\| + \|(d_s)_P\|) \quad \text{and} \quad \|\bar{v}\| = O(\|(d_x)_Z\| + \|(d_s)_P\|),$$

which together with Lemma 7.8 implies $\|\bar{u}\| = O(\mu)$ and $\|\bar{v}\| = O(\mu)$. Furthermore, from Lemma 7.11 and relations (7.5) and (7.6),

$$\begin{aligned}
&\|(d_x)_P\|^2 + \|(d_s)_Z\|^2 \\
&= \|D_P D_P^{-1}(d_x)_P\|^2 + \|D_Z^{-1} D_Z (d_s)_Z\|^2 \\
&\le \|D_P^2\| \, \|D_P^{-2}(d_x)_P\|^2 + \|D_Z^{-2}\| \, \|D_Z(d_s)_Z\|^2 \\
&= \|(X_P S_P)^{-1} X_P^2\| \, \|D_P^{-1}(d_x)_P\|^2 + + \|(X_Z S_Z)^{-1} S_Z^2\| \, \|D_Z(d_s)_Z\|^2 \\
&\le \left(\|(X_P S_P)^{-1} X_P^2\| + \|(X_Z S_Z)^{-1} S_Z^2\|\right) \left(\|D_P^{-1}(d_x)_P\|^2 + \|D_Z(d_s)_Z\|^2\right) \\
&\le \left(\|(X_P S_P)^{-1} X_P^2\| + \|(X_Z S_Z)^{-1} S_Z^2\|\right) \left(\|D_P^{-1}\bar{u}\|^2 + \|D_Z \bar{v}\|^2\right) \\
&\le \left(\|(X_P S_P)^{-1} X_P^2\| + \|(X_Z S_Z)^{-1} S_Z^2\|\right) \left(\|D_P^{-2}\| \, \|\bar{u}\|^2 + \|D_Z^2\| \, \|\bar{v}\|^2\right) \\
&\le O(1/\mu) \left(\|D_P^{-2}\| \, \|\bar{u}\|^2 + \|D_Z^2\| \, \|\bar{v}\|^2\right) \\
&= O(\mu) \left(\|D_P^{-2}\| + \|D_Z^2\|\right) \\
&= O(\mu) \left(\|(X_P S_P) X_P^{-2}\| + \|(X_Z S_Z) S_Z^{-2}\|\right) \\
&= O(\mu^2).
\end{aligned}$$

□

The above theorem leads to the result described in Theorem 7.6 for the predictor-corrector LCP algorithm. The following proposition concerns assumption 7.1.

Proposition 7.13 *There is a monotone LCP problem, where a strict complementarity solution does not exist, for which the predictor-corrector algorithm or affine scaling algorithm possesses no superlinear convergence.*

Proof. Consider the simple monotone LCP with $n = 1$, $M = 1$ and $q = 0$. The unique complementarity solution is $s = x = 0$, which is not strictly complementary. Note that the feasible solution $s = x = \epsilon$ is a perfectly centered pair for any $\epsilon > 0$. The direction in the predictor step (or affine scaling algorithm) is

$$d_x = -x/2 \quad \text{and} \quad d_s = -s/2.$$

Thus, even taking the step size $\theta = 1$, the new solution will be $s = x = \epsilon/2$. Thus, the complementarity slackness sequence is reduced at most linearly, with constant $1/4$, which proves the proposition.

□

7.4 Quadratically Convergent Algorithms

The predictor-corrector algorithm described in previous sections needs to solve two systems of linear equations or two least squares problems—one in the predictor step and one in the corrector step. If one counts each iteration as solving one system of linear equations, as is usually done in the classical analysis of interior-point algorithms, the average order of convergence of this algorithm is only $\sqrt{2}$. In this section we further show how to construct an algorithm whose order of convergence exactly equals 2. We also show that the solution sequence generated by the algorithm is a Cauchy, and therefore convergent, sequence.

7.4.1 Variant 1

An iteration of the variant proceeds as follows. Given $(x^k, s^k) \in \mathcal{N}_2(\beta)$, we perform $T(\geq 1)$ successive predictor steps followed by one corrector step, where in tth predictor step of these T steps we choose $\tau = \tau_t > 0$ where

$$\sum_{t=1}^{T} \tau_t = \beta. \tag{7.28}$$

In other words, at the tth predictor step of these T steps, we solve system (7.12) with $\mu' = (x')^T s'/n$ and $(x, s) = (x', s') \in \mathcal{N}(\beta + \tau_1 + ... + \tau_{t-1})$ (the initial $(x', s') = (x^k, s^k) \in \mathcal{N}(\beta)$) and $\gamma = 0$, i.e., $(d_x, d_s) = d(x', s', 0)$. For some $\theta > 0$ let

$$x(\theta) = x' + \theta d_x, \quad s(\theta) = s' + \theta d_s \quad \text{and} \quad \mu(\theta) = (x(\theta))^T s(\theta)/n.$$

Our specific choice, $\bar{\theta}$, for θ is similar as before: the largest θ such that $(x(\theta), s(\theta)) \in \mathcal{N}_2(\alpha)$ for

$$\alpha = \beta + \tau_1 + ... + \tau_{t-1} + \tau_t.$$

From the first inequality in (7.14), the fact $\bar{\theta} \leq 1$, (7.17), and Theorem 7.12 we have

$$\begin{aligned} \mu(\bar{\theta}) &\leq \frac{\|D_x d_s\|}{\tau_t} + \frac{\|D_x d_s\|}{\sqrt{n}} \\ &\leq \frac{M(\mu')^2}{\tau_t} \end{aligned} \tag{7.29}$$

for some fixed positive quantity M independent of k and t. Now update $x' := x(\bar{\theta})$ and $s' := s(\bar{\theta})$.

After T predictor steps we have $(x', s') \in \mathcal{N}_2(2\beta)$. Now we perform one corrector step as before to generate

$$(x^{k+1}, s^{k+1}) \in \mathcal{N}_2(\beta).$$

Based on the previous lemmas and results, each predictor step within an iteration achieves quadratic convergence order for any positive constant sequence $\{\tau_t\}$ satisfying (7.28). For example, one natural choice would be $\tau_t = \beta/T$ for $t = 1, 2, ..., T$. Since each iteration solves $T + 1$ systems of linear equations, the average order of the convergence of the complementary gap to zero in Variant 1 is $2^{T/(T+1)}$ per linear system solver for any constant $T \geq 1$.

Theorem 7.14 *Variant 1 generates a sequence $\{x^k, s^k\}$ such that the average convergence order is $2^{T/(T+1)}$ per linear system solver for any constant $T \geq 1$.*

7.4.2 Variant 2

Now we develop a new variant where we let $T = \infty$, that is, no corrector step is needed anymore in the rest of the iterations of the algorithm. The algorithm becomes the pure Newton method or the primal-dual affine scaling algorithm.

After $(x^K, s^K) \in \mathcal{N}(\beta)$ for some finite K, we perform only the predictor step, where we choose $\tau = \tau_t > 0$ satisfying (7.28). One natural choice will be

$$\tau_t = \beta(1/2)^t \quad \text{for} \quad t = 1, 2, ...$$

For simplicity, let us reset $K := 1$. Then, in the kth iteration we solve system (7.12) with

$$(x, s) = (x^k, s^k) \in \mathcal{N}(\beta + \sum_{t=1}^{k-1} \tau_t) \qquad \left(\text{where } \sum_{t=1}^{0} \tau_t := 0\right)$$

and $\gamma = 0$, i.e., $(d_x, d_s) = d(x^k, s^k, 0)$. For some $\theta > 0$ let

$$x(\theta) = x^k + \theta d_x, \quad s(\theta) = s^k + \theta d_s. \tag{7.30}$$

Our specific choice for θ is $\bar{\theta}$, the largest θ such that $(x(\theta), s(\theta)) \in \mathcal{N}_2(\alpha)$ for

$$\alpha = \beta + \sum_{t=1}^{k} \tau_t.$$

Now directly update

$$x^{k+1} := x(\bar{\theta}) \quad \text{and} \quad s^{k+1} := s(\bar{\theta}). \tag{7.31}$$

Theorem 7.15 *Let $(x^K)^T s^K$ be small enough. Then, Variant 2 generates a sequence $\{x^k, s^k\}$ with $k \geq K$ such that*

i) *the order of the convergence of the complementary gap to zero equals at least* 2,

ii) *$\{x^k, s^k\}$ is a Cauchy, and therefore convergent, sequence.*

Proof. At the kth iteration ($k \geq K := 1$) we have from (7.29)

$$(x^{k+1})^T s^{k+1} \leq \frac{M[(x^k)^T s^k]^2}{\tau_k} = M[(x^k)^T s^k]^2 2^k / \beta,$$

or

$$\log_2[(x^{k+1})^T s^{k+1}] \leq 2\log_2[(x^k)^T s^k] + \log_2(M/\beta) + k. \tag{7.32}$$

For $(x^K)^T s^K$ small enough, the inequality (7.32) implies that $\{\log_2[(x^k)^T s^k]\}$ is a geometric sequence (with base close to 2, say 1.5) tending to $-\infty$. Since k is just an arithmetic sequence and $\log_2(M/\beta)$ is fixed, we should have

$$\lim_{k\to\infty} \frac{k + \log_2(M/\beta)}{\log_2[(x^k)^T s^k]} \to 0, \tag{7.33}$$

geometrically. This implies that

$$\liminf_{k\to\infty} \frac{\log[(x^{k+1})^T s^{k+1}]}{\log[(x^k)^T s^k]} \geq 2,$$

which from Proposition 7.1 proves (i).

Now from Theorem 7.12, (7.30) and (7.31)

$$\|x^{k+1} - x^k\| = \bar{\theta}\|d_x^k\| < \|d_x^k\| = O(\mu^k) = O((x^k)^T s^k / n)$$

and

$$\|s^{k+1} - s^k\| = \bar{\theta}\|d_s^k\| < \|d_s^k\| = O(\mu^k) = O((x^k)^T s^k / n).$$

Hence, $\{x^k, s^k\}$ must be a Cauchy sequence, since $\{(x^k)^T s^k\}$ converges to zero superlinearly from (i). This proves (ii).

□

To actually achieve the order 2 of convergence of the primal-dual gap, we need to decide when to start the primal-dual affine scaling procedure described in Variant 2. Note from (7.32) that as long as $\{\log_2[(x^k)^T s^k]\}$ is a geometric sequence with base close to 1.5 tending to $-\infty$, we shall have the order 2 of convergence of $\{(x^k)^T s^k/(x^0)^T s^0\}$ to zero. Thus, we can start the procedure at any time when $(x^K)^T s^K < 1$. Again for simplicity, let $K := 1$. Then we add a *safety check* to see if for $k = 1, 2, ...$

$$\begin{array}{rcl} |\log[(x^{k+1})^T s^{k+1}]|/|\log[(x^k)^T s^k]| & \geq & 1.5 \\ (x^{k+1})^T s^{k+1}/(x^k)^T s^k & \leq & 1 - \Omega(1/\sqrt{n}). \end{array} \tag{7.34}$$

If both inequalities in (7.34) are satisfied, we *continue* the predictor step. Otherwise we conclude that $(x^K)^T s^K$ was not "small enough," and we do one *corrector* step and then *restart* the predictor procedure. This safety check will guarantee that the algorithm maintains the polynomial complexity $O(\sqrt{n}\log(R/\epsilon))$ and achieves the order 2 of the convergence of the complementary gap to zero, since eventually no corrector (or centering) step is needed anymore in the rest of the iterations, according to the theorem.

Thus, we have shown that after the complementary gap becomes smaller than a fixed positive number, the pure primal-dual Newton method with the step-size choice in Variant 2 generates an iteration sequence which not only polynomially converges to an optimal solution pair, but one whose convergence is actually quadratic.

In practice, the step size, θ^k, in the predictor step can be simply chosen as the bound given in Lemma 7.7. Thus, no quartic equation solver is needed to guarantee our theoretical results. Also we see that the step size in Variant 2 converges to 1 superlinearly while the solution sequence remains "centered," i.e., $(x^k, s^k) \in \mathcal{N}_2(2\beta)$, without any explicit centering. This may partially explain why the large step strategy does not hurt the convergence of the algorithm in practice.

7.5 Notes

The issue of the asymptotic convergence of interior-point algorithms was first raised in Iri and Imai [156]. They showed that their (product) barrier function method with an exact line search procedure possesses quadratic convergence for nondegenerate LP. Then, Yamashita [371] showed that a variant of this method possesses both polynomial $O(nL)$ complexity and quadratic convergence for nondegenerate LP, and Tsuchiya and Tanabe [343] showed that Iri and Imai's method possesses quadratic convergence under a weaker nondegeneracy assumption.

Zhang, Tapia and Dennis [391, 390] first showed that a primal-dual algorithm exhibits $O(nL)$ complexity, with superlinear convergence under the assumption of the convergence of the iteration sequence, and quadratic convergence under the assumption of nondegeneracy. Kojima, Megiddo and Mizuno [179], Ji, Potra and Huang [163], and Zhang, Tapia and Potra [392] also showed quadratic convergence of a path-following algorithm for linear complementarity problems under the nondegeneracy assumption. McShane [213] showed that a primal-dual algorithm exhibits $O(\sqrt{n}L)$ complexity, with superlinear convergence under the assumption of the convergence of the iteration sequence. Other algorithms, interior or exterior, with quadratic convergence for nondegenerate LP include Coleman and Li's [70]. Some negative results on the asymptotic convergence of Karmarkar's original algorithm and a potential reduction method (with separate primal and dual updates) were given by Bayer and Lagarias [39], and Gonzaga and Todd [130], respectively.

Quadratic convergence for general LP, assuming neither the convergence of the iteration sequence nor nondegeneracy, was first established by Ye, Güler, Tapia and Zhang [383], and independently by Mehrotra [220] and Tsuchiya [341]. The algorithm of Mehrotra, and Ye et al., is based on the predictor-corrector algorithm of Mizuno et al. (also see Barnes). As we mentioned before, if one counts each iteration as solving one system of linear equations, as is usually done in the analysis of interior-point algorithms, the (average) order of convergence of the algorithm is only $\sqrt{2}$. Tsuchiya's result is based on Iri and Imai's $O(nL)$ method which requires knowledge of the exact optimal objective value in advance. A standard way of dealing with this difficulty is to integrate the primal and dual problems into a single LP problem, whose size is twice that of the original problem. Thus, the (average) order of convergence would actually be below $\sqrt{2}$. The convergence order 2 algorithm for general LP, counting each iteration as solving one system of linear equations of the size of the original problem, was first given in Ye [378].

Quadratic convergence for the monotone LCP, described in section 7.3, is based on Ye and Anstreicher [382]. They also give an example to show that the predictor step cannot achieve superlinear convergence if the LCP has no a strictly complementary solution. Monteiro and Wright [247] further shows that any algorithm that behaves like Newton's method near the solution set cannot converge superlinearly when applied to an LCP which does not have a strictly complementary solution.

Recently, Mizuno [231] proposed a superlinearly convergent infeasible-interior-point algorithm for geometrical LCPs without the strictly complementary condition.

Most recently, Gonzaga and Tapia [129, 128] proved that the iteration sequence (x^k, y^k, s^k) generated by the predictor-corrector algorithm converges to an optimal solution on the interior of the optimal face. Consequently, Luo et al. [201] announced a genuine quadratically convergent algorithm. Bonnans and Gonzaga [59] developed a simplified predictor-corrector where the same Jacobian matrix is used in both the predictor and corrector steps within one iteration. The convergence order of the complementary gap to zero is $T + 1$ where T is the number of predictor steps in each iteration. Luo, Sturm and Zhang [199] analyzed the superlinear convergence behavior of the predictor-corrector algorithm for positive semi-definite programming.

In the analysis of interior-point algorithms, the error function is chosen as the primal-dual gap or complementary $x^T s$ which should be zero at an optimal solution pair. For an optimization problem that possesses a strict complementarity solution, this error bound will lead to the same convergence rate for distances from iterates to the solution set, see Hoffman [149], Mangasarian [207, 208], and Luo and Tseng [200], and references therein.

7.6 Exercises

7.1 *Prove Proposition 7.1.*

7.2 *Prove that the sequence with* $r^k = (1/k)^k$ *is of order unity, and it is superlinearly convergent.*

7.3 *Prove relation (7.6).*

7.4 *Prove Lemma (7.7).*

7.5 *Prove that the safety check described at the end of Section 7.4 works.*

Chapter 8

Nonlinear Convex Optimization

In this chapter, we discuss interior-point algorithms for solving nonlinear convex optimization problems. These algorithms illustrate how widely applicable potential functions and interior-point algorithms could be in solving broader optimization problems.

8.1 von Neumann Economic Growth Problem

Consider the von Neumann economic growth (NEG) problem:

$$\gamma^* := \max\{\gamma \mid \exists\, y \neq 0 : \quad y \geq 0, \quad (B - \gamma A)y \geq 0\},$$

where $A = \{a_{ij} \geq 0\}$ and $B = \{b_{ij} \geq 0\}$ are two given nonnegative matrices in $\mathcal{R}^{m \times n}$, and

Assumption 8.1 *A has no all-zero columns.*

Note that γ^* is bounded above based on this assumption. In fact,

$$\gamma^* \leq \max_j \frac{\sum_{i=1}^m b_{ij}}{\sum_{i=1}^m a_{ij}}.$$

(The NEG problem is a fractional program, and the results developed in this section are applicable to other fractional programs.)

The NEG problem has an economic interpretation. Each row-index i stands for a "good," while each column index j stands for a "process." Process j can convert a_{ij} units of good i, in one time period, into b_{ij} units of good i. So a process uses goods as materials or inputs, and gives goods as products or outputs. Matrix B is referred as output matrix, and A is the input matrix. Component y_j of y denotes the "intensity" by which we let process j work. Vector By gives the amounts of outputs produced, and Ay gives the amounts of inputs consumed, during one time period. Then, γ represents the growth factor at intensity y. So $By \geq \gamma Ay$ requires that, for each good i, the amount of good i produced in period t is at least the amount of good i required in period $(t+1)$ with the growth factor γ. The NEG problem is to find the largest growth factor using an optimal intensity vector.

There is a related dual NEG problem,

$$\eta^* := \min\{\eta \mid \exists\, x \neq 0 : \quad x \geq 0, \quad (\eta A - B)^T x \geq 0\}.$$

We further assume that

Assumption 8.2 *B has no all-zero rows.*

Then, η^* is also bounded below. But a duality overlap may exist, i.e., it is possible $\eta^* < \gamma^*$. However, under the following irreducibility assumption the model is well behaved: $\eta^* = \gamma^*$.

Assumption 8.3 *There is no (proper) subset S of the rows and no subset T of the columns such that $A_{ij} = 0$ for all $i \in S$ and all $j \in T$, and such that for all $i \in \{1, 2, ..., m\} \setminus S$, $B_{ij} > 0$ for some $j \in T$,*

Moreover, the γ-level set,

$$\Gamma(\gamma) := \{y \in R^n : \ e^T y = 1, \quad y \geq 0, \quad (B - \gamma A)y \geq 0\}, \tag{8.1}$$

has a nonempty interior for $\gamma < \gamma^*$, meaning in this paper that

$$\overset{\circ}{\Gamma}(\gamma) = \{y \in R^n : \ e^T y = 1, \quad y > 0, \quad (B - \gamma A)y > 0\}$$

is nonempty for $\gamma < \gamma^*$. Note that we have replaced $y \neq 0$ with $e^T y = 1$ in the NEG problem, where e is the vector of all ones. This is without loss of generality since the system is homogeneous in y.

Obviously, the underlying decision problem related to the NEG problem can be solved in polynomial time: Given matrices A and B, and a number γ, does the linear system

$$\{e^T y = 1, \quad y \geq 0, \quad (B - \gamma A)y \geq 0\}$$

has a feasible point? Let $0 \le \gamma^* \le R$ for some positive R. Then, one can use the *bisection* method to generate a $\bar{\gamma}$ such that $\gamma^* - \epsilon \le \bar{\gamma} \le \gamma^*$ in $O(\log(R/\epsilon))$ bisection steps where each step solves a linear feasibility problem with data A, B and γ. Therefore, the NEG problem is polynomially solvable.

In this section, we develop an interior-point algorithm to directly solve the NEG problem. The algorithm is in the same spirit of the earlier center-section algorithm for linear programming, that is, it reduces the max-potential of the γ-level set $\Gamma(\gamma)$ by increasing γ. We show that the algorithm generates an ϵ approximate solution in $O((m+n)(\log(R/\epsilon)+\log(m+n))$ iterations where each iteration solves a system of $(m+n)$ linear equations.

8.1.1 Max-potential of $\Gamma(\gamma)$

We apply the analytic center and the max-potential theory to the inequality system $\Gamma(\gamma)$ of (8.1) for a fixed $\gamma < \gamma^*$. Recall that the max-potential of $\Gamma(\gamma)$, if it has a nonempty interior, is defined as

$$\mathcal{B}(\gamma) := \mathcal{B}(\Gamma(\gamma)) = \max_{y \in \Gamma(\gamma)} \left(\sum_{i=1}^{m} \log(By - \gamma Ay)_i + \sum_{j=1}^{n} \log y_j \right).$$

In the following, we frequently use the slack vector $s := By - \gamma Ay$.

Clearly, since B is a nonnegative matrix, we must have $\gamma^* \ge 0$. Without loss of generality, we further assume that $\Gamma(0)$ of (8.1) has a nonempty interior. This fact automatically holds if Assumption 8.2 holds. We also need the system $\Gamma(\gamma)$ to have a bounded and nonempty interior for all $\gamma < \gamma^*$, so that the analytic center and the max-potential are well defined for all $\gamma < \gamma^*$. As we discussed earlier, this is true under Assumptions 8.1 and 8.3. In what follows, we replace Assumption 8.3 by a weaker assumption:

Assumption 8.4 *There exists an optimal intensity vector $y^* \in \Gamma(\gamma^*)$ such that*

$$\bar{\pi} := (B + A)y^* > 0.$$

Then, we prove the following lemma.

Lemma 8.1 *Let $0 \le \gamma < \gamma^*$. Then, the system $\Gamma(\gamma)$ under Assumptions 8.1, 8.2 and 8.4 has a bounded and nonempty interior. Moreover, the max-potential*

$$\mathcal{B}(\gamma) \ge (m+n)\log(\frac{\gamma^* - \gamma}{\gamma^* + 1}) + \sum_{i=1}^{m} \log(\bar{\pi}_i/2) + n\log(\bar{\xi}/n),$$

where

$$\bar{\xi} = \max\left\{z : \ (\frac{1}{\gamma^*+1}\bar{\pi} + Ae/n)z \leq \frac{1}{2(\gamma^*+1)}\bar{\pi}\right\} > 0.$$

Proof. $\Gamma(\gamma)$ being bounded is obvious since $e^T y = 1$ and $y \geq 0$. Let y^* be the one in $\Gamma(\gamma^*)$ such that

$$e^T y^* = 1, \quad y^* \geq 0, \quad (B - \gamma^* A)y^* \geq 0, \tag{8.2}$$

and it satisfies

$$(B + A)y^* = \bar{\pi} > 0.$$

Let

$$\delta = \frac{\gamma(1+\gamma^*)}{1+\gamma}.$$

Then, for $0 \leq \gamma < \gamma^*$ we must have

$$\gamma < \delta < \gamma^*.$$

The left inequality follows from $1 + \gamma^* > 1 + \gamma$, and the right inequality follows from $\gamma < \gamma^*$ which implies $\gamma(1+\gamma^*) < \gamma^*(1+\gamma)$, which further implies

$$\gamma^* > \frac{\gamma(1+\gamma^*)}{1+\gamma} = \delta.$$

Thus, we have

$$(\gamma^* - \delta)(B + A)y^* = (\gamma^* - \delta)\bar{\pi} > 0. \tag{8.3}$$

Adding two inequalities (8.2) and (8.3), we have

$$((1 + \gamma^* - \delta)B - \delta A)y^* \geq (\gamma^* - \delta)\bar{\pi} > 0,$$

or

$$(B - \frac{\delta}{1+\gamma^*-\delta}A)y^* \geq \frac{\gamma^* - \delta}{1+\gamma^*-\delta}\bar{\pi} > 0,$$

which implies

$$(B - \gamma A)y^* \geq \frac{\gamma^* - \gamma}{1+\gamma^*}\bar{\pi} > 0,$$

since

$$\gamma = \frac{\delta}{1+\gamma^*-\delta}.$$

Therefore, there is an $0 < \omega < 1$ such that

$$\bar{y} = (1-\omega)y^* + \omega e/n > 0$$

and

$$(B - \gamma A)\bar{y} > 0.$$

That is, $\bar{y}$ is in the interior of $\Gamma(\gamma)$.

Specifically, let $\omega = \bar{\xi}(\gamma^* - \gamma)/\gamma^*$. Then, we have

$$\begin{aligned}
&(B - \gamma A)\bar{y} \\
&= (B - \gamma A)((1-\omega)y^* + \omega e/n) \\
&= (B - \gamma A)\left((1 - \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*})y^* + \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*}e/n\right) \\
&= (1 - \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*})(B - \gamma A)y^* + \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*}(B - \gamma A)e/n \\
&\geq (1 - \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*})(B - \gamma A)y^* - \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*}\gamma Ae/n \\
&\geq (1 - \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*})(B - \gamma A)y^* - \bar{\xi}(\gamma^* - \gamma)Ae/n \\
&\geq (1 - \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*})\frac{\gamma^* - \gamma}{1+\gamma^*}\bar{\pi} - \bar{\xi}(\gamma^* - \gamma)Ae/n \\
&\geq (1 - \bar{\xi})\frac{\gamma^* - \gamma}{1+\gamma^*}\bar{\pi} - \bar{\xi}(\gamma^* - \gamma)Ae/n \\
&= \frac{\gamma^* - \gamma}{2(1+\gamma^*)}\bar{\pi} + \frac{\gamma^* - \gamma}{2(1+\gamma^*)}\bar{\pi} - \bar{\xi}\frac{\gamma^* - \gamma}{1+\gamma^*}\bar{\pi} - \bar{\xi}(\gamma^* - \gamma)Ae/n \\
&= \frac{\gamma^* - \gamma}{2(1+\gamma^*)}\bar{\pi} + (\gamma^* - \gamma)\left(\frac{1}{2(1+\gamma^*)}\bar{\pi} - \bar{\xi}(\frac{1}{1+\gamma^*}\bar{\pi} + Ae/n)\right) \\
&\geq \frac{\gamma^* - \gamma}{2(1+\gamma^*)}\bar{\pi}.
\end{aligned}$$

Furthermore, we have

$$\bar{y} \geq \omega e/n = \frac{\bar{\xi}(\gamma^* - \gamma)}{\gamma^*}e/n > \frac{\bar{\xi}(\gamma^* - \gamma)}{1+\gamma^*}e/n.$$

Note that the max-potential

$$\mathcal{B}(\gamma) \geq \sum_{i=1}^{m} \log(B\bar{y} - \gamma A\bar{y})_i + \sum_{j=1}^{n} \log \bar{y}_j,$$

which together with the above two bounds give the desired result.

$\square$

Note that the above lemma may not hold in general.

Example 8.1 *Let*

$$B = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \quad \textit{and} \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then, for this problem $\gamma^* = 2$. However, for any $1 < \gamma < 2$, it must be true that $(y_1 = 1,\ y_2 = 0)$ is on the boundary of $\Gamma(0)$.

It can be shown that Assumption 8.3 implies Assumption 8.4. However, Assumption (8.4) is weaker than Assumption 8.3. Consider

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This system is reducible but $\Gamma(\gamma)$ has a nonempty interior for all $\gamma < \gamma^* = 1$, and it satisfies Assumption 8.4.

There is also an economic interpretation for Assumption 8.4: if $\Gamma(\gamma)$ has an empty interior for some $\gamma < \gamma^*$, then at every optimal intensity vector y^*

$$(B + A)y^* \not> 0,$$

which implies that for some good i

$$b_i y^* = a_i y^* = 0$$

at every optimal intensity vector y^*, where b_i and a_i are the ith row of the matrices B and A, respectively. Thus, the ith good is irrelevant, as it is neither produced nor consumed, and so can be removed from further consideration. Therefore, we can set up a reduced NEG problem (both row and column dimensions may be reduced) such as

$$\max\{\gamma \mid \exists\, y \neq 0 : \quad y \geq 0, \quad (B_2 - \gamma A_2)y \geq 0, \quad (b_i + a_i)y = 0\}$$

where B_2 and A_2 are the remaining matrices of B and A after deleting b_i and a_i, respectively. Now the reduced γ-level set will have a nonempty interior for any $\gamma < \gamma^*$.

We now prove two more lemmas to indicate that the max-potential of $\Gamma(\gamma)$ is an effective *measure* for the γ-level set of the NEG problem.

Lemma 8.2 *Let $\gamma^0 < \gamma^1 < \gamma^*$. Then,*

$$\mathcal{B}(\gamma^0) > \mathcal{B}(\gamma^1).$$

Proof. Let y^1 be the analytic center of $\Gamma(\gamma^1)$. Then, since

$$(B-\gamma^0 A)y^1 = (B-\gamma^1 A)y^1 + (\gamma^1-\gamma^0)Ay^1$$

and $(\gamma^1-\gamma^0)Ay^1 \geq 0$, we must have

$$(B-\gamma^0 A)y^1 \geq (B-\gamma^1 A)y^1 > 0.$$

This shows that y^1 is in the interior of $\Gamma(\gamma^0)$. Moreover, $(\gamma^1-\gamma^0)Ay^1 \neq 0$, since A has no all-zero columns and $y^1 > 0$. Thus,

$$\begin{aligned}
\mathcal{B}(\gamma^0) &\geq \sum_{i=1}^m \log(By^1-\gamma^0 Ay^1)_i + \sum_{j=1}^n \log y_j^1 \\
&= \sum_{i=1}^m \log(By^1-\gamma^1 Ay^1 + (\gamma^1-\gamma^0)Ay^1)_i + \sum_{j=1}^n \log y_j^1 \\
&> \sum_{i=1}^m \log(By^1-\gamma^1 Ay^1)_i + \sum_{j=1}^n \log y_j^1 \\
&= \mathcal{B}(\gamma^1).
\end{aligned}$$

□

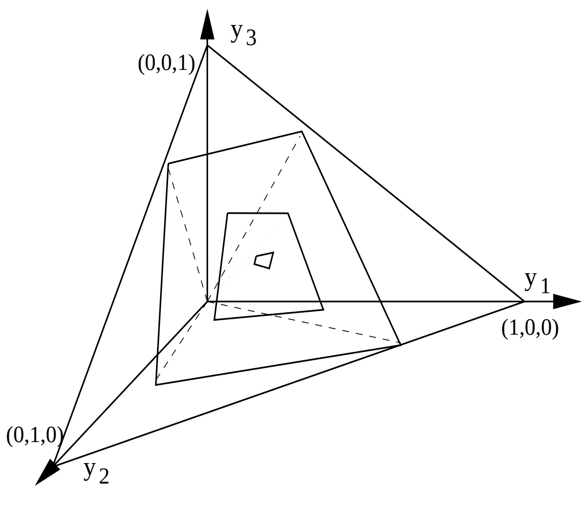

Figure 8.1: Illustration of the level set $\Gamma(\gamma)$ on the simplex polytope; the size of $\Gamma(\gamma)$ decreases as γ increases.

Lemma 8.3 *Let* $\min(\bar{\pi}) \geq 1/R$ *and* $\bar{\xi} \geq 1/R$ *for some positive* R*, and let* γ *satisfy* $(\gamma^* - \gamma)/(\gamma^* + 1) \geq \epsilon$*. Then, the max-potential*

$$\mathcal{B}(\gamma) \geq (m+n)\log(\epsilon/R) - m\log 2 - n\log n.$$

Proof. This is directly from Lemma 8.1.

□

Lemma 8.3 indicates that if we reduce

$$\mathcal{B}(\gamma) \leq -O((m+n)(\log(R/\epsilon) + \log n)),$$

then it must be true that $(\gamma^* - \gamma)/(1 + \gamma^*) < \epsilon$ and any $y \in \Gamma(\gamma)$ is an ϵ approximate solution. The algorithm actually returns an approximate (analytic) center of $\Gamma(\gamma)$.

8.1.2 Some technical results

Most of technical results presented in this section are related to those discussed earlier in Section 2.2. Let $0 \leq \gamma^0 < \gamma^*$. Then, the analytic center y^0 of $\Gamma(\gamma^0)$, or simply Γ^0, satisfies the following conditions.

$$\begin{pmatrix} X^0 s^0 \\ Z^0 y^0 \end{pmatrix} = \begin{pmatrix} e \\ e \end{pmatrix}, \tag{8.4}$$

where (s^0, y^0) is feasible with $\gamma = \gamma^0$ for the system

$$s - (B - \gamma A)y = 0, \quad e^T y = 1, \quad (s, y) \geq 0, \tag{8.5}$$

and (x^0, z^0) is feasible with $\gamma = \gamma^0$ for the system

$$(B - \gamma A)^T x + z = (m+n)e, \quad (x, z) \geq 0. \tag{8.6}$$

Now let y^1 be the analytic center of $\Gamma^1 = \Gamma(\gamma^1)$ with $\gamma^1 = \gamma^0 + \Delta\gamma$. For a suitable choice of $\Delta\gamma$, we can show that $\gamma^0 < \gamma^1 < \gamma^*$ and

$$P(\gamma^1) \leq P(\gamma^0) - \Omega(1).$$

We prove a lemma on how to select $\Delta\gamma$.

Lemma 8.4 *Let*

$$\Delta\gamma = \beta \min\{\frac{1}{\|X^0 A Y^0 e\|}, \frac{1}{\|Y^0 A^T X^0 e\|}\},$$

for some constant $0 < \beta < 1$ *and let* $\gamma^1 = \gamma^0 + \Delta\gamma$. *Furthermore, let*

$$\bar{s} = s^0 - \Delta\gamma A y^0 \quad \text{and} \quad \bar{y} = y^0,$$

and

$$\bar{x} = x^0 \quad \text{and} \quad \bar{z} = z^0 + \Delta\gamma A^T x^0.$$

Then, $(\bar{s}, \bar{y})$ *and* $(\bar{x}, \bar{z})$ *are feasible for systems (8.5) and (8.6) with* $\gamma = \gamma^1$, *respectively. Moreover,*

$$\left\| \begin{pmatrix} \bar{X}\bar{s} - e \\ \bar{Z}\bar{y} - e \end{pmatrix} \right\| \le \sqrt{2}\beta,$$

and

$$(\bar{s}, \bar{y}) > 0 \quad \text{and} \quad (\bar{x}, \bar{z}) > 0.$$

Proof. The two equations in systems (8.5) and (8.6) for $\gamma = \gamma^1$ can easily be verified. The inequality for the norm can be proved from relation (8.4) and the choice of $\Delta\gamma$.

$$\begin{aligned} \|\bar{X}\bar{s} - e\|^2 &= \|X^0(s^0 - \Delta\gamma A y^0) - e\|^2 \\ &= \|\Delta\gamma X^0 A Y^0 e\|^2 \\ &\le \beta^2. \end{aligned}$$

Similarly,

$$\begin{aligned} \|\bar{Z}\bar{y} - e\|^2 &= \|Y^0(z^0 + \Delta\gamma A^T x^0) - e\|^2 \\ &= \|\Delta\gamma Y^0 A^T X^0 e\|^2 \\ &\le \beta^2. \end{aligned}$$

These relations also imply that

$$\bar{s} > 0 \quad \text{and} \quad \bar{z} > 0,$$

since

$$\bar{x} = x^0 > 0 \quad \text{and} \quad \bar{y} = y^0 > 0.$$

$\square$

The above lemma establishes a fact that $(\bar{s}, \bar{y})$ and $(\bar{x}, \bar{z})$ are approximate centers for systems (8.5) and (8.6) with $\gamma = \gamma^1$, respectively. Thus, $(\bar{x}, \bar{z})$ or $(\bar{s}, \bar{y})$ can be used as an initial pair of Newton's method to generate the new center pair (x^1, z^1) and (s^1, y^1).

We now state another technical result.

Proposition 8.5 *Let* $H = \{h_{ij}\}$ *be a nonnegative* $m \times n$*-matrix. Then*

$$\|He\| \le e^T He,$$

and

$$\|H^T e\| \le e^T H^T e = e^T He.$$

8.1.3 Center-section algorithm

We now present a conceptual algorithm, which uses the perfect center, to illustrate the basic idea of our approach:

1. Let $\gamma^0 = 0$. Then generate the analytic center y^0 of $\Gamma^0 = \Gamma(\gamma^0)$. Set $k := 0$.

2. Let
$$\Delta\gamma^k = \beta \min\{\frac{1}{\|X^k A Y^k e\|}, \frac{1}{\|Y^k A^T X^k e\|}\}$$
for some constant $0 < \beta < 1$ and let $\gamma^{k+1} = \gamma^k + \Delta\gamma^k$. Then use Newton's method, which is described below, to generate the analytic center y^{k+1} of $\Gamma^{k+1} = \Gamma(\gamma^{k+1})$.

3. If $\mathcal{B}(\gamma^{k+1}) > -O((m+n)(\log(R/\epsilon) + \log n)$, then set $k := k+1$ and return to Step 2.

We can use the primal-dual Newton procedure to accomplish Step 1 of the algorithm: Let $(\bar{x}, \bar{z}, \bar{s}, \bar{y})$ be defined in Lemma 8.4 and repeatedly solve for (d_s, d_y) and (d_x, d_z):
$$\begin{array}{rcl} \bar{S}d_x + \bar{X}d_s & = & e - \bar{x}\bar{s}, \\ \bar{Y}d_z + \bar{Z}d_y & = & e - \bar{y}\bar{z}, \\ (B - \gamma^1 A)^T d_x + d_z & = & 0, \\ d_s - (B - \gamma^1 A)d_y & = & 0. \end{array} \tag{8.7}$$
Then, let
$$\bar{x} := \bar{x} + d_x \quad \text{and} \quad \bar{z} := \bar{z} + d_z,$$
and
$$\bar{s} := \bar{s} + d_s \quad \text{and} \quad \bar{y} := \bar{y} + d_y.$$

We now analyze the algorithm using approximate center pairs (s^0, y^0) and (x^0, z^0) that are feasible for systems (8.5) and (8.6) with $\gamma = \gamma^0$, respectively, and
$$\left\| \begin{pmatrix} X^0 s^0 - e \\ Z^0 y^0 - e \end{pmatrix} \right\| \le \delta. \tag{8.8}$$
As we proved in Theorem 3.2(iii),
$$\mathcal{B}(\gamma^0) \ge \mathcal{B}(s^0, y^0) := \sum_{i=1}^m \log s_i^0 + \sum_{j=1}^n \log y_j^0 \ge \mathcal{B}(\gamma^0) - \frac{\delta^2}{2(1-\delta)}.$$
Thus, $\mathcal{B}(s, y)$ is close to $\mathcal{B}(\gamma)$ when (s, y) is an approximate center of $\Gamma(\gamma)$, and it can be used to terminate the algorithm.

The following lemma is an analogue to Lemma 8.4.

Lemma 8.6 *Let positive constants δ and β satisfy $\delta + \sqrt{2}\beta < 1$. Let (x^0, z^0) and (s^0, y^0) be an approximate center pair for $\Gamma(\gamma^0)$. Let $\Delta\gamma$, γ^1, $(\bar{s}, \bar{y})$ and $(\bar{x}, \bar{z})$ be selected as in Lemma 8.4. Then, $(\bar{s}, \bar{y})$ and $(\bar{x}, \bar{z})$ are feasible for systems (8.5) and (8.6) with $\gamma = \gamma^1$, respectively. Moreover,*

$$\left\| \begin{pmatrix} \bar{X}\bar{s} - e \\ \bar{Z}\bar{y} - e \end{pmatrix} \right\| \le \delta + \sqrt{2}\beta < 1,$$

and

$$(\bar{s}, \bar{y}) > 0 \quad \textit{and} \quad (\bar{x}, \bar{z}) > 0.$$

Now, using $(\bar{x}, \bar{z})$ and $(\bar{s}, \bar{y})$ as the initial pair, we apply the Newton procedure described by (8.7) to generate a new approximate center pair (s^1, y^1) and (x^1, z^1) for $\Gamma(\gamma^1)$. Note that we terminate the procedure in *one* step, then assign

$$(s^1, y^1) = (\bar{s} + d_s, \bar{y} + d_y) \quad \text{and} \quad (x^1, z^1) = (\bar{x} + d_x, \bar{z} + d_z).$$

Note that (iv) of Theorem 3.2 and (8.6) indicate that

$$\left\| \begin{pmatrix} X^1 s^1 - e \\ Z^1 y^1 - e \end{pmatrix} \right\| \le \frac{\sqrt{2}}{4} \frac{(\delta + \sqrt{2}\beta)^2}{1 - \delta - \sqrt{2}\beta}, \tag{8.9}$$

and

$$\|(Y^0)^{-1}(Y^1 - Y^0)\| \le \frac{\delta + \sqrt{2}\beta}{1 - \delta - \sqrt{2}\beta}. \tag{8.10}$$

Next, for suitable constants δ and β, for example,

$$\delta = 1/12 \quad \text{and} \quad \beta = 1/4\sqrt{2}, \tag{8.11}$$

we prove that the potential value at (s^1, y^1) in $\Gamma^1 = \Gamma(\gamma^1)$ is reduced by a constant from the value at (s^0, y^0) in $\Gamma^0 = \Gamma(\gamma^0)$.

Theorem 8.7 *Let δ and β be chosen as in (8.11), and (x^0, y^0, s^0, z^0) satisfy (8.8). Let*

$$\Delta\gamma = \beta \min\left\{ \frac{1}{\|X^0 A Y^0 e\|}, \frac{1}{\|Y^0 A^T X^0 e\|} \right\}$$

and $\gamma^1 = \gamma^0 + \Delta\gamma$. Let (s^1, y^1) and (x^1, y^1) be generated in one step of the Newton procedure. Then,

$$\gamma^0 < \gamma^1 < \gamma^*,$$

$$\left\| \begin{pmatrix} X^1 s^1 - e \\ Z^1 y^1 - e \end{pmatrix} \right\| \le \delta,$$

and

$$\mathcal{B}(s^1, y^1) \le \mathcal{B}(s^0, y^0) - \Omega(1).$$

Proof. $\gamma^1 > \gamma^0$ because $\Delta\gamma > 0$ and $\gamma^1 < \gamma^*$ because $\Gamma^1 = \Gamma(\gamma^1)$ has a nonempty interior. From inequality (8.9) we have

$$\left\| \begin{pmatrix} X^1 s^1 - e \\ Z^1 y^1 - e \end{pmatrix} \right\| \le \frac{\sqrt{2}}{4}\frac{1}{6} < \delta.$$

Now we prove the potential reduction inequality. We have

$$\begin{aligned}
(x^0)^T s^1 + (z^0)^T y^1 &= (x^0)^T (B - \gamma^1 A) y^1 + (z^0)^T y^1 \\
&= (y^1)^T ((B - \gamma^0 A)^T x^0 + z^0) - \Delta\gamma (x^0)^T A y^1 \\
&= (m+n) e^T y^1 - \Delta\gamma (x^0)^T A y^1 \\
&= (m+n) - \Delta\gamma (x^0)^T A y^1.
\end{aligned}$$

Thus,

$$\begin{aligned}
\prod_{i=1}^m (x_i^0 s_i^1) \prod_{i=1}^n (z_i^0 y_i^1) &\le \left(\frac{(x^0)^T s^1 + (z^0)^T y^1}{m+n} \right)^{m+n} \\
&= \left(1 - \frac{\Delta\gamma (x^0)^T A y^1}{m+n} \right)^{m+n} \\
&\le \exp(-\Delta\gamma (x^0)^T A y^1).
\end{aligned}$$

Moreover, using $(x^0)^T A y^1 > 0$, Lemma 8.5 and relation (8.10), we have

$$\begin{aligned}
\Delta\gamma (x^0)^T A y^1 &= \Delta\gamma e^T X^0 A Y^1 e \\
&= \beta \min\{\frac{1}{\|X^0 A y^0\|}, \frac{1}{\|Y^0 A^T x^0\|}\} e^T X^0 A Y^1 e \\
&\ge \beta \frac{e^T X^0 A Y^1 e}{e^T X^0 A Y^0 e} \\
&\ge \beta (1 - \frac{\delta + \sqrt{2}\beta}{1 - \delta - \sqrt{2}\beta}) \\
&= \beta \frac{1 - 2(\delta + \sqrt{2}\beta)}{1 - \delta - \sqrt{2}\beta}.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
&\mathcal{B}(s^1, y^1) - \mathcal{B}(s^0, y^0) \\
&= \sum_{i=1}^m \log \frac{s_i^1}{s_i^0} + \sum_{i=1}^n \log \frac{y_i^1}{y_i^0} \\
&= \sum_{i=1}^m \log(x_i^0 s_i^1) + \sum_{i=1}^n \log(z_i^0 y_i^1) - \sum_{i=1}^m \log(x_i^0 s_i^0) - \sum_{i=1}^n \log(z_i^0 y_i^0)
\end{aligned}$$

$$
\begin{aligned}
&\leq -\beta\frac{1-2(\delta+\sqrt{2}\beta)}{1-\delta-\sqrt{2}\beta} - \sum_{i=1}^{m}\log(x_i^0 s_i^0) - \sum_{i=1}^{n}\log(z_i^0 y_i^0) \\
&\leq -\beta\frac{1-2(\delta+\sqrt{2}\beta)}{1-\delta-\sqrt{2}\beta} + \frac{\delta^2}{2(1-\delta)} \quad \text{(from Lemma 3.1 and (8.8))}.
\end{aligned}
$$

One can verify

$$\beta\frac{1-2(\delta+\sqrt{2}\beta)}{1-\delta-\sqrt{2}\beta} - \frac{\delta^2}{2(1-\delta)} = \frac{1}{8\sqrt{2}} - \frac{1}{264} > 0.$$

This gives the desired result.

□

We now formally state the algorithm.

Algorithm 8.1 *Let* $\gamma^0 = 0$*. Then generate a* δ*-approximate center* y^0 *of* $\Gamma^0 = \Gamma(\gamma^0)$*. Set* $k := 0$*.*
While $\mathcal{B}(s^{k+1}, y^{k+1}) > -O((m+n)(\log(R/\epsilon) + \log n))$ **do**

1. *Let*
$$\Delta\gamma^k = \beta\min\left\{\frac{1}{\|X^kAY^ke\|}, \frac{1}{\|Y^kA^TX^ke\|}\right\}$$
for some constant $0 < \beta < 1$ *and let* $\gamma^{k+1} = \gamma^k + \Delta\gamma^k$*. Let*
$$\bar{s} = s^k - \Delta\gamma A y^k \quad \text{and} \quad \bar{y} = y^k,$$
and
$$\bar{x} = x^k \quad \text{and} \quad \bar{z} = z^k + \Delta\gamma A^T x^k.$$

2. *Solve for* (d_s, d_y) *and* (d_x, d_z) *from (8.7) and let*
$$x^{k+1} = \bar{x} + d_x \quad \text{and} \quad z^{k+1} = \bar{z} + d_z,$$
and
$$s^{k+1} = \bar{s} + d_s \quad \text{and} \quad y^{k+1} = \bar{y} + d_y.$$

3. *Set* $k := k+1$ *and return to Step 1.*

It is well known that an initial δ-approximate center pair, (z^0, y^0) and (x^0, z^0), can be generated in no more than $O((m+n)(\log(R/\epsilon) + \log n))$ interior-point algorithm iterations. Thus, we conclude the following.

Theorem 8.8 *Algorithm 8.1, with a suitable choice of δ and β, terminates in $k = O((m+n)(\log(R/\epsilon) + \log n))$ iterations and each iteration solves a system of $(m+n)$ linear equations. The resulting $y^k \in \Gamma(\gamma^k)$ and γ^k satisfy*

$$0 < \frac{\gamma^* - \gamma}{1 + \gamma^*} < \epsilon.$$

The algorithm also generates the optimal dual vector. More precisely, we prove the following result.

Proposition 8.9 *Any limit point of*

$$\lim_{k \to \infty} \frac{x^k}{e^T x^k},$$

where x^k is generated by Algorithm 8.1, is a solution for the dual NEG problem with $\eta = \eta^*$ under Assumptions 8.1, 8.2 and 8.4.

Proof. For simplicity, we assume that (x^k, y^k, s^k, z^k) is exactly centered, i.e., it satisfies relations (8.4), (8.5) and (8.6). Since at least one component of s^k converges to zero, $e^T x^k = e^T (S^k)^{-1} e$ tends to $+\infty$. Moreover, from (8.6) we have

$$(B - \gamma^k A)^T \frac{x^k}{e^T x^k} + \frac{z^k}{e^T x^k} = (m+n) \frac{e}{e^T x^k}.$$

Thus, the right-hand vector of the above equation converges to zero. Since $z^k / e^T x^k > 0$ for all k,

$$\lim_{k \to \infty} (B - \gamma^k A)^T \frac{x^k}{e^T x^k} = - \lim_{k \to \infty} \frac{z^k}{e^T x^k} \le 0.$$

Furthermore, under Assumptions 8.1, 8.2 and 8.4 we have $\gamma^k \to \gamma^* = \eta^*$. Therefore, any limit point of the sequence of positive $x^k / e^T x^k$ is a solution for the dual.

$\Box$

Finally we turn our attention to the question raised earlier, that is, what happens if $\Gamma(\gamma)$ has an empty interior for some $\gamma < \gamma^*$. It turns out that there exists a nice duality theorem for the NEG problem, that is, under Assumptions 8.1 and 8.2, $\Gamma(\gamma)$ has a nonempty interior for all $\gamma < \eta^* \le \gamma^*$ (see Kemeny et al. and Gale [175]). Thus, the algorithm discussed in this paper will precisely generate η^* under only Assumptions 8.1 and 8.2. Similarly, the η-level set of the dual has an nonempty interior for all $\eta > \gamma^* \ge \eta^*$. Thus, one can apply the algorithm for solving the dual to generate γ^* in the same manner. Thus, we can solve the NEG problem under only Assumptions 8.1 and 8.2, which are basic assumptions for a meaningful economic growth model.

8.2 Convex Optimization

The problem studied is that of finding an interior point in a convex set Γ, where we assume $\Gamma \subset \mathcal{R}^m$ has a nonempty interior and is contained in the cube $\Omega^0 = \{y \in \mathcal{R}^m : 0 \le y \le e\} = [0,1]^m$. The set Γ is defined implicitly by a **separating oracle** which for every $\bar{y} \in \Omega^0$ either answers that $\bar{y}$ belongs to $\overset{\circ}{\Gamma}$ or generates a separating hyperplane $\{y \in \mathcal{R}^m : a^T y \le a^T \bar{y}\} \supset \Gamma$. Without loss of generality, we assume that a is normalized so that $\|a\| = 1$.

The problem of finding an interior point in a convex set defined by a system (finite or infinite) of convex inequalities

$$\Gamma = \{y \in \mathcal{R}^m : f_i(y) \le 0, \quad i = 1, 2, ...\},$$

where each $f_i : \mathcal{R}^m \mapsto \mathcal{R}$ is convex, can be cast in this manner. In particular, the separating oracle just needs to select a to be $g_i/\|g_i\|$, where g_i is an arbitrary subgradient of any of the functions f_i satisfying $f_i(\bar{y}) \ge 0$, i.e. $g_i \in \partial f_i(\bar{y})$ (the subdifferential of f_i). Note that for any $g_i \in \partial f_i(\bar{y})$, $f_i(y) \le f_i(\bar{y})$ implies $g_i^T(y - \bar{y}) \le 0$. Thus, if $f_i(\bar{y}) \ge 0$, $f_i(y) \le 0$ then $g_i^T(y - \bar{y}) \le 0$. (In fact, the requirement that $f_i(\bar{y})$ be computed exactly and $g_i \in \partial f_i(\bar{y})$ can be significantly relaxed.) All that is required from the oracle is that the following query be answered: Find any i such that $f_i(\bar{y}) \ge 0$, and a such that $a^T(y - \bar{y}) \le 0$ for all $y \in \{y \in \mathcal{R}^m : f_i(y) \le 0\}$.

The problem under investigation may also be cast as that of finding the solution to an infinite system of linear inequalities, which is defined implicitly by the oracle $\Gamma = \{y \in \mathcal{R}^m : \; G^T y \le g\}$ for some $G \in \mathcal{R}^{m \times d}$ and $g \in \mathcal{R}^d$ and d is infinite. The classical centering methods that have been suggested for the above convex feasibility problem include the center of gravity method, the max-volume sphere method, the ellipsoid method, the max-volume ellipsoid method, and the volumetric center method.

The column generation or cutting plane method computes $\bar{y}$ as the analytic center of the system of inequalities generated so far. In this section, we show that for any given convex feasibility problem with a nonempty interior, the algorithm is a fully polynomial-time approximation scheme that uses only linear inequalities to approximate the solution set. A fully polynomial-time approximation scheme means that for every ϵ, the accuracy at termination, the running time is a polynomial in the dimension m and ϵ.

8.2.1 Max-potential reduction

Now, we use an approximate center y^k to generate a cut, where (y^k, s^k) is an interior point in $\Omega = \{y \in \mathcal{R}^m : A^T y \le c\}$ and there is $x^k > 0$ with

$Ax^k = 0$ such that

$$\|X^k s^k - e\| \leq \eta \tag{8.12}$$

for some $0 < \eta < 1$. Let us place a cut exactly at y^k, that is, we add a new inequality $a^T y \leq a^T y^k$ to Ω, and consider the new set

$$\Omega^+ = \{y : A^T y \leq c, \quad a^T y \leq a^T y^k\}.$$

We now prove a lemma resembling Theorem 2.10.

Lemma 8.10 *Denote by $(\bar{y}, \bar{s})$ the analytic center of Ω and let*

$$\bar{r} = \sqrt{a^T (A\bar{S}^{-2}A^T)^{-1} a}.$$

Then the max-potential of Ω^+

$$\mathcal{B}(\Omega^+) \leq \mathcal{B}(\Omega) + \log(\bar{r}) - \alpha$$

for some constant α depending only on η. Moreover, if $0 < \eta < 1/100$, then we have $\alpha > 0$.

Proof. Denote by $\bar{y}^+$ the analytic center for Ω^+. Let $\bar{s}^+ = c - A^T\bar{y}^+ > 0$ and $\bar{s}^+_{n+1} = a^T y^k - a^T \bar{y}^+$. Then we have

$$\begin{aligned}
\bar{s}^+_{n+1} &= a^T(y^k - \bar{y}^+) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}(A\bar{S}^{-2}A^T)(y^k - \bar{y}^+) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-2}(A^T y^k - A^T\bar{y}^+) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-2}(-c + A^T y^k + c - A^T\bar{y}^+) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-2}(\bar{s}^+ - s^k) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-1}(\bar{S}^{-1}\bar{s}^+ - \bar{S}^{-1}s^k) \\
&\leq \|a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-1}\|\|\bar{S}^{-1}\bar{s}^+ - \bar{S}^{-1}s^k\| \\
&= \bar{r}\|\bar{S}^{-1}\bar{s}^+ - e + e - \bar{S}^{-1}s^k\| \\
&\leq \bar{r}(\|\bar{S}^{-1}\bar{s}^+ - e\| + \|e - \bar{S}^{-1}s^k\|) \\
&\leq \bar{r}(\|\bar{S}^{-1}\bar{s}^+ - e\| + \frac{\eta}{(1-\eta)^2}) \quad \text{(from Theorem 3.2).}
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\exp \mathcal{B}(\Omega^+)}{\bar{r}\exp \mathcal{B}(\Omega)} &= \frac{\bar{s}^+_{n+1}}{\bar{r}} \prod_{j=1}^{n} \frac{\bar{s}^+_j}{\bar{s}_j} \\
&\leq (\|\bar{S}^{-1}\bar{s}^+ - e\| + \frac{\eta}{(1-\eta)^2}) \prod_{j=1}^{n} \frac{\bar{s}^+_j}{\bar{s}_j}.
\end{aligned}$$

Note that we still have

$$e^T \bar{S}^{-1} \bar{s}^+ = n.$$

This can be seen as follows:

$$e^T \bar{S}^{-1} \bar{s}^+ = e^T \bar{X} \bar{s}^+ = e^T \bar{X}(c - A^T \bar{y}^+) = e^T \bar{X} c = e^T \bar{X} \bar{s} = n,$$

where the first and fifth equalities are due to $\bar{X}\bar{s} = e$, and the second and fourth equalities follow from $A\bar{x} = 0$.

Thus, from the proof of Theorem 2.10

$$(\|\bar{S}^{-1}\bar{s}^+ - e\| + \frac{\eta}{(1-\eta)^2}) \prod_{j=1}^{n} \frac{\bar{s}_j^+}{\bar{s}_j} \le 4\exp(\frac{8\eta - 3}{2(1-\eta)^2}).$$

Thus,

$$\mathcal{B}(\Omega^+) - \mathcal{B}(\Omega) \le \log(\bar{r}) + \log(4) + \frac{8\eta - 3}{2(1-\eta)^2}.$$

Let

$$\alpha = -\log(4) - \frac{8\eta - 3}{2(1-\eta)^2}.$$

Let $\eta = 1/100$. Then we have $\alpha > 0$ and the desired result.

□

8.2.2 Compute a new approximate center

In this section, we show how to construct a pair (x, y, s) from (x^k, y^k, s^k) such that $(A, a)x = 0$ with $x > 0$, (y, s) is in the interior of Ω^+, and

$$\|Xs - e\| < \eta < 1.$$

Suppose a pair (x^k, y^k, s^k) is given which satisfies (8.12), we use the **dual** scaling for the construction of (x, y, s). Let

$$\begin{aligned}
r^k &= \sqrt{a^T(A(S^k)^{-2}A^T)^{-1}a}, \\
\Delta y &= -(\beta/r^k)(A(S^k)^{-2}A^T)^{-1}a, \\
\Delta s &= (\beta/r^k)A^T(A(S^k)^{-2}A^T)^{-1}a, \\
\Delta x &= -(\beta/r^k)(S^k)^{-2}A^T(A(S^k)^{-2}A^T)^{-1}a.
\end{aligned}$$

Then we set

$$y = y^k + \Delta y$$

and

$$x = \begin{bmatrix} x^k + \Delta x \\ \beta/r^k \end{bmatrix} = \begin{bmatrix} x^k - (\beta/r^k)(S^k)^{-2}A^T(A(S^k)^{-2}A^T)^{-1}a \\ \beta/r^k \end{bmatrix}.$$

It can be readily verified that

$$s = \begin{bmatrix} c - A^T(y^k + \Delta y) \\ a^T y^k - a^T(y^k + \Delta y) \end{bmatrix} = \begin{bmatrix} s^k + \Delta s \\ \beta r^k \end{bmatrix}$$

$$= \begin{bmatrix} s^k + (\beta/r^k)A^T(A(S^k)^{-2}A^T)^{-1}a \\ \beta r^k \end{bmatrix}.$$

First, we have

$$(A, a)x = Ax^k - (\beta/r^k)a + (\beta/r^k)a = 0.$$

Second, we have

$$s^k + (\beta/r^k)A^T(A(S^k)^{-2}A^T)^{-1}a = (S^k)(e + p^k)$$

and

$$x^k - (\beta/r^k)(S^k)^{-2}A^T(A(S^k)^{-2}A^T)^{-1}a = (S^k)^{-1}(X^kS^ke - p^k),$$

where

$$p^k = (\beta/r^k)(S^k)^{-1}A^T(A(S^k)^{-2}A^T)^{-1}a.$$

Note that we have

$$\|p^k\| = \beta. \tag{8.13}$$

On the other hand, we have

$$x_j^k s_j^k \geq 1 - \eta.$$

Thus, if we select η and β such that

$$1 - \eta - \beta > 0, \tag{8.14}$$

then both

$$s = \begin{bmatrix} (S^k)(e + p^k) \\ \beta r^k \end{bmatrix} > 0$$

and

$$x = \begin{bmatrix} (S^k)^{-1}(X^kS^ke - p^k) \\ \beta/r^k \end{bmatrix} > 0.$$

A simple calculation yields

$$Xs - e = \begin{bmatrix} X^k s^k - e \\ 0 \end{bmatrix} - \begin{bmatrix} (p^k)^2 \\ 1-\beta^2 \end{bmatrix} + \begin{bmatrix} (X^k S^k - I)p^k \\ 0 \end{bmatrix}$$

where the vector

$$(p^k)^2 = ((p_1^k)^2, (p_2^k)^2, ...(p_n^k)^2)^T.$$

Therefore, we have

$$\begin{aligned} \|Xs - e\| &\le \|X^k s^k - e\| + \sqrt{\|(p^k)^2\|^2 + (1-\beta^2)^2} + \|X^k s^k - e\|\|p^k\| \\ &\le \eta + \sqrt{\beta^4 + (1-\beta^2)^2} + \eta\beta \end{aligned}$$

where the last step follows from (8.12) and (8.13). Let $\beta = 1/\sqrt{2}$ and $\eta = 0.15$, then

$$\|Xs - e\| \le \gamma = 0.15 + 1/\sqrt{2} + 0.15/\sqrt{2} < 1.$$

Or, we can let

$$\eta = 1/100, \quad \text{and have} \quad \gamma := 1/100 + 1.01/\sqrt{2} < 1.$$

Furthermore, it can also be easily verified that (8.14) holds.

Hence, using this (y, s) as a starting pair, we can apply the **dual** Newton procedure of Chapter 3 to generate a pair (y^{k+1}, s^{k+1}) and $x^{k+1} = x(y^{k+1})$ such that

$$\begin{aligned} (A, a)x^{k+1} &= 0, \quad x^{k+1} > 0, \\ s^{k+1} &= (c^T, a^T y^k)^T - (A, a)^T y^{k+1} > 0, \end{aligned}$$

and

$$\|X^{k+1} s^{k+1} - e\| \le \eta.$$

By Theorem 3.2(iv) and the above given values of γ and η, this can be accomplished in 4 **dual** Newton steps due to the fact $\gamma^{16} \le \eta$. This column generation process can be repeated, and from Lemma 8.10 the nested sequence of polyhedral sets Ω^k generated by the algorithm satisfies

$$\mathcal{B}(\Omega^{k+1}) \le \mathcal{B}(\Omega^k) + \log(\bar{r}^k) - \alpha \tag{8.15}$$

where α is some constant,

$$\bar{r}^k = \sqrt{a_{k+1}^T (A(\bar{S}^k)^{-2} A^T)^{-1} a_{k+1}},$$

$(\bar{y}^k, \bar{s}^k)$ is the analytic center of Ω^k, and a_{k+1} is the cut generated at the kth iteration. Note that $(\bar{y}^k, \bar{s}^k)$ is solely used for analysis, and the algorithm does not need any knowledge of $(\bar{y}^k, \bar{s}^k)$.

8.2.3 Convergence and complexity

Let the solution set Γ be contained in $\Omega^0 = \{y \in \mathcal{R}^m : 0 \le y \le e\}$, and $\overset{\circ}{\Gamma}$ contain a full dimensional closed ball with $\epsilon < \frac{1}{2}$ radius. We also assume that there exists an oracle which for every $\bar{y} \in \Omega^0$ either returns that $\bar{y} \in \overset{\circ}{\Gamma}$ or generates a separating hyperplane $\{y : a^T y \le a^T \bar{y}\} \supset \Gamma$, with $\|a\| = 1$ being assumed.

The column-generation, or cutting plane, algorithm from approximate analytic centers is as follows:

Algorithm 8.2 *Let*

$$A^0 = (I, -I) \in \mathcal{R}^{m \times 2m}, \quad c^0 = \begin{pmatrix} e \\ 0 \end{pmatrix} \in \mathcal{R}^{2m}, \tag{8.16}$$

$$y^0 = \frac{1}{2}e \in \mathcal{R}^m, \; s^0 = c^0 - (A^0)^T y^0 = \frac{1}{2}e \in \mathcal{R}^{2m}, \; x^0 = 2e \in \mathcal{R}^{2m}.$$

Set

$$k := 0.$$

While $y^k \notin \overset{\circ}{\Gamma}$ **do**

1. *Query the oracle to generate a hyperplane* $\{y : a_{k+1}^T y \le a_{k+1}^T y^k\} \supset \Gamma$ *with* $\|a_{k+1}\| = 1$, *and let*
$$\Omega^{k+1} = \{y \in \mathcal{R}^m : c^{k+1} - (A^{k+1})^T y \ge 0\}$$
where
$$A^{k+1} = (A^k, a_{k+1}) \quad \textit{and} \quad c^{k+1} = \begin{pmatrix} c^k \\ a_{k+1}^T y^k \end{pmatrix}$$

2. *Compute* $(y^{k+1}, s^{k+1}, x^{k+1})$ *such that* y^{k+1} *is an* η*-approximate analytic center of* Ω^{k+1}, *using the Newton method with the updating scheme of section 8.2.2 and starting from* (y^k, s^k, x^k), *an* η*-approximate of* Ω^k.

3. *Set* $k := k + 1$ *and return to Step 1.*

Let the potential function computed at the **exact** analytic center $\bar{y}^k$ be:

$$\mathcal{B}(\Omega^k) = \sum_{j=1}^{2m+k} \log(c^k - (A^k)^T \bar{y}^k)_j.$$

Clearly the following relations hold, provided that termination has not occurred:

$$\Gamma \subset \Omega^k \quad \forall k, \tag{8.17}$$

and

$$\mathcal{B}(\Omega^{k+1}) \le \mathcal{B}(\Omega^k) + \frac{1}{2}\log(\bar{r}^k)^2 - \alpha \quad \text{(by Lemma 8.10)} \tag{8.18}$$

where

$$(\bar{r}^k)^2 = a_{k+1}^T (A^k (\bar{S}^k)^{-2} (A^k)^T)^{-1} a_{k+1} \quad \text{and} \quad \bar{s}^k = c^k - (A^k)^T \bar{y}^k.$$

Lemma 8.11 *For all $k \ge 0$,*

$$\mathcal{B}(\Omega^k) \ge (2m+k)\log\epsilon.$$

Proof. From (8.17), $\Gamma \subset \Omega^k$. Thus Ω^k contains a full dimensional ball with ϵ radius. Let the center of this ball be $\bar{y}$. Then $c^k - (A^k)^T\bar{y} \ge \epsilon e$; thus

$$\mathcal{B}(\Omega^k) = \sum_{j=1}^{2m+k} \log(c^k - (A^k)^T\bar{y}^k)_j \ge \sum_{j=1}^{2m+k} \log(c^k - (A^k)^T\bar{y})_j \ge \sum_{j=1}^{2m+k} \log\epsilon$$

where $\bar{y}^k$ denotes the analytic center of Ω^k.

$\square$

Lemma 8.12 *Let $s = c^k - (A^k)^T y$ for any $y \in \Omega^k$. Then*

i) $0 \le s_j \le 1, \quad j = 1, \ldots, 2m$

ii) $0 \le s_j \le \sqrt{m}, \quad j = 2m+1, \ldots, 2m+k.$

Proof. For $j = 1, \ldots, m$, $s_j = 1 - y_j$; since $0 \le y_j \le 1$, $0 \le s_j \le 1$.
For $j = m+1, \ldots, 2m$, $s_j = y_{j-m}$; since $0 \le y_{j-m} \le 1$, $0 \le s_j \le 1$.
For $j = 2m+1, \ldots, 2m+k$,

$$s_j = a_{j-2m}^T y^{j-2m} - a_{j-2m}^T y \le \|a_{j-2m}\|\ \|y^{j-2m} - y\| = \|y^{j-2m} - y\| \le \sqrt{m}$$

The last inequality is due to the fact that $0 \le y^{j-2m} \le e$ and $0 \le y \le e$ or $y^{j-2m} \in \Omega^0$ and $y \in \Omega^0$.

$\square$

Lemma 8.11 indicates that in order to prove finite convergence, one needs to show that $\mathcal{B}(\Omega^k)$ grows more slowly than $2m+k$. By Lemma 8.10, this means finding upper bounds on $\bar{r}^k$. In the following Lemma this is achieved by using a construction which bounds $A^k(\bar{S}^k)^{-2}(A^k)^T$ from below by using a certain matrix B^k which is simple enough to handle.

Lemma 8.13 *Let $s = c^k - (A^k)^T y$ for any $y \in \Omega^k$ and $B^0 = 8I$, $B^{k+1} = B^k + \frac{1}{m} a_{k+1} a_{k+1}^T$. Then*

$$A^k S^{-2} (A^k)^T \succeq B^k;$$

that is,

$$A^k S^{-2} (A^k)^T - B^k$$

is positive semi-definite.

Proof. Let $Y = \operatorname{diag}(y)$. Then

$$\begin{aligned} A^k S^{-2} (A^k)^T &= Y^{-2} + (I - Y)^{-2} + \sum_{j=1}^{k} \frac{a_j a_j^T}{(s_{2m+j})^2} \\ &\succeq Y^{-2} + (I - Y)^{-2} + \frac{1}{m} \sum_{j=1}^{k} a_j a_j^T \quad \text{(by Lemma 8.12)} \\ &\succeq 8I + \frac{1}{m} \sum_{j=1}^{k} a_j a_j^T \quad (\text{as } 0 \le y \le e) \\ &= B^k \end{aligned}$$

□

Lemma 8.14 *Let $\bar{s}^k = c^k - (A^k)^T \bar{y}^k$ be the slack vector at the analytic center $\bar{y}^k$ of Ω^k and $(\omega^k)^2 = a_{k+1}^T (B^k)^{-1} a_{k+1}$, then*

$$(\omega^k)^2 \ge a_{k+1}^T (A^k (\bar{S}^k)^{-2} (A^k)^T)^{-1} a_{k+1} = (\bar{r}^k)^2.$$

This lemma implies that upper bounds on the series of $(\omega^k)^2$ will lead to upper bounds on the series $(\bar{r}^k)^2$.

Lemma 8.15

$$\sum_{j=0}^{k} (\omega^j)^2 \le 2m^2 \log(1 + \frac{k+1}{8m^2}).$$

Proof. Note that

$$\det B^{k+1} = \det(B^k + \frac{1}{m} a_{k+1} a_{k+1}^T) = (1 + \frac{(\omega^k)^2}{m}) \det B^k.$$

Thus

$$\log\det B^{k+1} = \log\det B^k + \log(1 + \frac{(\omega^k)^2}{m}).$$

But

$$\frac{(\omega^k)^2}{m} \le \frac{1}{8} a_{k+1}^T a_{k+1} = \frac{1}{8};$$

hence

$$\begin{aligned} \log(1 + \frac{(\omega^k)^2}{m}) &\ge \frac{(\omega^k)^2}{m} - \frac{(\frac{(\omega^k)^2}{m})^2}{2(1 - \frac{(\omega^k)^2}{m})} \\ &= \frac{(\omega^k)^2}{m}(1 - \frac{\frac{(\omega^k)^2}{m}}{2(1 - \frac{(\omega^k)^2}{m})}) \\ &\ge \frac{(\omega^k)^2}{2m}. \end{aligned}$$

Thus we have

$$\log\det B^{k+1} \ge \log\det B^0 + \sum_{j=0}^{k} \frac{(\omega^j)^2}{2m} = m\log 8 + \sum_{j=0}^{k} \frac{(\omega^j)^2}{2m}.$$

But

$$\frac{1}{m}\log\det B^{k+1} \le \log \frac{\text{trace } B^{k+1}}{m} = \log(8 + \frac{k+1}{m^2}).$$

Thus

$$\sum_{j=0}^{k} \frac{(\omega^j)^2}{2m} \le m\log(8 + \frac{k+1}{m^2}) - m\log 8$$

or

$$\sum_{j=0}^{k} (\omega^j)^2 \le 2m^2 \log(1 + \frac{k+1}{8m^2}).$$

□

Theorem 8.16 *The cutting plane algorithm stops with a feasible solution as soon as k satisfies:*

$$\frac{\epsilon^2}{m} \ge \frac{\frac{1}{2} + 2m\log(1 + \frac{k+1}{8m^2})}{2m+k+1} \exp(-2\alpha \frac{k+1}{k+1+2m}).$$

Proof. From relation (8.18) and Lemma 8.11,

$$\begin{aligned} (2m+k+1)\log\epsilon &\le \mathcal{B}(\Omega^{k+1}) \\ &\le \mathcal{B}(\Omega^0) + \frac{1}{2}\sum_{j=0}^{k} \log(\bar{r}^j)^2 - (k+1)\alpha \\ &= 2m\log\frac{1}{2} + \frac{1}{2}\sum_{j=0}^{k}\log(\bar{r}^j)^2 - (k+1)\alpha. \end{aligned}$$

Thus

$$
\begin{aligned}
&\log \epsilon + \frac{k+1}{2m+k+1}\alpha \\
\leq\;& \frac{1}{2(2m+k+1)}\left[2m\log\frac{1}{4} + \sum_{j=0}^{k}\log(\bar{r}^j)^2\right] \\
\leq\;& \frac{1}{2}\log\frac{2m\frac{1}{4} + \sum_{j=0}^{k}(\bar{r}^j)^2}{2m+k+1} \quad \text{(from the concavity of } \log\text{)} \\
\leq\;& \frac{1}{2}\log\frac{\frac{m}{2} + \sum_{j=0}^{k}(\omega^j)^2}{2m+k+1} \quad \text{(from Lemma 8.14)} \\
\leq\;& \frac{1}{2}\log\frac{\frac{m}{2} + 2m^2\log(1+\frac{k+1}{8m^2})}{2m+k+1} \quad \text{(from Lemma 8.15)}
\end{aligned}
$$

or

$$\frac{\epsilon^2}{m} \leq \frac{\frac{1}{2} + 2m\log(1+\frac{k+1}{8m^2})}{2m+k+1}\exp(-2\alpha\frac{k+1}{k+1+2m}).$$

□

Theorem 8.16 implies that the complexity of the column generation scheme, counted by the calls to the oracle, is $O^*(\frac{m^2}{\epsilon^2})$; the notation O^* means that lower order terms are ignored. The largest value of η that guarantees $\gamma^{16} \leq \eta$ (so that 4 dual Newton steps are enough to recenter) is about $\eta = .09$ with $\beta = .691$. In this case α may be negative; nonetheless the algorithm will still terminate after $O^*(\frac{m^2}{\epsilon^2})$ iterations.

Theorem 8.17 *The approximate analytic center algorithm, which uses the updating scheme of Section 8.2.2 and the Newton method, is, for appropriate values of η and β which depend on the exact mix of recentering and updating steps, a fully polynomial-time approximation scheme.*

8.3 Positive Semi-Definite Programming

Recall that $\mathcal{M}^n$ denotes the set of symmetric matrices in $\mathcal{R}^{n\times n}$. Let $\mathcal{M}^n_+$ denote the set of positive semi-definite matrices and $\overset{\circ}{\mathcal{M}}{}^n_+$ the set of positive definite matrices in $\mathcal{M}^n$. The goal of this section is to extend interior-point algorithms to solving the positive semi-definite programming problem:

$$
\begin{array}{lll}
(PSP) & \inf & C \bullet X \\
& \text{subject to} & A_i \bullet X = b_i, i = 1,2,...,m,\ X \succeq 0,
\end{array}
$$

where $C \in \mathcal{M}^n$, $A_i \in \mathcal{M}^n$, $i = 1, 2, ..., m$, and $b \in \mathcal{R}^m$. The dual problem can be written as

$$\begin{array}{lll} (PSD) & \sup & b^T y \\ & \text{subject to} & \sum_i^m y_i A_i + S = C, \; S \succeq 0, \end{array}$$

which is analogous to the dual of LP.

Denote the primal feasible set by $\mathcal{F}_p$ and the dual by $\mathcal{F}_d$. We assume that both $\overset{\circ}{\mathcal{F}}_p$ and $\overset{\circ}{\mathcal{F}}_d$ are nonempty. Thus, the optimal solution sets for both (PSP) and (PSD) are bounded. Let z^* denote the optimal value and $\mathcal{F} = \mathcal{F}_p \times \mathcal{F}_d$. In this section, we are interested in finding an ϵ approximate solution for the PSP problem:

$$C \bullet X - b^T y = S \bullet X \le \epsilon.$$

For simplicity, we assume that a central path pair (X^0, y^0, S^0), which satisfies

$$(X^0)^{.5} S^0 (X^0)^{.5} = \mu^0 I \quad \text{and} \quad \mu^0 = X^0 \bullet S^0/n,$$

is known. We will use it as our initial point throughout this section.

Define the "∞-norm" of $\mathcal{M}^n$ by

$$\|X\|_\infty = \max_{j \in \{1,...,n\}} \{|\lambda_j(X)|\},$$

where $\lambda_j(X)$ is the jth eigenvalue of X, and the "Euclidean" norm by

$$\|X\| = \sqrt{X \bullet X} = \sqrt{\sum_{j=1}^n (\lambda_j(X))^2}.$$

Furthermore, note that for $X \in \mathcal{M}^n$

$$\text{tr}(X) = \sum_{j=1}^n \lambda_j(X) \quad \text{and} \quad \det(I + X) = \prod_{j=1}^n (1 + \lambda_j(X)).$$

Then, we have the following lemma which resembles Lemma 3.1.

Lemma 8.18 *Let $X \in \mathcal{M}^n$ and $\|X\|_\infty < 1$. Then,*

$$tr(X) \ge \log\det(I + X) \ge tr(X) - \frac{\|X\|^2}{2(1 - \|X\|_\infty)}.$$

8.3.1 Potential reduction algorithm

Let $X \in \overset{\circ}{\mathcal{F}}_p$, $(y,S) \in \overset{\circ}{\mathcal{F}}_d$, and $z \le z^*$. Then consider the primal potential function

$$\mathcal{P}(X,z) = (n+\rho)\log(C \bullet X - z) - \log\det X,$$

and the primal-dual potential function

$$\psi(X,S) = (n+\rho)\log(S \bullet X) - \log\det XS,$$

where $\rho = \sqrt{n}$. Let $z = b^T y$. Then $S \bullet X = C \bullet X - z$, and we have

$$\psi(x,s) = \mathcal{P}(x,z) - \log\det S.$$

Consider a pair of $(X^k, y^k, S^k) \in \overset{\circ}{\mathcal{F}}$. Fix $z^k = b^T y^k$, then the gradient vector of the primal potential function at X^k is

$$\nabla\mathcal{P}(X^k, z^k) = \frac{n+\rho}{S^k \bullet X^k} C - (X^k)^{-1}.$$

The following corollary is an analog to inequality (3.15).

Corollary 8.19 *Let $X^k \in \overset{\circ}{\mathcal{M}}{}^n_+$ and $\|(X^k)^{-.5}(X - X^k)(X^k)^{-.5}\|_\infty < 1$. Then, $X \in \overset{\circ}{\mathcal{M}}{}^n_+$ and*

$$\mathcal{P}(X,z^k) - \mathcal{P}(X^k,z^k) \le \nabla\mathcal{P}(X^k,z^k) \bullet (X - X^k)$$
$$+\frac{\|(X^k)^{-.5}(X - X^k)(X^k)^{-.5}\|^2}{2(1 - \|(X^k)^{-.5}(X - X^k)(X^k)^{-.5}\|_\infty)}.$$

Let

$$\mathcal{A} = \begin{pmatrix} A_1 \\ A_2 \\ \dots \\ A_m \end{pmatrix}.$$

Then, define

$$\mathcal{A}X = \begin{pmatrix} A_1 \bullet X \\ A_2 \bullet X \\ \dots \\ A_m \bullet X \end{pmatrix} = b,$$

and

$$\mathcal{A}^T y = \sum_{i=1}^m y_i A_i.$$

Then, we directly solve the following problem

$$\begin{array}{ll} \text{minimize} & \nabla\mathcal{P}(X^k, z^k) \bullet (X - X^k) \\ \text{subject to} & \mathcal{A}(X - X^k) = 0, \\ & \|(X^k)^{-.5}(X - X^k)(X^k)^{-.5}\| \le \alpha < 1. \end{array}$$

Let $X' = (X^k)^{-.5}X(X^k)^{-.5}$. Note that for any symmetric matrices $Q, T \in \mathcal{M}^n$ and $X \in \overset{\circ}{\mathcal{M}}{}^n_+$,

$$Q \bullet X^{.5}TX^{.5} = X^{.5}QX^{.5} \bullet T \text{ and } \|XQ\|_. = \|QX\|_. = \|X^{.5}QX^{.5}\|_..$$

Then we transform the above problem into

$$\begin{array}{ll} \text{minimize} & (X^k)^{.5}\nabla\mathcal{P}(X^k, z^k)(X^k)^{.5} \bullet (X' - I) \\ \text{subject to} & \mathcal{A}'(X' - I) = 0, \ i = 1, 2, ..., i, \\ & \|X' - I\| \le \alpha, \end{array}$$

where

$$\mathcal{A}' = \begin{pmatrix} A'_1 \\ A'_2 \\ \ldots \\ A'_m \end{pmatrix} := \begin{pmatrix} (X^k)^{.5}A_1(X^k)^{.5} \\ (X^k)^{.5}A_2(X^k)^{.5} \\ \ldots \\ (X^k)^{.5}A_m(X^k)^{.5} \end{pmatrix}.$$

Let the minimizer be X' and let $X^{k+1} = (X^k)^{.5}X'(X^k)^{.5}$. Then

$$X' - I = -\alpha\frac{P^k}{\|P^k\|},$$

$$X^{k+1} - X^k = -\alpha\frac{(X^k)^{.5}P^k(X^k)^{.5}}{\|P^k\|}, \tag{8.19}$$

where

$$\begin{array}{rcl} P^k & = & \mathcal{P}_{\mathcal{A}'}(X^k)^{.5}\nabla\mathcal{P}(X^k, z^k)(X^k)^{.5} \\ & = & (X^k)^{.5}\nabla\mathcal{P}(X^k, z^k)(X^k)^{.5} - \mathcal{A}'^T y^k \end{array}$$

or

$$P^k = \frac{n + \rho}{S^k \bullet X^k}(X^k)^{.5}(C - \mathcal{A}^T y^k)(X^k)^{.5} - I,$$

and

$$y^k = \frac{S^k \bullet X^k}{n + \rho}(\mathcal{A}'\mathcal{A}'^T)^{-1}\mathcal{A}'(X^k)^{.5}\nabla\mathcal{P}(X^k, z^k)(X^k)^{.5}.$$

Here, $\mathcal{P}_{\mathcal{A}'}$ is the projection operator onto the null space of $\mathcal{A}'$, and

$$\mathcal{A}'\mathcal{A}'^T := \begin{pmatrix} A'_1 \bullet A'_1 & A'_1 \bullet A'_2 & \dots & A'_1 \bullet A'_m \\ A'_2 \bullet A'_1 & A'_2 \bullet A'_2 & \dots & A'_2 \bullet A'_m \\ \dots & \dots & \dots & \dots \\ A'_m \bullet A'_1 & A'_m \bullet A'_2 & \dots & A'_m \bullet A'_m \end{pmatrix} \in \mathcal{M}^m.$$

In view of Corollary 8.19 and

$$\begin{aligned} \nabla\mathcal{P}(X^k, z^k) \bullet (X^{k+1} - X^k) &= -\alpha \frac{\nabla\mathcal{P}(X^k, z^k) \bullet (X^k)^{.5} P^k (X^k)^{.5}}{\|P^k\|} \\ &= -\alpha \frac{(X^k)^{.5} \nabla\mathcal{P}(X^k, z^k)(X^k)^{.5} \bullet P^k}{\|P^k\|} \\ &= -\alpha \frac{\|P^k\|^2}{\|P^k\|} = -\alpha\|P^k\|, \end{aligned}$$

we have

$$\mathcal{P}(X^{k+1}, z^k) - \mathcal{P}(X^k, z^k) \le -\alpha\|P^k\| + \frac{\alpha^2}{2(1-\alpha)}.$$

Thus, as long as $\|P^k\| \ge \beta > 0$, we may choose an appropriate α such that

$$\mathcal{P}(X^{k+1}, z^k) - \mathcal{P}(X^k, z^k) \le -\delta$$

for some positive constant δ.

Now, we focus on the expression of P^k, which can be rewritten as

$$P(z^k) := P^k = \frac{n+\rho}{S^k \bullet X^k} (X^k)^{.5} S(z^k) (X^k)^{.5} - I \tag{8.20}$$

with

$$S(z^k) = C - \mathcal{A}^T y(z^k) \tag{8.21}$$

and

$$y(z^k) := y^k = y_2 - \frac{S^k \bullet X^k}{n+\rho} y_1 = y_2 - \frac{C \bullet X^k - z^k}{n+\rho} y_1 \tag{8.22}$$

where y_1 and y_2 are given by

$$\begin{aligned} y_1 &= (\mathcal{A}'\mathcal{A}'^T)^{-1}\mathcal{A}' I = (\mathcal{A}'\mathcal{A}'^T)^{-1} b, \\ y_2 &= (\mathcal{A}'\mathcal{A}'^T)^{-1}\mathcal{A}'(X^k)^{.5} C (X^k)^{.5}. \end{aligned} \tag{8.23}$$

Regarding $\|P^k\| = \|P(z^k)\|$, we have the following lemma resembling Lemma 4.8.

Lemma 8.20 *Let*

$$\mu^k = \frac{S^k \bullet X^k}{n} = \frac{C \bullet X^k - z^k}{n} \quad \textit{and} \quad \mu = \frac{S(z^k) \bullet X^k}{n}.$$

If

$$\|P(z^k)\| < \min(\beta\sqrt{\frac{n}{n+\beta^2}}, 1-\beta) \tag{8.24}$$

then the following three inequalities hold:

$$S(z^k) \succ 0, \quad \|(X^k)^{.5}S(z^k)(X^k)^{.5} - \mu e\| < \beta\mu, \quad \textit{and} \quad \mu < (1 - .5\beta/\sqrt{n})\mu^k. \tag{8.25}$$

Proof. The proof is by contradiction. For example, if the first inequality of (8.25) is not true, then $(X^k)^{.5}S(z^k)(X^k)^{.5}$ has at least one eigenvalue less than or equal to zero, and

$$\|P(z^k)\| \geq 1.$$

The proof of the second and third inequalities are similar to that of Lemma 4.8.

□

Based on this lemma, we have the following potential reduction theorem.

Theorem 8.21 *Given $X^k \in \overset{\circ}{\mathcal{F}}_p$ and $(y^k, S^k) \in \overset{\circ}{\mathcal{F}}_d$, let $\rho = \sqrt{n}$, $z^k = b^T y^k$, X^{k+1} be given by (8.19), and $y^{k+1} = y(z^k)$ in (8.22) and $S^{k+1} = S(z^k)$ in (8.21). Then, either*

$$\psi(X^{k+1}, S^k) \leq \psi(X^k, S^k) - \delta$$

or

$$\psi(X^k, S^{k+1}) \leq \psi(X^k, S^k) - \delta$$

where $\delta > 1/20$.

Proof. If (8.24) does not hold, i.e.,

$$\|P(z^k)\| \geq \min(\beta\sqrt{\frac{n}{n+\beta^2}}, 1-\beta),$$

then

$$\mathcal{P}(X^{k+1}, z^k) - \mathcal{P}(X^k, z^k) \leq -\alpha \min(\beta\sqrt{\frac{n}{n+\beta^2}}, 1-\beta) + \frac{\alpha^2}{2(1-\alpha)},$$

hence from the relation between $\mathcal{P}$ and ψ,

$$\psi(X^{k+1}, S^k) - \psi(X^k, S^k) \le -\alpha \min(\beta\sqrt{\frac{n}{n+\beta^2}}, 1-\beta) + \frac{\alpha^2}{2(1-\alpha)}.$$

Otherwise, from Lemma 8.20 the inequalities of (8.25) hold:

i) The first of (8.25) indicates that y^{k+1} and S^{k+1} are in $\overset{\circ}{\mathcal{F}}_d$.

ii) Using the second of (8.25) and applying Lemma 8.18 to matrix $(X^k)^{.5}S^{k+1}(X^k)^{.5}/\mu$, we have

$$\begin{aligned}
& n\log S^{k+1} \bullet X^k - \log\det S^{k+1}X^k \\
&= n\log S^{k+1} \bullet X^k/\mu - \log\det(X^k)^{.5}S^{k+1}(X^k)^{.5}/\mu \\
&= n\log n - \log\det(X^k)^{.5}S^{k+1}(X^k)^{.5}/\mu \\
&\le n\log n + \frac{\|(X^k)^{.5}S^{k+1}(X^k)^{.5}/\mu - I\|^2}{2(1-\|(X^k)^{.5}S^{k+1}(X^k)^{.5}/\mu - I\|_\infty)} \\
&\le n\log n + \frac{\beta^2}{2(1-\beta)} \\
&\le n\log S^k \bullet X^k - \log\det S^kX^k + \frac{\beta^2}{2(1-\beta)}.
\end{aligned}$$

iii) According to the third of (8.25), we have

$$\sqrt{n}(\log S^{k+1} \bullet X^k - \log S^k \bullet X^k) = \sqrt{n}\log\frac{\mu}{\mu^k} \le -\frac{\beta}{2}.$$

Adding the two inequalities in ii) and iii), we have

$$\psi(X^k, S^{k+1}) \le \psi(X^k, S^k) - \frac{\beta}{2} + \frac{\beta^2}{2(1-\beta)}.$$

Thus, by choosing $\beta = .43$ and $\alpha = .3$ we have the desired result.

$\square$

Theorem 8.21 establishes an important fact: the *primal-dual* potential function can be reduced by a constant no matter where X^k and y^k are. In practice, one can perform the line search to minimize the primal-dual potential function. This results in the following potential reduction algorithm.

Algorithm 8.3 *Given* $x^0 \in \overset{\circ}{\mathcal{F}}_p$ *and* $(y^0, s^0) \in \overset{\circ}{\mathcal{F}}_d$. *Let* $z^0 = b^T y^0$. *Set* $k := 0$.

While $S^k \bullet X^k \geq \epsilon$ **do**

1. *Compute* y_1 *and* y_2 *from (8.23).*

2. *Set* $y^{k+1} = y(\bar{z})$, $S^{k+1} = S(\bar{z})$, $z^{k+1} = b^T y^{k+1}$ *with*

$$\bar{z} = \arg\min_{z \geq z^k} \psi(X^k, S(z)).$$

If $\psi(X^k, S^{k+1}) > \psi(X^k, S^k)$ *then* $y^{k+1} = y^k$, $S^{k+1} = S^k$, $z^{k+1} = z^k$.

3. *Let* $X^{k+1} = X^k - \bar{\alpha}(X^k)^{.5} P(z^{k+1})(X^k)^{.5}$ *with*

$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi(X^k - \alpha(X^k)^{.5} P(z^{k+1})(X^k)^{.5}, S^{k+1}).$$

4. *Set* $k := k + 1$ *and return to Step 1.*

The performance of the algorithm results from the following corollary.

Corollary 8.22 *Let* $\rho = \sqrt{n}$. *Then, Algorithm 8.3 terminates in at most* $O(\sqrt{n} \log(C \bullet X^0 - b^T y^0)/\epsilon)$ *iterations with*

$$C \bullet X^k - b^T y^k \leq \epsilon.$$

Proof. In $O(\sqrt{n} \log(S^0 \bullet X^0/\epsilon))$ iterations

$$\begin{array}{rcl} -\sqrt{n} \log(S^0 \bullet X^0/\epsilon) & = & \psi(X^k, S^k) - \psi(X^0, S^0) \\ & \geq & \sqrt{n} \log S^k \bullet X^k + n \log n - \psi(X^0, S^0) \\ & = & \sqrt{n} \log(S^k \bullet X^k / S^0 \bullet X^0). \end{array}$$

Thus,

$$\sqrt{n} \log(C \bullet X^k - b^T y^k) = \sqrt{n} \log S^k \bullet X^k \leq \sqrt{n} \log \epsilon,$$

i.e.,

$$C \bullet X^k - b^T y^k = S^k \bullet X^k \leq \epsilon.$$

$\square$

8.3.2 Primal-dual algorithm

Once we have a pair $(X, y, S) \in \overset{\circ}{\mathcal{F}}$ with $\mu = S \bullet X/n$, we can apply the primal-dual Newton method to generate a new iterate X^+ and (y^+, S^+) as follows: Solve for d_X, d_y and d_S from the system of linear equations:

$$\begin{array}{rcl} D^{-1}d_X D^{-1} + d_S & = & R := \gamma\mu X^{-1} - S, \\ \mathcal{A}d_X & = & 0, \\ \mathcal{A}^T d_y + d_S & = & 0, \end{array} \tag{8.26}$$

where

$$D = X^{.5}(X^{.5}SX^{.5})^{-.5}X^{.5}.$$

Note that $d_S \bullet d_X = 0$.

This system can be written as

$$\begin{array}{rcl} d_{X'} + d_{S'} & = & R', \\ \mathcal{A}' d_{X'} & = & 0, \\ \mathcal{A}'^T d_y + d_{S'} & = & 0, \end{array} \tag{8.27}$$

where

$$d_{X'} = D^{-.5}d_X D^{-.5}, \quad d_{S'} = D^{.5}d_S D^{.5}, \quad R' = D^{.5}(\gamma\mu X^{-1} - S)D^{.5},$$

and

$$\mathcal{A}' = \begin{pmatrix} A'_1 \\ A'_2 \\ \ldots \\ A'_m \end{pmatrix} := \begin{pmatrix} D^{.5}A_1 D^{.5} \\ D^{.5}A_2 D^{.5} \\ \ldots \\ D^{.5}A_m D^{.5} \end{pmatrix}.$$

Again, we have $d_{S'} \bullet d_{X'} = 0$, and

$$d_y = (\mathcal{A}'\mathcal{A}'^T)^{-1}\mathcal{A}'R', \; d_{S'} = -\mathcal{A}'^T d_y, \text{ and } d_{X'} = R' - d_{S'}.$$

Then, assign

$$d_S = \mathcal{A}^T d_y \quad \text{and} \quad d_X = D(R - d_S)D.$$

Let

$$V^{1/2} = D^{-.5}XD^{-.5} = D^{.5}SD^{.5} \in \overset{\circ}{\mathcal{M}}{}_+^n.$$

Then, we can verify that $S \bullet X = I \bullet V$. We now present the following lemma whose proof is very similar to that for Lemmas 3.6 and 4.11 and will be omitted.

Lemma 8.23 *Let the direction d_X, d_y and d_S be generated by equation (8.26) with $\gamma = n/(n+\rho)$, and let*

$$\theta = \frac{\alpha}{\|V^{-1/2}\|_\infty \|\frac{I \bullet V}{n+\rho} V^{-1/2} - V^{1/2}\|}, \tag{8.28}$$

where α is a positive constant less than 1. Let

$$X^+ = X + \theta d_X, \quad y^+ = y + \theta d_y, \quad \text{and} \quad S^+ = S + \theta d_S.$$

Then, we have $(X^+, y^+, S^+) \in \overset{\circ}{\mathcal{F}}$ and

$$\psi(X^+, S^+) - \psi(X, S)$$

$$\leq -\alpha \frac{\|V^{-1/2} - \frac{n+\rho}{I \bullet V} V^{1/2}\|}{\|V^{-1/2}\|_\infty} + \frac{\alpha^2}{2(1-\alpha)}.$$

Applying Lemma 4.12 to $v \in \mathcal{R}^n$ as the vector of the n eigenvalues of V, we can prove the following lemma

Lemma 8.24 *Let $V \in \overset{\circ}{\mathcal{M}}{}^n_+$ and $\rho \geq \sqrt{n}$. Then,*

$$\frac{\|V^{-1/2} - \frac{n+\rho}{I \bullet V} V^{1/2}\|}{\|V^{-1/2}\|_\infty} \geq \sqrt{3/4}.$$

From these two lemmas we have

$$\psi(X^+, S^+) - \psi(X, S)$$

$$\leq -\alpha\sqrt{3/4} + \frac{\alpha^2}{2(1-\alpha)} = -\delta$$

for a constant δ. This leads to

Algorithm 8.4 *Given $(X^0, y^0, S^0) \in \overset{\circ}{\mathcal{F}}$. Set $\rho = \sqrt{n}$ and $k := 0$.*
While $S^k \bullet X^k \geq \epsilon$ do

1. *Set $(X, S) = (X^k, S^k)$ and $\gamma = n/(n+\rho)$ and compute (d_X, d_y, d_S) from (8.26).*

2. *Let $X^{k+1} = X^k + \bar{\alpha} d_X$, $y^{k+1} = y^k + \bar{\alpha} d_y$, and $S^{k+1} = S^k + \bar{\alpha} d_S$ where*

$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi(X^k + \alpha d_X, S^k + \alpha d_S).$$

3. *Set $k := k+1$ and return to Step 1.*

Theorem 8.25 *Let $\rho = \sqrt{n}$. Then, Algorithm 8.4 terminates in at most $O(\sqrt{n}\log(S^0 \bullet X^0/\epsilon))$ iterations with*

$$C \bullet X^k - b^T y^k \leq \epsilon.$$

Primal-dual adaptive path-following algorithms, the predictor-corrector algorithms and the wide-neighborhood algorithms similar to those in Section 4.5 can also be developed for solving (PSP).

8.4 Monotone Complementarity Problem

We present a generalization of the homogeneous self-dual linear programming (LP) algorithm to solving the monotone complementarity problem (MCP) in the form:

$$\begin{array}{lll} (MCP) & \text{minimize} & x^T s \\ & \text{subject to} & s = f(x), \ (x, s) \geq 0, \end{array}$$

where $(x, s) \in \mathcal{R}^{2n}$ and $f(x)$ is a continuous *monotone* mapping from $\mathcal{R}^n_+$ to $\mathcal{R}^n$. In other words, for every $x^1, x^2 \in \mathcal{R}^n_+$, we have

$$(x^1 - x^2)^T (f(x^1) - f(x^2)) \geq 0.$$

Denote by ∇f the Jacobian matrix of f, which is positive semi-definite in $\mathcal{R}^n_+$.

(MCP) is said to be (asymptotically) feasible if and only if there is a *bounded* sequence $(x^t > 0, s^t > 0)$, $t = 1, 2, ...$, such that

$$\lim_{t \to \infty} \quad s^t - f(x^t) \to 0,$$

where any limit point $(\hat{x}, \hat{s})$ of the sequence is called an (asymptotically) feasible point for (MCP). (MCP) has an interior feasible point if it has an (asymptotically) feasible point $(\hat{x} > 0, \hat{s} > 0)$. (MCP) is said to be (asymptotically) solvable if there is an (asymptotically) feasible $(\hat{x}, \hat{s})$ such that $\hat{x}^T \hat{s} = 0$, where $(\hat{x}, \hat{s})$ is called the "optimal" or "complementary" solution for (MCP). (MCP) is (strongly) infeasible if and only if there is no sequence $(x^t > 0, s^t > 0)$, $t = 1, 2, ...$, such that

$$\lim_{t \to \infty} \quad s^t - f(x^t) \to 0.$$

Denote the feasible set of (MCP) by $\mathcal{F}$ and the solution set by $\mathcal{S}$. Note that (MCP) being feasible does not imply that (MCP) has a solution. If (MCP) has a solution, then the solution set is convex and it contains a maximal solution (x^*, s^*) (the number positive components in (x^*, s^*) is maximal).

Consider a class of (MCP) where f satisfies the following condition. Let

$$\upsilon : (0,1) \rightarrow (1, \infty)$$

be a monotone increasing function such that

$$\|X(f(x+d_x) - f(x) - \nabla f(x) d_x)\|_1 \leq \upsilon(\alpha) d_x^T \nabla f(x) d_x \tag{8.29}$$

whenever

$$d_x \in \mathcal{R}^n, \quad x \in \mathcal{R}^n_{++}, \quad \left\|X^{-1} d_x\right\|_\infty \leq \alpha < 1.$$

Then, f is said to be scaled Lipschitz in $\overset{\circ}{\mathcal{R}}{}^n_+$.

Given $x^0 > 0$ and $s^0 = f(x^0) > 0$ one can develop an interior-point algorithm that generates a maximal complementary solution of the scaled Lipschitz (MCP) in $O(\sqrt{n} \log(1/\epsilon))$ interior-point iterations, where ϵ is the complementarity error.

However, the initial point x^0 is generally unknown. In fact, we don't even know whether such a point exists or not, that is, (MCP) might be infeasible or feasible but have no positive feasible point. To overcome this difficulty, in Section 5.3 we developed a homogeneous linear programming (LP) algorithm based on the construction of a homogeneous and self-dual LP model. In this section, we present a homogeneous model for solving the monotone complementarity problem. The algorithm again possesses the following desired features:

- It achieves $O(\sqrt{n} \log(1/\epsilon))$-iteration complexity if f satisfies the scaled Lipschitz condition.
- If (MCP) has a solution, the algorithm generates a sequence that approaches feasibility and optimality *simultaneously*; if the problem is (strongly) infeasible, the algorithm generates a sequence that converges to a certificate proving infeasibility.

8.4.1 A convex property

Let $f(x)$ be a continuous monotone mapping from $\overset{\circ}{\mathcal{R}}{}^n_+$ to $\mathcal{R}^n$. Consider the set of residuals

$$R_{++} = \{s - f(x) \in \mathcal{R}^n : \quad (x,s) > 0\},$$

and for a $r \in \mathcal{R}^n$ let

$$S_{++}(r) = \{(x,s) \in \mathcal{R}^{2n}_{++} : \ s = f(x) + r\}.$$

Since f is continuous in $\overset{\circ}{\mathcal{R}}{}^n_+$, we can easily verify that R_{++} is an open set. Furthermore, we have

Theorem 8.26 *Consider the mapping $F = (Xy, s - f(x)) \in \mathcal{R}^{2n}$ from $(x, s) \in \mathcal{R}^{2n}$. Then F maps $\mathcal{R}^{2n}_{++}$ onto $\overset{\circ}{\mathcal{R}}{}^n_+ \times R_{++}$ homeomorphically, that is, F is one-to-one on $\mathcal{R}^{2n}_{++}$, F maps $\mathcal{R}^{2n}_{++}$ onto $\overset{\circ}{\mathcal{R}}{}^n_+ \times R_{++}$, and the inverse mapping F^{-1} is continuous on $\overset{\circ}{\mathcal{R}}{}^n_+ \times R_{++}$.*

Simply using the monotone of f, we also have the following lemma

Lemma 8.27 *Let $r \in \mathcal{R}^n$. Assume that $(x^1, s^1) \in S_{++}(\theta^1 r)$ and $(x^2, s^2) \in S_{++}(\theta^2 r)$ where θ^1 and θ^2 are two real numbers. Then*

$$(\theta^2 - \theta^1) r^T (x^2 - x^1) \le (x^2 - x^1)^T (s^2 - s^1).$$

This lemma leads to the next lemma:

Lemma 8.28 *Let $r \in \mathcal{R}^n$ and $\theta^0 \le \theta^1$. Assume $S_{++}(\theta^0 r) \ne \emptyset$ and $S_{++}(\theta^1 r) \ne \emptyset$. Then, for every $\delta > 0$, the union of*

$$C_{++}(\theta r, \delta) = \{(x, y) \in S_{++}(\theta r) : \ x^T y \le \delta\}, \quad \theta \in [\theta^0, \theta^1]$$

is bounded.

Proof. Let $(x^0, s^0) \in S_{++}(\theta^0 r)$ and $(x^1, s^1) \in S_{++}(\theta^1 r)$, and

$$\max((s^0)^T x^0, (s^1)^T x^1) \le \delta.$$

Let $\theta \in [\theta^0, \theta^1]$ and $(x, s) \in C_{++}(\theta r, \delta)$. Then we have by Lemma 8.27 that

$$(s^1)^T x + (x^1)^T s \le (\theta^1 - \theta) r^T x + c^1$$

and

$$(s^0)^T x + (x^0)^T s \le (\theta^0 - \theta) r^T x + c^0,$$

where

$$c^1 = (\theta^1 - \theta^0)|r^T x^1| + 2\delta \text{ and } c^0 = (\theta^1 - \theta^0)|r^T x^0| + 2\delta.$$

Thus, if $\theta^0 < \theta^1$ then we have

$$\begin{aligned}(\theta - \theta^0)((s^1)^T x + (x^1)^T s) + (\theta^1 - \theta)((s^0)^T x + (x^0)^T s) \\ \le (\theta - \theta^0) c^1 + (\theta^1 - \theta) c^0.\end{aligned}$$

Thus, we have

$$e^T x + e^T s \le \frac{\max\{c^1, c^0\}}{\min\{(x^0, s^0, x^1, s^1)\}},$$

which implies that (x, s) is bounded. The lemma is obviously true if $\theta^0 = \theta^1$.

□

To prove the convexity of R_{++}, it suffices to prove that if the system

$$s = f(x) + r^0 + \theta r, \quad (x, s) > 0$$

has a solution at $\theta = 0$ and $\theta = 1$, then it has a solution for every $\theta \in [0, 1]$. Without loss of generality, we may assume $r^0 = 0$. Let $(x^0, s^0) \in S_{++}(0)$, $(x^1, s^1) \in S_{++}(r)$, and $\max((s^0)^T x^0, (s^1)^T x^1) \le \delta^*$. Now consider the system

$$Xs = (1-\theta)X^0 s^0 + \theta X^1 s^1 \quad \text{and} \quad s - f(x) = \theta r, \ (x, s) > 0. \qquad (8.30)$$

Let

$$\Theta = \{\theta \in \mathcal{R} : \text{ system (8.30) has a solution}\}.$$

Then, from the openness of R_{++} and Theorem 8.26 we can derive

Lemma 8.29 Θ *is an open set and system (8.30) has a unique solution* $(x(\theta), s(\theta))$ *for every* $\theta \in \Theta$. *Moreover,* $(x(\theta), s(\theta))$ *is continuous in* $\theta \in \Theta$.

We now ready to prove the following theorem.

Theorem 8.30 R_{++} *is an open convex subset of* $\mathcal{R}^n$.

Proof. The openness has been discussed earlier. Let

$$\theta^* = \inf\{\theta : [\theta, 1] \subset \Theta\}.$$

Since $1 \in \Theta$, we know by Lemma 8.29 that $\theta^* < 1$ and $\theta^* \notin \Theta$. If $\theta^* < 0$, $\theta r \in R_{++}$ for every $\theta \in [0, 1]$; hence the theorem follows. Suppose on the contrary that $\theta^* > 0$. Let $\{\theta^k \in (\theta^*, 1]\}$ be a sequence converging to θ^*. Then, for $k = 1, 2, ...$, we have

$$X(\theta^k)s(\theta^k) = (1-\theta^k)X^0 s^0 + \theta^k X^1 s^1,$$

$$s(\theta^k) - f(x(\theta^k)) = \theta^k r, \quad \text{and} \quad (x(\theta^k), s(\theta^k)) > 0,$$

which implies that

$$(x(\theta^k))^T s(\theta^k) = (1-\theta^k)(x^0)^T s^0 + \theta^k (x^1)^T s^1 \le \delta^*.$$

Thus, $(x(\theta^k), s(\theta^k))$ is in the union of $C_{++}(\theta r, \delta^*)$, $\theta \in [0, 1]$. Since the union is a bounded subset by Lemma 8.28, we may assume without loss of generality that the sequence $\{(x(\theta^k), s(\theta^k))\}$ converges to some $(\bar{x}, \bar{s}) \in$

$\mathcal{R}^{2n}_+$. By the continuity of the mapping $Xy: R^{2n}_+ \to R^n_+$, the point $(\bar{x}, \bar{s})$ satisfies

$$\bar{X}\bar{s} = (1-\theta^*)X^0 s^0 + \theta^* X^1 s^1 \in R^n_{++}.$$

By the continuity of $f: R^n_{++} \to R^n$, we then see that

$$\bar{s} - f(\bar{x}) = (1-\theta^*)0 + \theta^* r = \theta^* r.$$

This implies that $(\bar{x}, \bar{s}) \in S_{++}(\theta^* r)$ and $\theta^* \in \Theta$, the contradiction of which is $\theta^* \notin \Theta$. Thus we have shown that R_{++} is a convex set.

□

8.4.2 A homogeneous MCP model

Consider an augmented homogeneous model related to (MCP):

$$(HMCP) \quad \text{minimize} \quad x^T s + \tau\kappa$$

$$\text{subject to} \quad \begin{pmatrix} s \\ \kappa \end{pmatrix} = \begin{pmatrix} \tau f(x/\tau) \\ -x^T f(x/\tau) \end{pmatrix}, \; (x, \tau, s, \kappa) \geq 0.$$

Let

$$\psi(x, \tau) = \begin{pmatrix} \tau f(x/\tau) \\ -x^T f(x/\tau) \end{pmatrix} : \overset{\circ}{\mathcal{R}}{}^{n+1}_+ \to \mathcal{R}^{n+1}. \tag{8.31}$$

Then, it is easy to verify that $\nabla\psi$ is positive semi-definite as shown in the following lemma.

Lemma 8.31 *Let ∇f be positive semi-definite in $\mathcal{R}^n_+$. Then $\nabla\psi$ is positive semi-definite in $\overset{\circ}{\mathcal{R}}{}^{n+1}_+$, i.e. given $(x;\tau) > 0$,*

$$(d_x; d_\tau)^T \nabla\psi(x, \tau)(d_x; d_\tau) \geq 0$$

for any $(d_x; d_\tau) \in \mathcal{R}^{n+1}$, where

$$\nabla\psi(x, \tau) = \begin{pmatrix} \nabla f(x/\tau) & f(x/\tau) - \nabla f(x/\tau)(x/\tau) \\ -f(x/\tau)^T - (x/\tau)^T \nabla f(x/\tau)^T & (x/\tau)^T \nabla f(x/\tau)(x/\tau) \end{pmatrix}. \tag{8.32}$$

Proof.

$$\begin{array}{rl} & (d_x; d_\tau)^T \nabla\psi(x, \tau)(d_x; d_\tau) \\ = & d_x^T \nabla f(x/\tau) d_x - d_x^T \nabla f(x/\tau) x (d_\tau/\tau) \\ & -(d_\tau/tau) x^T \nabla f(x/\tau)^T d_x + d_\tau^2 x^T \nabla f(x/\tau) x/\tau^2 \\ = & (d_x - d_\tau x/\tau)^T \nabla f(x/\tau)(d_x - d_\tau x/\tau) \end{array} \tag{8.33}$$

□

Furthermore, we have the following theorem.

Theorem 8.32 *Let ψ be given by (8.31). Then,*

i. *ψ is a continuous homogeneous function in $\overset{\circ}{\mathcal{R}}{}_{+}^{n+1}$ with degree 1 and for any $(x;\tau) \in \overset{\circ}{\mathcal{R}}{}_{+}^{n+1}$*

$$(x;\tau)^T \psi(x,\tau) = 0$$

and

$$(x;\tau)^T \nabla\psi(x,\tau) = -\psi(x,\tau)^T.$$

ii. *If f is a continuous monotone mapping from $\mathcal{R}^n_+$ to $\mathcal{R}^n$, then ψ is a continuous monotone mapping from $\overset{\circ}{\mathcal{R}}{}_{+}^{n+1}$ to $\mathcal{R}^{n+1}$.*

iii. *If f is scaled Lipschitz with $\upsilon = \upsilon_f$, then ψ is scaled Lipschitz, that is, it satisfies condition (8.29) with*

$$\upsilon = \upsilon_\psi(\alpha) = \left(1 + \frac{2\upsilon_f(2\alpha/(1+\alpha))}{1-\alpha}\right)\left(\frac{1}{1-\alpha}\right).$$

iv. *$(HMCP)$ is (asymptotically) feasible and every (asymptotically) feasible point is an (asymptotically) complementary solution.*

Now, let $(x^*, \tau^*, s^*, \kappa^*)$ be a maximal complementary solution for $(HMCP)$. Then

v. *(MCP) has a solution if and only if $\tau^* > 0$. In this case, $(x^*/\tau^*, s^*/\tau^*)$ is a complementary solution for (MCP).*

vi. *(MCP) is (strongly) infeasible if and only if $\kappa^* > 0$. In this case, $(x^*/\kappa^*, s^*/\kappa^*)$ is a certificate to prove (strong) infeasibility.*

Proof. The proof of (i) is straightforward.

We leave the proof of (ii) as an exercise.

We now prove (iii). Assume $(x;\tau) \in R^{n+1}_{++}$ and let $(d_x; d_\tau)$ be given such that $\|(X^{-1}d_x; \tau^{-1}d_\tau)\|_\infty \le \alpha < 1$. To prove ψ is scaled Lipschitz we must bound

$$\left\| \begin{pmatrix} X & 0 \\ 0 & \tau \end{pmatrix} \left(\psi(x + d_x, \tau + d_\tau) - \psi(x,\tau) - \nabla\psi(x,\tau) \begin{pmatrix} d_x \\ d_\tau \end{pmatrix} \right) \right\|_1. \tag{8.34}$$

From (8.31) and (8.32), the upper part in (8.34) is identical to

$$\begin{array}{rl} & X\left(f(y+d_y)(\tau+d_\tau)-f(y)\tau-(\nabla f(y)d_x+f(y)d_\tau-\nabla f(y)xd_\tau/\tau)\right) \\ = & (\tau+d_\tau)X\left(f(y+d_y)-f(y)-\nabla f(y)d_y\right) \\ = & \tau(\tau+d_\tau)Y\left(f(y+d_y)-f(y)-\nabla f(y)d_y\right) \end{array} \tag{8.35}$$

where

$$y=x/\tau \quad \text{and} \quad y+d_y=\frac{x+d_x}{\tau+d_\tau}, \tag{8.36}$$

that is,

$$d_y=\frac{\tau d_x-xd_\tau}{\tau(\tau+d_\tau)}=\frac{d_x-(d_\tau/\tau)x}{\tau+d_\tau}. \tag{8.37}$$

Note

$$\begin{array}{rcl} \left\|Y^{-1}d_y\right\|_\infty & = & \left\|\tau X^{-1}(\tau d_x-d_\tau x)/(\tau(\tau+d_\tau))\right\|_\infty \\ & = & \left\|(\tau X^{-1}d_x-d_\tau e)/(\tau+d_\tau)\right\|_\infty \\ & \le & (\left\|X^{-1}d_x\right\|_\infty+\alpha)/(1-\alpha) \\ & \le & 2\alpha/(1-\alpha). \end{array} \tag{8.38}$$

Per the assumption that f is scaled Lipschitz with $\upsilon=\upsilon_f$, it follows for $\alpha\in[0,1)$ that

$$\begin{array}{rl} & \|\tau(\tau+d_\tau)Y(f(y+d_y)-f(y)-\nabla f(y)d_y)\|_1 \\ \le & \tau(\tau+d_\tau)\upsilon_f(2\alpha/(1-\alpha))d_y^T\nabla f(y)d_y \\ = & \frac{\tau\upsilon_f(2\alpha/(1-\alpha))}{\tau+d_\tau}(d_x-xd_\tau/\tau)^T\nabla f(y)(d_x-xd_\tau/\tau) \\ = & \frac{\upsilon_f(2\alpha/(1-\alpha))}{1+d_\tau/\tau}(d_x;d_\tau)^T\nabla\psi(x,\tau)(d_x;d_\tau) \\ \le & \frac{\upsilon_f(2\alpha/(1-\alpha))}{1-\alpha}(d_x;d_\tau)^T\nabla\psi(x,\tau)(d_x;d_\tau). \end{array} \tag{8.39}$$

Next we bound the lower part of (8.34). This part is equal to

$$\begin{array}{rl} & \tau\left(-f(y+d_y)^T(x+d_x)-(-f(y)^Tx)\right. \\ & \left.-[-f(y)^Td_x-x^T\nabla f(y)^Td_x/\tau+x^Tf(y)xd_\tau/\tau^2]\right) \\ = & \tau\left((x+d_x)^T(-f(y+d_y)+f(y)+\nabla f(y)d_y)\right. \\ & \left.-(x+d_x)^T\nabla f(y)d_y+(x/\tau)^T\nabla f(y)d_y(\tau+d_\tau)\right) \\ = & \tau\left((x+d_x)^T(-f(y+d_y)+f(y)+\nabla f(y)d_y)\right. \\ & \left.-(d_x-d_\tau x/\tau)^T\nabla f(y)d_y\right) \\ = & \tau^2(e+X^{-1}d_x)^TY\left(-f(y+d_y)+f(y)+\nabla f(y)d_y\right) \\ & -\tau(\tau+d_\tau)d_y^T\nabla f(y)d_y. \end{array}$$

Thus, using (8.33) and (8.38)

$$\begin{array}{ll} & |\tau\left(-f(y+d_y)^T(x+d_x)-(-f(y)^Tx)\right. \\ & -[-f(y)^Td_x - x^T\nabla f(y)^Td_x/\tau + x^Tf(y)xd_\tau/\tau^2])\,| \\ \leq & \tau^2\left\|e+X^{-1}d_x\right\|_\infty \|Y(-f(y+d_y)+(-f(y)-\nabla f(y)d_y))\|_1 \\ & +|\tau(\tau+d_\tau)|d_y^T\nabla f(y)d_y \\ \leq & \left(\tau^2(1+\alpha)\upsilon_f(2\alpha/(1-\alpha))+\tau(\tau+d_\tau)\right)d_y^T\nabla f(y)d_y \\ = & \frac{\tau^2(1+\alpha)\upsilon_f(2\alpha/(1-\alpha))+\tau(\tau+d_\tau)}{(\tau+d_\tau)^2}(d_x^T-d_\tau x/\tau)^T\nabla f(y)(d_x-d_\tau x/\tau) \\ = & \frac{(1+\alpha)\upsilon_f(2\alpha/(1-\alpha))+(1+d_\tau/\tau)}{(1+d_\tau/\tau)^2}(d_x;d_\tau)^T\nabla\psi(x,\tau)(d_x;d_\tau) \\ \leq & \left(\frac{(1+\alpha)\upsilon_f(2\alpha/(1-\alpha))}{(1-\alpha)^2}+\frac{1}{1-\alpha}\right)(d_x;d_\tau)^T\nabla\psi(x,\tau)(d_x;d_\tau). \end{array} \tag{8.40}$$

The sum of (8.39) and (8.40) is equal to

$$\upsilon_\psi(\alpha)(d_x;d_\tau)^T\nabla\psi(x,\tau)(d_x;d_\tau)$$

and it bounds the term in (8.34) leading to the desired result.

We leave the proof of (iv) as an exercise too.

We now prove (v). If $(x^*,\tau^*,s^*,\kappa^*)$ is a solution for $(HMCP)$ and $\tau^*>0$, then we have

$$s^*/\tau^* = f(x^*/\tau^*) \quad \text{and} \quad (x^*)^Ts^*/(\tau^*)^2 = 0,$$

that is, $(x^*/\tau^*, s^*/\tau^*)$ is a solution for (MCP). Let $(\hat{x},\hat{s})$ be a solution to (MCP). Then $\tau=1$, $x=\hat{x}$, $s=\hat{s}$, and $\kappa=0$ is a solution for $(HMCP)$. Thus, every maximal solution of $(HMCP)$ must have $\tau^*>0$.

Finally, we prove (vi). Consider the set

$$R_{++} = \{s - f(x) \in \mathcal{R}^n : \quad (x,s) > 0\}.$$

As proved in Theorem 8.30, R_{++} is an open convex set. If (MCP) is strongly infeasible, then we must have $0 \notin \bar{R}_{++}$ where $\bar{R}_{++}$ represents the closure of R_{++}. Thus, there is a hyperplane that separates 0 and $\bar{R}_{++}$, that is, there is a vector $a \in \mathcal{R}^n$ with $\|a\|=1$ and a positive number ξ such that

$$a^T(s-f(x)) \geq \xi > 0 \quad \forall\, x \geq 0,\ s \geq 0. \tag{8.41}$$

For $j=1,2,...,n$, set s_j sufficiently large, but fix x and the rest of s, $a_j \geq 0$ must be true. Thus,

$$a \geq 0, \quad \text{or} \quad a \in \mathcal{R}^n_+.$$

On the other hand, for any fixed x, we set $s=0$ and see that

$$-a^Tf(x) \geq \xi > 0 \quad \forall\, x \geq 0. \tag{8.42}$$

In particular,

$$-a^T f(ta) \geq \xi > 0 \quad \forall\, t \geq 0. \tag{8.43}$$

From the monotone of f, for every $x \in \mathcal{R}^n_+$ and any $t \geq 0$ we have

$$(tx - x)^T (f(tx) - f(x)) \geq 0.$$

Thus,

$$x^T f(tx) \geq x^T f(x) \tag{8.44}$$

and

$$\lim_{t \to \infty} x^T f(tx)/t \geq 0. \tag{8.45}$$

Thus, from (8.43) and (8.45)

$$\lim_{t \to \infty} a^T f(ta)/t = 0.$$

For an $x \in \mathcal{R}^n_+$, denote

$$f^\infty(x) := \lim_{t \to \infty} f(tx)/t,$$

where $f^\infty(x)$ represents the limit of any subsequence and its values may include ∞ or $-\infty$.

We now prove $f^\infty(a) \geq 0$. Suppose that $f^\infty(a)_j < -\delta$. Then consider the vector $x = a + \epsilon e_j$ where e_j is the vector with the jth component being 1 and zeros everywhere else. Then, for ϵ sufficiently small and t sufficiently large we have

$$\begin{array}{rcl}
x^T f(tx)/t & = & (a + \epsilon e_j)^T f(t(a + \epsilon e_j))/t \\
& = & a^T f(t(a + \epsilon e_j))/t + \epsilon e_j^T f(t(a + \epsilon e_j))/t \\
& < & \epsilon e_j^T f(t(a + \epsilon e_j))/t \quad \text{(from (8.42))} \\
& = & \epsilon \dfrac{f(t(a + \epsilon e_j))_j - f(ta)_j}{t} + \epsilon \dfrac{f(ta)_j}{t} \\
& \leq & \epsilon \left(O(\epsilon) + \dfrac{f(ta)_j}{t} \right) \quad \text{(from continuity of } f) \\
& = & \epsilon (O(\epsilon) - \delta/2) \\
& \leq & -\epsilon\delta/4.
\end{array}$$

But this contradicts relation (8.45). Thus, we must have

$$f^\infty(a) \geq 0.$$

We now further prove that $f^\infty(a)$ is bounded. Consider

$$\begin{array}{rcl} 0 & \leq & (ta-e)^T(f(ta)-f(e))/t \\ & = & a^Tf(ta) - e^Tf(ta)/t - a^Tf(e) + e^Tf(e)/t \\ & < & -e^Tf(ta)/t - a^Tf(e) + e^Tf(e)/t. \end{array}$$

Taking as a limit $t \to \infty$ from both sides, we have

$$e^Tf^\infty(a) \leq -a^Tf(e).$$

Thus, $f^\infty(a) \geq 0$ is bounded. Again, we have $a^Tf(ta) \leq -\xi$ from (8.43) and $a^Tf(ta) \geq a^Tf(a)$ from (8.44). Thus, $\lim a^Tf(ta)$ is bounded. To summarize, $(HMCP)$ has an asymptotical solution ($x^* = a, \tau^* = 0, s^* = f^\infty(a), \kappa^* = \lim -a^Tf(ta) \geq \xi$).

Conversely, if there is a bounded sequence ($x^k > 0, \tau^k > 0, s^k > 0, \kappa^k > 0$), then

$$\lim s^k = \lim \tau^k f(x^k/\tau^k) \geq 0, \quad \lim \kappa^k = \lim -(x^k)^Tf(x^k/\tau^k) \geq \xi > 0.$$

Then, we claim that there is no feasible point ($x \geq 0, s \geq 0$) such that $s - f(x) = 0$. We prove this fact by contradiction. If there is one, then

$$\begin{array}{rcl} 0 & \leq & ((x^k;\tau^k) - (x;1))^T(\psi(x^k,\tau^k) - \psi(x,1)) \\ & = & (x^k-x)^T(\tau^kf(x^k/\tau^k) - f(x)) + (\tau^k-1)^T(xf(x) - (x^k)^Tf(x^k/\tau^k)). \end{array}$$

Therefore,

$$(x^k)^Tf(x^k/\tau^k) \geq (x^k)^Tf(x) + \tau^kx^Tf(x^k/\tau^k) - \tau^kx^Tf(x).$$

Since the first two terms at the right-hand side are positive and $\lim \tau^k = 0$, we must have

$$\lim (x^k)^Tf(x^k/\tau^k) \geq 0,$$

which is a contradiction to $\kappa^k = -(x^k)^Tf(x^k/\tau^k) \geq \xi > 0$. Also, any limit of x^k is a separating hyperplane, i.e., a certificate proving infeasibility.

$\square$

8.4.3 The central path

Due to Theorem 8.32, we can solve (MCP) by finding a maximal complementary solution of $(HMCP)$. Select $x^0 > 0$, $s^0 > 0$, $\tau^0 > 0$ and $\kappa^0 > 0$ and let the residual vectors

$$r^0 = s^0 - \tau^0f(x^0/\tau^0), \quad z^0 = \kappa^0 + (x^0)^Tf(x^0/\tau^0).$$

Also let

$$\bar{n} = (r^0)^T x^0 + z^0 \tau^0 = (x^0)^T s^0 + \tau^0 \kappa^0.$$

For simplicity, we set

$$x^0 = e,\ \tau^0 = 1,\ s^0 = e,\ \kappa^0 = 1,\ \theta^0 = 1,$$

with

$$X^0 s^0 = e \quad \text{and} \quad \tau^0 \kappa^0 = 1.$$

Note that $\bar{n} = n + 1$ in this setting.

We present the next theorem.

Theorem 8.33 . *Consider $(HMCP)$.*

i. *For any $0 < \theta \leq 1$, there exists a strictly positive point $(x > 0, \tau > 0, s > 0, \kappa > 0)$ such that*

$$\begin{pmatrix} s \\ \kappa \end{pmatrix} - \psi(x,\tau) = \begin{pmatrix} s - \tau f(x/\tau) \\ \kappa + x^T f(x/\tau) \end{pmatrix} = \theta \begin{pmatrix} r^0 \\ z^0 \end{pmatrix}. \tag{8.46}$$

ii. *Starting from $(x^0 = e, \tau^0 = 1, s^0 = e, \kappa^0 = 1)$, for any $0 < \theta \leq 1$ there is a unique strictly positive point $(x(\theta), \tau(\theta), s(\theta), \kappa(\theta))$ that satisfies equation (8.46) and*

$$\begin{pmatrix} Xz \\ \tau\kappa \end{pmatrix} = \theta e. \tag{8.47}$$

iii. *For any $0 < \theta \leq 1$, the solution $(x(\theta), \tau(\theta), s(\theta), \kappa(\theta))$ in [ii] is bounded. Thus,*

$$\mathcal{C}(\theta) := \left\{ (x,\tau,s,\kappa) : \begin{pmatrix} s \\ \kappa \end{pmatrix} - \psi(x,\tau) = \theta \begin{pmatrix} r^0 \\ z^0 \end{pmatrix},\ \begin{pmatrix} Xz \\ \tau\kappa \end{pmatrix} = \theta e \right\} \tag{8.48}$$

for $0 < \theta \leq 1$ is a continuous bounded trajectory.

iv. *The limit point $(x(0), \tau(0), s(0), \kappa(0))$ is a maximal complementary solution for $(HMCP)$.*

Proof. We prove [i]. Again, the set

$$H_{++} := \left\{ \begin{pmatrix} s \\ \kappa \end{pmatrix} - \psi(x,\tau) : \quad (x,\tau,s,\kappa) > 0. \right\}$$

is open and convex. We have $(r^0; z^0) \in H_{++}$ by construction. On the other hand, $0 \in \bar{H}_{++}$ from Theorem 8.32. Thus,

$$\theta \begin{pmatrix} r^0 \\ z^0 \end{pmatrix} \in H_{++}.$$

The proof of [ii] is due to Theorem 8.26.

We now prove [iii]. Again, the existence is due to Theorem 8.26. We prove the boundedness. Assume $(x, \tau, s, \kappa) \in \mathcal{C}(\theta)$ then

$$\begin{array}{rl}
\multicolumn{2}{l}{(x;\tau)^T(r^0;z^0)} \\
= & (x;\tau)^T(s^0;\kappa^0) - (x;\tau)^T\psi(x^0;\tau^0) \\
= & (x;\tau)^T(s^0;\kappa^0) + (s;\kappa)^T(x^0;\tau^0) - (s;\kappa)^T(x^0;\tau^0) - (x;\tau)^T\psi(x^0;\tau^0) \\
= & (x;\tau)^T(s^0;\kappa^0) + (s;\kappa)^T(x^0;\tau^0) \\
 & -(x^0;\tau^0)^T(\theta(r^0;z^0) + \psi(x,\tau)) - (x;\tau)^T\psi(x^0;\tau^0) \\
= & (x;\tau)^T(s^0;\kappa^0) + (s;\kappa)^T(x^0;\tau^0) \\
 & -\theta(x^0;\tau^0)^T(r^0;z^0) - (x^0;\tau^0)^T\psi(x,\tau) - (x;\tau)^T\psi(x^0;\tau^0) \\
\geq & (x;\tau)^T(s^0;\kappa^0) + (s;\kappa)^T(x^0;\tau^0) \\
 & -\theta(x^0;\tau^0)^T(r^0;z^0) - (x;\tau)^T\psi(x,\tau) - (x^0;\tau^0)^T\psi(x^0;\tau^0) \\
= & (x;\tau)^T(s^0;\kappa^0) + (s;\kappa)^T(x^0;\tau^0) - \theta(x^0;\tau^0)^T(r^0;z^0) \\
= & (x;\tau)^T(s^0;\kappa^0) + (s;\kappa)^T(x^0;\tau^0) - \theta(x^0;\tau^0)^T((s^0;\kappa^0) - \psi(x^0,\tau^0)) \\
= & (x;\tau)^T(s^0;\kappa^0) + (s;\kappa)^T(x^0;\tau^0) - \theta(x^0;\tau^0)^T(s^0;\kappa^0).
\end{array}$$

Also for $0 < \theta \leq 1$,

$$\begin{aligned}
\theta(x;\tau)^T(r^0;z^0) = (x;\tau)^T((s;\kappa) - \psi(x,\tau)) &= (x;\tau)^T(s;\kappa) \\
&= \theta(n+1) = \theta(x^0;\tau^0)^T(s^0;\kappa^0).
\end{aligned}$$

From the above two relations, we have

$$(x;\tau)^T(s^0;\kappa^0) + (s;\kappa)^T(x^0;\tau^0) \leq (1+\theta)(x^0;\tau^0)^T(s^0;\kappa^0).$$

Thus, $(x;\tau;s;\kappa)$ is bounded.

Finally, we prove (iv). Let $(x^*, \tau^*, s^*, \kappa^*)$ be any maximal complementarity solution for $(HMCP)$ such that

$$(s^*;\kappa^*) = \psi(x^*;\tau^*) \quad \text{and} \quad (x^*)^T s^* + \tau^*\kappa^* = 0,$$

and it is normalized by

$$(r^0;z^0)^T(x^*;\tau^*) = (r^0;z^0)^T(x^0;\tau^0) = (s^0;\kappa^0)^T(x^0;\tau^0) = (n+1).$$

For any $0 < \theta \leq 1$, let (x, τ, s, κ) be the solution on the path. Then, we have

$$\begin{array}{rl}
\multicolumn{2}{l}{((x;\tau) - (x^*;\tau^*))^T((s;\kappa) - (s^*;\kappa^*))} \\
= & ((x;\tau) - (x^*;\tau^*))^T(\psi(x;\tau) - \psi(x^*;\tau^*)) + \theta(r^0;z^0)^T((x;\tau) - (x^*;\tau^*)) \\
\geq & \theta(r^0;z^0)^T((x;\tau) - (x^*;\tau^*)).
\end{array}$$

Therefore,

$$
\begin{array}{rl}
\multicolumn{2}{l}{(x;\tau)^T(s^*;\kappa^*) + (s;\kappa)^T(x^*;\tau^*)} \\
\leq & (x;\tau)^T(s;\kappa) - \theta(r^0;z^0)^T((x;\tau) - (x^*;\tau^*)) \\
= & (x;\tau)^T(s;\kappa) - (x;\tau)^T(s;\kappa) + \theta(r^0;z^0)^T(x^*;\tau^*) \\
= & \theta(r^0;z^0)^T(x^*;\tau^*) \\
= & \theta(n+1).
\end{array}
$$

Using $x_j s_j = \theta$ we obtain,

$$
\begin{array}{rl}
\multicolumn{2}{l}{(x;\tau)^T(s^*;\kappa^*) + (s;\kappa)^T(x^*;\tau^*)} \\
= & \theta \sum \frac{s_j^*}{s_j} + \frac{\kappa^*}{\tau} + \sum \frac{x_j^*}{x_j} + \frac{\tau^*}{\kappa} \\
\leq & \theta(n+1).
\end{array}
$$

Thus, we have

$$\frac{s_j^*}{s_j} \leq (n+1), \quad \text{and} \quad \frac{\kappa^*}{\kappa} \leq (n+1)$$

and

$$\frac{x_j^*}{x_j} \leq (n+1), \quad \text{and} \quad \frac{\tau^*}{\tau} \leq (n+1).$$

Thus, the limit point, $(x(0), \tau(0), s(0), \kappa(0))$, is a maximal complementarity solution for $(HMCP)$.

$\square$

We now present an interior-point algorithm that generates iterates within a neighborhood of $\mathcal{C}(\theta)$. For simplicity, in what follows we let $x := (x;\tau) \in \mathcal{R}^{n+1}$, $s := (s;\kappa) \in \mathcal{R}^{n+1}$, and $r^0 := (r^0; z^0)$. Recall that, for any $x, s > 0$

$$x^T\psi(x) = 0 \quad \text{and} \quad x^T\nabla\psi(x) = -\psi(x)^T. \tag{8.49}$$

Furthermore, ψ is monotone and satisfies the scaled Lipschitz. We will use these facts frequently in our analyses.

8.4.4 An interior-point algorithm

At iteration k with iterate $(x^k, s^k) > 0$, the algorithm solves a system of linear equations for direction (d_x, d_s) from

$$d_s - \nabla\psi(x^k)d_x = -\eta r^k \tag{8.50}$$

and

$$X^k d_s + S^k d_x \quad = \quad \gamma\mu^k e - X^k s^k, \tag{8.51}$$

where η and γ are proper given parameters between 0 and 1, and

$$r^k = s^k - \psi(x^k) \quad \text{and} \quad \mu^k = \frac{(x^k)^T s^k}{n+1}.$$

First we prove the following lemma.

Lemma 8.34 . *The direction (d_x, d_s) satisfies*

$$d_x^T d_s = d_x^T \nabla\psi(x^k) d_x + \eta(1-\eta-\gamma)(n+1)\mu^k.$$

Proof. Premultiplying each side of (8.50) by d_x^T gives

$$d_x^T d_s - d_x^T \nabla\psi(x^k) d_x = -\eta d_x^T (s^k - \psi(x^k)). \tag{8.52}$$

Multiplying each side of (8.50) by x^k and using (8.49) give

$$\begin{array}{rcl} (x^k)^T d_s + \psi(x^k) d_x & = & -\eta(x^k)^T r^k \\ & = & -\eta(x^k)^T (s^k - \psi(x^k)) \\ & = & -\eta(x^k)^T s^k \\ & = & -\eta(n+1)\mu^k. \end{array} \tag{8.53}$$

These two equalities in combination with (8.51) imply

$$\begin{array}{rcl} d_x^T d_s & = & d_x^T \nabla\psi(x^k) d_x - \eta(d_x^T s^k + d_s^T x^k + \eta(n+1)\mu^k) \\ & = & d_x^T \nabla\psi(x^k) d_x - \eta(-(n+1)\mu^k + \gamma(n+1)\mu^k + \eta(n+1)\mu^k) \\ & = & d_x^T \nabla\psi(x^k) d_x + \eta(1-\gamma-\eta)(n+1)\mu^k. \end{array}$$

□

For a step-size $\alpha > 0$, let the new iterate

$$x^+ := x^k + \alpha d_x > 0, \tag{8.54}$$

and

$$\begin{array}{rcl} s^+ & := & s^k + \alpha d_s + \psi(x^+) - \psi(x^k) - \alpha\nabla\psi(x^k) d_x \\ & = & \psi(x^+) + (s^k - \psi(x^k)) + \alpha(d_s - \nabla\psi(x^k) d_x) \\ & = & \psi(x^+) + (s^k - \psi(x^k)) - \alpha\eta(s^k - \psi(x^k)) \\ & = & \psi(x^+) + (1-\alpha\eta) r^k. \end{array} \tag{8.55}$$

The last two equalities come from (8.50) and the definition of r^k. Also let

$$r^+ = s^+ - \psi(x^+).$$

Then, we have

Lemma 8.35 . *Consider the new iterate (x^+, s^+) given by (8.54) and (8.55).*

a). $r^+ = (1-\alpha\eta)r^k$

b). $(x^+)s^+ = (x^k)^T s^k(1-\alpha(1-\gamma)) + \alpha^2\eta(1-\eta-\gamma)(n+1)\mu^k$

Proof. From (8.55)

$$\begin{array}{rcl} r^+ & = & s^+ - \psi(x^+) \\ & = & (1-\alpha\eta)r^k. \end{array}$$

Next we prove b). Using (8.49), (8.51), and Lemma 8.34, we have

$$\begin{array}{rl} (x^+)^T s^+ & \\ = & (x^+)^T(s^k + \alpha d_s + \psi(x^+) - \psi(x^k) - \alpha\nabla\psi(x^k)d_x) \\ = & (x^+)^T(s^k + \alpha d_s) - (x^+)^T(\psi(x^k) + \alpha\nabla\psi(x^k)d_x) \\ = & (x^+)^T(s^k + \alpha d_s) - (x^k + \alpha d_x)^T(\psi(x^k) + \alpha\nabla\psi(x^k)d_x) \\ = & (x^+)^T(s^k + \alpha d_s) - \alpha(x^k)^T\nabla\psi(x^k)d_x - \alpha d_x^T\psi(x^k) - \alpha^2 d_x^T\nabla\psi(x^k)d_x \\ = & (x^+)^T(s^k + \alpha d_s) - \alpha^2 d_x^T\nabla\psi(x^k)d_x \\ = & (x^k + \alpha d_x)^T(s^k + \alpha d_s) - \alpha^2 d_x^T\nabla\psi(x^k)d_x \\ = & (x^k)^T s^k + \alpha(d_x^T s^k + d_s^T x^k) + \alpha^2(d_x^T d_s - d_x^T\nabla\psi(x^k)d_x) \\ = & (x^k)^T s^k + \alpha(d_x^T s^k + d_s^T x^k) + \alpha^2\eta(1-\eta-\gamma)(n+1)\mu^k \\ = & (1-\alpha(1-\gamma))(x^k)^T s^k + \alpha^2\eta(1-\eta-\gamma)(n+1)\mu^k. \end{array}$$

□

This lemma shows that, for setting $\eta = 1-\gamma$, the infeasibility residual and the complementarity gap are reduced at exactly the same rate, as in the homogeneous linear programming algorithm. Now we prove the following.

Theorem 8.36 . *Assume that ψ is scaled Lipschitz with $v = v_\psi$ and at iteration k*

$$\|X^k s^k - \mu^k e\| \le \beta\mu^k, \quad \mu^k = \frac{(x^k)^T s^k}{n+1}$$

where

$$\beta = \frac{1}{3 + 4v_\psi(\sqrt{2}/2)} \le 1/3.$$

Furthermore, let $\eta = \beta/\sqrt{n+1}$, $\gamma = 1-\eta$, and $\alpha = 1$ in the algorithm. Then, the new iterate

$$x^+ > 0, \quad s^+ = \psi(x^+) + (1-\eta)r^k > 0,$$

and

$$\|X^+ s^+ - \mu^+ e\| \le \beta\mu^+, \quad \mu^+ = \frac{(x^+)^T s^+}{n+1}.$$

Proof. It follows from Lemma 8.35 that $\mu^+ = \gamma\mu^k$. From (8.51) we further have

$$S^k d_x + X^k d_s = -X^k s^k + \mu^+ e.$$

Hence,

$$D^{-1}d_x + Dd_s = -(X^k S^k)^{-1/2}(X^k s^k - \mu^+ e)$$

where $D = (X^k)^{1/2}(S^k)^{-1/2}$. Note

$$d_x^T d_s = d_x^T \nabla\psi(x^k) d_x \geq 0$$

from Lemma 8.34 and $\gamma = 1 - \eta$. This together with the assumption of the theorem imply

$$\|D^{-1}d_x\|^2 + \|Dd_s\|^2 \leq \|(X^k S^k)^{-1/2}(X^k s^k - \mu^+ e)\|^2 \leq \frac{\left\|X^k s^k - \mu^+ e\right\|^2}{(1-\beta)\mu^k}.$$

Also note

$$\begin{array}{rcl}\left\|X^k s^k - \mu^+ e\right\|^2 & = & \left\|X^k s^k - \mu^k e + (1-\gamma)\mu^k e\right\|^2 \\ & = & \left\|X^k s - \mu^k e\right\|^2 + ((1-\gamma)\mu^k)^2 \left\|e\right\|^2 \\ & \leq & (\beta^2 + \eta^2(n+1))(\mu^k)^2 \\ & = & 2\beta^2(\mu^k)^2.\end{array}$$

Thus,

$$\begin{array}{c}\left\|(X^k)^{-1}d_x\right\| = \left\|(X^k S^k)^{-1/2} D^{-1} d_x\right\| \leq \dfrac{\left\|D^{-1}d_x\right\|}{\sqrt{(1-\beta)\mu^k}} \\ \leq \dfrac{\sqrt{2}\beta\mu^k}{(1-\beta)\mu^k} = \dfrac{\sqrt{2}\beta}{1-\beta} \leq \dfrac{\sqrt{2}}{2},\end{array}$$

since $\beta \leq 1/3$. This implies that $x^+ = x^k + d_x > 0$. Furthermore, we have

$$\begin{array}{rcl}\|D_x d_s\| & = & \left\|D^{-1} D_x D d_s\right\| \\ & \leq & \left\|D^{-1}d_x\right\| \|Dd_s\| \\ & \leq & (\left\|D^{-1}d_x\right\|^2 + \|Dd_s\|^2)/2 \\ & \leq & \left\|(X^k S^k)^{-1/2}(X^k s^k - \mu^+)\right\|^2/2 \\ & \leq & \dfrac{\left\|X^k s^k - \mu^+ e\right\|^2}{2(1-\beta)\mu^k} \\ & \leq & \dfrac{2\beta^2(\mu^k)^2}{2(1-\beta)\mu^k} \\ & = & \dfrac{\beta^2\mu^k}{1-\beta},\end{array}$$

and

$$d_x^T d_s = d_x^T D^{-1} D d_s \le \left\| D^{-1} d_x \right\| \|D d_s\| \le \frac{\beta^2 \mu^k}{1-\beta}.$$

Consider

$$\begin{aligned}
& X^+ s^+ - \mu^+ e \\
= \;& X^+(s^k + d_s + \psi(x^+) - \psi(x^k) - \nabla\psi(x^k) d_x) - \mu^+ e \\
= \;& (X^k + D_x)(s^k + d_s) - \mu^+ e + X^+(\psi(x^+) - \psi(x^k) - \nabla\psi(x^k) d_x) \\
= \;& D_x d_s + X^+(\psi(x^+) - \psi(x^k) - \nabla\psi(x^k) d_x).
\end{aligned}$$

Using that ψ is the scaled Lipschitz, $d_x^T \nabla\psi(x^k) d_x = d_x^T d_s$ and the above four relations we obtain

$$\begin{aligned}
& \left\| X^+ s^+ - \mu^+ e \right\| \\
= \;& \left\| D_x d_s + X^+(\psi(x^+) - \psi(x^k) - \nabla\psi(x^k) d_x) \right\| \\
= \;& \left\| D_x d_s + (X^k)^{-1} X^+ X^k (\psi(x^+) - \psi(x^k) - \nabla\psi(x^k) d_x) \right\| \\
\le \;& \|D_x d_s\| + \left\| (X^k)^{-1} X^+ \right\|_\infty \left\| X^k (\psi(x^+) - \psi(x^k) - \nabla\psi(x^k) d_x) \right\|_1 \\
\le \;& \|D_x d_s\| + 2 \left\| X^k (\psi(x^+) - \psi(x^k) - \nabla\psi(x^k) d_x) \right\|_1 \\
\le \;& \|D_x d_s\| + 2 \upsilon_\psi(\sqrt{2}/2) d_x^T \nabla\psi(x^k) d_x \\
= \;& \|D_x d_s\| + 2 \upsilon_\psi(\sqrt{2}/2) d_x^T d_s \\
\le \;& \frac{\beta^2 \mu^k}{1-\beta} + 2 \upsilon_\psi(\sqrt{2}/2) \frac{\beta^2 \mu^k}{1-\beta} \\
\le \;& \frac{(1 + 2\upsilon_\psi(\sqrt{2}/2)) \beta^2 \mu^k}{1-\beta}.
\end{aligned}$$

Finally, $\beta = 1/(3 + 4\upsilon_\psi(\sqrt{2}/2))$ implies that

$$\frac{(1 + 2\upsilon_\psi(\sqrt{2}/2)) \beta^2 \mu^k}{1-\beta} \le \frac{\beta}{2}$$

and

$$\left\| X^+ s^+ - \mu^+ e \right\| \le \beta \mu^k / 2 < \beta \gamma \mu^k = \beta \mu^+.$$

It is easy to verify that $x^+ > 0$ and $\|X^+ s^+ - \mu^+ e\| < \beta \mu^+$ implies $s^+ > 0$.

□

The above theorem shows that the homogeneous algorithm will generate a sequence $(x^k, s^k) > 0$ with $(x^{k+1}, s^{k+1}) := (x^+, s^+)$ such that $s^k = \psi(x^k) + r^k$ and $\|X^k s^k - \mu^k\| \le \beta \mu^k$, where both $\left\| r^k \right\|$ and $(x^k)^T s^k$

converge to zero at a global rate $\gamma = 1-\beta/\sqrt{n+1}$. We see that if $\upsilon_\psi(\sqrt{2}/2)$ is a constant, or $\upsilon_f(2/(1+\sqrt{2}))$ is a constant in (MCP) due to (iii) of Theorem 8.32, then it results in an $O(\sqrt{n}\log(1/\epsilon))$ iteration algorithm with error ϵ. It generates a maximal solution for $(HMCP)$, which is either a solution or a certificate proving infeasibility for (MCP), due to (v) and (vi) of Theorem 8.32.

One more comment is that our results should hold for the case where $f(x)$ is a continuous *monotone* mapping from $\overset{\circ}{\mathcal{R}}{}^n_+$ to $\mathcal{R}^n$. In other words, $f(x)$ may not exist at the boundary of $\mathcal{R}^n_+$. In general, many convex optimization problems can be solved, either obtaining a solution, or proving infeasibility or unboundedness, by solving their KKT system which is a monotone complementarity problem.

8.5 Notes

For the NEG problem, see Gale [100], Kemeny, Morgenstern and Thompson [175], and Robinson [290]. A recent description of the problem can be found in Schrijver [298], and Tan and Freund [315]. It is easy to show that Assumption 8.3 implies Assumption 8.4, see Theorem 2 of Robinson [290].

The method described here is based on the paper of Ye [381]. Similar methods for the fractional programming over polyhedral and nonpolyhedral cones were developed by Boyd and Ghaoui [61], Freund and Jarre [95, 94], Nesterov and Nemirovskii [264], and Nemirovskii [257].

Various centers were considered for the center-section method as we mentioned earlier. Goffin, Haurie and Vial [109], Sonnevend [306], and Ye [377] were among the first to propose the analytic center-section or cutting plane method. Its complexity issues were addressed by Atkinson and Vaidya [31], Goffin, Luo and Ye [110, 111], and Nesterov [261]. In particular, Atkinson and Vaidya developed a scheme to delete unnecessary inequalities and managed to prove a polynomial analytic center-section algorithm. The analytic center-section method was used and tested for a variety of large scale problems, where they performed quite well, see, for example, Bahn, Goffin, Vial and Merle [32], Bahn, Merle, Goffin and Vial [33], and Mitchell [225, 226].

The primal potential reduction algorithm for positive semi-definite programming is due to Alizadeh [9, 8], in which Ye has "suggested studying the primal-dual potential function for this problem" and "looking at symmetric preserving scalings of the form $X_0^{-1/2}XX_0^{-1/2}$," and to Nesterov and Nemirovskii [263], and the primal-dual algorithm described here is due to Nesterov and Todd [265, 266]. One can also develop a dual potential reduction algorithm. In general, consider

$$\begin{array}{lll} (PSP) & \inf & C \bullet X \\ & \text{subject to} & A \bullet X = b,\ X \in K, \end{array}$$

and its dual

$$\begin{array}{lll} (PSD) & \sup & b^T y \\ & \text{subject to} & A^* \bullet Y + S = C,\ S \in K, \end{array}$$

where K is a convex homogeneous cone.

Interior-point algorithms compute a search direction (d_X, d_Y, d_S) and a new strictly feasible primal-dual pair X^+ and $(Y^+; S^+)$ is generated from

$$X^+ = X + \alpha d_X,\ Y^+ = Y + \beta d_Y,\ S^+ = S + \beta d_S,$$

for some step-sizes α and β.

The search direction (d_X, d_Y, d_S) is determined by the following equations.

$$A \bullet d_X = 0, \quad d_S = -A^* \bullet d_Y \quad \text{(feasibility)} \tag{8.56}$$

and

$$d_X + F''(S) d_S = -\frac{n+\rho}{X \bullet S} X - F'(S) \quad \text{(dual scaling)}, \tag{8.57}$$

or

$$d_S + F''(X) d_X = -\frac{n+\rho}{X \bullet S} S - F'(X) \quad \text{(primal scaling)}, \tag{8.58}$$

or

$$d_S + F''(Z) d_X = -\frac{n+\rho}{X \bullet S} S - F'(X) \quad \text{(joint scaling)}, \tag{8.59}$$

where Z is chosen to satisfy

$$S = F''(Z) X. \tag{8.60}$$

The differences among the three algorithms are the computation of the search direction and their theoretical close-form step-sizes. All three generate an ϵ-optimal solution (X, Y, S), i.e.,

$$X \bullet S \leq \epsilon$$

in a guaranteed polynomial time.

Other primal-dual algorithms for positive semi-definite programming are in Alizadeh, Haeberly and Overton [10, 11], Boyd, Ghaoui, Feron and Balakrishnan [62], Helmberg, Rendl, Vanderbei and Wolkowicz [144], Jarre

[161], Kojima, Shindoh and Hara [186], Vandenberghe and Boyd [351, 352], and references therein. Efficient interior-point algorithms are also developed for optimization over the second-order cone, see Andersen and Christiansen [16] and Xue and Ye [370]. These algorithms have established the best approximation complexity results for some combinatorial problems.

The scaled Lipschitz condition used in Section 8.4 was proposed by Kortanek and Zhu [190] for linearly constrained convex minimization, related to the work of Monteiro and Adler [241], and later extended by Potra and Ye [282] for the monotone complementary problem. This condition is included in a more general condition analyzed by Nesterov and Nemirovskii [263], den Hertog [145], den Hertog, Jarre, Roos and Terlaky [146], and Jarre [162].

Results in Section 8.4.1 are based on Kojima, Megiddo and Mizuno [179]. A similar augmented transformation in Section 8.4.2 has been discussed in Ye and Tse [387] and it is closely related to the recession function in convex analyses of Rockafellar [291]. All other results in Section 8.4 are based on Andersen and Ye [14]. Interior-point algorithms for convex programming include: Abhyankar, Morin and Trafalis [1] for multiple objective optimization, Anstreicher, den Hertog, Roos and Terlaky [24], Ben–Daya and Shetty [42], Bonnans and Bouhtou [58], Carpenter, Lustig, Mulvey and Shanno [63], Goldfarb and Liu [114], Jarre [160], Kapoor and Vaidya [172], Mehrotra and Sun [222], Pardalos, Ye and Han [275], Ponceleon [278], Ye [374], Ye and Tse [387], etc. for quadratic programming; Ding and Li [80], Güler [137], Harker and Xiao [142] Ji, Potra and Huang [163], Polak, Higgins and Mayne [276], Shanno and Simantiraki [300], Sun and Zhao [314], Tseng [337], etc. for the monotone complementarity problem; Ben–Tal and Nemirovskii [43], Faybusovich [89], Goldfarb and Scheinberg [115], Güler [138], Güler and Tuncel [139], Luo, Sturm and Zhang [199], Monteiro and Pang [244], Nesterov, Todd and Ye [267], Ramana, Tuncel and Wolkowicz [292], Vandenberghe, Boyd, and Wu [353], etc. for nonpolyhedral optimization; Anstreicher and Vial [27], Coleman and Li [70] Güler [138], den Hertog, Roos and Terlaky [148], Kortanek, Potra and Ye [187], Mehrotra and Sun [223], Monteiro [238], Nash and Sofer [252], Potra and Ye [281], Sun and Qi [313], Tanabe [154], Wang, Monteiro, and Pang [364], Zhang [388], etc. for nonlinear programming; Asic and Kovacevic-Vujcic [28], Ferris and Philpott [90], Todd [323], etc. for semi-infinite programming; Birge and Holmes [48], Birge and Qi [49], etc. for stochastic programming.

Applications, decompositions, inexact iteration, and special data structures of interior-point algorithms were described in Bixby, Gregory, Lustig, Marsten and Shanno [50], Choi and Goldfarb [65], Christiansen and Kortanek [68], Gondzio [120], Han, Pardalos and Ye [141], Ito, Kelley and Sachs [157], Kaliski [169], Pardalos, Ye and Han [275], Ponnambalam, Vannelli and Woo [279], Resende and Veiga [289], Tone [333], Wallacher and Zimmermann [363], etc.

8.6 Exercises

8.1 *Show that Assumption 8.3 implies Assumption 8.4.*

8.2 *Prove Lemma 8.6.*

8.3 *Prove Proposition 8.5.*

8.4 *Prove Lemma 8.14*

8.5 *Prove Corollary 8.19.*

8.6 *Prove Lemma 8.23.*

8.7 *Describe and analyze a dual potential algorithm for positive semi-definite programming in the standard form.*

8.8 *If* (MCP) *has a solution, then the solution set is convex and it contains a maximal solution* (x^*, s^*) *where the number positive components in* (x^*, s^*) *is maximal. Moreover, the indices of those positive components are invariant among all maximal solutions for* (MCP).

8.9 *Prove Theorem 8.26.*

8.10 *Prove Lemma 8.27.*

8.11 *Prove Lemma 8.29.*

8.12 *Prove (ii) and (iv) of Theorem 8.32.*

Chapter 9

Nonconvex Optimization

The aim of this chapter is to describe some results in interior-point algorithms for solving "hard" problems, such as the nonmonotone linear complementarity problem (LCP) and the quadratic programming (QP) problem, and to suggest some directions in which future progress might be made. These problems play important roles in optimization theory. In one sense they are continuous optimization and fundamental sub-problems for general nonlinear programming, but they are also considered the most challenging combinatorial optimization problems.

9.1 Linear Complementarity Problem

In this section, we are concerned with the LCP described in Section 1.3.8, where we assume, without loss of generality, that

$$\overset{\circ}{\mathcal{F}} = \{(x,s) : s = Mx + q,\ x > 0 \text{ and } s > 0\}$$

is nonempty. We also use $\mathcal{F}$ to denote the "feasible region," i.e.,

$$\mathcal{F} = \{(x,s) : s = Mx + q,\ x \geq 0 \text{ and } s \geq 0\}.$$

Thus, the LCP can be viewed as an optimization problem

$$\begin{array}{lll} (LCP) & \text{minimize} & x^T s \\ & \text{subject to} & (x,s) \in \mathcal{F}. \end{array}$$

Similar to solving the LP problem, we will describe a "condition-based" iteration complexity bound for solving the LCP. This condition number characterizes the degree of difficulty of the LCP solution when a potential reduction algorithm is used. We show how the condition number depends on the data (M, q).

9.1.1 Potential reduction algorithm

We again use the same potential function

$$\psi(x,s) = \psi_\rho(x,s) := \rho \ln(x^T s) - \sum_{j=1}^{n} \log(x_j s_j),$$

with $\rho \geq n$ as described in Chapter 4 (ρ here corresponds to $n+\rho$ there), for an interior feasible point (x,s). Starting from an interior point (x^0, s^0) with

$$\psi(x^0, s^0) =: \psi^0,$$

((x^0, s^0) can be chosen as an approximate analytic center of $\mathcal{F}$, see Chapter 4), the potential reduction algorithm generates a sequence of interior feasible points $\{x^k, s^k\}$ terminating at a point such that

$$\psi(x^k, s^k) \leq (\rho - n)\log \epsilon + n \log n.$$

From the arithmetic-geometric mean inequality,

$$n \ln((x^k)^T s^k) - \sum_{j=1}^{n} \log(x_j^k s_j^k) \geq n \ln(n) \geq 0.$$

Thus, we must have

$$(x^k)^T s^k \leq \epsilon.$$

To achieve a potential reduction, we again use the scaled gradient projection method. The gradient vector of the potential function with respect to x is

$$\nabla \psi_x = \frac{\rho}{\Delta} s - X^{-1} e$$

and the one to s is

$$\nabla \psi_s = \frac{\rho}{\Delta} x - S^{-1} e,$$

where $\Delta = x^T s$. Now, we solve, at the kth iteration, the following linear program subject to an ellipsoid constraint:

$$\begin{array}{lll} (EP) & \text{minimize} & \nabla^T \psi_{x^k} d_x + \nabla^T \psi_{s^k} d_s \\ & \text{subject to} & d_s = M d_x \\ & & \|(X^k)^{-1} d_x\|^2 + \|(S^k)^{-1} d_s\|^2 \leq \beta^2 < 1. \end{array}$$

Denote by $\bar{d}_x$ and $\bar{d}_s$ the minimal solution for (EP). Then, we have

$$\begin{pmatrix} (X^k)^{-1} \bar{d}_x \\ (S^k)^{-1} \bar{d}_s \end{pmatrix} = -\beta \frac{p^k}{\|p^k\|}, \tag{9.1}$$

where

$$p^k = \begin{pmatrix} p_x^k \\ p_s^k \end{pmatrix} = \begin{pmatrix} \frac{\rho}{\Delta^k} X^k(s^k + M^T\pi) - e \\ \frac{\rho}{\Delta^k} S^k(x^k - \pi) - e \end{pmatrix}, \tag{9.2}$$

and

$$\pi = ((S^k)^2 + M(X^k)^2 M^T)^{-1}(S^k - MX^k)(X^k s^k - \frac{\Delta^k}{\rho} e). \tag{9.3}$$

From the concavity of *log* function and Lemma 3.1,

$$\psi(x^k + \bar{d}_x, s^k + \bar{d}_s) - \psi(x^k, s^k) \le -\beta\|p^k\| + \frac{\beta^2}{2}(\rho + \frac{1}{1-\beta}). \tag{9.4}$$

Let

$$\beta = \min(\frac{\|p^k\|}{\rho+2}, \frac{1}{\rho+2}) \le 1/2. \tag{9.5}$$

Then we have

$$\psi(x^k + \bar{d}_x, s^k + \bar{d}_s) - \psi(x^k, s^k) \le -\min(\frac{\|p^k\|^2}{2(\rho+2)}, \frac{1}{2(\rho+2)}). \tag{9.6}$$

The algorithm can be described as follows:

Algorithm 9.1 *Given $x^0, s^0 > 0$ and $s^0 = Mx^0 + q$ and $k := 0$.*
While $(x^k)^T s^k \ge \epsilon$ **do**

1. *Compute π of (9.3) and p^k of (9.2), and select β of (9.5); construct $\bar{d}_x$ and $\bar{d}_s$ of (9.1).*
2. *Let $x^{k+1} = x^k + \Delta\bar{x}$ and $s^{k+1} = s^k + \Delta\bar{s}$.*
3. *Set $k := k + 1$ and return to Step 1.*

Clearly from inequality (9.6), $\|p^k\|^2$ can be used to measure the potential reduction at the kth iteration of the potential reduction algorithm. For any $x, s \in \overset{\circ}{\mathcal{R}}{}^n_+$, let

$$g(x, s) = \frac{\rho}{\Delta} Xs - e$$

and

$$H(x, s) = 2I - (XM^T - S)(S^2 + MX^2M^T)^{-1}(MX - S).$$

Note that $H(x, s)$ is positive semi-definite (PSD), and

$$\|p^k\|^2 = g^T(x^k, s^k)H(x^k, s^k)g(x^k, s^k). \tag{9.7}$$

Let us use $\|g(x,s)\|_H^2$ to denote $g^T(x,s)H(x,s)g(x,s)$. Then, we define a condition number for the LCP (M,q) as

$$\gamma(M,q,\epsilon) = \inf\{\|g(x,s)\|_H^2 : x^T s \geq \epsilon,\ \psi(x,s) \leq \psi^0 \text{ and } (x,s) \in \overset{\circ}{\mathcal{F}}\}. \quad (9.8)$$

We describe a sequence of propositions for $\gamma(M,q,\epsilon)$.

Proposition 9.1 *Let $\rho \geq 2n$. Then, for M being a diagonal and PSD matrix, and any $q \in \mathcal{R}^n$,*

$$\gamma(M,q,\epsilon) \geq n.$$

Proof. If M is diagonal and PSD, then the matrix

$$I - (XM^T - S)(S^2 + MX^2M^T)^{-1}(MX - S)$$

is diagonal. It is also PSD since the jth diagonal component is

$$1 - \frac{(M_{jj}x_j - s_j)^2}{s_j^2 + M_{jj}^2 x_j^2} = \frac{2M_{jj}x_j s_j}{s_j^2 + M_{jj}^2 x_j^2} \geq 0.$$

Therefore, for all $(x,s) \in \overset{\circ}{\mathcal{F}}$ and $\rho \geq 2n$,

$$\gamma(M,q,\epsilon) \geq \|g(x,s)\|^2 \geq \frac{(\rho - n)^2}{n} \geq n.$$

□

Proposition 9.2 *Let $\rho \geq 2n + \sqrt{2n}$. Then, for M being a PSD matrix and any $q \in \mathcal{R}^n$,*

$$\gamma(M,q,\epsilon) \geq 1.$$

We leave its proof to the reader.

Definition 9.1 *A matrix M is a P matrix if and only if its every principal submatrix has a positive determinant.*

Proposition 9.3 *Let $\rho \geq 3n + \sqrt{2n}$. Then, for M being a P-matrix and any $q \in \mathcal{R}^n$,*

$$\gamma(M,q,\epsilon) \geq \min(n\theta(M)/|\lambda(M)|, 1)$$

where $\lambda(M)$ is the least eigenvalue of $(M+M^T)/2$, and $\theta(M)$ is the positive P-matrix number of M^T, i.e.,

$$\theta(M) = \min_{x \neq 0}\{\max_j \frac{x_j(M^Tx)_j}{\|x\|^2}\}.$$

We will prove this proposition in the next section.

Definition 9.2 *A matrix M is row-sufficient if and only if for every vector ξ, diag$(\xi)M^T\xi \leq 0$ implies diag$(\xi)M^T\xi = 0$. A matrix M is column-sufficient if and only if M^T is row-sufficient. A matrix M is a sufficient matrix if and only if it is both row- and column-sufficient.*

Note that the class of row-sufficient matrices contains some popular matrices such as PSD and P matrices.

Proposition 9.4 *Let $\rho > n$ and be fixed. Then, for M being a row-sufficient matrix and $\{(x,s) \in \overset{\circ}{\mathcal{F}}: \psi(x,s) \leq \psi^0\}$ being bounded,*

$$\gamma(M,q,\epsilon) > 0.$$

Proof. It is easy to show that for any $(x,s) \in \overset{\circ}{\mathcal{F}}$,

$$\|g(x,s)\|_H^2 > 0.$$

Moreover, for all $(x,s) \in \overset{\circ}{\mathcal{F}}$, $x^Ts \geq \epsilon$ and $\psi(x,s) \leq \psi^0$,

$$\begin{aligned}
&\psi^0 \\
&\geq \psi(x,s) \\
&= \rho\ln(x^Ts) - \sum_{j=1}^n \log(x_js_j) \\
&= (\rho-n+1)\ln(x^Ts) + (n-1)\ln(x^Ts) - \sum_{j\neq i}\log(x_js_j) - \log(x_iy_i) \\
&\geq (\rho-n+1)\ln(x^Ts) + (n-1)\ln(x^Ts - x_iy_i) \\
&\quad - \sum_{j\neq i}\log(x_js_j) - \log(x_is_i) \\
&\geq (\rho-n+1)\ln(x^Ts) + (n-1)\ln(n-1) - \log(x_is_i) \\
&\geq -(\rho-n+1)\ln(1/\epsilon) + (n-1)\ln(n-1) - \log(x_is_i),
\end{aligned}$$

where $i \in \{1,2,\ldots,n\}$. Thus,

$$\ln(x_is_i) \geq -(\rho-n+1)\log(1/\epsilon) + (n-1)\ln(n-1) - \psi^0,$$

that is, x_is_i is bounded away from zero for every i. Since $\{(x,s) \in \overset{\circ}{\mathcal{F}}: \psi(x,s) \leq \psi^0\}$ is bounded, there must exist a positive number $\bar{\epsilon}$, independent of (x,s), such that

$$x_i \geq \bar{\epsilon} \quad \text{and} \quad s_i \geq \bar{\epsilon}, \quad i = 1,2,\ldots,n$$

for all (x,s) such that $x^Ts \geq \epsilon$, $\psi(x,s) \leq \psi^0$ and $(x,s) \in \overset{\circ}{\mathcal{F}}$. Therefore,

$$\begin{aligned}
&\gamma(M,q,\epsilon)\\
&= \inf\{\|g(x,s)\|_H^2 : x^Ts \geq \epsilon,\ \psi(x,s) \leq \psi^0 \text{ and } (x,s) \in \overset{\circ}{\mathcal{F}}\}\\
&\geq \inf\{\|g(x,s)\|_H^2 : x \geq \bar{\epsilon}e, \quad s \geq \bar{\epsilon}e,\ \psi(x,s) \leq \psi^0 \text{ and } (x,s) \in \mathcal{F}\}\\
&> 0.
\end{aligned}$$

The last inequality holds since the inf is taken in a compact set where $\|g(x,s)\|_H^2$ is always positive.

□

Note that $\psi(x,s) \leq \psi^0$ implies that $x^Ts \leq \psi^0/(\rho-n)$. Hence, the boundedness of $\{(x,s) \in \mathcal{F} : x^Ts \leq \psi^0/(\rho-n))\}$ guarantees the boundedness of $\{(x,s) \in\overset{\circ}{\mathcal{F}}: \psi(x,s) \leq \psi^0\}$. We now derive

Theorem 9.5 *The potential reduction algorithm with $\rho = \theta(n) > n$ solves the LCP for which $\gamma(M,q,\epsilon) > 0$ in*

$$O((\psi^0 + (\rho-n)\log(1/\epsilon) - n\log n)/\alpha(\gamma(M,q,\epsilon)))$$

iterations and each iteration solves a system of linear equations in at most $O(n^3)$ operations, where

$$\alpha(\gamma(M,q,\epsilon)) = \min(\frac{\gamma(M,q,\epsilon)}{2(\rho+2)}, \frac{1}{2(\rho+2)}).$$

Proof. Since $\overset{\circ}{\mathcal{F}}$ is nonempty, by solving a linear program in polynomial time, we can find an approximate analytic center (x^0,s^0) of $\mathcal{F}$. Due to (9.4), (9.5) and (9.6) the potential function is reduced by $O(\alpha(\gamma(M,q,\epsilon)))$ at each iteration. Hence, in total $O((\psi^0 + (\rho-n)\log(1/\epsilon) - n\log n)/\alpha(\gamma(M,q,\epsilon)))$ iterations we have $\psi(x^k,s^k) < (\rho-n)\log\epsilon + n\log n$ and $(x^k)^Ts^k < \epsilon$.

□

Corollary 9.6 *An instance (M,q) of the LCP is solvable in polynomial time if $\gamma(M,q,\epsilon) > 0$ and if $1/\gamma(M,q,\epsilon)$ is bounded above by a polynomial in $\log(1/\epsilon)$ and n.*

The condition number $\gamma(M,q,\epsilon)$ represents the degree of difficulty for the potential reduction algorithm in solving the LCP (M,q). The larger the condition number, the easier the LCP problem. We know that some LCPs are very hard, and some are easy. Here, the condition number builds a connection from easy LCPs to hard LCPs. In other words, the degree of difficulty continuously shifts from easy LCPs to hard LCPs.

9.1.2 A class of LCPs

We now further study $\|p^k\|$ by introducing the following lemma.

Lemma 9.7 $\|p^k\| < 1$ *implies*

$$s^k + M^T\pi^k > 0, \qquad x^k - \pi^k > 0$$

and

$$\frac{2n-\sqrt{2n}}{\rho}\Delta^k < \bar{\Delta} < \frac{2n+\sqrt{2n}}{\rho}\Delta^k,$$

where $\bar{\Delta} = (x^k)^T(s^k + M^T\pi^k) + (s^k)^T(x^k - \pi^k)$.

Proof. The proof is by contradiction. Let $\bar{s} = s^k + M^T\pi^k$ and $\bar{x} = x^k - \pi^k$. It is obvious that if $\bar{s} \not> 0$ or $\bar{x} \not> 0$, then

$$\|p^k\|^2 \geq 1.$$

On the other hand, we have

$$\begin{aligned}\|p^k\|^2 &= (\frac{\rho}{\Delta^k})^2 \| \begin{pmatrix} X^k\bar{s} \\ s^k\bar{x} \end{pmatrix} - \frac{\bar{\Delta}}{2n} e\|^2 + \|\frac{\rho\bar{\Delta}}{2n\Delta^k}e - e\|^2 \\ &\geq (\frac{\rho\bar{\Delta}}{2n\Delta^k} - 1)^2 2n.\end{aligned}$$

Hence, the following must be true

$$(\frac{\rho\bar{\Delta}}{2n\Delta^k} - 1)^2 2n < 1,$$

that is,

$$\frac{2n-\sqrt{2n}}{\rho}\Delta^k < \bar{\Delta} < \frac{2n+\sqrt{2n}}{\rho}\Delta^k.$$

$\square$

$\bar{\Delta}$ can be further expressed as

$$\bar{\Delta} = 2\Delta^k - q^T\pi^k.$$

Now let

$$\begin{aligned}&\Sigma^+(M,q) \\ &= \{\pi : x^T s - q^T\pi < 0,\ x - \pi > 0,\ \text{and } s + M^T\pi > 0 \text{ for some } (x,s) \in \overset{\circ}{\mathcal{F}}\}.\end{aligned}$$

Then, we can prove the following propositions.

Proposition 9.8 *Let $\Sigma^+(M,q)$ be empty for an LCP (M,q). Then, for $\rho \geq 2n + \sqrt{2n}$,*

$$\gamma(M,q,\epsilon) \geq 1.$$

Proof. The proof results directly from Lemma 9.7.

$\square$

Proposition 9.9 *Let*

$$\{\pi : x^T s - q^T \pi > 0,\ x - \pi > 0 \text{ and } s + M^T \pi > 0 \text{ for some } (x,s) \in \overset{\circ}{\mathcal{F}}\}$$

be empty for an LCP (M,q). Then, for $n < \rho \leq 2n - \sqrt{2n}$

$$\gamma(M,q,\epsilon) \geq 1.$$

Proof. The proof again results from Lemma 9.7.

$\square$

Now, let

$$\mathcal{G} = \{(M,q) : \overset{\circ}{\mathcal{F}} \text{ is nonempty and } \Sigma^+(M,q) \text{ is empty}\}.$$

It may not be possible in polynomial time to tell if an LCP problem (M,q) is an element of $\mathcal{G}$ (this is also true for some other LCP classes published so far). However, the co-problem, to tell whether an LCP problem (M,q) is not in $\mathcal{G}$, can be solved in polynomial time. We can simply run the potential reduction algorithm for the LCP problem. In polynomial time the algorithm either gives the solution or concludes that (M,q) is not in $\mathcal{G}$.

9.1.3 Some existing classes

We see that the new class $\mathcal{G}$ has the same bound on the condition number as the PSD class, that is, $\gamma(M,q,\epsilon) \geq 1$. Here, we list several existing classes of LCPs that belong to $\mathcal{G}$.

1. M is positive semi-definite and q is arbitrary.

   We have if Σ^+ is not empty, then

   $$0 < (x-\pi)^T(s + M^T\pi) = x^T s - q^T\pi - \pi^T M^T \pi$$

   which implies

   $$x^T s - q^T \pi > \pi^T M^T \pi \geq 0,$$

   a contradiction.

2. M is copositive and $q \geq 0$.

   We have
   $$x^T s - q^T \pi = x^T M x + q^T (x - \pi).$$
   Thus, $x > 0$ and $x - \pi > 0$ imply $x^T s - q^T \pi \geq 0$, that is, Σ^+ is empty.

3. M^{-1} is copositive and $M^{-1} q \leq 0$.

   We have
   $$x^T s - q^T \pi = s^T M^{-T} s - (M^{-1} q)^T (s + M^T \pi).$$
   Thus, $s > 0$ and $s + M^T \pi > 0$ implies $x^T s - q^T \pi \geq 0$, that is, Σ^+ is empty.

Although a trivial solution may exist for the last two classes (e.g., $x = 0$ and $s = q$ for the second class), our computational experience indicates that the potential reduction algorithm usually converges to a nontrivial solution if multiple solutions exist.

Example 9.1

$$M = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \quad \textit{and} \quad q = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

For this example the potential reduction algorithm constantly generates the solution
$$x = (2\ 2)^T \quad \text{and} \quad s = (0\ 0)^T$$
from virtually any interior starting point, avoiding the trivial solution $x = 0$ and $s = q$.

Another nonconvex LCP also belongs to $\mathcal{G}$.

Example 9.2

$$M = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix}, \quad \textit{and} \quad q = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

$\overset{\circ}{\mathcal{F}}$ is nonempty since $x = (3\ 1)^T$ is an interior feasible point; Σ^+ is empty since $x_1 - x_2 > 1$, $x_1 - \pi_1 > 0$, $x_2 - \pi_2 > 0$, $x_1 - x_2 - 1 + \pi_1 + 2\pi_2 > 0$ and $2x_1 - 1 - \pi_1 > 0$ imply

$$\begin{aligned}
& x^T s - q^T \pi \\
&= x^T (Mx + q) - q^T \pi \\
&= x_1 (x_1 - x_2) + 2 x_1 x_2 - x_1 - x_2 + \pi_1 + \pi_2 \\
&= x_1^2 + x_1 x_2 - 2x_1 - x_2 + 1 + (x_1 - x_2 - 1 + \pi_1 + 2\pi_2) + (x_2 - \pi_2) \\
&> x_1^2 + x_1 x_2 - 2x_1 - x_2 + 1 \\
&= (x_1 - 1)^2 + x_2 (x_1 - 1) > 0.
\end{aligned}$$

As a byproduct, we have

$$\mathcal{G} \subset \{(M,q) : |sol(M,q)| \geq 1\}.$$

In fact, any LCP (M,q) with $\gamma(M,q,\epsilon) > 0$ belongs to $\{(M,q) : |sol(M,q)| \geq 1\}$. Furthermore, if $\gamma(M,q,\epsilon) > 0$ for all $q \in \mathcal{R}^n$, then $M \in \mathcal{Q}$, a matrix class where the LCP (M,q) has at least one solution for all $q \in \mathcal{R}^n$. How to calculate $\gamma(M,q,\epsilon)$ or a lower bound for $\gamma(M,q,\epsilon)$ in polynomial time is a further research topic.

9.2 P-Matrix Linear Complementarity Problem

We now prove Proposition 9.3, that is,

$$\gamma(M,q,\epsilon) \geq \min(n\theta/|\lambda|, 1)$$

where $\lambda = \lambda(M)$ is the least eigenvalue of $(M+M^T)/2$, and $\theta = \theta(M)$ is the positive P-matrix number of M^T. Thus, the P-matrix LCP can be solved in

$$O(n^2 \max(\frac{-\lambda}{\theta n}, 1) \log(1/\epsilon))$$

iterations and each iteration solves a system of linear equations in at most $O(n^3)$ arithmetic operations. This bound indicates that the algorithm is a polynomial-time algorithm if $|\lambda|/\theta$ is bounded above by a polynomial in $\log(1/\epsilon)$ and n.

It has been shown that $\|p^k\|^2 \geq 1$ if M is positive semi-definite (that is, if $\lambda \geq 0$). Thus, in this section we assume that $\lambda < 0$. We also fix

$$\rho = 3n + \sqrt{2n}. \tag{9.9}$$

We first prove the following lemma.

Lemma 9.10

$$(a+b)^2 + (a+c)^2 \geq a^2 - 2bc.$$

Proof.

$$\begin{aligned}(a+b)^2 + (a+c)^2 &= 2a^2 + 2a(b+c) + b^2 + c^2 \\ &= a^2 - 2bc + a^2 + 2a(b+c) + (b+c)^2 \\ &= a^2 - 2bc + (a+b+c)^2 \geq a^2 - 2bc.\end{aligned}$$

□

Now, we have the following lemma.

Lemma 9.11 *Given any* $(x^k, s^k) \in \overset{\circ}{\mathcal{F}}$*, let* p^k *be the scaled gradient projection computed from (9.2) and (9.3). Then,*

$$\|p^k\|^2 \geq \min(\frac{\gamma n}{|\lambda|}, 1).$$

Proof. Let $\bar{s} = s^k + M^T\pi$, $\bar{x} = x^k - \pi$ and $\bar{\Delta} = (x^k)^T\bar{s} + (s^k)^T\bar{x}$. Then, it is obvious that if $\bar{s} \not> 0$ or $\bar{x} \not> 0$, then

$$\|p^k\|^2 \geq 1. \tag{9.10}$$

Therefore, we assume that $\bar{s} > 0$ and $\bar{x} > 0$ in the rest of the proof. Note that from Lemma 9.7

$$\|p^k\|^2 \geq (\frac{\rho\bar{\Delta}}{2n\Delta^k} - 1)^2 2n$$

and

$$\bar{\Delta} = \Delta^k + \bar{x}^T\bar{s} + \pi^T M^T \pi > \Delta^k + \pi^T M^T \pi.$$

Thus, if

$$\pi M^T \pi \geq -\frac{n\Delta^k}{\rho},$$

then from (9.9)

$$\|p^k\|^2 \geq (\frac{\rho\bar{\Delta}}{2n\Delta^k} - 1)^2 2n \geq 1. \tag{9.11}$$

Otherwise, we have

$$\lambda\|\pi\|^2 \leq \pi^T M^T \pi < -\frac{n\Delta^k}{\rho},$$

i.e.,

$$\|\pi\|^2 \geq \frac{n\Delta^k}{\rho|\lambda|}. \tag{9.12}$$

Since M is a P-matrix, there exists an index j such that $\pi_j(M^T\pi)_j \geq \gamma\|\pi\|^2 > 0$. Using Lemma 9.10 and (9.12), we have

$$\begin{aligned}
&\|p^k\|^2 \\
\geq\ & (\frac{\rho}{\Delta^k}x_j^k s_j^k - 1 + \frac{\rho}{\Delta^k}x_j^k(M^T\pi)_j)^2 + (\frac{\rho}{\Delta^k}x_j^k s_j^k - 1 - \frac{\rho}{\Delta^k}s_j^k\pi_j)^2 \\
\geq\ & (\frac{\rho}{\Delta^k}x_j^k s_j^k - 1)^2 + 2x_j^k s_j^k \pi_j (M^T\pi)_j (\frac{\rho}{\Delta^k})^2 \\
\geq\ & (\frac{\rho}{\Delta^k}x_j^k s_j^k - 1)^2 + 2\frac{\rho}{\Delta^k}x_j^k s_j^k \gamma\|\pi\|^2 \frac{\rho}{\Delta^k} \\
\geq\ & (\frac{\rho}{\Delta^k}x_j^k s_j^k - 1)^2 + 2\frac{\rho}{\Delta^k}x_j^k s_j^k \frac{\gamma n}{|\lambda|}.
\end{aligned} \tag{9.13}$$

If

$$\frac{\gamma n}{|\lambda|} \geq 1,$$

then again

$$\|p^k\|^2 \geq (\frac{\rho}{\Delta^k} x_j^k s_j^k)^2 + 1 \geq 1; \tag{9.14}$$

otherwise,

$$\|p^k\|^2 \geq 2\frac{\gamma n}{|\lambda|} - \frac{(\gamma n)^2}{\lambda^2} \geq \frac{\gamma n}{|\lambda|} \tag{9.15}$$

since the quadratic term of (9.13) yields the minimum at

$$\frac{\rho}{\Delta^k} x_j^k s_j^k = 1 - \frac{\gamma n}{|\lambda|}.$$

From (9.10), (9.11), (9.14) and (9.15), we have the desired result.

□

The result leads to the following theorem.

Theorem 9.12 *Let $\psi(x^0, s^0) \leq O(n \log n)$ and M be a P-matrix. Then, the potential reduction algorithm terminates at $(x^k)^T s^k \leq \epsilon$ in*

$$O(n^2 \max(|\lambda|/(\gamma n), 1) \log(1/\epsilon))$$

iterations and each iteration uses at most $O(n^3)$ arithmetic operations.

9.3 Generalized Linear Complementarity Problem

In this section we consider a generalized linear complementarity problem:

$$\begin{array}{lll} (GLCP) & \text{minimize} & x^T s \\ & \text{subject to} & Ax + Bs + Cz = q, \ (x, s, z) \geq 0. \end{array}$$

Let $\mathcal{F}$ denote the feasible set. It is evident that a solution, with $x^T s = 0$, to the GLCP may not exist even when the problem is feasible. However, a finite stationary or KKT point of the GLCP, which is defined as a point satisfying the first order optimality conditions of (GLCP), must exist, since the objective function is quadratic and bounded from below so that it has a finite minimum.

More precisely, a KKT point, $(\bar{x}, \bar{s}, \bar{z}) \in \mathcal{F}$, of the GLCP is represented by

$$\bar{s}^T \bar{x} + \bar{x}^T \bar{s} \leq \bar{s}^T x + \bar{x}^T s \quad \text{for all} \quad (x, s, z) \in \mathcal{F}.$$

In other words, $(\bar{x}, \bar{s}, \bar{z})$ is a minimal solution for the related linear program

$$\begin{array}{ll} \text{minimize} & \bar{s}^T x + \bar{x}^T s \\ \text{subject to} & Ax + Bs + Cz = q, \\ & (x, s, z) \geq 0, \end{array}$$

where its dual is

$$\begin{array}{ll} \text{maximize} & q^T \pi \\ \text{subject to} & \bar{s} - A^T \pi \geq 0, \\ & \bar{x} - B^T \pi \geq 0, \\ & -C^T \pi \geq 0. \end{array}$$

Thus, $(\bar{x}, \bar{s}, \bar{z}) \in \mathcal{F}$ is a KKT point if and only if there exists $\bar{\pi} \in \mathcal{R}^m$ such that

$$\bar{s} - A^T \bar{\pi} \geq 0, \quad \bar{x} - B^T \bar{\pi} \geq 0 \quad \text{and} \quad -C^T \bar{\pi} \geq 0,$$

and

$$\bar{x}^T (\bar{s} - A^T \bar{\pi}) = 0, \quad \bar{s}^T (\bar{x} - B^T \bar{\pi}) = 0 \quad \text{and} \quad \bar{z}^T (-C^T \bar{\pi}) = 0.$$

We see that finding such a KKT point itself is a GLCP. We also note that a solution to the GLCP, (x^*, s^*, z^*), can be viewed as a special KKT point with $\bar{\pi} = 0$.

The concept of the fully polynomial-time approximation scheme (FPTAS) was introduced in combinatorial optimization. Given an instance of an optimization problem and an $\epsilon > 0$, it returns an ϵ-approximate solution within a time period bounded by a polynomial both in the length of the instance and $1/\epsilon$. For some combinatorial optimization problems, the theory of NP-completeness can be applied to prove not only that they cannot be solved exactly by polynomial-time algorithms (unless $P = NP$), but also that they do not have ϵ-approximate algorithms, for various ranges of ϵ, again unless $P = NP$. Furthermore, approximation algorithms are widely used and accepted in practice.

In this paper, we develop a fully polynomial-time approximation scheme for generating an ϵ-KKT point of the GLCP—a point $(\hat{x}, \hat{s}, \hat{z}) \in \mathcal{F}$ and $\hat{\pi} \in \mathcal{R}^m$ with

$$\hat{s} - A^T \hat{\pi} \geq 0, \quad \hat{x} - B^T \hat{\pi} \geq 0 \quad \text{and} \quad -C^T \hat{\pi} \geq 0, \tag{9.16}$$

and

$$\frac{\hat{x}^T (\hat{s} - A^T \pi) + \hat{s}^T (\hat{x} - B^T \pi) + \hat{z}^T (-C^T \pi)}{\hat{x}^T \hat{s}} \leq \epsilon. \tag{9.17}$$

In other words, $(\hat{x}, \hat{s}, \hat{z}, \hat{\pi})$ is feasible and the sum of the complementary slackness vectors (or the primal-dual objective gap) relative to the (primal) objective value is less than ϵ. Thus, the algorithm is actually a polynomial approximation algorithm for solving a class of GLCPs in which every KKT point is a solution. This class includes the LCP with the row-sufficient matrix.

We assume at this moment that the interior of $\mathcal{F}$, $\overset{\circ}{\mathcal{F}} = \{(x, s, z) \in \mathcal{F} : x > 0,\ s > 0 \text{ and } z > 0\}$ is nonempty, and an approximate analytic center, (x^0, s^0, z^0) of $\mathcal{F}$, is available. We further assume that each component of z is bounded by R in the feasible set $\mathcal{F}$. Both assumptions will be *removed* later, so that the results should hold for general cases.

9.3.1 Potential reduction algorithm

Let $\rho \geq 2n + d$ and define the potential function

$$\psi(x, s, z) := \rho \ln(x^T s) - \sum_{j=1}^{n} \log(x_j) - \sum_{j=1}^{n} \log(s_j) - \sum_{j=1}^{d} \log(z_j)$$

to associate with a feasible point $(x, s, z) \in \overset{\circ}{\mathcal{F}}$. Using the arithmetic-geometric mean inequality, we have

$$(\rho - n) \ln(x^T s) - \sum_{j=1}^{d} \log(z_j) \leq \psi(x, s, z) - n \ln(n) \leq \psi(x, s, z).$$

On the other hand, $\sum_{j=1}^{d} \log(z_j) \leq d \log R$ from the boundedness assumption. Thus,

$$\psi(x, s, z) < n \log n - d \log R + (\rho - n) \log \epsilon \Longrightarrow x^T s < \epsilon. \tag{9.18}$$

The basic idea of the potential reduction algorithm works as follows. Given an approximate analytic center $(x^0, s^0, z^0) \in \overset{\circ}{\mathcal{F}}$, we generate a sequence of interior feasible solutions $\{x^k, s^k, z^k\} \in \overset{\circ}{\mathcal{F}}$ and $\{\pi^k\} \in \mathcal{R}^m$ with the following property: Unless

$$s^k - A^T \pi^k > 0, \quad x^k - B^T \pi^k > 0, \quad -C^T \pi^k > 0$$

and

$$\frac{(x^k)^T (s^k - A^T \pi^k) + (s^k)^T (x^k - B^T \pi^k) + (z^k)^T (-C^T \pi^k)}{(x^k)^T s^k} < (2n + d + \sqrt{2n + d})/\rho,$$

we always have

$$\psi(x^{k+1}, s^{k+1}, z^{k+1}) \leq \psi(x^k, s^k, z^k) - O(1/\rho).$$

Thus, if we choose $\rho = (2n+d+\sqrt{2n+d})/\epsilon$, the algorithm requires at most $\rho^2(\log(1/\epsilon) + \log R)$ iterations, which is a polynomial in $1/\epsilon$, to generate either a solution or an ϵ-KKT point of the GLCP. Note that each iteration involves $O(n^3)$ arithmetic operations.

There are many ways to achieve a potential reduction. We again use the scaled gradient projection method. The gradient vectors $\nabla\psi_x$ and $\nabla\psi_s$ are identical to those in the preceding section, and the one with respect to z is

$$\nabla\psi_z = -Z^{-1}e,$$

where $\Delta = x^T s$. Now, we solve the following linear program subject to an ellipsoidal constraint at the kth iteration:

$$\begin{array}{ll} \text{maximize} & \nabla^T\psi_{x^k} d_x + \nabla^T\psi_{s^k} d_s + \nabla^T\psi_{z^k} d_z \\ \text{subject to} & A d_x + B d_s + C d_z = 0, \\ & \|(X^k)^{-1} d_x\|^2 + \|(S^k)^{-1} d_s\|^2 + \|(Z^k)^{-1} d_z\|^2 \le \beta^2 < 1, \end{array}$$

and denote by $\bar d_x$, $\bar d_s$ and $\bar d_z$ its minimal solutions. Then, we have

$$\begin{pmatrix} (X^k)^{-1}\bar d_x \\ (S^k)^{-1}\bar d_s \\ (Z^k)^{-1}\bar d_z \end{pmatrix} = \frac{-\beta p^k}{\|p^k\|},$$

where p^k is the projection of the scaled gradient vector

$$(\nabla^T\psi_{x^k} X^k, \nabla^T\psi_{s^k} S^k, \nabla^T\psi_{z^k} Z^k)$$

onto the null space of the scaled constraint matrix (AX^k, BS^k, CZ^k), i.e.,

$$p^k = \begin{pmatrix} p_x^k \\ p_s^k \\ p_z^k \end{pmatrix} = \begin{pmatrix} \frac{\rho}{\Delta^k} X^k(s^k - A^T\pi^k) - e \\ \frac{\rho}{\Delta^k} S^k(x^k - B^T\pi^k) - e \\ \frac{\rho}{\Delta^k} Z^k(-C^T\pi^k) - e \end{pmatrix},$$

$$\pi^k = \frac{\Delta^k}{\rho}(\bar A \bar A^T)^{-1}\bar A(\nabla^T\psi_{x^k} X^k, \nabla^T\psi_{s^k} S^k, \nabla^T\psi_{z^k} Z^k)^T,$$

$$\bar A = (AX^k, BS^k, CZ^k), \quad \text{and} \quad \Delta^k = (x^k)^T s^k.$$

Let $x^{k+1} = x^k + \bar d_x$, $s^{k+1} = s^k + \bar d_s$ and $z^{k+1} = z^k + \bar d_z$. Then

$$\begin{aligned} &\rho\ln((x^{k+1})^T s^{k+1}) - \rho\ln((x^k)^T s^k) \\ &\le \frac{\rho}{\Delta^k}((s^k)^T\bar d_x + (x^k)^T\bar d_s + \bar d_x^T\bar d_s) \\ &= \frac{\rho}{\Delta^k}((s^k)^T\bar d_x + (x^k)^T\bar d_s) + \frac{\rho}{\Delta^k}\bar d_x^T(X^k)^{-1}(X^kS^k)(S^k)^{-1}\bar d_s \end{aligned}$$

$$
\begin{aligned}
&\le \frac{\rho}{\Delta^k}((s^k)^T\bar{d}_x + (x^k)^T\bar{d}_s) + \frac{\rho}{\Delta^k}\|\bar{d}_x^T(X^k)^{-1}\|\|(X^kS^k)\|\|(S^k)^{-1}\bar{d}_s\| \\
&\le \frac{\rho}{\Delta^k}((s^k)^T\bar{d}_x + (x^k)^T\bar{d}_s) + \rho\|\bar{d}_x^T(X^k)^{-1}\|\|(S^k)^{-1}\bar{d}_s\| \\
&\quad (\text{since } \|X^kS^k\| \le \Delta^k) \\
&\le \frac{\rho}{\Delta^k}((s^k)^T\bar{d}_x + (x^k)^T\bar{d}_s) + \frac{\rho}{2}(\|\bar{d}_x^T(X^k)^{-1}\|^2 + \|(S^k)^{-1}\bar{d}_s\|^2) \\
&\le \frac{\rho}{\Delta^k}((s^k)^T\bar{d}_x + (x^k)^T\bar{d}_s) + \frac{\rho\beta^2}{2}.
\end{aligned}
$$

Furthermore, from Lemma 3.1 of Chapter 3

$$
\begin{aligned}
&-\sum_{j=1}^n \log(x_j^{k+1}) - \sum_{j=1}^n \log(s_j^{k+1}) - \sum_{j=1}^d \log(z_j^{k+1}) \\
&\qquad + \sum_{j=1}^n \log(x_j^k) + \sum_{j=1}^n \log(s_j^k) + \sum_{j=1}^d \log(z_j^k) \\
&\le \ -e^T(X^k)^{-1}\bar{d}_x - e^T(S^k)^{-1}\bar{d}_s - e^T(Z^k)^{-1}\bar{d}_z + \frac{\beta^2}{2(1-\beta)}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\psi(x^{k+1}, s^{k+1}) - \psi(x^k, s^k) \\
&\le \frac{\rho}{\Delta^k}((s^k)^T\bar{d}_x + (x^k)^T\bar{d}_s) - e^T(X^k)^{-1}\bar{d}_x - e^T(S^k)^{-1}\bar{d}_s - e^T(Z^k)^{-1}\bar{d}_z \\
&\quad + \frac{\beta^2}{2}(\rho + \frac{1}{1-\beta}) \\
&= \nabla^T\psi_{x^k}\bar{d}_x + \nabla^T\psi_{s^k}\bar{d}_s + \nabla^T\psi_{z^k}\bar{d}_z + \frac{\beta^2}{2}(\rho + \frac{1}{1-\beta}) \\
&= (\nabla^T\psi_{x^k}X^k, \nabla^T\psi_{s^k}S^k, \nabla^T\psi_{z^k}Z^k)\frac{-\beta p^k}{\|p^k\|} + \frac{\beta^2}{2}(\rho + \frac{1}{1-\beta}) \\
&= \frac{-\beta\|p^k\|^2}{\|p^k\|} + \frac{\beta^2}{2}(\rho + \frac{1}{1-\beta}) \\
&= -\beta\|p^k\| + \frac{\beta^2}{2}(\rho + \frac{1}{1-\beta}).
\end{aligned}
$$

Therefore, choosing β as in (9.5), we have

$$
\psi(x^{k+1}, s^{k+1}, z^{k+1}) - \psi(x^k, s^k, z^k) \le -\min(\frac{\|p^k\|^2}{2(\rho+2)}, \frac{1}{2(\rho+2)}). \quad (9.19)
$$

In practice, the step-size β can be determined by a line search along the direction p^k to minimize the potential function.

9.3.2 Complexity Analysis

We further study $\|p^k\|$ by introducing the following lemma.

Lemma 9.13 *The scaled gradient projection $\|p^k\| < 1$ implies*

$$s^k - A^T\pi^k > 0, \quad x^k - B^T\pi^k > 0, \qquad -C^T\pi^k > 0,$$

and

$$\frac{2n+d-\sqrt{2n+d}}{\rho}\Delta^k < \bar{\Delta} < \frac{2n+d+\sqrt{2n+d}}{\rho}\Delta^k,$$

where $\bar{\Delta} = (x^k)^T(s^k - A^T\pi^k) + (s^k)^T(x^k - B^T\pi^k) + (z^k)^T(-C^T\pi^k)$.

Proof. The proof is by contradiction. It is obvious that if $s^k - A^T\pi^k \not> 0$ or $x^k - B^T\pi^k \not> 0$ or $-C^T\pi^k \not> 0$, then

$$\|p^k\| \geq \|p^k\|_\infty \geq 1.$$

On the other hand, we have

$$\begin{aligned}
\|p^k\|^2 &= (\frac{\rho}{\Delta^k})^2 \| \begin{pmatrix} X^k(s^k - A^T\pi^k) \\ S^k(x^k - B^T\pi^k) \\ Z^k(-C^T\pi^k) \end{pmatrix} - \frac{\bar{\Delta}}{2n+d} e\|^2 + \|\frac{\rho\bar{\Delta}}{(2n+d)\Delta^k}e - e\|^2 \\
&\geq \|\frac{\rho\bar{\Delta}}{(2n+d)\Delta^k}e - e\|^2 \\
&= (\frac{\rho\bar{\Delta}}{(2n+d)\Delta^k} - 1)^2(2n+d). \qquad (9.20)
\end{aligned}$$

(Note that the dimension of e is $2n+d$ here.) Hence, the following must be true

$$(\frac{\rho\bar{\Delta}}{(2n+d)\Delta^k} - 1)^2(2n+d) < 1,$$

that is,

$$\frac{2n+d-\sqrt{2n+d}}{\rho}\Delta^k < \bar{\Delta} < \frac{2n+d+\sqrt{2n+d}}{\rho}\Delta^k.$$

□

Now we can prove the following theorem. For simplicity, we assume that $\psi^0 = O(n\log n)$.

Theorem 9.14 *For any given $0 < \epsilon \leq 1$, let $\rho = (2n + d + \sqrt{2n+d})/\epsilon$. Then, under the assumptions (A1) and (A2), the potential reduction algorithm terminates in at most $O(\rho^2 \log(1/\epsilon) + \rho d \log R)$ iterations. The algorithm generates an ϵ-KKT point $(x^k, s^k, z^k) \in \mathcal{F}$ and $\pi^k \in \mathcal{R}^m$ of the GLCP, either*

$$(x^k)^T s^k \leq \epsilon$$

or

$$s^k - A^T \pi^k > 0, \quad x^k - B^T \pi^k > 0, \quad -C^T \pi^k > 0$$

and

$$\frac{(x^k)^T (s^k - A^T \pi^k) + (s^k)^T (x^k - B^T \pi^k) + (z^k)^T (-C^T \pi^k)}{(x^k)^T s^k} < \epsilon.$$

Proof. The proof directly follows Lemma 9.13. If $\|p^k\| \geq 1$ for all k, then from (9.19)

$$\psi(x^{k+1}, s^{k+1}, z^{k+1}) - \psi(x^k, s^k, z^k) \leq -O(1/\rho).$$

Therefore, in at most $O(\rho^2 \log(1/\epsilon) + \rho d \log R)$ iterations

$$\psi(x^k, s^k, z^k) \leq (\rho - n) \log \epsilon - d \log R + n \log n,$$

and from (9.18)

$$(x^k)^T s^k \leq \epsilon.$$

As we mentioned before, in this case (x^k, s^k, z^k) is a special stationary point with $\bar{\pi} = 0$. Otherwise, we have $\|p^k\| < 1$ for some $k \leq O(\rho^2 \log(1/\epsilon) + \rho d \log R)$. This implies the relations (9.16) and (9.17) from Lemma 9.13.

$\square$

Theorem 9.14 indicates that the potential reduction algorithm is a fully polynomial-time approximation scheme for computing an ϵ-approximate KKT point of the GLCP. In the following, we present a sufficient condition to show that a solution to the GLCP always exists, and the potential reduction algorithm solves it in polynomial time under the assumptions. Moreover, we have

Theorem 9.15 *Let BA^T be negative semi-definite. Furthermore, let $\rho = 2n + d + \sqrt{2n+d}$. Then, the potential reduction algorithm generates a solution to the GLCP in $O((2n+d)^2 \log(1/\epsilon) + (2n+d) d \log R)$ iterations.*

Proof. Basically, we show that $\|p^k\| \geq 1$ for all k if BA^T is negative semi-definite and $\rho \geq 2n + d + \sqrt{2n+d}$. We prove it by contradiction. Suppose $\|p^k\| < 1$, then

$$s^k - A^T\pi^k > 0 \quad x^k - B^T\pi^k > 0 \quad \text{and} \quad -C^T\pi^k > 0.$$

Thus,

$$(x^k - B^T\pi^k)^T(s^k - A^T\pi^k) > 0,$$

that is

$$(x^k)^T s^k - (x^k)^T A^T\pi^k - (s^k)^T B^T\pi^k + (\pi^k)^T BA^T\pi^k > 0.$$

Also note

$$-(z^k)^T C^T\pi^k > 0.$$

Combining the above two inequalities, we have

$$(x^k)^T s^k - (x^k)^T A^T\pi^k - (s^k)^T B^T\pi^k - (z^k)^T C^T\pi^k > -(\pi^k)^T BA^T\pi^k \geq 0$$

or

$$\begin{aligned}
\bar{\Delta} &= (x^k)^T(s^k - A^T\pi^k) + (s^k)^T(x^k - B^T\pi^k) + (z^k)^T(-C^T\pi^k) \\
&= (x^k)^T s^k + (x^k)^T s^k - (x^k)^T A^T\pi^k - (s^k)^T B^T\pi^k - (z^k)^T C^T\pi^k \\
&> (x^k)^T s^k = \Delta^k.
\end{aligned}$$

From (9.20), we have

$$\|p^k\|^2 \geq \left(\frac{\rho\bar{\Delta}}{(2n+d)\Delta^k} - 1\right)^2(2n+d) \geq 1,$$

which is a contradiction.

$\square$

Clearly, the result for the LCP with positive semi-definite matrix is a special application of Theorem 9.15.

9.3.3 Remove the assumptions

The above theorems hold under certain assumptions. We now remove some of them and show that our main results remain valid. Note first that the z-boundedness assumption is automatically unnecessary for LCP.

We now remove the assumption of availability of the initial point. We apply the linear programming Phase I procedure to find a feasible solution for the system

$$(A, B, C)u = q \quad \text{and} \quad u \geq 0.$$

In polynomial time, an interior-point algorithm either declares that the system is infeasible, or generates a max-feasible solution $\bar{u}$. Thus, we have detected those variables that must be zero in every feasible solutions of (GLCP) (in this case, the feasible region has empty interior). Then, we eliminate those variables from the system. For example, if x_1 is zero in every feasible solution of (GLCP), we can eliminate x_1 and then move s_1 into z; if both x_1 and s_1 are zero in every feasible solution, we can eliminate both of them. Thus, we will have a reduced system where the feasible region has a nonempty interior, and a feasible solution is at hand.

Hence, Theorems 9.14 and 9.15 hold without the assumption. It has been shown that every KKT point of the LCP with a row-sufficient matrix is a solution of the LCP. Therefore, we have

Corollary 9.16 *The potential reduction algorithm is a fully polynomial-time approximation scheme for generating an ϵ-approximate stationary point of the LCP with row-sufficient matrices, where every KKT point is a solution of the LCP.*

9.3.4 Numerical Examples

We give three examples of the GLCP and try to illustrate the convergence behavior of the algorithm. These experiments are very preliminary.

Example 9.3

$$A = \begin{pmatrix} 0 & 1 & 10 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad B = -I, \quad C = \emptyset, \quad \textit{and} \quad q = e.$$

The starting point is $x^0 = (2,2,2)^T$. The algorithm consistently generates the solution to the LCP, $x^* = (0, 0.5, 0.5)^T$ and $s^* = (4.5, 0, 0)^T$. In this example, A is a so called P_0 matrix, and it is indefinite.

Example 9.4

$$A = \begin{pmatrix} 0 & 1 & 10 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad B = -I, \quad C = \emptyset, \quad \textit{and} \quad q = e.$$

The starting point is again $x^0 = (2,2,2)^T$. The algorithm consistently generates a KKT point of the LCP, $\bar{x} = (0, \alpha, 1)^T$, $\bar{s} = (9 + \alpha, 0, 1)^T$ and $\bar{\pi} = (0, -3, 1)^T$ for some $\alpha > 0$. Note that there is no solution to the GLCP in this example.

Example 9.5

$$A = \begin{pmatrix} 0 & 1 & 10 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = -I, \quad C = \emptyset, \quad \textit{and} \quad q = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

The starting point is again $x^0 = (2,2,2)^T$. The algorithm consistently generates a KKT point of the LCP, $\bar{x} = (0,\alpha,1)^T$, $\bar{s} = (9+\alpha,0,1)^T$ and $\bar{\pi} = (0,-1,1)^T$ for some $\alpha > 0$. Again, there is no solution to the GLCP in this example.

9.4 Indefinite Quadratic Programming

Consider the quadratic programming (QP) problem and its dual (QD) in Section 1.3.7. For simplicity we assume that A has full row-rank. If Q is positive semi-definite in the null space of A, that is, for any given $d \in \mathcal{N}(A)$,

$$d^T Q d \geq 0,$$

then (QP) is a convex optimization problem and it can be solved in polynomial time. In this case, the standard duality theorems hold and the optimality condition becomes $x \in \mathcal{F}_p$, $(x,y) \in \mathcal{F}_d$, and $q(x) - d(x,y) = x^T(Qx + c - A^T y) = x^T s = 0$.

The algorithm presented in this section handles general QP problems: convex or nonconvex. For the simplicity of our analysis, throughout this section we let (QP) be a non-convex problem, e.g., Q have at least one negative eigenvalue in the null space of A. Then, (QP) becomes a hard problem, an NP-complete problem.

No time complexity bounds have been developed for various QP methods. (Of course, an enumerative search approach will solve (QP) but it possesses an exponential time bound.) These algorithms generally generate a sequence of points that converges to a stationary or KKT point associated with (QP), which satisfies

$$x^T(Qx + c - A^T y) = 0, \quad x \in \mathcal{F}_p, \quad \text{and} \quad (x,y) \in \mathcal{F}_d.$$

For convenience, we let $s = Qx + c - A^T y$. Then, $(x,y) \in \mathcal{F}_d$ implies $s \geq 0$. For any $x \in \mathcal{F}_p$ and $(x,y) \in \mathcal{F}_d$, the quantity $x^T s = q(x) - d(x,y)$ is the complementarity gap.

Here we assume that the feasible set, $\mathcal{F}_p$, of (QP) has a strictly positive feasible solution. For any given (A,b), to see if $\mathcal{F}_p$ possesses a strictly positive feasible solution can be solved as a single linear program in polynomial time, so that the above assumption is without of loss of any generality. We make an additional assumption that the feasible region is bounded. With this assumption (QP) has a minimizer and a maximizer. Let $\underline{z}$ and $\bar{z}$ be their minimal and maximal objective values, respectively.

An ϵ-minimal solution or ϵ-minimizer, $\epsilon \in (0,1)$, for (QP) is defined as an $x \in \mathcal{F}_p$ such that

$$\frac{q(x) - \underline{z}}{\bar{z} - \underline{z}} \leq \epsilon.$$

(Vavasis [360] discussed the importance to have the term $(\bar{z} - \underline{z})$ in the criterion for continuous optimization.) Similarly, we define an ϵ-KKT solution for (QP) as an (x, y, s) such that $x \in \mathcal{F}_p$, $(x, y, s) \in \mathcal{F}_d$, and

$$\frac{x^T s}{\bar{z} - \underline{z}} = \frac{q(x) - d(x,y)}{\bar{z} - \underline{z}} \leq \epsilon.$$

Note that the minimizer of (QP) is a special KKT point such that $q(x) = d(x, y) = \underline{z}$.

In this section we extend the potential reduction algorithm described in Section 9.3 to compute an ϵ-KKT point in $O((\frac{n^6}{\epsilon}\log\frac{1}{\epsilon} + n^4 \log n)\log(\log\frac{R}{\epsilon}))$ arithmetic operations. We also show that Q is positive semi-definite in the null space of the active set at the limit of this point, indicating that the limit satisfies the second-order necessary condition to be a local minimal solution. The result is the first approximation algorithm whose running time is almost linear in $\frac{1}{\epsilon}$, which was an open question in the area of nonlinear optimization complexity.

9.4.1 Potential reduction algorithm

We now describe the potential function and the potential reduction algorithm for solving (QP). Let $z \leq \underline{z}$. Then the potential function of $0 < x \in \mathcal{F}_p$ here is

$$\phi(x) := \rho \log(q(x) - z) - \sum_{j=1}^{n} \log(x_j),$$

where $\rho > n$.

Starting from a point $0 < x^0 \in \mathcal{F}_p$, the potential reduction algorithm will generate a sequence of $\{x^k\} \in \mathcal{F}_p$ such that $\phi(x^{k+1}) < \phi(x^k)$. For simplicity and convenience, we assume $x^0 = e$, and x^0 is the analytic center of $\mathcal{F}_p$. Our results hold even if x^0 is replaced by an approximate center. Therefore, this assumption is also without loss of generality.

Since x^0 is the analytic center of $\mathcal{F}_p$,

$$\{x \in \mathcal{F}_p : \ \|(X^0)^{-1}(x-x^0)\| \le 1\} \subset \mathcal{F}_p \subset \{x \in \mathcal{F}_p : \ \|(X^0)^{-1}(x-x^0)\| \le n\}. \tag{9.21}$$

In other words, $\mathcal{F}_p$ contains the ellipsoid $\{x \in \mathcal{F}_p : \ \|(X^0)^{-1}(x-x^0)\| \le 1\}$ and it is contained in the co-centered ellipsoid $\{x \in \mathcal{F}_p : \ \|(X^0)^{-1}(x-x^0)\| \le n\}$, where the ratio of their radii is n.

Thus, for any $x \in \mathcal{F}_p$

$$\sum_{j=1}^{n} \log(x_j^0/x_j) \ge -n\log(n+1). \tag{9.22}$$

Note that if

$$\rho\log(q(x^k)-z) - \rho\log(q(x^0)-z) \le \rho\log\epsilon$$

or

$$\phi(x^k) - \phi(x^0) \le \rho\log\epsilon + \sum_{j=1}^{n}\log(x_j^0/x_j) \tag{9.23}$$

we must have

$$\frac{q(x^k)-\underline{z}}{\bar{z}-\underline{z}} \le \frac{q(x^k)-\underline{z}}{q(x^0)-\underline{z}} \le \frac{q(x^k)-z}{q(x^0)-z} \le \epsilon,$$

which implies that x^k is an ϵ-minimizer and, thereby, an ϵ-KKT point. From relations (9.22) and (9.23), the above inequality holds if

$$\phi(x^k) - \phi(x^0) \le \rho\log\epsilon - n\log(n+1). \tag{9.24}$$

In other words, x^k becomes an ϵ-minimizer as soon as the net potential reduction exceeds $\rho\log(1/\epsilon) + n\log(n+1)$.

Given $0 < x \in \mathcal{F}_p$, let $\Delta = q(x) - z$ and let d_x, $Ad_x = 0$, be a vector such that $x^+ := x + d_x > 0$. Then

$$\begin{aligned}
&\rho\log(q(x^+)-z) - \rho\log(q(x)-z) \\
&= \ \rho\log(\Delta + \frac{1}{2}d_x^TQd_x + (Qx+c)^Td_x) - \rho\log\Delta \\
&= \ \rho\log(1 + (\frac{1}{2}d_x^TQd_x + (Qx+c)^Td_x)/\Delta) \\
&\le \ \frac{\rho}{\Delta}(\frac{1}{2}d_x^TQd_x + (Qx+c)^Td_x).
\end{aligned}$$

On the other hand, if $\|X^{-1}d_x\| \le \beta < 1$ then

$$-\sum_{j=1}^{n} \log(x_j^+) + \sum_{j=1}^{n} \log(x_j) \quad \le \quad -e^T X^{-1} d_x + \frac{\beta^2}{2(1-\beta)}.$$

Thus, if $\|X^{-1}d_x\| \le \beta < 1$ then $x^+ = x + d_x > 0$ and

$$\phi(x^+) - \phi(x) \le \frac{\rho}{\Delta}(\frac{1}{2}d_x^T Q d_x + (Qx + c - \frac{\Delta}{\rho}X^{-1}e)^T d_x) + \frac{\beta^2}{2(1-\beta)}. \quad (9.25)$$

To achieve a potential reduction, we minimize a quadratic function subject to an ellipsoid constraint. We solve the following problem at the kth iteration:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}d_x^T Q d_x + (Qx^k + c - \frac{\Delta^k}{\rho}(X^k)^{-1}e)^T d_x \\ \\ \text{subject to} & Ad_x = 0, \\ & \|(X^k)^{-1}d_x\|^2 \le \beta^2. \end{array}$$

Let

$$Q^k = X^k Q X^k,\ c^k = X^k Q x^k + X^k c - \frac{\Delta^k}{\rho}e,\ A^k = AX^k, \text{ and } d_x' = (X^k)^{-1}d_x.$$

Then the above problem becomes

$$\begin{array}{lll} (BQP) & \text{minimize} & q'(d_x') := \frac{1}{2}(d_x')^T Q^k d_x' + (c^k)^T d_x' \\ \\ & \text{subject to} & A^k d_x' = 0, \\ & & \|d_x'\|^2 \le \beta^2. \end{array}$$

Let $N^k \in R^{n\times(n-m)}$ be an orthonormal basis spanning the null space of A^k, where $(N^k)^T N^k = I$, and let $H^k = (N^k)^T Q^k N^k \in R^{(n-m)\times(n-m)}$ and $g^k = (N^k)^T c^k \in R^{n-m}$. Then $d_x' = N^k v$ for some $v \in R^{n-m}$ and problem (BQP) can be rewritten as

$$\begin{array}{lll} (BHP) & \text{minimize} & \frac{1}{2}v^T H^k v + (g^k)^T v \\ \\ & \text{subject to} & \|v\| \le \beta^2. \end{array}$$

This is the so-called ball-constrained quadratic problem. We assume that, for now, this problem can be solved efficiently (we will establish this fact later).

The solution d'_x of problem (BQP) satisfies the following necessary and sufficient conditions:

$$\begin{array}{lrcl} & (Q^k + \mu^k I)d'_x - (A^k)^T y(\mu^k) & = & -c^k \text{ for some } y(\mu^k), \\ & A^k d'_x & = & 0, \\ & \mu^k & \geq & \max\{0, -\lambda^k\}, \\ \text{and} & \|d'_x\| & = & \beta, \end{array} \tag{9.26}$$

or, equivalently, the solution v of (BHP) satisfies the following necessary and sufficient conditions:

$$\begin{array}{lrcl} & (H^k + \mu^k I)v & = & -g^k, \\ & \mu^k & \geq & \max\{0, -\lambda^k\}, \\ \text{and} & \|v\| & = & \beta, \end{array} \tag{9.27}$$

where $\lambda^k = \underline{\lambda}(H^k)$, and $\underline{\lambda}(H)$ denotes the least eigenvalue of matrix H. Since Q is not convex in the null space of A and x^k is strictly positive, we must have $\lambda^k < 0$.

Let

$$\begin{array}{c} s(\mu^k) = Q(x^k + d_x) + c - A^T y(\mu^k) \quad \text{and} \\ p^k = Q^k d'_x + c^k - (A^k)^T y(\mu^k)) = X^k s(\mu^k) - \frac{\Delta^k}{\rho} e. \end{array} \tag{9.28}$$

Then,

$$\mu^k = \|p^k\|/\beta, \qquad d'_x = -\frac{\beta p^k}{\|p^k\|} \tag{9.29}$$

and

$$\begin{aligned} q'(d'_x) &= \frac{1}{2}(d'_x)^T Q^k d'_x + (c^k)^T d'_x \\ &= (d'_x)^T (Q^k d'_x + c^k) - \frac{1}{2}(d'_x)^T Q^k d'_x \\ &= (d'_x)^T (Q^k d'_x + c^k - (A^k)^T y(\mu^k)) - \frac{1}{2}(d'_x)^T Q^k d'_x \\ &= -\beta^2 \mu^k - \frac{1}{2}(d'_x)^T Q^k d'_x \\ &= -\beta^2 \mu^k - \frac{1}{2} v^T (N^k)^T Q^k N^k (N^k)^T v \\ &\leq -\beta^2 \mu^k - \frac{1}{2}\underline{\lambda}((N^k)^T Q^k N^k)\|v\|^2 \end{aligned}$$

$$
\begin{aligned}
&= -\beta^2\mu^k + \frac{\beta^2}{2}|\lambda^k| \\
&\leq -\frac{\beta^2\mu^k}{2} = -\frac{\beta\|p^k\|}{2}.
\end{aligned} \tag{9.30}
$$

This implies that

$$
\frac{\rho}{\Delta}(\frac{1}{2}d_x^T Q d_x + (Qx^k + c - \frac{\Delta^k}{\rho}X^{-1}e)^T d_x) = \frac{\Delta^k}{\rho}q'(d'_x) \leq -\frac{\beta}{2}\frac{\rho}{\Delta^k}\|p^k\|.
$$

Thus, not only we have $x^{k+1} := x^k + d_x > 0$ but also, from (9.25),

$$
\phi(x^{k+1}) - \phi(x^k) \leq -\frac{\beta}{2}\frac{\rho}{\Delta^k}\|p^k\| + \frac{\beta^2}{2(1-\beta)}. \tag{9.31}
$$

Here we see that if

$$
\frac{\rho}{\Delta^k}\beta\mu^k = \frac{\rho}{\Delta^k}\|p^k\| \geq \frac{3}{4}
$$

and if β is chosen around $1/4$, then

$$
\phi(x^{k+1}) - \phi(x^k) \leq -\frac{5}{96}.
$$

Therefore, according to the implication of (9.24) we have

Theorem 9.17 *Let μ^k and β be in condition (9.26). Then, If $\frac{\rho}{\Delta^k}\beta\mu^k = \frac{\rho}{\Delta^k}\|p^k\| \geq \frac{3}{4}$ for all k, then in $O(\rho\log\frac{1}{\epsilon} + n\log n)$ iterations the algorithm returns an ϵ-minimal solution for (QP).*

Now, the question is what happens if $\frac{\rho}{\Delta^k}\|p^k\| < \frac{3}{4}$ for some k. In the next section we prove that then x^{k+1} must be an ϵ-KKT solution for a suitable z and $\rho = 2n^2(n+\sqrt{n})/\epsilon$. Thus, the total number of iterations to generate an ϵ-minimal or ϵ-KKT solution is bounded by $O(\frac{n^3}{\epsilon}\log\frac{1}{\epsilon} + n\log n)$. The algorithm can be simply stated as follows:

Algorithm 9.2 *Let $\rho = 2n^2(n+\sqrt{n})/\epsilon$, $\beta = 1/4$, $z \leq \underline{z}$, and x^0 be an approximate analytic center of $\mathcal{F}_p$. Set $k := 0$.*

While $\frac{\rho}{\Delta^k}\beta\mu^k = \frac{\rho}{\Delta^k}\|p^k\| < \frac{3}{4}$ *in (9.26) or* $\frac{q(x^k)-z}{q(x^0)-z} < \epsilon$ **do**

1. *Solve (BQP).*
2. *Let $x^{k+1} = x^k + X^k d'_x$.*
3. *Set $k := k+1$ and return to Step 1.*

9.4.2 Generating an ϵ-KKT point

We first illustrate how to compute a lower bound, z, for the (QP) minimal value $\underline{z}$. A z can be generated by solving, again, a ball-constrained problem:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}(x-e)^T Q(x-e) + (Qe+c)^T(x-e) \\ \text{subject to} & A(x-e) = 0, \\ & \|x-e\|^2 \leq 1. \end{array}$$

Let $\hat{x}$ be the minimizer. Then, according to Theorem 4 of [376] (we leave its proof as an exercise) and relation (9.21) (i.e., the ratio of the radii of the inscribing and circumscribing ellipsoids to $\mathcal{F}_p$ is $1/n$), we have

$$q(e) - q(\hat{x}) \geq \frac{1}{n^2}(q(e) - \underline{z}).$$

Thus, we can assign

$$z := q(e) - n^2(q(e) - q(\hat{x})). \tag{9.32}$$

Note that an approximate x, say $q(e) - q(x) \geq (q(e) - q(\hat{x}))/1.1$, would establish a lower bound $z := q(e) - 1.1n^2(q(e) - q(x))$. This bound is perfectly acceptable for establishing our complexity result as well.

We now back to the case that $\frac{\rho}{\Delta^k}\|p^k\| < \frac{3}{4}$. In fact, we address a weaker case where $\frac{\rho}{\Delta^k}\|p^k\| < 1$, that is,

$$\|\frac{\rho}{\Delta^k}X^k s(\mu^k) - e\| < 1.$$

First, we must have

$$s(\mu^k) = Q(x^k + d_x) + c - A^T y(\mu^k) = Qx^{k+1} + c - A^T y(\mu^k) > 0.$$

Furthermore,

$$\begin{array}{ll} & \|\frac{\rho}{\Delta^k}X^k s(\mu^k) - e\|^2 \\ = & (\frac{\rho}{\Delta^k})^2\|X^k s(\mu^k) - \frac{(x^k)^T s(\mu^k)}{n}e\|^2 + \|\frac{\rho(x^k)^T s(\mu^k)}{n\Delta^k}e - e\|^2 \\ \geq & \|\frac{\rho(x^k)^T s(\mu^k)}{n\Delta^k}e - e\|^2 \\ \geq & \left(\frac{\rho(x^k)^T s(\mu^k)}{n\Delta^k} - 1\right)^2 n. \end{array}$$

Hence, $\frac{\rho}{\Delta^k}\|p^k\| < 1$ implies

$$\frac{n-\sqrt{n}}{\rho} \le \frac{(x^k)^T s(\mu^k)}{\Delta^k} \le \frac{n+\sqrt{n}}{\rho}.$$

Moreover,

$$\begin{array}{rcl} (x^{k+1})^T s(\mu^k) & = & (x^k)^T (X^k)^{-1} X^{k+1} s(\mu^k) \\ & \le & \|(X^k)^{-1} X^{k+1}\| (x^k)^T s(\mu^k) \\ & \le & (1+\beta)(x^k)^T s(\mu^k) \le 2(x^k)^T s(\mu^k). \end{array}$$

Therefore, we have

$$\frac{(x^{k+1})^T s(\mu^k)}{\Delta^k} \le \frac{2(n+\sqrt{n})}{\rho} \le \frac{\epsilon}{n^2}$$

or

$$\frac{(x^{k+1})^T s(\mu^k)}{q(x^k) - q(x^0) + n^2(q(x^0) - q(\hat{x}))} \le \frac{\epsilon}{n^2}.$$

Consequently, if $q(x^k) \ge q(x^0)$ then

$$\frac{(x^{k+1})^T s(\mu^k)}{n^2(q(x^k) - q(\hat{x}))} \le \frac{\epsilon}{n^2};$$

otherwise

$$\frac{(x^{k+1})^T s(\mu^k)}{n^2(q(x^0) - q(\hat{x}))} \le \frac{\epsilon}{n^2}.$$

Both of them imply that

$$\frac{(x^{k+1})^T s(\mu^k)}{\bar{z} - q(\hat{x})} \le \epsilon,$$

which further implies that

$$\frac{(x^{k+1})^T s(\mu^k)}{\bar{z} - \underline{z}} \le \epsilon.$$

That is, x^{k+1} is an ϵ-KKT point for (QP). To summarize, we have

Theorem 9.18 *Let z and ρ be chosen as above and let $\beta < 1$ in condition (9.26). Then, $\mu^k\|d'_x\| = \|p^k\| < \frac{\Delta^k}{\rho}$ implies that $x^{k+1} := x^k + X^k d'_x$ is an ϵ-KKT point for (QP).*

9.4.3 Solving the ball-constrained QP problem

We now analyze solving (BQP), or (BHP) equivalently, in each iteration of the algorithm. Consider the necessary and sufficient conditions (9.26) or (9.27). It has been shown that μ^k is unique in these conditions and

$$\mu^k \leq |\underline{\lambda}(H^k)| + \frac{\|g^k\|}{\beta} \leq |\underline{\lambda}(Q^k)| + \frac{\|c^k\|}{\beta}.$$

It is also known that any eigenvalue

$$|\lambda(Q^k)| \leq n \max\{|q^k_{ij}|\},$$

where q^k_{ij} is the (i,j)th component of matrix Q^k. Thus,

$$0 \leq \mu^k \leq R^k := n \max\{|q^k_{ij}|\} + \frac{\|c^k\|}{\beta},$$

where R^k is a computable upper bound at each iteration.

Since $Q^k = X^k Q X^k$, therefore

$$\max\{|q^k_{ij}|\} \leq \max\{|q_{ij}|\} \max\{x^k_i x^k_j\} \leq n^2 \max\{|q_{ij}|\},$$

which implies that

$$|\underline{\lambda}(Q^k)| \leq n^3 \max\{|q_{ij}|\}.$$

Similarly, we can show

$$\|c^k\| \leq n^{3.5} \max\{|q_{ij}|\} + n\|c\| + (\bar{z} - \underline{z}).$$

Let

$$R := \frac{(n^{3.5} + n^3) \max\{|q_{ij}|\} + n\|c\| + (\bar{z} - \underline{z})}{\beta(q(e) - z)}. \tag{9.33}$$

Then, R^k is uniformly bounded by $R(q(e) - z)$, which only depends on the problem data. Note that R is invariant if the objective function or (Q, c) is scaled by a positive number.

For any given μ, denote solutions of the top linear equations by $d'_x(\mu)$ in conditions (9.26) and $v(\mu)$ in conditions (9.27). It can be also shown that $\|d'_x(\mu)\| = \|v(\mu)\|$ is a decreasing function for any $\mu \geq |\lambda^k|$. Besides, for any given μ we can check to see if $\mu \geq |\lambda^k|$ by checking the definiteness of matrix $H^k + \mu I$, which can be solved as a LDL^T decomposition. These facts lead to a bisection method to search for the root of $\|d'_x(\mu)\| = \|v(\mu)\| = \beta$ while $\mu \in [|\lambda^k|, R^k] \subset [0, R^k]$. Obviously, for a given $\epsilon' \in (0,1)$, a μ, such that $0 \leq \mu - \mu^k \leq \epsilon'$, can be obtained in $\log(R^k/\epsilon') \leq \log(R(q(e) - z)/\epsilon')$ bisection steps, and the cost of each step is $O(n^3)$ arithmetic operations.

The remaining question is what ϵ' would be sufficient to carry out our main algorithm. In what follows we let

$$\Delta^k = q(x^k) - z \geq \epsilon(q(e) - z), \tag{9.34}$$

since otherwise x^k is already an ϵ-minimizer for (QP). We discuss the condition of ϵ' in three cases.

(A) The termination iteration.

Let us consider the termination iteration where

$$\|p^k\| < \frac{3\Delta^k}{4\rho}.$$

From relation (9.29) we see

$$0 < \mu^k < \frac{3\Delta^k}{4\rho\beta}.$$

Let the right endpoint of the interval generated by the bisection search be denoted as μ. If we require μ such that

$$0 \leq \mu - \mu^k \leq \frac{\Delta^k}{4\rho\beta},$$

then

$$\mu^k \leq \mu < \frac{\Delta^k}{\rho\beta}.$$

Also note $\|d'_x(\mu)\| \leq \beta$ since $\mu \geq \mu^k$. This leads to that

$$\|\mu d'_x(\mu)\| \leq \frac{\Delta^k}{\rho},$$

and, from Theorem 9.18, $x^+ := x^k + X^k d'_x(\mu)$ must be an ϵ-KKT point. Thus, noticing the choice of ρ, $\beta = \frac{1}{4}$ and (9.34), in this case we can choose

$$\epsilon' \leq \frac{\epsilon^2(q(e) - z)}{2n^2(n + \sqrt{n})} \tag{9.35}$$

to meet the requirement.

(B) The non-termination iteration.

Now let us consider the non-termination iteration, i.e.,

$$\|p^k\| \geq \frac{3\Delta^k}{4\rho}.$$

In this iteration from (9.29) and (9.30) we have

$$\mu^k \geq \frac{3\Delta^k}{4\rho\beta} \quad \text{and} \quad q'(d'_x(\mu^k)) \leq -\frac{3\Delta^k\beta}{8\rho}.$$

Thus, if we could generate a d'_x such that

$$q'(d'_x) - q'(d'_x(\mu^k)) \leq \frac{3\Delta^k\beta}{24\rho} \leq \frac{\beta\|p^k\|}{6} = \frac{\beta^2\mu^k}{6} \tag{9.36}$$

and

$$\|d'_x\| \leq \beta,$$

then

$$q'(d'_x) \leq -\frac{\beta\|p^k\|}{2} + \frac{\beta\|p^k\|}{6} = -\frac{\beta\|p^k\|}{3},$$

which implies that

$$\frac{\rho}{\Delta}(\frac{1}{2}d_x^TQd_x + (Qx^k + c - \frac{\Delta^k}{\rho}X^{-1}e)^Td_x) \leq -\frac{\beta}{3}\frac{\rho}{\Delta^k}\|p^k\|,$$

and we would still have $x^k + X^kd'_x > 0$ and, from (9.25),

$$\phi(x^k + X^kd'_x) - \phi(x^k) \leq -\frac{2}{96}.$$

In other words, d'_x is an acceptable approximation to $d'_x(\mu^k)$ to reduce the potential function by a constant.

We analyzes the complexity bounds to compute such an approximation. Again, let the right endpoint of the interval generated by the bisection search be denoted as μ. Then, $\mu \geq \mu^k$. If $\mu = \mu^k$, then we get an exact solution. Thus, we assume $\mu > \mu^k \geq |\lambda^k|$. We consider two sub-cases,

(B.1) in the non-termination iteration. In this case, we let

$$\mu^k \geq |\lambda^k| + 5\epsilon'.$$

Note that

$$\begin{aligned}
&\|d'_x(\mu^k)\|^2 - \|d'_x(\mu)\|^2 \\
&= \|v(\mu^k)\|^2 - \|v(\mu)\|^2 \\
&= v^T(\mu^k)(I - (H^k + \mu^kI)(H^k + \mu I)^{-2}(H^k + \mu^kI))v(\mu^k) \\
&= v^T(\mu^k)\left(2(\mu - \mu^k)(H^k + \mu I)^{-1} - (\mu - \mu^k)^2(H^k + \mu I)^{-2}\right)v(\mu^k) \\
&\leq \|v(\mu^k)\|^2\left(\frac{2(\mu - \mu^k)}{(\mu - |\lambda^k|)} - \frac{(\mu - \mu^k)^2}{((\mu - |\lambda^k|))^2}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \|v(\mu^k)\|^2 \left(\frac{2(\mu-\mu^k)}{(\mu-\mu^k)+(\mu^k-|\lambda^k|)} - \frac{(\mu-\mu^k)^2}{((\mu-\mu^k)+(\mu^k-|\lambda^k|))^2} \right) \\
&= \beta^2 \frac{(\mu-\mu^k)^2 + 2(\mu-\mu^k)(\mu^k-|\lambda^k|)}{((\mu-\mu^k)+(\mu^k-|\lambda^k|))^2} \\
&= \beta^2 \left(1 - \frac{(\mu^k-|\lambda^k|)^2}{((\mu-\mu^k)+(\mu^k-|\lambda^k|))^2} \right) \\
&\leq \beta^2 \left(1 - \frac{(5\epsilon')^2}{((\mu-\mu^k)+5\epsilon')^2} \right).
\end{aligned}
$$

In $O(\log(R^k/\epsilon'))$ bisection steps, we have $\mu - \mu^k \leq \epsilon'$. Then,

$$\|d'_x(\mu^k)\|^2 - \|d'_x(\mu)\|^2 \leq \frac{11\beta^2}{36}.$$

On the other hand,

$$
\begin{aligned}
&q'(d'_x(\mu)) - q'(d'_x(\mu^k)) \\
&\quad = \frac{1}{2}v(\mu)^T H^k v(\mu) + (g^k)^T v(\mu) - \frac{1}{2}v(\mu^k)^T H^k v(\mu^k) - (g^k)^T v(\mu^k) \\
&\quad = \frac{1}{2}(H^k v(\mu) + g^k)^T (v(\mu) - v(\mu^k)) + \frac{1}{2}(H^k v(\mu^k) + g^k)^T (v(\mu) - v(\mu^k)) \\
&\quad = -\frac{1}{2}\mu v(\mu)^T (v(\mu) - v(\mu^k)) - \frac{1}{2}\mu^k v(\mu^k)^T (v(\mu) - v(\mu^k)) \\
&\quad = -\frac{1}{2}(\mu-\mu^k) v(\mu)^T (v(\mu) - v(\mu^k)) - \frac{1}{2}\mu^k (\|v(\mu)\|^2 - \|v(\mu^k)\|^2) \\
&\quad \leq \epsilon'\beta^2 + \frac{11\beta^2\mu^k}{72}.
\end{aligned}
$$

Thus, if we condition

$$\epsilon'\beta^2 \leq \frac{1}{72}\frac{3\Delta^k\beta}{4\rho} \leq \frac{\beta^2\mu^k}{72},$$

then (9.36) holds, i.e.,

$$q'(d'_x(\mu)) - q'(d'_x(\mu^k)) \leq \frac{\beta^2\mu^k}{72} + \frac{11\beta^2\mu^k}{72} = \frac{\beta^2\mu^k}{6}$$

and $d'_x(\mu)$ will be an acceptable approximation to $d'_x(\mu^k)$. Note that choosing

$$\epsilon' \leq \frac{\epsilon^2(q(e)-z)}{48n^2(n+\sqrt{n})} \tag{9.37}$$

will meet the condition, due to (9.34) and the selection of ρ and β.

(B.2) in the non-termination iteration. In this case, we have

$$\mu^k < |\lambda^k| + 5\epsilon'.$$

Thus, in $O(\log(R^k/\epsilon'))$ bisection steps, we have $\mu - \mu^k < \epsilon'$ so that $\mu - |\lambda^k| < 6\epsilon'$. However, unlike Case (B.1) we find $d'_x(\mu)$ (or $v(\mu)$) is not a sufficient approximation to $d'_x(\mu^k)$ (or $v(\mu^k)$). When we observe this fact, we do the following computation.

Let q^k, $\|q^k\| = 1$, be an eigenvector corresponding the λ^k, the least eigenvalue of matrix H^k. Then, one of the unit vector e_j, $j = 1, ..., n-m$, must have $|e_j^T q^k| \geq 1/\sqrt{n-m}$. (In fact, we can use any unit vector q to replace e_j as long as $q^T q^k \geq 1/\sqrt{n-m}$.) Now solve for y from

$$(H^k + \mu I)y = e_j$$

and let

$$v = v(\mu) + \alpha y,$$

where α is chosen such that $\|v\| = \beta$. Note we have

$$(H^k + \mu I)v = -g^k + \alpha e_j,$$

and in the computation of $v(\mu)$ and y, matrix $H^k + \mu I$ only needs to be factorized once.

We have that

$$\|y\| \geq \frac{1}{\sqrt{n-m}(\mu - |\lambda^k|)}$$

and

$$|\alpha| \leq 2\beta(\mu - |\lambda^k|)\sqrt{n-m} \leq 2\beta(6\epsilon')\sqrt{n-m}.$$

Let $d'_x = N^k v$. Then, we have

$$\begin{aligned}
&q'(d'_x) - q'(d'_x(\mu^k)) \\
&= \frac{1}{2}v^T H^k v + (g^k)^T v - \frac{1}{2}v(\mu^k)^T H^k v(\mu^k) - (g^k)^T v(\mu^k) \\
&= \frac{1}{2}(H^k v + g^k)^T (v - v(\mu^k)) + \frac{1}{2}(H^k v(\mu^k) + g^k)^T (v - v(\mu^k)) \\
&= \frac{1}{2}(H^k v + g^k - \alpha e_j)^T (v - v(\mu^k)) + \frac{1}{2}\alpha e_j^T (v - v(\mu^k)) \\
&\quad -\frac{1}{2}\mu^k v(\mu^k)^T (v - v(\mu^k)) \\
&= -\frac{1}{2}\mu v^T (v - v(\mu^k)) + \frac{1}{2}\alpha e_j^T (v - v(\mu^k)) - \frac{1}{2}\mu^k v(\mu^k)^T (v - v(\mu^k)) \\
&= -\frac{1}{2}(\mu v + \mu^k v(\mu^k))^T (v - v(\mu^k)) + \frac{1}{2}\alpha e_j^T (v - v(\mu^k))
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2}(\mu - \mu^k) v^T (v - v(\mu^k)) + \frac{1}{2}\alpha e_j^T (v - v(\mu^k)) \\
&\le \epsilon' \beta^2 + 2(6\epsilon')\beta^2 \sqrt{n-m}.
\end{aligned}
$$

Thus, if we condition

$$\epsilon' \beta^2 + 12\epsilon' \beta^2 \sqrt{n-m} \le \frac{3\Delta^k \beta}{24\rho},$$

then (9.36) holds, i.e.,

$$q'(d_x') - q'(d_x'(\mu^k)) \le \frac{3\Delta^k \beta}{24\rho} \le \frac{\beta \|p^k\|}{6},$$

and d_x' must be an acceptable approximate to $d_x'(\mu^k)$. Note that choosing

$$\epsilon' \le \frac{\epsilon^2 (q(e) - z)}{52 n^2 (n + \sqrt{n})\sqrt{n-m}}, \tag{9.38}$$

will meet the condition, due to (9.34) and the selection of ρ and β.

Comparing (9.35), (9.37), and (9.38), we choose

$$\epsilon' = \frac{\epsilon^2 (q(e) - z)}{52 n^2 (n + \sqrt{n})\sqrt{n-m}},$$

will meet the conditions for all three cases. Hence, our bisection method will terminate in at most $O(\log(R/\epsilon) + \log n)$ steps, where R is defined by (9.33), and it either finds an ϵ-KKT point in Case A of the termination iteration, or calculates a sufficient approximation $d_x'(\mu)$ in Case (B.1) of the non-termination iteration, or spends additional $O((n-m)^3)$ arithmetic operations to generate a sufficient approximation d_x' in Case (B.2) of the non-termination iteration. Normally, $n < R/\epsilon$. Thus, the running time of the bisection method in each iteration is bounded by $O(n^3 \log(R/\epsilon))$ arithmetic operations. To summarize

Theorem 9.19 *The total running time of the potential reduction algorithm is bounded by*

$$O\left(\left(\frac{n^6}{\epsilon}\log\frac{1}{\epsilon} + n^4 \log n\right)\log\frac{R}{\epsilon}\right)$$

arithmetic operations.

One comment is about z, the lower bound for $\underline{z}$, used in our algorithm. If we somehow know $\underline{z}$ (we probably do in solving many combinatorial optimization problems), then we can choose $\rho = 2(n + \sqrt{n})/\epsilon$ and reduce the overall complexity bound by a factor n^2.

Finally, when $\epsilon \to 0$, we must have $\mu^k \to 0$, so that $|\lambda^k| \to 0$. At the limit, λ^k represent the least eigenvalue of Q in the null space of all active constraints: $Ax = b$ plus $x_j = 0$ for every j that $x_j^k \to 0$. This implies that Q is positive semi-definite in the null space of all active constraints of the limit. This is the second order necessary condition for the limit being a local minimal solution.

9.5 Notes

In this chapter we have extended a potential reduction algorithm to solving the general linear complementarity problem and quadratic programming. We have shown that the algorithm is a fully polynomial-time approximation scheme for computing an ϵ-approximate stationary or KKT point, which itself is a (nonconvex) linear complementarity problem. (The concept of the fully polynomial-time approximation scheme (FPTAS) was introduced in combinatorial optimization, for example, see Papadimitriou and Steiglitz [272].) The result is the first approximation algorithm whose running time is almost linear in $\frac{1}{\epsilon}$, which was an open question in the area of nonlinear optimization complexity, see Vavasis [359]. We would also like to mention that algorithms, similar to the one described in this paper, have actually been used in practice, and they seem to work very well (e.g., Kamarth, Karmarkar, Ramakrishnan, and Resende [170, 171]).

If (QP) is a convex optimization problem, then it can be solved in polynomial time, e.g., see Vavasis [359] and references therein. If Q have at least one negative eigenvalue in the null space of A, then (QP) becomes a hard problem–an NP-complete problem. ([101], [273], [294], [359], and etc). Some proposed algorithms for solving general QP problems include the principal pivoting method of Lemke-Cottle-Dantzig (e.g., [71]), the active-set method (e.g., [105]), the interior-point algorithm (e.g., [170, 171] and [376]), and other special-case methods (e.g., [249] and [273]). Other interior-point methods for nonconvex optimization can be found in Bonnans and Bouhtou [58] and Pardalos and Ye [376, 385].

Even finding a local minimal and checking the existence of a KKT point are NP-complete problems (see, e.g., Murty and Kabadi [251], Horst, Pardalos and Thoai [150], Johnson, Papadimitriou and Yannakakis [166], and Pardalos and Jha [274]). Finding even an ϵ-minimal or ϵ-KKT point are hard problems. Bellare and Rogaway [41] showed that there exists a constant, say $\frac{1}{4}$, such that no polynomial-time algorithm exists to compute an $\frac{1}{4}$-minimal solution for (QP), unless $P = NP$. Vavasis [360] and Ye [376] developed a polynomial-time algorithm to compute an $(1-\frac{1}{n^2})$-minimal solution. Using a steepest-descent-type method, Vavasis [359] also proved an arithmetic operation upper bound, $O(n^3(\frac{R}{\epsilon})^2)$, for computing an ϵ-KKT point of a box-constrained QP problem, where R is a fixed number depending on the problem data.

The LCP potential algorithm described in this chapter is due to Kojima, Megiddo and Ye [181]. Other algorithms can be found in Kojima, Mizuno and Noma [182] and Noma [268]. The description of $\mathcal{G}$ is technically similar to Eaves' class and Garcia's class. These two classes and some others have been extensively studied, see Cottle, Pang and Stone [72]. Since every KKT point of the LCP with a row-sufficient matrix is a solution (Cottle et al. [72]), the algorithm is actually a polynomial approximation algorithm for solving the class of LCPs with row-sufficient matrices.

Now consider (BQP), or (BHP) equivalently. First, a brief history of this problem. There is a class of nonlinear programming algorithms called model trust region methods. In these algorithms, a quadratic function is used as an approximate model of the true objective function around the current iterate. The next step is to minimize the model function. In general, however, the model is expected to be accurate or trusted only in a neighborhood of the current iterate. Accordingly, the quadratic model is minimized in a 2-norm neighborhood, which is a ball, around the current iterate.

The model-trust region problem, (BQP), is due to Levenberg [196] and Marquardt [209]. These authors considered only the case when Q^k is positive definite. Moré [248] proposed an algorithm with a convergence proof for this case. Gay [102] and Sorenson [309] proposed algorithms for the general case, also see Dennis and Schnable [76]. These algorithms work very well in practice, but no complexity result was established for this problem then.

A simple polynomial bisection method was proposed by Ye [376] and Vavasis and Zippel [361]. Recently, Rendl and Wolkowicz [285] showed that (BQP) can be reformulated as a positive-semidefinite problem, which is a *convex* nonlinear problem. There are polynomial interior-point algorithms (see Nesterov and Nemirovskii [263]) to compute an d'_x such that $q'(d'_x) - q(d'_x(\mu^k)) \le \epsilon'$ in $O(n^3\log(R^k/\epsilon'))$ arithmetic operations. This will also establish an

$$O((\frac{n^6}{\epsilon}\log\frac{1}{\epsilon} + n^4\log n)\log\frac{R}{\epsilon})$$

arithmetic operation bound for our algorithm.

In addition, Ye [380] further developed a Newton-type method for solving (BQP) and established an arithmetic operation bound $O(n^3\log(\log(R^k/\epsilon')))$ to yield a μ such that $0 \le \mu - \mu^k \le \epsilon'$. The method can be adapted into our each iteration. We first find an approximate $\underline{\mu}$ to the absolute value of the least eigenvalue $|\lambda^k|$ and an approximate eigenvector q to the true q^k, such that $0 \le \underline{\mu} - |\lambda^k| \le \epsilon'$ and $q^T q^k \ge 1/\sqrt{n-m}$. This approximation can be done in $O(\log(\log(R^k/\epsilon')))$ arithmetic operations. Then, we will use q to replace e_j in Case (B.2) of the non-termination iteration (i.e., $\|v(\underline{\mu})\| < \beta$) to enhance $v(\underline{\mu})$ and generate a desired approximation. Otherwise, we know $\mu^k > \underline{\mu}$ and, using the method in Ye [380], we will generate a $\mu \in (\underline{\mu}, R^k)$ such that $|\mu - \mu^k| \le \epsilon'$ in $O(n^3 \log(\log(R^k/\epsilon')))$ arithmetic operations. This shall establish an

$$O((\frac{n^6}{\epsilon}\log\frac{1}{\epsilon} + n^4 \log n)\log(\log\frac{R}{\epsilon}))$$

arithmetic operation bound for our algorithm.

9.6 Exercises

9.1 *Verify inequality (9.4).*

9.2 *For any $x, s \in \overset{\circ}{\mathcal{R}}{}^n_+$, prove that*

$$H(x,s) = 2I - (XM^T - S)(S^2 + MX^2M^T)^{-1}(MX - S)$$

is positive semi-definite.

9.3 *Prove Proposition 9.2.*

9.4 *In Proposition 9.4 show that for any $(x,s) \in \overset{\circ}{\mathcal{F}}$*

$$\|g(x,s)\|_H^2 > 0.$$

9.5 *Given $r > 0$ and let $d(r)$ be the minimizer for*

$$\begin{array}{ll} \textit{minimize} & q(d) := \frac{1}{2}d^TQd + c^Td \\ \textit{subject to} & Ad = 0, \\ & \|d\|^2 \le r^2. \end{array}$$

Then, for $0 < r \le R$

$$q(0) - q(d(r)) \ge \frac{r^2}{R^2}(q(0) - q(d(R))).$$

9.6 *Given $\underline{z}$, the exact global minimal objective value of a QP problem, develop an approximation algorithm for finding an KKT point of the problem and analyze its complexity.*

Bibliography

[1] S. S. Abhyankar, T. L. Morin, and T. B. Trafalis. Efficient faces of polytopes : Interior point algorithms, parameterization of algebraic varieties, and multiple objective optimization. In J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming*, volume 114 of *Contemporary Mathematics*, pages 319–341. American Mathematical Society, Providence, RI, 1990.

[2] I. Adler, R. M. Karp and R. Shamir. A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivot steps. *J. Complexity*, 3:372–387, 1987.

[3] I. Adler and N. Megiddo. A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *Journal of the ACM*, 32:871–895, 1985.

[4] I. Adler, N. K. Karmarkar, M. G. C. Resende, and G. Veiga. An implementation of Karmarkar's algorithm for linear programming. *Mathematical Programming*, 44:297–335, 1989. Errata in *Mathematical Programming*, 50:415, 1991.

[5] I. Adler and R. D. C. Monteiro. Limiting behavior of the affine scaling continuous trajectories for linear programming problems. *Mathematical Programming*, 50:29–51, 1991.

[6] I. Adler and R. D. C. Monteiro. A geometric view of parametric linear programming. *Algorithmica*, 8:161–176, 1992.

[7] M. Akgül. On the exact solution of a system of linear homogeneous equations via a projective algorithm. *Arabian Journal for Science and Engineering*, 15(4):753–754, 1990.

[8] F. Alizadeh. *Combinatorial optimization with interior point methods and semi–definite matrices.* PhD thesis, University of Minnesota, Minneapolis, MN, 1991.

[9] F. Alizadeh. Optimization over the positive semi-definite cone: Interior–point methods and combinatorial applications. In P. M. Pardalos, editor, *Advances in Optimization and Parallel Computing*, pages 1–25. North Holland, Amsterdam, The Netherlands, 1992.

[10] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5:13–51, 1995.

[11] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. New York University Computer Science Dept Report 721, 1996.

[12] A. Altman and J. Gondzio. An efficient implementation of a higher order primal–dual interior point method for large sparse linear programs. Technical Report ZTSW–1/A214/92, Systems Research Institute, Polish Academy of Sciences, Newelska6, PL–01–447 Warshaw, Poland, June 1992. Revised September 1992. To appear in *Archives of Control Sciences.*

[13] E. D. Andersen and Y. Ye. Combining interior-point and pivoting algorithms for linear programming. Working Paper, Department of Management Sciences, The University of Iowa, IA, 1994.

[14] E. D. Andersen and Y. Ye. On a homogeneous algorithm for the monotone complementarity problem. Working Paper, Department of Management Sciences, The University of Iowa, IA, 1995.

[15] K. D. Andersen. An infeasible dual affine scaling method for linear programming. *COAL Newsletter* 22:19–27, 1993.

[16] K. D. Andersen and E. Christiansen. A newton barrier method for minimizing a sum of Euclidean norms subject to linear equality constraints. Preprint 95-07, Department of Mathematics and Computer Science, Odense University, Denmark, 1995.

[17] K. M. Anstreicher. A monotonic projective algorithm for fractional linear programming. *Algorithmica*, 1(4):483–498, 1986.

[18] K. M. Anstreicher. On long step path following and SUMT for linear and quadratic programming. Technical Report, Yale School of Management, Yale University, New Haven, CT, 1990.

[19] K. M. Anstreicher. A standard form variant and safeguarded linesearch for the modified Karmarkar algorithm. *Mathematical Programming*, 47:337–351, 1990.

[20] K. M. Anstreicher. A combined phase I – phase II scaled potential algorithm for linear programming. *Mathematical Programming*, 52:429–439, 1991.

[21] K. M. Anstreicher. Analysis of interior point algorithms for linear programming using a volumetric barrier. Working Paper, Department of Management Science, University of Iowa, Iowa City, IA, 1994.

[22] K. M. Anstreicher. On the performance of Karmarkar's algorithm over a sequence of iterations. *SIAM Journal on Optimization*, 1(1):22–29, 1991.

[23] K. M. Anstreicher and R. A. Bosch. Long steps in a $O(n^3L)$ algorithm for linear programming. *Mathematical Programming*, 54:251–265, 1992.

[24] K. M. Anstreicher, D. den Hertog, C. Roos, and T. Terlaky. A long step barrier method for convex quadratic programming. Technical Report 90–53, Faculty of Mathematics and Informatics, TU Delft, NL–2628 BL Delft, The Netherlands, 1990. To appear in *Algorithmica*.

[25] K. M. Anstreicher, J. Ji, F. A. Potra, and Y. Ye. Average performance of a self–dual interior–point algorithm for linear programming. in P. Pardalos, editor, *Complexity in Numerical Optimization* page 1–15, World Scientific, New Jersey, 1993.

[26] K. M. Anstreicher, J. Ji, F. A. Potra, and Y. Ye. Probabilistic analysis of an infeasible primal–dual algorithm for linear programming. Reports on Computational Mathematics 27, Dept. of Mathematics, University of Iowa, Iowa City, IA, 1992.

[27] K. M. Anstreicher and J. P. Vial. On the convergence of an infeasible primal–dual interior–point method for convex programming. Technical Report 93–34, Faculty of Technical Mathematics and Informatics, TU Delft, NL–2628 BL Delft, The Netherlands, 1993.

[28] M. D. Asic and V. V. Kovacevic-Vujcic. An interior semi–infinite programming method. *Journal of Optimization Theory and Applications*, 59:369–390, 1988.

[29] M. D. Asic, V. V. Kovacevic-Vujcic, and M. D. Radosavljevic-Nikolic. Asymptotic behavior of Karmarkar's method for linear programming. *Mathematical Programming*, 46:173–190, 1990.

[30] G. Astfalk, I. J. Lustig, R. E. Marsten, and D. Shanno. The interior–point method for linear programming. *IEEE Software*, 9(4):61–68, 1992.

[31] D. S. Atkinson and P. M. Vaidya. A cutting plane algorithm for convex programming that uses analytic centers. *Mathematical Programming*, 69:1–44, 1995.

[32] O. Bahn, J. L. Goffin, J. P. Vial, and O. Du Merle. Implementation and behavior of an interior point cutting plane algorithm for convex programming: An application to geometric programming. *Discrete Applied Mathematics*, 49:3–23, 1993.

[33] O. Bahn, O. Du Merle, J. L. Goffin, and J. P. Vial. A cutting plane method from analytic centers for stochastic programming. *Mathematical Programming*, 69:45–74, 1995.

[34] I. Bárány and Füredi. Computing the volume is difficult. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, page 442-447, ACM, New York, 1986.

[35] E. R. Barnes. A variation on Karmarkar's algorithm for solving linear programming problems. *Mathematical Programming*, 36:174–182, 1986.

[36] E. R. Barnes, S. Chopra, and D. J. Jensen. The affine scaling method with centering. Technical Report, Dept. of Mathematical Sciences, IBM T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY, 1988.

[37] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming, Part I: Affine and projective scaling trajectories. *Transactions of the American Mathematical Society*, 314(2):499–526, 1989.

[38] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming, Part II: Legendre transform coordinates. *Transactions of the American Mathematical Society*, 314(2):527–581, 1989.

[39] D. A. Bayer and J. C. Lagarias. Karmarkar's linear programming algorithm and Newton's method. *Mathematical Programming*, 50:291–330, 1991.

[40] M. S. Bazaraa, J. J. Jarvis, and H. F. Sherali. *Linear Programming and Network Flows*, chapter 8.4 : Karmarkar's projective algorithm, pages 380–394, chapter 8.5 : Analysis of Karmarkar's algorithm, pages 394–418. John Wiley & Sons, New York, second edition, 1990.

[41] M. Bellare and P. Rogaway. The complexity of approximating a nonlinear program. *Mathematical Programming*, 69:429-442, 1995.

[42] M. Ben–Daya and C. M. Shetty. Polynomial barrier function algorithm for convex quadratic programming. *Arabian Journal for Science and Engineering*, 15(4):657–670, 1990.

[43] A. Ben–Tal and A. S. Nemirovskii. Interior point polynomial time method for truss topology design. *SIAM Journal on Optimization*, 4:596–612, 1994.

[44] D. P. Bertsekas and J. Tsitsiklis. *Linear Programming.* Athena Scientific, Belmont, MA, 1996.

[45] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, MA, 1995.

[46] D. Bertsimas and X. Luo. On the worst case complexity of potential reduction algorithms for linear programming. Working Paper 3558-93, Sloan School of Management, MIT, Cambridge, MA, 1993.

[47] U. Betke and P. Gritzmann. Projection algorithms for linear programming. *European Journal of Operational Research*, 60:287–295, 1992.

[48] J. R. Birge and D. F. Holmes. Efficient solution of two stage stochastic linear programs using interior point methods. *Computational Optimization and Applications*, 1:245–276, 1992.

[49] J. R. Birge and L. Qi. Computing block–angular Karmarkar projections with applications to stochastic programming. *Management Science*, 34:1472–1479, 1988.

[50] R. E. Bixby, J. W. Gregory, I. J. Lustig, R. E. Marsten, and D. F. Shanno. Very large–scale linear programming : A case study in combining interior point and simplex methods. *Operations Research*, 40:885–897, 1992.

[51] R. E. Bixby and M. J. Saltzman. Recovering an optimal LP basis from an interior point solution. *Operations Research Letters* 15,:169–178, 1994.

[52] C. E. Blair. The iterative step in the linear programming algorithm of N. Karmarkar. *Algorithmica*, 1(4):537–539, 1986.

[53] R. G. Bland, D. Goldfarb and M. J. Todd. The ellipsoidal method: a survey. *Operations Research* 29:1039–1091, 1981.

[54] L. Blum. Towards an asymptotic analysis of Karmarkar's algorithm. *Information Processing Letters*, 23:189–194, 1986.

[55] L. Blum. A new simple homotopy algorithm for linear programming I. *Journal of Complexity*, 4:124–136, 1988.

[56] L. Blum, M. Shub and S. Smale. On a theory of computations over the real numbers: NP-completeness, recursive functions and universal machines. *Proc. 29th Symp. Foundations of Computer Science*, 387–397, 1988.

[57] P. T. Boggs, P. D. Domich, J. R. Donaldson, and C. Witzgall. Algorithmic enhancements to the method of center for linear programming. *ORSA Journal on Computing*, 1:159–171, 1989.

[58] F. Bonnans and M. Bouhtou. The trust region affine interior point algorithm for convex and nonconvex quadratic programming. Technical Report, Institute National de Recherche en Informatique et Automatique (INRIA), F–78153 Roquencourt, France, 1993.

[59] F. Bonnans and C. Gonzaga. Technical Report, Institute National de Recherche en Informatique et Automatique (INRIA), F–78153 Roquencourt, France, 1994.

[60] K. H. Borgwardt. *The Simplex Method: A Probabilistic Analysis.* Springer, Berlin, 1987.

[61] S. Boyd and L. E. Ghaoui. Methods of centers for minimizing generalized eigenvalues. *Linear Algebra and Its Applications, special issue on Linear Algebra in Systems and Control*, 188:63–111, 1993.

[62] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Science.* SIAM Publications. SIAM, Philadelphia, 1994.

[63] T. J. Carpenter, I. J. Lustig, J. M. Mulvey, and D. F. Shanno. Higher order predictor–corrector interior point methods with application to quadratic objectives. *SIAM Journal on Optimization*, 3:696–725, 1993.

[64] T. M. Cavalier and A. L. Soyster. Some computational experience and a modification of the Karmarkar algorithm. Working Paper 85–105, Dept. of Industrial and Management Systems Engineering, Pennsylvania State University, University Park, PA, 1985.

[65] I. C. Choi and D. Goldfarb. Exploiting special structure in a primal–dual path–following algorithm. *Mathematical Programming*, 58:33–52, 1993.

[66] I. C. Choi, C. L. Monma, and D. F. Shanno. Computational experience with a primal–dual interior point method for linear programming. Technical Report, RUTCOR Center of Operations Research, Rutgers University, New Brunswick, NJ, 1989.

[67] I. C. Choi, C. L. Monma, and D. F. Shanno. Further development of a primal–dual interior point method. *ORSA Journal on Computing*, 2:304–311, 1990.

[68] E. Christiansen and K. O. Kortanek. Computation of the collapse state in limit analysis using the LP primal affine scaling algorithm. *Journal of Computational and Applied Mathematics*, 34:47–63, 1991.

[69] S. A. Cook. The complexity of theorem-proving procedures. *Proc. 3rd ACM Symposium on the Theory of Computing*, 151–158, 1971.

[70] T. F. Coleman and Y. Li. A globally and quadratically convergent affine scaling method for linear L_1 problems. *Mathematical Programming*, 52:189–222, 1992.

[71] R. W. Cottle and G. B. Dantzig. Complementary pivot theory in mathematical programming. *Linear Algebra and its Applications*, 1:103-125, 1968.

[72] R. Cottle, J. S. Pang, and R. E. Stone. *The Linear Complementarity Problem*, chapter 5.9 : Interior–point methods, pages 461–475. Academic Press, Boston, 1992.

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

[74] L. Danzer, B. Grünbaum and V. Klee. Helly's theorem and its relatives. in V. Klee, editor, *Convexity*, page 101–180, AMS, Providence, RI, 1963.

[75] J. Dennis, A. M. Morshedi, and K. Turner. A variable metric variant of the Karmarkar algorithm for linear programming. *Mathematical Programming*, 39:1–20, 1987.

[76] J. E. Dennis and R. E. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice–Hall, Englewood Cliffs, New Jersey, 1983.

[77] Z. Y. Diao. Karmarkar's algorithm and its modification. *Chinese Journal on Operations Research*, 7(1):73–75, 1988.

[78] I. I. Dikin. Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk SSSR*, 174:747–748, 1967. Translated in : *Soviet Mathematics Doklady*, 8:674–675, 1967.

[79] I. I. Dikin. On the convergence of an iterative process. *Upravlyaemye Sistemi*, 12:54–60, 1974. (In Russian).

[80] J. Ding and T. Y. Li. A polynomial–time predictor–corrector algorithm for a class of linear complementarity problems. *SIAM Journal on Computing*, 1(1):83–92, 1991.

[81] P. D. Domich, P. T. Boggs, J. E. Rogers, and C. Witzgall. Optimizing over 3–d subspaces in an interior point method. *Linear Algebra and Its Applications*, 152:315–342, 1991.

[82] M. E. Dyer and A. M. Frieze. The complexity of computing the volume of a polyhedron. *SIAM J. Comput.*, 17:967–974, 1988.

[83] M. E. Dyer, A. M. Frieze and R. Kannan. A random polynomial time algorithm for estimating volumes of convex bodies. *J. Assoc. Comp. Mach.*, 38:1–17, 1991.

[84] J. Edmonds. Systems of distinct representatives and linear algebra. *J. Res. Nat. Bur. Standards*, 71B, 1967, pages 241–245.

[85] A. S. El–Bakry, R. A. Tapia, and Y. Zhang. A study of indicators for identifying zero variables in interior–point methods. *Siam Review*, 36:45–72, 1994.

[86] G. Elekes. A geometric inequality and the complexity of computing volume. *Discr. Comput. Geom.*, 1:289–292, 1986.

[87] J. Elzinga and T. G. Moore. A central cutting plane algorithm for the convex programming problem. *Mathematical Programming*, 8:134–145, 1975.

[88] S. C. Fang and S. Puthenpura. *Linear Optimization and Extensions.* Prentice Hall, Englewood Cliffs, NJ, 1994.

[89] L. Faybusovich. Jordan algebras, Symmetric cones and Interior-point methods. Research Report, University of Notre Dame, Notre Dame, 1995.

[90] M. C. Ferris and A. B. Philpott. On affine scaling and semi–infinite programming. *Mathematical Programming*, 56:361–364, 1992.

[91] S. Filipowski. On the complexity of solving feasible systems of linear inequalities specified with approximate data. *Mathematical Programming*, 71:259–288, 1995.

[92] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming : Sequential Unconstrained Minimization Techniques.* John Wiley & Sons, New York, 1968. Reprint : Volume 4 of *SIAM Classics in Applied Mathematics*, SIAM Publications, Philadelphia, PA, 1990.

[93] R. Fourer and S. Mehrotra. Solving symmetric indefinite systems in an interior–point method for linear programming. *Mathematical Programming* 62:15–39, 1993.

[94] R. Freund and F. Jarre. An interior–point method for fractional programs with convex constraints. *Mathematical Programming*, 67:407–440, 1994.

[95] R. Freund and F. Jarre. An interior–point method for convex multi–fractional programming. Technical Report (in preparation), AT&T Bell Laboratories, Murray Hill, NJ, 1993.

[96] R. M. Freund. Polynomial–time algorithms for linear programming based only on primal scaling and projected gradients of a potential function. *Mathematical Programming*, 51:203–222, 1991.

[97] R. M. Freund. A potential-function reduction algorithm for solving a linear program directly from an infeasible "warm start." *Mathematical Programming*, 52:441–466, 1991.

[98] R. M. Freund and J. R. Vera. some characterizations and properties of the 'distance to ill-posedness' and the condition measure of a conic linear system. Technical Report, MIT, 1995.

[99] K. R. Frisch. The logarithmic potential method for convex programming. Unpublished manuscript, Institute of Economics, University of Oslo, Oslo, Norway, 1955.

[100] D. Gale. *The Theory of Linear Economic Models.* McGraw-Hill, New York, 1960.

[101] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman, San Francisco, 1979.

[102] D. M. Gay. Computing optimal locally constrained steps. *SIAM J. Sci. Statis. Comput.*, 2:186-197, 1981.

[103] D. M. Gay. A variant of Karmarkar's linear programming algorithm for problems in standard form. *Mathematical Programming*, 37:81–90, 1987. Errata in *Mathematical Programming*, 40:111, 1988.

[104] G. de Ghellinck and J. P. Vial. A polynomial Newton method for linear programming. *Algorithmica*, 1(4):425–453, 1986.

[105] P. E. Gill, W. M. Murray, and M. H. Wright. *Practical Optimization.* Academic Press, London, 1981.

[106] P. E. Gill, W. Murray, M. A. Saunders, J. A. Tomlin, and M. H. Wright. On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method. *Mathematical Programming*, 36:183–209, 1986.

[107] V. L. Girko. On the distribution of solutions of systems of linear equations with random coefficients. *Theor. Probability and Math. Statist.*, 2:41–44, 1974.

[108] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. Working Paper, MIT, Cambridge, MA, 1994. To appear in *Journal of Assoc. Comput. Mach.*.

[109] J. L. Goffin, A. Haurie, and J. P. Vial. Decomposition and nondifferentiable optimization with the projective algorithm. *Management Science*, 38(2):284–302, 1992.

[110] J. L. Goffin, Z. Q. Luo, and Y. Ye. On the complexity of a column generation algorithm for convex or quasiconvex feasibility problems. In *Large Scale Optimization: State–of–the–Art*, W. W. Hager, D. W. Hearn and P. M. Pardalos (eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993, pages 182–191.

[111] J. L. Goffin, Z. Q. Luo, and Y. Ye. Complexity analysis of an interior-point cutting plane method for convex feasibility problem. To appear in *SIAM J. Optimization*, 1995.

[112] A. V. Goldberg, S. A. Plotkin, D. B. Shmoys, and E. Tardos. Interior–point methods in parallel computation. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, NC, USA, 1989*, pages 350–355. IEEE Computer Society Press, Los Alamitos, CA, USA, 1990.

[113] A. V. Goldberg, S. A Plotkin, D. B. Shmoys, and E. Tardos. Using interior–point methods for fast parallel algorithms for bipartite matching and related problems. *SIAM Journal on Computing*, 21(1):140–150, 1992.

[114] D. Goldfarb and S. Liu. An $O(n^3L)$ primal interior point algorithm for convex quadratic programming. *Mathematical Programming*, 49:325–340, 1990/91.

[115] D. Goldfarb and K. Scheinberg. Interior point trajectories in semidefinite programming. Technical Report, Department of IE/OR, Columbia University, New York, 1996.

[116] D. Goldfarb and M. J. Todd. Linear Programming. In G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, editors, *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science*, pages 141–170. North Holland, Amsterdam, The Netherlands, 1989.

[117] D. Goldfarb and D. Xiao. A primal projective interior point method for linear programming. *Mathematical Programming*, 51:17–43, 1991.

[118] A. J. Goldman and A. W. Tucker. Polyhedral convex cones. In H. W. Kuhn and A. W. Tucker, Editors, *Linear Inequalities and Related Systems*, page 19–40, Princeton University Press, Princeton, NJ, 1956.

[119] G. H. Golub and C. F. Van Loan. *Matrix Computations*, 2nd Edition. Johns Hopkins University Press, Baltimore, MD, 1989.

[120] J. Gondzio. Splitting dense columns of constraint matrix in interior point methods for large scale linear programming. *Optimization*, 24:285–297, 1992.

[121] J. Gondzio. Multiple centrality corrections in a primal-dual method for linear programming. Technical Report 1994.20, Logilab, HEC Geneva, Section of Management Studies, University of Geneva, Bd. Carl-Vogt 102, 1211 Geneva, Switzerland, 1994, revised May 1995.

[122] C. C. Gonzaga. An algorithm for solving linear programming problems in $O(n^3L)$ operations. In N. Megiddo, editor, *Progress in Mathematical Programming : Interior Point and Related Methods*, Springer Verlag, New York, 1989, pages 1–28.

[123] C. C. Gonzaga. Conical projection algorithms for linear programming. *Mathematical Programming*, 43:151–173, 1989.

[124] C. C. Gonzaga. Polynomial affine algorithms for linear programming. *Mathematical Programming*, 49:7–21, 1990.

[125] C. C. Gonzaga. Large steps path–following methods for linear programming, Part I : Barrier function method. *SIAM Journal on Optimization*, 1:268–279, 1991.

[126] C. C. Gonzaga. Large steps path–following methods for linear programming, Part II : Potential reduction method. *SIAM Journal on Optimization*, 1:280–292, 1991.

[127] C. C. Gonzaga. Path following methods for linear programming. *SIAM Review*, 34(2):167–227, 1992.

[128] C. C. Gonzaga and R. A. Tapia. On the convergence of the Mizuno–Todd–Ye algorithm to the analytic center of the solution set. Technical Report 92–36, Dept. of Mathematical Sciences, Rice University, Houston, TX, 1992.

[129] C. C. Gonzaga and R. A. Tapia. On the quadratic convergence of the simplified Mizuno–Todd–Ye algorithm for linear programming. Technical Report 91–41, Dept. of Mathematical Sciences, Rice University, Houston, TX, 1992. Revised November 1992.

[130] C. C. Gonzaga and M. J. Todd. An $O(\sqrt{n}L)$–iteration large–step primal–dual affine algorithm for linear programming. *SIAM Journal on Optimization*, 2:349–359, 1992.

[131] H. J. Greenberg. The use of the optimal partition in a linear programming solution for postoptimal analysis. Technical Report, Mathematics Department, University of Colorado, March 1993.

[132] P. Gritzmann and V. Klee. Mathematical programming and convex geometry. In P. Gruber and J. Wills, editors, *Hand Book of Convex Geometry*, North Holland, Amsterdam, 1993, page 627–674.

[133] M. Grötschel, L. Lovász and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, Berlin, 1988.

[134] B. Grünbaum. *Convex Polytopes*. Wiley, New York, NY, 1967.

[135] B. Grünbaum. Partitions of mass-distributions and of convex bodies by hyperplanes. *Pacific Journal of Mathematics*, 10:1257–1261, 1960.

[136] O. Güler. Limiting behavior of weighted central paths in linear programming. *Mathematical Programming*, 65:347–363, 1994.

[137] O. Güler. Existence of interior points and interior paths in nonlinear monotone complementarity problems. *Mathematics of Operations Research*, 18(1):128–147, 1993.

[138] O. Güler. On the self-concordance of the universal barrier function. Technical Report GU 95-2, Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD, 1995.

[139] O. Güler and L. Tuncel. Characterizations of the barrier parameter of homogeneous convex cones. Research Report CORR 95-14, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, 1995.

[140] O. Güler and Y. Ye. Convergence behavior of interior point algorithms. *Mathematical Programming*, 60:215–228, 1993.

[141] C. G. Han, P. M. Pardalos, and Y. Ye. Implementation of interior–point algorithms for some entropy optimization problems. *Optimization Methods and Software*, 1(1):71–80, 1992.

[142] P. T. Harker and B. Xiao. A polynomial–time algorithm for affine variational inequalities. *Applied Mathematics Letters*, 4(2):31–34, 1991.

[143] J. Hartmanis and R. E. Stearns, On the computational complexity of algorithms. *Trans. A.M.S*, 117: 285–306, 1965.

[144] C. Helmberg, F. Rendl, R. J. Vanderbei and H. Wolkowicz. An interior point method for semidefinite programming. Research Report, University of Waterloo, 1994.

[145] D. den Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming, Algorithms and Complexity.* PhD thesis, Faculty of Mathematics and Informatics, TU Delft, NL–2628 BL Delft, The Netherlands, 1992.

[146] D. den Hertog, F. Jarre, C. Roos, and T. Terlaky. A sufficient condition for self-concordance, with application to some classes of structured convex programming problems. *Mathematical Programming*, 69:75–88, 1995.

[147] D. den Hertog and C. Roos. A survey of search directions in interior point methods for linear programming. *Mathematical Programming*, 52:481–509, 1991.

[148] D. den Hertog, C. Roos, and T. Terlaky. A large–step analytic center method for a class of smooth convex programming problems. *SIAM Journal on Optimization*, 2:55–70, 1992.

[149] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49:263–265, 1952.

[150] R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to Global Optimization.* Kluwer Academic Publishers, Boston, 1995.

[151] S. Huang and K. O. Kortanek. A note on a potential reduction algorithm for LP with simultaneous primal–dual updating. *Operations Research Letters*, 10:501–507, 1991.

[152] P. Huard. Resolution of mathematical programming with nonlinear constraints by the method of centers. In J. Abadie, editor, *Nonlinear Programming*, pages 207–219. North Holland, Amsterdam, The Netherlands, 1967.

[153] P. Huard and B. T. Liêu. La méthode des centres dans un espace topologique. *Numeriche Methematik*, 8:56-67, 1966.

[154] H. Fujita, H. Konno, and K. Tanabe. *Optimization Method.* Iwanami Applied Mathematics Book Series, Iwanami-Shoten, Tokyo, 1994.

[155] P-F. Hung. *An asymptotical $O(\sqrt{n}L)$-Iteration Path-Following Linear Programming Algorithm that Uses Wider Neighborhood and its Implementation.* Ph. D. Thesis, Applied Mathematical and Computational Sciences, University of Iowa, 1994.

[156] M. Iri and H. Imai. A multiplicative barrier function method for linear programming. *Algorithmica*, 1(4):455–482, 1986.

[157] S. Ito, C. T. Kelley, and E. W. Sachs. Inexact primal-dual interior point iteration for linear programs in function spaces. *Computational Optimization and Applications*, 4:189–201, 1995.

[158] B. Jansen. *Interior Point Techniques in Optimization.* PhD thesis, Faculty of Mathematics and Informatics, TU Delft, NL–2628 BL Delft, The Netherlands, 1995.

[159] B. Jansen, C. Roos, and T. Terlaky. A polynomial primal-dual Dikin-type algorithm for linear programming. To appear in *Mathematics of Operations Research*, 1994.

[160] F. Jarre. On the convergence of the method of analytic centers when applied to convex quadratic programs. *Mathematical Programming*, 49:341–358, 1990/91.

[161] F. Jarre. An interior–point method for minimizing the largest eigenvalue of a linear combination of matrices. *SIAM Journal on Control and Optimization* 31:1360–1377, 1993.

[162] F. Jarre. Interior–point methods for convex programming. *Applied Mathematics & Optimization*, 26:287–311, 1992.

[163] J. Ji, F. Potra, and S. Huang. A predictor-corrector method for linear complementarity problems with polynomial complexity and superlinear convergence. Department of Mathematics, University of Iowa, Iowa City, IA, 1991.

[164] J. Ji and Y. Ye. A complexity analysis for interior–point algorithms based on Karmarkar's potential function. *SIAM Journal on Optimization*, 4:512–520, 1994.

[165] F. John. Extremum problems with inequalities as subsidiary conditions. In *Sduties and Essays, presented to R. Courant on his 60th Birthday, January 8, 1948*, Interscience, New York, 1948, page 187-204.

[166] D. S. Johnson, C. Papadimitriou and M. Yannakakis. How easy is local search. *J. Comp. Syst. Sci.*, 37:79-100, 1988.

[167] M. J Kaiser, T. L. Morin and T. B. Trafalis. Centers and invariant points of convex bodies. In P. Gritzmann and B. Sturmfels, editors, *Applied Geometry and Discrete Mathematics: The Victor Klee*

Festschrift, Amer. Math. Soc. and Ass. Comput. Mach., Providence, RI, 1991.

[168] B. Kalantari. Karmarkar's algorithm with improved steps. *Mathematical Programming*, 46:73–78, 1990.

[169] J. A. Kaliski and Y. Ye. A short-cut potential reduction algorithm for linear programming. *Management Science*, 39:757–776, 1993.

[170] A. P. Kamath and N. K. Karmarkar. A continuous method for computing bounds in integer quadratic optimization problems. *Journal of Global Optimization*, 2(3):229–241, 1992.

[171] A. P. Kamath, N. K. Karmarkar, K. G. Ramakrishnan, and M. G. C. Resende. A continuous approach to inductive inference. *Mathematical Programming*, 57:215–238, 1992.

[172] S. Kapoor and P. M. Vaidya. Fast algorithms for convex quadratic programming and multicommodity flows. *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, 1986, pages 147–159.

[173] N. K. Karmarkar. A new polynomial–time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[174] R. M. Karp. Reducibility among combinatorial problems. R. E. Miller and J. W. Thatcher, eds., *Complexity of Computer Computations*, Plenum Press, New York, 1972, pages 85–103.

[175] J. Kemeny, O. Morgenstern, and G. Thompson. A generalization of the von Neumann model of an expanding economy. *Econometrica*, 24:115–135, 1956.

[176] L. G. Khachiyan. A polynomial algorithm for linear programming. *Doklady Akad. Nauk USSR*, 244:1093–1096, 1979. Translated in *Soviet Math. Doklady*, 20:191–194, 1979.

[177] L. G. Khachiyan and M. J. Todd. On the complexity of approximating the maximal inscribed ellipsoid for a polytope. *Mathematical Programming*, 61:137–160, 1993.

[178] V. Klee and G. J. Minty. How good is the simplex method. In O. Shisha, editor, *Inequalities III*, Academic Press, New York, NY, 1972.

[179] M. Kojima, N. Megiddo, and S. Mizuno. A general framework of continuation methods for complementarity problems. *Mathematics of Operations Research*, 18:945–963, 1993.

[180] M. Kojima, N. Megiddo, and S. Mizuno. A primal–dual infeasible–interior–point algorithm for linear programming. *Mathematical Programming*, 61:263–280, 1993.

[181] M. Kojima, N. Megiddo, and Y. Ye. An interior point potential reduction algorithm for the linear complementarity problem. *Mathematical Programming*, 54:267–279, 1992.

[182] M. Kojima, S. Mizuno, and T. Noma. A new continuation method for complementarity problems with uniform P–functions. *Mathematical Programming*, 43:107–113, 1989.

[183] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial–time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–26, 1989.

[184] M. Kojima, S. Mizuno, and A. Yoshise. A primal–dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming : Interior Point and Related Methods*, Springer Verlag, New York, 1989, pages 29–47.

[185] M. Kojima, S. Mizuno, and A. Yoshise. An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems. *Mathematical Programming*, 50:331–342, 1991.

[186] M. Kojima, S. Shindoh, and S. Hara Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. Research Reports on Information Sciences, No. B-282, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Japan, 1994, Revised April 1995.

[187] K. O. Kortanek, F. Potra, and Y. Ye. On some efficient interior point methods for nonlinear convex programming. *Linear Algebra and Its Applications*, 152:169–189, 1991.

[188] K. O. Kortanek and M. Shi. Convergence results and numerical experiments on a linear programming hybrid algorithm. *European Journal of Operational Research*, 32:47–61, 1987.

[189] K. O. Kortanek and J. Zhu. New purification algorithms for linear programming. *Naval Research Logistics Quarterly*, 35:571–583, 1988.

[190] K. O. Kortanek and J. Zhu. A polynomial barrier algorithm for linearly constrained convex programming problems. *Mathematics of Operations Research*, 18(1):116–127, 1993.

[191] V. V. Kovacevic–Vujcic. Improving the rate of convergence of interior point methods for linear programming. *Mathematical Programming*, 52:467–479, 1991.

[192] E. Kranich. Interior–point methods bibliography. *SIAG/OPT Views–and–News, A Forum for the SIAM Activity Group on Optimization*, 1:11, 1992.

[193] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, University of California Press, Berkeley and Los Angeles, CA, 1961.

[194] J. C. Lagarias. The nonlinear geometry of linear programming, Part III : Projective Legendre transform coordinates and Hilbert geometry. *Transactions of the American Mathematical Society*, 320:193–225, 1990.

[195] J. C. Lagarias. A collinear scaling interpretation of Karmarkar's linear programming algorithm. *SIAM Journal on Optimization*, 3:630–636, 1993.

[196] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly Appl. Math.*, 2:164-168, 1963.

[197] A. Levin. On an algorithm for the minimization of convex functions. *Soviet Math. Doklady*, 6:286-290, 1965.

[198] D. G. Luenberger. *Linear and Nonlinear Programming*, 2nd Edition. Addison–Wesley, Menlo Park, CA, 1984.

[199] Z. Q. Luo, J. Sturm and S. Zhang. Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. Report 9607/A, Econometric Institute, Erasmus University, Rotterdam, 1995.

[200] Z. Q. Luo and P. Tseng. Error bounds and convergence analysis of matrix splitting algorithms for the affine variational inequality problem. *SIAM J. on Optimization*, 2:43–54, 1992.

[201] Z. Q. Luo and Y. Ye. A genuine quadratically convergent polynomial interior-point algorithm for linear programming. In Ding-Zhu Du and Jie Sun, editors, *Advances in Optimization and Approximation*, pages 235–246, Kluwer Academic, Dordrecht, 1994

[202] I. J. Lustig. Feasibility issues in a primal–dual interior point method for linear programming. *Mathematical Programming*, 49:145–162, 1990/91.

[203] I. J. Lustig, R. E. Marsten, and D. F. Shanno. Computational experience with a primal–dual interior point method for linear programming. *Linear Algebra and Its Applications*, 152:191–222, 1991.

[204] I. J. Lustig, R. E. Marsten, and D. F. Shanno. Computational experience with a globally convergent primal–dual predictor–corrector algorithm for linear programming. *Mathematical Programming*, 66:123–135, 1994.

[205] I. J. Lustig, R. E. Marsten, and D. F. Shanno. Interior point methods : Computational state of the art. *ORSA Journal on Computing*, 6:1–14, 1994.

[206] I. J. Lustig, R. E. Marsten, and D. F. Shanno. On implementing Mehrotra's predictor–corrector interior point method for linear programming. *SIAM Journal on Optimization*, 2:435–449, 1992.

[207] O. L. Mangasarian. Global error bounds for monotone affine variational inequality problems. Technical Report 998, Computer Sciences Department, University of Wisconsin, Madison, WI, 1991, to appear in *Linear Algebra and Applications.*

[208] O. L. Mangasarian and T.-H. Shiau. Error bounds for monotone linear complementarity problems. *Mathematical Programming*, 36:81–89, 1986.

[209] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *J. SIAM*, 11:431-441, 1963.

[210] W. F. Mascarenhas. The affine scaling algorithm fails for $\lambda = 0.999$. To appear in *SIAM J. Optimization*, 1996.

[211] C. McDiarmid. On the improvement per iteration in Karmarkar's method for linear programming. *Mathematical Programming*, 46:299–320, 1990.

[212] L. McLinden. The analogue of Moreau's proximation theorem, with applications to the nonlinear complementarity problem. *Pacific Journal of Mathematics*, 88:101–161, 1980.

[213] K. A. McShane. Superlinearly convergent $O(\sqrt{n}L)$-iteration interior-point algorithms for linear programming and the monotone linear complementarity problem. *SIAM Journal on Optimization*, 4(2):247–261, 1994.

[214] 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.

[215] N. Megiddo. Linear programming (1986). *Annual Review of Computer Science*, 2:119–145, 1987.

[216] 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*, 1986, pages 1–35.

[217] N. Megiddo. On finding primal– and dual–optimal bases. *ORSA Journal on Computing*, 3:63–65, 1991.

[218] N. Megiddo, S. Mizuno and T. Tsuchiya. A modified layered-step interior-point algorithm for linear programming. IBM Technical Report, San Jose, 1996.

[219] N. Megiddo and M. Shub. Boundary behavior of interior point algorithms in linear programming. *Mathematics of Operations Research*, 14:97–146, 1989.

[220] S. Mehrotra. Quadratic convergence in a primal–dual method. *Mathematics of Operations Research*, 18:741-751, 1993.

[221] S. Mehrotra. On the implementation of a primal–dual interior point method. *SIAM Journal on Optimization*, 2(4):575–601, 1992.

[222] S. Mehrotra and J. Sun. An algorithm for convex quadratic programming that requires $O(n^{3.5}L)$ arithmetic operations. *Mathematics of Operations Research*, 15:342–363, 1990.

[223] S. Mehrotra and J. Sun. A method of analytic centers for quadratically constrained convex quadratic programs. *SIAM Journal on Numerical Analysis*, 28(2):529–544, 1991.

[224] S. Mehrotra and Y. Ye. Finding an interior point in the optimal face of linear programs. *Mathematical Programming*, 62:497–515, 1993.

[225] J. E. Mitchell. An interior point column generation method for linear programming using shifted barriers. *SIAM Journal on Optimization*, 4(2):423–440, 1994.

[226] J. E. Mitchell and M. J. Todd. Solving combinatorial optimization problems using Karmarkar's algorithm. *Mathematical Programming*, 56:245–284, 1992.

[227] B. S. Mityagin. Two inequalities for volumes of convex bodies. *Matematicheskie Zametki*, 1:99-106, 1969.

[228] S. Mizuno. $O(n^\rho L)$ iteration $O(n^3 L)$ potential reduction algorithms for linear programming. *Linear Algebra and Its Applications*, 152:155–168, 1991.

[229] S. Mizuno. A new polynomial time method for a linear complementarity problem. *Mathematical Programming*, 56:31–43, 1992.

[230] S. Mizuno. Polynomiality of infeasible–interior–point algorithms for linear programming. *Mathematical Programming*, 67:109–119, 1994.

[231] S. Mizuno. A superlinearly convergent infeasible-interior-point algorithm for geometrical LCPs without a strictly complementary condition. *Mathematics of Operations Research*, 21(2):382–400, 1996.

[232] S. Mizuno, M. Kojima, and M. J. Todd. Infeasible–interior–point primal–dual potential–reduction algorithms for linear programming. *SIAM Journal on Optimization*, 5:52–67, 1995.

[233] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive step primal–dual interior–point algorithms for linear programming. *Mathematics of Operations Research*, 18:964–981, 1993.

[234] S. Mizuno, M. J. Todd, and Y. Ye. A surface of analytic centers and primal–dual infeasible–interior–point algorithms for linear programming. *Mathematics of Operations Research*, 20:135–162, 1995.

[235] C. L. Monma. Successful implementations of interior algorithms. *SIAM News*, 22(2):14–16, 1989.

[236] R. D. C. Monteiro. On the continuous trajectories for a potential reduction algorithm for linear programming. *Mathematics of Operations Research*, 17:225–253, 1992.

[237] R. D. C. Monteiro. Convergence and boundary behavior of the projective scaling trajectories for linear programming. *Mathematics of Operations Research*, 16:842–858, 1991.

[238] R. D. C. Monteiro. A globally convergent primal-dual interior point algorithm for convex programming. *Mathematical Programming*, 64:123–147, 1994.

[239] R. D. C. Monteiro and I. Adler. Interior path following primal–dual algorithms : Part I : Linear programming. *Mathematical Programming*, 44:27–41, 1989.

[240] R. D. C. Monteiro and I. Adler. Interior path following primal–dual algorithms : Part II : Convex quadratic programming. *Mathematical Programming*, 44:43–66, 1989.

[241] R. D. C. Monteiro and I. Adler. An extension of Karmarkar–type algorithm to a class of convex separable programming problems with global linear rate of convergence. *Mathematics of Operations Research*, 15:408–422, 1990.

[242] R. D. C. Monteiro, I. Adler, and M. G. C. Resende. A polynomial–time primal–dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. *Mathematics of Operations Research*, 15:191–214, 1990.

[243] R. D. C. Monteiro and S. Mehrotra. A general parametric analysis approach and its implication to sensitivity analysis in interior point methods. *Mathematical Programming*, 47:65–82, 1996.

[244] R. D. C. Monteiro and J-S. Pang. On two interior-point mappings for nonlinear semidefinite complementarity problems. Working Paper, John Hopkins University, 1996.

[245] R. D. C. Monteiro and T. Tsuchiya. Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem. To appear in *Mathematics of Operations Research*, 1996.

[246] R. D. C. Monteiro, T. Tsuchiya, and Y. Wang. A simplified global convergence proof of the affine scaling algorithm. *Ann. Oper. Res.*, 47:443–482, 1993.

[247] R. D. C. Monteiro and S. J. Wright. Local convergence of interior–point algorithms for degenerate monotone LCP. *Computational Optimization and Applications*, 3:131–155, 1994.

[248] J. J. Moré. The Levenberg-Marquardt algorithm: implementation and theory. In G. A. Watson, editor: *Numerical Analysis* Springer-Verlag, New York, 1977.

[249] K. G. Murty. *Linear Complementarity, Linear and Nonlinear Programming.* Heldermann, Verlag, Berlin, 1988.

[250] K. G. Murty. *Linear Complementarity, Linear and Nonlinear Programming*, volume 3 of *Sigma Series in Applied Mathematics*, chapter 11.4.1 : The Karmarkar's algorithm for linear programming, pages 469–494. Heldermann Verlag, Nassauische Str. 26, D–1000 Berlin 31, Germany, 1988.

[251] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117-129, 1987.

[252] S. G. Nash and A. Sofer. A barrier method for large–scale constrained optimization. *ORSA Journal on Computing*, 5:40–53, 1993.

[253] J. L. Nazareth. Homotopy techniques in linear programming. *Algorithmica*, 1(4):529–535, 1986.

[254] J. L. Nazareth. The homotopy principle and algorithms for linear programming. *SIAM Journal on Optimization*, 1:316–332, 1991.

[255] D. J. Newman. Location of the maximum on unimodal surfaces. *J. Ass. Comput. Math.*, 12:395-398, 1965.

[256] A. S. Nemirovskii. A new polynomial algorithm for linear programming. *Doklady Akademii Nauk SSSR*, 298(6):1321–1325, 1988. Translated in : *Soviet Mathematics Doklady*, 37(1):264–269, 1988.

[257] A. S. Nemirovskii. On polynomiality of the method of analytic centers for fractional problems. *Mathematical Programming*, 73(2):175–198, 1996.

[258] A. S. Nemirovsky and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization.* John Wiley and Sons, Chichester, 1983. Translated by E. R. Dawson from *Slozhnost' Zadach i Effektivnost' Metodov Optimizatsii*, 1979, Glavnaya redaktsiya fizikomatematicheskoi literatury, Izdatelstva "Nauka".

[259] J. von Neumann. On a maximization problem. Manuscript, Institute for Advanced Studies, Princeton University, Princeton, NJ 08544, USA, 1947.

[260] Yu. E. Nesterov. Polynomial–time dual algorithms for linear programming. *Kibernetika*, 25(1):34–40, 1989. Translated in : *Cybernetics*, 25(1):40–49, 1989.

[261] Yu. E. Nesterov. Complexity estimates of some cutting plane methods based on the analytic barrier. *Mathematical Programming*, 69:149–176, 1995.

[262] Yu. E. Nesterov and A. S. Nemirovskii. Acceleration and parallelization of the path–following interior point method for a linearly constrained convex quadratic problem. *SIAM Journal on Optimization*, 1(4):548–564, 1991.

[263] Yu. E. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms.* SIAM Publications. SIAM, Philadelphia, 1993.

[264] Yu. E. Nesterov and A. S. Nemirovskii. An interior-point method for generalized linear-fractional programming. *Mathematical Programming*, 69:177–204, 1995.

[265] Yu. E. Nesterov and M. J. Todd. Self-scaled cones and interior-point methods in nonlinear programming. Technical Report No. 1091, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1994.

[266] Yu. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. Technical Report No. 1125, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1995.

[267] Yu. Nesterov, M. J. Todd and Y. Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. Technical Report No. 1156, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1996.

[268] T. Noma. *A globally convergent iterative algorithm for complementarity problems : A modification of interior point algorithms for linear complementarity problems.* PhD thesis, Dept. of Information Science, Tokyo Institute of Technology, 2–12–1 Oh–Okayama, Meguro–ku, Tokyo 152, Japan, 1991.

[269] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, New York, 1970.

[270] M. W. Padberg. A different convergence proof of the projective method for linear programming. *Operations Research Letters*, 4:253–257, 1986.

[271] V. Pan. The modified barrier function method for linear programming and its extension. *Computers and Mathematics with Applications*, 20(3):1–14, 1990.

[272] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Prentice Hall, Englewood Cliffs, New Jersey, 1982.

[273] P. M. Pardalos and J. B. Rosen. *Constrained Global Optimization: Algorithms and Applications.* Springer-Verlag, Lecture Notes in Computer Sciences 268, 1987.

[274] P. M. Pardalos and S. Jha. Complexity of uniqueness and local search in quadratic 0-1 programming. *Operations Research Letters*, 11:119-123, 1992.

[275] P. M. Pardalos, Y. Ye, and C. G. Han. Algorithms for the solution of quadratic knapsack problems. *Linear Algebra and Its Applications*, 152:69–91, 1991.

[276] E. Polak, J. E. Higgins, and D. Q. Mayne. A barrier function method for minimax problems. *Mathematical Programming*, 54:155–176, 1992.

[277] R. Polyak. Modified barrier functions (theory and methods). *Mathematical Programming*, 54:177–222, 1992.

[278] D. B. Ponceleon. *Barrier methods for large–scale quadratic programming.* PhD thesis, Computer Science Department, Stanford University, Stanford, CA, 1990.

[279] K. Ponnambalam, A. Vannelli, and S. Woo. An interior point method implementation for solving large planning problems in the oil refinery industry. *Canadian Journal of Chemical Engineering*, 70:368–374, 1992.

[280] F. A. Potra. A quadratically convergent predictor–corrector method for solving linear programs from infeasible starting points. *Mathematical Programming*, 67:383–406, 1994.

[281] F. A. Potra and Y. Ye. A quadratically convergent polynomial algorithm for solving entropy optimization problems. *SIAM Journal on Optimization*, 3:843–860, 1993.

[282] F. A. Potra and Y. Ye. Interior point methods for nonlinear complementarity problems. Reports on Computational Mathematics 15, Dept. of Mathematics, The University of Iowa, Iowa City, IA, 1991.

[283] M. J. D. Powell. On the number of iterations of Karmarkar's algorithm for linear programming. *Mathematical Programming*, 62:153–197, 1993.

[284] S. Ramaswamy and J. E. Mitchell. A long step cutting plane algorithm that uses the volumetric barrier. Department of Mathematical Science, RPI, Troy, NY, 1995.

[285] F. Rendl and H. Wolkowicz. A semidefinite framework for trust region subproblems with applications to large scale minimization. CORR Report 94-32, Department of Combinatorics and Optimization, University of Waterloo, Ontario, Canada, 1994.

[286] J. Renegar. A polynomial–time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, 40:59–93, 1988.

[287] J. Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65:73–91, 1994.

[288] J. Renegar and M. Shub. Unified complexity analysis for Newton LP methods. *Mathematical Programming*, 53:1–16, 1992.

[289] M. G. C. Resende and G. Veiga. An implementation of the dual affine scaling algorithm for minimum cost flow on bipartite uncapacitated networks. *SIAM Journal on Optimization*, 3:516–537, 1993.

[290] S. Robinson. A linearization technique for solving the irreducible von Neumann economic model. In Łoś and Łoś, eds., *Mathematical Models in Economics*, page 139–150, PWN-Polish Scientific Publishers, Warszawa, 1974.

[291] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.

[292] M. Ramana, L. Tuncel and H. Wolkowicz. Strong duality for semidefinite programming. CORR 95-12, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, 1995.

[293] C. Roos and J. P. Vial. A polynomial method of approximate centers for linear programming. *Mathematical Programming*, 54:295–305, 1992.

[294] S. Sahni. Computationally related problems. *SIAM J. Compu.*, 3:262-279, 1974.

[295] R. Saigal. A simple proof of primal affine scaling methods. Tech. Report, Dept. of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, 1993.

[296] R. Saigal. *Linear Programming: Modern Integrated Analysis.* Kluwer Academic Publisher, Boston, 1995.

[297] B. K. Schmidt and T. H. Mattheiss. The probability that a random polytope is bounded. *Mathematics of Operations Research*, 2:292–296, 1977.

[298] A. Schrijver. *Theory of Linear and Integer Programming.* John Wiley & Sons, New York, 1986.

[299] R. Seidel. Small-dimensional linear programming and convex hulls made easy. *Discrete Comput. Geom.*, 6:423–434, 1991.

[300] D. Shanno and E. Simantiraki. An infeasible interior-point method for linear complementarity problems. Rutcor Research Report 7–95, March 1995 (revised, February, 1996).

[301] H. D. Sherali. Algorithmic insights and a convergence analysis for a Karmarkar–type of algorithm for linear programming problems. *Naval Research Logistics Quarterly*, 34:399–416, 1987.

[302] H. D. Sherali, B. O. Skarpness, and B. Kim. An assumption–free convergence analysis of the scaling algorithm for linear programs, with application to the L_1 estimation problem. *Naval Research Logistics Quarterly*, 35:473–492, 1988.

[303] N. Z. Shor. Cut-off method with space extension in convex programming problems. *Kibernetika*, 6:94-95, 1977.

[304] M. Shub. On the asymptotic behavior of the projective rescaling algorithm for linear programming. *Journal of Complexity*, 3:258–269, 1987.

[305] S. Smale. On the average number of steps of the simplex method of linear programming. *Mathematical Programming*, 27:241–262, 1983.

[306] G. Sonnevend. An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In A. Prekopa, J. Szelezsan, and B. Strazicky, editors, *System Modelling and Optimization : Proceedings of the 12th IFIP–Conference held in Budapest, Hungary, September 1985*, volume 84 of *Lecture Notes in Control and Information Sciences*, pages 866–876. Springer Verlag, Berlin, Germany, 1986.

[307] G. Sonnevend, J. Stoer, and G. Zhao. On the complexity of following the central path by linear extrapolation in linear programming. *Methods of Operations Research*, 62:19–31, 1990.

[308] G. Sonnevend, J. Stoer, and G. Zhao. On the complexity of following the central path of linear programs by linear extrapolation II. *Mathematical Programming*, 52:527–553, 1991.

[309] D. C. Sorenson. Newton's method with a model trust region modification. *SIAM J. Numer. Anal.*, 19:409-426, 1982.

[310] J. F. Sturm and S. Zhang. New complexity results for the Iri-Imai method. To appear in *Annals of Operations Research*, 1995.

[311] J. F. Sturm and S. Zhang. On a wide region of centers and primal-dual interior point algorithms for linear programming. Tinbergen Institute Rotterdam, Erasmus University Rotterdam, The Netherlands, 1995.

[312] J. Sun. A convergence proof for an affine–scaling algorithm for convex quadratic programming without nondegeneracy assumptions. *Mathematical Programming*, 60:69–79, 1993.

[313] J. Sun and L. Qi. An interior point algorithm of $O(\sqrt{n}|\ln\varepsilon|)$ iterations for C^1–convex programming. *Mathematical Programming*, 57:239–257, 1992.

[314] J. Sun and G. Zhao. A quadratically convergent polynomial long-step algorithm for a class of nonlinear monotone complementarity problems. Working Paper, National Univ. of Singapore, 1996.

[315] K. C. Tan and R. M. Freund. Newton's method for the general parametric center problem with applications. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, 1991.

[316] K. Tanabe. Complementarity–enforced centered Newton method for mathematical programming. In K. Tone, editor, *New Methods for Linear Programming*, pages 118–144, The Institute of Statistical Mathematics, 4–6–7 Minami Azabu, Minatoku, Tokyo 106, Japan, 1987.

[317] K. Tanabe. Centered Newton method for mathematical programming. In M. Iri and K. Yajima, editors, *System Modeling and Optimization*, Proceedings of the 13th IFIP Conference, pages 197–206, Springer-Verlag, Tokyo, Japan, 1988.

[318] R. A. Tapia. Current research in numerical optimization. *SIAM News*, 20:10–11, March 1987.

[319] S. P. Tarasov, L. G. Khachiyan, and I. I. Érlikh. The method of inscribed ellipsoids. *Soviet Mathematics Doklady*, 37:226–230, 1988.

[320] T. Terlaky, editor. *Interior-Point Methods in Mathematical Programming* Kluwer Academic Publisher, Boston, 1996.

[321] M. J. Todd, Improved bounds and containing ellipsoids in Karmarkar's linear programming algorithm. *Mathematics of Operations Research*, 13:650-659, 1988.

[322] M. J. Todd. Recent developments and new directions in linear programming. In M. Iri and K. Tanabe, editors, *Mathematical Programming : Recent Developments and Applications*, pages 109–157. Kluwer Academic Press, Dordrecht, The Netherlands, 1989.

[323] M. J. Todd. Interior–point algorithms for semi–infinite programming. *Mathematical Programming*, 65:217–245, 1994.

[324] M. J. Todd. A low complexity interior point algorithm for linear programming. *SIAM Journal on Optimization*, 2:198–209, 1992.

[325] M. J. Todd. Polynomial expected behavior of a pivoting algorithm for linear complementarity and linear programming problems. *Mathematical Programming*, 35:173–192, 1986.

[326] M. J. Todd. Probabilistic models for linear programming. *Mathematics of Operations Research*, 16:671-693, 1991.

[327] M. J. Todd. A lower bound on the number of iterations of primal–dual interior–point methods for linear programming. In G.A. Watson and D.F. Griffiths, editors, *Numerical Analysis 1993*, Volume 303 of *Pitman Research Notes in Mathematics*, pages 237–259, Longman Press, Burnt Hill, UK, 1994.

[328] M. J. Todd. Potential-reduction methods in mathematical programming. Technical Report 1112, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1995.

[329] M. J. Todd and B. P. Burrell. An extension of Karmarkar's algorithm for linear programming using dual variables. *Algorithmica*, 1(4):409–424, 1986.

[330] M. J. Todd and Y. Ye. A centered projective algorithm for linear programming. *Mathematics of Operations Research*, 15:508–529, 1990.

[331] M. J. Todd and Y. Ye. A lower bound on the number of iterations of long-step and polynomial interior-point linear programming algorithms. To appear in *Annals of Operations Research*, 1995.

[332] M. J. Todd and Y. Ye. Approximate Farkas lemmas and stopping rules for iterative infeasible interior-point linear programming. Technical Report, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1994.

[333] K. Tone. An active–set strategy in interior point method for linear programming. *Mathematical Programming*, 59:345–360, 1993.

[334] J. F. Traub. Computational complexity of iterative processes. *SIAM J. Comput.*, 1:167–179, 1972.

[335] J. F. Traub, G. W. Wasilkowski and H. Wozniakowski. *Information-Based Complexity*. Academic Press, San Diego, CA, 1988.

[336] P. Tseng. A simple complexity proof for a polynomial–time linear programming algorithm. *Operations Research Letters*, 8:155–159, 1989.

[337] P. Tseng. Complexity analysis of a linear complementarity algorithm based on a Lyapunov function. *Mathematical Programming*, 53:297–306, 1992.

[338] P. Tseng and Z. Q. Luo. On the convergence of the affine–scaling algorithm. *Mathematical Programming*, 52:301–319, 1992.

[339] T. Tsuchiya. Global convergence of the affine scaling methods for degenerate linear programming problems. *Mathematical Programming*, 52:377–404, 1991.

[340] T. Tsuchiya. Global convergence property of the affine scaling methods for primal degenerate linear programming problems. *Mathematics of Operations Research*, 17(3):527–557, 1992.

[341] T. Tsuchiya. Quadratic convergence of the Iri-Imai algorithm for degenerate linear programming problems. *Journal of Optimization Theory and Applications*, 87(3):703–726, 1995.

[342] T. Tsuchiya and M. Muramatsu. Global convergence of a long–step affine scaling algorithm for degenerate linear programming problems. *SIAM J. Optimization*, 5:525–551, 1995.

[343] T. Tsuchiya and K. Tanabe. Local convergence properties of new methods in linear programming. *Journal of the Operations Research Society of Japan*, 33:22–45, 1990.

[344] A. W. Tucker. Dual systems of homogeneous linear relations, in H. W. Kuhn and A. W. Tucker, Editors, *Linear Inequalities and Related Systems*, page 3–, Princeton University Press, Princeton, NJ, 1956.

[345] L. Tunçel. Constant potential primal–dual algorithms : A framework. *Mathematical Programming*, 66:145–159, 1994.

[346] R. Tutuncu. An infeasible-interior-point potential-reduction algorithm for linear programming. Technical Report No. 1136, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1995.

[347] P. M. Vaidya. A locally well–behaved potential function and a simple Newton–type method for finding the center of a polytope. In N. Megiddo, editor, *Progress in Mathematical Programming : Interior Point and Related Methods*, pages 79–90. Springer Verlag, New York, 1989.

[348] P. M. Vaidya. An algorithm for linear programming which requires $O((m+n)n^2 + (m+n)^{1.5}nL)$ arithmetic operations. *Mathematical Programming*, 47:175–201, 1990. Condensed version in : *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, 1987, pages 29–38.

[349] P. M. Vaidya. A new algorithm for minimizing a convex function over convex sets. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, NC, USA, 1989*, pages 338–343. IEEE Computer Society Press, Los Alamitos, CA, 1990.

[350] P. M. Vaidya and D. S. Atkinson. A technique for bounding the number of iterations in path following algorithms. In P. Pardalos, editor, *Complexity in Numerical Optimization* page 462–489, World Scientific, New Jersey, 1993.

[351] L. Vandenberghe and S. Boyd. A primal–dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, 69:205–236, 1995.

[352] L. Vandenberghe and S. Boyd. Semidefinite programming. Technical Report, Information Systems Laboratory, Dept. of Electrical Engineering, Stanford University, Stanford, CA, 1994.

[353] L. Vandenberghe, S. Boyd, and S.P. Wu. Determinant maximization with linear matrix inequality constraints. Technical Report, Information Systems Laboratory, Stanford University, CA, 1996.

[354] R. J. Vanderbei. LOQO: An Interior Point Code for Quadratic Programming. Program in Statistics & Operations Research, Princeton University, NJ, 1995.

[355] R. J. Vanderbei. Symmetric quasidefinite matrices. *SIAM Journal on Optimization*, 5:100–113, 1995.

[356] R. J. Vanderbei and T. J. Carpenter. Symmetric indefinite systems for interior point methods. *Mathematical Programming*, 58:1–32, 1993.

[357] R. J. Vanderbei and J. C. Lagarias. I. I. Dikin's convergence result for the affine–scaling algorithm. In J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming*, volume 114 of *Contemporary Mathematics*, pages 109–119. American Mathematical Society, Providence, RI, 1990.

[358] R. J. Vanderbei, M. S. Meketon, and B. A. Freedman. A modification of Karmarkar's linear programming algorithm. *Algorithmica*, 1(4):395–407, 1986.

[359] S. A. Vavasis. *Nonlinear Optimization: Complexity Issues.* Oxford Science, New York, NY, 1991.

[360] S. A. Vavasis. Polynomial time weak approximation algorithms for quadratic programming. In C. A. Floudas and P. M. Pardalos, editors, *Complexity in Numerical Optimization*. World Scientific, New Jersey, 1993.

[361] S. A. Vavasis and R. Zippel. Proving polynomial time for sphere-constrained quadratic programming. Technical Report 90-1182, Department of Computer Science, Cornell University, Ithaca, NY, 1990.

[362] S. A. Vavasis and Y. Ye. An primal-dual accelerated interior-point method whose running time depends only on A. Working Paper, Department of Management Sciences, The University of Iowa, Iowa City, IA, 1994.

[363] C. Wallacher and U. Zimmermann. A combinatorial interior point method for network flow problems. *Mathematical Programming*, 52:321–335, 1992.

[364] T. Wang, R. D. C. Monteiro, and J.-S. Pang. An interior point potential reduction method for constrained equations. To appear in *Mathematical Programming*, 1996.

[365] Z. L. Wei. An interior point method for linear programming. *Journal of Computational Mathematics*, 5:342–351, 1987.

[366] M. H. Wright. A brief history of linear programming. *SIAM News*, 18(3):4, November 1985.

[367] S. J. Wright. An infeasible–interior–point algorithm for linear complementarity problems. *Mathematical Programming*, 67:29–51, 1994.

[368] S. Q. Wu and F. Wu. New algorithms for linear programming. *Acta Mathematica Applicate Sinica (English Series)*, 8(1):18–26, 1992.

[369] X. Xu, P-F. Hung and Y. Ye. A simplification of the homogeneous and self-dual linear programming algorithm and its implementation. To appear in *Annals of Operations Research*, 1995.

[370] G. Xue and Y. Ye. Efficient algorithms for minimizing a sum of Euclidean norms with applications. Working Paper, Department of Management Sciences, The University of Iowa, 1995.

[371] H. Yamashita. A polynomially and quadratically convergent method for linear programming. Mathematical Systems Institute, Tokyo, Japan, 1986.

[372] Y. Ye. Karmarkar's algorithm and the ellipsoid method. *Operations Research Letters*, 6:177–182, 1987.

[373] Y. Ye. An $O(n^3L)$ potential reduction algorithm for linear programming. *Mathematical Programming*, 50:239–258, 1991.

[374] Y. Ye. Further development on the interior algorithm for convex quadratic programming. Working Paper, Dept. of Engineering Economic Systems, Stanford University, Stanford, CA 94305, USA, 1987.

[375] Y. Ye. A class of projective transformations for linear programming. *SIAM Journal on Computing*, 19:457–466, 1990.

[376] Y. Ye. On an affine scaling algorithm for nonconvex quadratic programming. *Mathematical Programming*, 52:285–300, 1992.

[377] Y. Ye. A potential reduction algorithm allowing column generation. *SIAM Journal on Optimization*, 2:7–20, 1992.

[378] Y. Ye. On the Q-order of convergence of interior-point algorithms for linear programming. In F. Wu, editor, *Proc. of the 1992 Symp. on Applied Mathematics.* Institute of Applied Mathematics, Chinese Academy of Sciences, 1992, 534-543.

[379] Y. Ye. Toward probabilistic analysis of interior–point algorithms for linear programming. *Mathematics of Operations Research*, 19:38–52, 1994.

[380] Y. Ye. Combining binary search and Newton's method to compute real roots for a class of real functions. *Journal of Complexity*, 10:271–280, 1994.

[381] Y. Ye. On the von Neumann economic growth problem. *Mathematics of Operations Research*, 20:617–633, 1995.

[382] Y. Ye and K. M. Anstreicher. On quadratic and $O(\sqrt{n}L)$ convergence of a predictor–corrector algorithm for LCP. *Mathematical Programming*, 62:537–551, 1993.

[383] Y. Ye, O. Güler, R. A. Tapia, and Y. Zhang. A quadratically convergent $O(\sqrt{n}L)$–iteration algorithm for linear programming. *Mathematical Programming*, 59:151–162, 1993.

[384] Y. Ye and M. Kojima. Recovering optimal dual solutions in Karmarkar's polynomial algorithm for linear programming. *Mathematical Programming*, 39:305–317, 1987.

[385] Y. Ye and P. M. Pardalos. A class of linear complementarity problems solvable in polynomial time. *Linear Algebra and Its Applications*, 152:3–17, 1991.

[386] 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.

[387] Y. Ye and E. Tse. An extension of Karmarkar's projective algorithm for convex quadratic programming. *Mathematical Programming*, 44:157–179, 1989.

[388] S. Zhang. Convergence property of the Iri-Imai algorithm for some smooth convex programming problems. *Journal of Optimization Theory and Applications*, 82:121–138, 1994.

[389] Y. Zhang. On the convergence of a class of infeasible interior-point methods for the horizontal linear complementarity problem. *SIAM Journal on Optimization*, 4:208–227, 1994.

[390] Y. Zhang and R. A. Tapia. A superlinearly convergent polynomial primal–dual interior–point algorithm for linear programming. *SIAM Journal on Optimization*, 3:118–133, 1993.

[391] Y. Zhang, R. A. Tapia, and J. E. Dennis. On the superlinear and quadratic convergence of primal–dual interior point linear programming algorithms. *SIAM Journal on Optimization*, 2(2):304–324, 1992.

[392] Y. Zhang, R. A. Tapia, and F. Potra. On the superlinear convergence of interior point algorithms for a general class of problems. *SIAM Journal on Optimization*, 3(2):413–422, 1993.

[393] Y. Zhang and D. Zhang. On polynomiality of the Mehrotra-type predictor-corrector interior-point algorithms. *Mathematical Programming*, 68:303–317, 1995.

[394] G. Zhao and J. Stoer. Estimating the complexity of path–following methods for solving linear programs by curvature integrals. *Applied Mathematics & Optimization*, 27(1):85–103, 1993.

[395] U. Zimmermann. Search directions for a class of projective methods. *Zeitschrift für Operations Research—Methods and Models of Operations Research*, 34:353–379, 1990.

Index