WORST-CASE EXAMPLES OF INTERIOR-POINT TRAJECTORIES

WORST-CASE EXAMPLES OF INTERIOR-POINT TRAJECTORIES

by

EISSA NEMATOLLAHI, M.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2008)
(Mathematics and Statistics)

McMaster University
Hamilton, Ontario

TITLE: Worst-Case Examples of Interior-Point Trajectories
AUTHOR: Eissa Nematollahi, M.Sc.
SUPERVISOR: Dr. Tamás Terlaky
NUMBER OF PAGES: xvii, 105

Abstract

In this thesis, we present some pathological linear optimization problems (LPs) for central-path-following interior-point methods (IPMs). The problems are constructed based on Klee and Minty's worst-case example for the simplex method. Redundancy is playing an important role in the constructions.

We first review some pathological examples for pivot algorithms applied to LPs and linear complementarity problems. We also describe some unfavorable linear and nonlinear convex optimization problems for IPMs, in particular for central-path-following variants. In some examples, the central path has a zigzagging pattern with "sharp" turns, which is difficult for central-path-following IPMs to follow.

We next introduce a redundant Klee-Minty (KM) problem in which an exponential number of redundant constraints are placed parallel to the facets of the KM cube at uniform distances. We show that the central path of this problem follows the simplex path, the edge-path followed by the simplex method, visiting a small neighborhood of each of the 2^n vertices of the KM cube. The well-known iteration complexity upper bound for this problem is $\mathcal{O}\big(n\sqrt{n}\log n\big)$, where n is the number of inequalities. This upper bound is significantly reduced to $\mathcal{O}\big(\sqrt{n}\log n\big)$ by allowing the distances to decay geometrically. Furthermore, we argue that the number of iteration required by central-path-following IPMs that operate in a small neighborhood of the central path is bounded below by $\Omega\big(\sqrt[3]{n/(\log n)^2}\big)$.

We simplify the previous constructions by placing the redundant constraints parallel to the coordinate hyperplanes. Besides easing the analysis, this construction provides a tighter iteration complexity lower bound $\Omega\big(\sqrt{n/\log n}\big)$, while retaining the upper bound $\mathcal{O}\big(\sqrt{n}\log n\big)$. Hence, it yields a nearly worst-case example for central-path-following IPMs since the gap between the lower and upper bounds is essentially closed.

Acknowledgements

First, I would like to express my gratitude to all of those who gave me the possibility to start, pursue, and successfully complete my doctoral program.

In particular, I wish to thank my supervisor Professor Tamás Terlaky for his guidance and unremitting support without which my Ph.D. degree would not have been accomplished. I am indebted to my friend and colleague Dr. Maziar Salahi who encouraged me to start my Ph.D. at McMaster University.

I am thankful to Professor Antoine Deza for the interesting discussions on a research project we jointly worked on that led to two journal publications. I would also like to extend my special thanks to my friend and colleague Dr. Reza Peyghami whose amazing enthusiasm and perseverance in research inspired me. Many special thanks to all AdvOL members for making such a friendly research environment. I also wish to thank my supervisory committee members Professors Fred M. Hoppe, Jiming Peng, and Gail S.K. Wolkowicz.

As a holder of the Marrie-Curie research fellowship, I traveled to the Center for Operations Research and Econometrics (CORE) at the Université Catholique de Louvain in Belgium for a period of three months starting from September 2005. I would like to give my sincere thanks to Professors Yurii Nesterov and François Glineur who welcomed me at CORE and helped me broaden my knowledge in optimization. I also wish to acknowledge the Selection Committee of the tri-annual Young Researchers Competition on continuous optimization for selecting me as a finalist for the competition in 2007 organized by the Mathematical Programming Society (MPS).

I would like to seize this opportunity to express my deepest gratitude to my family and friends for their constant kindness and love.

Finally, I am pleased to thank NSERC (Discovery Grant #5-48923) and MITACS for their financial support during the course of my studies.

Contents

List of Figures

List of Algorithms

Abbreviations

IPC	interior-point condition
IPM	interior-point method
LCP	linear complementarity problem
LP	linear optimization problem
MPC	Mehrotra's predictor-corrector
MTYPC	Mizuno-Todd-Ye's predictor-corrector
KM	Klee-Minty
NP	non-polynomial

Notation and Symbols

$\arg\max_{x\in\mathcal{E}} \psi(x)$	the unique maximizer of a function ψ over a convex set $\mathcal{E}$.
$\arg\min_{x\in\mathcal{E}} \phi(x)$	the unique minimizer of a function ϕ over a convex set $\mathcal{E}$.
A	a known matrix in $\mathbb{R}^{m\times n}$.
A_{ij}	the i^{th} component of the j^{th} column of A.
$A_{i\cdot}$	the i^{th} row of A.
$A_{\cdot j}$	the j^{th} column of A.
$A_{\mathcal{I}\mathcal{J}}$	a submatrix of A whose entries are A_{ij} where $i\in\mathcal{I}$ and $j\in\mathcal{J}$.
α,β,γ	real constants.
$\lfloor\alpha\rfloor$	the biggest integer number smaller than or equal to α.
$\lceil\alpha\rceil$	the smallest integer number bigger than or equal to α.
$\mathcal{B}$	the set $\{i \mid x_i^* > 0\}$, where $(x^*;y^*;s^*)\in\mathcal{F}_*$ with $x^*+s^*>0$.
b,c	known column vectors in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively.
b_i	the i^{th} component of b.
$(b_1;\cdots;b_m)$	the column vector $b\in\mathbb{R}^m$.
$b_{\mathcal{I}}$	a subvector of b whose components are b_i where $i\in\mathcal{I}$.
$\mathcal{C}$	the central path.
χ	the analytic center.
$\mathcal{D}$	the dual affine subspace $\big\{(y;s) \mid A^Ty+s=c\big\}$.
$\mathcal{D}_0$	the translation $\big\{(y;s) \mid A^Ty+s=0\big\}$ of the dual affine subspace.
$\mathcal{D}_\oplus$	the dual feasible set $\big\{(y;s) \mid A^Ty+s=c, s\ge 0\big\}$.
$\mathcal{D}_+$	the relative interior $\big\{(y;s) \mid A^Ty+s=c, s>0\big\}$ of $\mathcal{D}_\oplus$.
$\mathcal{D}_*$	the set of dual optimal solutions.
d	a distance vector in redundant Klee-Minty examples.
$(\Delta x;\Delta y;\Delta s)$	the Newton direction.
$(\Delta^a x;\Delta^a y;\Delta^a s)$	the affine-scaling direction.
e	the all-one vector.
$\varepsilon,\delta,\tau,\rho,\sigma$	parameters.
$\mathcal{F}$	the primal-dual affine subspace $\{(x;y;s) \mid x\in\mathcal{P}, (y;s)\in\mathcal{D}\}$.
$\mathcal{F}_\oplus$	the primal-dual feasible set $\{(x;y;s) \mid x\in\mathcal{P}_\oplus, (y;s)\in\mathcal{D}_\oplus\}$.

$\mathcal{F}_+$	the relative interior $\{(x;y;s) \mid x \in \mathcal{P}_+, (y;s) \in \mathcal{D}_+\}$ of $\mathcal{F}_\oplus$.
$\mathcal{F}_*$	the set of primal-dual optimal solutions.
$F_\mathcal{P}(x)$	the primal logarithmic barrier function $-e^T \log x$.
$F_\mathcal{D}(y)$	the dual logarithmic barrier function $e^T \log(c - A^T y)$.
g	the duality gap $x^T s = c^T x - b^T y$.
h	a repetition m-vector in redundant Klee-Minty examples.
I	the identity matrix.
$\mathcal{I}, \mathcal{J}, \mathcal{K}$	generic index subsets.
i, j, k	indices.
$\mathcal{KM}$	the feasible region of the Klee-Minty problem
L	the bit-length of the input-data (A, b, c) of an LP.
$\mathcal{L}$	a level set.
$\log x$	the column vector $(\log x_1; \cdots; \log x_n)$.
$\log(xs)$	the column vector $\big(\log(x_1 s_1); \cdots; \log(x_n s_n)\big)$.
$\min x$	$\min\{x_i \mid i = 1, \ldots, n\}$.
$\max x$	$\max\{x_i \mid i = 1, \ldots, n\}$.
μ	the central-path parameter.
μ_g	the gap-μ defined by $\mu_g = x^T s/n$, where $x \in \mathcal{P}_\oplus$ and $(y;s) \in \mathcal{D}_\oplus$.
$\mathscr{N}$	a neighborhood of the central path.
$\mathcal{N}$	the set $\{i \mid s_i^* > 0\}$, where $(x^*; y^*; s^*) \in \mathcal{F}_*$ with $x^* + s^* > 0$.
n	the number of the inequalities of an LP.
$\mathcal{O}$	$\alpha_k = \mathcal{O}(\beta_k)$ means $\lvert\alpha_k\rvert \leq \gamma \lvert\beta_k\rvert$ for some $\gamma > 0$ and large enough k.
Ω	$\alpha_k = \Omega(\beta_k)$ means $\lvert\alpha_k\rvert \geq \gamma \lvert\beta_k\rvert$ for some $\gamma > 0$ and large enough k.
$\mathcal{P}$	the affine subspace $\{x \mid Ax = b\}$.
$\mathcal{P}_\oplus$	the primal feasible set $\{x \mid Ax = b, x \geq 0\}$.
$\mathcal{P}_+$	the relative interior $\{x \mid Ax = b, x > 0\}$ of $\mathcal{P}_\oplus$.
$\mathcal{P}_*$	the set of primal optimal solutions.
$\mathbb{R}^n$	the set of n-dimensional real-valued vectors.
$\mathbb{R}^n_\oplus$	the nonnegative orthant $\{x \in \mathbb{R}^n \mid x \geq 0\}$.
$\mathbb{R}^n_+$	the positive orthant $\{x \in \mathbb{R}^n \mid x > 0\}$.
S	$\operatorname{diag} s$.
$\mathcal{S}^n$	the n-simplex $\big\{x \in \mathbb{R}^n_\oplus \mid e^T x \leq 1\big\}$.
s	an unknown column vector in $\mathbb{R}^n$.
τ	a positive number by which the Klee-Minty cube is tilted.
v^0	the vector $(0; \cdots; 0; 1) \in \mathbb{R}^m$.
X	$\operatorname{diag} x$.
x	an unknown column vector in $\mathbb{R}^n$.
x_i	the i^{th} component of x.
$x_{i:j}, (x_k)_{i:j}$	the subvector $(x_i; \cdots; x_j)$ of x.
x^T	the transpose of x.

x^{-1}	the vector $({x_1}^{-1}; \cdots ; {x_n}^{-1})$.
$\sqrt{x}$	the vector $(\sqrt{x_1}; \cdots ; \sqrt{x_n})$.
$\dot{x}(t)$	the vector $\big(\frac{dx_1(t)}{dt}; \cdots ; \frac{dx_n(t)}{dt}\big)$.
$\lfloor x \rfloor$	the vector $(\lfloor x_1 \rfloor ; \cdots ; \lfloor x_n \rfloor)$.
$\lceil x \rceil$	the vector $(\lceil x_1 \rceil ; \cdots ; \lceil x_n \rceil)$.
xs	the vector $(x_1 s_1; \cdots ; x_n s_n)$.
$x \geq s$	$x - s \in \mathbb{R}^n_{\oplus}$.
$x > s$	$x - s \in \mathbb{R}^n_{+}$.
$x(\mu)$	the μ-center in the primal space.
x^*	a primal optimal solution.
y	an unknown column vector in $\mathbb{R}^m$.
$(y; s)$	a simple notation for $(y^T, s^T)^T$.
$\big(y(\mu); s(\mu)\big)$	the μ-center in the dual space.
$(y^*; s^*)$	a dual optimal solution.
z	the column vector $(x; y; s) \in \mathbb{R}^{m+2n}$.
$z(\mu)$	the μ-center $\big(x(\mu); y(\mu); s(\mu)\big)$ in the primal-dual space.
z^*	a primal-dual optimal solution $(x^*; y^*; s^*)$.
$\mathbb{Z}^n$	the set of n-dimensional integer-valued vectors.
$\mathbb{Z}^n_{\oplus}$	the set $\{0, 1, 2, \ldots\}$.
$\mathbb{Z}^n_{+}$	the set $\{1, 2, 3, \ldots\}$.

Chapter 1

Introduction

In this chapter, we begin with introducing linear optimization problems in the standard primal and dual forms. Then, we briefly review the basics of the theory of interior-point methods, in particular central-path-following variants. Finally, we give an overview of the thesis and introduce some notation that are used throughout the thesis.

1.1 Linear Optimization Problems

We refer to the problem of optimizing a linear function over a finite set of linear constraints as a *linear optimization problem* (LP). A *primal LP* in the standard form is given as

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{1.1.1}$$

where $A \in \mathbb{R}^{m \times n}$ has full row rank, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$ are known vectors, and $x \in \mathbb{R}^n$ is an unknown vector. The associated *dual LP* is

$$\begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & A^T y + s = c, \\ & s \geq 0, \end{aligned} \tag{1.1.2}$$

where $y \in \mathbb{R}^m$ and $s \in \mathbb{R}^n$ are unknown vectors. We refer to the data set (A, b, c) as the *input-data* of the LP. We denote the feasible regions represented by the constraints of the primal and dual problems by $\mathcal{P}_\oplus$ and $\mathcal{D}_\oplus$, respectively;

i.e., we have

$$\begin{aligned}\mathcal{P}_\oplus &= \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}\,,\\ \mathcal{D}_\oplus &= \left\{(y;s) \in \mathbb{R}^{m+n} \mid A^T y + s = c, s \geq 0\right\}.\end{aligned}$$

For the sake of convenience, we define the set

$$\mathcal{F}_\oplus = \left\{(x;y;s) \in \mathbb{R}^{m+2n} \mid x \in \mathcal{P}_\oplus, (y;s) \in \mathcal{D}_\oplus\right\},$$

which is called the primal-dual feasible region. We denote the sets of optimal solutions of the primal and dual problems by $\mathcal{P}_*$ and $\mathcal{D}_*$, respectively; then, we define the primal-dual optimal set as

$$\mathcal{F}_* = \left\{(x;y;s) \in \mathbb{R}^{m+2n} \mid x \in \mathcal{P}_*, (y;s) \in \mathcal{D}_*\right\}.$$

We now recall two classical results from the linear optimization literature, known as the *weak duality* and *strong duality* theorems. For proofs, the reader is referred to [38]. For every $(x;y;s) \in \mathcal{F}_\oplus$, we define the *duality gap* as

$$g = c^T x - b^T y.$$

It is easy to verify that $g = x^T s$. Since the vectors x and s are nonnegative, the duality gap is always nonnegative and vanishes when $x^T s = 0$. These observations prove the following theorem known as the weak duality theorem.

Theorem 1.1.1. **[38]** *Let $\mathcal{F}_\oplus$ be nonempty. Then, for every $(x;y;s) \in \mathcal{F}_\oplus$, we have*

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

Moreover, the equality holds if and only if $x^T s = 0$.

This theorem implies that every feasible solution of the primal (dual) problem yields an upper (lower) bound for the dual (primal) objective value. The following theorem (strong duality theorem) shows that there exists a primal-dual optimal solution if the primal-dual feasible region is nonempty, and the duality gap vanishes only at optimality.

Theorem 1.1.2. **[38]** *If $\mathcal{F}_\oplus$ is nonempty then so is the primal-dual optimal set $\mathcal{F}_*$. Moreover, for every $(x;y;s) \in \mathcal{F}_\oplus$, we have*

$$(x;y;s) \in \mathcal{F}_* \quad \Longleftrightarrow \quad c^T x = b^T y.$$

This theorem provides necessary and sufficient optimality conditions. In other words, a vector $(x; y; s)$ is an optimal solution of the *primal-dual problem*[1] if and only if it satisfies

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

This linear system is under-determined. Therefore, we replace the linear equation $c^T x = b^T y$ by an equivalent set of n nonlinear equations $xs = 0$, where xs is the component-wise product of the vectors x and s. Thus, a vector $(x; y; s)$ is a primal-dual optimal solution if and only if it satisfies

$$\begin{aligned} Ax = b, &\quad x \geq 0, \\ A^T y + s = c, &\quad s \geq 0, \\ xs = 0. & \end{aligned} \tag{1.1.3}$$

In other words, we have

$$\mathcal{F}_* = \{(x; y; s) \in \mathcal{F}_\oplus \mid xs = 0\}.$$

The equation $xs = 0$ is often referred to as the *complementarity condition*. Moreover, every $(x; y; s) \in \mathcal{F}_\oplus$ that satisfies the complementarity condition $xs = 0$ is called a *complementary solution*. As a consequence, every complementary solution is an optimal solution.

In 1947, Dantzig [10] introduced the *simplex method* for solving LPs. To explain the simplex method, let us consider the primal problem (1.1.1). Let $\mathcal{P}_\oplus$ be nonempty. The simplex method is an iterative approach that starts from a feasible basis and updates the current basis, according to the pivot rule it is using, while monotonically decreasing the objective value at each iteration. A geometric interpretation of this method is that it starts from a vertex of the feasible region and moves along the edges to one of the adjacent vertices so that the objective value is monotonically decreasing. Under non-degeneracy condition[2] for the primal problem, the primal simplex method finds an optimal vertex or detects unboundedness in a finite number of iterations [7], regardless of the pivot rule it uses. For degenerate problems, on the other hand, the

[1] The problem of minimizing the duality gap over $\mathcal{F}_\oplus$ is referred to as the primal-dual problem.

[2] An LP is called primal degenerate when the feasible region has a degenerate vertex, i.e., a vertex that has more than $n - m$ zero components.

simplex method terminates in a finite number of iteration only when a finite pivot rule[3] is employed [8, 44].

1.2 Notation

The set of all n-dimensional real-valued column vectors is denoted by $\mathbb{R}^n$. Assume that $x, s \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and $\alpha \in \mathbb{R}$. The transpose of x is denoted by x^T. We may denote a vector x either by $(x_1; \cdots; x_n)$ or by $(x_k)_{1:n}$. The notation $(y; s)$ is a simple notation for $(y^T, s^T)^T \in \mathbb{R}^{m+n}$. We denote by $x_{i:j}$ the subvector $(x_i; \cdots; x_j)$, where $1 \leq i \leq j \leq n$. Subsets of an index set are denoted by $\mathcal{I}$, $\mathcal{J}$, $\mathcal{K}$. We denote by $x_{\mathcal{I}}$ a subvector of x whose components are x_i, where $i \in \mathcal{I}$.

The largest integer number less than or equal to α is denoted by $\lfloor \alpha \rfloor$, and the smallest integer number larger than or equal to α is denoted by $\lceil \alpha \rceil$. The unique maximizer of a strictly concave function ψ over a convex set $\mathcal{E}$ is denoted by $\arg\max_{x \in \mathcal{E}} \psi(x)$. Similarly, the unique minimizer of a strictly convex function ϕ over a convex set $\mathcal{E}$ is denoted by $\arg\min_{x \in \mathcal{E}} \phi(x)$.

We denote by $\log x$ the vector $(\log x_1; \cdots; \log x_n)$. Similar component-wise operations follow for

$$\sqrt{x}, x^{-1}, x + s, xs, \frac{x}{s}, \lfloor x \rfloor, \lceil x \rceil .$$

We also define $\alpha + x = (\alpha + x_1; \cdots; \alpha + x_n)$. The all-one vector is denoted by e, whose dimension is dictated by an appropriate context. By $x \geq s$ ($x > s$), we mean that $x_i \geq s_i$ ($x_i > s_i$) for $i = 1, \ldots, n$.

The set of vectors satisfying $x \geq 0$ is called the nonnegative orthant and denoted by $\mathbb{R}^n_{\oplus}$. Similarly, the set of all vectors $x > 0$ is called the positive orthant and denoted by $\mathbb{R}^n_{+}$. The set of all $m \times n$ matrices is denoted by $\mathbb{R}^{m \times n}$. The diagonal matrix whose diagonal elements are the components of a vector x is denoted by X, i.e., $X = \operatorname{diag} x$. Therefore, we have $XSe = Xs = Sx = xs$ each of which may be used interchangeably.

We denote the sets of feasible, strictly feasible, and optimal points of the primal, dual, and primal-dual problems according to Table 1.1. The primal and dual standard logarithmic barrier functions are, respectively, denoted by

[3] A pivot rule is said to be finite if the number of pivot steps to solve the problem is proven to be finite, i.e., no loop may occur during the procedure. We refer the reader to [44] for more discussion on pivot rules.

Table 1.1

	primal	dual	primal-dual
the set of feasible points	$\mathcal{P}_\oplus$	$\mathcal{D}_\oplus$	$\mathcal{F}_\oplus$
the set of strictly feasible points	$\mathcal{P}_+$	$\mathcal{D}_+$	$\mathcal{F}_+$
the set of optimal points	$\mathcal{P}_*$	$\mathcal{D}_*$	$\mathcal{F}_*$

$F_\mathcal{P}(x)$ and $F_\mathcal{D}(x)$. We denote the analytic center by χ, the cental path by $\mathcal{C}$, the central-path parameter by μ, and the duality gap by g. A primal and dual feasible solutions are often denoted by x and $(y; s)$, respectively, where $y \in \mathbb{R}^m$ and $x, s \in \mathbb{R}^n$. For the sake of simplicity and by abuse of notation, we denote a μ-center by $x(\mu)$ and $\big(y(\mu); s(\mu)\big)$ in the primal and dual spaces, respectively.

For real sequences $\{\alpha_k\}$ and $\{\beta_k\}$, we write $\alpha_k = \mathcal{O}\left(\beta_k\right)$ whenever there exist a positive integer k_0 and a $\gamma > 0$ such that $|\alpha_k| \leq \gamma\,|\beta_k|$ for $k \geq k_0$. An absolute constant may then be denoted by $\mathcal{O}(1)$. We also write $\alpha_k = \Omega(\beta_k)$ whenever there exist a positive integer k_0 and a $\gamma > 0$ such that $|\alpha_k| \geq \gamma\,|\beta_k|$ for $k \geq k_0$.

In order to unify some formulas, we follow MATLAB[4] standards, and by convention, we define

$$\sum_{i=m}^{n} x_i = 0, \quad \prod_{i=m}^{n} x_i = 1, \qquad \text{for} \;\; m > n.$$

For instance, we can write that $n! = \prod_{k=1}^{n} k$ is true for every $n \in \mathbb{Z}_\oplus$.

1.3 Interior-Point Methods

After Klee and Minty [23] proved that the simplex method may require an exponential number of pivot steps to solve an LP, the question

Is there any algorithm that solves LPs in *polynomial time*?

became the central open problem in linear optimization. By "polynomial time", we roughly mean that the number of arithmetic operations required by the algorithm to find an optimal solution of a problem is bounded above by a polynomial in the dimensions and the bit-length of the input-data of the

[4] http://www.mathworks.com/

problem. Note that the binary encoding of an integer α requires a string of length $l(\alpha) = 1 + \lceil \log_2 (1 + |\alpha|) \rceil$.

In 1979, Khachiyan [22] proved that the *ellipsoid method* solves LPs in polynomial time. His method was based on the techniques developed by Shor [40] and Yudin and Nemirovsky [51] for nonlinear optimization problems. He proved that the ellipsoid method enjoys $\mathcal{O}(n^2 L)$ iteration complexity upper bound with the total arithmetic operations $\mathcal{O}(n^4 L)$, where n is the number of inequalities and L is the bit-length of the input-data (A, b, c) defined as

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n} l(A_{ij}) + \sum_{i=1}^{m} l(b_i) + \sum_{j=1}^{n} l(c_j). \tag{1.3.1}$$

Although theoretically a polynomial-time algorithm, the ellipsoid method did not appear to be efficient in computational practice.

In 1984, Karmarkar [21] proposed a more efficient polynomial-time algorithm that sparked the research on *interior-point methods* (IPMs). His approach, known also as Karmarkar's projective method, enjoys better iteration complexity upper bound $\mathcal{O}(nL)$ with the total arithmetic operations $\mathcal{O}(n^{3.5} L)$. He claimed that his algorithm is more efficient, in practice, than the simplex method for large-scale LPs.

As opposed to the simplex method which moves along the boundary of the feasible region to find an optimal vertex, IPMs generate a sequence of strictly feasible points that converges to an optimal solution. Note that IPMs require the existence of both a primal and a dual strictly feasible points. In other words, the sets

$$\begin{aligned} \mathcal{P}_+ &= \{x \mid Ax = b, x > 0\}, \\ \mathcal{D}_+ &= \left\{(y; s) \mid A^T y + s = c, s > 0\right\}, \end{aligned}$$

are required to be nonempty. This condition is known as the *interior-point condition* (IPC) and is stated in the following assumption.

Assumption 1.3.1. **[IPC]** *The primal-dual problem has a strictly feasible solution. That is, the following set is nonempty:*

$$\mathcal{F}_+ = \{(x; y; s) \mid x \in \mathcal{P}_+, (y; s) \in \mathcal{D}_+\}.$$

This assumption can be made without loss of generality since we can always find an embedding model [38] of the original problem in a higher dimension for which the all-one vector e is a strictly feasible point. In addition, it yields an optimal solution of the original problem.

In the following, we briefly describe the following variants of IPMs: affine-scaling, potential-reduction, and central-path-following methods. The focus will be on central-path-following IPMs as the most popular and efficient algorithms, which are implemented in most professional software packages such as CPLEX[5], MOSEK[6], and XPRESS-MP[7].

1.3.1 Affine-Scaling Methods

Proposed originally by Dikin [15] in 1967, the *affine-scaling method* was independently reintroduced almost twenty years later by Barnes [5] and Vanderbei, Meketon, and Freedman [49]. In the following, we explain the algorithm applied to an LP in the dual form (1.1.2).

Let $(\bar{y}; \bar{s}) \in \mathcal{D}_+$ be an initial point for the dual affine-scaling method. We first consider the mapping $w \mapsto \bar{S}^{-1}s$ that maps $\bar{s}$ to the all-one vector e. The dual problem in the $(y; w)$-space is given as

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + \bar{S}w = c, \\ & \qquad\qquad w \geq 0, \end{array}$$

The ball $\left\{w \in \mathbb{R}^n \mid \|w - e\|_2^2 \leq 1\right\}$ is obviously inscribed in the nonnegative orthant $\mathbb{R}^n_\oplus$. The transformation of this ball from the w-space to the s-space is the *Dikin ellipsoid*

$$\left\{s \in \mathbb{R}^n \mid \left\|\bar{S}^{-1}(s - \bar{s})\right\|_2^2 \leq 1\right\},$$

centered around $\bar{s}$, which is also inscribed in the nonnegative orthant. We solve the dual problem (1.1.2) by approximating the nonnegative orthant with the inscribed Dikin ellipsoid at $\bar{s}$. We denote the unique solution of this problem by $(\bar{y}^+; \bar{s}^+)$.

The next iterate of the dual affine-scaling method is obtained by moving from the current point toward the point $(\bar{y}^+; \bar{s}^+)$ with a proper step-size; see Figure 1.1 for an illustration in the y-space. The dual affine-scaling direction may

[5] http://www.ilog.com/
[6] http://www.mosek.com/
[7] http://www.dashoptimization.com/

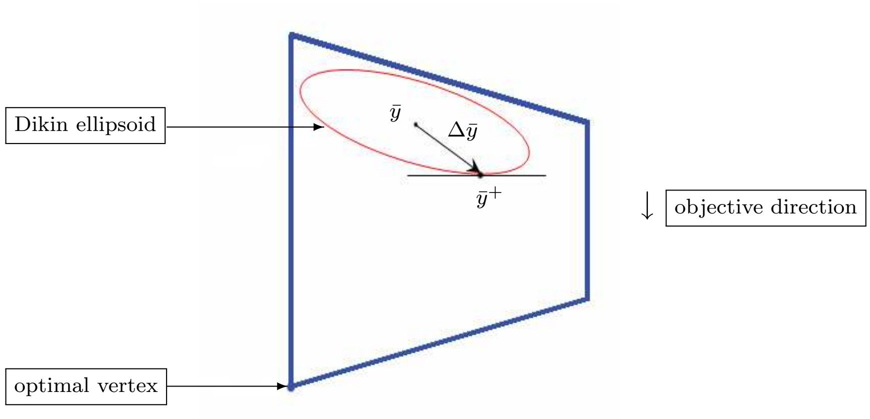


Figure 1.1: The Dikin ellipsoid and one iteration of the dual affine-scaling method.

be obtained by solving the problem

$$\begin{aligned} \max \quad & b^T \Delta y \\ \text{s.t.} \quad & A^T \Delta y + \Delta s = 0, \\ & \left\| \bar{S}^{-1} \Delta s \right\|_2^2 \leq 1. \end{aligned}$$

This problem is equivalent to

$$\begin{aligned} \max \quad & b^T \Delta y \\ \text{s.t.} \quad & \left\| \bar{S}^{-1} A^T \Delta y \right\|_2^2 \leq 1, \end{aligned}$$

whose solution is explicitly given as $\Delta\bar{y} = \left(A\bar{S}^{-2}A^T\right)^{-1} b$. In addition, we get $\Delta\bar{s} = -A^T\left(A\bar{S}^{-2}A^T\right)^{-1} b$.

There are several examples [9, 19, 25, 43] for which the (primal/dual) affine-scaling method with large step-sizes fails to converge to an optimal point. However, in [46] it has been shown that for shorter step-sizes, the algorithm is convergent to an optimal solution. Interestingly, the affine-scaling direction at a point on the *central path*[8] is tangent to it. Therefore, the continuous trajectory of the vector field induced by these directions initiated at a point on the central path lies on the central path.

[8]See page 12 for the definition of the central path.

On the other hand, Megiddo and Shub [27] observed that a continuous trajectory of the (primal/dual) affine-scaling vector field may visit all the vertices of the Klee-Minty cube before reaching the optimal solution. A similar result has been established by the authors for the (primal/dual) affine-scaling method with tiny step-sizes. Their results have strengthened the belief that the (primal/dual) affine-scaling method is an exponential-time algorithm in its worst case. In Chapter 3, we explain Megiddo and Shub's observation [27] in details as a pathological example for trajectories of the affine-scaling vector field.

Analogous to other IPMs, there is a primal-dual variant of the affine-scaling method. In the primal-dual affine-scaling method, we use Newton's method to solve the nonlinear system (1.1.3). Recall that every solution of system (1.1.3) is an optimal solution of problem (1.1.1). The Newton direction at a strictly feasible point $(\bar{x}; \bar{y}; \bar{s}) \in \mathcal{F}_+$ is obtained by solving the following linear system for $(\Delta x; \Delta y; \Delta s)$:

$$\begin{aligned} A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \\ \bar{S}\Delta x + \bar{X}\Delta s &= -\bar{x}\bar{s}. \end{aligned} \tag{1.3.2}$$

This system has a unique solution since, by assumption, the matrix A has full row rank. The direction $(\Delta x; \Delta y; \Delta s)$ is often referred to as the *primal-dual affine-scaling direction*. The solution of system (1.3.2) is explicitly given as

$$\begin{aligned} \Delta\bar{x} &= \bar{S}^{-1}\bar{X}A^T(A\bar{S}^{-1}\bar{X}A^T)^{-1}b - \bar{x}, \\ \Delta\bar{y} &= (A\bar{S}^{-1}\bar{X}A^T)^{-1}b, \\ \Delta\bar{s} &= -A^T(A\bar{S}^{-1}\bar{X}A^T)^{-1}b, \end{aligned}$$

To compute the next iterate, we move in the direction $(\Delta\bar{x}; \Delta\bar{y}; \Delta\bar{s})$ with a proper step-size. Therefore, the primal-dual affine-scaling method generates a sequence of strictly feasible points of the primal-dual problem. This sequence is convergent [46] to a solution of system (1.1.3), which yields an optimal solution for the primal-dual problem. Jensen, Roos, and Terlaky [20] proved that a family of primal-dual affine-scaling methods solves LPs in polynomial time providing that the starting point is well centered, i.e., it is close enough to the central path.

1.3.2 Potential-Reduction Methods

The key setting in the potential-reduction method is to define a *potential function* to measure the progress of the method. Potential functions for LPs were first introduced for the primal problem. Karmarkar [21] used a type of potential function for his projective method. However, the success of primal-dual variants of IPMs inspired researchers to use potential functions in the primal-dual space. Kojima, Mizuno, and Yoshise [24] obtained an $\mathcal{O}(\sqrt{n}L)$ iteration complexity upper bound for their primal-dual potential-reduction method, which is the best-known complexity bound for IPMs. For a detailed discussion on potential-reduction methods the reader may consult [2].

We consider the primal-dual potential function of Tanabe [41] and Todd and Ye [45] defined as

$$\phi_\rho(x, s) = \rho \log(x^T s) - e^T \log(xs), \tag{1.3.3}$$

for some parameter $\rho > n$. We can easily verify that

$$\phi_\rho(x, s) = (\rho - n) \log(x^T s) + \phi_n(x, s).$$

From the arithmetic-geometric mean inequality, we have

$$\frac{1}{n} x^T s \geq \prod_{i=1}^{n} (x_i s_i)^{\frac{1}{n}}.$$

Taking the logarithm of both sides we obtain $\phi_n(x, s) \geq n \log n$. Therefore, we have

$$\phi_\rho \to -\infty \quad \Longleftrightarrow \quad x^T s \to 0.$$

This idea forms the basis of potential-reduction methods. In fact, we generate a sequence of strictly feasible primal-dual points $(x^k; y^k; s^k) \in \mathcal{F}_+$ such that $\phi_\rho(x^k, s^k) \to -\infty$. In potential-reduction methods, the potential function is reduced by at least a constant amount at each iteration, hence, pushing the duality gap $x^T s$ to zero while keeping primal and dual feasibility. Therefore, by Theorem 1.1.2, the sequence $(x^k; y^k; s^k)$ converges to an optimal solution of the primal-dual problem.

1.3.3 Central-Path-Following IPMs

Central-path-following IPMs have become the most popular methods among all IPMs. In the last decade, the theory of the primal-dual approach has been well developed, and techniques based on primal-dual central-path-following IPMs have emerged as the basis of modern optimization software packages. In this subsection, we define the notions of the *analytic center*, the *central path*, and *proximity measures*. Then, we describe the *small-update*, *large-update*, and *predictor-corrector* variants of central-path-following IPMs. The interested reader is referred to [38] for a detailed discussion.

Central-path-following IPMs are closely related to the classical *barrier function methods*; see [16] and the references therein. Consider the primal problem (1.1.1). Defining the standard *primal logarithmic barrier function* $F_{\mathcal{P}}(x) = -e^T \log x$ over $\mathbb{R}^n_+$, we consider the primal barrier problem

$$\begin{aligned} \min \quad & c^T x + \mu F_{\mathcal{P}}(x) \\ \text{s.t.} \quad & Ax = b, \\ & x > 0, \end{aligned} \tag{1.3.4}$$

in which a strictly convex function is minimized over an affine subspace intersected by the positive orthant. The positive parameter μ is called the *central-path parameter* for the reason that will become clear later. Assume that A has a full row rank and Assumption 1.3.1 (IPC) holds. Then, for a given $\mu > 0$, problem (1.3.4) has a unique minimizer. This unique minimizer is referred to as the *μ-center* of the primal problem and is denoted by $x(\mu)$. From the necessary and sufficient optimality conditions we have

$$\begin{aligned} Ax = b, &\quad x > 0, \\ A^T y + s = c, &\quad s > 0, \\ xs = \mu e. & \end{aligned} \tag{1.3.5}$$

The unique solution of this system is called the μ-center of the primal-dual problem and is denoted by $z(\mu)$. System (1.3.5) can be interpreted as perturbed optimality conditions (1.1.3) in which the right-hand side of the complementarity condition $xs = 0$ is perturbed by $\mu > 0$. Interestingly, the same optimality conditions are achieved if we consider the dual barrier problem

$$\begin{aligned} \max \quad & b^T y + \mu F_{\mathcal{D}}(y) \\ \text{s.t.} \quad & A^T y + s = c, \\ & s > 0, \end{aligned} \tag{1.3.6}$$

where $F_{\mathcal{D}}(y) = e^T \log(c - A^T y)$ is the standard dual logarithmic barrier function. This problem has also a unique maximizer which is called the μ-center of the dual problem and is denoted by $\big(y(\mu); s(\mu)\big)$. The following theorem summarizes these observations and shows the relation between the solutions of barrier problems (1.3.4) and (1.3.6) and system (1.3.5). For the proof, we refer the reader to [38].

Theorem 1.3.2. *Assume that $\mu > 0$ and A has full row rank. Then, the following statements are equivalent:*

1. *Assumption 1.3.1* (IPC) *holds.*
2. *Primal barrier problem* (1.3.4) *has a unique minimizer $x(\mu)$.*
3. *Dual barrier problem* (1.3.6) *has a unique maximizer $\big(y(\mu); s(\mu)\big)$.*
4. *Primal-dual system* (1.3.5) *has a unique solution $z(\mu)$.*

Moreover, we have $z(\mu) = \big(x(\mu); y(\mu); s(\mu)\big)$.

According to this theorem, the mappings

$$\begin{aligned} &\mu \mapsto x(\mu) = \arg\min \left\{ c^T x + \mu F_{\mathcal{P}}(x) \mid x \in \mathcal{P}_+ \right\}, \\ &\mu \mapsto \big(y(\mu); s(\mu)\big) = \arg\max \left\{ b^T y + \mu F_{\mathcal{D}}(y) \mid (y; s) \in \mathcal{D}_+ \right\}, \\ &\mu \mapsto z(\mu) = \big(x(\mu); y(\mu); s(\mu)\big), \end{aligned} \tag{1.3.7}$$

are well-defined functions from $(0, +\infty)$ into, respectively, $\mathcal{P}_+$, $\mathcal{D}_+$, and $\mathcal{F}_+$. Recall that the value of each of these functions at a given μ is referred to as the μ-center of the corresponding space. We now define the central path, see Figure 1.2, which plays a vital role in the theory of central-path-following IPMs.

Definition 1.3.3. The set of all μ-centers is called the *central path* of the corresponding space. More precisely, the central path is defined as follows:

the central path in the primal space	$\{x(\mu) \mid \ \mu > 0\}$,
the central path in the dual space	$\left\{\big(y(\mu); s(\mu)\big) \mid \ \mu > 0\right\}$,
the central path in the primal-dual space	$\{z(\mu) \mid \ \mu > 0\}$.

We may also refer to the set $\{y(\mu) \mid \ \mu > 0\}$ as the central path in the y-space. Furthermore, for the sake of simplicity, each of these sets might be referred to as the central path if the space is clear from the context. We denote the central path by $\mathcal{C}$.

Now, let the set $\mathcal{P}_\oplus$ be bounded. As $\mu \to \infty$, the primal central path converges to a point which is called the *analytic center*, see Figure 1.2, of the feasible region $\mathcal{P}_\oplus$. The analytic center is not defined for unbounded sets. Moreover, it is independent of the objective function and can be defined as follows.

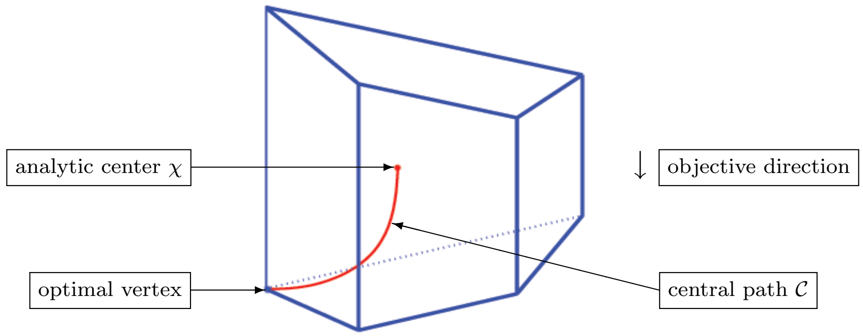


Figure 1.2: The analytic center and the central path.

Definition 1.3.4. If $\mathcal{P}_{\oplus}$ is bounded, then its *analytic center* is defined as

$$\chi^{\mathcal{P}} = \arg\min \{F_{\mathcal{P}}(x) \mid x \in \mathcal{P}_{+}\}.$$

Similarly, if $\mathcal{D}_{\oplus}$ is bounded, then its *analytic center* is defined as

$$\chi^{\mathcal{D}} = \arg\min \{F_{\mathcal{D}}(y) \mid (y; s) \in \mathcal{D}_{+}\}.$$

The main idea of central-path-following IPMs is to follow the central path. To do this, we define a neighborhood of the central path and generate a sequence of points within the neighborhood that converges to an optimal solution. Let us first show how we compute a search direction at a given point and then define a neighborhood of the central path.

We compute a search direction at every point by applying Newton's method to system (1.3.5). In other words, given a strictly feasible point $(x; y; s) \in \mathcal{F}_{+}$, we would like to find a direction $(\Delta x; \Delta y; \Delta s)$ such that $(x+\Delta x; y+\Delta y; s+\Delta s) \in \mathcal{F}_{+}$. Therefore, the direction is required to satisfy

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

or equivalently

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

Neglecting the second-order term $\Delta x\Delta s$ in the third equation, we obtain a search direction $(\Delta x; \Delta y; \Delta s)$ by solving the linear system

$$\begin{aligned} A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \\ S\Delta x + X\Delta s &= \mu e - xs. \end{aligned} \tag{1.3.8}$$

Definition 1.3.5. We refer to system (1.3.8) as the *Newton system*. Since A has full row rank, the Newton system has a unique solution, which is referred to as the *Newton direction*. Moreover, the Newton direction at $\mu = 0$ is referred to as the *primal-dual affine-scaling direction*.

To define a neighborhood of the central path, we need a measure for the distance of an iterate to the central path. The following reformulation of system (1.3.8) helps us to define a *proximity measure* to measure the distance of a point to the central path. We scale $(\Delta x; \Delta y; \Delta s)$ to $(d_x; d_y; d_s)$ as

$$d_x = \sqrt{\frac{s}{\mu x}}\Delta x, \quad d_y = \frac{1}{\sqrt{\mu}}\Delta y, \quad d_s = \sqrt{\frac{x}{\mu s}}\Delta s.$$

Hence, the Newton system (1.3.8) is equivalent to the scaled system

$$\begin{aligned} \bar{A}d_x &= 0, \\ \bar{A}^T d_y + d_s &= 0, \\ d_x + d_s &= \sqrt{\frac{\mu}{xs}} - \sqrt{\frac{xs}{\mu}}, \end{aligned}$$

where $\bar{A} = A\,\mathrm{diag}(\sqrt{x/s})$. A natural way of defining a proximity measure is

$$\psi(x, s, \mu) = \frac{1}{2}\left\|\sqrt{\frac{xs}{\mu}} - \sqrt{\frac{\mu}{xs}}\right\|. \tag{1.3.9}$$

We define $\mu_g = x^T s/n$, which is called the gap-μ. For the sake of simplicity, we denote $\psi(x, s, \mu_g)$ by $\psi(x, s)$.

Definition 1.3.6. For a given $\delta > 0$, the set

$$\mathcal{N}_\psi(\delta) = \{(x; y; s) \in \mathcal{F}_+ \mid \psi(x, s) \leq \delta\},$$

is called the *δ-neighborhood of the central path*, in which the proximity measure ψ is given by (1.3.9).

Small-Update Central-Path-Following IPMs

In the small-update[9] variant, we start with a point $(x^0; y^0; s^0) \in \mathcal{F}_+$ in a small neighborhood of the central path. We set the initial central-path parameter to $\mu^0 = \mu_g^0$ and update it as $\mu^+ = \sigma\mu$, where the μ-update parameter σ is a constant depending only on the size of the problem. Then, we take a full Newton step (the step-size equal to one) to compute the next iterate. Algorithm 1.1 describes the small-update variant. The output of the algorithm is a primal-dual strictly feasible point with the duality gap smaller than a given precision ε. This algorithm is well defined as shown in the following theorem.

Theorem 1.3.7. [38] *If $\delta = 1/\sqrt{2}$ and $\sigma = 1 - 0.5/\sqrt{n}$ then Algorithm 1.1 terminates at no more than*

$$\left\lceil 2\sqrt{n} \log \frac{n\mu_0}{\varepsilon} \right\rceil$$

iterations. The output of the algorithm is a primal-dual strictly feasible point $(x; y; s)$ that satisfies $x^T s < \varepsilon$.

Algorithm 1.1 always terminates at an interior point that is as close to an optimal solution as we specify by choosing an accuracy parameter ε. To obtain an exact optimal solution of an LP, we first choose $\varepsilon \leq \mathcal{O}(\frac{1}{n})2^{-5L}$ (see Theorem 2.2.4) and apply Algorithm 1.1 to the problem to find an interior point with the duality gap smaller than ε. Then, we apply a *rounding procedure*[10] to obtain an exact optimal solution. Having had an exact optimal solution, we can find [26] an optimal vertex solution of an LP in strongly polynomial time. With this choice of ε, the number of iterations required by the algorithm is $\mathcal{O}(\sqrt{n}L)$.

[9] It might be referred to as the *short-step* variant as well.

[10] See § 2.2 for details.

Algorithm 1.1: Small-update central-path-following IPMs.

parameters: the size of the central-path neighborhood $\delta = 1/\sqrt{2}$
the μ-update parameter $\sigma = 1 - 0.5/\sqrt{n}$
input data : a given accuracy $\varepsilon > 0$
a primal-dual point $(x^0; y^0; s^0) \in \mathcal{N}_\psi(\delta)$
output data: a point $(x; y; s) \in \mathcal{F}_+$ that satisfies $x^T s \leq \varepsilon$

begin
$(x; y; s) = (x^0; y^0; s^0)$
$\mu = \mu_g$
while $n\mu \geq \varepsilon$ **do**
$\mu = \sigma\mu$
solve system (1.3.8) for $(\Delta x; \Delta y; \Delta s)$
$(x; y; s) = (x; y; s) + (\Delta x; \Delta y; \Delta s)$
end
end

Large-Update Central-Path-Following IPMs

Despite enjoying the best-known complexity result for IPMs, the small-update variant turns out to be slow as the μ-update parameter σ is only slightly smaller than one, particularly for large-scale problems. In the *large-update*[11] variant, we start with a point $(x^0; y^0; s^0) \in \mathcal{F}_+$ in a large neighborhood of the central path. We set the initial central-path parameter to $\mu^0 = \mu_g^0$ and update it as $\mu^+ = \sigma\mu$, where the μ-update parameter σ is a small positive constant, e.g., $\sigma = 10^{-3}$. Then, we do a line search for the proximity measure in the Newton direction to compute the next iterate and repeat this procedure until entering the neighborhood of the central path. Algorithm 1.2 describes the large-update variant. The output of the algorithm is a primal-dual strictly feasible point with the duality gap smaller than a given precision ε. This algorithm is well defined as shown in the following theorem.

Theorem 1.3.8. [**38**] *Let* $0 < \sigma < 1$ *and*

$$\delta = \frac{\sqrt{R}}{2\sqrt{1+\sqrt{R}}}, \quad \text{where } R = \left(\frac{1}{\sigma} - 1\right)\sqrt{n}.$$

[11] It might be referred to as the *long-step* variant as well.

Algorithm 1.2: Large-update central-path-following IPMs.

parameters: the size of the central-path neighborhood δ, e.g., $\delta = 3\sqrt[8]{n}$
the μ-update parameter σ, e.g., $\sigma = 10^{-3}$
a damping factor, e.g., $\eta = 0.99$
input data : a given accuracy $\varepsilon > 0$
a primal-dual point $(x^0; y^0; s^0) \in \mathcal{N}_\psi(\delta)$
output data: a point $(x; y; s) \in \mathcal{F}_+$ that satisfies $x^T s \leq \varepsilon$

begin
  $(x; y; s) = (x^0; y^0; s^0)$
  $\mu = \mu_g$
  while $n\mu \geq \varepsilon$ **do**
    $\mu = \sigma\mu$
    while $(x; y; s) \notin \mathcal{N}_\psi(\delta)$ **do**
      solve system (1.3.8) for $(\Delta x; \Delta y; \Delta s)$
      $\alpha_{\max} = \max\{\alpha \mid (x + \alpha\Delta x; y + \alpha\Delta y; s + \alpha\Delta s) \geq 0\}$
      $\bar{\alpha} = \min\{\eta\alpha_{\max}, 1\}$
      $\alpha = \arg\min\{\psi(x + \alpha\Delta x, s + \alpha\Delta s) \mid 0 \leq \alpha \leq \bar{\alpha}\}$
      $(x; y; s) = (x; y; s) + \alpha(\Delta x; \Delta y; \Delta s)$
    end
  end
end

Then, Algorithm 1.2 terminates at no more than

$$\left\lceil \frac{r}{1-\sigma} \log \frac{n\mu_0}{\varepsilon} \right\rceil$$

iterations, where $r = \lceil 2(1 + \sqrt{R})^4 \rceil$. *The output of the algorithm is a primal-dual strictly feasible point* $(x; y; s)$ *that satisfies* $x^T s < \varepsilon$.

Corollary 1.3.9. *In the large-update variant* σ *is an absolute constant. Therefore, we have* $r = \mathcal{O}(n)$, *implying that the number of iteration required by the large-update variant to solve an LP is*

$$\mathcal{O}\left(n \log \frac{n\mu_0}{\varepsilon}\right).$$

Predictor-Corrector Central-Path-Following IPMs

We describe two variants of predictor-corrector central-path-following IPMs introduced by Mizuno, Todd, and Ye [30] and Mehrotra [28]. We refer to the first variant as MTYPC and to the latter one as MPC.

In the MTYPC variant, we use two nested neighborhoods of the central path and start with a point $(x^0; y^0; s^0)$ inside the inner neighborhood. Every iteration of the MTYPC method is composed of two steps: the predictor step and the corrector step. In the predictor step, we compute the Newton direction with $\mu = 0$. The resulting direction is called the *affine-scaling direction* and is denoted by $(\Delta^a x; \Delta^a y; \Delta^a s)$. Then, we move along the affine-scaling direction to the boundary of the outer neighborhood. In the corrector step, we compute the Newton direction with $\mu = \mu_g$. The resulting direction aims to the μ_g-center and is called the *centering direction*. The next iterate is obtained by taking a full Newton step which is ensured [30] to belong to the inner neighborhood.

Algorithm 1.3 describes the MTYPC method. The output of the algorithm is a primal-dual strictly feasible point with the duality gap smaller than a given precision ε. The following theorem provides conditions upon which Algorithm 1.3 is well defined

Theorem 1.3.10. [30] *If $\delta_1 = 1/2$ and $\delta_2 = 1/4$ then Algorithm 1.3 terminates at no more than*

$$\left\lceil 2\sqrt{n} \log \frac{n\mu_0}{\varepsilon} \right\rceil$$

iterations. The output of the algorithm is a primal-dual strictly feasible point $(x; y; s)$ that satisfies $x^T s < \varepsilon$.

The MPC variant has appeared to be the most efficient algorithm for solving large-scale LPs and is implemented in most state-of-the-art optimization software packages. One of the advantages of this method is that unlike the MTYPC method that solves two linear systems with different coefficient matrices in every step, the MPC variant solves two linear systems with identical coefficient matrices but different right-hand sides at each iteration. Therefore, only one Cholesky or Bunch-Parlett factorization with two back-substitutions is required. In the MPC variant, we use the so-called negative-infinity-norm neighborhood of the central path, defined as

$$\mathcal{N}_{-\infty}(\delta) = \{(x; y; s) \in \mathcal{F}_+ \mid xs \geq \delta\mu_g e\},$$

Algorithm 1.3: Mizuno-Todd-Ye's predictor-corrector variant.

parameters: the size of the outer neighborhood $\delta_1 = 1/2$
the size of the inner neighborhood $\delta_2 = 1/4$
input data : a given accuracy $\varepsilon > 0$
a primal-dual point $(x^0; y^0; s^0) \in \mathcal{N}_\psi(\delta_2)$
output data: a point $(x; y; s) \in \mathcal{F}_+$ that satisfies $x^T s \le \varepsilon$

begin
  $(x; y; s) = (x^0; y^0; s^0)$
  while $n\mu \ge \varepsilon$ **do**
    `// predictor step.`
    $\mu = 0$
    solve system (1.3.8) for $(\Delta^a x; \Delta^a y; \Delta^a s)$
    $\alpha_{\max} = \max\{\alpha \mid (x + \alpha\Delta^a x; y + \alpha\Delta^a y; s + \alpha\Delta^a s) \in \mathcal{N}_\psi(\delta_1)\}$
    $(x; y; s) = (x; y; s) + \alpha_{\max}(\Delta^a x; \Delta^a y; \Delta^a s)$
    `// corrector step.`
    $\mu = \mu_g$
    solve system (1.3.8) for $(\Delta x; \Delta y; \Delta s)$
    $(x; y; s) = (x; y; s) + (\Delta x; \Delta y; \Delta s)$
  end
end

where $\delta \in (0, 1)$. Note that as δ approaches to zero, the neighborhood covers almost the whole primal-dual feasible region.

In the MPC variant, we start with a point $(x^0; y^0; s^0) \in \mathcal{N}_{-\infty}(\delta)$, where δ is a small positive number, e.g., $\delta = 10^{-3}$. Then, we compute the affine-scaling direction $(\Delta^a x; \Delta^a y; \Delta^a s)$ without taking any step. Let α_a be the largest step-size along this direction from the current iterate to the boundary of the feasible region. We define

$$\mu_a = \frac{(x + \alpha_a \Delta^a x)^T (s + \alpha_a \Delta^a s)}{n}$$

and solve the system

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

with $\mu = \mu_a^3/\mu_g^2$ for the direction $(\Delta x; \Delta y; \Delta s)$. Then, we move in this direction to the boundary of the neighborhood $\mathcal{N}_{-\infty}(\delta)$. Algorithm 1.4 describes the

Algorithm 1.4: Mehrotra's predictor-corrector variant.

parameters: the size of the neighborhood δ, e.g., $\delta = 10^{-3}$
input data : a given accuracy $\varepsilon > 0$
a primal-dual point $(x^0; y^0; s^0) \in \mathcal{N}_{-\infty}(\delta)$
output data: a point $(x; y; s) \in \mathcal{F}_+$ that satisfies $x^T s \leq \varepsilon$

begin
$(x; y; s) = (x^0; y^0; s^0)$
while $n\mu \geq \varepsilon$ **do**
`// predictor step (no actual step is taken.)`
$\mu = 0$
solve system (1.3.8) for $(\Delta^a x; \Delta^a y; \Delta^a s)$
$\alpha_a = \max\{\alpha \mid (x + \alpha\Delta^a x; y + \alpha\Delta^a y; s + \alpha\Delta^a s) \geq 0\}$
`// corrector step.`
$\mu_a = (x + \alpha_a \Delta^a x)^T (s + \alpha_a \Delta^a s)/n$
$\mu = \mu_a^3 / \mu_g^2$
solve system (1.3.8) for $(\Delta x; \Delta y; \Delta s)$
$\alpha = \max\{\alpha \mid (x + \alpha\Delta x; y + \alpha\Delta y; s + \alpha\Delta s) \in \mathcal{N}_{-\infty}(\delta)\}$
$(x; y; s) = (x; y; s) + \alpha(\Delta x; \Delta y; \Delta s)$
end
end

MPC variant. The output of the algorithm is a primal-dual strictly feasible point with the duality gap smaller than a given precision ε.

1.4 An Overview of the Thesis

In Chapter 2, we study some important properties of the central path and the effects of redundancy on it. This study assists us to understand the basic idea behind the pathological examples for central-path-following IPMs that are described in Chapters 4–7. We also describe the rounding procedure for obtaining an exact optimal solution.

In Chapter 3, we review some pathological linear optimization problems and linear complementarity problems for some popular simplex-type methods. In addition, we review some unfavorable interior-point trajectories of linear and nonlinear convex optimization problems.

In Chapter 4, we provide an example, based on the Klee-Minty (KM) problem, whose central path has an exponential number of "sharp" turns with respect to the dimension of the problem. More precisely, by carefully adding an exponential number of redundant constraints to the KM example (3.1.1), the central path is forced to follow the simplex path, the edge-path followed by the simplex method, visiting a small neighborhood of each of the 2^n vertices of the KM cube. Note that the redundant constraints are placed parallel to the facets of the KM cube containing the origin at uniform distances.

In Chapter 5, we refine the construction by allowing the distances to decay geometrically. This construction tightens the gap between the iteration complexity lower and upper bounds for central-path-following IPMs.

In Chapter 6, we present a different construction of the redundant KM example in which the redundant constraints are put parallel to the coordinate hyperplanes. We study the following two cases: uniform distances and geometrically decaying distances. It turns out that the latest construction provides even tighter iteration complexity bounds, yielding a nearly worst-case example for central-path-following IPMs.

In Chapter 7, we show that the distances of the redundant constraints to the corresponding facets can be set to zero, but the number of the redundant constraints dramatically increases.

Finally, in Chapter 8, we conclude the thesis by providing a summary of the whole thesis and outlining directions for the future work.

Chapter 2

Properties of the Central Path

In Chapter 1, we introduced an LP in both primal and dual forms and defined the analytic center and the central path as key settings for central-path-following IPMs. In this chapter, we state a well-known result [38] that the central path of an LP has a unique limit point in the optimal set. Therefore, since an exact solution cannot be obtained in a finite number of iterations, we employ some finite-termination techniques that yield an exact optimal solution. In addition, we observe that the central path is an analytic curve and that both the analytic center and the central path are analytic concepts not geometric ones. In other words, their definitions depend on the representation of the feasible region. We also study the effects of redundancy on the analytic center and the central path.

2.1 Convergence of the Central Path

In this section, we present some results about the limiting behavior of the central path as the central-path parameter μ approaches to zero. We assume that Assumption 1.3.1 (IPC) holds so that the central path is well defined as shown in § 1.3.3. We start with the following theorem that shows that the central path is an analytic curve; i.e., it is infinitely many times differentiable. Moreover, its derivatives are bounded.

Theorem 2.1.1. [38] *All derivatives of $x(\mu)$, $y(\mu)$, and $s(\mu)$ defined as* (1.3.7) *with respect to μ exist and are bounded.*

Next theorem shows that the primal-dual optimal set is nonempty and bounded and contains a strictly complementary solution.

Theorem 2.1.2. [**38**] *If Assumption 1.3.1* (IPC) *holds then*

1. *the primal-dual optimal set $\mathcal{F}_*$ is nonempty and bounded.*
2. *A strictly complementary solution exists. In other words, there exists a vector $(x^*; y^*; s^*) \in \mathcal{F}_*$ such that $x^* + s^* > 0$.*

Let $(x^*; y^*; s^*)$ be a strictly complementary solution. Since $x^* s^* = 0$ and $x^* + s^* > 0$, the sets

$$\mathcal{B} = \{i \in \mathcal{I} \mid x_i^* > 0\}, \qquad \mathcal{N} = \{i \in \mathcal{I} \mid s_i^* > 0\},$$

form a partition for the index set $\mathcal{I} = \{1, 2, \ldots, n\}$. This partition is called *the optimal partition*[1] and is denoted by $(\mathcal{B}, \mathcal{N})$. Recall that the optimal set of the primal-dual problem is the set of all solutions of system (1.1.3). We can easily see that for a given $(x; y; s) \in \mathcal{F}_\oplus$, we have

$$xs = 0 \quad \Longleftrightarrow \quad x_\mathcal{N} = 0 \ \text{ and } \ s_\mathcal{B} = 0.$$

Therefore, the primal and dual *optimal sets* are, respectively, given by

$$\begin{aligned} \mathcal{P}_* &= \{x \mid A_{\cdot\mathcal{B}} x_\mathcal{B} = b, x_\mathcal{B} > 0, x_\mathcal{N} = 0\}, \\ \mathcal{D}_* &= \left\{(y; s) \mid A_{\cdot\mathcal{B}}^T y + s_\mathcal{N} = c_\mathcal{N}, s_\mathcal{B} = 0, s_\mathcal{N} > 0\right\}, \end{aligned}$$

and the primal-dual optimal set is given by

$$\mathcal{F}_* = \{(x; y; s) \mid x \in \mathcal{P}_*, \quad (y; s) \in \mathcal{D}_*\}.$$

In the following, we define the analytic centers of the primal, dual, and primal-dual optimal sets.

Definition 2.1.3. We define the *analytic centers of the optimal sets $\mathcal{P}_*$, $\mathcal{D}_*$, and $\mathcal{F}_*$* as

$$\begin{aligned} \chi^{\mathcal{P}_*} &= \arg\max_{x \in \mathcal{P}_*} \ e^T \log x_\mathcal{B}, \\ \chi^{\mathcal{D}_*} &= \arg\max_{(y;s) \in \mathcal{D}_*} \ e^T \log s_\mathcal{N}, \\ \chi^{\mathcal{F}_*} &= \left(\chi^{\mathcal{P}_*}; \chi^{\mathcal{D}_*}\right). \end{aligned}$$

[1] Note that $\mathcal{B}$ and $\mathcal{N}$ remain invariant for every strictly complementary solution.

The following theorem shows that the central path, as a parameterized curve with respect to μ, converges to the analytic center of the optimal set as $\mu \to 0$.

Theorem 2.1.4. [38] *As $\mu \to 0$, the μ-centers converge to the analytic center of the optimal set. In other words, we have*

$$\big(x(\mu); y(\mu); s(\mu)\big) \to \chi^{\mathcal{F}_*},$$

where $x(\mu)$, $y(\mu)$, and $s(\mu)$ are given by (1.3.7).

Example 2.1.5. Consider the problem

$$\begin{array}{ll} \max & e^T x \\ \text{s. t.} & e^T x \le 1, \\ & \quad x \ge 0, \end{array} \tag{2.1.1}$$

where $x \in \mathbb{R}^n$. The feasible region represented by the constraints of this problem is called the *n-simplex* and is denoted by $\mathcal{S}^n$. The optimal set of this problem is $\left\{x \in \mathbb{R}^n_\oplus \mid e^T x = 1\right\}$. Problem (2.1.1) in the standard primal form is given as

$$\begin{array}{ll} \min & -e^T x \\ \text{s. t.} & e^T x + x_0 = 1, \\ & x, \quad x_0 \ge 0. \end{array}$$

The dual form of this problem is

$$\begin{array}{ll} \max & y_0 \\ \text{s. t.} & y_0 e \ +s = -e, \\ & y_0 \ +s_0 = 0, \\ & s_0, \quad s \ge 0, \end{array}$$

whose unique optimal solution is given by $y_0 = -1$, $s_0 = 1$, and $s = 0$. Therefore, the sets $\mathcal{B} = \{1, \ldots, n\}$ and $\mathcal{N} = \{0\}$ form the optimal partition since $x_0 = 0$ and $x = (1/n; \cdots; 1/n)$ yield a strictly complementary solution. By definition, the analytic center of the optimal set of problem (2.1.1) in the x-space is the unique minimizer of the problem

$$\begin{array}{ll} \min & -e^T \log x \\ \text{s. t.} & e^T x = 1, \\ & \quad x > 0, \end{array}$$

which is $x^* = (1/n; \cdots; 1/n)$. On the other hand, for a given $\mu > 0$, the μ-center $x(\mu)$ is the unique minimizer of the problem

$$\begin{array}{ll} \min & -e^T x - \mu\big(\log(1-e^T x) + e^T \log x\big) \\ \text{s.t.} & e^T x < 1, \\ & x > 0. \end{array}$$

From the optimality condition, we get

$$\Big(1 + \frac{\mu}{1-e^T x}\Big)e = \mu x^{-1},$$

which implies that $x = \alpha e$ for some $0 < \alpha < 1/n$. Thus, we get

$$\frac{\alpha(n+1)-1}{\alpha(1-\alpha n)}\mu = 1,$$

implying, since $\mu > 0$, that $\alpha \to 1/n$ as $\mu \to 0$. As a consequence, we have

$$x(\mu) \to \left(\frac{1}{n}; \cdots; \frac{1}{n}\right) \qquad \text{as} \qquad \mu \to 0,$$

confirming the statement of Theorem 2.1.4. □

2.2 Rounding Procedure

As we have seen in § 1.3, IPMs terminate at an interior point of the feasible region with the duality gap smaller than a given precision ε. In this section, we present how an *exact* optimal solution can be obtained. First, we see that the *optimal partition* can be identified in polynomial time. Then, we discuss standard *rounding procedures*[2] that yield an exact optimal solution. A rounding procedure is a finite algorithm that produces a solution in the optimal set.

The discussion of this section considers polynomial-time IPMs that generate a sequence of strictly feasible points $(x; y; s) \in \mathcal{F}_+$ such that

$$xs > \frac{\beta}{n}(x^T s)e. \tag{2.2.1}$$

[2]It might be referred to as *finite-termination* procedures.

for some absolute constant $\beta > 0$. Güler and Ye [18] showed that many polynomial-time IPMs possess this property. We also recall the definition (1.3.1) of the bit-length L of the input-data, which is used throughout this section.

2.2.1 Identification of the Optimal Set

In this subsection, we show that IPMs that possess property (2.2.1) find the optimal partition in a finite number of iterations. We start with the following lemma that shows that there exists a strictly complementary solution whose positive components are well separated from zero.

Lemma 2.2.1. *There exist a positive* γ *and a strictly complementary solution* $(x^*; y^*; s^*) \in \mathcal{F}_\oplus$ *such that*

$$\begin{aligned} x^*_{\mathcal{B}} &\geq \gamma e > 2^{-L} e, & s^*_{\mathcal{B}} &= 0, \\ x^*_{\mathcal{N}} &= 0, & s^*_{\mathcal{N}} &\geq \gamma e > 2^{-L} e. \end{aligned} \tag{2.2.2}$$

Proof. The reader may consult [38, 39] for the proof. □

Theorem 2.2.2. *Let* $(x; y; s) \in \mathcal{F}_+$ *satisfy* (2.2.1)*. Then, we have*

$$\begin{aligned} x_{\mathcal{B}} &\geq \frac{\beta\gamma}{n} e, & s_{\mathcal{B}} &\leq \frac{x^T s}{\gamma} e, \\ x_{\mathcal{N}} &\leq \frac{x^T s}{\gamma} e, & s_{\mathcal{N}} &\geq \frac{\beta\gamma}{n} e. \end{aligned} \tag{2.2.3}$$

Proof. By Lemma 2.2.1, there exists a strictly complementary solution $(x^*; y^*; s^*) \in \mathcal{F}_\oplus$ that satisfies (2.2.2). Since $s - s^* = -A^T(y - y^*)$ and $A(x - x^*) = 0$, we have

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

Recalling that $s^*_{\mathcal{B}} = 0$ and $x^*_{\mathcal{N}} = 0$, we get

$$x^T_{\mathcal{N}} s^*_{\mathcal{N}} + s^T_{\mathcal{B}} x^*_{\mathcal{B}} = x^T s,$$

implying that $s^T_{\mathcal{B}} x^*_{\mathcal{B}} \leq x^T s$. Hence, from the fact that $x^*_{\mathcal{B}} \geq \gamma e$, we obtain

$$s_{\mathcal{B}} \leq \frac{x^T s}{\gamma} e.$$

Since $(x;y;s)$ satisfies (2.2.1), we have $x_{\mathcal{B}}s_{\mathcal{B}} > \frac{\beta}{n}(x^Ts)e$. Combining with $x^Ts \geq s_{\mathcal{B}}^T x_{\mathcal{B}}^*$, we get $x_{\mathcal{B}} \geq \frac{\beta}{n} x_{\mathcal{B}}^*$, implying $x_{\mathcal{B}} \geq \frac{\beta\gamma}{n}e$. Inequalities for $x_{\mathcal{N}}$ and $s_{\mathcal{N}}$ can be proven similarly. □

Corollary 2.2.3. *Let* $(x;y;s) \in \mathcal{F}_+$ *satisfy* (2.2.1). *If* $x^Ts < \beta\gamma^2/n$ *then the sets* $\{i \mid x_i \geq s_i\}$ *and* $\{i \mid x_i < s_i\}$ *form the optimal partition. In particular, since* $\gamma > 2^{-L}$, *the statement holds if* $x^Ts < \beta 2^{-2L}/n$.

Proof. It suffices to show that $x_{\mathcal{B}} \geq s_{\mathcal{B}}$ and $x_{\mathcal{N}} < s_{\mathcal{N}}$ which follow immediately from Theorem 2.2.2 and the fact that $\beta\gamma/n > x^Ts/\gamma$. □

2.2.2 Finding an Exact Optimal Solution

In this subsection, we discuss two rounding procedures for finding an *exact optimal solution*–a point in the relative interior of the optimal set. Having this solution, we can obtain [26] a feasible basis solution in strongly polynomial time.

Ye's Approach

Ye [50] suggests an orthogonal projection procedure that produces an exact optimal solution. This method can be applied at each iteration to test if the optimal partition is obtained, and if an exact optimal solution is reached. To explain this method, we let $(x^k;y^k;s^k) \in \mathcal{F}_+$ be the current iterate and set $\mathcal{B}_k = \left\{i \mid x_i^k \geq s_i^k\right\}$ and $\mathcal{N}_k = \left\{i \mid x_i^k < s_i^k\right\}$. Then, we solve the problems

$$\begin{array}{ll} \min & \left\| x_{\mathcal{B}_k} - x_{\mathcal{B}_k}^k \right\| \\ \text{s.t.} & A_{\cdot\mathcal{B}_k} x_{\mathcal{B}_k} = b, \end{array}$$

and

$$\begin{array}{ll} \min & \left\| y - y^k \right\| \\ \text{s.t.} & A_{\cdot\mathcal{B}_k}^T y = c_{\mathcal{B}_k}, \end{array}$$

whose solutions can be obtained by projecting $x_{\mathcal{B}_k}^k$ and y^k orthogonally onto the hyperplanes $\{x_{\mathcal{B}_k} \mid A_{\cdot\mathcal{B}_k} x_{\mathcal{B}_k} = b\}$ and $\left\{y \mid A_{\cdot\mathcal{B}_k}^T y = c_{\mathcal{B}_k}\right\}$, respectively. If the resulting solutions $x_{\mathcal{B}_k}^*$ and y^* satisfy

$$x_{\mathcal{B}_k}^* > 0, \qquad s_{\mathcal{N}_k}^* = c_{\mathcal{N}_k} - A_{\cdot\mathcal{N}_k}^T y^* > 0, \tag{2.2.4}$$

then, $(x^*_{\mathcal{B}_k}; 0; y^*; 0; s^*_{\mathcal{N}_k})$ is a strictly complementary solution in the primal-dual space. Hence $(\mathcal{B}_k, \mathcal{N}_k)$ is the optimal partition. However, if inequalities (2.2.4) are violated, we continue running the algorithm until they are satisfied. This procedure is well defined since $\left\|x_{\mathcal{B}_k} - x^k_{\mathcal{B}_k}\right\|$ and $\left\|y - y^k\right\|$ converge to zero as $(x^k)^T s^k \to 0$.

Mehrotra-Ye's Approach

To obtain an exact optimal solution, Mehrotra and Ye [29] propose a procedure in which two related linear systems are solved using only one factorization. Their algorithm is as follows. At the current iterate $(x^k; y^k; s^k) \in \mathcal{F}_+$, we set $\mathcal{B}_k = \left\{i \mid x^k_i \geq s^k_i\right\}$ and $\mathcal{N}_k = \left\{i \mid x^k_i < s^k_i\right\}$. Then, we solve the linear systems

$$\begin{aligned} A_{\cdot\mathcal{B}_k} \Delta x_{\mathcal{B}_k} &= b - A_{\cdot\mathcal{B}_k} x^k_{\mathcal{B}_k} = A_{\cdot\mathcal{N}_k} x^k_{\mathcal{N}_k}, \\ A^T_{\cdot\mathcal{B}_k} \Delta y &= c_{\mathcal{B}_k} - A^T_{\cdot\mathcal{B}_k} y^k = s^k_{\mathcal{B}_k}, \end{aligned} \tag{2.2.5}$$

by the Gaussian elimination for $\Delta x_{\mathcal{B}_k}$ and Δy. Note that linearly dependent rows and/or columns of $A_{\cdot\mathcal{B}_k}$ are deleted during the elimination procedure. In addition, the components of the $\Delta x_{\mathcal{B}_k}$ (Δy) corresponding to the linearly dependent columns (rows) are set to zero. Then, a point $(x^*; y^*; s^*)$ generated as

$$\begin{aligned} x^*_{\mathcal{B}_k} &= x^k_{\mathcal{B}_k} + \Delta x_{\mathcal{B}_k}, & y^* &= y^k + \Delta y, \\ x^*_{\mathcal{N}_k} &= 0, & s^* &= c - A^T y^*, \end{aligned} \tag{2.2.6}$$

is considered. If $(x^*; y^*; s^*)$ satisfies

$$\begin{aligned} A_{\cdot\mathcal{B}_k} x^*_{\mathcal{B}_k} &= b, & s^*_{\mathcal{B}_k} &= c_{\mathcal{B}_k} - A^T_{\cdot\mathcal{B}_k} y^* = 0, \\ x^*_{\mathcal{B}_k} &> 0, & s^*_{\mathcal{N}_k} &= c_{\mathcal{N}_k} - A^T_{\cdot\mathcal{N}_k} y^* > 0, \end{aligned} \tag{2.2.7}$$

then, it is a strictly complementary solution. Hence $(\mathcal{B}_k, \mathcal{N}_k)$ is the optimal partition. Otherwise, we continue running the algorithm until conditions (2.2.7) are satisfied. The following theorem ensures that Mehrotra-Ye's approach is well defined.

Theorem 2.2.4. *Suppose that at iteration k we have*

$$(x^k)^T s^k < \frac{\beta\gamma^2}{n} 2^{-3L}.$$

Then $(x^*; y^*; s^*)$ defined as (2.2.6) using Mehrotra-Ye's approach satisfies (2.2.7). In particular, the statement is true if $(x^k)^T s^k < \beta 2^{-5L}/n$.

Proof. Since $\beta\gamma^2 2^{-3L}/n < \beta\gamma^2$, from Corollary 2.2.3, the partition $(\mathcal{B}_k, \mathcal{N}_k)$ is readily optimal, and we have

$$A_{\cdot\mathcal{B}_k} x^*_{\mathcal{B}_k} = b, \qquad s^*_{\mathcal{B}_k} = c_{\mathcal{B}_k} - A^T_{\cdot\mathcal{B}_k} y^* = 0.$$

Let $A_{\mathcal{J}\mathcal{K}}$ be a largest invertible submatrix of $A_{\cdot\mathcal{B}_k}$. In the Mehrotra-Ye's approach, we first solve the system

$$A_{\mathcal{J}\mathcal{K}} \Delta x_{\mathcal{K}} = b_{\mathcal{J}} - A_{\mathcal{J}\mathcal{K}} x^k_{\mathcal{K}} = A_{\cdot\mathcal{N}_k} x^k_{\mathcal{N}_k},$$

for $\Delta x_{\mathcal{K}}$. Then, we extend this solution to $\Delta x_{\mathcal{B}_k}$ by filling the rest of the components by zeros. An upper bound for this solution may be calculated as

$$\|\Delta x_{\mathcal{B}_k}\| = \|\Delta x_{\mathcal{K}}\| \leq \left\|A^{-1}_{\mathcal{J}\mathcal{K}}\right\| \|A_{\cdot\mathcal{N}_k}\| \left\|x^k_{\mathcal{N}_k}\right\|,$$

where $\|\cdot\|$ denotes the maximum of the absolute value of the components of a vector or a matrix. Using (1.3.1), the definition of L, we get $\|A\| \leq 2^L$, and Theorem 2.2.2 gives $\left\|x^k_{\mathcal{N}_k}\right\| \leq (x^k)^T s^k/\gamma$. It is shown [39] that $\left\|A^{-1}_{\mathcal{J}\mathcal{K}}\right\| \leq 2^{2L}$. Therefore, we obtain

$$\|\Delta x_{\mathcal{B}_k}\| \leq \frac{(x^k)^T s^k}{\gamma} 2^{3L}.$$

Since $x_{\mathcal{B}_k} \geq \frac{\beta\gamma}{n} e$ from Theorem 2.2.2 and $\beta\gamma/n > (x^k)^T s^k 2^{3L}/\gamma$, we have

$$x^*_{\mathcal{B}_k} = x^k_{\mathcal{B}_k} + \Delta x_{\mathcal{B}_k} > 0.$$

Similarly, we can prove that $s^*_{\mathcal{N}_k} > 0$. □

2.3 Redundancy and the Central Path

In this section, we first show that the concepts of the analytic center and the central path are analytic not geometric. By this, we mean that they depend on the representation of an LP. We then provide equivalent definitions for the analytic center and the central path. The effects of shifting, adding, and deleting constraints on the central path are studied in [11]. Here, we investigate the effect of "redundancy" on the central path.

2.3.1 The Central Path is Representation-Dependent

We provide some examples to illustrate that the analytic center and the central path of a problem depend on the representations of the problem. We start with the definition of redundancy.

Definition 2.3.1. A *redundant constraint* in an LP is a constraint that can be removed without changing the geometry of the feasible region of the LP. An LP is called *redundant* if it contains some redundant constraints.

The following example shows how the analytic center of the n-simplex[3] varies as its representation changes.

Example 2.3.2. The analytic center χ^n of $\mathcal{S}^n$ is the unique maximizer of the barrier function $\log(1-e^Tx)+e^T\log x$, i.e.,

$$\chi^n = \left(\frac{1}{n+1};\cdots;\frac{1}{n+1}\right).$$

Now, let the constraint $x_1 \geq 0$ be repeated h_1 times. The geometry of the set $\mathcal{S}^n$ remains unchanged while its representation is altered. The new analytic center $\chi^n(h_1)$ is the unique maximizer of $\log(1-e^Tx)+p^T\log x$, where $p = (h_1;1;\cdots;1)$. Therefore, we have

$$\chi^n(h_1) = \left(\frac{h_1}{h_1+n};\frac{1}{h_1+n};\cdots;\frac{1}{h_1+n}\right),$$

hence, $\chi^n(h_1) \to (1;0;\cdots;0)$ as $h_1 \to \infty$. □

In the following example, we show how the central path is changing as we use different representations of the feasible region.

Example 2.3.3. Consider the problem of minimizing the function e^Tx over $\mathcal{S}^n$. This problem in standard primal form is given as

$$\begin{aligned} \min \quad & e^Tx \\ \text{s.t.} \quad & e^Tx + x_{n+1} = 1, \\ & x, \quad x_{n+1} \geq 0. \end{aligned} \tag{2.3.1}$$

[3]See Example 2.1.5 for the definition of the n-simplex.

The optimality conditions given by (1.3.5) for the corresponding barrier problem can be written as

$$\begin{aligned} e^T x + x_{n+1} &= 1, \\ y + s_k &= 1 \qquad \text{for } k = 1, \ldots, n, \\ y + s_{n+1} &= 0, \\ x_k s_k &= \mu \qquad \text{for } k = 1, \ldots, n+1, \\ x_k, s_k &> 0 \qquad \text{for } k = 1, \ldots, n+1. \end{aligned}$$

It follows that the central path is a line segment given as the set

$$\Big\{ x \in \mathbb{R}^n \mid x_1 = x_2 = \cdots = x_n = \frac{1 - x_{n+1}}{n} = w(\mu), \mu > 0 \Big\},$$

where

$$w(\mu) = \frac{1 + \mu + n\mu - \sqrt{(1 + \mu + n\mu)^2 - 4n\mu}}{2n}$$

maps $(0, +\infty)$ onto $\big(0, \frac{1}{n+1}\big)$.

Now, let the constraint $x_1 \geq 0$ be repeated h_1 times. The optimality conditions (1.3.5) for the corresponding barrier function of the redundant problem become

$$\begin{aligned} e^T x + x_{n+1} &= 1, \\ \frac{1}{h_1} y + s_k &= \frac{1}{h_1} \qquad \text{for } k = 1, \ldots, h_1, \\ y + s_k &= 1 \qquad \text{for } k = h_1 + 1, \ldots, h_1 + n - 1, \\ y + s_{h_1+n} &= 0, \\ x_1 s_k &= \mu \qquad \text{for } k = 1, \ldots, h_1, \\ x_k s_{h_1+k-1} &= \mu \qquad \text{for } k = 2, \ldots, n+1, \\ x_k &> 0 \qquad \text{for } k = 1, \ldots, n+1, \\ s_k &> 0 \qquad \text{for } k = 1, \ldots, h_1 + n. \end{aligned}$$

Therefore, the central path of the redundant problem is a line segment given as the set

$$\Big\{ x \in \mathbb{R}^n \mid \frac{x_1}{h_1} = x_2 = \cdots = x_n = \frac{1 - x_{n+1}}{h_1 + n - 1} = w(\mu), \mu > 0 \Big\},$$

where

$$w(\mu) = \frac{1 + (h_1 + n)\mu - \sqrt{\big(1 + (h_1 + n)\mu\big)^2 - 4\mu(h_1 + n - 1)}}{2(h_1 + n - 1)}$$

maps $(0, +\infty)$ onto $\left(0, \frac{1}{h_1+n}\right)$. □

As Examples 2.3.2 and 2.3.3 show, an LP can have different analytic centers and central paths, depending on its representations.

2.3.2 Equivalent Definitions for the Central Path

In § 1.3.3, we defined the central path $\mathcal{C}$ as the set of the solutions of system (1.1.3) for all $\mu > 0$. In this subsection, we provide some equivalent definitions for the central path using the concepts of level sets and redundancy. We start with the definition of a level set.

Definition 2.3.4. Let $\alpha \in \mathbb{R}$ and $\varphi : \mathcal{E} \to \mathbb{R}$ be a function where $\mathcal{E} \subseteq \mathbb{R}^n$. Then, the set $\mathcal{L}_\varphi(\alpha) = \{x \in \mathcal{E} \mid \varphi(x) \le \alpha\}$ is called the *level set* of φ corresponding to α.

Consider the primal problem (1.1.1) whose central path is defined (see page 12) as the set $\mathcal{C} = \{x(\mu) \mid \mu > 0\}$. Suppose that Assumption 1.3.1 (IPC) holds. Let $x^* \in \mathcal{P}_*$. For a given $d_0 > c^T x^*$, we define

$$\mathcal{L}(d_0, h_0) = \left\{x \in \mathcal{P}_\oplus \mid c^T x \le d_0 \;\text{ repeated } h_0 \text{ times}\right\}. \tag{2.3.2}$$

The analytic center of $\mathcal{L}(d_0, h_0)$ is the unique maximizer of the problem

$$\begin{array}{ll} \max & h_0 \log(d_0 - c^T x) + e^T \log x \\ \text{s.t.} & Ax = b, \\ & c^T x < d_0, \\ & x > 0, \end{array} \tag{2.3.3}$$

where h_0 can be interpreted as a weight for the constraint $c^T x \le d_0$. We show that the analytic center of $\mathcal{L}(d_0, h_0)$ moves on the central path toward the analytic center of the optimal set whenever $h_0 \to \infty$ or $d_0 \to c^T x^*$. Figure 2.1 illustrates this observation for the following problem[4] in 2D:

$$\begin{array}{ll} \min & y_2 \\ \text{s.t.} & 0 \le y_1 \le 1, \\ & \frac{1}{4} y_1 \le y_2 \le 1 - \frac{1}{4} y_1, \end{array} \tag{2.3.4}$$

[4] This is the Klee-Minty problem [23] in 2D.

with the level sets $y_2 \leq d_0$ repeated h_0 times.

For a fixed d_0 chosen such that $d_0 > c^T x$ for all $x \in \mathcal{P}_\oplus$, the set of the minimizers of problem (2.3.3) for all $h_0 \in \mathbb{R}_+$ forms the central path of the primal problem (1.1.1) as shown in the following theorem.

Theorem 2.3.5. *Suppose that Assumption 1.3.1* (IPC) *holds. Let* $x^* \in \mathcal{P}_*$ *and* $d_0 \in \mathbb{R}$ *such that* $d_0 > c^T x$ *for all* $x \in \mathcal{P}_\oplus$. *Define a function from* $\mathbb{R}_+$ *to* $\mathcal{P}_+$ *as*

$$\chi^d(h_0) = \arg\max \left\{ h_0 \log(d_0 - c^T x) + e^T \log x \mid x \in \mathcal{P}_+, c^T x < d_0 \right\}.$$

Then, we have $\mathcal{C} = \left\{ \chi^d(h_0) \mid h_0 > 0 \right\}$. *Moreover, we have* $\chi^d(h_0) \to \chi^{\mathcal{P}_*}$ *as* $h_0 \to \infty$.

Proof. Let $x \in \left\{ \chi^d(h_0) \mid h_0 > 0 \right\}$. From the optimality conditions, there exists a y such that

$$\begin{aligned} Ax &= b, \qquad x > 0, \\ \frac{h_0}{d_0 - c^T x} c - \frac{1}{x} - A^T y &= 0, \\ c^T x &< d_0. \end{aligned}$$

Hence x is the μ-center with $\mu = (d_0 - c^T x)/h_0 > 0$ implying that $x \in \mathcal{C}$. It also implies that $\mu \to 0$ as $h_0 \to \infty$ since $0 < d_0 - c^T x < d_0 - c^T x^*$ for all $x \in \mathcal{P}_\oplus$. Therefore, by Theorem 2.1.4, we have $\chi^d(h_0) \to \chi^{\mathcal{P}_*}$ as $h_0 \to \infty$.

Now let $x \in \mathcal{C}$. There exist a $\mu > 0$ and a vector $(y; s)$ such that

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

It follows that x is the analytic center of the level set $\mathcal{L}(d_0, h_0)$ with $h_0 = (d_0 - c^T x)/\mu > 0$. □

Now let us fix $h_0 \in \mathbb{R}_+$. In the following theorem, we show that the set of the minimizers of problem (2.3.3) for all $d_0 \in (c^T x^*, \infty)$ is identical to the central path of the primal problem (1.1.1) (see Figure 2.1).

Theorem 2.3.6. *Suppose that Assumption 1.3.1* (IPC) *holds. Let* $x^* \in \mathcal{P}_*$ *and* $h_0 \in \mathbb{R}_+$. *Define a function from* $(c^T x^*, \infty)$ *to* $\mathcal{P}_+$ *as*

$$\chi^h(d_0) = \arg\max \left\{ h_0 \log(d_0 - c^T x) + e^T \log x \mid x \in \mathcal{P}_+, c^T x < d_0 \right\}.$$

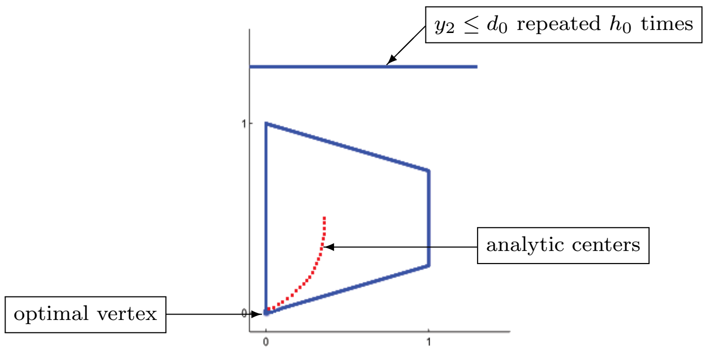


Figure 2.1: The analytic center moves on the central path toward the optimal solution as (1) $d_0 \to 0$ with a fixed h_0 or (2) $h_0 \to \infty$ with a fixed d_0.

Then, we have $\mathcal{C} = \left\{\chi^h(d_0) \mid d_0 > c^T x^*\right\}$. *Moreover, we have* $\chi^h(d_0) \to \chi^{\mathcal{P}_*}$ *as* $d_0 \to c^T x^*$.

Proof. We skip the proof as it is similar to the one of Theorem 2.3.5. □

Example 2.3.7. Consider problem (2.1.1) described in Example 2.3.3. According to Theorem 2.3.5 with $d_0 = 1$, the central path of this problem is the set of the maximizers of the problem

$$\begin{aligned} \max \quad & h_0 \log(1 - e^T x) + e^T \log x \\ \text{s.t.} \quad & e^T x < 1, \\ & x > 0, \end{aligned}$$

for all $h_0 \in [1, \infty)$. Thus, the central path is a line segment (the solid line in Figure 2.2) represented by the set

$$\left\{x \mid x_1 = \cdots = x_n = \frac{1}{h_0 + n}, h_0 \in [1, \infty)\right\}.$$

Now, let the constraint $x_1 \geq 0$ be repeated h_1 times. According to Theorem 2.3.6, the central path is identical to the set of the maximizers of the

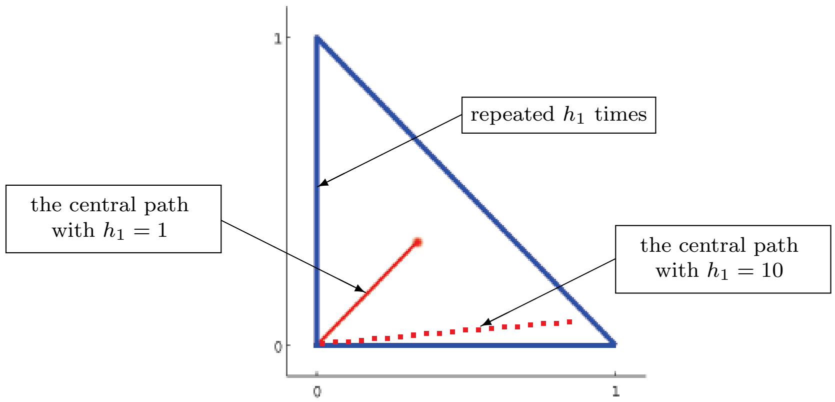


Figure 2.2: The central paths of the problem of minimizing $x_1 + x_2$ over $\mathcal{S}^2$ with and without redundancy.

problem

$$\begin{array}{ll} \max & h_0 \log(1 - e^T x) + p^T \log x \\ \text{s.t.} & e^T x < 1, \\ & x > 0, \end{array}$$

for all $h_0 \in [1, \infty)$ where $p = (h_1; 1; \cdots; 1)$. In other words, the central path is a line segment (the dotted line in Figure 2.2) represented by the set

$$\left\{ x \mid \frac{x_1}{h_1} = x_2 = \cdots = x_n = \frac{1}{h_0 + h_1 + n}, h_0 \in [1, \infty) \right\}.$$

Note that how the central path changes as we repeat the constraint $x_1 \geq 0$ for a fixed h_1 number of times. □

2.3.3 The Central Path of Redundant Problems

In the previous subsection, we studied the effect of repeating and shifting a constraint parallel to the objective function. Based on this study, we gave equivalent definitions for the central path. In this subsection, we study the effect of repeating an arbitrary redundant constraint on the central path.

Assume that $\mathcal{P}_\oplus$ is bounded and consider the primal problem (1.1.1) with a redundant constraint $u^T x \le d_0$ repeated h_0 times added to the constraints. In other words, we consider the problem

$$\begin{array}{lll} \min & c^T x & \\ \text{s.t.} & Ax = b, & \\ & u^T x \le d_0 & \text{repeated } h_0 \text{ times}, \\ & x \ge 0, & \end{array} \tag{2.3.5}$$

whose feasible region is geometrically identical to that of problem (1.1.1). The following theorem shows how the analytic center $\chi^u(h_0)$ of the feasible region represented by the constraints of problem (2.3.5) moves as $h_0 \to \infty$.

Theorem 2.3.8. *Assume that $\mathcal{P}_\oplus$ is bounded. As $h_0 \to \infty$, the analytic center $\chi^u(h_0)$ of the feasible region represented by the constraints of problem* (2.3.5) *converges to the analytic center of the optimal set of the problem*

$$\begin{array}{ll} \min & u^T x \\ \text{s.t.} & Ax = b, \\ & x \ge 0. \end{array} \tag{2.3.6}$$

Proof. Analogous to the proof of Theorem 2.3.5, we can show that for every h_0, the analytic center $\chi^u(h_0)$ is the μ-center of problem (2.3.6) with

$$\mu = \frac{d_0 - u^T \chi^u(h_0)}{h_0}.$$

Since $\mathcal{P}_\oplus$ is bounded, we have $\mu \to 0$ as $h_0 \to \infty$. Therefore, from Theorem 2.1.4, the analytic center $\chi^u(h_0)$ converges to the analytic center of the optimal set of problem (2.3.6). □

The following theorem shows how the central path of problem (2.3.5) changes as $h_0 \to \infty$.

Theorem 2.3.9. *Assume that $\mathcal{P}_\oplus$ is bounded. As $h_0 \to \infty$, every point on the central path of the redundant problem* (2.3.5) *converges to an optimal point of problem* (2.3.6).

Proof. Let $\mu > 0$ be given. For every $h_0 \ge 0$, let us denote the μ-center of problem (2.3.5) by $\chi^{u,\mu}(h_0)$. We show that the μ-center $\chi^{u,\mu}(h_0)$ converges to

Table 2.1

	α	h_1	h_2	d_1	d_2
Figure 2.3a	–	0	0	–	–
Figure 2.3b	–	20	0	0	–
Figure 2.3c	1	0	100	–	0
Figure 2.3d	0	0	100	–	0
Figure 2.3e	1	1300	300	8	0
Figure 2.3f	0	60	200	2	0

an optimal point of problem (2.3.6) as $h_0 \to \infty$. By definition, the μ-center $\chi^{u,\mu}(h_0)$ is the unique minimizer of the problem

$$\begin{array}{ll} \min & c^T x - \mu\big(e^T \log x + h_0 \log(d_0 - u^T x)\big) \\ \text{s.t.} & Ax = b, \\ & u^T x < d_0, \\ & x > 0, \end{array}$$

whose necessary and sufficient optimality conditions are

$$\begin{aligned} Ax &= b, & x &> 0 \\ A^T y + s &= \frac{d_0 - u^T x}{\mu h_0} c + u, & s &> 0 \\ xs &= \frac{d_0 - u^T x}{h_0} e, \\ u^T x &< d_0. \end{aligned}$$

Since $d_0 - u^T x$ is bounded, the solution $\chi^{u,\mu}(h_0)$ of this system converges to an optimal point of problem (2.3.6) as $h_0 \to \infty$. □

Figure 2.3 illustrates the effect of redundancy on the central path of the Klee-Minty (KM) problem (2.3.4). Different types and numbers of redundant constraints are added to obtain an interesting collection of the central paths. In exact words, we consider the problem

$$\begin{array}{lll} \min & y_2 & \\ \text{s.t.} & 0 \le y_1 \le 1, & \\ & \frac{1}{4} y_1 \le y_2 \le 1 - \frac{1}{4} y_1, & \\ & 0 \le y_1 + d_1 & \text{repeated } h_1 \text{ times}, \\ & \frac{1}{4}\alpha y_1 \le y_2 + d_2 & \text{repeated } h_2 \text{ times}, \end{array} \quad (2.3.7)$$

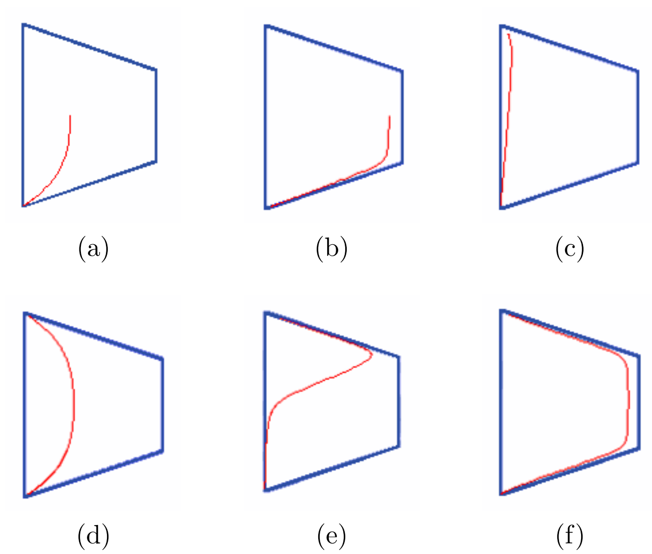


Figure 2.3: Effect of redundancy on the central path.

in which the parameter α is either zero or one. If $\alpha = 0$ then the corresponding redundant constraint is parallel to the coordinate hyperplane $y_2 = 0$; if $\alpha = 1$ then it is parallel to a facet of the KM cube. Note that zero repetition means no constraint. Figure 2.3 shows the central path of problem (2.3.7) with the data chosen according to Table 2.1. Note that similar shapes of the central paths may also be obtained with different choice of data. For instance, the central path of problem (2.3.7) with $h_1 = 1300$, $h_2 = 1000$, $d_1 = 8$, $d_2 = 2$, and $\alpha = 1$ or with $h_1 = 20$, $h_2 = 1000$, $d_1 = d_2 = 0$, and $\alpha = 0$ is similar to the one in Figure 2.3f.

Chapter 3

Pathological Examples in Optimization

In this chapter, we review some pathological examples in optimization. We first review some exponential examples for pivot algorithms used for solving LPs and linear complementarity problems (LCPs). In particular, we will see that variants of the Klee-Minty (KM) construction are used to show that some simplex-type methods may take an exponential number of steps to converge to the optimal point. We also review some unfavorable examples for IPMs, including Megiddo and Shub's [27] result that shows that a trajectory of the affine-scaling vector field may visit an arbitrarily small neighborhood of every vertex of the KM cube. Finally, we present some linear and nonlinear convex optimization problems with ill-behaved central paths.

3.1 Exponential Examples for Pivot Algorithms

In this section, we study some exponential examples for pivot algorithms used for solving LPs and LCPs.

3.1.1 Linear Optimization Problems

One of the interesting features of the simplex method is its flexibility in choosing the pivot element during the procedure. To learn more about the pivot

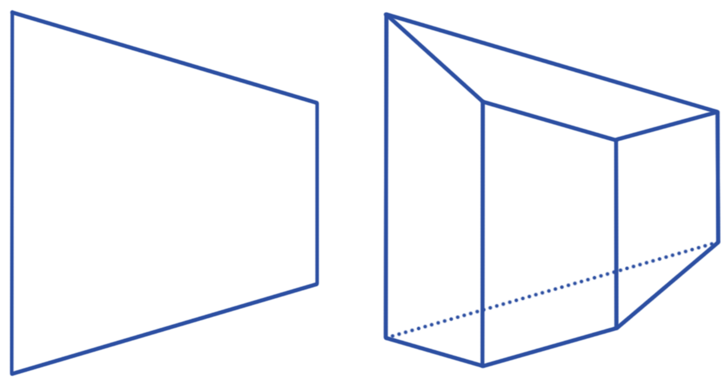


Figure 3.1: KM cubes in 2D and 3D.

rules, the reader is referred to a survey by Terlaky and Zhang [44]. The efficiency of pivot rules is analyzed from two different points of view: finiteness and iteration complexity.

Dantzig's pivot rule [10], known also as the most-negative-cost pivot rule, is not finite, i.e., a cycle might happen during the procedure. Bland [8] has introduced the minimal-index pivot rule to avoid cycling. Therefore, the simplex method combined with minimal-index pivot rule is a finite procedure. Zionts [54] has presented the criss-cross simplex method for solving LPs, and finiteness of his approach has yet been an open question. However, Terlaky [42] has proposed a finite criss-cross algorithm.

The iteration complexity of most pivot rules used in the simplex method has been proven to be exponential in the worst case. Klee and Minty [23] gave an example for which the simplex method with Dantzig's pivot rule takes an exponential number of steps to reach the optimal point. More precisely, they introduced the problem

$$\begin{array}{lll} \min & y_m & \\ \text{s. t.} & 0 \le y_1 \le 1, & \\ & \tau\, y_{k-1} \le y_k \le 1 - \tau\, y_{k-1} & \text{for } k = 2, \ldots, m, \end{array} \tag{3.1.1}$$

whose feasible region is an m-dimensional unit cube tilted by a positive factor $0 < \tau < 1/2$. This tilted cube (see Figure 3.1) is referred to as the *KM cube* and is denoted by $\mathcal{KM}(\tau)$. Then, they showed that the simplex method with Dantzig's pivot rule, initiated at the feasible basis solution $(0; 0; \cdots; 1)$, visits all the 2^m vertices of the KM cube. Such a pivot rule is called exponential, and most pivot rules have been proven to be exponential too; see [44] and the references therein. Therefore, the KM problem proves that the simplex method belongs to the class of non-polynomial (NP) algorithms.

Avis and Chvátal [4] used a variant[1] of the KM problem to construct a worst-case example for Bland's pivot rule. They introduced the problem

$$\begin{aligned} \max \quad & \sum_{j=1}^{m} \tau^{m-j} x_j \\ \text{s.t.} \quad & 2\sum_{j=1}^{i-1} \tau^{i-j} x_j + x_i + x_{m+i} = 1 \quad && \text{for } i = 1, \ldots, m, \\ & x_j \geq 0 \quad && \text{for } j = 1, \ldots, 2m, \end{aligned}$$

where $0 < \tau < 1/2$ and showed that Bland's minimal-index pivot rule takes

$$\alpha_m = \frac{2}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{m+1} - \frac{2}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{m+1} - 1$$

steps[2] to find the optimal solution.

Murty [32] presented a parametric LP given as

$$\begin{aligned} \max \quad & \sum_{j=1}^{m} 4^{m-j} x_j \\ \text{s.t.} \quad & 2\sum_{j=1}^{i-1} x_j + x_i - x_{m+i} = \lambda - 2^{m-i+1} \quad && \text{for } i = 1, \ldots, m, \\ & x_j \geq 0 \quad && \text{for } j = 1, \ldots, 2m, \end{aligned} \tag{3.1.2}$$

with the parameter λ and showed that the computational effort required by the simplex method for solving this problem is not bounded above by a polynomial in the size of the problem. Roos [37] showed that Terlaky's finite criss-cross algorithm takes an exponential number of pivot steps to converge to the optimal solution of the KM example. A bad network problem was designed by Zadeh [52] for which the simplex method and some other minimum-cost flow algorithms take an exponential number of steps to reach the optimal solution.

[1] The transformation $y_k = \sum_{i=1}^{k} \tau^{k-j} x_j$ for $k = 1, \ldots, m$ will revert the Avis and Chvátal's problem to the KM problem given by (3.1.1).

[2] For every positive integer m, α_m is the m^{th} Fibonacci number which is integer.

3.1.2 Linear Complementarity Problems

Lemke's complementary pivot algorithm is a popular method to solve LCPs; see for example [33]. Murty [31, 33] constructed a class of examples with integer data and proved that Lemke's algorithm requires an exponential number of pivot steps, with respect to the dimension of the problem, to find the optimal solution. More precisely, he defined a matrix $M \in \mathbb{R}^{n\times n}$ as

$$M = \begin{pmatrix} 1 & & & & \\ 2 & 1 & & & \\ \vdots & \vdots & \ddots & & \\ 2 & 2 & \cdots & 1 & \\ 2 & 2 & \cdots & 2 & 1 \end{pmatrix}$$

and considered the following LCP[3]: find $\bar{x}$ and $\widehat{x}$ in $\mathbb{R}^n$ such that they satisfy

$$\begin{aligned} \bar{x} &= M\widehat{x} + q, & \widehat{x} &\geq 0, \\ \bar{x}\widehat{x} &= 0, & \bar{x} &\geq 0, \end{aligned}$$

where $q \in \mathbb{R}^n$ is given as $q = (2^n - 2^{n+1}; 2^{n-1} - 2^{n+1}; \cdots ; 2 - 2^{n+1})$. Then, he showed that Lemke's complementary pivot algorithm requires 2^n pivot steps before termination.

3.2 Klee-Minty Cubes and Affine-Scaling IPMs

The KM example has been used to design ill-behaved examples for IPMs as well. As shown in [27], for a given $\delta > 0$, a trajectory of the affine-scaling vector field may visit the δ-neighborhood of each vertex of the KM cube providing that the starting point is properly chosen in the δ-neighborhood of the vertex $v^0 = (0; 0; \cdots ; 1)$. We describe this observation in the sequel.

We have briefly explained the basic structure of affine-scaling IPMs in § 1.3.1. The analysis in this section is based on the dual affine-scaling method. Let us begin with a review of the notions of a vector field and the trajectories induced

[3]The equation $\bar{x} = M\widehat{x} + q$ in the LCP is obtained from the equality constraints of problem (3.1.2) by setting $\widehat{x} = x_{1:m}$, $\bar{x} = x_{m+1:2m}$, and q being the right-hand side of the equation with $\lambda = 2^{n+1}$.

by it. We first define the sets

$$\mathcal{D} = \left\{(y;s) \mid A^T y + s = c\right\}, \quad \mathcal{D}_0 = \left\{(y;s) \mid A^T y + s = 0\right\}.$$

The set $\mathcal{D}$ is the dual affine subspace and $\mathcal{D}_0$ is its translation to the origin.

Definition 3.2.1. [1] Let O be a relatively open subset in $\mathcal{D}$. Then, a mapping $\Psi : O \to \mathcal{D}_0$ is called a *vector field* in O.

Definition 3.2.2. [1] A *trajectory of a vector field* Ψ passing through $y^0 \in O$ at an initial point $t^0 \in \mathbb{R}$ is defined to be the solution curve of the differential equation

$$\begin{cases} \dot{y}(t) = \Psi\left(y(t)\right), \\ y(t^0) = y^0. \end{cases}$$

We are interested in the affine-scaling vector field defined by the mapping

$$\begin{cases} \Psi : \mathcal{D}_+ \to \mathcal{D}_0, \\ \Psi = (\widehat{\Psi}; \bar{\Psi}), \end{cases}$$

where $\widehat{\Psi} = -\left(AS^{-2}A^T\right)^{-1} b$ and $\bar{\Psi} = A^T\left(AS^{-2}A^T\right)^{-1} b$.

In the following, we study the vector field induced by the affine-scaling directions on the KM problem (3.1.1). The KM problem (3.1.1) is in the standard dual form (1.1.2) with the input-data given by

$$A = \left(\widehat{A}, \bar{A}\right), \quad b = (0; \cdots; 0; 1), \quad c = \begin{pmatrix} \widehat{c} \\ \bar{c} \end{pmatrix}, \tag{3.2.1}$$

where the $m \times m$ matrices $\widehat{A}$ and $\bar{A}$ and the m-vectors $\widehat{c}$ and $\bar{c}$ are defined as

$$\widehat{A} = \begin{pmatrix} -1 & \tau & & \\ & \ddots & \ddots & \\ & & -1 & \tau \\ & & & -1 \end{pmatrix}, \bar{A} = \begin{pmatrix} 1 & \tau & & \\ & \ddots & \ddots & \\ & & 1 & \tau \\ & & & 1 \end{pmatrix}, \widehat{c} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \bar{c} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Therefore, we have

$$AS^{-2}A^T = \begin{pmatrix} p_1 & q_1 & & & \\ q_1 & p_2 & q_2 & & \\ & \ddots & \ddots & \ddots & \\ & & q_{m-2} & p_{m-1} & q_{m-1} \\ & & & q_{m-1} & p_m \end{pmatrix}$$

where

$$p_k = \frac{1}{\widehat{s}_k^2} + \frac{1}{\bar{s}_k^2} + \frac{\tau^2}{\widehat{s}_{k+1}^2} + \frac{\tau^2}{\bar{s}_{k+1}^2} \qquad \text{for } k = 1, \ldots, m-1,$$

$$p_m = \frac{1}{\widehat{s}_m^2} + \frac{1}{\bar{s}_m^2},$$

$$q_k = \frac{-\tau}{\widehat{s}_{k+1}^2} + \frac{\tau}{\bar{s}_{k+1}^2} \qquad \text{for } k = 1, \ldots, m-1.$$

It can be verified that the affine-scaling vector field for the KM problem is given by $\Psi = (\widehat{\Psi}; \bar{\Psi})$ where

$$\widehat{\Psi} = -\left(AS^{-2}A^T\right)^{-1} b = \left((-1)^{m-k+1} \frac{\omega_{k-1}}{\omega_m} \prod_{i=k}^{m-1} q_i \right)_{1:m}$$

and $\bar{\Psi} = -A^T\widehat{\Psi}$. Note that ω_k denotes the determinant of the k^{th} principal minor of the positive-definite Hessian matrix $AS^{-2}A^T$. Obviously $\omega_k > 0$ for $k = 1, \ldots, m$, and the following recursive formula holds:

$$\begin{cases} \omega_0 = 1, \quad \omega_1 = p_1, \\ \omega_k = p_k\omega_{k-1} - q_k^2\omega_{k-2} \quad \text{for } k = 2, \ldots, m. \end{cases}$$

The unique trajectory of the vector field Ψ corresponding to a given initial point $(y^0; s^0) \in \mathcal{D}_+$ is determined by the solution curve of the differential equation

$$\begin{cases} \big(\dot{y}(t); \dot{s}(t)\big) = \Psi\big(y(t); s(t)\big), \\ \big(y(t^0); s(t^0)\big) = (y^0; s^0), \end{cases} \tag{3.2.2}$$

and, according to [1, 27], converges to the unique optimal point, which is the origin. The following theorem shows how bad these trajectories can be.

Theorem 3.2.3. **[27]** *For a given $\delta > 0$, there exists a point $(y^0; s^0) \in \mathcal{D}_+$ such that the solution of system* (3.2.2) *visits the δ-neighborhood of each vertex of the KM cube.*

Interestingly, if $(y^0; s^0)$ is on the central path of the KM problem, then the affine-scaling trajectory defined by the solution of system (3.2.2) coincides the central path [1].

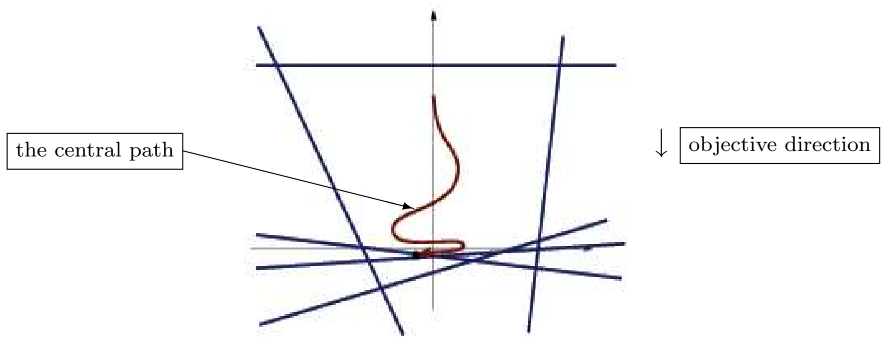


Figure 3.2: A zigzagging central path of an LP.

3.3 Examples of Ill-Behaved Central Paths

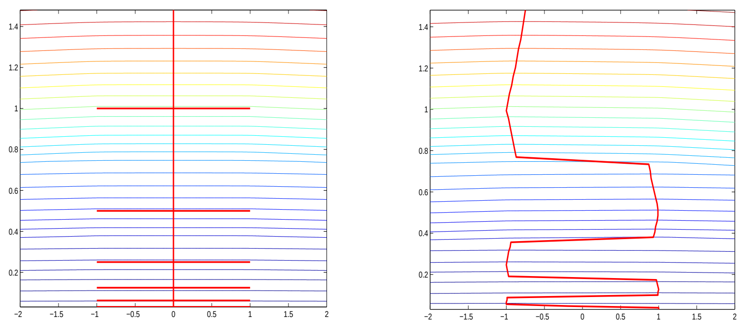
In this section, we describe some examples with ill-behaved central paths. One of the examples is an LP while others are nonlinear convex optimization problems. First, let us consider [14] the problem

$$\begin{aligned} \max \quad & -y_2 \\ \text{s.t.} \quad & 2y_2 \leq 1, \\ & -3y_1 - y_2 \leq 1, \\ & 10y_1 - y_2 \leq 5, \\ & (-1)^{k+1} 11 y_1 - 10^{k+1} y_2 \leq 5 \quad \text{for } k = 1, \ldots, m-3. \end{aligned}$$

This problem has a zigzagging central path, and the number of almost 180° turns in the zigzagging pattern is $m-3$, where m is the number of inequalities. The length of almost linear segments in the zigzagging central path vanishes as we approach to the optimal solution. Figure 3.2 shows the central path of this LP with a zigzagging shape.

In [17], some interesting nonlinear convex optimization problems with ill-behaved central paths are introduced. The authors show that the central path may have "antenna-like" or zigzagging shapes with almost-straight line segments with fixed lengths repeated infinitely many times; see Figures 8.1a and 8.1b. In exact words, they first define a convex function $\varphi^0 : \mathbb{R}^2 \to \mathbb{R}$ as

$$\varphi^0(x) = \frac{1}{2} \|x - \bar{x}\|_2 + \frac{1}{2} \|x - \widehat{x}\|_2 - 1,$$



(a) Antenna-like central path (b) Zigzagging central path

Figure 3.3: Ill-behaved central paths of nonlinear convex problems.

where $\bar{x} = (-1; 1)$ and $\widehat{x} = (1; 1)$. This function is differentiable everywhere but $\bar{x}$ and $\widehat{x}$, and its set of minimizers is the line segment

$$\left\{x \in \mathbb{R}^2 \mid -1 \leq x_1 \leq 1, x_2 = 1\right\}.$$

Next, they define a function ϕ as

$$\phi(x) = \alpha(x_2 - 1) + \beta + \gamma\varphi(x),$$

where α, β, and γ are positive parameters, and the function φ is a smoothing[4] function of φ^0 satisfying $0 \leq \varphi(x) \leq \varphi^0(x)$. Then, they show that the central path of minimizing ϕ over $x_2 \geq 0$ is the cross defined by

$$\left\{x \in \mathbb{R}^2 \mid x_1 = 0, x_2 > 0\right\} \cup \left\{x \in \mathbb{R}^2 \mid -1 \leq x_1 \leq 1, x_2 = 1\right\}.$$

By producing a sequence of similar functions, they constructed an optimization problem whose central path is an antenna-like shape having infinitely many almost-straight line segments with a fixed length as shown in Figure 8.1a.

Their second example is obtained by a slight perturbation of the first one, which was just discussed, whose central path has a zigzagging shape with infinite variation as shown in Figure 8.1b. They conclude that convexity combined with even differentiability of any degree does not suffice to ensure having a reasonably well-behaved central path.

[4] A smoothing function of φ_0 is a differentiable function φ such that $\varphi_0 = \varphi$ everywhere except in a small neighborhood of $\bar{x}$ and $\widehat{x}$.

Chapter 4

A Klee-Minty Type Example for Central-Path-Following IPMs

The Klee-Minty (KM) problem is a well-known worst-case example on which simplex methods take an exponential number of pivot steps to converge to the optimal point. Variants of the simplex method visit all the 2^m vertices of the m-dimensional KM cube before termination. Following the so-called central path, central-path-following IPMs converge to an optimal solution of an LP in polynomial time. In this chapter, we show that by systematically adding an exponential number of redundant inequalities at a proper distance to the feasible region, the central path can be bent [12] along the edges of the KM cube visiting all the 2^m vertices of that cube. Therefore, central-path-following IPMs take $2^m - 2$ "sharp" turns as the central path passes through within an arbitrarily small neighborhood of every vertex of the cube before converging to the optimal solution. In other words, the central path closely follows the *simplex path*, the edge-path followed by the simplex method; see Figure 4.1.

4.1 A Redundant KM Problem

The KM problem (3.1.1) has m variables and $2m$ constraints, and its feasible region is an m-dimensional tilted cube denoted by $\mathcal{KM}(\tau)$ as previously mentioned. Some variants of the simplex method take $2^m - 1$ number of pivot steps to solve problem (3.1.1) as they visit all the 2^m vertices of the cube ordered by the decreasing values of the last coordinate y_m starting from the vertex $v^0 = (0; \cdots; 0; 1)$ until the optimal value $y_m^* = 0$ is reached at the origin.

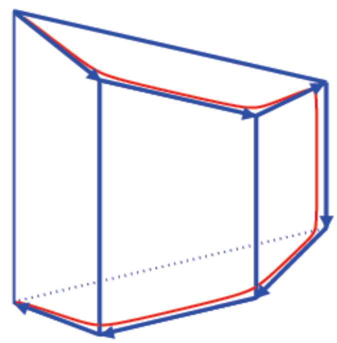


Figure 4.1: The central path is bent along the simplex path.

Central-path-following IPMs start at a neighborhood of the analytic center of $\mathcal{KM}(\tau)$, follow the central path, and converge to the origin. The goal here is to provide an example whose central path is bent along the simplex path, see Figure 4.1, visiting an arbitrarily small neighborhood of every vertex of the KM cube. We show that this can be done by adding an exponential number of inequalities to the problem as described in the following. For $k = 1, \ldots, m$, with the convention $y_0 = 0$, the inequality $\tau y_{k-1} \le y_k + d_0$ is added h_k times, where d_0 and h_k are to be determined. So, the redundant KM problem is defined as

$$\begin{array}{lll} \min & y_m & \\ \text{s.t.} & \tau y_{k-1} \le y_k \le 1 - \tau y_{k-1} & \text{for } k = 1, \ldots, m, \\ & \tau y_{k-1} \le y_k + d_0 & \text{repeated } h_k \text{ times for } k = 1, \ldots, m. \end{array} \tag{4.1.1}$$

The new representation of the feasible region of this problem is denoted by $\mathcal{KM}_1(\tau, d_0, h)$, where $h = (h_1; \cdots ; h_m)$. Note that the sets $\mathcal{KM}(\tau)$ and $\mathcal{KM}_1(\tau, d_0, h)$ are geometrically identical for all $\tau > 0$, $d_0 \ge 0$, and h, while they differ in the algebraic representation. The vector h should be an integer-valued vector since its components represent the number of inequalities parallel to the corresponding facets. Yet, the components can be interpreted as weights for the redundant constraints in case they are not integer valued.

4.1.1 The Central Path of the Redundant KM Problem

We denote the slack variables of problem (3.1.1) as

$$\begin{cases} \widehat{s}_k = \quad y_k - \tau y_{k-1} \\ \bar{s}_k = 1 - y_k - \tau y_{k-1} \end{cases} \quad \text{for } k = 1, \ldots, m, \tag{4.1.2}$$

and let $\widehat{s} = (\widehat{s}_1; \cdots ; \widehat{s}_m)$ and $\bar{s} = (\bar{s}_1; \cdots ; \bar{s}_m)$. The logarithmic barrier function for problem (4.1.1) is defined as

$$F_1(y) = -e^T \log \widehat{s} - e^T \log \bar{s} - h^T \log(d_0 + \widehat{s}).$$

By Definition 1.3.4, the analytic center of $\mathcal{KM}_1(\tau, d_0, h)$ is then given as

$$\chi^1(\tau, d_0, h) = \arg\min \{F_1(y) \mid (\widehat{s}; \bar{s}; d_0 + \widehat{s}) > 0\}.$$

The necessary and sufficient optimality conditions for this convex minimization problem are

$$\begin{cases} \frac{1}{\widehat{s}_k} - \frac{1}{\bar{s}_k} - \frac{\tau}{\widehat{s}_{k+1}} - \frac{\tau}{\bar{s}_{k+1}} + \frac{h_k}{d_0 + \widehat{s}_k} - \frac{h_{k+1}\tau}{d_0 + \widehat{s}_{k+1}} = 0, \quad k = 1, \ldots, m-1, \\ \frac{1}{\widehat{s}_m} - \frac{1}{\bar{s}_m} + \frac{h_m}{d_0 + \widehat{s}_m} = 0, \\ \widehat{s}_k > 0, \bar{s}_k > 0, d_0 + \widehat{s}_k > 0, \quad k = 1, \ldots, m. \end{cases} \tag{4.1.3}$$

Hence, the analytic center $\chi^1(\tau, d_0, h)$ is the unique solution of system (4.1.3). By Definition 1.3.3, the central path of problem (4.1.1) is the set

$$\mathcal{C}_1(\tau, d_0, h) = \{y(\mu) \mid \mu > 0\},$$

in which the function $y(\mu)$ is defined as

$$y(\mu) = \arg\min \{y_m + \mu F_1(y) \mid (\widehat{s}; \bar{s}; d_0 + \widehat{s}) > 0\}.$$

Therefore, from the optimality conditions, every $y(\mu)$ satisfies the first $m-1$ equations of system (4.1.3).

4.1.2 The Simplex Path of the KM Cube

To ease the analysis, we give a mathematical definition for the simplex path of problem (4.1.1). In addition, for a given $\delta > 0$, we define a δ-neighborhood of the simplex path. For this purpose, we begin with defining the sets

$$\begin{aligned} T_\delta^k &= \{y \in \mathcal{KM}_1(\tau, d_0, h) \mid \bar{s}_k \le \delta\}, \\ C_\delta^k &= \{y \in \mathcal{KM}_1(\tau, d_0, h) \mid \bar{s}_k > \delta, \widehat{s}_k > \delta\}, \\ B_\delta^k &= \{y \in \mathcal{KM}_1(\tau, d_0, h) \mid \widehat{s}_k \le \delta\}, \end{aligned}$$

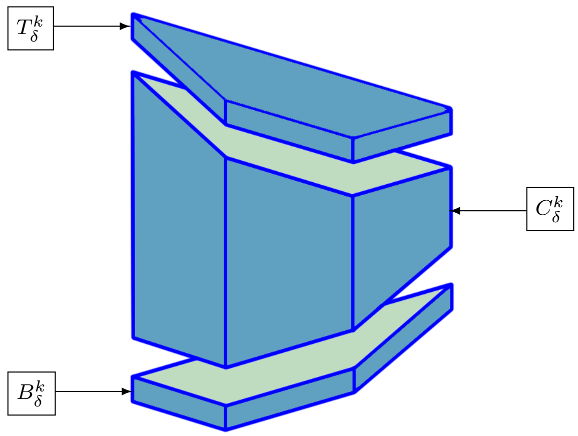


Figure 4.2: The sets T_δ^k, C_δ^k, and B_δ^k.

for $k = 2, \ldots, m$, and

$$\widehat{C}_\delta^k = \{ y \in \mathcal{KM}_1(\tau, d_0, h) \mid \bar{s}_k \le \delta, \widehat{s}_{k-1} \le \delta, \ldots, \widehat{s}_1 \le \delta \ \} \, ,$$

for $k = 1, \ldots, m$; see Figure 4.3. Visually, the sets T_δ^k, C_δ^k, and B_δ^k can be considered as the top, central, and bottom parts of $\mathcal{KM}_1(\tau, d_0, h)$, and obviously

$$\mathcal{KM}_1(\tau, d_0, h) = T_\delta^k \cup C_\delta^k \cup B_\delta^k \quad \text{ for } k = 2, \ldots, m.$$

Interestingly, the set $\widehat{C}_\delta^m$ is a δ-neighborhood of the vertex v^0; see Figure 4.3. Setting

$$A_\delta^k = T_\delta^k \cup \widehat{C}_\delta^{k-1} \cup B_\delta^k \quad \text{ for } k = 2, \ldots, m,$$

we define a *δ-neighborhood of the simplex path* as

$$P_\delta = \bigcap_{k=2}^{m} A_\delta^k.$$

Figure 4.4 illustrates how the intersection of the sets A_δ^2 and A_δ^3 forms the δ-neighborhood P_δ of the simplex path in 3D. The *simplex path* itself is precisely determined by

$$P_0 = \bigcap_{k=2}^{m} A_0^k.$$

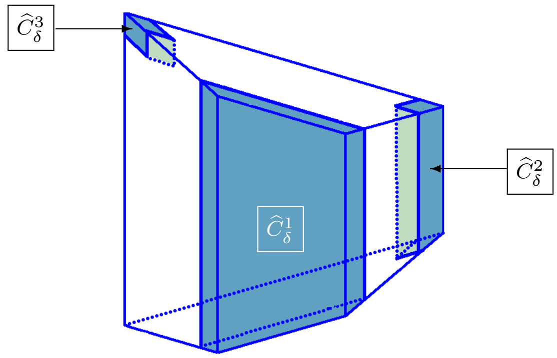


Figure 4.3: The sets $\widehat{C}_\delta^k$, for $k = 1, 2, 3$.

4.2 The Central Path Follows the Simplex Path of the KM Cube

In this section, we assume that $m \geq 2$, $\tau > 0$, and $\delta > 0$ with $\tau + \delta < 1/2$. We seek for a $d_0 \in \mathbb{R}_+$ and a repetition vector $h \in \mathbb{Z}_\oplus^m$. For the sake of simplicity, we define, for $k = 1, \ldots, m-1$,

$$\begin{aligned} \ell_k &= \frac{h_k}{d_0 + 1} - \frac{h_{k+1}\tau}{d_0}, \\ u_k &= \frac{h_k}{d_0} - \frac{h_{k+1}\tau}{d_0 + 1}, \end{aligned}$$

as well as $\ell_m = h_m/(d_0+1)$ and $u_m = h_m/d_0$. In the following lemmas, we show that there exist a $d_0 \in \mathbb{R}_+$ and a vector $h \in \mathbb{Z}_\oplus^m$ satisfying, for $k = 1, \ldots, m$,

$$\ell_k \tau^{k-1} - \sum_{i=1}^{k-1} u_i \tau^{i-1} \geq \frac{2}{\delta}. \tag{4.2.1}$$

Lemma 4.2.1. *System* (4.2.1) *holds for*

$$\begin{cases} d_0 \geq m 2^{m+1}, \\ h_k = \dfrac{d_0 + 1}{\tau^{k-1}\delta} \left(2^{m+2} - 2^{k+1}\right) \qquad \textit{for } k = 1, \ldots, m. \end{cases}$$

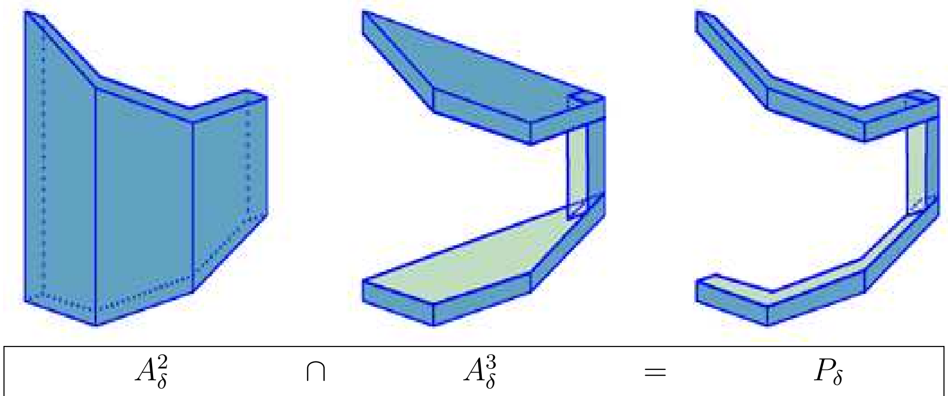


Figure 4.4: The δ-neighborhood P_δ of the simplex path for the KM 3-cube.

Proof. For every $k = 1, \ldots, m$, substituting the value of h_k in the definitions of ℓ_k and u_k, we can see that $\ell_k > 0$, $u_k > 0$, and

$$\begin{aligned} \ell_k \tau^{k-1} &= \frac{2^{k+1}}{\delta} - \frac{2^{m+2} - 2^{k+2}}{\delta d_0} \geq \frac{2^{k+1}}{\delta} - \frac{2^{m+2}}{\delta d_0}, \\ u_i \tau^{i-1} &= \frac{2^{i+1}}{\delta} + \frac{2^{m+2} - 2^{i+1}}{\delta d_0} \leq \frac{2^{i+1}}{\delta} + \frac{2^{m+2}}{\delta d_0} \qquad \text{for } i = 1, \ldots, k-1. \end{aligned}$$

Therefore,

$$\begin{aligned} \ell_k \tau^{k-1} - \sum_{i=1}^{k-1} u_i \tau^{i-1} &\geq \frac{2^{k+1}}{\delta} - \frac{2^{m+2}}{\delta d_0} - \frac{2^{k+1} - 4}{\delta} - \frac{(k-1)2^{m+2}}{\delta d_0}, \\ &\geq \frac{4}{\delta} - \frac{k 2^{m+2}}{\delta d_0}, \\ &\geq \frac{2}{\delta}\left(2 - \frac{m 2^{m+1}}{d_0}\right), \\ &\geq \frac{2}{\delta}. \end{aligned}$$

The last inequality holds since $d_0 \geq m 2^{m+1}$. □

Corollary 4.2.2. *Let d_0 and h be given as in Lemma 4.2.1. Then, the integer-valued vector $\widetilde{h} = \lfloor 2h \rfloor \in \mathbb{Z}^m_\oplus$ satisfies system* (4.2.1).

Proof. Let us use $\ell_k(h)$ and $u_k(h)$, for $k = 1, \ldots, m$, to emphasize that they are (linear) functions of h. Obviously, we have

$$\ell_k(2h)\tau^{k-1} - \sum_{i=1}^{k-1} u_i(2h)\tau^{i-1} \geq \frac{4}{\delta},$$

which implies, since $0 \leq \omega = 2h - \widetilde{h} < 1$, that

$$\begin{aligned}
\ell_k(\widetilde{h})\tau^{k-1} - \sum_{i=1}^{k-1} u_i(\widetilde{h})\tau^{i-1} &\geq \frac{4}{\delta} - \ell_k(\omega)\tau^{k-1} + \sum_{i=1}^{k-1} u_i(\omega)\tau^{i-1}, \\
&\geq \frac{4}{\delta} - \ell_k(\omega)\tau^{k-1}, \\
&\geq \frac{4}{\delta} - \frac{\omega_k \tau^{k-1}}{d_0 + 1}, \\
&\geq \frac{2}{\delta}.
\end{aligned}$$

Therefore, $\widetilde{h}$ satisfies system (4.2.1). □

It is easy to verify, for $k = 1, \ldots, m$, that the k^{th} inequality of (4.2.1) implies

$$\begin{cases} \ell_k \geq \dfrac{2}{\delta}, \\ \ell_k \tau^{k-j} - \displaystyle\sum_{i=j}^{k-1} u_i \tau^{i-j+1} \geq \dfrac{2}{\delta} \qquad \text{for } j = 1, \ldots, k-1. \end{cases} \tag{4.2.2}$$

The following lemma, which is the core of the main theorem of this chapter, states that the central path of problem (4.1.1) in a "central" part C_δ^{k+1} of the KM cube is pushed to a "corner" $\widehat{C}_\delta^k$ of the cube for $k = 1, \ldots, m-1$.

Lemma 4.2.3. *Let $d_0 \in \mathbb{R}_+$ and $h \in \mathbb{Z}_\oplus^m$ satisfy inequalities* (4.2.1). *Then, for problem* (4.1.1)*, we have*

$$\mathcal{C}_1(\tau, d_0, h) \cap C_\delta^{k+1} \subseteq \widehat{C}_\delta^k \qquad \textit{for } k = 1, \ldots, m-1.$$

Proof. We show, for every $k = 1, \ldots, m-1$, that every point on the central path $\mathcal{C}_1(\tau, d_0, h)$ that satisfies $\widehat{s}_{k+1} > \delta$ and $\bar{s}_{k+1} > \delta$ also satisfies

$$\bar{s}_k \leq \delta, \widehat{s}_{k-1} \leq \delta, \ldots, \widehat{s}_1 \leq \delta.$$

Recall that every point on the central path $\mathcal{C}_1(\tau, d_0, h)$ satisfies the first $m-1$ equations of system (4.1.3). From the k^{th} equation of system (4.1.3), we have

$$\frac{h_k}{d_0+\widehat{s}_k} - \frac{h_{k+1}\tau}{d_0+\widehat{s}_{k+1}} = -\frac{1}{\widehat{s}_k} + \frac{1}{\bar{s}_k} + \frac{\tau}{\widehat{s}_{k+1}} + \frac{\tau}{\bar{s}_{k+1}} \le \frac{1}{\bar{s}_k} + \frac{2\tau}{\delta},$$

which implies that

$$\ell_k = \frac{h_k}{d_0+1} - \frac{h_{k+1}\tau}{d_0} \le \frac{1}{\bar{s}_k} + \frac{2\tau}{\delta} \le \frac{1}{\bar{s}_k} + \frac{1}{\delta}.$$

Thus, using inequalities (4.2.2), we get $\bar{s}_k \le \delta$. Let $j = 1, \dots, k-1$. Adding the k^{th} equation of system (4.1.3) multiplied by τ^{k-j} to the i^{th} equation of system (4.1.3) multiplied by $-\tau^{i-j}$ for all $i = j, \dots, k-1$, we have

$$\frac{2h_k\tau^{k-j}}{d_0+\widehat{s}_k} - \frac{h_{k+1}\tau^{k-j+1}}{d_0+\widehat{s}_{k+1}} - \frac{h_j}{d_0+\widehat{s}_j} \le \frac{1}{\widehat{s}_j} + \frac{2\tau^{k-j+1}}{\delta},$$

which implies that

$$\ell_k\tau^{k-j} - \sum_{i=j}^{k-1} u_i\tau^{i-j+1} \le \frac{2h_k\tau^{k-j}}{d_0+1} - \frac{h_{k+1}\tau^{k-j+1}}{d_0} - \frac{h_j}{d_0} \le \frac{1}{\widehat{s}_j} + \frac{1}{\delta}.$$

Thus, using inequalities (4.2.2), we get $\widehat{s}_j \le \delta$. □

We now provide the main result of this chapter, which consists of the following two statements.

1. The analytic center $\chi^1(\tau, d_0, h)$ of $\mathcal{KM}_1(\tau, d_0, h)$ belongs to $\widehat{C}^m_\delta$, a δ-neighborhood of the vertex v^0.

2. The central path $\mathcal{C}_1(\tau, d_0, h)$ of problem (4.1.1) stays in P_δ, a δ-neighborhood of the simplex path.

Theorem 4.2.4. *Let $d_0 \in \mathbb{R}_+$ and $h \in \mathbb{Z}^m_\oplus$ satisfy conditions* (4.2.1)*. Then, for problem* (4.1.1)*, we have*

1. $\chi^1(\tau, d_0, h) \in \widehat{C}^m_\delta$.

2. $\mathcal{C}_1(\tau, d_0, h) \subseteq P_\delta$.

Proof. The following inequalities will be used throughout the proof:

$$d_0 \leq d_0 + \widehat{s}_i \leq d_0 + 1 \quad \text{for } i = 1, \ldots, m.$$

To prove the first statement, we show that the unique solution $\chi^1(\tau, d_0, h)$ of system (4.1.3) belongs to $\widehat{C}^m_\delta$, or equivalently, it satisfies

$$\bar{s}_m \leq \delta, \widehat{s}_{m-1} \leq \delta, \ldots, \widehat{s}_1 \leq \delta.$$

From the m^{th} equation of system (4.1.3), we have

$$\frac{h_m}{d_0 + \widehat{s}_m} = \frac{1}{\bar{s}_m} - \frac{1}{\widehat{s}_m} \leq \frac{1}{\bar{s}_m},$$

which implies that

$$\ell_m = \frac{h_m}{d_0 + 1} \leq \frac{1}{\bar{s}_m}.$$

Thus, using inequalities (4.2.2), we get $\bar{s}_m \leq \delta$. For $j = 1, \ldots, m-1$, adding the m^{th} equation of system (4.1.3) multiplied by τ^{m-j} to the i^{th} equation of system (4.1.3) multiplied by $-\tau^{i-j}$ for all $i = j, \ldots, m-1$, we have

$$\frac{2h_m\tau^{m-j}}{d_0 + \widehat{s}_m} - \frac{h_j}{d_0 + \widehat{s}_j} = \frac{1}{\widehat{s}_j} - \frac{1}{\bar{s}_j} - \sum_{i=j+1}^{m} \frac{2\tau^{i-j}}{\bar{s}_i} \leq \frac{1}{\widehat{s}_j},$$

which implies that

$$\ell_m\tau^{m-j} - \sum_{i=j}^{m-1} u_i\tau^{i-j+1} \leq \frac{2h_m\tau^{m-j}}{d_0 + 1} - \frac{h_j}{d_0} \leq \frac{1}{\widehat{s}_j}.$$

Thus, using inequalities (4.2.2), we get $\widehat{s}_j \leq \delta$.

We now prove the second statement of the theorem. For the sake of simplicity, we let $\mathcal{C}_1 = \mathcal{C}_1(\tau, d_0, h)$ and $\mathcal{KM}_1 = \mathcal{KM}_1(\tau, d_0, h)$. For $k = 2, \ldots, m$, we have

$$\begin{aligned}
\mathcal{C}_1 &= \mathcal{C}_1 \cap \mathcal{KM}_1, \\
&= \mathcal{C}_1 \cap \left(T^k_\delta \cup C^k_\delta \cup B^k_\delta\right), \\
&= T^k_\delta \cup \left(\mathcal{C}_1 \cap C^k_\delta\right) \cup B^k_\delta, \\
&\subseteq T^k_\delta \cup \widehat{C}^{k-1}_\delta \cup B^k_\delta, \qquad \text{by Lemma 4.2.3} \\
&= A^k_\delta,
\end{aligned}$$

which implies that $\mathcal{C}_1 \subseteq \bigcap_{k=2}^m A^k_\delta = P_\delta$. □

4.3 Bit-Length Based Complexity of LPs

Let (RP_1^m) denote the redundant KM problem (4.1.1) in which the parameters τ, δ, d_0, and h are chosen as

$$\begin{cases} \tau = \dfrac{m}{2(m+1)}, \\ \delta = \dfrac{1}{4(m+1)}, \\ d_0 = m2^{m+1} - 1, \\ h_k = \left\lfloor 2m^2\left(1+\dfrac{1}{m}\right)^k \left(2^{2m+k+5} - 2^{m+2k+4}\right) \right\rfloor \qquad \text{for } k = 1, \ldots, m. \end{cases}$$

The following theorem provides the iteration complexity bounds for problem (RP_1^m). Recall that the theoretical iteration complexity upper bound for LPs is known to be $\mathcal{O}(\sqrt{n}L)$, where n is the number of inequalities and L is the bit-length of the input-data given by formula (1.3.1).

Theorem 4.3.1. *The iteration complexity lower and upper bounds for problem* (RP_1^m) *are* $\Omega(2^m)$ *and* $\mathcal{O}\left(m^4 2^{4.5m}\right)$.

Proof. The number of the inequalities of problem (RP_1^m) is

$$\begin{aligned} n &= 2m + e^T h, \\ &\le 2m + 2m^2 \sum_{k=1}^{m} \left(1+\frac{1}{m}\right)^k \left(2^{2m+k+5} - 2^{m+2k+4}\right), \\ &\le 2m + m^2 \sum_{k=1}^{m} 2^{2m+k+8}, \end{aligned}$$

or $n = \mathcal{O}\left(m^2 2^{3m}\right)$. Using formula (1.3.1), we can calculate the bit-length L of the input-data of problem (RP_1^m) as

$$\begin{aligned} L &= m + \lceil \log 2 \rceil + 4m\left(1 + \lceil \log 2 \rceil\right) + 3m\left(1 + \lceil \log \tau \rceil\right) + \\ &\quad + \left(1 + \lceil \log\left(1 + d_0\right) \rceil\right) \sum_{k=1}^{m} h_k, \\ &= \mathcal{O}\left(n \log d_0\right), \end{aligned}$$

implying that $L = \mathcal{O}(nm) = \mathcal{O}\left(m^3 2^{3m}\right)$. Therefore, the iteration complexity upper bound for problem (RP_1^m) is $\mathcal{O}\left(m^4 2^{4.5m}\right)$.

On the other hand, since the central path closely follows the simplex path, it has $2^m - 2$ "sharp" turns[1] and $2^m - 1$ almost-straight line segments; see Figure 4.1. As a consequence, central-path-following IPMs that work in a small neighborhood of the central path take at least $2^m - 1$ iterations to find the optimal solution. This implies that the iteration complexity lower bound for problem (RP_1^m) is $\Omega\left(2^m\right)$. □

Corollary 4.3.2. *The iteration complexity bounds for problem* (RP_1^m) *are* $\Omega\big(\sqrt[3]{n/(\log n)^2}\big)$ *and* $\mathcal{O}\big(n\sqrt{n}\log n\big)$.

4.4 Summary

1. We showed that, without changing the geometry of the feasible set $\mathcal{KM}(\tau)$, the central path can be forced to follow the simplex path so that it visits an arbitrarily small neighborhood of each vertex of the KM m-cube by carefully adding redundant constraints.

2. This result highlights that although the central path is a smooth analytic curve in the interior of the feasible set, it might be severely distracted by redundancy. In particular, exponentially many redundant constraints, inter-playing with the geometry of the problem, may force the central path to take exponentially many and arbitrarily sharp turns.

3. The redundant KM example gives a lower bound for the number of iterations required by central-path-following IPMs to reach the optimal point. We have shown that the iteration complexity bounds for this example are $\Omega\big(\sqrt[3]{n/(\log n)^2}\big)$ and $\mathcal{O}\big(n\sqrt{n}\log n\big)$, where n is the number of inequalities.

4. The result of this chapter is true at least for 1^{st}-order central-path-following IPMs. However, it can be easily verified, with similar arguments, that the result is true even for 2^{nd}-order central-path-following IPMs as at least $2^{m-1} - 1$ iterations are needed to reach the optimal point. Also, it might be extended to k^{th}-order

[1] As stated in Theorem 2.1.1, the central path is an analytic curve, i.e., infinitely many times differentiable. However, the simplex path is a continuous piece-wise linear path, and the angle between every two adjacent linear pieces is at most $\pi/2 + \arctan\tau$. Here, by a sharp turn of the central path, we mean that the central path can be arbitrarily close to a break point of the simplex path.

method, which requires at least $2^{m-k+1}-1$ iterations to converge to the optimal solution.

5. State-of-the-art preprocessing tools in modern linear optimization software packages would eliminate the added redundant inequalities. Therefore, IPM-based codes would solve the preprocessed redundant KM problem (RP_1^m) efficiently, just as commercial simplex codes do solve the problem in only one pivot.

Chapter 5

Tight Complexity Bounds of Central-Path-Following IPMs

In this chapter, we provide another redundant LP based on the KM problem (3.1.1). In Chapter 4, we used uniform distances for the redundant constraints which are placed parallel to the corresponding facets. In the construction presented in this chapter, we allow the distances to decay geometrically and show that fewer number of redundant constraints are enough to bend the central path along the simplex path. This construction provides better iteration complexity bounds. In addition, we utilize the special structure of the KM cube to tighten the gap between the iteration complexity lower and upper bounds.

5.1 Introduction

We consider the redundant KM problem

$$\begin{array}{lll} \min & y_m & \\ \text{s.t.} & \tau y_{k-1} \leq y_k \leq 1 - \tau y_{k-1} & \text{for } k = 1, \ldots, m, \\ & \tau y_{k-1} \leq y_k + d_k & \text{repeated } h_k \text{ times for } k = 1, \ldots, m, \end{array} \tag{5.1.1}$$

with the convention $y_0 = 0$. The distances d_k for $k = 1, \ldots, m$ are allowed to be distinct in contrast to problem (4.1.1) in which all the distances are equal to d_0. We denote by $\mathcal{KM}_2(\tau, d, h)$ the feasible set represented by the constraints of problem (5.1.1), where $d = (d_1; \cdots; d_m)$ and $h = (h_1; \cdots; h_m)$. Note that the sets $\mathcal{KM}(\tau)$ and $\mathcal{KM}_2(\tau, d, h)$ are geometrically identical for all $\tau > 0$,

$d \geq 0$, and h, while they differ in the algebraic representation. The analysis given here is analogous to the one presented in Chapter 4.

The logarithmic barrier function for problem (5.1.1) is defined as

$$F_2(y) = -e^T \log \widehat{s} - e^T \log \bar{s} - h^T \log(d + \widehat{s}).$$

where the slack vectors $\widehat{s}_k$ and $\bar{s}_k$ are defined as (4.1.2). By Definition 1.3.4, the analytic center of $\mathcal{KM}_2(\tau, d, h)$ is then given as

$$\chi^2(\tau, d, h) = \arg\min \left\{ F_2(y) \mid (\widehat{s}; \bar{s}; d + \widehat{s}) > 0 \right\}.$$

Therefore, the analytic center $\chi^2(\tau, d, h)$ is the unique solution of the necessary and sufficient optimality conditions

$$\begin{cases} \frac{1}{\widehat{s}_k} - \frac{1}{\bar{s}_k} - \frac{\tau}{\widehat{s}_{k+1}} - \frac{\tau}{\bar{s}_{k+1}} + \frac{h_k}{d_k + \widehat{s}_k} - \frac{h_{k+1}\tau}{d_{k+1} + \widehat{s}_{k+1}} = 0, & k = 1, \ldots, m-1, \\ \frac{1}{\widehat{s}_m} - \frac{1}{\bar{s}_m} + \frac{h_m}{d_m + \widehat{s}_m} = 0, & \\ \widehat{s}_k > 0, \bar{s}_k > 0, d_k + \widehat{s}_k > 0, & k = 1, \ldots, m. \end{cases} \tag{5.1.2}$$

By Definition 1.3.3, the central path of problem (5.1.1) is the set

$$\mathcal{C}_2(\tau, d, h) = \left\{ y(\mu) \mid \mu > 0 \right\},$$

in which the function $y(\mu)$ is defined as

$$y(\mu) = \arg\min \left\{ y_m + \mu F_2(y) \mid (\widehat{s}; \bar{s}; d + \widehat{s}) > 0 \right\}.$$

Therefore, from the optimality conditions, every $y(\mu)$ satisfies the first $m - 1$ equations of system (5.1.2).

5.2 Reducing the Number of Redundant Constraints

In this section, we assume that $m \geq 2$, $\tau > 0$, and $\delta > 0$ with $\tau + \delta < 1/2$. The analysis provided in this section is similar to, but not extension of, the one given in § 4.2. We define, for $k = 1, \ldots, m-1$,

$$\ell_k = \frac{h_k \tau^{k-1}}{d_k + 1} - \frac{h_{k+1}\tau^k}{d_{k+1}},$$

$$u_{k+1} = \frac{h_1}{d_1} - \frac{h_{k+1}\tau^k}{d_{k+1} + 1},$$

as well as $\ell_m = h_m\tau^{m-1}/(d_m+1)$ and $u_1 = 0$. In the following lemmas, we show that there exist $d \in \mathbb{R}^m_+$ and $h \in \mathbb{Z}^m_\oplus$ such that they satisfy

$$\begin{cases} \ell_k - u_k \geq \dfrac{2}{\delta} \\ \ell_k, \quad u_k \geq 0 \end{cases} \quad \text{for } k = 1,\dots,m. \tag{5.2.1}$$

Lemma 5.2.1. *System* (5.2.1) *holds for*

$$\begin{cases} d_k = m\left(2+\dfrac{2}{m}\right)^{m-k+2} \\ h_k = \dfrac{d_k+1}{\tau^{k-1}\delta}\left[\left(2+\dfrac{2}{m}\right)^{m+1} - \left(2+\dfrac{2}{m}\right)^{k}\right] \end{cases} \quad \textit{for } k = 1,\dots,m.$$

Proof. For every $k = 1,\dots,m$, substituting the value of h_k in the definitions of ℓ_k and u_k, we have

$$\ell_k = \frac{1}{\delta}\left[\left(2+\frac{2}{m}\right)^{k+1} - \left(2+\frac{2}{m}\right)^{k} - \frac{\left(2+\frac{2}{m}\right)^{m+1} - \left(2+\frac{2}{m}\right)^{k+1}}{d_{k+1}}\right],$$
$$u_k = \frac{1}{\delta}\left[\left(2+\frac{2}{m}\right)^{k} - \left(2+\frac{2}{m}\right) + \frac{\left(2+\frac{2}{m}\right)^{m+1} - \left(2+\frac{2}{m}\right)}{d_1}\right],$$

implying that $\ell_k \geq 0$, $u_k \geq 0$, and

$$\ell_k \geq \frac{1}{\delta}\left[\left(2+\frac{2}{m}\right)^{k+1} - \left(2+\frac{2}{m}\right)^{k} - \frac{\left(2+\frac{2}{m}\right)^{m+1}}{d_{k+1}}\right],$$
$$u_k \leq \frac{1}{\delta}\left[\left(2+\frac{2}{m}\right)^{k} - \left(2+\frac{2}{m}\right) + \frac{\left(2+\frac{2}{m}\right)^{m+1}}{d_1}\right].$$

Substituting the values of d_1 and d_{k+1}, we obtain

$$\ell_k - u_k \geq \frac{1}{\delta}\left[\frac{1}{m}\left(2+\frac{2}{m}\right)^{k} + 2 + \frac{1}{m}\right] \geq \frac{2}{\delta},$$

and the result follows. □

Corollary 5.2.2. *Let d and h be given as in Lemma 5.2.1. Then, the integer-valued vector $\widetilde{h} = \lfloor 2h \rfloor \in \mathbb{Z}^m_\oplus$ satisfies system* (5.2.1).

Proof. We skip the proof as it is similar to the proof of Corollary 4.2.2. □

The following lemma plays an important role in proving the main result which will be stated in the subsequent theorem.

Lemma 5.2.3. *Assume that vectors* $d \in \mathbb{R}^m_+$ *and* $h \in \mathbb{Z}^m_\oplus$ *satisfy conditions* (5.2.1)*. Then, for problem* (5.1.1)*, we have*

$$\mathcal{C}_2(\tau, d, h) \cap C^{k+1}_\delta \subseteq \widehat{C}^k_\delta \quad \textit{for } k = 1, \ldots, m-1.$$

Proof. From the fact that $\ell_k \geq 0$ for $k = 1, \ldots, m$, we have

$$\frac{h_1}{d_1} \geq \frac{h_k \tau^{k-1}}{d_k} \quad \text{for } k = 1, \ldots, m. \tag{5.2.2}$$

We show, for every $k = 1, \ldots, m-1$, that every point on the central path $\mathcal{C}_2(\tau, d, h)$ that satisfies $\widehat{s}_{k+1} > \delta$ and $\bar{s}_{k+1} > \delta$ also satisfies

$$\bar{s}_k \leq \delta, \widehat{s}_{k-1} \leq \delta, \ldots, \widehat{s}_1 \leq \delta.$$

Recall that every point on the central path $\mathcal{C}_2(\tau, d, h)$ satisfies the first $m-1$ equations of system (5.1.2). From the k^{th} equation of system (5.1.2) multiplied by τ^{k-1}, we have

$$\begin{aligned}\frac{h_k \tau^{k-1}}{d_k + \widehat{s}_k} - \frac{h_{k+1} \tau^k}{d_{k+1} + \widehat{s}_{k+1}} &= -\frac{\tau^{k-1}}{\widehat{s}_k} + \frac{\tau^{k-1}}{\bar{s}_k} + \frac{\tau^k}{\widehat{s}_{k+1}} + \frac{\tau^k}{\bar{s}_{k+1}} \\ &\leq \frac{\tau^{k-1}}{\bar{s}_k} + \frac{2\tau^k}{\delta},\end{aligned}$$

which implies that

$$\begin{aligned}\ell_k = \frac{h_k \tau^{k-1}}{d_k + 1} - \frac{h_{k+1} \tau^k}{d_{k+1}} &\leq \frac{\tau^{k-1}}{\bar{s}_k} + \frac{2\tau^k}{\widehat{s}_{k+1}} \\ &\leq \frac{1}{\bar{s}_k} + \frac{1}{\delta}.\end{aligned}$$

Thus, using inequalities (5.2.1), we get $\bar{s}_k \leq \delta$. Let $j = 1, \ldots, k-1$. Adding the k^{th} equation of system (5.1.2) multiplied by τ^{k-1} to the i^{th} equation of system (5.1.2) multiplied by $-\tau^{i-1}$ for all $i = j, \ldots, k-1$, we have

$$\frac{2h_k \tau^{k-1}}{d_k + \widehat{s}_k} - \frac{h_{k+1} \tau^k}{d_{k+1} + \widehat{s}_{k+1}} - \frac{h_j \tau^{j-1}}{d_j + \widehat{s}_j} \leq \frac{\tau^{j-1}}{\widehat{s}_j} + \frac{2\tau^k}{\delta},$$

which implies, using inequalities (5.2.2), that

$$\begin{aligned}\ell_k - u_k &\leq \frac{2h_k\tau^{k-1}}{d_k+1} - \frac{h_{k+1}\tau^k}{d_{k+1}} - \frac{h_j\tau^{j-1}}{d_j}\\ &\leq \frac{\tau^{j-1}}{\widehat{s}_j} + \frac{1}{\delta}.\end{aligned}$$

Thus, using inequalities (5.2.1), we get $s_j \leq \delta$. □

We now provide the main result of this chapter, which consists of the following two statements.

1. The analytic center $\chi^2(\tau,d,h)$ of $\mathcal{KM}_2(\tau,d,h)$ belongs to $\widehat{C}^m_\delta$, a δ-neighborhood of the vertex v^0.
2. The central path $\mathcal{C}_2(\tau,d,h)$ of problem (5.1.1) stays in P_δ, a δ-neighborhood of the simplex path.

Theorem 5.2.4. *Assume that vectors $d \in \mathbb{R}^m_+$ and $h \in \mathbb{Z}^m_\oplus$ satisfy conditions* (5.2.1). *Then, for problem* (5.1.1), *we have*

1. $\chi^2(\tau,d,h) \in \widehat{C}^m_\delta$.
2. $\mathcal{C}_2(\tau,d,h) \subseteq P_\delta$.

Proof. The second statement follows from Lemma 5.2.3, with a similar argument as in the proof of Theorem 4.2.4. To prove the first statement, we show that the unique solution $\chi^2(\tau,d,h)$ of system (5.1.2) belongs to $\widehat{C}^m_\delta$, or equivalently, it satisfies

$$\bar{s}_m \leq \delta, \widehat{s}_{m-1} \leq \delta, \ldots, \widehat{s}_1 \leq \delta.$$

First, we note that

$$d_i \leq d_i + \widehat{s}_i \leq d_i + 1 \quad \text{for } i = 1, \ldots, m.$$

From the m^{th} equation of system (5.1.2) multiplied by τ^{m-1}, we have

$$\frac{h_m\tau^{m-1}}{d_m+\widehat{s}_m} = \frac{\tau^{m-1}}{\bar{s}_m} - \frac{\tau^{m-1}}{\widehat{s}_m} \leq \frac{\tau^{m-1}}{\bar{s}_m},$$

which implies that

$$\ell_m = \frac{h_m \tau^{m-1}}{d_m + 1} < \frac{\tau^{m-1}}{\bar{s}_m}.$$

Thus, using inequalities (5.2.1), we get $\bar{s}_m < \delta$. For all $j = 1, \dots, m-1$, adding the m^{th} equation of system (5.1.2) multiplied by τ^{m-1} to the i^{th} equation of system (5.1.2) multiplied by $-\tau^{i-1}$ for all $i = j, \dots, m-1$, we have

$$\frac{2h_m \tau^{m-1}}{d_m + \widehat{s}_m} - \frac{h_j \tau^{j-1}}{d_j + \widehat{s}_j} = \frac{\tau^{j-1}}{\widehat{s}_j} - \frac{\tau^{j-1}}{\bar{s}_j} - \sum_{i=j+1}^{m} \frac{2\tau^{i-1}}{\bar{s}_i} \leq \frac{\tau^{j-1}}{\widehat{s}_j},$$

which implies, using inequalities (5.2.2), that

$$\ell_m - u_m \leq \frac{2h_m \tau^{m-1}}{d_m + 1} - \frac{h_j \tau^{j-1}}{d_j} \leq \frac{\tau^{j-1}}{\widehat{s}_j}.$$

Thus, using inequalities (5.2.1), we get $\widehat{s}_j \leq \delta$. □

5.3 Tight Iteration Complexity Bounds

As mentioned in § 4.3, the theoretical iteration complexity upper bound for LPs is known to be $\mathcal{O}(\sqrt{n}L)$, where n is the number of inequalities and L is the bit-length of the input-data given by formula (1.3.1). Alternatively, we have a well-known iteration complexity upper bound based on a given accuracy parameter. For a given accuracy ε, the number of iterations required by central-path-following IPMs to reach the duality gap smaller than ε is bounded above by

$$\mathcal{O}\left(\sqrt{n} \log \frac{g_0}{\varepsilon}\right),$$

where g_0 is the duality gap at the starting point.

Let (RP_2^m) denote the redundant KM problem (5.1.1) in which the parameters

τ, δ, d, and h are chosen as

$$\begin{cases} \tau = \dfrac{m}{2(m+1)}, \\ \delta = \dfrac{1}{4(m+1)}, \\ d_k = m\left(2+\dfrac{2}{m}\right)^{m-k+2} & \text{for } k=1,\dots,m, \\ h_k = \left\lfloor 4m^2\left(\left(2+\dfrac{2}{m}\right)^{2m+3} - \left(2+\dfrac{2}{m}\right)^{m+k+2}\right)\right\rfloor & \text{for } k=1,\dots,m. \end{cases}$$

Now, using problem (RP_2^m), we give lower and upper bounds for the iteration complexity of central-path-following IPMs. Then, we tighten the bounds by utilizing the special structure of the KM cube.

Theorem 5.3.1. *The iteration complexity lower and upper bounds for problem* (RP_2^m) *are* $\Omega(2^m)$ *and* $\mathcal{O}\left(m^{5.5}2^{3m}\right)$.

Proof. Similar to the discussion in Theorem 4.3.1, we can show that the iteration complexity for problem (RP_2^m) is at least $\Omega\left(2^m\right)$. On the other hand, since

$$h_k \le 4m^2\left(2+\frac{2}{m}\right)^{2m+3} = m^2 2^{2m+5}\left(1+\frac{1}{m}\right)^{2m+3} \le m^2 2^{2m+12},$$

the number of inequalities is $n = 2m + e^T h = \mathcal{O}\left(m^3 2^{2m}\right)$. Therefore, since $L = \mathcal{O}\left(nm\right)$, see the discussion in the proof of Theorem 4.3.1, the iteration complexity upper bound is $\mathcal{O}\left(m^{5.5}2^{3m}\right)$. □

The following lemma shows which part of the central path belongs to the set $\widehat{C}_\delta^m$, which is a δ-neighborhood of the vertex v^0.

Lemma 5.3.2. *Let* $\mu \ge \tau^{m-1}\delta$ *and* $y(\mu)$ *be the* μ*-center of problem* (RP_2^m)*. Then, we have* $y(\mu) \in \widehat{C}_\delta^m$*.*

Proof. From Theorem 5.2.4, the set $\left\{\mu \mid y(\mu) \in \widehat{C}_\delta^m\right\}$ is nonempty. Thus, there exist some $\mu > 0$ such that $y(\mu)$ satisfies

$$\widehat{s}_1 = \delta, \widehat{s}_2 \le \delta, \dots, \widehat{s}_{m-1} \le \delta, \bar{s}_m \le \delta. \tag{5.3.1}$$

Let μ' be the maximum of such μ's. Therefore, we have $\{y(\mu) \mid \mu \geq \mu'\} \subseteq \widehat{C}_\delta^m$, and it suffices to prove that $\mu' \leq \tau^{m-1}\delta$.

We consider the dual formulation of problem (RP_2^m). For $k = 1, \ldots, m$, let us denote the dual variables corresponding to the slack variables $\widehat{s}_k$ and $\bar{s}_k$ by $\widehat{x}_k$ and $\bar{x}_k$, respectively. Since the k^{th} redundant inequality $d_k + \widehat{s}_k \geq 0$ is repeated h_k times, we denote the dual variables corresponding to each such inequality by $\widetilde{x}_{kj}$, for $j = 1, \ldots, h_k$. The dual formulation of problem (RP_2^m) is given by

$$
\begin{aligned}
\min \quad & \sum_{k=1}^{m} \left(\bar{x}_k + d_k \sum_{j=1}^{h_k} \widetilde{x}_{kj} \right) \\
\text{s. t.} \quad & \bar{x}_k - \widehat{x}_k + \tau \bar{x}_{k+1} + \tau \widehat{x}_{k+1} - \\
& \qquad - \sum_{j=1}^{h_k} \widetilde{x}_{kj} + \tau \sum_{j=1}^{h_{k+1}} \widetilde{x}_{k+1,j} = 0 \qquad \text{for } k = 1, \ldots, m-1, \\
& \qquad \widehat{x}_m - \bar{x}_m + \sum_{j=1}^{h_m} \widetilde{x}_{mj} = 1, \\
& \qquad \widehat{x}_k \geq 0, \quad \bar{x}_k \geq 0 \qquad \text{for } k = 1, \ldots, m, \\
& \qquad \widetilde{x}_{kj} \geq 0 \qquad \text{for } \begin{cases} j = 1, \ldots, h_k, \\ k = 1, \ldots, m. \end{cases}
\end{aligned}
$$

Multiplying the k^{th} equality constraint of this problem by τ^{k-1} for all $k = 1, \ldots, m$ and adding them up, we obtain

$$
\bar{x}_1 - \widehat{x}_1 + 2 \sum_{k=2}^{m-1} \tau^{k-1} \bar{x}_k + 2\tau^{m-1} \widehat{x}_m - \sum_{j=1}^{h_1} \widetilde{x}_{1j} + 2 \sum_{j=1}^{h_m} \tau^{m-1} \widetilde{x}_{mj} = \tau^{m-1},
$$

which implies that

$$
\bar{x}_1 \leq \widehat{x}_1 - 2 \sum_{j=1}^{h_m} \tau^{m-1} \widetilde{x}_{mj} + \sum_{j=1}^{h_1} \widetilde{x}_{1j} + \tau^{m-1}. \tag{5.3.2}
$$

The μ'-center satisfies (5.3.1) as well as

$$
\begin{aligned}
\widehat{x}_1(\mu') \widehat{s}_1(\mu') &= \mu', \\
\widetilde{x}_{1j}(\mu') \big(d_1 + \widehat{s}_1(\mu') \big) &= \mu' \qquad \text{for } j = 1, \ldots, h_1, \\
\widetilde{x}_{mj}(\mu') \big(d_m + \widehat{s}_m(\mu') \big) &= \mu' \qquad \text{for } j = 1, \ldots, h_m.
\end{aligned}
$$

Therefore, we get

$$\begin{aligned}
\widehat{x}_1(\mu') &= \frac{\mu'}{\delta}, \\
\widetilde{x}_{1j}(\mu') &\le \frac{\mu'}{d_1} && \text{for } j = 1, \ldots, h_1, \\
\widetilde{x}_{mj}(\mu') &\ge \frac{\mu'}{d_m + 1} && \text{for } j = 1, \ldots, h_m.
\end{aligned}$$

Combining with inequality (5.3.2) that holds for every feasible point, we have

$$\begin{aligned}
\bar{x}_1(\mu') &\le \left(\frac{1}{\delta} - \frac{2h_m\tau^{m-1}}{d_m + 1} + \frac{h_1}{d_1}\right)\mu' + \tau^{m-1}, \\
&= \left(\frac{1}{\delta} - \ell_m + u_m\right)\mu' + \tau^{m-1},
\end{aligned}$$

which implies, using inequalities (5.2.1) and $\bar{x}_1(\mu') \ge 0$, that $\mu' \le \tau^{m-1}\delta$. □

Now, we are in position to present the main theorem of this section that provides tighter iteration complexity bounds for problem (RP_2^m).

Theorem 5.3.3. *The iteration complexity lower and upper bounds for problem* (RP_2^m) *are* $\Omega(2^m)$ *and* $\mathcal{O}(m^{2.5}2^m)$.

Proof. It has already been discussed that the iteration complexity lower bound for problem (RP_2^m) is $\Omega(2^m)$. We show that the number of iterations to solve problem (RP_2^m) is bounded above by $\mathcal{O}(m^{2.5}2^m)$.

The iteration complexity upper bound is given as

$$\mathcal{O}\left(\sqrt{n}\log\frac{n\mu^0}{\varepsilon}\right),$$

where μ^0 is the central path parameter at the starting point, $n = 2m + e^T h$ is the number of inequalities, and ε is a given accuracy. Lemma 5.3.2 guarantees that the μ^0-center with $\mu^0 = \tau^{m-1}\delta$ belongs to $\widehat{C}_\delta^m$. It can be verified that the primal-dual solution given as

$$\begin{aligned}
y^* = 0, \quad \widehat{x}^* &= (\tau^{m-1}; \cdots; \tau; 1), \quad & \widehat{s}^* &= 0, \\
\bar{x}^* &= 0, & \bar{s}^* &= e,
\end{aligned}$$

is the unique optimal solution of problem (RP_2^m). It follows that Lemma 2.2.1 holds with $\gamma = \tau^{m-1}$. Therefore, by Corollary 2.2.3, we can set $\varepsilon = \beta\gamma^2/n$ to reveal the optimal partition, hence, the unique optimal point. As a consequence, the iteration complexity upper bound is

$$\mathcal{O}\left(\sqrt{n}\log\frac{n\mu^0}{\varepsilon}\right) = \mathcal{O}\left(\sqrt{n}\log\frac{n^2\delta}{\beta\tau^{m-1}}\right) = \mathcal{O}\left(m^{2.5}2^m\right),$$

since β is an absolute positive constant, $\delta = \frac{1}{4(m+1)}$, $\tau < \frac{1}{2}$, and $n = \mathcal{O}(m^3 2^{2m})$, as shown in the proof of Theorem 5.3.1. □

Corollary 5.3.4. *The iteration complexity bounds for problem* (RP_2^m) *are* $\Omega\big(\sqrt{n/(\log n)^3}\big)$ *and* $\mathcal{O}\big(\sqrt{n}\log n\big)$.

5.4 Summary

1. We provided another variant of the redundant KM problem in which the distances of the redundant constraints are allowed to decay geometrically. In comparison with the previous variant discussed in Chapter 4, the number of redundant constraints significantly reduced from $\mathcal{O}\left(m^2 2^{3m}\right)$ to $\mathcal{O}\left(m^3 2^{2m}\right)$. As a consequence, the iteration complexity bounds became tighter.

2. Besides gaining a significant reduction in the number of inequalities, we also utilized the special structure of the KM cube to adjust the termination criterion that resulted to further tightening of the complexity bounds. We obtained the iteration complexity bounds $\Omega\big(\sqrt{n/(\log n)^3}\big)$ and $\mathcal{O}\big(\sqrt{n}\log n\big)$, where n is the number of inequalities.

Chapter 6

A Worst-Case Example for Central-Path-Following IPMs

In this chapter, we simplify the constructions presented in Chapters 4 and 5 by placing redundant constraints parallel to the coordinate hyperplanes. We discuss two cases; the redundant constraints are placed at uniform distances or at geometrically decaying distances. It turns out that in this construction, fewer redundant inequalities are needed to bend the central path along the simplex path. In particular, this redundant KM example appears to be the worst-case example for central-path-following IPMs as it yields the tightest iteration complexity bounds.

6.1 Introduction

We consider the following variant of the redundant KM example in which the redundant constraints are put parallel to the coordinate hyperplanes.

$$\begin{array}{lll} \min & y_m & \\ \text{s.t.} & \tau y_{k-1} \leq y_k \leq 1 - \tau y_{k-1} & \text{for } k = 1, \ldots, m, \\ & \quad 0 \leq y_k + d_k & \text{repeated } h_k \text{ times for } k = 1, \ldots, m, \end{array} \tag{6.1.1}$$

with the convention $y_0 = 0$. We denote by $\mathcal{KM}_*(\tau, d, h)$ the feasible set represented by the constraints of problem (6.1.1), where $d = (d_1; \cdots; d_m)$ and $h = (h_1; \cdots; h_m)$. Note that the sets $\mathcal{KM}(\tau)$ and $\mathcal{KM}_*(\tau, d, h)$ are geometrically identical for all $\tau > 0$, $d \geq 0$, and h, while they differ in the algebraic representation.

The logarithmic barrier function for problem (6.1.1) is defined as

$$F_*(y) = -e^T \log \widehat{s} - e^T \log \bar{s} - h^T \log(d + y).$$

where the slack vectors $\widehat{s}_k$ and $\bar{s}_k$ are defined as (4.1.2). By Definition 1.3.4, the analytic center of $\mathcal{KM}_*(\tau, d, h)$ is then given as

$$\chi^*(\tau, d, h) = \arg\min \{F_*(y) \mid (\widehat{s}; \bar{s}; d + y) > 0\}.$$

Therefore, the analytic center $\chi^*(\tau, d, h)$ is the unique solution of the necessary and sufficient optimality conditions

$$\begin{cases} \frac{1}{\widehat{s}_k} - \frac{1}{\bar{s}_k} - \frac{\tau}{\widehat{s}_{k+1}} - \frac{\tau}{\bar{s}_{k+1}} + \frac{h_k}{d_k+y_k} - \frac{h_{k+1}\tau}{d_{k+1}+y_{k+1}} = 0, & k = 1, \dots, m-1, \\ \frac{1}{\widehat{s}_m} - \frac{1}{\bar{s}_m} + \frac{h_m}{d_m+y_m} = 0, & \\ \widehat{s}_k > 0, \bar{s}_k > 0, d_k + y_k > 0, & k = 1, \dots, m. \end{cases} \tag{6.1.2}$$

By Definition 1.3.3, the central path of problem (6.1.1) is the set

$$\mathcal{C}_*(\tau, d, h) = \{y(\mu) \mid \mu > 0\},$$

in which the function $y(\mu)$ is defined as

$$y(\mu) = \arg\min \{y_m + \mu F_*(y) \mid (\widehat{s}; \bar{s}; d + y) > 0\}.$$

Therefore, from the optimality conditions, every $y(\mu)$ satisfies the first $m-1$ equations of system (6.1.2).

6.2 Optimal Repetition and Distance Vectors

In this section, we assume that $m \geq 2$, $\tau > 0$, and $\delta > 0$ with $\tau + \delta < 1/2$. We define, for $k = 1, \dots, m$,

$$\ell_k = \frac{h_k \tau^{k-1}}{d_k + 1},$$

$$u_k = \sum_{i=1}^{k-1} \frac{h_i \tau^{i-1}}{d_i}.$$

Let us also define the set $\mathscr{D} = \{d \in \mathbb{R}^m \mid\ d_1 > 0, \dots, d_{m-1} > 0, d_m \geq 0\}$. In the following lemmas, we show how to choose $d \in \mathscr{D}$ and $h \in \mathbb{Z}^m_{\oplus}$ such that they satisfy

$$\ell_k - u_k \geq \frac{2}{\delta} \quad \text{for } k = 1, \dots, m. \tag{6.2.1}$$

Lemma 6.2.1. *Let $d \in \mathscr{D}$ and define*

$$h_k = \frac{2(d_k+1)}{\tau^{k-1}\delta} \prod_{i=1}^{k-1} \left(2 + \frac{1}{d_i}\right) \quad \text{for } k = 1, \ldots, m. \tag{6.2.2}$$

Then, we have $\ell_k - u_k = 2/\delta$ for $k = 1, \ldots, m$.

Proof. It is easy to verify that $\ell_1 - u_1 = 2/\delta$. For all $k = 2, \ldots, m$, we have

$$\begin{aligned}
\ell_k - u_k =& \frac{h_k \tau^{k-1}}{d_k + 1} - \sum_{i=1}^{k-1} \frac{h_i \tau^{i-1}}{d_i}, \\
=& \frac{2}{\delta} \prod_{i=1}^{k-1} \left(2 + \frac{1}{d_i}\right) - \frac{2}{\delta} \sum_{i=1}^{k-1} \left(1 + \frac{1}{d_i}\right) \prod_{i=1}^{i-1} \left(2 + \frac{1}{d_i}\right), \\
=& \frac{2}{\delta} \prod_{i=1}^{k-1} \left(2 + \frac{1}{d_i}\right) - \frac{2}{\delta} \left(1 + \frac{1}{d_{k-1}}\right) \prod_{i=1}^{k-2} \left(2 + \frac{1}{d_i}\right) - \\
& \quad \frac{2}{\delta} \sum_{i=1}^{k-2} \left(1 + \frac{1}{d_i}\right) \prod_{i=1}^{i-1} \left(2 + \frac{1}{d_i}\right), \\
=& \frac{2}{\delta} \prod_{i=1}^{k-2} \left(2 + \frac{1}{d_i}\right) - \frac{2}{\delta} \sum_{i=1}^{k-2} \left(1 + \frac{1}{d_i}\right) \prod_{i=1}^{i-1} \left(2 + \frac{1}{d_i}\right), \\
=& \ell_{k-1} - u_{k-1},
\end{aligned}$$

hence, the proof is completed by induction. □

Corollary 6.2.2. *Given a vector $d \in \mathscr{D}$, we define h as* (6.2.2)*. Then, the integer-valued vector $\widetilde{h} = \lfloor 2h \rfloor \in \mathbb{Z}^m_{\oplus}$ satisfies* (6.2.1)*.*

Proof. For $k = 1, \ldots, m$, let us use $\ell_k(h)$ and $u_k(h)$ to emphasize that they are (linear) functions of h. Therefore, since $0 \le \omega = 2h - \widetilde{h} < 1$, we have

$$\ell_k(\widetilde{h}) - u_k(\widetilde{h}) = \ell_k(2h) - u_k(2h) - \ell_k(\omega) + u_k(\omega) \ge \frac{4}{\delta} - \ell_k(\omega) \ge \frac{2}{\delta},$$

which completes the proof. □

The following lemma, which plays a vital role in proving the main theorem, roughly indicates how close the central path is to the boundary of the KM cube.

Lemma 6.2.3. *Let $d \in \mathscr{D}$ and $h \in \mathbb{Z}_{\oplus}^m$ be chosen such that they satisfy conditions* (6.2.1)*. Then, for problem* (6.1.1)*, we have*

$$\mathcal{C}_*(\tau, d, h) \cap C_\delta^{k+1} \subseteq \widehat{C}_\delta^k \quad \text{for } k = 1, \ldots, m-1.$$

Proof. For $k = 1, \ldots, m-1$, we show that every point on the central path $\mathcal{C}_*(\tau, d, h)$ that satisfies $\widehat{s}_{k+1} > \delta$ and $\bar{s}_{k+1} > \delta$ also satisfies

$$\bar{s}_k \leq \delta, \widehat{s}_{k-1} \leq \delta, \ldots, \widehat{s}_1 \leq \delta.$$

Recall that every point on the central path $\mathcal{C}_*(\tau, d, h)$ satisfies the first $m-1$ equations of system (6.1.2). From the k^{th} equation of system (6.1.2) multiplied by τ^{k-1}, we have

$$\frac{h_k \tau^{k-1}}{d_k + y_k} = -\frac{\tau^{k-1}}{\widehat{s}_k} + \frac{\tau^{k-1}}{\bar{s}_k} + \frac{\tau^k}{\widehat{s}_{k+1}} + \frac{\tau^k}{\bar{s}_{k+1}} \leq \frac{\tau^{k-1}}{\bar{s}_k} + \frac{2\tau^k}{\delta},$$

which implies that

$$\ell_k = \frac{h_k \tau^{k-1}}{d_k + 1} \leq \frac{1}{\bar{s}_k} + \frac{1}{\delta}.$$

Thus, using inequalities (6.2.1), we get $\bar{s}_k \leq \delta$. For $j = 1, \ldots, k-1$, adding the k^{th} equation of system (6.1.2) multiplied by τ^{k-1} to the i^{th} equation of system (6.1.2) multiplied by $-\tau^{i-1}$ for all $i = j, \ldots, k-1$, we have

$$\frac{h_k \tau^{k-1}}{d_k + y_k} - \sum_{i=j}^{k-1} \frac{h_i \tau^{i-1}}{d_i + y_i} \leq \frac{\tau^{j-1}}{\widehat{s}_j} + \frac{2\tau^k}{\delta},$$

which implies that

$$\ell_k - u_k \leq \frac{h_k \tau^{k-1}}{d_k + 1} - \sum_{i=j}^{k-1} \frac{h_i \tau^{i-1}}{d_i} \leq \frac{1}{\widehat{s}_j} + \frac{1}{\delta}.$$

Thus, using inequalities (6.2.1), we get $\widehat{s}_j \leq \delta$. □

We now provide the main result of this chapter, which consists of the following two statements.

1. The analytic center $\chi^*(\tau, d, h)$ of $\mathcal{KM}_*(\tau, d, h)$ belongs to $\widehat{C}_\delta^m$, a δ-neighborhood of the vertex v^0.

2. The central path $\mathcal{C}_*(\tau, d, h)$ of problem (6.1.1) stays in P_δ, a δ-neighborhood of the simplex path.

Theorem 6.2.4. *Assume that $d \in \mathscr{D}$ and $h \in \mathbb{Z}_\oplus^m$ satisfy conditions (6.2.1). Then, for problem (6.1.1), we have*

1. $\chi^*(\tau, d, h) \in \widehat{C}_\delta^m$.
2. $\mathcal{C}_*(\tau, d, h) \subseteq P_\delta$.

Proof. The second statement follows from Lemma 6.2.3, with a similar argument as in the proof of Theorem 4.2.4. To prove the first statement, we show that the unique solution $\chi^*(\tau, d, h)$ of system (6.1.2) belongs to $\widehat{C}_\delta^m$, or equivalently, it satisfies

$$\bar{s}_m \leq \delta, \widehat{s}_{m-1} \leq \delta, \ldots, \widehat{s}_1 \leq \delta.$$

Obviously, we have

$$d_i \leq d_i + \widehat{s}_i \leq d_i + 1 \quad \text{for } i = 1, \ldots, m.$$

From the m^{th} equation of system (6.1.2) multiplied by τ^{m-1}, we have

$$\frac{h_m \tau^{m-1}}{d_m + y_m} = \frac{\tau^{m-1}}{\bar{s}_m} - \frac{\tau^{m-1}}{\widehat{s}_m} \leq \frac{\tau^{m-1}}{\bar{s}_m},$$

which implies that

$$\ell_m = \frac{h_m \tau^{m-1}}{d_m + 1} \leq \frac{1}{\bar{s}_m}.$$

Thus, using inequalities (6.2.1), we get $\bar{s}_m \leq \delta$. For $j = 1, \ldots, m-1$, adding the m^{th} equation of system (6.1.2) multiplied by τ^{m-1} to the i^{th} equation of system (6.1.2) multiplied by $-\tau^{i-1}$ for all $i = j, \ldots, m-1$, we have

$$\frac{h_m \tau^{m-1}}{d_m + y_m} - \sum_{i=j}^{m-1} \frac{h_i \tau^{i-1}}{d_i + y_i} = \frac{\tau^{j-1}}{\widehat{s}_j} - \frac{\tau^{j-1}}{\bar{s}_j} - \sum_{i=j+1}^{m} \frac{2\tau^{i-1}}{\bar{s}_i} \leq \frac{\tau^{j-1}}{\widehat{s}_j},$$

which implies that

$$\ell_m - u_m \leq \frac{h_m \tau^{m-1}}{d_m + 1} - \sum_{i=j}^{m-1} \frac{h_i \tau^{i-1}}{d_i} \leq \frac{1}{s_j}.$$

Thus, using inequalities (6.2.1), we get $\widehat{s}_j \leq \delta$. □

6.3 Choosing a Proper Distance Vector

In this section, we describe how to choose a distance vector d of problem (6.1.1) so that the number of redundant constraints n is minimized. First, we assume that all the distances are chosen to be equal. Then, we allow the distances to vary and show that to minimize n, they have to decay geometrically.

6.3.1 Uniform Distances

Let d_0 denote the uniform distances of the redundant constraints to the corresponding hyperplanes. For $k = 1, \dots, m$, from (6.2.2), we have

$$h_k = \frac{2(d_0+1)}{\tau^{k-1}\delta}\left(2 + \frac{1}{d_0}\right)^{k-1}.$$

Since $2 + 1/d_0 \leq 2^{1+1/d_0}$, we have

$$h_k \leq \frac{d_0+1}{\tau^{k-1}\delta} 2^{k+\frac{k-1}{d_0}}.$$

Hence, minimizing the right-hand side of this inequality suggests choosing $d_0 = m/2$ that yields an upper bound $\mathcal{O}\big(\frac{m2^k}{\tau^k\delta}\big)$ for h_k. Now, consider the redundant KM example (6.1.1) with the data given as

$$\begin{cases} \tau = \dfrac{m}{2(m+1)}, \\ \delta = \dfrac{1}{4(m+1)}, \\ d_0 = \dfrac{m}{2}, \\ h_k = m(m+2)\left(1+\dfrac{1}{m}\right)^{2k-1} 2^{2k} \quad \text{for } k = 1, \dots, m. \end{cases}$$

On can verify, using the inequality $\big(1 + 1/m\big)^m \leq 2^2$, that

$$n = 2m + \sum_{k=1}^{m} h_k \leq 2m + \sum_{k=1}^{m} m(m+2)2^{2k+4} = \mathcal{O}(m^2 2^{2m}).$$

In Chapter 7, using a different approach, we show that $d_0 = 0$ is also possible. In this case, we need to add a significantly larger number of redundant constraints to the KM cube to force the central path to follow the simplex path.

6.3.2 Geometrically Decaying Distances

Let us denote the number of redundant inequalities at a given distance vector d by $n(d)$. We seek for

$$\widehat{d} = \arg\min_{d\in\mathscr{D}} n(d) = \arg\min_{d\in\mathscr{D}} \sum_{k=1}^{m} \frac{d_k+1}{\tau^{k-1}} \prod_{i=1}^{k-1}\left(2+\frac{1}{d_i}\right) \tag{6.3.1}$$

The following lemma provides a solution for this optimization problem.

Lemma 6.3.1. *The vector* $\widehat{d}$ *whose components are given by the recursive formula*

$$\begin{cases} \widehat{d}_m = 0, \\ \tau \widehat{d}_k^2 = \widehat{d}_{k+1}^2 + (1+\widehat{d}_{k+1})^2 \quad \text{for } k = 1,\dots,m-1, \end{cases} \tag{6.3.2}$$

is the unique solution of problem (6.3.1).

Proof. Let us define

$$\varphi(d) = \sum_{k=1}^{m} \frac{\pi_{1:k-1}(d)}{\tau^{k-1}}(d_k+1) \quad \text{where } \pi_{1:k-1}(d) = \prod_{i=1}^{k-1}\left(2+\frac{1}{d_i}\right).$$

The function φ is linear with respect to d_m with a positive slope. Therefore, its minimum with respect to d_m happens at $d_m = 0$. We now redefine the problem by reducing the dimension as follows. Let $\bar{d} = d_{1:m-1}$ and define a new differentiable function $\bar{\varphi} : \mathbb{R}_+^{m-1} \to \mathbb{R}$ as

$$\bar{\varphi}(\bar{d}) = \varphi(d)|_{d_m=0} = \sum_{k=1}^{m-1} \frac{\pi_{1:k-1}(\bar{d})}{\tau^{k-1}}(\bar{d}_k+1) + \frac{\pi_{1:m-1}(\bar{d})}{\tau^{m-1}}.$$

For $k = 1,\dots,m$, the k^{th} component of the gradient of $\bar{\varphi}$ is

$$\frac{\partial \bar{\varphi}}{\partial \bar{d}_k} = \frac{\pi_{1:k-1}(\bar{d})}{\tau^{k-1}} \left[1 - \frac{1}{\bar{d}_k^2}\left(\sum_{i=k+1}^{m-1} \frac{\pi_{k+1:i-1}(\bar{d})}{\tau^{i-k}}(\bar{d}_i+1) + \frac{\pi_{k+1:m-1}(\bar{d})}{\tau^{m-k}}\right)\right].$$

Therefore, $\bar{\varphi}$ has a unique stationary point $\widetilde{d}$, i.e., $\nabla\bar{\varphi}(\widetilde{d}) = 0$, given by

$$\widetilde{d} = \left(\sum_{i=k+1}^{m-1} \frac{\pi_{k+1:i-1}(\widetilde{d})}{\tau^{i-k}}(\widetilde{d}_i+1) + \frac{\pi_{k+1:m-1}(\widetilde{d})}{\tau^{m-k}}\right)_{1:m-1}^{\frac{1}{2}}, \tag{6.3.3}$$

We claim that $\widetilde{d}$ is the global minimizer of the function $\bar{\varphi}$ over $\mathbb{R}^{m-1}_{+}$. Let us consider a generalized function $\widetilde{\varphi} : \mathbb{R}^{m-1} \to \mathbb{R}$ defined as

$$\widetilde{\varphi}(\bar{d}) = \begin{cases} \bar{\varphi}(\bar{d}) & \text{if } \bar{d} \in \mathbb{R}^{m-1}_{+}, \\ +\infty & \text{otherwise .} \end{cases}$$

Since $\widetilde{\varphi}$ is a closed proper coercive function, by Weierstrass' theorem [6], it has a minimum in $\mathbb{R}^{m-1}$. Since $\widetilde{d}$ is the unique stationary point of $\widetilde{\varphi}$, it is the global minimizer of $\widetilde{\varphi}$, thus, the global minimizer of $\bar{\varphi}$ over $\mathbb{R}^{m-1}_{+}$. Now, we prove that the unique minimizer of $\bar{\varphi}$, given by (6.3.3), satisfies the recursive formula (6.3.2). Since

$$\begin{aligned}
\widetilde{d}^2_{k+1} &= \sum_{i=k+2}^{m-1} \frac{\pi_{k+2:i-1}(\widetilde{d})}{\tau^{i-k-1}}(\widetilde{d}_i + 1) + \frac{\pi_{k+2:m-1}(\widetilde{d})}{\tau^{m-k-1}}, \\
&= \frac{\tau}{2 + \frac{1}{\widetilde{d}_{k+1}}} \left(\sum_{i=k+2}^{m-1} \frac{\pi_{k+1:i-1}(\widetilde{d})}{\tau^{i-k}}(\widetilde{d}_i + 1) + \frac{\pi_{k+1:m-1}(\widetilde{d})}{\tau^{m-k}} \right), \\
&= \frac{\tau \widetilde{d}_{k+1}}{2\widetilde{d}_{k+1} + 1} \left(\widetilde{d}^2_k - \frac{\widetilde{d}_{k+1} + 1}{\tau} \right),
\end{aligned}$$

the result immediately follows. □

Neglecting lower-order terms in the recursive formula (6.3.2), we get $\tau \widehat{d}^2_k = 2\widehat{d}^2_{k+1}$, which results to

$$\widehat{d}_k = \frac{1}{\sqrt{2}} \left(\frac{2}{\tau} \right)^{\frac{m-k}{2}} \qquad \text{for } k = 1, \ldots, m-1.$$

Let (RP^m_3) denote problem (6.1.1) with the data given as

$$\begin{cases}
\tau = \dfrac{m}{2(m+1)}, \\
\delta = \dfrac{1}{4(m+1)}, \\
d_k = 2^{m-k} - 1 & \text{for } k = 1, \ldots, m, \\
h_k = \left\lfloor m\left(1 + \dfrac{1}{m}\right)^k 2^{m+3} \displaystyle\prod_{i=1}^{k-1} \left(2 + \dfrac{1}{2^{m-i} - 1}\right) \right\rfloor & \text{for } k = 1, \ldots, m.
\end{cases}$$

Note that since $\tau = \frac{m}{2(m+1)}$, we have

$$d_k = \frac{1}{\sqrt{2}}\left(\frac{2}{\tau}\right)^{\frac{m-k}{2}} = \mathcal{O}\left(2^{m-k}\right) \quad \text{for } k = 1, \ldots, m-1.$$

Therefore, for simplicity, we have chosen $d_k = 2^{m-k} - 1$ for $k = 1, \ldots, m$ which also satisfies $d_m = 0$.

6.4 The Iteration Complexity $\mathcal{O}(\sqrt{n}L)$ Cannot Be Improved

In this section, using problem (RP_3^m), we provide our tightest lower and upper bounds for the iteration complexity of central-path-following IPMs. We first provide a range for μ in which μ-centers belong to the set $\widehat{\mathcal{C}}_\delta^m$.

Lemma 6.4.1. *Let $\mu \geq \tau^{m-1}\delta$ and $y(\mu)$ be the μ-center of problem* (RP_3^m). *Then, we have $y(\mu) \in \widehat{\mathcal{C}}_\delta^m$.*

Proof. We skip the proof as it is similar to that of Lemma 5.3.2. □

The following theorem gives the tightest iteration complexity bounds for central-path-following IPMs.

Theorem 6.4.2. *The iteration complexity lower and upper bounds for problem* (RP_3^m) *are $\Omega(2^m)$ and $\mathcal{O}\left(m^{1.5}2^m\right)$.*

Proof. We leave the proof to the reader as it is similar to that of Theorem 5.3.3. We only remark that extending the definition of d_k in formula (6.3.2) to d_0, we have $n = 2m + e^T h \leq 2m + 2\tau d_0^2/\delta = \mathcal{O}\left(m2^{2m}\right)$. □

Corollary 6.4.3. *The iteration complexity bounds for problem* (RP_3^m) *are $\Omega\left(\sqrt{n/\log n}\right)$ and $\mathcal{O}\left(\sqrt{n}\log n\right)$.*

6.5 Summary

1. Unlike previous variants of the redundant KM examples, in which the redundant constraints are put parallel to the facets containing the origin, in the KM problem (RP_3^m), we placed the redundant constraints

parallel to the coordinate hyperplanes. We considered both uniform and geometrically decaying distances.

2. It turned out that, in the constructions of this chapter, fewer number of redundant constraints are enough to force the central path to follow the simplex path and visit all the vertices of the KM cube. The reduction in the number of redundant constraints resulted in achieving the tightest complexity bounds. For problem (RP_3^m), we showed that the iteration complexity bounds are $\Omega\big(\sqrt{n/\log n}\big)$ and $\mathcal{O}\big(\sqrt{n}\log n\big)$.

Chapter 7

All Redundant Constraints Touch the Klee-Minty Cube

In this chapter, we use the same construction as the one described in Chapter 6 with all the distances set to zero. In other words, we consider the redundant KM problem in which the coordinate hyperplanes are repeated exponentially many times. In this case, a significantly larger number of redundant constraints are required to bend the central path along the simplex path.

7.1 Introduction

We consider the following variant of the redundant KM example in which the redundant constraints are repetitions of the coordinate hyperplanes.

$$\begin{array}{lll} \min & y_m & \\ \text{s.t.} & \tau y_{k-1} \leq y_k \leq 1 - \tau y_{k-1} & \text{for } k = 1, \ldots, m, \\ & 0 \leq y_k \quad \text{repeated } h_k \text{ times for } k = 1, \ldots, m, & \end{array} \tag{7.1.1}$$

with the convention $y_0 = 0$. We denote by $\mathcal{KM}(\tau, h)$ the feasible set represented by the constraints of problem (7.1.1), where $h = (h_1; \cdots; h_m)$. Note that the sets $\mathcal{KM}(\tau)$ and $\mathcal{KM}(\tau, h)$ are geometrically identical for all $\tau > 0$ and h, while they differ in the algebraic representation. We seek for an integer-valued vector h by which the central path of problem (7.1.1) is bent along the simplex path visiting a small neighborhood of each vertex of the KM cube.

The logarithmic barrier function for problem (7.1.1) is defined as

$$F_3(y) = -e^T \log \widehat{s} - e^T \log \bar{s} - h^T \log y.$$

where the slack vectors $\widehat{s}_k$ and $\bar{s}_k$ are defined as (4.1.2). By Definition 1.3.4, the analytic center of $\mathcal{KM}_3(\tau, h)$ is then given as

$$\chi^3(\tau, h) = \arg\min \{F_3(y) \mid (\widehat{s}; \bar{s}; y) > 0\}.$$

Therefore, the analytic center $\chi^3(\tau, h)$ is the unique solution of the necessary and sufficient optimality conditions

$$\begin{cases} \frac{1}{\widehat{s}_k} - \frac{1}{\bar{s}_k} - \frac{\tau}{\widehat{s}_{k+1}} - \frac{\tau}{\bar{s}_{k+1}} + \frac{h_k}{y_k} - \frac{h_{k+1}\tau}{y_{k+1}} = 0, \quad k = 1, \ldots, m-1, \\ \frac{1}{\widehat{s}_m} - \frac{1}{\bar{s}_m} + \frac{h_m}{y_m} = 0, \\ \widehat{s}_k > 0, \bar{s}_k > 0, y_k > 0, \quad k = 1, \ldots, m. \end{cases} \tag{7.1.2}$$

By Definition 1.3.3, the central path of problem (7.1.1) is the set

$$\mathcal{C}_3(\tau, h) = \{y(\mu) \mid \mu > 0\},$$

in which the function $y(\mu)$ is defined as

$$y(\mu) = \arg\min \{y_m + \mu F_3(y) \mid (\widehat{s}; \bar{s}; y) > 0\}.$$

Therefore, from the optimality conditions, every $y(\mu)$ satisfies the first $m-1$ equations of system (7.1.2).

7.2 A Zero-Distance Redundant KM Problem

In this section, we assume that $m \geq 2$, $\tau > 0$, and $\delta > 0$ with $\tau + \delta < 1/2$. In the following lemmas, we show how to choose $h \in \mathbb{Z}^m_{\oplus}$ such that it satisfies

$$h_k \tau^{k-1} \delta \geq 1 + 2\tau^k + \sum_{i=1}^{k-1} h_i \quad \text{for } k = 1, \ldots, m. \tag{7.2.1}$$

Lemma 7.2.1. *For any $k \in \mathbb{Z}_+$, we have*

$$1 + 2\tau + \sum_{i=1}^{k} \frac{i}{\tau^{i(i-1)/2}\delta^i} \leq \frac{k+1}{\tau^{k(k-1)/2}\delta^k}.$$

Proof. The proof is by induction on k. The statement is obviously true for $k = 1$ as $\tau + \delta < 1/2$. Assuming that the inequality holds for all $k = 1, \ldots, j$, we have

$$\begin{aligned} 1 + 2\tau + \sum_{i=1}^{j+1} \frac{i}{\tau^{i(i-1)/2}\delta^i} &\leq \frac{j+1}{\tau^{j(j-1)/2}\delta^j} + \frac{j+1}{\tau^{j(j+1)/2}\delta^{j+1}} \\ &= \frac{(j+1)(1+\tau^j\delta)}{\tau^{j(j+1)/2}\delta^{j+1}} \\ &\leq \frac{j+2}{\tau^{j(j+1)/2}\delta^{j+1}}. \end{aligned}$$

Therefore, the inequality holds for $j = k+1$, so the proof is completed. □

The following lemma provides a vector $h \in \mathbb{R}^m_\oplus$ that satisfies (7.2.1).

Lemma 7.2.2. *The vector* $h = \left(\frac{1}{\delta}; \frac{2}{\tau\delta^2}; \cdots; \frac{m}{\tau^{m(m-1)/2}\delta^m}\right)$ *satisfies* (7.2.1).

Proof. Obviously $h_1 \geq 1/\delta$. Therefore, it suffices to show, for $k = 2, \ldots, m$, that

$$\frac{k\tau^{k-1}\delta}{\tau^{k(k-1)/2}\delta^k} \geq 1 + 2\tau + \sum_{i=1}^{k-1} \frac{i}{\tau^{i(i-1)/2}\delta^i}.$$

This inequality is equivalent to

$$\frac{k}{\tau^{(k-1)(k-2)/2}\delta^{k-1}} \geq 1 + 2\tau + \sum_{i=1}^{k-1} \frac{i}{\tau^{i(i-1)/2}\delta^i},$$

which hold as shown in Lemma 7.2.1. □

Next lemma shows the existance of an integer-valued vector $h \in \mathbb{Z}^m_\oplus$ that satisfies (7.2.1).

Lemma 7.2.3. *If* h *is a solution of* (7.2.1)*, then so is the integer-valued vector* $\widetilde{h} = \lfloor 2h \rfloor$.

Proof. We have $0 \leq \omega = 2h - \widetilde{h} < 1$. Multiplying each side of (7.2.1) by 2 and substituting $2h_k$ by $\widetilde{h}_k + \omega_k$ for all $k = 1, \ldots, m$, we get

$$\widetilde{h}_k \tau^{k-1} + \omega_k \tau^{k-1}\delta \geq 2 + 4\tau^k + \sum_{i=1}^{k-1} \widetilde{h}_i + \sum_{i=1}^{k-1} \omega_i,$$

implying that

$$\begin{aligned}\widetilde{h}_k\tau^{k-1}\delta &\geq 2+4\tau^k-\omega_k\tau^{k-1}\delta+\sum_{i=1}^{k-1}\widetilde{h}_i+\sum_{i=1}^{k-1}\omega_i\\ &\geq 2+4\tau^k-\delta+\sum_{i=1}^{k-1}\widetilde{h}_i\\ &\geq 1+2\tau^k+\sum_{i=1}^{k-1}\widetilde{h}_i.\end{aligned}$$

Therefore, the integer-valued vector $\widetilde{h}$ satisfies (7.2.1). □

In the following lemma, we prove some implications of inequalities (7.2.1).

Lemma 7.2.4. *For any $k = 2,\dots,m$, the k^{th} inequality of* (7.2.1) *implies that*

$$h_k\tau^{k-1}\delta \geq 2\tau^k+\tau^{j-1}\left(1+\sum_{i=j}^{k-1}h_i\right) \qquad \text{for } j=1,\dots,k-1.$$

Proof. The proof immediately follows as $\tau^{j-1} < 1$ and $h_i > 0$ for $i = 1,\dots,j-1$. □

The following lemma, which plays a vital role in proving the main theorem, roughly indicates how close the central path is to the boundary of the KM cube.

Lemma 7.2.5. *Let $h \in \mathbb{Z}^m_{\oplus}$ satisfy conditions* (7.2.1). *Then, for problem* (7.1.1), *we have*

$$\mathcal{C}(\tau,h)\cap C^{k+1}_{\delta} \subseteq \widehat{C}^k_{\delta} \qquad \text{for } k=1,\dots,m-1.$$

Proof. For $k = 1,\dots,m-1$, we show that every point on the central path $\mathcal{C}(\tau,h)$ that satisfies $\widehat{s}_{k+1} > \delta$ and $\bar{s}_{k+1} > \delta$ also satisfies

$$\bar{s}_k \leq \delta, \widehat{s}_{k-1} \leq \delta, \dots, \widehat{s}_1 \leq \delta.$$

Recall that every point on the central path $\mathcal{C}(\tau,h)$ satisfies the first $m-1$ equations of system (7.1.2). From the k^{th} equation of system (7.1.2), we have

$$\frac{h_k}{y_k} = -\frac{1}{\widehat{s}_k}+\frac{1}{\bar{s}_k}+\frac{\tau}{\widehat{s}_{k+1}}+\frac{\tau}{\bar{s}_{k+1}} \leq \frac{1}{\bar{s}_k}+\frac{2\tau}{\delta}.$$

Since $h_k \geq (1+2\tau)/\delta$ from (7.2.1), we get $\bar{s}_k \leq \delta$. For $j = 1, \ldots, k-1$, adding the k^{th} equation of system (7.1.2) multiplied by τ^{k-1} to the i^{th} equation of system (7.1.2) multiplied by $-\tau^{i-1}$ for all $i = j, \ldots, k-1$, we have

$$\frac{h_k \tau^{k-1}}{y_k} - \sum_{i=j}^{k-1} \frac{h_i \tau^{i-1}}{y_i} \leq \frac{\tau^{j-1}}{\widehat{s}_j} + \frac{2\tau^k}{\delta}.$$

Using the facts that $y_i \geq \tau^{i-j} y_j$ for any $i = j, \ldots, m$ and $y_j \geq s_j$, we obtain

$$h_k \tau^{k-1} - \sum_{i=j}^{k-1} \frac{h_i \tau^{j-1}}{s_j} \leq \frac{\tau^{j-1}}{\widehat{s}_j} + \frac{2\tau^k}{\delta},$$

or equivalently

$$h_k \tau^{k-1} \leq \left(1 + \sum_{i=j}^{k-1} h_k\right) \frac{\tau^{j-1}}{s_j} + \frac{2\tau^k}{\delta}.$$

Therefore, from the inequalities of Lemma 7.2.4, we get $s_j \leq \delta$. □

We now provide the main result of this chapter, which consists of the following two statements.

1. The analytic center $\chi(\tau, h)$ of $\mathcal{KM}(\tau, h)$ belongs to $\widehat{C}^m_\delta$, a δ-neighborhood of the vertex v^0.
2. The central path $\mathcal{C}(\tau, h)$ of problem (7.1.1) stays in P_δ, a δ-neighborhood of the simplex path.

Theorem 7.2.6. *Assume that $h \in \mathbb{Z}^m_\oplus$ satisfies conditions* (7.2.1)*. Then, for problem* (7.1.1)*, we have*

1. $\chi(\tau, h) \in \widehat{C}^m_\delta$.
2. $\mathcal{C}(\tau, h) \subseteq P_\delta$.

Proof. The second statement follows from Lemma 7.2.5, with a similar argument as in the proof of Theorem 4.2.4. To prove the first statement, we show that the unique solution $\chi(\tau, h)$ of system (7.1.2) belongs to $\widehat{C}^m_\delta$, or equivalently, it satisfies

$$\bar{s}_m \leq \delta, \widehat{s}_{m-1} \leq \delta, \ldots, \widehat{s}_1 \leq \delta.$$

From the m^{th} equation of system (7.1.2), we have

$$\frac{h_m}{y_m} = \frac{1}{\bar{s}_m} - \frac{1}{\widehat{s}_m} \leq \frac{1}{\bar{s}_m},$$

which implies that

$$h_m \leq \frac{1}{\bar{s}_m}.$$

Since $h_m \geq 1/\delta$ from inequalities (7.2.1), we get $\bar{s}_m \leq \delta$. For every $j = 1,\ldots,m-1$, adding the m^{th} equation of system (7.1.2) multiplied by τ^{m-1} to the i^{th} equation of system (7.1.2) multiplied by $-\tau^{i-1}$ for all $i = j,\ldots,m-1$, we have

$$\frac{h_m\tau^{m-1}}{y_m} - \sum_{i=j}^{m-1}\frac{h_i\tau^{i-1}}{y_i} = \frac{\tau^{j-1}}{\widehat{s}_j} - \frac{\tau^{j-1}}{\bar{s}_j} - \sum_{i=j+1}^{m}\frac{2\tau^{i-1}}{\bar{s}_i} \leq \frac{\tau^{j-1}}{\widehat{s}_j}.$$

Using the facts that $y_i \geq \bar{\tau}^{i-j}y_j$ for any $i = j,\ldots,m$ and $y_j \geq s_j$, we obtain

$$h_m\tau^{m-1} - \sum_{i=j}^{m-1}\frac{h_i\tau^{j-1}}{s_j} \leq \frac{\tau^{j-1}}{s_j} + \frac{2\tau^m}{\delta},$$

or equivalently

$$h_m\tau^{m-1} \leq \left(1 + \sum_{i=j}^{m-1} h_m\right)\frac{\tau^{j-1}}{s_j} + \frac{2\tau^m}{\delta},$$

which implies, using the inequalities of Lemma 7.2.4, that $s_j \leq \bar{\delta}$. □

Now, let us consider problem (7.1.1) with the data given as

$$\begin{cases} \tau = \dfrac{m}{2(m+1)}, \\ \delta = \dfrac{1}{4(m+1)}, \\ h_k = \left\lfloor \dfrac{2k}{\tau^{k(k-1)/2}\delta^k} \right\rfloor \quad \text{for } k = 1,\ldots,m. \end{cases}$$

Hence, the number of the inequalities for this problem is

$$\begin{aligned} n &= 2m + e^T h, \\ &\leq 2m + \sum_{k=1}^{m} 2k\left(\frac{2m+2}{m}\right)^{k(k-1)/2}(4m+4)^k, \\ &\leq 2m + 2m(m+1)^m 2^{2m}\left(1+\frac{1}{m}\right)^{m(m-1)/2}\sum_{k=1}^{m} 2^{k(k-1)/2}. \end{aligned}$$

Since $(1+1/m)^m \leq 2^2$ and $\sum_{k=1}^{m} 2^{k(k-1)/2} \leq 2^{m^2/2}$, we get

$$n \leq 2m + (m+1)^{m+1} 2^{\frac{m^2}{2}+3m},$$

implying that $n = O\big(2^{m^2}\big)$. As seen, in this case, we need to add a significantly larger number of redundant constraints to force the central path to closely follow the simplex path.

7.3 Summary

1. In this chapter, we introduced a redundant KM problem whose redundant constraints are repetitions of the coordinate hyperplanes.
2. The number of redundant constraints in the previous KM constructions are $n = \mathcal{O}(m^2 2^{3m})$ and $n = \mathcal{O}(m^k 2^{2m})$ with $k = 1, 2, 3$. For the construction of this chapter, where all the redundant constraints touch the feasible region, the number of redundant constraints is required to be $n = O\big(2^{m^2}\big)$, which is remarkably larger than the number of redundant constraints in the previous constructions.

Chapter 8

Conclusions and Future Work

In this chapter, we first summarize the contents of the previous chapters and then outline some directions for future work for the interested reader.

8.1 Conclusions

In Chapter 1, we have started with an introduction to LPs as well as the simplex method. Then, we have briefly discussed some variants of IPMs, including affine-scaling, potential-reduction, and central-path-following methods. Unlike the simplex method that solves an LP by moving on the boundary of the feasible region from one vertex to another, IPMs pass through the interior of the feasible region toward optimality. We have defined the analytic center and the central path as key settings of the theory of central-path-following IPMs. Furthermore, we have provided pseudo-codes for the small-update, large-update, and predictor-corrector variants of central-path-following IPMs as well as their iteration complexity results.

In Chapter 2, we have observed that the central path is an analytic curve (infinitely many times differentiable) that converges to the analytic center of the optimal set. IPMs terminate at an interior point with the duality gap smaller than a given precision. To obtain an exact optimal solution, we may employ a rounding procedure which is a finite algorithm. We have discussed Ye's [50] and Mehrotra and Ye's [29] rounding procedures. In addition, we have shown that the central path depends on the representation of the feasible region and is highly affected by redundancy.

In Chapter 3, we have reviewed some pathological examples presented in optimization. We have learnt that variants of the simplex method may take an exponential number of pivot steps to reach an optimal solution of an LP. In particular, we have addressed the following results.

We have observed that the simplex method with Dantzig's pivot rule [10] may take an exponential number of pivot steps on the Klee-Minty (KM) problem [23]. In contrast to Dantzig's pivot rule [10], Bland's pivot rule [8] is proven to be finite, i.e., no cycle may happen during the procedure of finding an optimal solution. Using a variant of the KM problem, Avis and Chvátal [4] have shown that the simplex method with Bland's pivot rule is also exponential in the worst case. Roos [37] has shown that Terlaky's [42] finite criss-cross method may also take an exponential number of steps to find the optimal solution of the KM example. Variants of the KM problem have been introduced for almost all pivot rules, except Zadeh's [53] pivot rule whose complexity has not been clarified yet [44]. Murty has presented exponential examples [31, 32] for LCPs and parametric LPs. Megiddo and Shub [27] have observed that a trajectory of the vector field induced by the affine-scaling directions may trace the simplex path of the KM cube. As a consequence, affine-scaling methods do not seem to be "good" algorithms for solving LPs.

Next, we have reviewed some examples [14, 17] for linear and nonlinear convex optimization problems that exhibit ill-behaved central paths. We have noticed that the central path may have an "antenna-like" or "zigzagging" shape with almost-straight line segments of fixed lengths with infinite variations.

In Chapters 4, we have presented a redundant KM construction in which an exponential number of redundant constraints are added parallel to the facets of the KM cube at uniform distances. Then, we have shown that the central path of this redundant problem arbitrarily closely follows the simplex path (the edge-path followed by the simplex method) visiting all 2^m vertices of the KM cube. Moreover, we have proven that the construction yields a lower bound for the number of iterations.

In Chapters 5, we have provided another redundant KM problem in which the distances of the redundant constraints to the corresponding facets are allowed to decay geometrically. In this construction, we have significantly reduced the number of redundant constraints. This reduction has led to tighter iteration complexity lower and upper bounds. Furthermore, we have customized the termination criterion by exploiting the special structure of the KM cube that has resulted to further tightening of the gap between the iteration complexity bounds.

In Chapters 6, we have refined the previous redundant KM constructions by placing the redundant constraints parallel to the coordinate hyperplanes. We have shown that the best iteration complexity bounds for this construction are achieved if the distances are chosen to decay geometrically. This construction has emerged to provide the tightest iteration complexity bounds and, as a consequence, has yielded a nearly worst-case example for central-path-following IPMs.

In Chapters 7, by providing a different analysis, we have proven that the distances can be set to zero. In other words, the redundant constraints are simply the repetitions of the nonnegativity constraints. We have seen that we need a significantly larger number of redundant constraints to bend the central path along the simplex path.

8.2 Future Work

In this section, we outline some directions for future work by explaining some interesting open problems, each of which can be a fruitful research project in the near future.

Generalizing the Redundant KM Construction

We have shown, by utilizing the dependence of the central path on redundancy, that the central path of the Klee-Minty problem can be forced to follow the simplex path arbitrarily closely. We believe that this result can be generalized to every LP. In other words, we believe that by adding redundant constraints to an LP, we can bend the central path along the simplex path, or at least some of the simplex paths, of the LP. This expectation is supported by the following argument.

Consider the primal problem as given by (1.1.1). Assume that $\mathcal{P}_{\oplus}$ is nonempty and bounded. We define vectors $c^1, \ldots, c^n$ as follows. Starting from $c^n = c$, we use the Gram-Schmidt orthogonalization process to compute vectors c^i, for $i = 1, \ldots, n-1$. Now, for every $i = 1, \ldots, n$, let $\underline{x}^i$ be an optimal solution of the problem

$$\begin{aligned} \min \quad & (c^i)^T x \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0. \end{aligned}$$

We define a redundant problem as

$$\begin{array}{lll} \min & c^T x & \\ \text{s.t.} & Ax = b, & \\ & x \geq 0, & \\ & (c^i)^T x \geq (c^i)^T \underline{x}^i - d_i & \text{repeated } h_i \text{ times for } i = 1, \ldots, n, \end{array}$$

We conjecture that the central path of this redundant problem, for some repetition vector $h = (h_1; \cdots; h_n)$ and some distance vector $d = (d_1; \cdots; d_n)$, closely follows a simplex path.

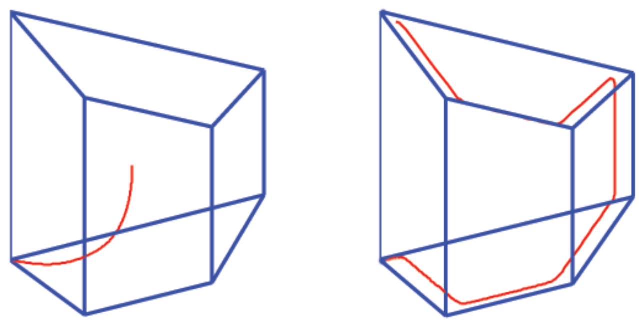
Discrete Algorithms on the Redundant KM Problems

As we have previously seen, the central path may have $2^m - 2$ sharp turns in dimension m with $2^m - 1$ almost-straight line segments of a fixed length. Therefore, central-path-following IPMs that operate in a small neighborhood must closely follow the central path and thus must have at least $2^m - 1$ iterations. However, in practice, we usually use a large neighborhood of the central path. It is an interesting research direction to investigate whether or not the number of iterations of central-path-following IPMs that operate in a large neighborhood of the central path is bounded below by $\Omega(2^m)$. Our experiments in lower dimensions indicate that the result is true for small-update, large-update, and Mizuno-Todd-Ye's predictor-corrector variants.

Another interesting problem is to construct similar worst-case models for other IPMs, such as primal-dual affine-scaling and potential-reduction methods. As we mentioned in Chapter 3, a trajectory of the affine-scaling vector field may visit all the vertices of the KM cube. However, the behavior of the primal-dual affine-scaling and potential-reduction methods in the worst case is still an open problem.

Worst-Case Examples for Other Path-Following IPMs

The redundant Klee-Minty example introduced in Chapter 6 is a worst-case example for central-path-following IPMs. The central path followed by these algorithms is defined using the standard barrier function as discussed in Chapter 1. Using different barrier functions, such as the universal barrier [36] or the volumetric barrier [48], we can define different paths which also lead to an



(a) Without redundancy (b) With redundancy

Figure 8.1: Volumetric path of the Klee-Minty problem in 3D.

optimal solution. Constructing worst-case examples for these paths is an interesting research topic. In the following, we give some illustrations to convince the reader that redundancy can also be used to design worst-case examples for volumetric-path-following IPMs.

Consider the dual problem (1.1.2). Recalling from Chapter 1, we define the dual standard barrier function for the dual problem as $F_{\mathcal{D}}(y) = e^T \log s$, where $s = c - A^T y$, whose gradient and Hessian are given by

$$\nabla F_{\mathcal{D}}(y) = -AS^{-1}e,$$
$$\nabla^2 F_{\mathcal{D}}(y) = AS^{-2}A^T.$$

The volumetric barrier function [48] of the dual problem is defined by

$$V_{\mathcal{D}}(y) = \frac{1}{2} \log \det \nabla^2 F_{\mathcal{D}}(y).$$

Now consider the redundant Klee-Minty construction given is Chapter 6. The volumetric path of problem (6.1.1) in dimension 3 with and without the presence of redundancy is shown in Figure 8.1. In the redundant case, we have chosen $h = 3 \times 10^6(1000; 1; 1)$ and $d = (64; 8; 0)$.

This illustration in 3D suggests that by adding a huge number of redundant constraints, we may bend the volumetric path along the edges of the Klee-Minty cube so that it visits the δ-neighborhood of each vertex of the cube.

Tight Iteration Complexity Bounds for Other Path-Following IPMs

In Chapter 6, we provided a lower bound for the number of iterations required by central-path-following IPMs which almost matches the well-known iteration complexity upper bound $\mathcal{O}\big(\sqrt{n}\log\frac{n\mu^0}{\varepsilon}\big)$. Therefore, the iteration complexity upper bound is tight. A question that may naturally arise here is as follows. How tight are the proven iteration complexity upper bounds [36] for other path-following IPMs, e.g., volumetric- or universal-path-following variants?

The well-known iteration complexity upper bound required by volumetric-path-following IPMs [47] to solve problem (1.1.1) is

$$\mathcal{O}\left(\sqrt{m}\sqrt[4]{n}\log\frac{n\mu^0}{\varepsilon}\right). \tag{8.2.1}$$

To the best of our knowledge, there is no proven iteration complexity lower bound for these algorithms. In the previous subsection, we have provided some evidence indicating that the volumetric path may be forced to follow the simplex path of the Klee-Minty cube. As a follow-up result, this construction would lead to a lower bound for the number of iterations required by volumetric-path-following IPMs. We would also argue that the number of redundant inequalities required to bend the volumetric path along the simplex path of the Klee-Minty cube has to be at least $\Omega(2^{4m})$. The reason is that the volumetric-path-following IPMs would require at least $2^m - 1$ iterations which has to be smaller than or equal to the iteration complexity upper bound given by (8.2.1).

Efficient Polynomial Algorithms Not Suffering from Redundancy

As we have observed throughout the thesis, the central path is highly affected by redundancy. In exact words, the redundant KM constructions demonstrate that the central path is not "central" in a geometrical sense. Since many real-world problems appear to be redundant, it is an interesting and important research direction to design a polynomial-time algorithm that does not suffer, or at least suffers less, from redundancy.

The simplex method, as a non-polynomial algorithm, is not affected by repetition of constraints, but it suffers from degeneracy. One research direction may be to consider a mixed simplex-type method together with IPMs to take the advantages of both at the same time and design an algorithm that suffers neither from redundancy nor from degeneracy.

Another direction can be to develop a path-following IPM that follows a path that is not affected or at least less affected by redundancy. Theoretically, such a path exists. For instance, the path induced by the universal barrier [36] is not affected by redundancy. However, with current knowledge, it is hopeless to implement a universal-path-following IPM efficiently. It is worthwhile mentioning that the complexity of the volumetric-center cutting-plane algorithm [3] depends only on the dimension of the problem, but not on the number of inequalities, as well as the cost of calling a *separation oracle*[1]. Therefore, finding a separation oracle whose cost per call is reasonably low may lead to a practical algorithm that does not suffer from redundancy.

[1]For a given convex set $\mathcal{E} \in \mathbb{R}^n$ and a point $\bar{x} \in \mathbb{R}^n$, a separation oracle either reports that $\bar{x} \in \mathcal{E}$ or returns a vector $u \in \mathbb{R}^n$ such that $u^T x > u^T \bar{x}$ for every $x \in \mathcal{E}$.

Bibliography

[1] I. Adler and R. D. C. Monteiro, *Limiting behavior of the affine scaling continuous trajectories for linear programming problems*, Mathematical Programming, 50 (1991), pp. 29–51.

[2] K. M. Anstreicher, *Potential reduction algorithms*, in Interior Point Methods of Mathematical Programming, T. Terlaky, ed., Kluwer Academic Publishers, 1996, pp. 125–158.

[3] ——, *Towards a practical volumetric cutting plane method for convex programming*, SIAM Journal on Optimization, 9 (1999), pp. 190–206.

[4] D. Avis and V. Chvátal, *Notes on Bland's pivoting rule*, Mathematical Programming Study, 8 (1978), pp. 24–34.

[5] E. R. Barnes, *A variation on Karmarkar's algorithm for solving linear programming problems*, Mathematical Programming, 36 (1986), pp. 174–182.

[6] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*, no. 1 in Athena Scientific Optimization and Computation Series, Athena Scientific, Belmont, Massachusetts, 2003.

[7] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, 1997.

[8] R. G. Bland, *New finite pivoting rules for the simplex method*, Mathematics of Operations Research, 2 (1977), pp. 103–107.

[9] I. Castillo and E. Barnes, *Chaotic behavior of the affine scaling algorithm for linear programming*, SIAM Journal on Optimization, 11 (2000), pp. 781–795.

[10] G. B. Dantzig, *Maximization of a linear function of variables subject to linear inequalities*, in Activity Analysis of Production and Allocation, T. C. Koopmans, ed., John Wiley, 1951, pp. 339–347.

[11] D. den Hertog, *Interior Point Approach to Linear, Quadratic, and Convex Programming*, vol. 277 of Mathematics and its Applications, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[12] A. Deza, E. Nematollahi, R. Peyghami, and T. Terlaky, *The central path visits all the vertices of the Klee-Minty cube*, Optimization Methods and Software, 21 (2006), pp. 851–865.

[13] A. Deza, E. Nematollahi, and T. Terlaky, *How good are interior-point methods? Klee-Minty cubes tighten iteration-complexity bounds*, Mathematical Programming, 113 (2008), pp. 1–14.

[14] A. Deza, T. Terlaky, and Y. Zinchenko, *Polytopes and arrangements: diameter and curvature*, Operations Research Letters, 36, pp. 215–222.

[15] I. I. Dikin, *Iterative solution of problems of linear and quadratic programming*, Soviet Mathematics Doklady, 8 (1967), pp. 674–675.

[16] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons, New York, 1968. Reprint: vol. 4 of SIAM Classics in Applied Mathematics, SIAM Publications, Philadelphia, PA 191042688, USA, 1990.

[17] J. C. Gilbert, C. C. Gonzaga, and E. Karas, *Examples of ill-behaved central paths in convex optimization*, Mathematical Programming, 103 (2005), pp. 63–94.

[18] O. Güler and Y. Ye, *Convergence behavior of interior-point algorithms*, Mathematical Programming, 60 (1993), pp. 215–228.

[19] L. A. Hall and R. J. Vanderbei, *Two-third is sharp for affine scaling*, Operations Research Letters, 13 (1993), pp. 197–201.

[20] B. Jansen, C. Roos, and T. Terlaky, *A polynomial primal-dual Dikin-type algorithm for linear programming*, Mathematics of Operations Research, 21 (1996), pp. 341–353.

[21] N. K. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395.

[22] L. G. Khachiyan, *A polynomial algorithm in linear programming*, Soviet Mathematics Doklady, 20 (1979), pp. 191–194.

[23] V. Klee and G. J. Minty, *How good is the simplex algorithm?*, in Inequalities III, O. Shisha, ed., Academic Press, 1972, pp. 159–175.

[24] M. Kojima, S. Mizuno, and A. Yoshise, *An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems*, Mathematical Programming, 50 (1991), pp. 331–342.

[25] W. F. Mascarenhas, *The affine scaling algorithm fails for stepsize 0.999*, SIAM Journal on Optimization, 7 (1997), pp. 34–46.

[26] N. Megiddo, *On finding primal– and dual–optimal bases*, ORSA Journal on Computing, 3 (1991), pp. 63–65.

[27] N. Megiddo and M. Shub, *Boundary behavior of interior point algorithms in linear programming*, Mathematics of Operations Research, 14 (1989), pp. 97–146.

[28] S. Mehrotra, *On finding a vertex solution using interior-point methods*, Linear Algebra and Its Applications, 152 (1991), pp. 233–253.

[29] S. Mehrotra and Y. Ye, *Finding an interior point in the optimal face of linear programs*, Mathematical Programming, 62 (1993), pp. 497–515.

[30] S. Mizuno, M. Todd, and Y. Ye, *On adaptive-step primal-dual interior-point algorithms for linear programming*, Mathematics of Operations Research, 18 (1993), pp. 964–981.

[31] K. Murty, *Computational complexity of complementary pivot methods*, Mathematical Programming Study, 7 (1978), pp. 61–73.

[32] ——, *Computational complexity of parametric linear programming*, Mathematical Programming, 19 (1980), pp. 213–219.

[33] K. G. Murty, *Linear Complementarity, Linear and Nonlinear Programming*, vol. 3 of Sigma Series in Applied Mathematics, Heldermann Verlag, Berlin, Germany, 1988.

[34] E. Nematollahi and T. Terlaky, *A redundant Klee-Minty construction with all the redundant constraints touching the feasible region*, Operations Research Letters. To appear.

[35] ——, *A simpler and tighter redundant Klee-Minty construction*, Optimization Letters. To appear.
`http://www.springerlink.com/content/x013qq7g16kg3347/`.

[36] Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 1994.

[37] C. Roos, *An exponential example for Terlaky's pivoting rule for the criss-cross simplex method*, Mathematical Programming, 46 (1990), pp. 79–84.

[38] C. Roos, T. Terlaky, and J.-P. Vial, *Theory and Algorithms for Linear Optimization: An Interior Point Approach*, Springer, New York, NY, second ed., 2006.

[39] A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley & Sons, Inc., New York, NY, USA, 1986.

[40] N. Shor, *Utilization of the operation of space dilatation in the minimization of convex functions*, Kibernetica, 1 (1970), pp. 6–12. English Translation: *Cybernetics* 6, pp. 7–15.

[41] K. Tanabe, *Centered Newton method for mathematical programming*, in System Modelling and Optimization: Proceedings of the 13th IFIP Conference, Tokyo, Japan, Aug./Sept. 1987, M. Iri and K. Yajima, eds., vol. 113 of Lecture Notes in Control and Information Sciences, Springer Verlag, 1988, pp. 197–206.

[42] T. Terlaky, *A finite criss-cross method for oriented matroids*, Journal of Combinatorial Theory (B), 42 (1987), pp. 319–327.

[43] T. Terlaky and T. Tsuchiya, *A note on Mascarenhas' counterexample about global convergence of the affine scaling algorithm*, Applied Mathematics and Optimization, 40 (1999), pp. 287–314.

[44] T. Terlaky and S. Zhang, *Pivot rules for linear programming – a survey*, Annals of Operations Research, 46 (1993), pp. 203–233.

[45] M. J. Todd and Y. Ye, *A centered projective algorithm for linear programming*, Mathematics of Operations Research, 15 (1990), pp. 508–529.

[46] T. Tsuchiya, *Affine scaling algorithm*, in Interior Point Methods of Mathematical Programming, T. Terlaky, ed., Kluwer Academic Publishers, 1996, pp. 35–82.

[47] P. Vaidya and D. Atkinson, *A technique for bounding the number of iterations in path following algorithms*, in Complexity in Numerical Optimization, P. Pardalos, ed., World Scientific, Singapore, 1993, pp. 462–489.

[48] P. M. Vaidya, *A new algorithm for minimizing convex functions over convex sets*, Mathematical Programming, 73 (1996), pp. 291–341.

[49] R. J. Vanderbei, M. S. Meketon, and B. A. Freedman, *A modification of Karmarkar's linear programming algorithm*, Algorithmica, 1 (1986), pp. 395–407.

[50] Y. Ye, *On the finite convergence of interior-point algorithms for linear programming*, Mathematical Programming, 57 (1992), pp. 325–335.

[51] D. Yudin and A. S. Nemirovsky, *Informational complexity and efficient methods for the solution of convex extremal problems*, Ékonomika i Mathematicheskie Metody, 12 (1976), pp. 357–369. English Translation: *Matekon* 13, pp. 3–25.

[52] N. Zadeh, *A bad network problem for the simplex method and other minimum cost flow algorithms*, Mathematical Programming, 5 (1973), pp. 255–266.

[53] ——, *What is the worst case behavior of the simplex algorithm?*, Technical Report No. 27, Department of operations research, Stanford University, 1980.

[54] S. Zionts, *The criss-cross method for solving linear programming problems*, Management Science, 15 (1969), pp. 426–445.

Index