Interior Point Method for Linear and Convex Optimizations

by

Shiu-Tung NG

Thesis
Submitted to the Faculty of the Graduate School of
The Chinese University of Hong Kong
(Division of Mathematics)

In partial fulfillment of the requirements
for the Degree of
Master of Philosophy

July, 1998

Abstract

This thesis is a survey on the interior point method for solving linear optimization (LO) and convex optimization (CO) problems. The major materials of the thesis come from [7], [14], [26], and [27]. The target of LO and CO are to find the optimal value of an objective function subject to some constraints. LO and CO problems are widely used in applications of many areas such as statistics, engineering, economics, etc. We will discuss some simple interior point algorithms for LO and CO problems. The interior point algorithm is an iterative method which seeks the solution inside the feasible set of an optimization problem.

In chapter 1, we state the LO and CO problems to be studied. We also give some definitions and theorems which will be used later.

In chapter 2, we work on the LO problem. We will show the relation between the primal and dual problem and then give a simple algorithm which can solve the self-dual problem (combining the primal and dual problem) by Newton steps. The solution of the LO problem is then given by the self-dual problem and we will illustrate some LO problems at the end of this chapter.

In chapter 3, we give the *Logarithmic Barrier Algorithm* and *Center Algorithm* to solve the CO problem. These two iterative methods solve a sequence of unconstrained minimization problems associated with each CO problem instead of solving the CO problem directly. In order to apply the two algorithms, some assumptions have to be satisfied. Firstly, the objective function and the constraints function should be twice continuously differentiable. Secondly, the interior of the feasible set should be bounded and nonempty. Finally, the functions of the unconstrained minimization problems are κ-*self-concordant*. We will finally apply these two algorithms to some CO problems from the real applications.

摘要

線性最優化的內點法的理論已趨於成熟，非線性最優化的內點法理論則尚屬研究初期。本畢業論文將已有的線性優化及凸優化問題的內點法理論加以系統的綜述。內點法是一種迭代法，在優化問題的可行集中搜索求解。

在第一章，我們提出本論文中將討論的線性最優化及凸最優化問題，同時我們將介紹一些後面將要用到的定義及定理。

第二章討論線性最優化問題，我們先說明最優化問題與及對偶問題的關係，再將它們結合形成一個自對偶問題。之後介紹用於求解這對偶問題的牛頓算法，而原來問題的解是自對偶問題的解之一部份。最後我們應用這個算法到某些線性最優化問題上。

在第三章，我們討論兩個求解凸最優化問題的算法，即對數障礙法和中心法。這兩個迭代法能求出連帶於每一個凸最優化問題之無約束極小化問題序列的解。而且解之極限正是原來問題的解。最後我們會應用這兩個算法求解一些實際的凸最優化問題。

ACKNOWLEDGMENTS

I wish to express my sincere and deepest gratitude to my supervisor, Prof. Zou Jun, for his inspired guidance, constant encouragement and help throughout the period of my M.Phil. studies and in the preparation of this thesis. I would like to thank my colleagues Y.L.Keung, W.C.Leung, W.C.Tang, C.F.Chan and K.W.Mak for their many helpful discussions.

Contents

Chapter 1

Preliminary

1.1 Linear and Convex Optimization Model

This thesis will focus on the general *linear optimization* problem:

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

where A is a $m \times n$ matrix, $b \in \mathbb{R}^m, c \in \mathbb{R}^n$ are two given vectors, and $x \in \mathbb{R}^n$ is the unknown vector, and focus on the *convex optimization* problem:

$$\begin{aligned} \max \quad & f_0(y) \\ s.t. \quad & f_i(y) \leq 0, \quad i = 1, 2, \cdots, n, \end{aligned}$$

where $y \in \mathbb{R}^m$ and $f_i(y), i = 0, 1, \cdots, n$ are convex functions. We will not discuss the whole class of convex functions but the smooth convex functions which satisfy some conditions. These conditions are needed in the theoretical analysis and will be given later.

1.2 Notations for Linear Optimization

The following are some definitions which will be used in chapter 2.

- A *symmetric matrix* is a matrix M such that $M^t = M$.
- A matrix M is called *skew-symmetric* if $M^t = -M$.
- The *generalized inverse* of a square matrix $A_{n\times n}$ of rank r is defined by

$$A^+ = A_2^T(A_2A_2^T)^{-1}(A_1^TA_1)^{-1}A_1^T$$

  where A_1 is a $n \times r$ matrix and A_2 is a $r \times n$ matrix and they are of rank r such that $A = A_1A_2$.
- A set $S \in \mathbb{R}^n$ is called a *compact set* if it is closed and bounded.
- For any $x, y \in \mathbb{R}^n$, xy denotes the componentwise product.
  That is $xy = (x_1y_1, x_2y_2, \cdots, x_ny_n)^T$.
- For any $x, y \in \mathbb{R}^n$, x and y are called *strictly complementary* if $xy = 0$ and $x + y > 0$.

Lemma 1.2.1 *If S and X are positive diagonal matrix, M is skew-symmetric and they are of the same size, then $S + XM$ is nonsingular.*

Proof: By considering

$$(S + XM)y = 0.$$

Since X is a positive diagonal matrix, there is a vector z with the relation $y = Xz$. Thus we have

$$(S + XM)Xz = 0.$$

By multiply both sides by z^T, we get

$$z^T(S + XM)Xz = 0.$$

This is equivalent to

$$z^TSXz + (Xz)^TMXz = 0.$$

Since M is skew symmetric and X, S are positive diagonal matrix, we must have $z = 0$. This proves that $S + XM$ is nonsingular. □

1.3 Definition and Properties of Convexities

We now introduce some basic concepts and definitions.

- A set of points T is a *convex set* if every convex combination of points in T is also in T. That is, $\forall x_1, x_2 \in T$ and $\lambda \in [0, 1]$,

$$\lambda x_1 + (1 - \lambda)x_2 \in T.$$

- A function f is a *convex function* in a nonempty convex set S if $\forall x_1, x_2 \in S$ and $\lambda \in [0, 1]$,

$$f[\lambda x_1 + (1 - \lambda)x_2] \leq \lambda f(x_1) + (1 - \lambda)f(x_2). \tag{1.1}$$

  The function f is *strictly convex* if a strict inequality holds in (1.1) when $\lambda \in (0, 1)$ and $x_1 \neq x_2$. If the convex set S is replaced by the entire space, we simply say that f is a convex function.

We have two equivalent definitions for a continuously differentiable convex function f.

Lemma 1.3.1 *If f is continuously differentiable for $x \in S$, then an equivalent definition of a convex function is that*

$$f(x_2) \geq f(x_1) + (x_2 - x_1)^T \bigtriangledown f(x_1), \qquad \forall x_1, x_2 \in S. \tag{1.2}$$

f is strictly convex in S if strict inequality holds in (1.2) whenever $x_1 \neq x_2$.

Proof: For each $x_1, x_2 \in S$ and $0 < \lambda < 1$, we have

$$f[x_1 + \lambda(x_2 - x_1)] = f[(1 - \lambda)x_1 + \lambda x_2] \leq (1 - \lambda)f(x_1) + \lambda f(x_2).$$

Setting $h = x_2 - x_1$, we get

$$f(x_1 + \lambda h) - f(x_1) \leq \lambda[f(x_2) - f(x_1)]$$

Subtracting $\nabla f(x_1)^T(\lambda h)$ from both sides and dividing by λ gives

$$\frac{f(x_1 + \lambda h) - f(x_1) - \nabla f(x_1)^T(\lambda h)}{\lambda} \leq f(x_2) - f(x_1) - \nabla f(x_1)^T(x_2 - x_1)$$

Now as $\lambda \to 0$, the left side goes to zero while the right side, being independent of λ, remains constant. This implies

$$f(x_2) \geq f(x_1) + (x_2 - x_1)^T \nabla f(x_1),$$

which proves the first part of the lemma.

Now suppose that f is continuously differentiable for $x \in S$ and satisfies (1.2). Given $x_1, x_2 \in S$, $\lambda \in (0,1)$, we set $x_0 = \lambda x_1 + (1 - \lambda)x_2$. Then

$$f(x_0) = f(x_0) + \nabla f(x_0)^T[\lambda(x_1 - x_0) + (1 - \lambda)(x_2 - x_0)]$$

and using the linearity of $\nabla f(x_0)$ we can write this as

$$f(x_0) = \lambda[f(x_0) + \nabla f(x_0)^T(x_1 - x_0)] + (1 - \lambda)[f(x_0) + \nabla f(x_0)^T(x_2 - x_0)].$$

Inequality (1.2) holds for $x = x_1$ and $x = x_2$, so

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2),$$

that is, f is convex.

Replacing the inequalities by strict inequalities, the last part of the lemma follows. □

Lemma 1.3.2 *If f is twice continuously differentiable, another way to define a convex function is that $\nabla^2 f(x)$ is positive semidefinite for all $x \in S$. If $\nabla^2 f(x)$ is positive definite, then $f(x)$ is strictly convex.*

Proof: For each $x_1 \neq x_2 \in S$, considering the expansion of f at x_2,

$$f(x_2) = f(x_1) + (x_2 - x_1)^T \bigtriangledown f(x_1) + (x_2 - x_1)^T \bigtriangledown^2 f(x_0)(x_2 - x_1),$$

where $x_0 \in (x_1, x_2)$.

Since $\bigtriangledown^2 f(x)$ is positive semidefinite, $(x_2 - x_1)^T \bigtriangledown^2 f(x_0)(x_2 - x_1) \geq 0$. Therefore, $f(x_2) \geq f(x_1) + (x_2 - x_1)^T \bigtriangledown f(x_1)$. By lemma 1.3.1 , f is convex in S.

On the other hand, suppose that f is convex in S. For $x, h \in S$ and $t \in \mathbb{R}$ such that $x + th \in S$, we define $g(t) = f(x + th)$. Then g is convex in a neighborhood of the origin and

$$\begin{aligned} g'(t) &= \bigtriangledown f(x + th)^T h, \\ g''(t) &= h^T \bigtriangledown^2 f(x + th)^T h. \end{aligned}$$

The convexity of g implies that $g''(0) \geq 0$, i.e. $h^T \bigtriangledown^2 f(x)^T h \geq 0$. Since h is arbitrary, $\bigtriangledown f(x)$ is positive semidefinite. □

We now state two more lemmas about the properties of a convex function.

Lemma 1.3.3 *If $f(x)$ is a convex function in a convex set S, then for $k \in \mathbb{R}$, $\hat{S} = \{x \mid f(x) \leq k,\ x \in S\}$ is a convex set.*

Proof: Choose any $x_1, x_2 \in S$ such that $f(x_1) \leq k$ and $f(x_2) \leq k$. Since $f(x)$ is convex in S, $\forall \lambda \in [0, 1]$, $\lambda x_1 + (1 - \lambda) x_2 \in S$ and

$$\begin{aligned} f[\lambda x_1 + (1 - \lambda) x_2] &\leq \lambda f(x_1) + (1 - \lambda) f(x_2) \\ &\leq \lambda k + (1 - \lambda) k \\ &= k. \end{aligned}$$

This completes the proof. □

The next section gives two well-known results about the unconstrained minimization which will be used in chapter 3.

1.4 Useful Theorem for Unconstrained Minimization

The following are two main theorems for unconstrained minimization.

Theorem 1.4.1 *A necessary condition that a differentiable function f has an unconstrained local minimum at a point x^* is that*

$$\nabla f(x^*) = 0.$$

Theorem 1.4.2 *Sufficient conditions that a point x^* be an isolated local unconstrained minimum of the twice-differentiable function f are that*

$$\nabla f(x^*) = 0$$

and

$$y^T \nabla^2 f(x^*) y > 0, \qquad \forall y \neq 0.$$

The proof of these two theorems can be found in many optimization textbook (see [1]).

Chapter 2

Linear Optimization

Before introducing algorithm for solving LO, we first give the background of the algorithm.

2.1 Self-dual Linear Optimization Model

We can find in many textbook that every LO problem can be rewritten in the canonical form given by (P), the primal problem

$$(P) \qquad \min\{c^T x \mid \ Ax \geq b,\ x \geq 0\}.$$

According to (P), we can construct the dual problem of (P)

$$(D) \qquad \max\{b^T y \mid \ A^T y \leq c,\ y \geq 0\}.$$

Lemma 2.1.1 *(weak duality) If x is feasible for (P) and y is feasible for (D) then $c^T x \geq b^T y$. Furthermore, if $c^T x = b^T y$ then x and y are optimal.*

Proof: Let x is feasible for (P) and y is feasible for (D). Then $x \geq 0, y \geq 0, Ax \geq b$ and $A^T y \leq c$. As a consequence we have

$$b^T y \leq (Ax)^T y = x^T (A^T y) \leq c^T x.$$

Hence, any y that is feasible for (D) provides a lower bound $b^T y$ for the value of $c^T x$, whenever x is feasible for (P). Conversely, any x that is feasible for (P) provides an upper bound $c^T x$ for the value of $b^T y$, whenever y is feasible for (D). This phenomenon is known as the *weak duality property.* Thus, if $c^T x = b^T y$ for a feasible pair (x, y) then x is optimal for (P) and y is optimal for (D). □

One of the most important results in LO states that the converse statement of lemma 2.1.1 is also true. This is known as the *strong duality property.*

Lemma 2.1.2 *(strong duality) A feasible solution x to the primal problem is optimal if and only if there exists a feasible solution y to the dual problem such that*

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

In particular, y is an optimal solution to the dual.

The proof of the *Strong Duality Theorem* can be found in many LO textbooks (see [28]).

Now, we introduce the *self-dual LO model* (SP)

$$(SP) \qquad \min\{q^T x \mid \ s(x) := Mx + q \geq 0,\ x \geq 0\}$$

where M is a $n \times n$ skew symmetric matrix, $q \geq 0$. The meaning of self-dual is that the dual of (SP) is exactly the same problem as (SP) itself. The argument is straightforward by writing down the dual of (SP)

$$(DSP) \qquad \max\{-q^T y \mid \ M^T y \leq q,\ y \geq 0\}.$$

Since M is skew symmetric $(M = -M^T)$ and a maximizing problem can be written as a minimizing problem by changing the sign of the cost function, it is equivalent to write (DSP) as

$$(DSP) \qquad \min\{q^T y \mid \ -My \leq q,\ y \geq 0\}.$$

Therefore, the self-duality of (SP) is concluded.

The reason we consider the skew symmetric model is because every LO problem can be embedded in a self-dual problem. By the *Strong Duality Theorem*, whenever both (P) and (D) have an optimal solution then the system

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

has a solution, and any such solution gives optimal solutions for (P) and (D). Writing the above system in matrix form,

$$\begin{pmatrix} 0 & A \\ -A^T & 0 \\ b^T & -c^T \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} \geq \begin{pmatrix} b \\ -c \\ 0 \end{pmatrix} \qquad x \geq 0,\ y \geq 0. \tag{2.1}$$

Now introducing a homogenizing variable κ, the coefficient matrix becomes skew symmetric as follows:

$$\begin{pmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{pmatrix} \begin{pmatrix} y \\ x \\ \kappa \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad x \geq 0,\ y \geq 0,\ \kappa \geq 0. \tag{2.2}$$

Note that this system is completely equivalent to (2.1) if $\kappa = 1$. Actually, given any solution (x, y, κ) of (2.2) with $\kappa > 0$, $(\frac{x}{\kappa}, \frac{y}{\kappa}, 1)$ also solves the system, since the system is homogeneous. If $\kappa = 0$ for every solution of (2.2), then the first system (2.1) has no solution. The reason is that if (2.1) has a solution (x, y), then $(x, y, 1)$ solves (2.2). Thus it becomes natural to consider the following self-dual problem:

$$\min \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}^T \begin{pmatrix} y \\ x \\ \kappa \end{pmatrix} : \quad M \begin{pmatrix} y \\ x \\ \kappa \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} y \\ x \\ \kappa \end{pmatrix} \geq 0 \right\}$$

where

$$M = \begin{pmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{pmatrix}.$$

In the next section, we will introduce some definitions and theorems which are useful in the further analysis.

2.2 Definitions and Main Theorems

The following are some important definitions:

- The *feasible set* of the self-dual LO model (SP) is defined by

$$SP := \{x \ : \ x \geq 0, \ s(x) := Mx + q \geq 0\}.$$

- The vector $s(x) = Mx + q$ is called the *surplus vector* of vector x.

Lemma 2.2.1 *(SP) admits the zero vector $x = 0$ as a feasible solution, and this solution is optimal.*

Proof: Since M is skew symmetric, we have

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

Consequently, given any $x_0 \geq 0$ such that $s(x_0) \geq 0$ in SP, the cost function $q^T x_0 \geq 0$ in SP. Therefore $x = 0$ with *surplus vector* $s(x) = q \geq 0$ is obvious feasible and is optimal since the cost function $q^T x = 0$. □

Remark: The optimal solution $x = 0$ may be not unique in general as the following example shows:

$$M = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad q = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Here, $(0 \ x_2)$ is an optimal solution for any $x_2 \geq 0$. □

- We define the *optimal set* of (SP) by

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

As the cost function $q^T x = e^T(xs)$ and $x = 0$ is optimal, the optimal set indeed shows every optimal solution must be in SP^* and every $x \in SP^*$ must be an optimal solution.

Lemma 2.2.2 *Let x and y be feasible for (SP). Then x and y are optimal if and only if*

$$xs(y) = ys(x) = xs(x) = ys(y) = 0.$$

Proof: Since M is skew symmetric

$$\begin{aligned}(x-y)^T(s(x)-s(y)) &= (x-y)^T[(Mx+q)-(My+q)] \\ &= (x-y)^T M(x-y) \ = \ 0.\end{aligned}$$

Hence

$$x^T s(y) + y^T s(x) = x^T s(x) + y^T s(y)$$

and this vanishes if and only if x and y are optimal, by noting that $x, s(x), y, s(y) \geq 0$. Thus optimal solutions are complementary. □

- (SP) satisfies the so-called *Interior Point Condition* (*IPC*) if there exists a positive vector x such that $s(x)$ is positive.

Theorem 2.2.1 *The next statements are equivalent.*
(i) (SP) satisfies the IPC.
(ii) For $\mu > 0$, there exists $(x, s) > 0$ such that

$$\begin{cases} Mx + q &= s, \\ xs &= \mu e. \end{cases}$$

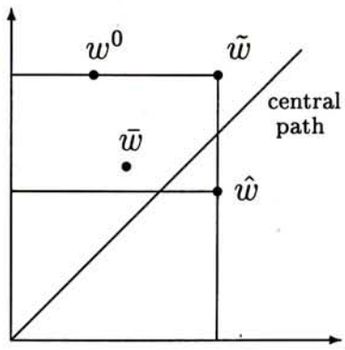


Figure 2.1: $L_{\hat{w}} \neq \emptyset$

(iii) For $w > 0$, there exists $(x, s) > 0$ such that

$$\begin{cases} Mx + q &= s, \\ xs &= w. \end{cases}$$

Proof: $(iii) \rightarrow (ii) \rightarrow (i)$ is clearly hold.

$(i) \rightarrow (iii)$ Suppose that there is a positive vector $x^0 \in SP$ with positive surplus vector $s^0 := s(x^0)$ such that $w^0 = x^0 s^0$. Let

$$L_w := \{x \in SP : \ xs(x) \leq w\}.$$

First we want to prove by contradiction that for each $\hat{w} > 0, L_{\hat{w}} \neq \emptyset$.

If $\hat{w} > x^0 s(x^0)$, then $x^0 \in L_{\hat{w}}$ and so $L_{\hat{w}} \neq \emptyset$.

If $\hat{w} \not\geq x^0 s(x^0)$, then we assume that $L_{\hat{w}} = \emptyset$ and let $\tilde{w} := \max\{\hat{w}, w^0\}$,

$$g(w) := \begin{cases} \| \max\{w - \hat{w}, 0\}\}_\infty & if \ L_w \neq \emptyset \\ \infty & if \ L_w = \emptyset \end{cases}$$

further let $\bar{w}$ be the minimizer of $g(w)$ in the set $L_{\tilde{w}}$.

If $L_{\tilde{w}}$ is compact, then the vector $\bar{w} \not\leq \hat{w}$ exists. We can get more idea in the figure 2.1. In fact for each $w > 0$, L_w is compact. First L_w is closed by its definition. Second if there is $x \in L_w$, that is $(x, s(x)) \geq 0$ and $x^T s(x) \leq e^T w$,

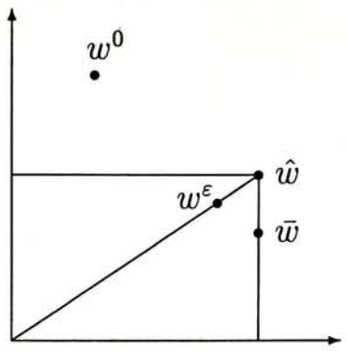


Figure 2.2: (*iii*) is solvable

then we have

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

It implies for each j,

$$\begin{aligned} x_j s_j^0 \leq x^T s^0 &\leq x^T s^0 + s^T x^0 \\ &\leq x^T s + (x^0)^T s^0 \\ &= e^T w + (x^0)^T s^0. \end{aligned}$$

Therefore,

$$x_j \leq \frac{e^T w + (x^0)^T s^0}{s_j^0}.$$

Therefore L_w is bounded and so L_w is compact. Since $L_{\bar{w}}$ is compact, we know that there is a vector $\bar{w} \not\leq \hat{w}$ such that $g(\bar{w}) \leq g(w)$ for all $w \in L_{\bar{w}}$. We now want to find a now vector w_α such that $g(w_\alpha) < g(w)$. Now we apply the Newton step for a sufficient small α from $\bar{w}$ to $\hat{w}$ and we denote it by

$$\begin{aligned} w_\alpha = x_\alpha s_\alpha &= (\bar{x} + \alpha\Delta x)(\bar{s} + \alpha\Delta s) \\ &= \bar{x}\bar{s} + \alpha(\hat{w} - \bar{x}\bar{s}) + \alpha^2\Delta x\Delta s \\ &= \bar{w} + \alpha(\hat{w} - \bar{w}) + \alpha^2\Delta x\Delta s. \end{aligned}$$

By the definition of $\bar{w}$ and $g(w)$, we suppose that $g(\bar{w}) = (\bar{w} - \hat{w})_j$, that is the jth coordinate of $\hat{w} - \bar{w}$ is less than zero. Thus for

$$\alpha < \frac{-(\hat{w} - \bar{w})_j}{(\Delta x \Delta s)_j},$$

we find that the jth coordinate of w_α,

$$\begin{aligned}(w_\alpha)_j &= \bar{w}_j + \alpha(\hat{w} - \bar{w})_j + \alpha^2(\Delta x \Delta s)_j \\ &< \bar{w}_j.\end{aligned}$$

Thus, w_α is a point closer to $\hat{w}$ than $\bar{w}$ in $\|\cdot\|_\infty$ which gives a contradiction to the definition of $\bar{w}$. Thus we have $L_{\hat{w}} \neq \emptyset$ and $L_{\hat{w}}$ is compact for each $\hat{w} > 0$. Finally, we use the same technique to prove that for each $\hat{w} > 0$, there exists $(x, s) > 0$ such that

$$\begin{cases} Mx + q &= s, \\ xs &= \hat{w}. \end{cases}$$

Suppose that this problem is not solvable, we define

$$g(w) = \| \max\{\hat{w} - w, 0\} \|_\infty$$

for $w \in L_{\hat{w}}$. We also define $\bar{w}$ is the minimizer of $g(w)$. Since $L_{\hat{w}}$ is nonempty and compact, $\bar{w}$ is well defined. we apply the Newton step with small enough α from $\bar{w}$ to $(1 - \varepsilon)\hat{w}$ gives a contradiction to the definition of $\bar{w}$. This completes the proof. □

In fact, the solutions of systems (ii) and (iii) are unique. Suppose that $(x, s) > 0$ and $(y, t) > 0$ be two solutions of (ii), we consider

$$\begin{aligned} 0 &= (x - y)^T M(x - y) \\ &= (x - y)^T (s - t) \\ &= x^T s + y^T t - x^T t - y^T s \\ &= 2n\mu - \sum_{i=1}^{n} x_i \frac{\mu}{y_i} - \sum_{i=1}^{n} y_i \frac{\mu}{x_i}. \end{aligned}$$

This is equivalent to

$$2n = \sum_{i=1}^{n} \frac{x_i^2 + y_i^2}{x_i y_i} \geq \sum_{i=1}^{n} \frac{2x_i y_i}{x_i y_i} = 2n.$$

The equality holds if and only if $x_i = y_i$ for all i. By the definition of s and t, we also get $s = t$. Therefore (x, s) is equal to (y, t), that is to say the solutions of (ii) are unique.

- The solution of (ii) is denoted by $x(\mu)$ and $s(\mu)$.
- The *central path* is defined by $\{x(\mu) : \mu > 0\}$.

The concept of central path will play a key role both in the development of the theory and in the design of algorithms.

- The *optimal partition* is defined by two subsets B and N of the index set $\{1, 2, \cdots, n\}$ for

$$\begin{aligned} B &:= \{i : x_i > 0, \ for\ some\ x \in SP^*\}, \\ N &:= \{i : s(x)_i > 0, \ for\ some\ x \in SP^*\}. \end{aligned}$$

The importance of the two subsets B and N will be shown after the following theorem.

Theorem 2.2.2 *Assume that (SP) satisfies the IPC. Then the limit of $x(\mu) \to x^*$ exists when $\mu \downarrow 0$ and is optimal. Moreover, $x_B^* > 0$, $s_N^* > 0$ and $x^* + s^* > 0$ where $s^* = s(x^*)$.*

Proof: We first prove for $\bar{\mu} > 0$, the set

$$\{(x(\mu), s(\mu)) : 0 < \mu \leq \bar{\mu}\}$$

is bounded. By the interior-point condition, there exists $(x^0, s^0) > 0$. Since

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

this leads to

$$\begin{aligned} s_j^0 x_j(\mu) \leq x(\mu)^T s^0 + (x^0)^T s(\mu) &= (x^0)^T s^0 + x(\mu)^T s(\mu) \\ &= n\mu + (x^0)^T s^0 \\ &\leq n\bar{\mu} + (x^0)^T s^0. \end{aligned}$$

This shows $x_j(\mu) \leq (n\bar{\mu} + (x^0)^T s^0)/s_j^0$. The set $\{x(\mu) : 0 < \mu \leq \bar{\mu}\}$ is therefore bounded. The proof for the boundedness of the set $\{s(\mu) : 0 < \mu \leq \bar{\mu}\}$ is similar. Using $x_j^0 s_j(\mu) \leq n\bar{\mu} + (x^0)^T s^0$, we get $s_j(\mu) \leq (n\bar{\mu} + (x^0)^T s^0)/x_j^0$. Now, let $\{\mu_k\}_{k=1}^{\infty}$ be a positive sequence such that $\mu_k \to 0$ for $k \to \infty$. So $\{(x(\mu_k), s(\mu_k))\}$ is bounded and contains a sequence converging to a point (x^*, s^*). Then (x^*, s^*) is feasible and from theorem 2.2.1(ii), $(x^*)^T s^* = 0$. So x^* is an optimal solution. We next show that (x^*, s^*) are strictly complementary. Since M is skew symmetric, we have

$$\begin{aligned} (x(\mu_k) - x^*)^T (s(\mu_k) - s^*) &= (x(\mu_k) - x^*)^T ((Mx(\mu_k) + q) - (Mx^* + q) \\ &= (x(\mu_k) - x^*)^T M(x(\mu_k) - x^*) = 0. \end{aligned}$$

Expanding the terms and using $(x^*)^T s^* = 0$ and $x_i(\mu_k)s_i(\mu_k) = \mu_k$, for all i, we get

$$(x^*)^T s(\mu_k) + x(\mu_k)^T s^* = n\mu_k,$$

writing in summation form, this implies

$$\sum_{i \in \{j:\, x_j^* > 0\}} x_i^* s_i(\mu_k) + \sum_{i \in \{j:\, s_j^* > 0\}} x_i(\mu_k) s_i^* = n\mu_k,$$

dividing both sides by μ_k and using $x_i(\mu_k)s_i(\mu_k) = \mu_k$, for all i, we have

$$\sum_{i \in \{j:\, x_j^* > 0\}} \frac{x_i^* s_i(\mu_k)}{x_i(\mu_k) s_i(\mu_k)} + \sum_{i \in \{j:\, s_j^* > 0\}} \frac{x_i(\mu_k) s_i^*}{x_i(\mu_k) s_i(\mu_k)} = n,$$

which is equivalent to

$$\sum_{i \in \{j:\, x_j^* > 0\}} \frac{x_i^*}{x_i(\mu_k)} + \sum_{i \in \{j:\, s_j^* > 0\}} \frac{s_i^*}{s_i(\mu_k)} = n.$$

Letting $k \to \infty$, this shows that the total number of positive terms in x^* and s^* is equal to n. So, $(x^*, s(x^*))$ is strictly complementary. □

From the proof of theorem 2.2.2, we know $B \bigcup N = \{1, 2, \cdots, n\}$ since $x^* s^* = 0$. Later, we will give an iterative method for solving (SP) and applying the partition, we can separate such iterative solution for small and large variables so that we can apply the rounding procedure to get the exact solution.

Next, we prove that this accumulation point x^* mentioned in theorem 2.2.2 is actually unique. To do so, we introduce the analytic center.

- If $\bar{x} \in SP^*, \bar{s} = s(\bar{x})$ maximize the product

$$\prod_{i \in B} x_i \prod_{i \in N} s_i$$

over $x \in SP^*$, then $\bar{x}$ is called an *analytic center* of SP^*.

Theorem 2.2.3 *Assuming the IPC is satisfied, then the limit point of the central path x^* is the analytic center of SP^*.*

Proof: Let (x^*, s^*) be a limit point of the central path, where $s^* = s(x^*)$. According to theorem 2.2.2, let $\{\mu_k\}_{k=1}^{\infty}$ be a positive sequence such that $\mu_k \to 0$ and such that $(x(\mu_k), s(\mu_k))$, converges to (x^*, s^*). Then x^* is optimal, which means $(x^*)^T s^* = 0$, and strictly complementary. Now let $\bar{x}$ be optimal in (SP) and let $\bar{s} = M\bar{x} + q$ be its surplus vector. Applying the orthogonality property to the points $\bar{x}$ and $x(\mu)$ we obtain

$$(x(\mu_k) - \bar{x})^T (s(\mu_k) - \bar{s}) = 0.$$

Rearranging terms and using $x(\mu_k)^T s(\mu_k) = n\mu_k$ and $(\bar{x})^T \bar{s} = 0$, we get

$$\sum_{i=1}^{n} \bar{x}_i s_i(\mu_k) + \sum_{i=1}^{n} x_i(\mu_k) \bar{s}_i = n\mu_k.$$

Since the pair (x^*, s^*) is strictly complementary and $(\bar{x}, \bar{s})$ is an arbitrary optimal pair, we have for each coordinate i:

$$x_i^* = 0 \Rightarrow \bar{x}_i = 0, \qquad s_i^* = 0 \Rightarrow \bar{s}_i = 0.$$

Thus we may write

$$\sum_{i \in \{j:\, x_j^* > 0\}} \bar{x}_i s_i(\mu_k) + \sum_{i \in \{j:\, s_j^* > 0\}} x_i(\mu_k) \bar{s}_i = n\mu_k,$$

dividing both sides by $\mu_k = x_i(\mu_k) s_i(\mu_k)$, we get

$$\sum_{i \in B} \frac{\bar{x}_i}{x_i(\mu_k)} + \sum_{i \in N} \frac{\bar{s}_i}{s_i(\mu_k)} = n,$$

letting $k \to \infty$,it follows that

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

Applying the arithmetic-geometric-mean inequality, we obtain

$$\left(\prod_{i \in B} \frac{\bar{x}_i}{x_i^*} \prod_{i \in N} \frac{\bar{s}_i}{s_i^*} \right)^{\frac{1}{n}} \leq \frac{1}{n} \left(\sum_{i \in B} \frac{\bar{x}_i}{x_i^*} + \sum_{i \in N} \frac{\bar{s}_i}{s_i^*} \right) = 1,$$

this leads to

$$\prod_{i \in B} \bar{x}_i \prod_{i \in N} \bar{s}_i \leq \prod_{i \in B} x_i^* \prod_{i \in N} s_i^*.$$

We have completed the proof. □

In the next section we show that the self-dual problem defined in section 2.1 can be further embedded into another self-dual problem which satisfies the *IPC* automatically.

2.3 Self-dual Embedding and Simple Example

In this section, we assume $q = 0$ in the self-dual model (SP). This is general enough as we see previously that any LO problem can be embedded in a self-dual

problem with $q = 0$. Now we introduce another self-dual problem which satisfied the *IPC*:

$$\min \left\{ (n+1)\vartheta : \begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix} \begin{pmatrix} x \\ \vartheta \end{pmatrix} + \begin{pmatrix} 0 \\ n+1 \end{pmatrix} \geq 0, \ \begin{pmatrix} x \\ \vartheta \end{pmatrix} \geq 0 \right\} \quad (2.3)$$

where $r = e - Me$. This problem satisfies the *IPC* automatically and gives the solutions for the original problem. We show this in the following: First, the *IPC* is satisfied since the vector $(e^T \ 1)^T$ with surplus vector $(e^T \ 1)^T$ is feasible:

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

Second, by theorem 2.2.2, there exists a strictly complementary solution $(\tilde{x}^T \ \tilde{\vartheta})^T$. Meanwhile, by lemma 2.2.1, the optimal value of the self-dual problem is equal to 0. Thus, the optimal value of the new cost function

$$(n+1)\tilde{\vartheta} = 0.$$

Since $n + 1 > 0$, it follows that $\tilde{\vartheta} = 0$. On the other hand, the surplus vector of $(\tilde{x}^T \ \tilde{\vartheta})^T$ is greater than or equal to zero, this gives

$$\begin{aligned} \begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ n+1 \end{pmatrix} &= \begin{pmatrix} M\tilde{x} \\ -r^T\tilde{x} + n + 1 \end{pmatrix} \\ &= \begin{pmatrix} s(\tilde{x}) \\ n + 1 - r^T\tilde{x} \end{pmatrix} \geq 0 \end{aligned}$$

where $s(\tilde{x})$ is the surplus vector of $\tilde{x}$ to the original self-dual problem. Therefore $\tilde{x} \geq 0$ and $s(\tilde{x}) \geq 0$, it implies that $\tilde{x}$ is optimal for the original problem. Finally, such $(\tilde{x}, s(\tilde{x}))$ is strictly complementary for the original problem since $(\tilde{x}^T \ \tilde{\vartheta})^T$ is a strictly complementary solution of the new problem.

Conclusion: From above, we know that every LO problem can be embedded in another self-dual problem of the form (SP) with $(x, s(x)) = (e, e)$ as an *interior*

point (*IP*). By theorem 2.2.2, a strictly complementary solution of the new problem can be found. As a consequence, the first n unknowns of the new self-dual problem solve the original problem. Moreover, if the n^{th} entry value κ is greater than zero, then an optimal solution of the original LO problem is found. On the other hand, if $\kappa = 0$ in all solutions then the original LO problem does not have an optimal solution. We use the following simple problem to illustrate the embedding procedure. Consider the problem

$$(EP) \quad \max\{y_1 + y_2 : \ -1 \leq y_1 \leq 1, \ y_2 \leq 1\}.$$

Introducing nonnegative slack variables s_1, s_2, s_3 and eliminating the free variables y_1 and y_2 by using

$$\begin{aligned} y_1 &= s_1 - 1, \\ y_2 &= 1 - s_3, \end{aligned}$$

we obtain the following canonical problem

$$\max\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}^T \begin{pmatrix} s_1 \\ s_3 \end{pmatrix} : \begin{pmatrix} 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} s_1 \\ s_3 \end{pmatrix} \geq 2, \begin{pmatrix} s_1 \\ s_3 \end{pmatrix} \geq 0\right\},$$

which is the dual problem of the following primal problem:

$$\min\left\{2x_1 : \begin{pmatrix} 1 \\ 0 \end{pmatrix} x_1 \geq \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \ x_1 \geq 0\right\}.$$

Now,

$$A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad c = (2), \quad b = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

We apply the embedding technique to this problem, yielding the self-dual problem

$$\min\left\{0 : \quad M\begin{pmatrix} s_1 \\ s_3 \\ x_1 \\ \kappa \end{pmatrix} \geq 0, \quad \begin{pmatrix} s_1 \\ s_3 \\ x_1 \\ \kappa \end{pmatrix} \geq 0\right\},$$

where x_1 is the dual variable and κ the homogenizing variable and

$$M = \begin{pmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 2 \\ 1 & -1 & -2 & 0 \end{pmatrix}.$$

The problem is further embedded into

$$\min \left\{ (n+1)\vartheta : \quad \begin{pmatrix} M & r \\ -r^T & 0 \end{pmatrix} \begin{pmatrix} \xi \\ \vartheta \end{pmatrix} \geq - \begin{pmatrix} 0 \\ n+1 \end{pmatrix}, \quad \begin{pmatrix} \xi \\ \vartheta \end{pmatrix} \geq 0 \right\}$$

where $n = 4$ and

$$r = e - Me = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 3 \end{pmatrix}.$$

We arrive, finally, at

$$\min \left\{ 5\vartheta : \begin{pmatrix} 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 2 & 0 \\ 1 & -1 & -2 & 0 & 3 \\ -1 & 0 & 0 & -3 & 0 \end{pmatrix} \begin{pmatrix} s_1 \\ s_3 \\ x_1 \\ \kappa \\ \vartheta \end{pmatrix} \geq - \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 5 \end{pmatrix}, \begin{pmatrix} s_1 \\ s_3 \\ x_1 \\ \kappa \\ \vartheta \end{pmatrix} \geq 0 \right\}.$$

Here, the all one vector is feasible and its surplus vector is also the all one vector.

2.4 Newton step

Now, we have the embedding problem (2.3) which satisfies the *IPC*. Then by theorem 2.2.1, the following system

$$\begin{cases} Mx + q &= s, \\ xs &= w \end{cases}$$

has a solution $(x, s) > 0$ for $w > 0$.

We are in fact interested in the case $w = \mu e$ and then letting $\mu \to 0$. We will continue to use w instead of μe for generality in the following argument. Now, we apply the Newton's method to find a displacements $\Delta x, \Delta s$ for a given (x, s) such that $x + \Delta x, s + \Delta s$ satisfy the system for the general new target w:

$$\begin{cases} M(x + \Delta x) + q &= s + \Delta s > 0, \qquad x + \Delta x > 0 \\ (x + \Delta x)(s + \Delta s) &= w. \end{cases}$$

Since $Mx + q = s$, it gives

$$\begin{cases} M\Delta x &= \Delta s, \\ x\Delta s + s\Delta x + \Delta x\Delta s &= w - xs. \end{cases} \tag{2.4}$$

By neglecting the quadratic term $\Delta x \Delta s$, we obtain the linear system

$$\begin{cases} M\Delta x &= \Delta s, \\ x\Delta s + s\Delta x &= w - xs. \end{cases} \tag{2.5}$$

Then the Newton direction at x to the target w is given by

$$\Delta x = (S + XM)^{-1}(w - xs),$$

where X and S are positive diagonal matrices with their diagonal entries being the components of x and s. Since M is skew-symmetric, by lemma 1.2.1, $S + XM$ is nonsingular. Hence Δx is defined uniquely. Such displacements $\Delta x, \Delta s$ will be exact for (2.4) if the quadratic term $\Delta x \Delta s = 0$. Newton step at (x, s) by (x^+, s^+):

$$x^+ = x + \Delta x, \qquad s^+ = s + \Delta s.$$

We now see that we can use the Newton step to find a direction at x to a new target, say in the central path. Therefore if we apply it repeatly by decreasing the scalar μ, the problem is solved and the algorithm is the so-called *centering method.* But we should be careful about the feasibility in each step. In practice, μ cannot be decreased too fast.

We will show the conditions for feasibility of each Newton step in the next lemma.

Lemma 2.4.1 *The Newton step is feasible if and only if* $w + \Delta x \Delta s \geq 0$ *and strictly feasible if and only if* $w + \Delta x \Delta s > 0$.

Proof: By considering $x^+ s^+$ and expanding the terms, we have

$$\begin{aligned} x^+ s^+ &= (x + \Delta x)(s + \Delta s) \\ &= xs + (s\Delta x + x\Delta s) + \Delta x \Delta s. \end{aligned}$$

Since the Newton step satisfies

$$s\Delta x + x\Delta s = w - xs,$$

so,

$$x^+ s^+ = w + \Delta x \Delta s.$$

Thus, if x^+ and s^+ are nonnegative (positive) , then their product is nonnegative (positive) as well. As a consequence, x^+ and s^+ are feasible implies $w + \Delta x \Delta s$ is nonnegative (positive).
Conversely, we define

$$x^\alpha = x + \alpha \Delta x, \qquad s^\alpha = s + \alpha \Delta s, \qquad 0 \leq \alpha \leq 1.$$

We have $x^0 s^0 = xs > 0$.

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

Using $s\Delta x + x\Delta s = w - xs$ gives

$$x^\alpha s^\alpha = xs + \alpha(w - xs) + \alpha^2 \Delta x \Delta s.$$

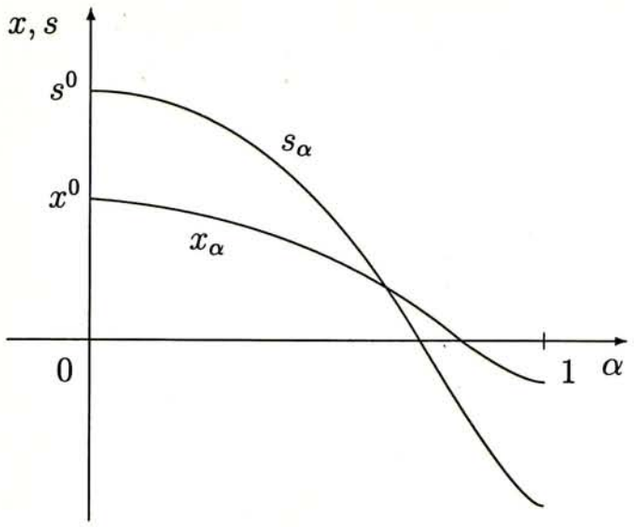


Figure 2.3: To show that $(x^+, s^+) \geq 0$

Now suppose $w + \Delta x \Delta s \geq 0$. Then it follows that

$$\begin{aligned} x^\alpha s^\alpha &\geq xs + \alpha(w - xs) - \alpha^2 w \\ &= (1-\alpha)(xs + \alpha w) \\ &> 0 \end{aligned}$$

for $0 \leq \alpha < 1$. Since x^0 and s^0 are positive and $x^\alpha s^\alpha$ are all positive for all $\alpha \in [0, 1)$, the vector $x^+ = x^1 = x + \Delta x$ and $s^+ = s^1 = s + \Delta s$ cannot have negative entries by the continuity of x^+ and s^+. We observe this property in figure 2.3 with one dimension. Similarly we can prove for the case $w + \Delta x \Delta s > 0$. This completes the proof. □

If the Newton step is feasible then we have the following result:

$$(x^+)^T s^+ = e^T(x^+ s^+) = e^T(w + \Delta x \Delta s) = e^T w + \Delta x^T M \Delta x = e^T w,$$

since $\Delta s = M \Delta x$.

In the following we analyze the convergence of the Newton step.

2.5 Rescaling and Definition of $\delta(xs, w)$

We use scaling vectors d and u defined below to adapt the new situation:

$$d := \sqrt{\frac{x}{s}}, \qquad u := \sqrt{\frac{xs}{w.}}$$

If we let

$$v = \sqrt{xs}.$$

Then $x, s, \Delta x$ and Δs are rescaled as follows:

$$x = vd, \qquad s = \frac{v}{d,}$$
$$d_x = \frac{\Delta x}{d,} \qquad d_s = d\Delta s,$$

where the division here is defined to be the componentwise division.

From (2.5) we see that

$$0 = \Delta x^T M \Delta x = \Delta x^T \Delta s = d_x^T d_s.$$

Therefore, the orthogonality of Δx and Δs is reserved. Meanwhile,

$$v(d_x + d_s) = s\Delta x + x\Delta s = w - xs := \Delta w.$$

Let

$$d_w := \frac{\Delta w}{v}$$

Then we have

$$d_x + d_s = d_w,$$

since $d_x^T d_s = 0$, we get

$$\|d_x\|^2 + \|d_s\|^2 = \|d_w\|^2.$$

This shows that the displacement d_x, d_s are zero if and only if $d_w = 0$. Thus, we have the definition of 'distance'

$$\delta(xs, w) := \frac{\|d_w\|}{2\sqrt{\min(w)}} = \frac{1}{2\sqrt{\min(w)}} \left\| \frac{xs - w}{\sqrt{xs}} \right\|.$$

From the definition, $\delta(xs, w) = 0$ if and only if $xs = w$.

Before stating the convergence theorem of the Newton step, we present a useful lemma.

Lemma 2.5.1 *We have* $\|d_x d_s\|_\infty \leq \frac{1}{4}\|d_w\|^2$ *and* $\|d_x d_s\| \leq \frac{1}{2\sqrt{2}}\|d_w\|^2$.

Proof: Note that,

$$d_x d_s = \frac{1}{4}((d_x + d_s)^2 - (d_x - d_s)^2),$$

so we have

$$-\frac{1}{4}(d_x - d_s)^2 \leq d_x d_s \leq \frac{1}{4}(d_x + d_s)^2.$$

This implies

$$-\frac{1}{4}\|d_x - d_s\|^2 e \leq d_x d_s \leq \frac{1}{4}\|d_x + d_s\|^2 e,$$

which yields the first inequality. On the other hand,

$$\begin{aligned} \|d_x d_s\|^2 = e^T (d_x d_s)^2 &= \frac{1}{16} e^T ((d_x + d_s)^2 - (d_x - d_s)^2)^2 \\ &\leq \frac{1}{16} e^T ((d_x + d_s)^4 + (d_x - d_s)^4) \\ &\leq \frac{1}{16} (\|d_x + d_s\|^4 + \|d_x - d_s\|^4) \\ &= \frac{1}{8}\|d_x + d_s\|^4, \end{aligned}$$

that gives the second inequality and the proof of lemma is complete. □

Now we can state the theorem.

Theorem 2.5.1 *The Newton step is feasible if* $\delta(xs, w) \leq 1$. *Moreover, if* $\delta(xs, w) < 1$ *then*

$$\delta(x^+s^+, w) \leq \frac{\delta^2}{\sqrt{2(1-\delta^2)}}$$

where $\delta := \delta(xs, w)$.

Proof: Consider

$$\left\| \frac{d_x d_s}{w} \right\|_\infty \leq \frac{\|d_x d_s\|_\infty}{\min(w)}.$$

From lemma 2.5.1, we obtain

$$\begin{aligned} \left\| \frac{d_x d_s}{w} \right\|_\infty &\leq \frac{\|d_w\|^2}{4\min(w)} \\ &= \delta(xs, w)^2 \\ &\leq 1. \end{aligned}$$

This implies $w + \Delta x \Delta s = w + d_x d_s \geq 0$. By lemma 2.4.1, the Newton step is now feasible. On the other hand, by using lemma 2.5.1 and the definition of δ,

$$\begin{aligned} \min(x^+s^+) &= \min(w + d_x d_s) \\ &\geq \min(w) - \|d_x d_s\|_\infty \\ &\geq \min(w) - \frac{1}{4}\|d_w\|^2 \\ &= \min(w)(1-\delta^2). \end{aligned} \tag{2.6}$$

Now

$$\begin{aligned} \delta(x^+s^+, w)^2 &= \frac{1}{4\min(w)} \left\| \frac{w - x^+s^+}{\sqrt{x^+s^+}} \right\|^2 \\ &= \frac{\|w - x^+s^+\|^2}{4\min(w)e^T(x^+s^+)} \\ &\leq \frac{\|w - x^+s^+\|^2}{4\min(w)\min(x^+s^+)}. \end{aligned}$$

By the previous inequality (2.6), we know $\min(x^+s^+) \geq \min(w)(1-\delta^2)$, so

$$\begin{aligned} \delta(x^+s^+, w)^2 &\leq \frac{\|w - x^+s^+\|^2}{4(1-\delta^2)\min(w)^2} \\ &= \frac{\|d_x d_s\|^2}{4(1-\delta^2)\min(w)^2}. \end{aligned}$$

Applying the inequality in lemma 2.5.1, then we find,

$$\begin{aligned}\delta(x^+s^+,w)^2 &\leq \frac{\|d_w\|^4}{32(1-\delta^2)\min(w)^2}\\ &= \frac{\delta^4}{2(1-\delta^2)},\end{aligned}$$

the result follows. □

From above, we can conclude that the algorithm with full Newton steps converges only if (x,s) is close enough to the target vector w. By applying theorem 2.5.1 and then restricting

$$\delta(x^+s^+,w) \leq \frac{\delta^2}{\sqrt{2(1-\delta^2)}} \leq \delta \quad \text{for convergence, and}$$
$$\delta(x^+s^+,w) \leq \frac{\delta^2}{\sqrt{2(1-\delta^2)}} \leq \delta^2 \quad \text{for quadratically convergence,}$$

we derive after simple calculations that

Algorithm with full Newton steps converges for $\delta \leq \sqrt{\frac{2}{3}}$

Algorithm with full Newton steps converges quadratically for $\delta \leq \frac{1}{\sqrt{2}}$

where δ denotes $\delta(xs,w)$ in every step.

- The quadratically convergent region for a specified target w is defined by

$$\left\{(x,s) : x \in SP,\ s = Mx+q,\ \delta(xs,w) \leq \frac{1}{\sqrt{2}}\right\}.$$

For larger values of $\delta(xs,w)$, we need a different analysis called damped Newton steps. Next we will suggest a algorithm for solving the embedding problem (2.3) and give an iteration bound for such algorithm.

2.6 An Interior Point Method

In this section, we give the algorithm with full Newton steps. This algorithm will search the optimal solution inside the feasible set. Therefore it is an *interior point method.*

2.6.1 Algorithm with Full Newton Steps

The following is the algorithm with full Newton steps:

Algorithm with full Newton Steps

`Declare` `accuracy parameter` $\varepsilon > 0$,
`barrier update parameter` $\theta \in (0, 1)$,
$x := e,\ s := e,\ \mu := 1$

`While` $n\mu \geq \varepsilon$
  $\mu := (1 - \theta)\mu$;
  `Solve` (2.5) `for` $\Delta x, \Delta s$ `by the Newton step`;
  $x := x + \Delta x$;
  $s := s + \Delta s$;
`End`

By a suitable choice of θ, the iterated x and $s = Mx + q$ are positive vectors. It is equivalent to say that x is in the interior of the feasible set.

2.6.2 Iteration Bound

We can bound the maximum number of iterations for this algorithm. To do, so we need the following lemma

Lemma 2.6.1 *Let* $(x, s) > 0$ *be feasible and* $\mu > 0$ *such that* $x^T s = n\mu$. *Moreover, let* $\delta := \delta(xs, \mu e)$ *and* $\mu^+ = (1 - \theta)\mu$. *Then*

$$(\delta^+)^2 := \delta(xs, \mu^+ e)^2 = (1 - \theta)\delta^2 + \frac{\theta^2 n}{4(1 - \theta)}.$$

Proof: Let $u := \sqrt{\frac{xs}{\mu e}}$. By the definition,

$$\delta = \frac{1}{2\sqrt{\min(\mu e)}} \left\| \frac{xs - \mu e}{\sqrt{xs}} \right\| = \frac{1}{2\sqrt{\mu}} \left\| \sqrt{\frac{xs}{e}} - \sqrt{\frac{\mu^2 e}{xs}} \right\| = \frac{1}{2} \|u - u^{-1}\|.$$

Since $\delta^+ = \delta(xs, \mu^+ e)$, it implies

$$\begin{aligned} 4(\delta^+)^2 &= \left\| \sqrt{1-\theta} u^{-1} - \frac{u}{\sqrt{1-\theta}} \right\|^2 \\ &= \left\| \sqrt{1-\theta}(u^{-1} - u) - \frac{\theta u}{\sqrt{1-\theta}} \right\|^2 . \end{aligned} \tag{2.7}$$

From $x^T s = n\mu$ it follows that $\|u\|^2 = n$. Hence, u is orthogonal to $u^{-1} - u$:

$$u^T(u^{-1} - u) = n - \|u\|^2 = 0.$$

Applying this orthogonal property to (2.7), we obtain

$$\begin{aligned} 4(\delta^+)^2 &= (1-\theta)\|u^{-1} - u\|^2 + \frac{\theta^2 \|u\|^2}{1-\theta}, \\ (\delta^+)^2 &= (1-\theta)\delta^2 + \frac{\theta^2 n}{4(1-\theta)}. \end{aligned}$$

The desired result follows. □

Now, we can prove the following iteration bound of the algorithm with full Newton steps.

Theorem 2.6.1 *If* $\theta = \dfrac{1}{2\sqrt{n}}$, *then the algorithm with full Newton steps requires at most*

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

iterations. The output is a feasible pair (x, s) *such that*

$$x^T s = n\mu \quad \textit{and} \quad \delta(xs, \mu e) \leq \frac{1}{\sqrt{2}}$$

for some $\mu \leq \dfrac{\varepsilon}{n}$.

Proof: We know $xs = \mu_0 e = e$, so $x^T s = n\mu_0$ and $0 = \delta(xs, \mu_0 e) \leq \dfrac{1}{\sqrt{2}}$. When we update $\mu^+ = (1-\theta)\mu_0$ with $\theta = \dfrac{1}{2\sqrt{n}}$, lemma 2.6.1 gives

$$\begin{aligned} \delta(xs, \mu^+ e)^2 &\leq \frac{1-\theta}{4} + \frac{1}{16(1-\theta)} \\ &\leq \frac{1}{4} + \frac{1}{8} = \frac{3}{8} < \frac{1}{2}. \end{aligned}$$

Hence, by theorem 2.5.1, this Newton step is feasible and $\delta(x^+s^+, \mu^+) \leq \frac{1}{2}$. As this Newton step is feasible, we know that $(x^+)^T s^+ = n\mu^+$. Now we repeat this process and want to find k which is the number of iterations such that

$$n\mu = n(1-\theta)^k \mu_0 = n(1-\theta)^k \leq \varepsilon.$$

Since,

$$n(1-\theta)^{\frac{1}{\theta}\log\frac{n}{\varepsilon}} \leq ne^{-\log\frac{n}{\varepsilon}} = \varepsilon.$$

By eliminating k, we find

$$k \leq \left\lceil 2\sqrt{n}\log\frac{n}{\varepsilon} \right\rceil.$$

This proves the theorem. □

In conclusion, the full Newton steps algorithm with $\theta = \frac{1}{2\sqrt{n}}$ proceeds in the following way:

Given an accuracy parameter $\varepsilon > 0$, an update parameter $\theta = \frac{1}{2\sqrt{n}}$ and $\mu^+ = (1-\theta)\mu$.

Step I:	$x^T s = n\mu, \quad \delta(xs, \mu e) < \frac{1}{\sqrt{2}}$
Step II:	Since $\delta(xs, \mu^+ e) \leq \frac{1}{\sqrt{2}} \leq 1$ where $\mu^+ = (1-\theta)\mu$, the Newton step is feasible. Solve for $\Delta x, \Delta s$. Set $x^+ = x + \Delta x$ and $s^+ = s + \Delta s$.
Step III:	$\delta(x^+s^+, \mu^+ e) \leq \frac{1}{2} < \frac{1}{\sqrt{2}}$, since feasible $(x^+)^T s^+ = n\mu^+$ Goto Step II until $n\mu < \varepsilon$

In each iteration, the Newton step is feasible so we will go along this direction $(x^+ = x + \Delta x)$ and on the other hand $(x^+)^T s^+ = n\mu^+$. We repeat this process until $x^T s = n\mu$ for some $\mu \leq \frac{\varepsilon}{n}$. If ε is small enough, $xs < n\mu e < \varepsilon e \approx 0$ at the end of iterations. Here, the algorithm with full Newton steps is quadratically convergent since $\delta \leq \frac{1}{\sqrt{2}}$ is preserved.

2.7 Background and Rounding Procedure for Interior-point Solution

After performing the algorithm with full Newton steps, we find a feasible pair $(x, s(x)) > 0$ such that $xs(x) < \varepsilon e$. As our goal is to find a strictly complementary solution, we need a rounding procedure in order to round off $(x, s(x))$ to an exact solution. In the following analysis of the rounding procedure, we have to separate the 'large' and 'small' entries of $x(\mu)$. We first define the *condition number* which will be used in later separation.

- The *condition number* of (SP) is defined by

$$\sigma := \min\{\sigma^x, \sigma^s\}$$

where

$$\sigma^x := \min_{i\in B} \max_{x\in SP^*} \{x_i\}, \qquad \sigma^s := \min_{i\in N} \max_{x\in SP^*} \{s(x)_i\}.$$

The condition number can also be written as

$$\sigma := \min_{i\in B\cup N} \max_{x\in SP^*} \{x_i + s(x)_i\}.$$

Since for any $x \in SP^*$, $xs(x) = 0$, we have

$$x_i = x_i + s(x)_i,\ i \in B, \qquad s(x)_i = x_i + s(x)_i,\ i \in N.$$

So the condition number

$$\begin{aligned} \sigma &= \min\left\{\min_{i\in B} \max_{x\in SP^*} \{x_i + s(x)_i\},\ \min_{i\in N} \max_{x\in SP^*} \{x_i + s(x)_i\}\right\} \\ &= \min_{i\in B\cup N} \max_{x\in SP^*} \{x_i + s(x)_i\}. \end{aligned}$$

We use the previous example to make clear the concept of the condition number.

Let $z := (s_1, s_3, x, \kappa, \vartheta)^T$. Also suppose that we know a strictly complementary solution $z = (1.6, 0, 0.8, 0.8, 0)$ with $s(z) = (0, 0.8, 0, 0, 1)$. Therefore, the optimal partition is

$$B = \{s_1, x, \kappa\}, \qquad N = \{s_3, \vartheta\}.$$

For any $z \in SP$, z is optimal if and only if $s_3 = \vartheta = 0$ and $s(s_1) = s(x) = s(\kappa) = 0$. This gives

$$\begin{aligned} x - \kappa &= 0, \\ -s_1 + 2\kappa &= 0, \\ s_1 - 2x &= 0. \end{aligned}$$

Thus the optimal solution is of the form

$$z = (2\kappa, 0, \kappa, \kappa, 0)$$

and

$$s(z) = Mz + q = (0, \kappa, 0, 0, -5\kappa + 5).$$

Since, all these values should be nonnegative. This occur if and only if $0 \leq \kappa \leq 1$. Hence the maximal values of the variables in B are $\{2, 1, 1\}$, and their minimum is 1. The maximal surplus values for the variables in N are $\{1, 5\}$ and its minimum is 1. So the condition number is 1 which is the minimum value of these two *min_max* values. In the above illustration, we can find the condition number if a strictly complementary solution is given. But in general, to calculate the condition number of (SP) is as difficult as to solve the problem itself. In the next lemma, we can find a lower bound for σ if the matrix M is integral.

Lemma 2.7.1 *The condition number σ of (SP) with integral matrix M satisfies*

$$\sigma \geq \frac{1}{\prod_{j=1}^{n} \|M_j\|}$$

where M_j denotes the j-th column of M.

Proof: See [14], P.62-64 □

We observe from the proof in [14] that the lower bound of σ is not so sharp.

- For any $x \in SP$, we define a function $\varphi(x)$ by

$$\varphi(x) := \frac{\max(xs(x))}{\min(xs(x))}$$

  which will be used later.

The following lemma gives the upper bound and lower bound for 'small' and 'large' entries, respectively, for $x(\mu)$ and $s(\mu)$. Meanwhile, we can observe from the lemma that the upper bound for the small coordinates is of magnitude $O(\mu)$.

Lemma 2.7.2 *Let $x \in SP$ such that $\varphi(x) \leq \tau$. Then we have*

$$x_i \geq \frac{\sigma}{\tau n},\ i \in B, \qquad x_i \leq \frac{x^T s}{\sigma},\ i \in N,$$
$$s_i \leq \frac{x^T s}{\sigma},\ i \in B, \qquad s_i \geq \frac{\sigma}{\tau n},\ i \in N$$

where $s := s(x)$.

Proof: From the definition of $\varphi(x)$, we can find two positive numbers τ_1, τ_2 such that $\tau = \frac{\tau_2}{\tau_1}$ and $\tau_1 \leq x_i s_i \leq \tau_2$. Let $i \in N$ and let $\tilde{x}$ be an optimal solution of (SP) such that $\tilde{s}_i := s_i(\tilde{x})$ is maximal. By considering,

$$(x - \tilde{x})^T (s - \tilde{s}) = 0,$$

we get,

$$x^T \tilde{s} + \tilde{x}^T s = x^T s.$$

Since $(x, s) \geq 0$ and $(\tilde{x}, \tilde{s}) \geq 0$, it implies

$$x_i \tilde{s}_i \leq x^T \tilde{s} \leq x^T s.$$

By the definition of condition number that $\tilde{s}_i \geq \sigma$, we first obtain

$$x_i \leq \frac{x^T s}{\tilde{s}_i} \leq \frac{x^T s}{\sigma}.$$

Meanwhile, using $x_i s_i \geq \tau_1$ and $x^T s \leq n\tau_2$, we get

$$s_i \geq \frac{\tau_1}{x_i} \geq \frac{\tau_1 \sigma}{x^T s} \geq \frac{\tau_1 \sigma}{n\tau_2} = \frac{\sigma}{n\tau}.$$

Similarly, we can prove

$$x_i \geq \frac{\sigma}{\tau n}, \; i \in B, \qquad s_i \leq \frac{x^T s}{\sigma}, \; i \in B.$$

This completes the proof. □

The above lemma has the following important consequence.

Corollary 2.7.1 *If a point $x(\mu)$ on the central path is given so that*

$$\mu < \frac{\sigma^2}{n^2}$$

then we can determine the optimal partition (B, N) of (SP).

Proof: Let $x(\mu)$ be on the central path. Then $x(\mu) \in SP$ and $x(\mu)s(\mu) = \mu e$. Thus $\varphi(x(\mu)) = 1$. We summarize the result of lemma 2.7.2 in the following table

	$i \in B,$	$i \in N$
$x_i(\mu)$	$\geq \frac{\sigma}{n},$	$\leq \frac{n\mu}{\sigma}$
$s_i(\mu)$	$\leq \frac{n\mu}{\sigma},$	$\geq \frac{\sigma}{n}$

Therefore, we have a complete separation of the small and large variables if

$$\frac{n\mu}{\sigma} \leq \frac{\sigma}{n}.$$

It is equivalent to

$$\mu \leq \frac{\sigma^2}{n^2}.$$

That means if a point $x(\mu)$ on the central path such that μ is small enough, less than $\frac{\sigma^2}{n^2}$, then we can determine the optimal partition (B, N) of (SP). □

From corollary 2.7.1, we can find μ by applying the lower bound of σ in lemma

2.7.1 in order to form the optimal partition which is useful in the following rounding procedure.

Assuming that we have a solution pair $(x, s(x)) > 0$ and the optimal partition (B, N) has been determined, we can always rewrite the system of equations in (SP) as

$$\begin{pmatrix} s_B \\ s_N \end{pmatrix} = \begin{pmatrix} M_{BB} & M_{BN} \\ M_{NB} & M_{NN} \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} + \begin{pmatrix} q_B \\ q_N \end{pmatrix}. \tag{2.8}$$

Here $q_B = 0$, since $0 = q^T x = q_B^T x_B$ where $x_B > 0$. Now considering the vector $\bar{x}_B$ and $\bar{x}_N$ defined by

$$\bar{x}_B = x_B - \xi, \qquad \bar{x}_N = 0$$

where ξ is any solution of

$$M_{BB}\xi = s_B - M_{BN}x_N = M_{BB}x_B. \tag{2.9}$$

The solution ξ exists because M_{BB} is singular. M_{BB} is singular since (2.3) satisfies the *IPC* which implies the existence of strictly complementary solution. Let's say

$$\begin{pmatrix} \tilde{x}_B > 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \tilde{s}_N > 0 \end{pmatrix}.$$

Substituting into (2.8), it gives $M_{BB}\tilde{x}_B = 0$ and so M_{BB} is singular. Now equation (2.8) becomes

$$\begin{pmatrix} 0 \\ \bar{s}_N \end{pmatrix} = \begin{pmatrix} M_{BB} & M_{BN} \\ M_{NB} & M_{NN} \end{pmatrix} \begin{pmatrix} \bar{x}_B \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ q_N \end{pmatrix}$$

since $\bar{s}_B = M_{BB}\bar{x}_B = M_{BB}(x_B - \xi) = 0$. Thus $\bar{x}$ and $\bar{s}$ are clearly complementary. However, it gives no information on the positivity of

$$\begin{aligned} \bar{x}_B &= x_B - \xi, \\ \bar{s}_N &= M_{NB}\bar{x}_B + q_N = M_{NB}(x_B - \xi) + q_N \\ &= s_N - M_{NN}x_N - M_{NB}\xi. \end{aligned} \tag{2.10}$$

Next we show that $\bar{x}_B$ and $\bar{s}_N$ in (2.10) are positive vectors if we solve (2.9) for ξ by Gaussian elimination. That is to say, $(\bar{x}, \bar{s})$ gives a strictly complementary solution to (SP).

In the rounding procedure, we need to solve $M_{BB}\xi = M_{BB}x_B$ for ξ. The solution of following minimization problem

$$\min\{\|x_B^{-1}\xi\| : \ M_{BB}\xi = M_{BB}x_B\},$$

which is equivalent to

$$\min\{\|\eta\| : \ M_{BB}X_B\eta = M_{BB}x_B\},$$

solves our problem in which ξ is given by

$$\xi = X_B\eta = X_B(M_{BB}X_B)^+M_{BB}x_B.$$

where $(M_{BB}X_B)^+$ is the generalized inverse of $M_{BB}X_B$. For a square linear system $A_nx = b$ with $rank(A_n) = r < n$, matrix A can be decomposed as $A_n = A_{n\times r}A_{r\times n}$ where these two matrices have rank r (for example the LU decomposition where L is a lower triangular matrix and U is an upper triangular matrix, see [29] for details). We first obtain $A_{r\times n}x = (A_{n\times r}^TA_{n\times r})^{-1}A_{n\times r}^Tb$ in which $A_{n\times r}$ is nonsingular since $A_{n\times r}$ has full column rank. As we want to find a least square solution x, we know that such a solution is orthogonal to the null space of $A_{r\times n}$. Therefore, x belongs to the row space of $A_{r\times n}$ and hence there exists $y \in \mathbb{R}^r$ such that $x = A_{r\times n}^Ty$. Now, $A_{r\times n}A_{r\times n}^Ty = (A_{n\times r}^TA_{n\times r})^{-1}A_{n\times r}^Tb$ and here $A_{r\times n}A_{r\times n}^T$ is invertible. As a result, we get $x = A_{r\times n}^Ty = A_{r\times n}^T(A_{r\times n}A_{r\times n}^T)^{-1}(A_{n\times r}^TA_{n\times r})^{-1}A_{n\times r}b$. We define the generalized inverse $A^+ = A_{r\times n}^T(A_{r\times n}A_{r\times n}^T)^{-1}(A_{n\times r}^TA_{n\times r})^{-1}A_{n\times r}^T$.

Actually, if the algorithm with full Newton steps create a feasible pair (x, s), in which the distance between the small and large variable is sufficiently large,

satisfies

$$x^T s(x) \leq \frac{\sigma_{SP}^2}{4n\omega^2 |B|^{\frac{3}{2}} \pi_B},$$

where $\omega := \|M\|_\infty$ and π_B is the product of 2-norm of the nonzero column in M_{BB} then we can obtain a solution ξ of (2.9) by Gaussian elimination in $O(|B^*|^3)$ arithmetical operations where B^* denotes the subset of B with nonzero column in M_{BB}. Then this guarantees

$$\bar{x}_B = x_B - \xi > 0, \qquad \bar{s}_N = s_N - M_{NN} x_N - M_{NB} \xi > 0.$$

We illustrate the algorithm with full Newton steps by some linear optimization problems.

2.8 Solving Some LP problems

In this section, we first summarize the procedure of setting up the (SP), which satisfies the *IPC*, from a general LO problem:

$$(P) \quad \begin{cases} \min & c^T x \\ s.t. & Ax \geq b \\ & x \geq 0 \end{cases}$$

where

$$c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, \; x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \; b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \; A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

We then construct the self-dual embedding:

$$(SD1)\quad \begin{cases} \min & \begin{pmatrix} 0_m \\ 0_n \\ 0 \end{pmatrix}^T \begin{pmatrix} y \\ x \\ \kappa \end{pmatrix} \\ s.t. & \begin{pmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{pmatrix} \begin{pmatrix} y \\ x \\ \kappa \end{pmatrix} \geq \begin{pmatrix} 0_m \\ 0_n \\ 0 \end{pmatrix} \\ & y \geq 0_m,\ x \geq 0_n,\ \kappa \geq 0 \end{cases}$$

Let $M = \begin{pmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{pmatrix}$ and $r = e_{m+n+1} - Me_{m+n+1}$. Then we can further embed $(SD1)$ into

$$(SD2)\quad \begin{cases} \min & \begin{pmatrix} 0_m \\ 0_n \\ 0 \\ n+1 \end{pmatrix}^T \begin{pmatrix} y \\ x \\ \kappa \\ \vartheta \end{pmatrix} \\ s.t. & \begin{pmatrix} M & r \\ -r & 0 \end{pmatrix} \begin{pmatrix} 0_{m+n+1} \\ \vartheta \end{pmatrix} + \begin{pmatrix} 0_{m+n+1} \\ n+1 \end{pmatrix} \geq 0 \\ & y \geq 0_m,\ x \geq 0_n,\ \kappa \geq 0,\ \vartheta \geq 0 \end{cases}$$

which satisfies the *IPC* automatically. Now let

$$\overline{M} = \begin{pmatrix} M & r \\ -r & 0 \end{pmatrix}, \quad z = \begin{pmatrix} y \\ x \\ \kappa \\ \vartheta \end{pmatrix}, \quad q = \begin{pmatrix} 0_m \\ 0_n \\ 0 \\ n+1 \end{pmatrix},$$

and solve

$$\begin{cases} \overline{M}z^{(i)} + q &= s^{(i)} \\ z^{(i)}s^{(i)} &= \mu e \end{cases}$$

for $\mu \to 0$ by applying the Newton steps sequentially in the following algorithm:

Algorithm with full Newton Step

`Declare` $\varepsilon > 0$,

$$\theta = \frac{1}{2\sqrt{n}} \in (0,1),$$

$$z := e,\ s := e,\ \mu := 1$$

`Do while` $n\mu \geq \varepsilon$

$\mu := (1-\theta)\mu$;

$\Delta z = (S + Z\overline{M})^{-1}(\mu e - zs)$;

$z = z + \Delta z$;

$s = s + \overline{M}\Delta z$;

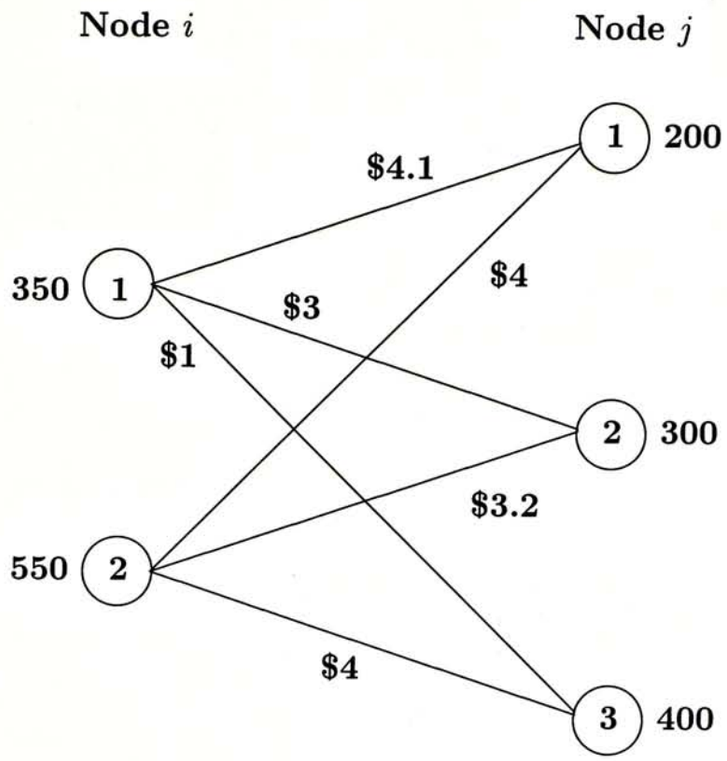
`End`

Then the solution of (P) is given by the $(m+1)^{th}$ to $(n+m)^{th}$ entries of the iterated z in the above algorithm.

We now illustrate the algorithm with full Newton steps for some linear optimization problems.

1. **A Transportation problem**

   Let i denote the node which represent a cannery at San Diego for $i = 1$, at Seattle for $i = 2$. Let j denote the node which represent the warehouses at New York for $j = 1$, at Chicago for $j = 2$, at San Francisco for $j = 3$.



Let x_{ij} denote the number of cases which are sent from node i to j.
We have some restrictions on this optimal shipping problem.

- The unit sending costs form node i to node j are showed on the edge connecting the corresponding nodes.
- Each cannery sends no more cases than it produces.
- Each warehouse can obtain as many cases as it can sell.

We can set up the following LP to minimize the total shipping cost:

$$\begin{aligned}
\min \quad & z = 4.1x_{11} + 3x_{12} + x_{13} + 4x_{21} + 3.2x_{22} + 1.3x_{23} \\
s.t. \quad & \begin{cases}
x_{11} + x_{12} + x_{13} \leq 350 \\
x_{21} + x_{22} + x_{23} \leq 550 \\
x_{11} + x_{21} \geq 200 \\
x_{12} + x_{22} \geq 300 \\
x_{13} + x_{23} \geq 400 \\
x_{ij} \geq 0 \qquad i = 1,2,\ j = 1,2,3.
\end{cases}
\end{aligned}$$

We then write it in the form of (P), i.e.

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

and further embed it into $(SD2)$ so that we can apply the algorithm with full Newton steps in section 2.8.

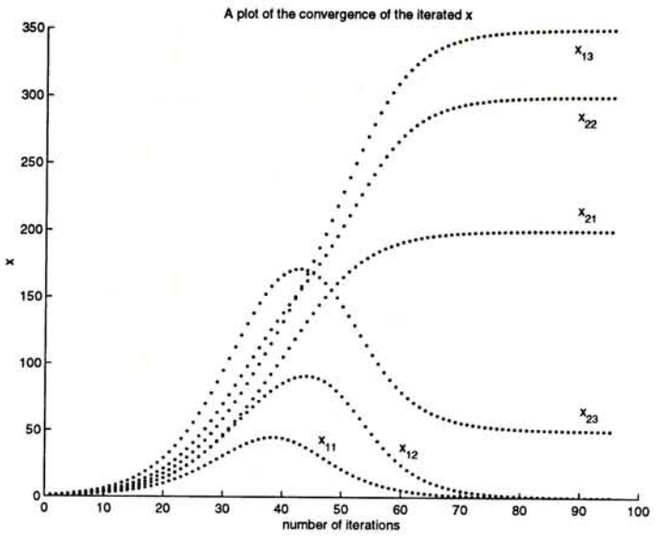
The following table shows the number of iterations required, the errors in 2-norm and maximum norm for some given ε.

The optimal solution of the problem is $x^* = (0, 0, 350, 200, 300, 50)$, and the optimal cost is 2175.

ε	number of iterations	$\|x - x^*\|_2$	$\|x - x^*\|_\infty$
1E-3	64	32.4811	19.1360
1E-5	95	0.2939	0.1749
1E-7	126	2.9E-3	1.7E-3
1E-10	172	3.0E-6	1.8E-6

For $\varepsilon = 10^{-10}$, the numerical optimal solution of the problem is $x = (0.0000,$ 0.0000, 350.0000, 200.0000, 300.0000, 50.0000), and the optimal shipping cost is 2174.999998.

The following graph shows the convergence of the solution. The x-axis shows the number of iterations. When the number of iterations increase, each entry of the iterated solution x converge to the corresponding entry in the true solution x.



2. **A blending problem**

The following table shows the percentage of lead, zinc and tin and also the cost of 9 different alloys.

Alloy i		1	2	3	4	5	6	7	8	9
Percentage of	lead	10	10	40	60	30	30	30	50	20
	zinc	10	30	50	30	30	40	20	40	30
	tin	80	60	10	10	40	30	50	10	50
Cost per *lb.*		4.1	4.3	5.8	6.0	7.6	7.5	7.3	6.9	7.3

Let x_i be the quality of alloy i to be purchased. We want to set up a LP to find a combination of alloys in order to produce $100lbs$ of an alloy which is 30% lead, 30% zinc and 40% tin such that the cost is minimum. The LP is

as below:

$$\min\ z = 4.1x_1 + 4.3x_2 + 5.8x_3 + 6x_4 + 7.6x_5 + 7.5x_6 + 7.3x_7 + 6.9x_8 + 7.3x_9$$

$$s.t. \begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 = 100 \\ .1x_1 + .1x_2 + .4x_3 + .6x_4 + .3x_5 + .3x_6 + .3x_7 + .5x_8 + .2x_9 = 30 \\ .1x_1 + .3x_2 + .5x_3 + .3x_4 + .3x_5 + .4x_6 + .2x_7 + .4x_8 + .3x_9 = 30 \\ .8x_1 + .6x_2 + .1x_3 + .1x_4 + .4x_5 + .3x_6 + .5x_7 + .1x_8 + .5x_9 = 40 \\ x_i \geq 0 \qquad i = 1, 2, \cdots, 9. \end{cases}$$

We then write it in the form of (P), i.e.

$$\begin{aligned} \min_x \quad & c^T x \\ s.t. \quad & Ax \geq b, \\ & x \geq 0, \end{aligned}$$

and further embed it into $(SD2)$ so that we can apply the algorithm with full Newton steps.

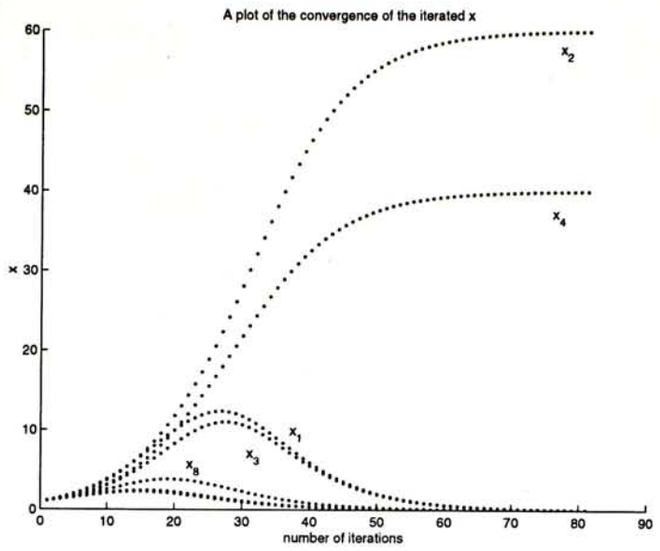
The following table shows the number of iterations required, the errors in 2-norm and maximum norm for some given ε.

The optimal solution of the problem is $x^* = (0, 60, 0, 40, 0, 0, 0, 0, 0)$, and optimal cost is 498.

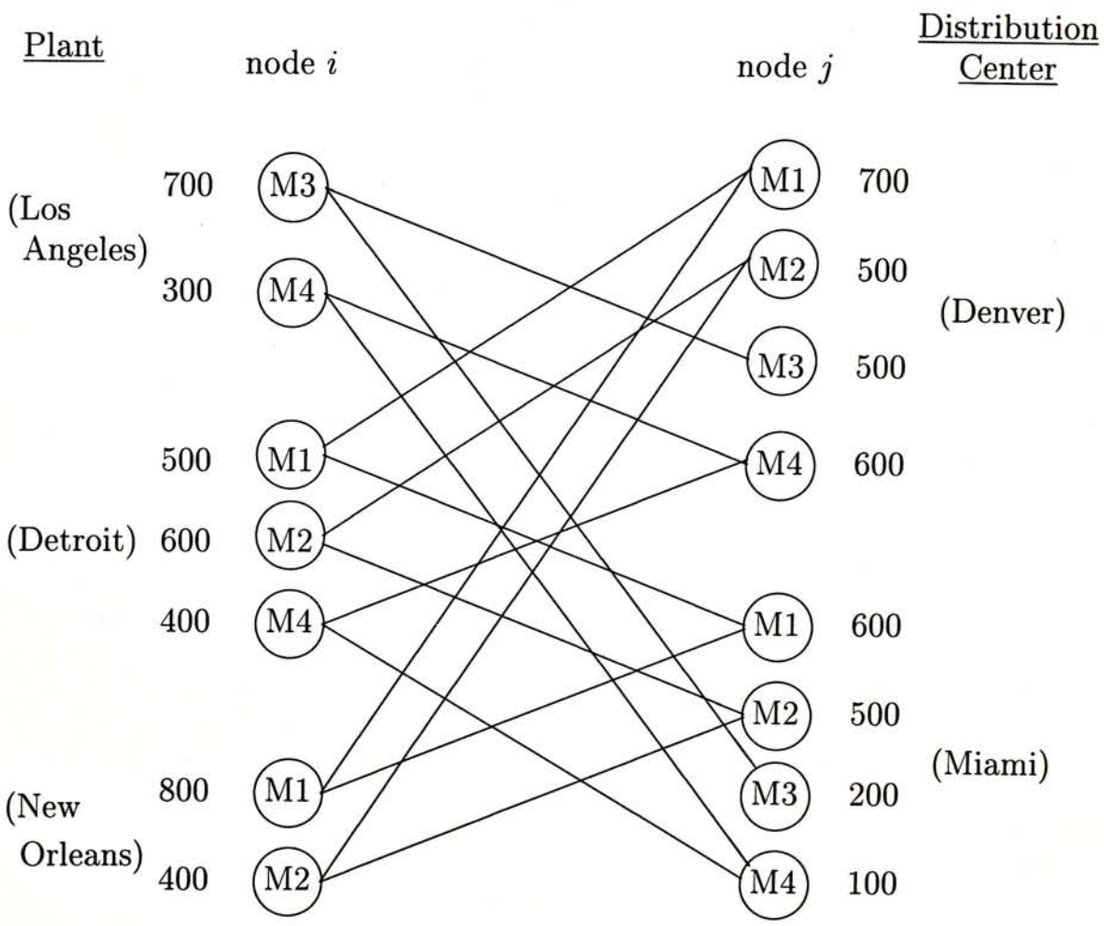
ε	number of iterations	$\|x - x^*\|_2$	$\|x - x^*\|_\infty$
1E-2	62	1.4033	1.0819
1E-3	81	0.1400	0.1079
1E-5	119	1.4E-3	1.1E-3
1E-7	157	1.3E-5	1.0E-5

For $\varepsilon = 10^{-7}$, the numerically found solution is $x = (0.0000,\ 60.0000,\ 0.0000,\ 40.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000)$, and optimal cost is 497.999979.

Similar to example 1, the following graph shows the convergence of the solution.



3. Multicommodity Transportation Model



The train transportation cost per car between the plant and distribution center is:

		Distribution center	
		Denver	Miami
Plant	Los Angeles	80	215
	Detroit	100	108
	New Orleans	102	68

To minimize the transportation cost, we set up the following LP:

$$\begin{aligned}
\min\ z \quad = \quad & 80(x_{13}+x_{24})+215(x_{17}+x_{28})+100(x_{31}+x_{42}+x_{54}) \\
& +108(x_{35}+x_{46}+x_{58})+102(x_{61}+x_{72})+68(x_{65}+x_{76}) \\
s.t. \quad & \begin{cases}
x_{13}+x_{17} \le 700 & x_{31}+x_{61} \ge 700 \\
x_{24}+x_{28} \le 300 & x_{42}+x_{71} \ge 500 \\
x_{31}+x_{35} \le 500 & x_{13} \ge 500 \\
x_{42}+x_{46} \le 600 & x_{24}+x_{54} \ge 600 \\
x_{54}+x_{58} \le 400 & x_{35}+x_{65} \ge 600 \\
x_{61}+x_{65} \le 800 & x_{46}+x_{76} \ge 500 \\
x_{72}+x_{76} \le 400 & x_{17} \ge 200 \\
 & x_{28}+x_{58} \ge 100
\end{cases} \\
& x_{ij} \ge 0
\end{aligned}$$

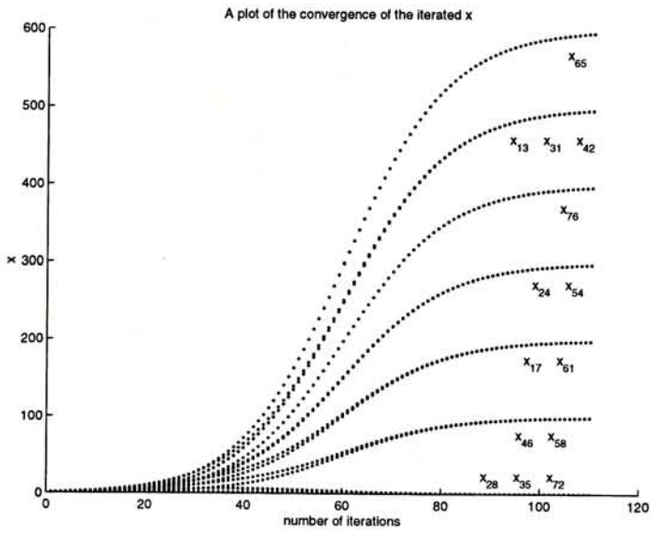
We then write it in the form of (P) and further embed it into $(SD2)$ so that we can apply the algorithm with full Newton steps.

The optimal solution is $x^* = (500,\ 200,\ 300,\ 0,\ 500,\ 0,\ 500,\ 100,\ 300,\ 100,\ 200,\ 600,\ 0,\ 400)$, the optimal cost is 347000.

The following table shows the number of iterations required, the errors in 2-norm and maximum norm for some given ε.

ε	number of iterations	$\|x - x^*\|_2$	$\|x - x^*\|_\infty$
1E-3	110	10.7786	5.2132
1E-5	159	0.1081	5.2E-2
1E-7	208	1.1E-3	5.2E-4
1E-10	282	1.02E-6	4.9E-7

For $\varepsilon = 10^{-10}$, the numerical found solution is x = (500.0000, 200.0000, 300.0000, 0.0000, 500.0000, 0.0000, 500.0000, 100.0000, 300.0000, 100.0000, 200.0000, 600.0000, 0.0000, 400.0000), the optimal cost is 346999.999719. The following graph shows the convergence of the solution.



2.9 Remarks

We can summarize the results of the three illustrated LP problems in the following table.

Example	*Size of A*	n	ε	*Number of iterations*	$\|x - x^*\|_2$
1	$A_{5\times6}$	6	1E-10	172	3.0E-6
2	$A_{8\times9}$	9	1E-7	157	1.3E-5
3	$A_{15\times14}$	14	1E-10	282	1.02E-6

We observe from the table that although the number of variables n in problem 3 is more than twice of that in problem 1, the number of iterations in problem 3 is less than twice of that in problem 1. That is to say the number of iterations is not proportional to the number of variables. However, this does not mean we have gained something here. The cost of calculating Δz increases very fast with the increasing of the number of variables n.

When we apply the algorithm with full Newton steps to any LP problem, we prefer to set ε =1E-10. By this choice, the algorithm will give a numerical solution with $\|x - x^*\|_2$ <1E-5 in general. By observing the graphs, although initially the numerical solution converges not so good, the iteration becomes very stable afterwards. The reason of this unstability can be explained by starting the algorithm with all one vector.

We also remark here that we have no restriction on the application of the algorithm with full Newton steps to any LP. The first example show that LP with equality constraints can also be solved by rewriting into the form (P).

In next chapter, we will study two iterative methods for solving convex optimization problems.

Chapter 3

Convex Optimization

In this chapter, we consider the *interior point methods* for solving the *convex optimization problem.*

3.1 Introduction

3.1.1 Convex Optimization Problem

The CO problem to be discussed is a problem of the following type

$$(CP) \quad \begin{cases} \max & f_0(y) \\ s.t. & f_i(y) \leq 0, \quad i = 1, 2, \cdots, n. \end{cases}$$

where $-f_0(y)$ and $f_i(y), 1 \leq i \leq n$ are convex functions and $y \in \mathbb{R}^m$. The feasible set of (CP) is denoted by PF, and the interior of this set by PF^o. The dual problem of (CP) formulated by Wolfe [10] is given by

$$(CD) \quad \begin{cases} \min_{x,y} & f_0(y) - \sum_{i=1}^n x_i f_i(y) \\ s.t. & \sum_{i=1}^n x_i \bigtriangledown f_i(y) = \bigtriangledown f_0(y), \\ & x_i \geq 0, \qquad i = 1, 2, \cdots, n. \end{cases}$$

We denote the feasible set of (CD) by DF and the interior of DF by DF^o.

In our analysis, we need the following assumptions:

- $-f_0(y)$ and $f_i(y)$, $1 \leq i \leq n$ are twice continuously differentiable.
- PF^0 is a nonempty bounded set.

The next two theorem give some relations between the prime and dual problems. First we prove that the primal objective function value is always bounded by the dual objective function value.

Theorem 3.1.1 *If $\bar{y} \in PF$ and $(x, y) \in DF$, then*

$$f_0(\bar{y}) \leq f_0(y) - \sum_{i=1}^{n} x_i f_i(y).$$

Proof: Since $-f_0(y)$ and $f_i(y)$, $1 \leq i \leq n$ are convex functions, we have

$$\begin{aligned} -f_0(y) &\leq -f_0(\bar{y}) - (y - \bar{y})^T \bigtriangledown f_0(y), \\ f_i(y) &\leq f_i(\bar{y}) + (y - \bar{y})^T \bigtriangledown f_i(y), \quad i = 1, 2, \cdots, n. \end{aligned}$$

By using these inequalities and the feasibility of solutions, we find

$$\begin{aligned} -f_0(y) + \sum_{i=1}^{n} x_i f_i(y) &\leq \left[-f_0(\bar{y}) - (y - \bar{y})^T \bigtriangledown f_0(y)\right] \\ &\quad + \left[\sum_{i=1}^{n} x_i f_i(\bar{y}) + \sum_{i=1}^{n} x_i (y - \bar{y})^T \bigtriangledown f_i(y)\right] \\ &= -f_0(\bar{y}) + \sum_{i=1}^{n} x_i f_i(\bar{y}). \\ &\leq -f_0(\bar{y}), \end{aligned}$$

the result follows. □

Theorem 3.1.2 *If PF^o is nonempty, then (CD) has an optimal solution and the optimal values of (CP) and (CD) are equal.*

Proof: We will give the proof in section 3.2.1. □

3.1.2 Idea of Interior Point Method

We will give two interior point methods to solve (CP) later. We now introduce the idea of the two interior point methods. These two methods solve a sequence of functions (say $\{\phi_n(y)\}_{n=1,2,...}$) approximately. Each $\phi_n(y)$ are associated with (CP). They involve the logarithm function to prevent the iterates from going outside the feasible set PF as n increase. On the other hand, the primal objective function $f_0(y)$ is minimized since its role becomes important in the sequence of functions as n increase . These two properties are the main reason that the two algorithm work well for solving (CP).

In next section, we will give the first algorithm which solves (CP) by a sequence of unconstrained minimizations.

3.2 Logarithmic Barrier Method

3.2.1 Basic Concepts and Properties

The *logarithmic barrier function* (see [24], [25]) associated with (CP) is defined by

$$\phi(y,\mu) = -\frac{f_0(y)}{\mu} - \sum_{i=1}^{n} \ln(-f_i(y)). \tag{3.1}$$

In the later proposed *Logarithm Barrier Algorithm*, we minimize a sequence of barrier function $\phi(y,\mu)$ approximately for decreasing μ from ∞ to 0. We observe from the definition of $\phi(y,\mu)$ that the singularity of the algorithm at zero will prevent the iterates from going outside the feasible set PF. On the other hand, the primal objective function $f_0(y)$ is minimized since its role becomes important in $\phi(y,\mu)$ as μ decreasing from ∞ to 0. Therefore, the algorithm seems to work well for solving (CP). We will give the proof later (see Theorem 3.1.2).

Before stating the algorithm, we introduce some terminology and useful definitions. The first and second order derivatives of $\phi(y,\mu)$ are given by

- $g(y,\mu) := \nabla\phi(y,\mu) = -\dfrac{\nabla f_0(y)}{\mu} + \sum_{i=1}^{n} \dfrac{\nabla f_i(y)}{-f_i(y)}$,

- $H(y,\mu) := \nabla^2\phi(y,\mu) = -\dfrac{\nabla^2 f_0(y)}{\mu} + \sum_{i=1}^{n} \left[\dfrac{\nabla^2 f_i(y)}{-f_i(y)} + \dfrac{\nabla f_i(y) \nabla f_i(y)^T}{f_i(y)^2}\right]$.

We will write $H = H(y,\mu)$ and $g = g(y,\mu)$ when there is no confusion caused.

We will need to measure the distance between any two points, to do so, we define a norm as follows:

$$\|z\|_H = \sqrt{z^T H z}$$

where $H := H(y,\mu)$. We observe that $H(y,\mu)$ is only positive semidefinite due to the convexities of $-f_0(y)$ and $f_i(y)$, $i = 1, 2, \cdots, n$, so $\|\cdot\|_H$ cannot define a norm in general. However, we will prove in lemma 3.2.2 that $H(y,\mu)$ is positive definite on PF^o if (CP) satisfies the so-called *self-concordance condition.* Therefore in such case $\|\cdot\|_H$ defines a norm.

For each μ, the minimizer $y(\mu)$ of $\phi(y,\mu)$ is called *the μ-center.* It is defined uniquely because $\phi(y,\mu)$ is strictly convex in PF^o and go to infinity on the boundary of PF. The fact that $\phi(y,\mu)$ is strictly convex will be proved in lemma 3.2.2 by adding one more assumption on $\phi(y,\mu)$.
We have

Theorem 3.2.1 *The necessary and sufficient Karush-Kuhn-Tucker (KKT) conditions for $y(\mu)$ to be a minimizer of $\phi(y,\mu)$ are as follows:*

$$\begin{aligned} f_i(y) \le 0,\ x \ge 0, &\qquad 1 \le i \le n \\ \sum_{i=1}^{n} x_i \nabla f_i(y) = \nabla f_0(y), & \\ -f_i(y)x_i = \mu, &\qquad 1 \le i \le n. \end{aligned} \tag{3.2}$$

Proof: We first show the necessary KKT conditions. For a given $\mu > 0$, let $y(\mu)$ be the minimizer of $\phi(y,\mu)$. By the first-order necessary conditions for

unconstrained minimization (cf.[1]),

$$\nabla\phi(y(\mu),\mu) = 0.$$

That implies

$$\nabla f_0(y(\mu)) = \sum_{i=1}^{n} \frac{-\mu}{f_i(y(\mu))} \nabla f_i(y(\mu)).$$

Let $x_i = \dfrac{-\mu}{f_i(y(\mu))}, 1 \leq i \leq n$. By the feasibility of $y(\mu)$, each $x_i > 0$, and so we have proved the necessary KKT conditions.

We now prove the sufficient KKT conditions. Substituting the KKT conditions into the gradient of $\phi(y,\mu)$, we get

$$\nabla\phi(y,\mu) = -\frac{\nabla f_0(y)}{\mu} + \sum_{i=1}^{n} \frac{\nabla f_i(y)}{-f_i(y)} = 0.$$

In addition, since $\nabla^2\phi(y,\mu)$ is positive definite, we always have $z^T\nabla^2\phi(y,\mu)z > 0$, for all nonzero $z \in \mathbb{R}^m$. Combining with (3.2), the second-order sufficiency conditions for unconstrained minimization (cf.[1]) are satisfied, i.e. y is a minimizer of $\phi(y,\mu)$. □

Now we introduce some terminology which are needed in the later analysis.

- We define the *primal central path* as $\{y(\mu) : \mu > 0\}$.
- We also define the *dual central path* as $\{(x(\mu), y(\mu)) : \mu > 0\}$ where $x(\mu)$ is defined in the KKT conditions. From the definition of $x(\mu)$, $(x(\mu), y(\mu))$ is dual feasible.

We now prove Theorem 3.1.2 which gives us more information on the relation between (CP) and (CD).

Proof of Theorem 3.1.2: (CP) has optimal solutions since PF^0 is nonempty. From [1], we know that $x(\mu)$ and $y(\mu)$ are continuously differentiable. Hence,

$(x(\mu), y(\mu))$ converges to some point (x^*, y^*) as $\mu \to 0$. Since the *duality gap*:

$$\begin{aligned} f_0(y(\mu)) - \sum_{i=1}^{n} x_i(\mu) f_i(y(\mu)) - f_0(y(\mu)) &= -\sum_{i=1}^{n} x_i(\mu) f_i(y(\mu)) \\ &= n\mu \to 0 \quad as\ \mu \to 0, \end{aligned}$$

the second-order sufficient conditions (cf.[1]) holds. Then y^* is an optimal solution of (CP). As their duality gap is zero, by theorem 3.1.1, (x^*, y^*) is an optimal solution of (CD) and its optimal value is equal to the primal optimal value. □

Now we prove the monotonicity of the primal and dual objectives along the central path under the following weak assumptions:

- $-f_0(y)$ and $f_i(y)$, $1 \le i \le n$ are continuously differentiable.
- PF^0 is a nonempty bounded set.

Before stating the theorem, we introduce the logarithmic barrier function for (CD) up to a constant:

$$\phi_d(x, y, \mu) = -\frac{f_0(y)}{\mu} + \frac{1}{\mu}\sum_{i=1}^{n} x_i f_i(y) + \sum_{i=1}^{n} \ln x_i + n(1 - \ln \mu).$$

We also need the following lemma in proving the theorem:

Lemma 3.2.1 *If $\bar{y} \in PF$ and $(x, y) \in DF$, then $\phi(\bar{y}, \mu) \ge \phi_d(x, y, \mu)$. Moreover, $\phi(y(\mu), \mu) = \phi_d(x(\mu), y(\mu), \mu)$.*

Proof: Since $-f_0(y)$ and $f_i(y)$, $i = 1, 2, \cdots, n$, are convex and $\bar{y} \in PF$ and $(x, y) \in DF$, we know that

$$\begin{aligned} -f_0(\bar{y}) + \sum_{i=1}^{n} x_i f_i(\bar{y}) &\ge -f_0(y) - \nabla f_0(y)^T(\bar{y} - y) \\ &\quad + \sum_{i=1}^{n} x_i f_i(y) + \sum_{i=1}^{n} x_i \nabla f_i(y)^T(\bar{y} - y) \\ &= -f_0(y) + \sum_{i=1}^{n} x_i f_i(y). \end{aligned}$$

Using this inequality, we get

$$\begin{aligned}\phi(\bar{y},\mu)-\phi_d(x,y,\mu) &= \left[-\frac{f_0(\bar{y})}{\mu}-\sum_{i=1}^{n}\ln(-f_i(\bar{y}))\right]+\left[\frac{f_0(y)}{\mu}\right. \\ &\quad \left.-\frac{1}{\mu}\sum_{i=1}^{n}x_i f_i(y)-\sum_{i=1}^{n}\ln x_i-n(1-\ln\mu)\right] \\ &\geq -\frac{1}{\mu}\sum_{i=1}^{n}x_i f_i(\bar{y})-\sum_{i=1}^{n}\ln\left(-x_i f_i(\bar{y})\right)-n(1-\ln\mu).\end{aligned}$$

Now considering the right hand side of this inequality, we observe that it is convex in x_i. We then take its derivatives with respect to x_i and get $x_i=-\dfrac{\mu}{f_i(\bar{y})}$. Therefore such x_i will minimize the right hand side of this inequality with optimal value 0. Thus the inequality becomes

$$\phi(\bar{y},\mu)-\phi_d(x,y,\mu)\geq 0.$$

The equality holds if and only if all the inequalities become equalities, or equivalently $\bar{y}=y=y(\mu)$ and $x=x(\mu)$. □

The monotonicity of the primal and dual objective along the central path is given by the next theorem.

Theorem 3.2.2 *The objective function $f_0(y(\mu))$ of the primal problem (CP) is monotonically increasing, and the objective function of the dual problem (CD), i.e.*

$$f_0(y(\mu))-\sum_{i=1}^{n}x_i f_i(y(\mu)),$$

is monotonically decreasing if μ decreases, where $x(\mu)$ and $y(\mu)$ are defined by the KKT conditions in the system of equations (3.2).

Proof: Let $\mu_1<\mu_2$. Since $y(\mu)$ minimizes $\phi(y,\mu)$, we have

$$\phi(y(\mu_1),\mu_1)\leq\phi(y(\mu_2),\mu_1)$$

and

$$\phi(y(\mu_2), \mu_2) \leq \phi(y(\mu_1), \mu_2).$$

By adding these two inequalities and using the definition of $\phi(y, \mu)$, we get

$$\phi(y(\mu_1), \mu_1) + \phi(y(\mu_2), \mu_2) \leq \phi(y(\mu_2), \mu_1) + \phi(y(\mu_1), \mu_2),$$

or

$$-\frac{f_0(y(\mu_1))}{\mu_1} - \frac{f_0(y(\mu_2))}{\mu_2} \leq -\frac{f_0(y(\mu_2))}{\mu_1} - \frac{f_0(y(\mu_1))}{\mu_2}.$$

It implies

$$0 < \left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right) [f_0(y(\mu_1)) - f_0(y(\mu_2))].$$

Since $\mu_1 < \mu_2$, the first part of the theorem follows.

We now prove the second part of the lemma using the similar argument. By lemma 3.2.1, we know that $(x(\mu), y(\mu))$ maximizes $\phi_d(x, y, \mu)$. Suppose $\mu_1 < \mu_2$, we have

$$\begin{aligned} \phi_d(x(\mu_1), y(\mu_1), \mu_1) &\geq \phi_d(x(\mu_2), y(\mu_2), \mu_1), \\ \phi_d(x(\mu_2), y(\mu_2), \mu_2) &\geq \phi_d(x(\mu_1), y(\mu_1), \mu_2). \end{aligned}$$

Adding the two inequalities gives

$$-\frac{G(\mu_1)}{\mu_1} - \frac{G(\mu_2)}{\mu_2} \geq -\frac{G(\mu_2)}{\mu_1} - \frac{G(\mu_1)}{\mu_2}$$

where $G(\mu) = \dfrac{f_0(y(\mu))}{\mu} - \dfrac{1}{\mu}\displaystyle\sum_{i=1}^{n} x_i(\mu) f_i(y(\mu))$. Rearranging the terms gives

$$(G(\mu_2) - G(\mu_1)) \left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right) \geq 0.$$

Since $\mu_1 < \mu_2$, it proves the second part of the theorem. □

Now we introduce a general *Logarithmic Barrier Algorithm* which gives an approximated solution y of (CP) such that $f_0(y^*) - f_0(y) \leq \varepsilon$.

Logarithmic Barrier Algorithm

Declear

$\varepsilon > 0$ is the accuracy parameter;

$\tau > 0$ is the proximity parameter;

$\theta \in (0,1)$ is the reduction parameter;

$\mu_0 > 0$ is the initial barrier value;

y_0 is a given interior feasible point such that

$\|p(y^0, \mu_0)\|_{H(y^0,\mu_0)} \leq \tau$.

$y := y_0,\ \mu := \mu_0$

Do while $\mu > \frac{\varepsilon}{4n}$

Reduce μ;

Do while $\|p\|_H \geq \tau$

Calculate the search direction p;

$\tilde{\alpha} := \arg\min_{\alpha>0}\{\phi(y + \alpha p, \mu) :\ y + \alpha p \in PF^o\}$;

$y := y + \tilde{\alpha} p$;

End

End

There are some variations of the above algorithm.

	Our choice
Minimizing method for $\phi(y, \mu)$	Newton direction $p = -H^{-1}g$
Stopping criterion	$\|p\|_H \leq \tau < 1$
Updating scheme of μ	μ is reduced by a constant

In our choice, p is associated with $H(y, \mu)$ and $g(y, \mu)$ and we will write p instead of $p(y, \mu)$ if there is no confusion. Clearly, $\|p\|_H = 0$ if and only if $y = y(\mu)$. If $\|p\|_H = 0$ then $0 = p = -H^{-1}g$. The definition of Newton direction p is well-defined because H is positive definite if $\phi(y, \mu)$ satisfies the self-concordance condition and so its inverse always exists. This gives $\nabla\phi(y, \mu) = g = -Hp = 0$.

Combining with the fact that $H = \nabla^2\phi(y,\mu)$ is positive definite, the sufficient conditions for an unconstrained minimum is satisfied, thus y is a minimizer of $\phi(y,\mu)$. However $\phi(y,\mu)$ has a unique minimizer $y(\mu)$, so we know $y = y(\mu)$. It is simpler for the converse part. If $y = y(\mu)$, then by the necessary condition for an unconstrained minimum, we have $g = \nabla\phi(y,\mu) = 0$. Therefore by the definition of p, $p = 0$ and hence $\|p\|_H = 0$.

For finding an initial interior feasible point, we need an initialization algorithm which will be presented later.

The performance of this algorithm depends on two factors, the performance of the Newton method and the reduction rate of μ. We need an extra assumption, which is suggested in next section, on $\phi(y,\mu)$ in our analysis.

3.2.2 κ-Self-Concordance Condition

To deal with the first problem, i.e. the performance of the Newton's method, we further assume an additional smoothness condition on $\phi(y,\mu)$, namely the *self-concordance condition*, which is defined as follows:

- A function $\varphi : PF^0 \to \mathbb{R}$ is called *κ-self-concordant* on PF^0, with $\kappa \geq 0$, if $\varphi \in C^3(PF^0)$ and for all $y \in PF^0$ and $h \in \mathbb{R}^n$ it satisfies the condition

$$|\nabla^3 \varphi(y)[h,h,h]| \leq 2\kappa(h^T \nabla^2 \varphi(y)h)^{\frac{3}{2}}, \tag{3.3}$$

  where $\nabla^3\varphi(y)[h,h,h]$ denotes the third differential of φ at y and h.

The exponent $\frac{3}{2}$ ensures the independence of the norm of h. Clearly, linear and convex quadratic functions are 0-self-concordant. This condition gives us some information on the shape of the function φ because the third differential of φ is small relative to the second differential of φ. We can imagine that large κ implies that the third differential of φ may be large. So φ cannot be well approximated by a quadratic function and the Newton's method will not perform very well.

It is clear that if φ is $\frac{1}{2}$-self-concordant, then 4φ is 1-self-concordant. In [9], this self-concordance condition is replaced by something like requiring the supremum

$$\sup_{y \in PF^0,\, h \in \mathbb{R}^n} \frac{|\nabla^3 \varphi(y)[h,h,h]|}{(h^T \nabla^2 \varphi(y) h)^{\frac{3}{2}}}$$

to be finite. We will not use this definition here but we will assume that $\kappa \geq 1$ in our analysis without loss of generality.

In general, it is difficult to check whether a logarithmic barrier function is self-concordant. But in [6], we can find many problems which can be reformulated such that the new logarithmic barrier function is self concordant, and in [5] the logarithmic barrier function for some classes of problems satisfy the self-concordance condition.

In our analysis, we will assume the logarithmic barrier function for (CP) to be κ-self-concordant. In [3], such function which is self-concordant and goes to infinity as y approaches the boundary of PF^0 is called *strong self-concordant.*

We show in the next lemma that the boundedness of PF and the self-concordance of $\phi(y,\mu)$ implies that $\phi(y,\mu)$ is strictly convex.

Lemma 3.2.2 *If PF^0 is bounded and $\phi(y,\mu)$ is self-concordance for each $\mu > 0$ then $\phi(y,\mu)$ is strictly convex.*

Proof: Recall that

$$\nabla^2 \phi(y,\mu) = -\frac{\nabla^2 f_0(y)}{\mu} + \sum_{i=1}^{n} \left[\frac{\nabla^2 f_i(y)}{-f_i(y)} + \frac{\nabla f_i(y) \nabla f_i(y)^T}{f_i(y)^2} \right] \qquad (3.4)$$

is positive semidefinite. Now suppose that $\nabla^2 \phi(y,\mu)$ is not positive definite. Then there exist $h \neq 0$ such that $h^T \nabla^2 \phi(y,\mu) h = 0$. From (3.4), we get

$$\begin{aligned} h^T \nabla^2 f_i(y) h &= 0, \\ \nabla f_i(y)^T h &= 0, \qquad \forall i = 1, \cdots, n. \end{aligned}$$

On the other hand, since $\phi(y, \mu)$ is self-concordance, it satisfies

$$| \bigtriangledown^3 \phi(y, \mu)[d, d, d]| \leq 2\kappa(d^T \bigtriangledown^2 \phi(y, \mu)d)^{\frac{3}{2}} = 0,$$

for all $d \in \mathbb{R}^n$. Then this implies $d^T \bigtriangledown^2 \phi(y + d)d = 0$. Therefore, for all $d \in \mathbb{R}^n$, we have

$$\begin{aligned} d^T \bigtriangledown^2 f_i(y + d)d &= 0, \\ \bigtriangledown f_i(y + d)^T d &= 0, \qquad \forall i = 1, \cdots, n. \end{aligned}$$

Considering the Taylor expansion of $f_i(y + th)$ for each $i = 1, \cdots, n$,

$$f_i(y + th) = f_i(y) + t \bigtriangledown f_i(y)^T h + \frac{t^2}{2} h^T \bigtriangledown^2 f_i(y + sh)h$$

where $s \in (0, t)$. This gives,

$$f_i(y + th) = f_i(y) + \frac{t^2}{2s^2}(sh)^T \bigtriangledown^2 f_i(y + sh)(sh) = f_i(y)$$

which contradicts to the boundedness of PF^0 by letting $t \to \infty$. This completes the proof by applying lemma 1.3.2 to the positive definite matrix $\bigtriangledown^2\phi(y, \mu)$. □

3.2.3 Short-step Logarithmic Barrier Algorithm

Before stating the *Short-step Logarithmic Barrier Algorithm*, we state some theoretical results about the *Logarithmic Barrier Algorithm* in [7]. The first result is that the difference of the barrier function value of the approximately centered iterate and the exact μ-center is bounded. The following lemma gives a convergent result of the Newton's method for the target $y(\mu)$.

Lemma 3.2.3 *If $y \in PF^0$ and $\|p\|_H < \frac{1}{\kappa}$ then $y^+ := y + p \in PF^0$ and*

$$\|p(y^+, \mu)\|_{H(y^+,\mu)} \leq \frac{\kappa}{(1 - \kappa\|p\|_H)^2}\|p\|_H^2.$$

Proof: One can show that the following inequalities hold for some $t \in [0,1]$ with arbitrary $v \in \mathbb{R}^m$ (cf. [7])

$$(1 - t\kappa\|d\|_H)\|v\|_H \leq \|v\|_{H(y+td)} \leq \frac{1}{1 - t\kappa\|d\|_H}\|v\|_H. \tag{3.5}$$

If $\|p\|_H < \frac{1}{\kappa}$, then $H(y+tp)$ is bounded for all $0 \leq t \leq 1$ and so is $\phi(y+tp, \mu)$. Thus $y + p \in PF^0$ as ϕ goes to infinity on the boundary of PF, see the details in [7]. □

In order to have the convergence of the Newton's method, the Newton direction has to satisfy $\|p\|_H < \frac{3-\sqrt{5}}{2\kappa}$. If $\|p\|_H \leq \frac{1}{3\kappa}$ then $\|p(y^+, \mu)\|_{H(y^+,\mu)} \leq \frac{9}{4}\kappa\|p\|_H^2$.

Lemma 3.2.4 *If* $\|p\|_H \leq \dfrac{1}{3\kappa}$, *then*

$$\phi(y, \mu) - \phi(y(\mu), \mu) \leq \frac{\|p\|_H^2}{1 - \left(\frac{9}{4}\kappa\|p\|_H\right)^2}.$$

Proof: See [7] □

The following lemma shows the difference in the objective value at y and $y(\mu)$ in (CP) is bounded (cf. [7]):

Lemma 3.2.5 *If* $\|p\|_H \leq \dfrac{1}{3\kappa}$, *then*

$$|f_0(y) - f_0(y(\mu))| \leq \frac{\|p\|_H}{1 - \frac{9}{4}\kappa\|p\|_H}\frac{1 + \kappa\|p\|_H^2}{1 - \kappa\|p\|_H}\mu\sqrt{n} \leq \left(1 + \frac{1}{9\kappa}\right)\frac{2\mu\sqrt{n}}{\kappa}.$$

Furthermore, one can also prove that the total number of Newton iterations in *Logarithmic Barrier Algorithm* is at most

$$\left[\frac{22}{(1-\theta)^2}\left(\theta\kappa^2 n + \frac{5}{2}\kappa\sqrt{n}\right) + \frac{22}{3\theta}\right]\ln\frac{4n\mu_0}{\varepsilon}$$

to give an ε-optimal solution ($f_0(y^*) - f_0(y) \leq \varepsilon$) (cf. [7]).

We next give some terminology for different step lengths with the order of Newton iterations:

- A step length is called

Term	Value of θ	# of Newton iterations
long-step	$0 < \theta < 1$	$O\left(\kappa^2 n \ln \frac{n\mu_0}{\epsilon}\right)$
medium-step	$\theta = \frac{\nu}{\sqrt{n}},\ \nu > 0$	$O\left(\kappa^2 \sqrt{n} \ln \frac{n\mu_0}{\epsilon}\right)$
short-step	$\theta = \frac{\nu}{\kappa\sqrt{n}},\quad \frac{\nu}{\kappa}$ is small	$O\left(\kappa \sqrt{n} \ln \frac{n\mu_0}{\varepsilon}\right)$

The term ' small $\frac{\nu}{\kappa}$ ' means that only one unit Newton step is required in inner loop such that the iterate is close to the new μ-center after the reduction of barrier value μ.

The following theorem provides a θ with this property (cf. [7]):

Theorem 3.2.3 *If* $\|p\|_H \leq \frac{1}{3\kappa}$, *then*

$$\|p(y^+, \mu^+)\|_{H(y^+,\mu^+)} \leq \frac{1}{3\kappa}$$

where $y^+ := y + p$ *and* $\mu^+ := (1-\theta)\mu$ *with* $\theta = \frac{1}{30\kappa\sqrt{n}}$.

From this theorem, we know that the new iterate y^+ is keeping close to the new μ^+-center $y(\mu^+)$ (in $\|\cdot\|_{H(y^+,\mu^+)}$) and then the Newton's method converges quadratically with constant at most $\frac{9}{4}\kappa$. The following algorithm gives an ε-optimal solution.

Short-step Logarithmic Barrier Algorithm

Declear

$\varepsilon > 0$ is the accuracy parameter;

$\tau = \frac{1}{3\kappa}$ is the proximity parameter;

$\theta = \frac{1}{30\kappa\sqrt{n}}$ is the reduction parameter;

$\mu_0 > 0$ is the initial barrier value;

y_0 is a given interior feasible point such that

$$\|p(y^0, \mu_0)\|_{H(y^0,\mu_0)} \leq \tau.$$

$y := y_0$, $\mu := \mu_0$

Do while $\mu > \frac{\varepsilon}{4n}$

$\mu = (1 - \theta)\mu$;

$p = -H^{-1}g$;

$y := y + p$;

End

In the next section, we present an initialization algorithm which can generate an interior feasible point y_0 that satisfies the input assumption of the Short-step Logarithmic Barrier Algorithm.

3.2.4 Initialization Algorithm

The initialization algorithm will generate an y_0, approximating

$$\bar{y} = \arg\max_{y \in PF} \prod_{i=1}^{n} (-f_i(y)).$$

With this y_0, we can choose $\mu_0 < \infty$ large enough so that y_0 is close to the μ_0-center,

$$y(\mu_0) = \arg\min \left[\frac{-f_0(y)}{\mu_0} - \sum_{i=1}^{n} \ln(-f_i(y)) \right].$$

To do so, we introduce some notation:

$$\begin{aligned} g_i(y,\eta) &= f_i(y) - \eta[f_i(0)+1], \quad 1 \le i \le n, \\ g_{n+1}(y,\nu) &= -\sum_{i=1}^{n} \nabla f_i(y)^T \Big|_{y=0} \, y - \nu, \end{aligned}$$

and

$$\begin{aligned} I(\eta,\nu) &= \{y \mid g_{n+1}(y,\nu) \le 0,\ g_i(y,\eta) \le 0,\ 1 \le i \le n\}, \\ \Psi(y,\eta,\nu) &= -\ln(-g_{n+1}(y,\nu)) - \sum_{i=1}^{n} \ln(-g_i(y,\eta)), \end{aligned}$$

where $\Psi(y,\eta,\nu)$ is defined on the interior of $I(\eta,\nu)$. The idea of this algorithm is to find a sequence of approximated minimizers of $\Psi(y,\eta,\nu)$. Since $I(\eta,\nu) \to I(0,\infty) = PF$ as $\eta \to 0$ and $\nu \to \infty$, we may find the approximation of $\bar{y}$.

We now show that $I(1,1)$ is nonempty, and 0 is the minimizer of $\Psi(y,1,1)$. If PF is bounded, so is $I(\eta,\nu)$, $\forall \eta \in \mathbb{R}$, $\nu \le \infty$. The fact that $0 \in I(1,1)$ follows from

$$\begin{aligned} g_i(0,1) &= -1, \qquad \forall\, 1 \le i \le n, \\ g_{n+1}(0,1) &= -1. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} \nabla_y \Psi(y,1,1)|_{y=0} &= -\frac{\nabla_y(-g_{n+1}(y,1))}{-g_{n+1}(y,1)}\Big|_{y=0} - \sum_{i=1}^{n} \frac{\nabla_y(-g_i(y,1))}{-g_i(y,1)}\Big|_{y=0} \\ &= -\sum_{i=1}^{n} \nabla f_i(y)\Big|_{y=0} - \sum_{i=1}^{n} \frac{-\nabla f_i(y)}{1}\Big|_{y=0} \\ &= 0. \end{aligned}$$

Since $\Psi(y,\eta,\nu)$ is a convex function and the corresponding feasible set is a convex set, the minimizer is unique. Therefore 0 is the minimizer of $\Psi(y,\eta_0,\nu_0) = \Psi(y,1,1)$. We use the minimizer 0 of $\Psi(y,1,1)$ as a starting point in the following initialization algorithm.

Initialization Algorithm

For (i) $\eta_j = 1, \cdots$ (decreasing to 0)

(ii) $\nu_k = 1, \cdots$ (increasing integer to ∞)

$p = -\hat{H}^{-1}\hat{g}$;

$y = y + p$;

If $\|p\|_{\hat{H}} < \tau$ then break

End

Choose large enough $\mu_0 < \infty$ such that $\|p(y, \mu_0)\|_{H(y,\mu_0)} < \tau$

For particular y, η_j, ν_k,

$$\begin{aligned}
\hat{g} = \nabla_y \Psi(y, \eta_j, \nu_k) &= -\frac{\nabla_y(-g_{n+1}(y, \nu_k))}{-g_{n+1}(y, \nu_k)} - \sum_{i=1}^{n} \frac{\nabla_y(-g_i(y, \eta_j))}{-g_i(y, \eta_j)} \\
&= -\frac{b}{b^T y + \nu_k} - \sum_{i=1}^{n} \frac{\nabla f_i(y)}{a_i(y, \eta_j)}, \\
\hat{H} = \nabla_y^2 \Psi(y, \eta_j, \nu_k) &= \frac{bb^T}{(b^T y + \nu_k)^2} - \sum_{i=1}^{n} \left[\frac{\nabla^2 f_i(y)}{a_i(y, \eta_j)} - \frac{\nabla f_i(y) \nabla f_i(y)^T}{a_i(y, \eta_j)^2} \right],
\end{aligned}$$

where $a_i(y, \eta_j) = f_i(y) - \eta_j[f_i(0) + 1]$ and $b = \sum_{i=1}^{n} \nabla f_i(y) \Big|_{y=0}$.

In the above algorithm, by a suitable choice of η_j and ν_k such that $\eta_j \to 0$ and $\nu_k \to \infty$ slow enough, we can find the approximated minimizer of $\Psi(y, \eta_j, \nu_k)$ by one Newton step in each iteration. The iteration is terminated if y is close enough to $\bar{y}$. Then we take $y_0 = y$ and a large enough number $\mu_0 < \infty$ is chosen such that y_0 is close enough to the μ_0-center. Using this algorithm, the initialization requirement in the Short-step Logarithmic Barrier Algorithm is satisfied.

3.3 Center Method

3.3.1 Basic Concepts and Properties

In this section, we will introduce another method for solving (CP) called the *center method* (see [3], [19], [21], [22]). We first define the *distance function* associated with (CP)

$$\psi(y,z) = -q\ln(f_0(y) - z) - \sum_{i=1}^{n} \ln(-f_i(y)), \tag{3.6}$$

where z is a lower bound for the optimal value z^* and q is a given positive integer. We assume that

- $-f_0(y)$ and $f_i(y)$, $1 \leq i \leq n$ are twice continuously differentiable.
- PF^0 is a nonempty bounded set.
- $\psi(y,z)$ is κ-self-concordant

The necessary and sufficient KKT conditions for $y(z)$ to be a minimizer of $\psi(y,z)$ is the same as (3.2) with $\mu = \dfrac{f_0(y) - z}{q}$, i.e.

$$\begin{aligned} f_i(y) \leq 0,\ x \geq 0, &\qquad 1 \leq i \leq n, \\ \sum_{i=1}^{n} x_i \bigtriangledown f_i(y) = \bigtriangledown f_0(y), & \\ -f_i(y)x_i = \frac{f_0(y) - z}{q}, &\qquad 1 \leq i \leq n. \end{aligned}$$

The proof is the same as the one of theorem 3.2.1 by letting $\mu = \dfrac{f_0(y) - z}{q}$.

By letting $-f_i(y) = f_0(y) - z$, $n+1 \leq i \leq n+q$, we can rewrite $\psi(y,z)$ as

$$\psi(y,z) = -\sum_{i=1}^{n+q} \ln(-f_i(y)).$$

We present some terminology before stating the algorithm.

- The bounded convex set F_z is defined by

$$F_z = \{y : \ f_i(y) \leq 0, \quad 1 \leq i \leq n+q\}.$$

This set is bounded because $F_z \subseteq PF$ and PF^0 is assumed to be bounded. The interior of F_z is denoted by F_z^0.

Since $\psi(y,z)$ is κ-self-concordant, $\psi(y,z)$ is strictly convex in F_z^0. Therefore $y(z)$ is the unique minimizer of $\psi(y,z)$.

- The *analytic center* of F_z is the point which maximizes

$$\prod_{i=1}^{n+q}(-f_i(y)). \tag{3.7}$$

It is clear that $y(z)$ is the analytic center of F_z because $y(z)$ minimizes $\psi(y,z)$ and so it maximizes

$$e^{-\psi(y,z)} = \prod_{i=1}^{n+q}(-f_i(y)).$$

Since F_z is a bounded convex set, the analytic center of F_z is defined uniquely by $y(z)$.

Thus the necessary and sufficient KKT conditions for (3.7) is

$$\begin{aligned} f_i(y) \leq 0, \ \tilde{x}_i \geq 0, \quad & 1 \leq i \leq n+q \\ \sum_{i=1}^{n+q} \tilde{x}_i \bigtriangledown f_i(y) = 0, \quad & \\ -f_i(y)\tilde{x}_i = 1, \quad & 1 \leq i \leq n+q, \end{aligned} \tag{3.8}$$

which is equivalent to the necessary and sufficient KKT conditions for (3.6).

In the *center algorithm*, we need the first and second order derivatives of $\psi(y,z)$.

- $g(y,z) := \bigtriangledown\psi(y,z) = \sum_{i=1}^{n+q} \dfrac{\bigtriangledown f_i(y)}{-f_i(y)}$.

- $H(y,z) := \nabla^2\psi(y,z) = \sum_{i=1}^{n+q}\left[\frac{\nabla^2 f_i(y)}{-f_i(y)} + \frac{\nabla f_i(y)\,\nabla\, f_i(y)^T}{f_i(y)^2}\right]$.

By the self-concordant property of $\psi(y,z)$, $\psi(y,z)$ is strictly convex in F_z^0, and so we use the following norm again to measure the distance between two points:

$$\|v\|_H = \sqrt{v^T H v}.$$

We observe that $\psi(y,z)$ is in fact the logarithmic barrier function for the problem

$$\max\{0 :\ y \in F_z\}. \tag{3.9}$$

In order to analyze the center algorithm, we introduce the dual problem of (3.9).

$$\min\left\{-\sum_{i=1}^{n+q}\tilde{x}_i f_i(y) :\ \sum_{i=1}^{n+q}\tilde{x}_i\,\nabla\, f_i(y) = 0,\ x \geq 0\right\}.$$

The logarithmic barrier function associated with this dual problem is given by

$$\psi_d(\tilde{x},y,z) = \sum_{i=1}^{n+q}\tilde{x}_i f_i(y) + \sum_{i=1}^{n+q}\ln\tilde{x}_i + (n+q)$$

up to a constant. By a similar argument as used in the logarithmic barrier algorithm, the minimizer of $\psi_d(x,y,z)$ must satisfy the necessary and sufficient *KKT* conditions (3.8). Thus $(x(z),y(z))$ is the unique maximizer of $\psi_d(x,y,z)$.

The following theorem proves the monotonicity of the primal and dual objectives along the analytic center.

Theorem 3.3.1 *The primal objective $f_0(y(z))$ is monotonically increasing, the dual objective $f_0(y(z)) - \sum_{i=1}^{n} x_i(z)f_i(y(z))$ and $f_0(y(z)) - z$ are monotonically decreasing when z increases.*

Proof: For $\bar{z} > z$, we have

$$\psi_d(\tilde{x}(z),y(z),z) \geq \psi_d(\tilde{x}(\bar{z}),y(\bar{z}),z)$$

and

$$\psi_d(\tilde{x}(\bar{z}),y(\bar{z}),\bar{z}) \geq \psi_d(\tilde{x}(z),y(z),\bar{z})$$

since $(\tilde{x}(z), y(z))$ maximizes $\psi_d(\tilde{x}, y, z)$ and $(\tilde{x}(\bar{z}), y(\bar{z}))$ maximizes $\psi_d(\tilde{x}, y, \bar{z})$. By adding these two inequalities, we get

$$q[z\tilde{x}_{n+1}(z) + \bar{z}\tilde{x}_{n+1}(\bar{z})] \geq q[z\tilde{x}_{n+1}(\bar{z}) + \bar{z}\tilde{x}_{n+1}(z)],$$

which is equivalent to

$$(z - \bar{z})(\tilde{x}_{n+1}(z) - \tilde{x}_{n+1}(\bar{z})) \geq 0.$$

This implies

$$\tilde{x}_{n+1}(z) \leq \tilde{x}_{n+1}(\bar{z}).$$

Meanwhile, from the *KKT* conditions, we have $\tilde{x}_{n+1}(z) = \dfrac{1}{f_0(y(z)) - z}$ and a similar expression for $\tilde{x}_{n+1}(\bar{z})$. Therefore, $f_0(y(z)) - z$ is monotonically decreasing if z increases. Finally, by letting $\mu = \dfrac{f_0(y(z)) - z}{q}$ and similarly for $\bar{\mu}$, we know $\mu \geq \bar{\mu}$. Thus from theorem 3.2.2 it follows that $f_0(y(z)) - \sum_{i=1}^{n} x_i(z) f_i(y(z))$ is monotonically decreasing and $f_0(y(z))$ is monotonically increasing if z increases. □

Now we introduce a general *Center Algorithm* which can give an ε-optimal solution.

Center Algorithm

Declear

$\varepsilon > 0$ is the accuracy parameter;

$\tau > 0$ is the proximity parameter;

$\theta \in (0,1)$ is the updating factor;

$z^0 < f_0(y^0)$ is a lower bound for the optimal value;

y_0 is a given interior feasible point such that

$$\|p(y^0, z_0)\|_{H(y^0,z_0)} \leq \tau.$$

$y := y_0,\ z := z_0,\ \Delta = 4\left(1 + \frac{n}{q}\right)$

Do while $f_0(y) - z > \frac{\varepsilon}{\Delta}$

Enlarge z

Do while $\|p\|_H > \tau$

Calculate p;

$\tilde{\alpha} := \arg\min_{\alpha>0}\{\psi(y + \alpha p, z) :\ y + \alpha p \in F_z\}$;

$y := y + \tilde{\alpha} p$;

End

End

Since $0 < \theta < 1$, we have $z < z + \theta(f_0(y) - z) < f_0(y) < z^*$. The aim of this algorithm is to find a sequence $z_k < z_{k+1} \uparrow z^*$ and approximation $y_k \in F_z$ for the analytic center $y(z_k)$ of F_z.

In [7], an upper bound for the gap $(z^* - z)$ was given, and also the difference $\psi(y,z) - \psi(y(z),z)$ and the difference $(f_0(y(z)) - z)$.

Lemma 3.3.1 *We have*

$$z^* - z \leq (1 + \frac{n}{q})(f_0(y(z)) - z).$$

Lemma 3.3.2 *If* $\|p\|_H \leq \frac{1}{3\kappa}$, *then*

$$\psi(y,z) - \psi(y(z),z) \leq \frac{\|p\|_H^2}{1 - \left(\frac{9}{4}\kappa\|p\|_H\right)^2}.$$

Lemma 3.3.3 *If* $\|p\|_H := \|p(y,z)\|_H \leq \dfrac{1}{3\kappa}$ *and* $q \geq \dfrac{\sqrt{n}}{\kappa}$, *then*

$$f_0(y(z)) - z \leq \left(1 + \frac{2\sqrt{n}}{q}\frac{\|p\|_H}{1 - \frac{9}{4}\kappa\|p\|_H}\right)(f_0(y) - z).$$

The upper bound for the total number of the Newton iterations in order to give an ε-optimal solution is given by

$$88(1 + \frac{n}{q})\left(\frac{1}{3\theta} + q\kappa^2\left(\frac{\theta}{1-\theta} + \frac{3\sqrt{n}}{q\kappa + 3\sqrt{n}}\right)\right)\ln\frac{4(1+\frac{n}{q})(z^* - z^0)}{\varepsilon}.$$

The following table concludes the order of Newton iterations for providing an ε-optimal solution with different step sizes.

Term	Value of θ	# of Newton iterations
long-step	$0 < \theta < 1$	$O\left(\kappa^2 n \ln \frac{z^* - z^0}{\epsilon}\right)$
medium-step	$\theta = \dfrac{\nu}{\sqrt{n}}$, $\nu > 0$	$O\left(\kappa^2\sqrt{n} \ln \frac{z^* - z^0}{\epsilon}\right)$
short-step	$\theta = \dfrac{\nu}{\kappa\sqrt{n}}$, $\dfrac{\nu}{\kappa}$ is small	$O\left(\kappa\sqrt{n} \ln \frac{z^* - z^0}{\varepsilon}\right)$

We have the following theorem for short-step algorithm.

Theorem 3.3.2 *If* $\|p\|_H \leq \dfrac{1}{3\kappa}$, *then*

$$\|p(y^+, \mu^+)\|_{H(y^+,\mu^+)} \leq \frac{1}{3\kappa}.$$

where $y^+ := y + p$ *and* $z^+ := z + \theta(f_0(y) - z)$ *with* $\theta = \dfrac{1}{22\kappa\sqrt{q}}$.

From this theorem for short-step, we know that the iterate y^+ is keeping close to the new analytic-center $y(z^+)$ (in $\|\cdot\|_{H(y^+,z^+)}$) after one Newton iteration. We then give a Short-step Center Algorithm in the next section.

3.3.2 Short-step Center Algorithm

The following is a short-step center algorithm.

Short-step Center Algorithm

Declear

$\varepsilon > 0$ is the accuracy parameter;

$\tau = \frac{1}{3\kappa}$ is the proximity parameter;

$\theta = \frac{1}{22\kappa\sqrt{q}}$ is the updating factor;

$z^0 < f_0(y^0)$ is a lower bound for the optimal value;

y_0 is a given interior feasible point such that

$$\|p(y^0, z_0)\|_{H(y^0,z_0)} \leq \tau.$$

$y := y_0,\ z := z_0,$

Do while $f_0(y) - z > \frac{\varepsilon}{4\left(1+\frac{n}{q}\right)}$

$z := z + \theta(f_0(y) - z)$

$p = -H^{-1}g;$

$y := y + p$

End

In the next section, we present an initialization algorithm which can generate the interior feasible point y_0 so that we can apply the Short-step Center Algorithm.

3.3.3 Initialization Algorithm

The following algorithm can generate an initial guess y_0, an approximated analytic center $\bar{y}$ of PF, that is y_0 is an approximation of

$$\bar{y} = \arg\max \prod_{i=1}^{n} (-f_i(y)).$$

Then we can choose $z_0 > -\infty$ small enough such that y_0 is close to the analytic center $y(z_0)$ of F_{z_0}. For this purpose, we introduce

$$\begin{aligned} g_i(y,\eta) &= f_i(y) - \eta[f_i(0) + 1], \quad 1 \leq i \leq n, \\ g_{n+1}(y,\nu) &= -\sum_{i=1}^{n} \nabla f_i(y)^T \Big|_{y=0} \; y - \nu. \end{aligned}$$

Also let

$$\begin{aligned} I(\eta,\nu) &= \{y \mid g_{n+1}(y,\nu) \le 0,\ g_i(y,\eta) \le 0,\ 1 \le i \le n\}, \\ \Psi(y,\eta,\nu) &= -q\ln(-g_{n+1}(y,\nu)) - \sum_{i=1}^{n} \ln(-g_i(y,\eta)). \end{aligned}$$

The idea of this algorithm is to find the approximated analytic center of a sequence of set $I(\eta,\nu)$. Since $I(\eta,\nu) \to I(0,\infty) = PF$ as $\eta \to 0$ and $\nu \to \infty$, we may find the analytic center of $PF = I(0,\infty)$.

We now show that $I(1,q)$ is nonempty, and 0 is its analytic center. If PF is bounded, so is $I(\eta,\nu)$, $\forall \eta \in \mathbb{R},\ \nu \le \infty$. $I(1,q)$ is nonempty, since

$$\begin{aligned} g_i(0,1) &= -1, \qquad \forall 1 \le i \le n, \\ g_{n+1}(0,q) &= -q. \end{aligned}$$

That is to say $0 \in I(1,q)$. On the other hand, we have

$$\begin{aligned} \nabla_y \Psi(y,1,q)|_{y=0} &= -q\left.\frac{\nabla_y(-g_{n+1}(y,q))}{-g_{n+1}(y,q)}\right|_{y=0} - \left.\sum_{i=1}^{n}\frac{\nabla_y(-g_i(y,1))}{-g_i(y,1)}\right|_{y=0} \\ &= -q\frac{\sum_{i=1}^{n}\nabla f_i(y)|_{y=0}}{q} - \left.\sum_{i=1}^{n}\frac{-\nabla f_i(y)}{1}\right|_{y=0} \\ &= 0. \end{aligned}$$

Therefore the following initialization algorithm can be started at the analytic center 0 of $I(\eta_0,\nu_0)$ with $\eta_0 = 1, \nu_0 = q$.

Initialization Algorithm

For (i) $\eta_i = 1, \cdots$ (decreasing to 0)

(ii) $\nu_j = q, \cdots$ (increasing integer to ∞)

$p = -\hat{H}^{-1}\hat{g}$;

$y = y + p$;

If $\|p\|_{\hat{H}} < \tau$ then break

End

Choose small enough $z_0 > -\infty$ such that $\|p(y,z_0)\|_{H(y,z_0)} < \tau$

For particularly y, η_j, ν_k,

$$\begin{aligned}
\hat{g} = \nabla_y \Psi(y, \eta_j, \nu_k) &= -q\frac{\nabla_y(-g_{n+1}(y,\nu_k))}{-g_{n+1}(y,\nu_k)} - \sum_{i=1}^{n}\frac{\nabla_y(-g_i(y,\eta_j))}{-g_i(y,\eta_j)} \\
&= -q\frac{b}{b^T y + \nu_k} - \sum_{i=1}^{n}\frac{\nabla f_i(y)}{a_i(y,\eta_j)}, \\
\hat{H}^{-1} = \nabla_y^2 \Psi(y, \eta_j, \nu_k) &= q\frac{bb^T}{(b^T y + \nu_k)^2} - \sum_{i=1}^{n}\left[\frac{\nabla^2 f_i(y)}{a_i(y,\eta_j)} - \frac{\nabla f_i(y) \nabla f_i(y)^T}{a_i(y,\eta_j)^2}\right]
\end{aligned}$$

with $a_i(y, \eta_j) = f_i(y) + \eta_j[f_i(0) + 1]$ and $b = \sum_{i=1}^{n} \nabla f_i(y)\Big|_{y=0}$.

In the above algorithm, by a suitable choice of η_j and ν_k such that $\eta_j \to 0$ and $\nu_k \to \infty$ slow enough, we can find the approximated minimizer of $\Psi(y, \eta_j, \nu_k)$ by one Newton step in each iteration. The iteration is terminated if y is close enough to $\bar{y}$. Then we take $y_0 = y$ and a small enough number $z_0 > -\infty$ is chosen such that y_0 is still close to the analytic center $y(z_0)$ of F_{z_0}. Using this algorithm, the initialization requirement in the Short-step Center Algorithm is satisfied.

3.4 Properties and Examples on Self-Concordance

First let's recall the self-concordant condition:

- A functions $\varphi : PF^0 \to \mathbb{R}$ is called *κ-self-concordant* on PF^0, with $\kappa \geq 0$, if φ is three times continuously differentiable in PF^0 and for all $y \in PF^0$ and $h \in \mathbb{R}^n$ it satisfies the condition

$$|\nabla^3 \varphi(y)[h,h,h]| \leq 2\kappa(h^T \nabla^2 \varphi(y) h)^{\frac{3}{2}}, \tag{3.10}$$

where $\nabla^3\varphi(y)[h,h,h]$ denotes the third differential of φ at y and h, i.e.

$$\nabla^3\varphi(y)[h,h,h] = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\frac{\partial^3\varphi(y)}{\partial y_i \partial y_j \partial y_k} h_i h_j h_k.$$

It is clear that a linear function $l(y)$ and a convex quadratic function $q(y)$ is self-concordant. Since

$$\nabla^3 l(y)[h,h,h] = \nabla^2 l(y)[h,h] = 0,$$

and

$$\nabla^3 q(y)[h,h,h] = 0, \qquad \nabla^2 q(y)[h,h] \geq 0,$$

$l(y), q(y)$ satisfy the self-concordant condition if we take $\kappa = 1$ in both cases.

We now give some examples and properties of the self-concordant functions in the following:

- **The positive real axis.** The function $\varphi(y) = -\ln y$ is 1-self-concordant on the set $\{y \in \mathbb{R} : y > 0\}$.
  Proof: We have $\varphi'(y) = -\frac{1}{y}$, $\varphi''(y) = \frac{1}{y^2}$, $\varphi'''(y) = -\frac{2}{y^3}$.
  For all $y \in \{y : y > 0\}$, we get

$$|\varphi'''(y)| = \frac{2}{y^3} = 2\left(\frac{1}{y^2}\right)^{\frac{3}{2}} = 2\varphi''(y)^{\frac{3}{2}}.$$

  This gives the result. □

- **Affine transformations.** The self-concordant condition (3.3) is invariant under affine transformations. Let $\varphi(y)$ be κ-self-concordant on PF^0 and let $\mathcal{A}(x) := Ax + b$ be an affine mapping with some matrix $A \in \mathbb{R}^{m \times q}$ and some vector $b \in \mathbb{R}^m$ such that $\mathcal{A}(\mathbb{R}^q) \cap PF^0 \neq \emptyset$. Then $\varphi(\mathcal{A}(x))$ is κ-self-concordant on $\{x : \mathcal{A}(x) \in PF^0\}$.
  Proof: See [6, 9]. □

- **Linear functions.** The function

$$\varphi(y) = -\ln(a^T y + \beta)$$

  is 1-self-concordant on $\{y \in \mathbb{R}^m : a^T y + \beta > 0\}$.
  Proof: By the property of invariant under affine transformation and $-\ln y$

is 1-self-concordant on $\{y \in \mathbb{R} : \ y > 0\}$, the linear function is 1-self-concordant. □

- **Convex quadratic functions.** The function

$$\varphi(y) = -\ln(-q(y))$$

is 1-self-concordant on $\{y \in \mathbb{R}^m : \ q(y) < 0\}$ where $q(y)$ is convex quadratic.

Proof: $\forall y \in \{y \in \mathbb{R}^m : \ q(y) < 0\}$, let

$$d_1 = \frac{\nabla q(y)[h]}{-q(y)}, \qquad d_2 = \frac{\nabla^2 q(y)[h,h]}{-q(y)}.$$

Here $d_2 \geq 0$ since $q(y)$ is a convex function.

Then the derivatives of $\varphi(y)$ can be written in terms of d_1, d_2 and d_3 as follow:

$$\begin{aligned}
\nabla\varphi(y)[h] &= \frac{\nabla q(y)[h]}{-q(y)} := d_1, \\
\nabla^2\varphi(y)[h,h] &= \frac{\nabla^2 q(y)[h,h]}{-q(y)} + \frac{\nabla q(y)[h]^2}{q(y)^2} := d_1^2 + d_2, \\
\nabla^3\varphi(y)[h,h,h] &= \frac{\nabla^3 q(y)[h,h,h]}{-q(y)} + \frac{\nabla^2 q(y)[h,h] \nabla q(y)[h]}{q(y)^2} \\
&\quad + \frac{2 \nabla q(y)[h] \nabla^2 q(y)[h,h]}{q(y)^2} - \frac{2 \nabla q(y)[h]^2 \nabla q(y)[h]}{q(y)^3} \\
&= \frac{3 \nabla^2 q(y)[h,h] \nabla q(y)[h]}{q(y)^2} - \frac{2 \nabla q(y)[h]^3}{q(y)^3} \\
&:= 2d_1^3 + 3d_1 d_2,
\end{aligned}$$

since $q(y)$ is quadratic, $\nabla^3 q(y)[h,h,h] = 0$.

Now consider

$$\begin{aligned}
|\nabla^3 \varphi(y)[h,h,h]|^2 &= |2d_1^3 + 3d_1 d_2|^2 \\
&\leq 4d_1^6 + 12d_1^4 d_2 + 9d_1^2 d_2^2 \\
&\leq 4d_1^6 + 12d_1^4 d_2 + 12d_1^2 d_2^2 + 4d_2^3 \\
&= 4(d_1^2 + d_2)^3 \\
&= 4 \nabla^2 \varphi(y)[h,h]^3.
\end{aligned}$$

Taking square root on both sides, it shows $\varphi(y) = -\ln(-q(y))$ is 1-self-concordant. □

- **Summation.** Suppose that $\varphi_1(y)$ is κ_1-self-concordant on PF_1^0 and $\varphi_2(y)$ is κ_2-self-concordant on PF_2^0. Then $\varphi(y) := (m_1\varphi_1 + m_2\varphi_2)(y)$ is κ-self-concordant on $PF_1^0 \cap PF_2^0$ where $m_1, m_2 \in \mathbb{R}_+$ and $\kappa = \max\left\{\frac{\kappa_1}{\sqrt{m_1}}; \frac{\kappa_2}{\sqrt{m_2}}\right\}$.

  Proof: We have

$$|\nabla^3 \varphi_1(y)[h,h,h]| \leq 2\kappa_1 (h^T \nabla^2 \varphi_1(y) h)^{\frac{3}{2}}, \quad \forall y \in PF_1^0,$$

and

$$|\nabla^3 \varphi_2(y)[h,h,h]| \leq 2\kappa_2 (h^T \nabla^2 \varphi_2(y) h)^{\frac{3}{2}}, \quad \forall y \in PF_2^0.$$

Therefore $\forall y \in PF_1^0 \cap PF_2^0$, we have

$$\begin{aligned}
\left|\nabla^3\varphi(y)[h,h,h]\right| &= \left|\nabla^3(m_1\varphi_1 + m_2\varphi_2)(y)[h,h,h]\right| \\
&= \left|m_1 \nabla^3 \varphi_1(y)[h,h,h] + m_2 \nabla \varphi_2(y)[h,h,h]\right| \\
&\leq m_1 \left|\nabla^3\varphi_1(y)[h,h,h]\right| + m_2 \left|\nabla^3\varphi_2(y)[h,h,h]\right| \\
&\leq 2m_1\kappa_1 (h^T \nabla^2 \varphi_1(y) h)^{\frac{3}{2}} + 2m_2\kappa_2 (h^T \nabla^2 \varphi_2(y) h)^{\frac{3}{2}} \\
&= \frac{2\kappa_1}{\sqrt{m_1}} (h^T \nabla^2 m_1\varphi_1(y) h)^{\frac{3}{2}} + \frac{2\kappa_2}{\sqrt{m_2}} (h^T \nabla^2 m_2\varphi_2(y) h)^{\frac{3}{2}} \\
&\leq 2\kappa (h^T \nabla^2 (m_1\varphi_1 + m_2\varphi_2)(y) h)^{\frac{3}{2}}
\end{aligned}$$

where $\kappa = \max\{\frac{\kappa_1}{\sqrt{m_1}}; \frac{\kappa_2}{\sqrt{m_2}}\}$. In the last inequality, we have used the inequality that $a^{\frac{3}{2}} + b^{\frac{3}{2}} \leq (a+b)^{\frac{3}{2}}$, $\forall\ a, b \geq 0$. This completes the proof. □

In next section, we present some problems that the logarithmic barrier function is proved to satisfy the self-concordance condition in [5], [7].

3.5 Examples of Convex Optimization Problem

3.5.1 Self-concordant Logarithmic Barrier and Distance Function

Before stating the problems, we recall that a CO Problem (CP) is solvable by the Logarithmic Barrier Algorithm if it satisfies the following three conditions:

- $-f_0(y)$ and $f_i(y), 1 \leq i \leq n$ are twice continuously differentiable.
- PF^0 is a nonempty bounded set.
- The logarithmic barrier function associated with (CP) is κ-self-concordant.

The following are some problems with self-concordant logarithmic barrier function:

1. The linear programming (LP) problem:

$$(LP): \quad \begin{cases} \min & c^T y \\ s.t. & Ay \leq b, \\ & y \geq 0 \end{cases}$$

can be solved by the Logarithmic Barrier Algorithm and the Center Algorithm, if the feasible set PF is bounded and its interior is nonempty. The reason is that $\frac{c^T y}{\mu}$, $-\ln y_j$ and $-\ln(a_i y - b_i)$, where a_i is the i-th row of A, are 1-self-concordant and by the summation property the sum of these functions is 1-self-concordant in PF^0. Therefore the Logarithmic Barrier function

$$\phi(y, \mu) = \frac{c^T y}{\mu} - \sum_{i=1}^{n} \ln(a_i y - b_i) - \sum_{j=1}^{m} \ln y_j$$

associated with (LP) is 1-self-concordant. For $q \geq 1$, the distance function

$$\psi(y, \mu) = -q \ln(c^T y - z) - \sum_{i=1}^{n} \ln(a_i y - b_i) - \sum_{j=1}^{m} \ln y_j$$

associated with (LP) is $\max\left\{\frac{1}{\sqrt{q}}, 1\right\}$ =1-self-concordant.

If a (CP) consists of a system of equality constraints, the interior of the feasible set is then empty and so the initial assumption is not satisfied. To deal with this case, we can eliminate some of the variables by an affine transformation so that the equality constraints are removed. If the original logarithmic barrier function or distance function is self-concordant, the one after applying the affine transformation is also self-concordant. It is because the self-concordant property is invariant under affine transformations.

In [19], it gives some numerical experiments of Center Method for some (LP). In this thesis, We will give some numerical results for some CO problems later. We first apply the algorithms to (LP) which comes from [20]:

We Consider an economy involving the following activities: agriculture; transportation by truck; transportation by rail; manufacturing. The unit requirements for each activity are listed in the table below.

		Transportation		
	Agriculture	*by Truck*	*by Rail*	*Manufacturing*
Land	2	0	0	0
Energy resources	4.5	12	10	0.5
Capital goods	3	4	6	1.5
Transportation services	0	0	0	0.4
Final: Food	1	0	0	0
Manufactured Goods	0	0	0	1
Intermediate: Transportation	0	1	1	0
Activity levels	x_3	x_4	x_5	x_6

There are three primary goods used as inputs: land, energy resources and capital goods. The dimension of each input coefficient is a quantity per unit level of operation of the activity. There is one intermediate good: transportation services. These services are themselves an input used in the

manufacturing industry. There are two categories of final goods: food and manufactured goods.

Let the given endowments of primary goods be: land 100; energy 357.5; capital goods 227.5. The prices of the final goods are: food 36; manufactured goods 29.2. The final demand of food and manufactured goods are represented by x_1 and x_2 respectively.

Then, the optimal production is modeled by a (LP) with 6 variables and 6 constraints excluding those $y_i \geq 0$.

$$
\begin{array}{lrrrrrrl}
\max & 36y_1 + 29.2y_2 & & & & & & \\
& y_1 & -y_3 & & & & \leq & 0, \\
& & y_2 & & & -y_6 & \leq & 0, \\
& & & -y_4 & -y_5 & +0.4y_6 & = & 0, \\
s.t. & & 2y_3 & & & & \leq & 100, \\
& & 4.5y_3 & +12y_4 & +20y_5 & +0.5y_6 & \leq & 357.5, \\
& & 3y_3 & +4y_4 & +6y_5 & +1.5y_6 & \leq & 227.5, \\
& y_i \geq 0, & & i = 1, 2, \cdots, 6. & & & &
\end{array}
$$

The optimal solution $y^* = (50, 25, 50, 10, 0, 25)$, the optimal cost is $f_0(y^*) = 2530$. Here it contains an equality $-y_4 - y_5 + 0.4y_6 = 0$ and so the interior is empty. We need to remove the equality by eliminating one of the variable existing in $-y_4 - y_5 + 0.4y_6 = 0$, say $y_6 = 2.5y_4 + 2.5y_5$. We thus get the

new (LP) with 5 variables and 6 constraints.

$$
\begin{array}{lrrrrcl}
\min & -36y_1 - 29.2y_2 & & & & & \\
 & y_1 \quad -y_3 & & & & \leq & 0, \\
 & \quad y_2 & -2.5y_4 & -2.5y_5 & & \leq & 0, \\
 & & 4.5y_3 & +13.25y_4 & +21.25y_5 & \leq & 357.5, \\
s.t. & & 2y_3 & & & \leq & 100, \\
 & & 3y_3 & +7.75y_4 & +9.75y_5 & \leq & 227.5, \\
 & & & -2.5y_4 & -2.5y_5 & \leq & 0, \\
 & y_i \geq 0, \quad i = 1, 2, \cdots, 5. & & & & &
\end{array}
$$

We observe that the interior of the feasible set is now nonempty. For example, $(1, 1, 2, 2, 2)$ is an interior point. On the other hand, from the third inequality of the new (LP), variables y_3, y_4, y_5 are bounded and then from the first and second inequalities, variables y_1, y_2 are also bounded. Therefore the feasible set is bounded. Now, we can apply the short-step logarithmic barrier and center algorithm. Also we try larger step size (a long-step) in the two algorithms with replacement of stopping criterion by $f_0(y) \geq f_0(y^{short-step\ solution})$. The corresponding parameters in the algorithms are shown in the following table.

	Logarithmic Barrier		Center	
	Short-step	Long-step	Short-step	Long-step
q	-	-	$\sqrt{n}$	n
θ	$\frac{1}{30\sqrt{10}}$	$\frac{1}{\sqrt{10}}$	$\frac{1}{22\sqrt{q}}$	$\frac{5}{11}$
Initialization Iterations	24	24	24	24
μ_0	2048	2048	-	-
z_0	-	-	-1819	-8478
Algorithm Iterations	1720	48	1850	51

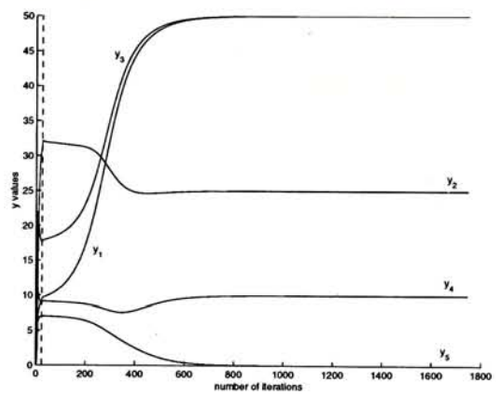
The numerically found optimal solutions by all these algorithms are $y =$ (50.0000, 25.0000, 50.0000, 10.0000, 0.0000, 25.0000), the optimal costs are

all $f_0(y) = 2529.9999$ which are correct to 4 decimal places.

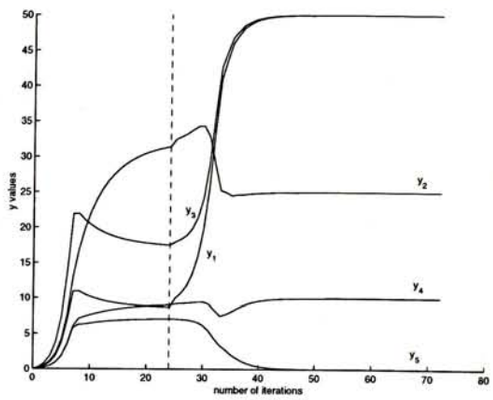
The following graphs show the convergence of the values y_i, $i = 1, 2, \cdots, 5$ in the two algorithm with the parameters given in the above table. The convergence of the initialization algorithms are shown in the left hand side of the vertical dashed line and the convergence of the main algorithms are shown in the right hand side. In order to show the clear convergence, the graphs are plotted by connecting the values of each y_i in each iteration .

Logarithmic Barrier Algorithm

Short-step

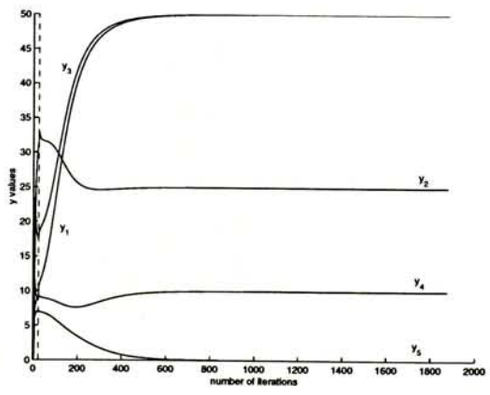


Long-step

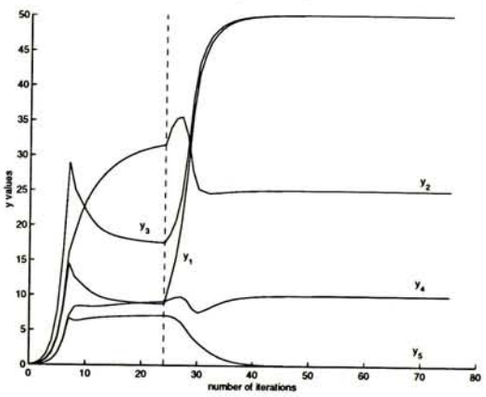


Center Algorithm

Short-step



Long-step



In the initialization algorithm of logarithmic barrier method, we reduce η by a factor of $\frac{1}{2}$ and enlarge ν by a factor of 2 simultaneously until $\nu \geq 10^7$ and $\|newton\ step\| < \frac{1}{3}$. Then we choose large enough μ_0 such

that $\|newton\ step\| < \frac{1}{3}$. In the initialization algorithm of center method, the stopping criteria are replaced by $\nu \geq \sqrt{n} \times 10^7$ and $\nu \geq n \times 10^7$ with $\|newton\ step\| < \frac{1}{3}$. Therefore we see that the number of initialization iterations are all 24. On the other hand, when comparing the two z_0 values of the center algorithm, the long-step one is more negative than the short-step one. It is because the weight of $-q\ln(c^T y - z)$ in the distance function for the long-step with $q = n$ is more important than for the short-step one with $q = \sqrt{n}$. Then we can start the two algorithms and get the result.

Although the curves are not very smooth at the beginning of the two main algorithms for long-step (i.e. they converge not very well initially), they converge in fact. □

Instead of (LP), the logarithmic barrier function and the center function of the convex quadratic programs with convex quadratic functions are also 1-self-concordant. The complexity analysis of this kind of problems are fully discussed (see [21, 22]).

2. We consider the convex quadratic programs with convex quadratic functions:

$$(CQP): \quad \begin{cases} \min & q_0(y) \\ s.t. & q_i(y) \leq 0, \qquad 1 \leq i \leq n. \end{cases}$$

where $q_i(y)$, $0 \leq i \leq n$ are convex quadratic functions

$$q_i(y) = \frac{1}{2} y^T Q_i y + b_i^T y + c_i$$

with $Q_i \in \mathbb{R}^{n \times n}$ positive semidefinite, $b_i \in \mathbb{R}^n$, $c_i \in \mathbb{R}$. Also we assume that the feasible set PF is bounded and its interior is nonempty. We can prove that the logarithmic barrier function

$$\phi(y, \mu) = \frac{q_0(y)}{\mu} - \sum_{i=1}^{n} \ln(-q_i(y))$$

associated with (CQP) is 1-self-concordant, and the distance function

$$\psi(y,z) = -\ln(-q_0(y) - z) - \sum_{i=1}^{n} \ln(-q_i(y))$$

associated with (CQP) is 1-self-concordant since $\frac{q_0(y)}{\mu}$, $-\ln(-q_0(y) - z)$ and $-\ln(-q_i(y))$ is 1-self-concordant in PF^0 and thus the sum of these functions is also 1-self-concordant in PF^0 by the summation property.

The following is a (CQP) problem given in [23]:

$$\begin{array}{ll} \min & y^T A y - 2y_1 \\ s.t. & y_i \geq 0, \qquad i = 1, 2, \cdots, 50, \end{array}$$

where A is a 50×50 symmetric positive definite matrix given by

$$A = \begin{bmatrix} 1 & -1 & & & & 0 \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ 0 & & & & -1 & 2 \end{bmatrix}.$$

Since the interior of the feasible set is unbounded, we add one more constraint

$$\sum_{i=1}^{50} y_i \leq 2500,$$

to the original problem. We can do so because we know the optimal solution of this test problem. The problem now becomes

$$\begin{array}{ll} \min & y^T A y - 2y_1 \\ s.t. & \sum_{i=1}^{50} y_i \leq 2500 \\ & y_i \geq 0, \qquad i = 1, 2, \cdots, 50. \end{array}$$

Now the feasible set becomes bounded and nonempty and we apply the Logarithmic Barrier and Center Method to this problem for some particular parameters given in the following table.

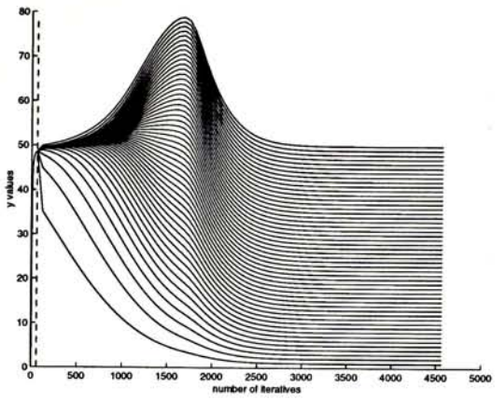
	Logarithmic Barrier		Center	
	Short-step	Long-step	Short-step	Long-step
q	-	-	$n = 51$	$n = 51$
θ	$\frac{1}{30\sqrt{51}}$	$\frac{1}{\sqrt{51}}$	$\frac{1}{22\sqrt{q}}$	$\frac{5}{11}$
Initialization Iterations	62	62	62	62
μ_0	8192	8192	-	-
z_0	-	-	-141264	-141264
Algorithm Iterations	4540	141	3396	37

In order to compare the difference between short-step and long-step algorithm, the stopping criterion of the long-step algorithms are replaced by $f_0(y) \leq f_0(y^{short-step\ solution})$.

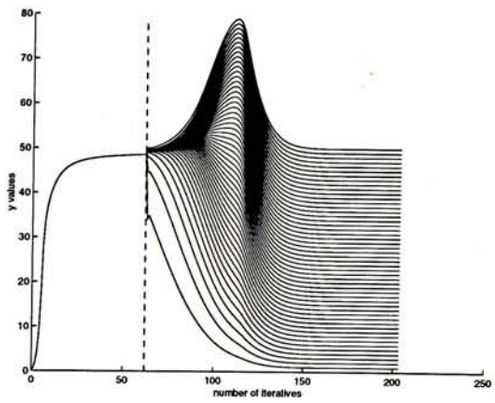
The following figures show the convergence of each values y_i in the iterated solution with different step sizes. In order to show clear the convergence, we connect the new value and the previous values for each y_i. The convergence of the values y_i in the initialization algorithm are shown in the left hand side of the vertical dashed line and the main algorithms in the right hand side of the vertical dashed line. The iterated solution $y_1, y_2, \cdots, y_{50}$ are shown by the tails from the top one to the bottom one respectively in each figure. Thus, $y_1 \approx 50, y_2 \approx 49, \cdots, y_{50} \approx 1$ and the optimal solution for this problem is $y^* = (50, 49, 48, 47, \cdots, 2, 1)$ and the optimal cost value is $f_0(y^*) = -50$.

Logarithmic Barrier Algorithm

Short-step

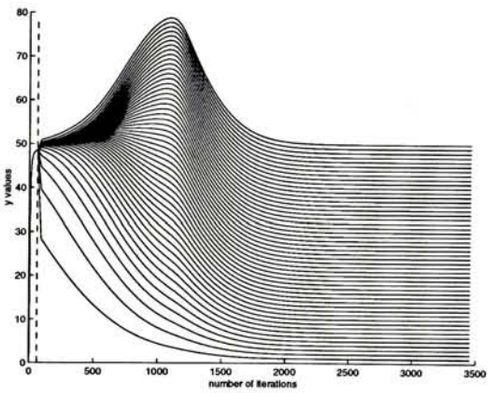


Long-step

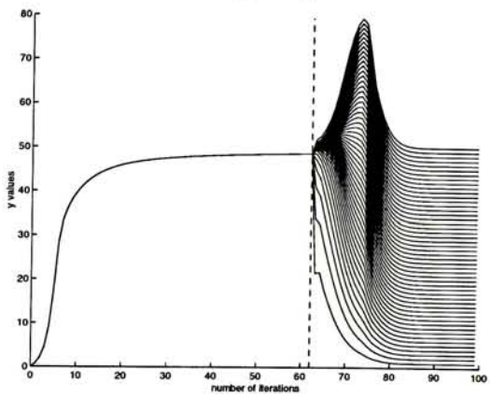


Center Algorithm

Short-step



Long-step



In the initialization algorithm, we reduce η by a factor of $\frac{1}{1.3}$ and enlarge ν by a factor of 1.3 simultaneously until $\nu \geq 10^7$ with $\|newton\ step\| < 1$ for the logarithmic barrier algorithm and $\nu \geq 5.1 \times 10^8$ with $\|newton\ step\| < 1$ for the center algorithm. Then we choose large enough μ_0 such that $\|newton\ step\| < 1$ for the logarithmic barrier algorithm and small enough z_0 such that $\|newton\ step\| < 1$ for the center algorithms. The reason we change $\|newton\ step\| < \frac{1}{3}$ by $\|newton\ step\| < 1$ is that the amplitude of z_0 and μ_0 to ensure $\|newton\ step\| < \frac{1}{3}$ is too large and $\|newton\ step\| < 1$ is good enough to apply the algorithms.

The numerical optimal cost values found are $f_0(y) = -50.000000000$ to 9 decimal places. □

3. Let $\{I_k\}_{k=1,\cdots,r}$ be a partition of $\{1,\cdots,n\}$ (i.e. $\cup_{k=1}^{r} I_k = \{1,\cdots,n\}$ and $I_k \cap I_l = \emptyset$ for $k \neq l$). The dual geometric programming problem is then given by (see [13])

$$(DGP): \begin{cases} \min & c^T y + \sum_{k=1}^{r} \left[\sum_{i\in I_k} y_i \ln y_i - \left(\sum_{i\in I_k} y_i\right) \ln \left(\sum_{i\in I_k} y_i\right)\right] \\ s.t. & Ay = b, \\ & y \geq 0. \end{cases}$$

The logarithmic barrier function of (DGP) is 2-self-concordant(see proof in [7]). □

4. The extended entropy programming problem is defined as (see [11, 12])

$$(EEP): \begin{cases} \min & c^T y + \sum_{i=1}^{n} f_i(y_i) \\ s.t. & Ay = b, \\ & y \geq 0. \end{cases}$$

where $|f_i'''(y_i)| \leq \kappa_i \dfrac{f_i''(y_i)}{y_i}$. The logarithmic barrier function of (EEP) is $(1 + \dfrac{1}{3}\max_i\{\kappa_i\})$-self-concordant(see proof in [7]). □

We know that the Logarithmic Barrier Algorithm and the Short-step Center Algorithm can solve CO problem which satisfies some conditions including the κ-self-concordance however it is difficult to check whether a general CO problem is κ-self-concordant or not. We will apply the Logarithmic Barrier and Center Algorithm to some CO problems without proving the self-concordance property. It always works because the barrier function $\phi(y,\mu)$ and the distance function $\psi(y,z)$ are convex in the feasible sets and so the minimum always exist. The inverse of $H(y,\mu)$ and $H(y,z)$ always exist and so the Newton steps are well-defined.

3.5.2 General Convex Optimization Problems

The following are some CO problems:

1. The following is a convex optimization problem about the maximum likelihood for interval censored data which comes from [15, 17, 18].

$$\begin{aligned} \max_{p\in\mathbb{R}^{14}} \quad & \sum_{i=1}^{46} \log\left(\sum_{j=1}^{14} \alpha_{ij} p_j\right) \\ s.t. \quad & p_1 + p_2 + \cdots + p_{14} = 1 \\ & p_i \geq 0, \qquad i = 1, 2, \cdots, 14 \end{aligned}$$

where $A = (\alpha_{ij})$ is a 46×14 matrix defined in next page.

Since the interior of the feasible set is empty, we need to transform the 14 variables problem into the following 13 variables problem.

$$\begin{aligned} \max_{p\in\mathbb{R}^{13}} \quad & \sum_{i=1}^{46} \log\left(\alpha_{i,14} + \sum_{j=1}^{13} (\alpha_{ij} - \alpha_{i,14}) p_j\right) \\ s.t. \quad & p_1 + p_2 + \cdots + p_{13} \leq 1 \\ & p_i \geq 0, \qquad i = 1, 2, \cdots, 13 \end{aligned}$$

Now the interior of the feasible becomes nonempty and bounded (e.g. $p = (0.05, 0.05, \cdots, 0.05)$ is feasible). We now apply the Logarithmic Barrier and Center Algorithm with different step sizes. Here the number of inequalities $n = 14$ and number of variables $m = 13$.

The following table shows the result from [15, 17] and our results. We apply the two algorithms by replacing the stopping criterion by $f_0(p^*) - f_0(p^{iterate}) \leq 4.6 \times 10^{-5}$. In the short-step algorithm, we assume $\kappa = 1$, $q = \sqrt{n}$ and in the long-step algorithm, we arbitrary choose θ and q.

$$A = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}$$

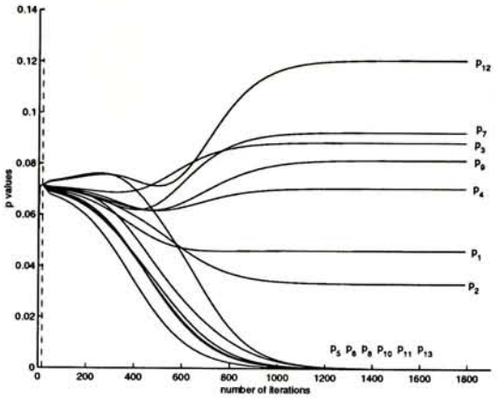
		Logarithmic Barrier		*Center*	
		Short-step	*Long-step*	*Short-step*	*Long-step*
q	-	-	-	$\sqrt{n}$	n
θ	-	$\frac{1}{30\sqrt{14}}$	$\frac{1}{\sqrt{14}}$	$\frac{1}{22\sqrt{q}}$	$\frac{5}{11}$
k	p_k^*	p_k			
1	0.0463	0.0463	0.0463	0.0463	0.0463
2	0.0334	0.0334	0.0334	0.0334	0.0334
3	0.0887	0.0887	0.0887	0.0887	0.0887
4	0.0708	0.0708	0.0708	0.0708	0.0708
5	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0926	0.0926	0.0926	0.0926	0.0926
8	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0818	0.0818	0.0818	0.0818	0.0818
10	0.0000	0.0000	0.0000	0.0000	0.0000
11	0.0000	0.0000	0.0000	0.0000	0.0000
12	0.1209	0.1209	0.1209	0.1209	0.1209
13	0.0000	0.0000	0.0000	0.0000	0.0000
14	0.4656	0.4656	0.4656	0.4656	0.4656
Initialization Iterations	-	17	17	17	17
μ_0	-	64	64	-	-
z_0	-	-	-	-199	-479
Algorithm Iterations	-	1786	52	1386	36
objective value	-58.06002	-58.06007	-58.06006	-58.06007	-58.06005

In [15], it states that using a EM algorithm to solve this problem which is easy to implement but converges slowly. It needs a couple of hundred steps to reach the bound. There are some alternative methods which can also solve this problem. The Logarithmic Barrier and Center Algorithm with long-step are easy to implement and needs less than hundred steps to reach the bound.

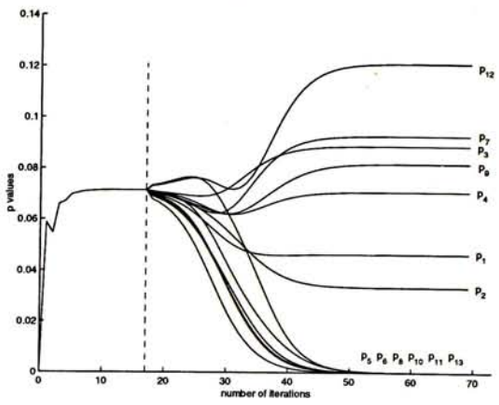
The following figures show the convergence of each values p_k in the iterated solution with different step sizes. The convergence of the p_k values in the initialization algorithm are shown in the left hand side of the vertical dashed line and the main algorithm in the right hand side of the vertical dashed line.

Logarithmic Barrier Algorithm

Short-step

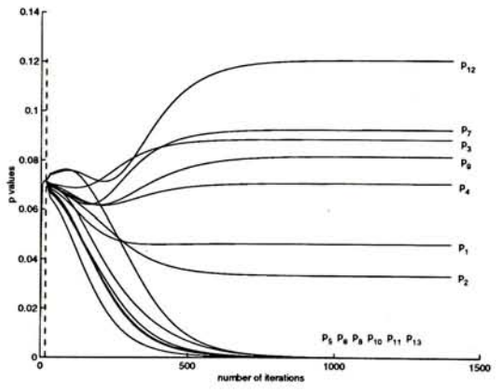


Long-step

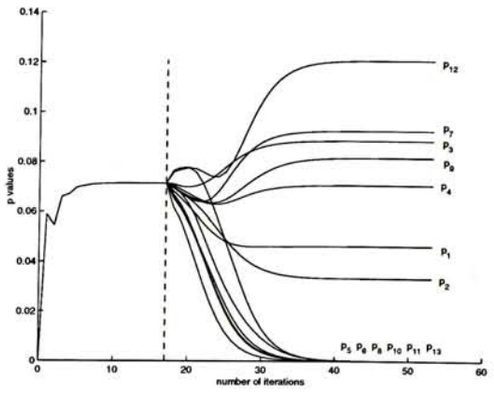


Center Algorithm

Short-step



Long-step



In the initialization algorithm of logarithmic barrier method, we reduce η by a factor of $\frac{1}{2}$ and enlarge ν by a factor of 2 simultaneously until $\nu \geq 10^5$ and $\|newton\ step\| < \frac{1}{3}$. Then we choose large enough μ_0 such that $\|newton\ step\| < \frac{1}{3}$, we get $\mu_0 = 64$. For the center initialization algorithm, $\nu \geq 10^5$ is replaced by $\nu \geq q \times 10^5$, the number of iterations are the same with the logarithmic barrier one. Thus, $\|newton\ step\| < \frac{1}{3}$ occur in those case for the larger enough ν. Then we can start the two algorithms.

□

The next two kind of CO problem are studied in Microeconomics.

2. The Cost minimization problem (see [16]):

$$\begin{aligned} \min \quad & \sum_{1=1}^{m} p_i x_i \\ s.t. \quad & f(x) \leq b, \\ & x \geq 0, \end{aligned}$$

where the production function $f(x)$ is strictly convex on $\mathbb{R}^m_{++}$. □

3. The utility minimization problem (see [16]):

$$\begin{aligned} \min \quad & U(x) \\ s.t. \quad & \textstyle\sum_{i=1}^{m} p_i x_i \leq m, \\ & x \geq 0, \end{aligned}$$

where the utility function $U(x)$ is strictly convex and the price vector $p > 0$.

□

4. The following is a CO problem collected in [23]:

$$\begin{aligned} \min \quad & \sum_{i=1}^{50} i(x_i^2 + x_i^4) \\ s.t. \quad & \sum_{i=1}^{50} x_i^2 = 1 \end{aligned}$$

By letting $y_i = x_i^2$, $i = 1, 2, \cdots, 50$, and then eliminating y_{50} by writing

$$y_{50} = 1 - \sum_{i=1}^{49} y_i,$$

the above problem becomes

$$\begin{aligned} \min \quad & \sum_{i=1}^{49} i(y_i + y_i^2) + 50 \left[(1 - \sum_{i=1}^{49} y_i) + (1 - \sum_{i=1}^{49} y_i)^2 \right] \\ s.t. \quad & \textstyle\sum_{i=1}^{49} y_i \leq 1, \\ & y_i \geq 0. \end{aligned}$$

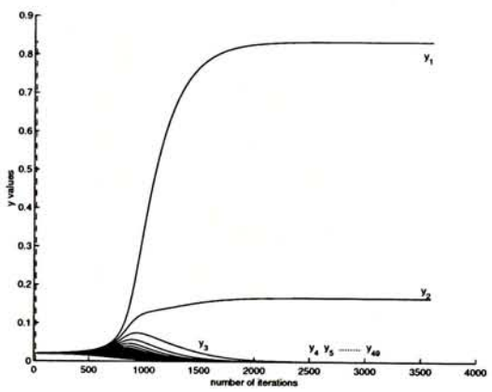
Now the feasible set becomes bounded and nonempty and we apply the Logarithmic Barrier and Center Method to this problem for some particular parameters given in the following table.

	Logarithmic Barrier		Center	
	Short-step	Long-step	Short-step	Long-step
q	-	-	$n = 50$	$n = 50$
θ	$\frac{1}{30\sqrt{50}}$	$\frac{1}{\sqrt{50}}$	$\frac{1}{22\sqrt{q}}$	$\frac{5}{11}$
Initialization Iterations	17	17	17	17
μ_0	16	16	-	-
z_0	-	-	-358	-358
Algorithm Iterations	3588	112	4434	60

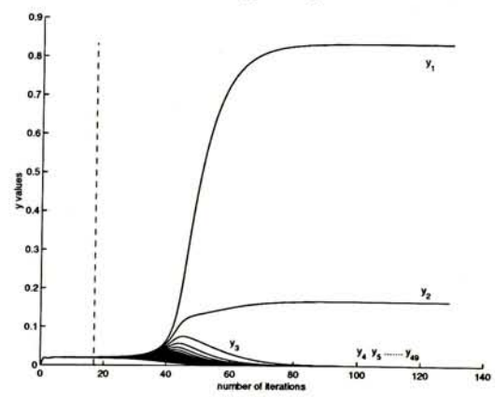
The following figures show the convergence of each values y_i in the iterated solution with different step sizes. The convergence of the y_i values in the initialization algorithm are shown in the left hand side of the vertical dashed line and the main algorithm in the right hand side of the vertical dashed line.

Logarithmic Barrier Algorithm

Short-step

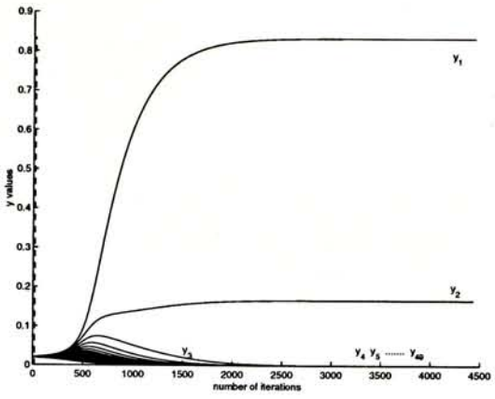


Long-step

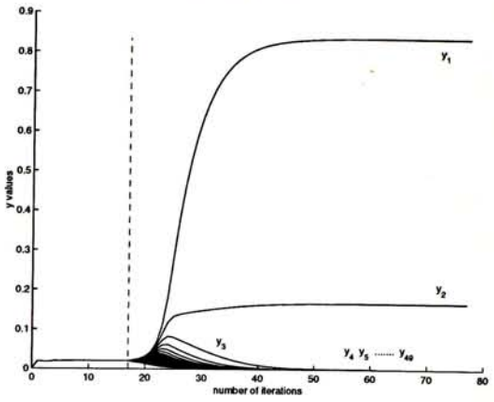


Center Algorithm

Short-step



Long-step



In the initialization algorithm of logarithmic barrier method, we reduce η by a factor of $\frac{1}{2}$ and enlarge ν by a factor of 2 simultaneously until $\nu \geq 10^5$ and $\|newton\ step\| < \frac{1}{3}$. Then we choose large enough μ_0 such that $\|newton\ step\| < \frac{1}{3}$, we get $\mu_0 = 16$.

For the center one, we reduce η by a factor of $\frac{1}{2}$ and enlarge ν by a factor of 2 simultaneously until $\nu \geq 5 \times 10^6$ and $\|newton\ step\| < \frac{1}{3}$. Then we choose small enough $z_0 = -358$ such that $\|newton\ step\| < \frac{1}{3}$. Then we can start the two algorithms. □

3.6 Remarks

We can see in the numerical experiments that the long-step algorithm is preferable for solving a general CO though the theoretical complexity of short-step algorithm is better than the long one. The long-step logarithmic barrier algorithm with $\theta = \frac{1}{\sqrt{n}}$ needs less than 200 iterations in all cases to get the same order of accuracy for the objective function in short-step. The long-step center algorithm with $\theta = \frac{5}{11}$ needs less than 50 iterations in most cases for the same targets.

In the initialization algorithm for logarithmic barrier method $\nu = 10^7$ is large enough to give $\|newton\ step\| < \frac{1}{3}$ in many cases. For center initialization algo-

rithm, $\nu = q \times 10^7$ is large enough to ensure $\|newton\ step\| < \frac{1}{3}$ in many cases. We remark that ν should be enlarged simultaneously with the reduction of η.

Bibliography

[1] Fiacco, A.V. and McCormick, G.P. (1968), Nonlinear Programming, Sequential Unconstrained Minimization Techniques, *John Wiley & Sons, New York.*

[2] Jarre, F. (1989), On the Method of Analytic Centers for Solving Smooth Convex Programs, In: S. Dolecki, ed., *Lecture Notes in Mathematics* 1405, Optimization, Springer, 69-85.

[3] Jarre, F. (1992), Interior-point Methods for Convex Programming, *Applied Mathematics & Optimization* 26, 287-311.

[4] Mehrotra, S. and Sun, J. (1990), An Interior Point Algorithm for Solving Smooth Convex Programs Based on Newton's Method, *Mathematical Developments Arising from Linear Programming,* J.C. Lagarias and M.J. Todd, eds., *Contemporary Mathematics* 114, 265-284.

[5] Den Hertog, D., Jarre, F., Roos, C. and Terlaky, T. (1995), A Sufficient Condition for Self-Concordance, with Application to Some Classes of Structured Convex Programming Problems, *Mathematical Programming, Series B* 69, 1, 75-88.

[6] Nesterov, Y.E. and Nemirovsky, A.S. (1989), Self-Concordant Functions and Polynomial Time Methods in Convex Programming, Report, Central Economical and Mathematical Institute, USSR Academy of Science, Moscow, USSR.

[7] Den Hertog, D. (1994), Interior Point Approach to Linear, Quadratic and Convex Programming, *Kluwer Academic Publishers, Dordrecht, Boston, London.*

[8] Jarre, F. (1992), Interior-point Methods for Convex Programming, *Applied Mathematics and Optimization* 26, 287-311.

[9] Terlaky, T. (1996), Interior Point Methods of Mathematical Programming, *Kluwer Academic Publishers, Dordrecht, Boston, London.*

[10] Wolfe, Ph. (1961), A Duality Theorem for Nonlinear Program, *Quarterly of Applied Mathematics* 19, 239-244.

[11] Ye, Y. and Potra, F. (1993), A Quadratically Convergent Polynomial Algorithm for Solving Entropy Optimization Problems. *SIAM Journal on Optimization* 3, 4, 843-860.

[12] Han, C.-G., Pardalos, P.M. and Ye, Y. (1991), On Interior-Point Algorithms for Some Entropy Optimization Problems, Working Paper, Computer Science Department, The Pennsylvania State University, University Park, Pennsylvania.

[13] Duffin, R.J., Peterson, E.L. and Zener, C. (1967), Geometric Programming, *John Wiley & Sons, New York.*

[14] Roos, C., Terlaky, T., Vial, J. -Ph. (1997), Theory and Algorithms for Linear Optimization, An Interior Point Approach, *John Wiley & Sons.*

[15] Böhning, D., Schlattmann, P., Dietz, E. (1996), Interval Censored Data: A note on the Nonparametric Maximum Likelihood Estimator of the Distribution Function, *Biometrika* 83, 2, 462-466.

[16] Madden, P., (1986), Concavity & Optimization in Microeconomics, *Basil Blackwell.*

[17] Gentleman, R., Geyer, C.J. (1994), Maximum Likelihood for Interval Censored Data: Consistency and Computation, *Biometrika* 81, 618-623.

[18] Finkelstein, D.M., Lakagos, S.W. (1985), A Semiparametric Model for Regression Analysis of Interval Censored Failure Time Data, *Biometrics* 41, 933-945.

[19] Jarre, F., Sonnevend, G., Stoer, J. (1988), An Implementation of the Method of Analytic Centers, *Lecture Notes in Control and Information Sciences* 111, *Springer, Berlin-New York.*

[20] Thore, S., Thompson, G.L. (1992), Computational Economics, *The Scientific Press.*

[21] Jarre, F. (1991), On the Convergence of the Method of Analytic Centers when applied to Convex Quadratic Programs, *Mathematical Programming* 49, 341-358.

[22] Mehrotra, S., Sun, J., (1991), A Method of Analytic Centers for Quadratically Constrained Convex Quadratic Programs, *SIAM Journal on Numerical Analysis* 28, 2, 529-544.

[23] Schittkowski, K. (1987), More Test Examples for Nonlinear Programming Codes, *Springer-Verlag.*

[24] Fletcher, R. (1987), Practical Methods of Optimization 2^{nd} Edition, *John Wiley & Sons Ltd.*

[25] Jansen, B. (1997), Interior Point Techniques in Optimization: Complementarity, Sensitivity and Algorithms, *Kluwer Academic Publishers.*

[26] Roos, C. (1997), Interior Point Algorithms for Linear Optimization, *Proceedings of HPOPT II held at the World Trade Center, Rotterdam,* August 20, 1997.

[27] Terlaky, T. (1997), Duality Theory for Linear Optimization, *Proceedings of HPOPT II held at the World Trade Center, Rotterdam,* August 20, 1997.

[28] Saigal, R. (1995), Linear Programming, *Kluwer Academic Publishers.*

[29] Strang, G. (1988), Linear Algebra and its Applications. *Harcourt Brace Jovanovich, Orlando, Florida, USA.*