# GRADUATE SEMINAR REPORT

# ON

# Path following methods for linear, quadratic and convex Programming

BY

YOSEPH TEKLIE

ADVISOR:

DR. SEMU MITIKU

School of Graduate Studies

ADDIS ABABA UNIVERSITY

## Preface

The seminar is devoted to the path following methods for linear, quadratic, and convex programming. The logarithmic barrier and center methods are used as path following methods. The central ideas in the analysis for both the Logarithmic barrier and the center methods are the same. First the 'Hessian norm' of the Newton search direction is always used to measure the distance to the current reference point on the central path. Second, the quadratic convergence result in the vicinity of the central path enables us to prove many other basic properties. The analysis used for linear programming problems and convex programming problems is somewhat different. The first topic discusses on the general frame work of optimization problems and the second topic is devoted on using path following methods for solving linear, quadratic, and convex programming problems. The last topic discusses on non-path following methods like Affine scaling method, Karmarkars method for linear programming and the steepest descent methods.

## Glossary of Symbols and Notations

$x_i$ : i-th coordinate of the vector x

$x^T$ : transpose of the vector x

e : vector of all 1's of appropriate length, i.e. $e^T=(1,\ldots,1)^T$

$R^n$ : n-dimensional Euclidean space

$\|x\|$ : 2-norm of the vector x

(LP): primal linear programming problem

(LD): dual linear programming problem

(QD): primal quadratic programming problem

(QD): Dual quadratic programming problem

$\mu$ : barrier parameter

$\mu_0$ : initial barrier parameter

$\phi_B$ : logarithmic barrier function

$\phi_D$ : distance function of Huard

$x(\mu), y(\mu)$ : minimizing point for $\phi_B$

$\delta(y,\mu)$ : distance measure to $y(\mu)$

CONTENTS

## 1.1 Problem statement

**Linear programming:**

Linear inequality systems were studied by Fourier Motzkin and Farkas (114) but only in the late 1940s did the work of Dantizig in the USA and Kantorovich make linear programming a leading research topic with wide research topic with wide ranging applications to problems of planning and scheduling as found in business, industry, and government. Dantizig developed the celebrated simplex method (In the tabulae form) for finding the maximum or minimum of a linear function subject to linear restrictions. The resulting optimization problem because of its planning applications was called linear programming (LP)

The primal formulation of linear program is as follows. Given a finite number of linear equations and inequalities.

$$\begin{aligned} a_i^T x &= b_i \\ a_j^T x &\le b_j \end{aligned} \qquad i = 1,2,\ldots,k,\ j = k+1,\ldots,m$$

and a linear function $C^T x$, $c, x \in R^n$

A primal formulation of linear program is

$$\text{(LP)} \qquad \begin{aligned} & C^T x \to \min \\ & Ax \le b \qquad \text{where } A=(mxn) \\ & x \ge 0 \end{aligned} \tag{1.1}$$

Where $b \in R^m$ A slightly simplified (the mathematically equivalent) form of the primal problem i.e. the Dual problem can be stated as

# 1. General framework of optimization problems

Broadly speaking, finding the optimizing (maximizing or minimizing) point of a given function over a given domain is an optimization problem. An optimization problem can be classified as constrained and unconstrained. An unconstrained optimization problem is the problem of finding the minimizing point of a non-linear function of n real variables.

$f : R^n \to R$ And it is denoted as

$$\min_{x \in R^n} f(x)$$

A constrained optimization problem involves minimization of a function subject to constraints on a finite set of continuous variables, i.e. to determine $x^* = (x_1^*, x_2^*, \ldots, x_n^*)^T$ that solves the problem

$$\min f(x) \text{ Subject to. } \quad \text{(A)}$$

$$g_i(x) \geq 0, i = 1,2,\ldots n$$

$$h_i(x) = 0, j = 1,\ldots p$$

When the function $f, \{g_i\}, and \{h_i\}$ are all linear, Problem A is called a linear programming problem. If any of the function is non-linear the problem A is called a non-linear programming problem. There are other terms such as convex, quadratic, concave, separable and factorable, which may apply to special cases of problem A and this will be defined in the following subsections.

$$b^T y \rightarrow \max$$

$$\text{(DP)} \quad A^T y \leq c \tag{1.2}$$

(1.2) can be converted into equations as

$$\begin{aligned} & b^T y \rightarrow \max \\ & A^T y + z = c \\ & z \geq 0 \end{aligned} \tag{1.3}$$

Where Z is an n-vector of dual slack variables.

**Convex programming problem**

A set, which contains the entire line through any two of its points, is an affine set. Mathematically we can rewrite this statement as: a subset M of $R^n$ is called affine if $(1-\lambda)x + \lambda y \in M, \forall x, y \in M$ and $\lambda \in R$. A subset K of $R^n$ is said to be convex if it contains the segment joining any two of its points i.e. $(1-\lambda)x + \lambda y \in K, \forall x, y \in K$ and $0 < \lambda < 1$. All affine sets are convex while the converse may not always be true. A function $f : R^n \rightarrow R$ is said to be convex if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ $, \forall x, y \in R^n, \lambda \in R$.

It is recognized that in general given information at a point, one can only make statements about what is true in a neighborhood about that point. For an important class of problems called convex programming problems, local information is also global information. Convex programming problems have a desirable property that local minima are global minima.

A convex programming problem is written as

$$
\text{(CP)} \qquad \begin{array}{ll} \max f(x) & \\ g_i(x) \le 0 & \quad i=1,2,\ldots,n \\ x \in R^m & \end{array}
$$

the feasible region is denoted by S, and the interior of this region by $S^o$.

One special case of convex programming problems is quadratic programming problem. Such a problem in its standard form is given as:

$$
\text{(QP)} \qquad \begin{array}{l} \min Q(x) = c^T x + \frac{1}{2} x^T Q x \\ Ax = b, x \ge 0 \end{array}
$$

Where Q : is symmetric, positive semi definite nxn matrix

A: mxn matrix, $b \in R^m$, $c \in R^n$, and $x \in R^n$ is the variable in which the minimization carried out.

## 1.2. Methods of solving constrained optimization problems

The classical method of solving a non-linear constrained optimization problems.

$$
\text{(P)} \quad \text{Max}\{f(x) : x \in R^m, f_i(x) \le 0, i = 1,..,n\}
$$

can roughly be distinguished into two classes interior point methods and exterior point methods . The interior point methods operate in the interior of the feasible region in searching the optimum point where as the exterior point method try to obtain the solution by operating from the outside the feasible region but approaching to it.

### 1.2.1 Simplex method

Linear programming (LP) was always be considered as distinct from non linear programming because of its combinatorial structure. Dantizig's simplex algorithm for solving linear optimization had always been an unbeaten tool for solving the linear programming till the emergence of Karmarks algorithm in 1984. The simplex method uses explicitly the combinatorial structure: it obtains an optimum by moving along the edges of the feasible region. The simplex method is an algorithm, which determines an optimal vertex of a pointed polyhedron: $S = \{x \in R^n : Ax \leq b\}$ provided the linear program $\max\{c^T x : Ax \leq b\}$ has an optimal solution.

The geometrical idea of a simplex algorithm is to move from vertex to vertex by successively increasing the objective function value $c^T x$ along the sides of the simplex. It consists of two main phases. In the first phase one determines a vertex of a feasible region or reports that the feasible region is empty. In the second phase, if the feasible region is non-empty, one moves from vertex to vertex-using simplex steps. Since a pointed polyhedron has only a finite number of vertices and since the objective function value is increased while moving from vertex to vertex. The second phase stops after a finite number of moves with either an optimal vertex or with an edge of the polyhedron along which the objective function can be arbitrarily increased.

### 1.2.2 Interior point methods

The philosophy of interior point methods is quite different from the simplex method. In general, they construct successively an infinite sequence of points in the interior of the feasible region. This method can be explained in the content of linear programs of the form:

$$\text{(LP)} \quad \min\{c^T x : Ax = b, x \geq 0\}.$$

If the feasible region S= $\{x \in R^n_+ : Ax \leq b, x \geq 0\}$. is non empty; then its relative interior is the set denoted by $S^o = \{x \in R^n_+ : Ax = b, x > 0\}$.

A point $x \in S_0$ is called a strictly feasible point of LP. Interior point algorithms start from a strictly feasible point and construct sequence of strictly feasible points, which eventually converge to an optimal solution of LP. A specific stopping criterion is used to interrupt the construction of the sequence after a finite number of iterations and to jump from the current iterate to an optimal solution.

Shortly, the general idea of interior point algorithms is, if $x_0 \in S_0$, then the set of all feasible directions at $x_0$ is the linear subspace $\{y \in R^n : Ay = 0\}$. So at first glance a reasonable search direction may be calculated as the solution $y_o$ of the problem:

$$\min\{y \in R^n : Ay = 0, \|y\|_2 \leq 1\}.$$

the direction $y_o$ is called the steepest descent direction corresponding to the problem LP.

Another important property of interior point methods is that some of the methods can also be applied to non-linear optimization problems. In fact the main idea of this method was developed for the non-linear programming problems already in 1960's. This leaded to the rebirth of some of the old methods, the logarithmic barrier method of Frisch and Fiacco and MC cormick, the center method of Huard and the affine scaling method of Dikin. The logarithmic barrier methods and the center methods are called path following the central path of the problem. We will discuss path following methods for linear, quadratic and convex programming problems respectively in the following sections.

## 2.0 PATH FOLLOWING METHODS

Many interior point algorithms may be thought of as homotopy methods. The general idea underlying all homotopy methods is quite natural and mathematically appealing. In order to determine the solution $x_o$ of a numerical problem we first embed the problem $P_o$ into a suitable family of problems (possibly simpler) $P_\mu$ $\mu \in (0, \mu_o)$ with solution $x(\mu)$ in such a way that

$$\lim_{\mu \to 0^+} x(\mu) = x_o$$

And then follow the solution path $x(\mu)$ as t approaches 0. To be more precise, given such a family of problems $P_\mu$ the solution $P_o$ is approximated better and better by successively constructing a decreasing null sequence $\mu_o, \mu_1, \ldots$ together with a sequence $\hat{x}_o, \hat{x}_1, \ldots$ of approximate solutions of problems $P_o, P_1, \ldots$. This method is referred as path following method. Path following methods can be short step, medium and long step method methods. Short Step path following methods follow the central path closely, where as the medium and the long step methods follow the central path loosely. The short step methods are unattractive in practice since they use fixed short steps and small updates in the parameter, and therefore require many iterations. On the other hand medium and long step path following methods are more much flexible since they allow to do large updates in the parameter and line searches. I will discuss path following methods for linear, convex and quadratic problems based on the classical barrier method and Huard's distance function. Both functions parameterize the so called central path, use Newton steps to reach the vicinity of the current sequence point on the central path, when after the barrier parameter denoted by $\mu$ or the lower bound parameter denoted by z are updated, which means that a new reference point is defined

## 2.1. LOGARITHMIC BARRIER METHOD

The logarithmic barrier method is a path following interior point method in which the objective function is combined in a complete function, called the barrier function, such that it contains singularities at the boundary. Because of these singularities all iterates will remain strictly feasible while approaching the optimal solution. Frisch(152) introduced the barrier method for linear programming.

In this method, the logarithmic barrier function is minimized exactly for decreasing values of the barrier parameter. This minima form the so-called central path of the problem, which is a smooth curve in the interior of the feasible region and ends in an optimal solution.

Consider the following convex programming

$$\text{(CP)} \qquad \begin{aligned} &\max f_o(y) \\ &f_i(y) \le 0, i = 1,2,\ldots n \\ &y \in R^m \end{aligned} \qquad (2.1)$$

the feasible region is denoted by S and the interior of the feasible region by S°. We will assume that:

1. The functions $-f_o(y)$ and $f_i(y)$, $1 \le i \le n$ are convex functions with continuous first and second derivatives in S°.
2. S° is non empty
3. S° is bounded.

The Wolfe's [152]. Formulation of the dual problem will be

$$\text{(CD)} \qquad \begin{aligned} &\min f_o(y) - \sum_{i=1}^{n} x_i f_i(y) \\ &\sum_{i=1}^{n} x_i \nabla f_i(y) \\ &x_i \geq 0 \end{aligned} \qquad (2.2)$$

Now the logarithmic barrier function ($\phi_\beta$) associated with (CP) is

$$\phi_\beta(y,\mu) = \frac{-f_o(y}{\mu} - \sum_{i=1}^{n} \ln(-f_i(y)$$

Where $\mu > 0$ is the barrier parameter. Because of the singularity of the logarithm at zero, this barrier function will prevent the iterates from going outside the feasible region. It is because f this that we call the logarithmic barrier method an interior point method.

For the description of the method we need the first and second derivatives of $\phi_\beta(y,\mu)$

$$g(y,\mu) := \nabla\phi_\beta(y,\mu) = \frac{-\nabla f_o(y)}{\mu} + \sum_{i=1}^{n} \frac{\nabla f_i(y)}{-f_i(y)} \qquad (2.3)$$

$$H(y,\mu) := \nabla^2\phi_\beta(y,\mu) = \frac{-\nabla^2 f_o(y)}{\mu} + \sum_{i=1}^{n} [\frac{\nabla^2 f_i(y)}{-f_i(y)} + \frac{\nabla f_i(y)\nabla f_i(y)^T}{f_i(y)^2}] \qquad (2.4)$$

For shortness sake we use H instead of $H(y,\mu)$ and $g$ instead of $g(y,\mu)$

Clearly H is positive definite and hence $\phi_\beta(y,\mu)$ is strictly convex on its bounded domain S° and takes infinite values on the boundary of S This function achieves the minimal value in its domain (for fixed $\mu$) at a unique point which is denoted by $y(\mu)$ and called $\mu$-center. The necessary and sufficient Karush -Kuhn - Tucker conditions are for the optimality of $y(\mu)$ are:

$$
\begin{aligned}
& f_i(y) \leq 0, 1 \leq i \leq n \\
& \sum_{i=1}^{n} x_i \nabla f_i(y) = \nabla f_o(y), x \geq 0 \\
& -f_i(y) x_i = \mu, 1 \leq i \leq n
\end{aligned}
\tag{2.5}
$$

### 2.1.1 CENTRAL PATH

The primal (dual) central path (trajectory) is defined as the set of centers $y(\mu),(x(\mu),y(\mu))$, where $\mu$ runs from $\infty$ to 0. Here not only $y(\mu)$ but also $(x(\mu),y(\mu))$, is primal feasible.

It is well known that $x(\mu)$ and $y(\mu)$ are continuously differentiable. Hence, it holds that $x(\mu)$ and $y(\mu)$ will converge to optimal solutions of (CD) and (CP) if $\mu$ tends to 0 from the right. This means that the central trajectory ends in an optimal solution of the problem. Moreover note that if the feasible region is bounded then the central path starts in a unique point where $\sum_{i=1}^{n} \ln(-f_i(y)$ is maximized since

$$\lim_{\mu \to \infty} \phi_\beta(y,\mu) = \sum_{i=1}^{n} \ln(-f_i(y)$$

This point is called the analytic center of the feasible region.

In logarithmic barrier method the original constrained problem is replaced by a sequence of unconstrained optimization problems i.e. minimizing $\phi_\beta(y,\mu)$ successively for a sequence of positive decreasing values of the barrier parameter $\mu$. Loosely speaking, this method follows the central path approximately to an optimal point.

There are some important elements in the design of this method.

Mainly we need to get:

1. The lan algorithm used to (approximately) minimize $\phi_\beta(y,\mu)$.
2. The criterion to terminate the approximate minimization.
3. The updating scheme for the barrier parameter $\mu$.

Before we proceed in to the details of the method, we require the following definition.

**Definition:** $\|.\|_H$ is used to measure distances between points. The measure is defined as follows:

$$\|Z\|_H = \sqrt{Z^T HZ}$$

Where H is the Hessian matrix $\phi_\beta$. $\|.\|_H$ is a norm sice H is positive definite.

**LOGARITHMIC BARRIER ALGORITHM**

The algorithm does a line search along the Newton direction. The Newton direction is defined as $P = -H^{-1}g$. The criterion to terminate the approximation minimization of $\phi_\beta(y,\mu)$ is $\|P\|_H \le \tau \le 1$. Where $\tau$ is the proximity parameter. Note that $\|P\|_H = 0$ if and only if $y = y(\mu)$.

Algorithm for finding an $\varepsilon$-optimal solution.

Input: $\varepsilon$ : is the accuracy parameter.

$\tau$ : is the proximity parameter.

$\theta$ : is the reduction parameter

$\mu_o$ : is the initial barrier value

$y^o$ : is a given interior feasible point such that $\left\|P(y^o,\mu_o)\right\|_{H(y^o,\mu_o)} \le \tau;$

Begin

$y := y_o; \mu := \mu_o;$

While $\mu > \dfrac{\varepsilon}{4n}$ do;

begin(outer step);

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

while $\|P\|_H \ge \tau$ do

Begin (inner step)

$\hat{\alpha} := \arg\min_{\alpha>0}\{\phi_\beta(y+\alpha p,\mu) : y+\alpha p \in S^o\}$

$y := y + \hat{\alpha} p$

end (inner step)

end (outer step).

We call an algorithm long step if $\theta$ is constant $(0 < \theta < 1)$,

If $\theta = \dfrac{v}{\sqrt{n}}$, where v>0 and independent of n & $\varepsilon$, we call the algorithm a short step algorithm.

**Example 1.** Let's consider the following simple linear programming problem (m=2 and n=4):

$$\begin{cases} \max 2y_1 + y_2) \\ y_1 \le 1 \\ -y_1 \le 1 \\ y_2 \le 1 \\ -y_2 \le 1 \end{cases}$$

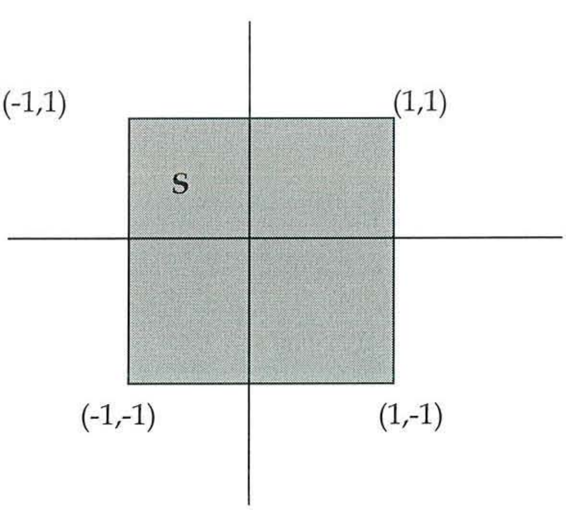



Fig 2.1

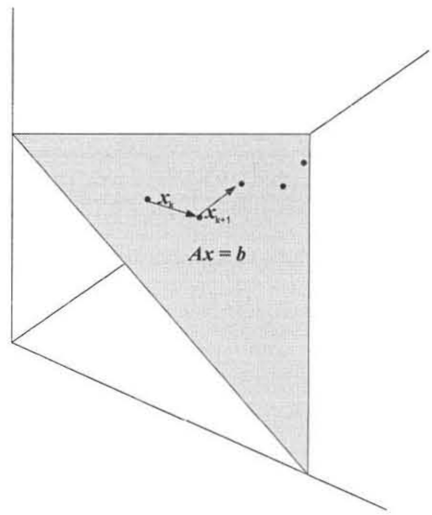
From the barrier function for the above problem we can calculate the first derivative of the barrier function to be equal to zero with respect to $y_1$ and $y_2$ .

i.e. $y_1(\mu) = \dfrac{1}{2}\sqrt{\mu^2 + 4} - \dfrac{1}{2}\mu$

$$y_2(\mu) = \sqrt{\mu^2 + 1} - \mu$$

Consequently by taking the limit at 0 for y1 and $y_2$ we can find the optimal solution to be at (1,1). Moreover, the central path starts in (0,0), which is the analytic center of the feasible region.

**Example 2.** If the feasible region of a linear program is a cone then the central path is a line, ending in the top of the cone. Again suppose that the constraints of the linear




problem describe a cone in $R^m$.

More over, let the objective function be such that the top of the cone is an optimal solution. Let the constraint matrix be $A^T$, which is $m \times n$ matrix and let $b \in R^m$ so the problem is as follows;

$$\begin{cases} \max b^T y \\ A^T y \leq c \end{cases}$$

Note that $A^{-T}c=t$ is the top of the come. Te distance function for this problem is

$$\phi_D(y,z) = -\frac{b^T y}{\mu} - \sum_{i=1}^{m} \ln(c_i - a_i^T y)$$

Where $a_i$ is the $i^{th}$ row of $A^T$. It can be verified the Karush-kuhn-Tucker condition for the Dual problem (LD) as given in (2.5) can show

$$x(\mu) = A^{-1}b$$

$$s(\mu) = \mu X^{-1}e$$

We can calculate $y(\mu)$ in a similar way as we calculated $y(\mu)$

$$y(\mu) = A^{-T}(c - s) = t - \mu v \text{ , Where } v_i = \frac{1}{(A^{-1}b)_i}$$

Consequently the path of centers is a line ending at the optimal point.

### 2.1.1 LINEAR OPTIMIZATION PROBLEMS

The logarithmic barrier method was original introduced by Frisch for linear programming and for convex programming. However he apparently minimized this function sequentially, but only used the gradient of it to steer away from the boundary. . Parisot further developed the logarithmic barrier method and devised sequentially unconstrained minimization techniques for linear programming.

Consider the dual linear program

$$\text{(DP)} \qquad \begin{cases} \max b^T y \\ A^T y \le c \end{cases}$$

Where A is an mxn matrix, $b \in R^m, c \in R^n, y \in R^m$. Y is the variable in which minimization carried out.

The slack variable s is defined as $s := c - A^T y$. The primal problem for (DP) will be:

$$(LP) \qquad \begin{cases} \min c^T x \\ Ax = b, x \geq 0 \end{cases}$$

We assume that all coefficients are integral and the feasible set for (LP) has a non-empty interior and that the optimal set is bounded.

The logarithmic barrier function for (DP) is

$$\phi_\beta(y,\mu) = \frac{-b^T y}{\mu} - \sum_{i=1}^{n} \ln s_i \qquad 2.6$$

Where $\mu$ is the positive parameter. The first and second derivatives of $\phi_\beta$ are

$$g =: \nabla\phi_\beta(y,\mu) = \frac{-b}{\mu} + As^{-1}e, \quad \text{Where} \quad s^{-1} = \begin{bmatrix} s_1^{-1} & & & \\ & s_2^{-1} & & \\ & & \cdot & \\ & & & s_n^{-1} \end{bmatrix}$$

$$\text{and } H =: \nabla^2\phi_\beta(y,\mu) = As^{-2}A^T, \text{ where } s^{-2} = \begin{bmatrix} s_1^{-2} & & & \\ & s_2^{-2} & & \\ & & \cdot & \\ & & & s_n^{-2} \end{bmatrix}$$

Thus, $\phi_\beta$ achieves a minimum value over S at a unique point. The Karush-Kuhn-Tucker Conditions are:

$$\begin{cases} A^T y + s = c, s \geq 0 \\ Ax = b, x \geq 0 \\ Sx = \mu e \end{cases} \qquad 2.7$$

The primal logarithmic barrier function is given by

$\phi_\beta(y,\mu) = \frac{c^T x}{\mu} - \sum_{i=1}^{n} \ln x_i$ We denote $(x(\mu), y(\mu), s(\mu))$ is the unique solution of (2.7).

**<u>Lemma 2.1.1</u>** If $\mu$ decreases then the objective function $c^T x(\mu)$ of the primal problem (LP) is monotonically decreasing and the objective function $b^T y(\mu)$ of the dual problem (LD) is monotonically increasing.

We introduce the following measure for the distance of an interior of feasible point y to $y(\mu)$

$$\delta(y,\mu) := \min_x \{ \left\| \frac{s_x}{\mu} - e \right\| : Ax = b \} \quad \text{where } s_x := s(\mu) \text{ and } e=(1,1,\ldots,1).$$

The unique solution of the minimization problem in the definition of $\delta(y,\mu)$ is denoted by $x(y,\mu)$ so we can also write

$$\delta(y,\mu) := \left\| \frac{s_x(y,\mu)}{\mu} - e \right\|$$

It can be verified that $y = y(\mu)$ if and only if $\delta(y,\mu) = 0$ and moreover,

$\delta(y,\mu) = 0$ if and only if $x(y,\mu) = x(\mu)$.

The Newton direction $p(y,\mu)$ in y with respect to $\phi_\beta(y,\mu)$ is given by

$$p(y,\mu) = -H^{-1}g = (As^{-2}A^T)^{-1}(b/\mu - As^{-1}e).$$

The next lemma states that there is a close relationship between the $\delta$-measure and the Hessian norm of the Newton direction $p(y,\mu)$. It can be shown that for a given y and $\mu$ we have

$$\delta(y,\mu) := \|p(y,\mu)\|_H = \|s^{-1}A^T p(y,\mu)\| = \|s^{-1}A^T p(y,\mu)\| \qquad \textbf{2.8}$$

**<u>Lemma 2.1.2</u>** If $\delta(y,\mu) \le 1$, then $y^+ = y + p$ is a strictly feasible point for LD.

Moreover $\delta(y^+,\mu) =\le \delta(y,\mu)^2$

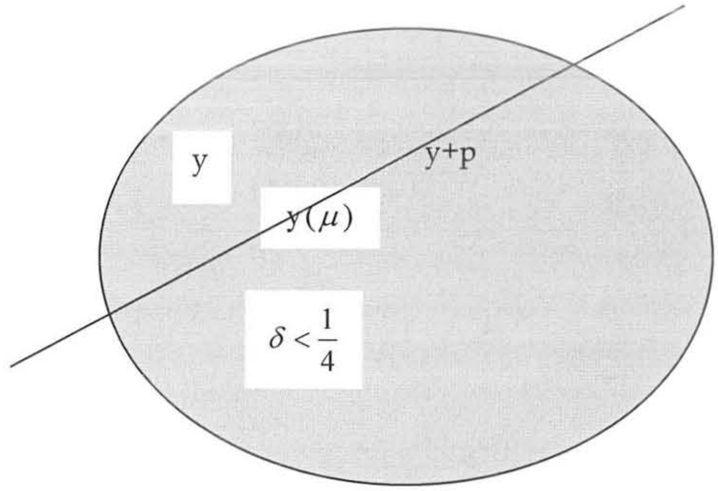



Fig 2.3 Quadratic convergence in the vicinity of the $\mu$-center.

**Theorem 2.1.1** If $\delta = \delta(y,\mu) \le 1$ then $\phi_\beta(y,\mu) - \phi_\beta(y(\mu),\mu) \le \dfrac{\delta^2}{1-\delta^2}$

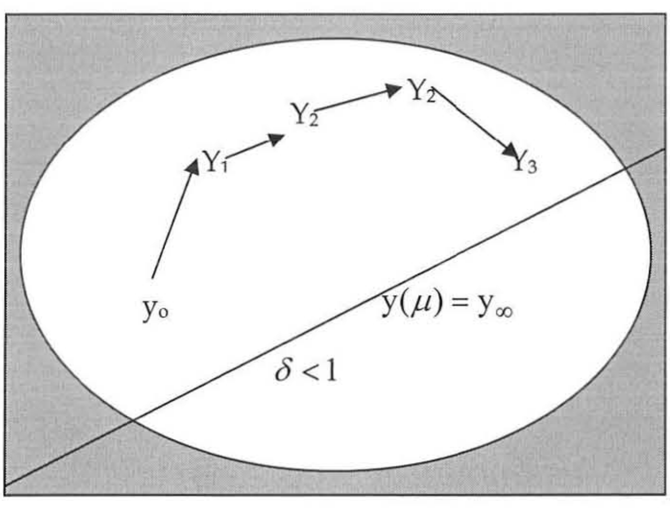



Fig. 2.4 A sequence of Newton steps converging to the $\mu-$center.

**Proof:** The barrier function $\phi_\beta$ is convex in y, hence

$$\phi_\beta(y+p,\mu) \ge \phi_\beta(y,\mu) + p^T g$$

Now, using $p = -H^{-1}g$, we have

$p^T g = -p^T Hp = -\delta^2$, Substituting this gives

$\phi_\beta(y+p,\mu) - \phi_\beta(y,\mu) \le \delta^2$ (*)

Now let $y_o := y$ and let $y_1, y_2,..$ denote the sequence of points obtained by repeating the Newton steps, starting at $y_o$. By Lemma 2.1.2 we have

$\delta(y_i,\mu) \le \delta(y_i,\mu)^{2^i} = \delta^{2^i}$ (**)

This means that $\delta(y_i,\mu) \to 0$ if $i \to \infty$. Since all $y_i$ lie in a bounded region and $\delta(y,\mu) = 0$ if and only if $y = y(\mu)$ we have that all the limit points of the sequence

are $y(\mu)$. Hence the sequence of Newton iterates converge to $y(\mu)$.

So using (*) and (**), we may write:

$$\phi_\beta(y,\mu) - \phi_\beta(y(\mu),\mu) = \sum_{i=0}^{\infty} \phi_\beta(y_i,\mu) - \phi_\beta(y^{i+1},\mu)$$

$$\le \sum_{i=0}^{\infty} \delta(y_i,\mu)^2 \le \sum_{i=0}^{\infty} \delta^{2^{i+1}} \le \frac{\delta^2}{1-\delta^2} \qquad //$$

The following proposition gives an upper bound for the difference in the objective value

**Proposition 2.1.1** If $\delta = \delta(y,\mu) < 1$, then $\left|b^T y - b^T y(\mu)\right| \le \dfrac{\delta(1+\delta)}{1-\delta}\mu\sqrt{n}$

Note that if $\delta(y,\mu) = 0$. Then according to Theorem 2.1.1 and Proposition 2.1.4

We have $\phi_\beta(y,\mu) = \phi_\beta(y(\mu),\mu)$ and $b^T y = b^T y(\mu)$ which is obviously true s since $y = y(\mu)$.

## ILLUSTRATION OF THE NEWTON PROCESS

Lets see the Newton process by means of an example. Consider the following simple linear programming problem.

$$\max y_1 + 2y_2$$

$$y_{1\le 0}$$

$$y_2 \le 0$$

It can easily be verified that the central path is given by $y(\mu) = -\mu\begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}$.

The Newton direction is given by:

$$p(y,\mu) = (As^{-2}A^T)^{-1}(\frac{b}{\mu} - As^{-1}e) = \begin{pmatrix} \frac{y_1}{\mu}+1 \\ \frac{2y_2}{\mu}+1 \end{pmatrix} \qquad (*)$$

According to (2.8) we have that

$$\delta(y,\mu)^2 = \left\|s^{-1}p_s\right\|^2 = (\frac{y_1}{\mu}+1)^2 + (\frac{2y_2}{\mu}+1)^2 \qquad (**)$$

Consequently for this example the boundary of the quadratic convergence region $\delta(y,\mu) < 1$ is an ellipsoid and its center is $y(\mu) = (-\mu, \frac{-\mu}{2})$. Let $\mu = 1$ with $y_1(1)$=-1 and $y_2(1)$=$\frac{-1}{2}$. Let $y^o = \begin{pmatrix} \frac{-3}{2} \\ \frac{-3}{4} \end{pmatrix}$ since, using (**) $\delta(y^o,1) = \sqrt{\frac{1}{4}+\frac{1}{4}} = \sqrt{\frac{1}{2}}$

we have that $y^0$ is the quadratic convergence region y(1). Lets us verify that by taking Newton steps and calculating the $\delta$-measure. Let $y^1$ and $y^2$ denote the points obtained by doing one and two full Newton steps respectively. For $y^1$ we obtain using (*).

$$y^1 = y^o + p(y^o,1)$$

$$= \begin{pmatrix} 2y_1^o + (y_1^o)^2 \\ 2y_2^0 + 2(y_2^0)^2 \end{pmatrix} = \begin{pmatrix} \frac{-3}{4} \\ \frac{-3}{8} \end{pmatrix}$$

Now we calculate the $\delta$-measure in $y^{2:}$

$$\delta(y^1,1) = \sqrt{\frac{1}{16} + \frac{1}{16}} = \frac{1}{2}\sqrt{\frac{1}{2}} \ .$$

Note that $\delta(y^1,1) = \delta(y^o,1)^3 < \delta(y^0,1)^2$

Then the next Newton iterate is :

$$y^2 = y^1 + p(y^1,1) = \begin{pmatrix} 2y_1^1 + (y_1^1)^2 \\ 2y_2^1 + 2(y_2^1)^2 \end{pmatrix} = \begin{pmatrix} -\frac{15}{16} \\ -\frac{15}{32} \end{pmatrix}$$

Now we calculate the $\delta$ – measure in $y^2$

$$\delta(y^1,1) = \sqrt{\frac{1}{256} + \frac{1}{256}} = \frac{1}{8}\sqrt{\frac{1}{2}}$$

Again, note that $\delta(y^2,1) = \sqrt{\frac{1}{2}}\delta(y^1,1)^2 < \delta(y^1,1)^2$. Which is again in accordance with Lemma 2.1.2. Let us now consider a point the quadratic convergence region, say $y^o$ =(-5,-2). For which we have $\delta(y^o,1) = 5$. Let's take a step $\alpha$ along the Newton direction. So for the next point $y^1$ we have Newton direction. So for the next point $y_1$ we have:

$$y^1 = y^0 + \alpha p(y^o,1) = \begin{pmatrix} y_1^o \\ y_2^o \end{pmatrix} + \alpha \begin{pmatrix} 2y_1^o + (y_1^o)^2 \\ 2y_2^o + 2(y_2^o)^2 \end{pmatrix} = \begin{pmatrix} -5 + 20\alpha \\ -2 + 6\alpha \end{pmatrix}$$

Then the Logarithmic barrier function value will decrease by at least 5-ln(1+5) (which is about 3.21) by taking $\alpha = \frac{1}{6}$. Since we have

$$\phi_\beta(y,1) = -y_1 - 2y_2 - \ln(-y_1) - \ln(-y_2),$$

the difference is logarithmic barrier function value between $y^o$ and $y^1$ is :

$$\phi_\beta(y^o,1) - \phi_\beta(y^1,1) = 32\alpha + \ln(1-4\alpha) + \ln(1-3\alpha),$$

## Acknowledgement

First, I want to thank my Advisor Dr. Semu Mitiku for stimulating and enthusiastic collaboration, and for improving this work by injecting many of his ideas and correcting some mistakes in the paper. I also would like to express my gratitude for my examiner and Instructor Prof. Dr. rer.nat.habil.R.Deumlich for valuable cooperation during the seminar work. A heart felt thank is also to my families for their continuous encouragement. Lastly, I would like to extend many thanks all friends and classmates.

by differentiating the right hand side we obtain the logarithmic barrier function is minimized

for $\alpha = \frac{20}{9}$ , given a reduction of about 3.84.

### 2.1.3 Logarithmic Barrier Method for Quadratic optimization problems

We consider the convex quadratic optimization problem:

$$\text{(QP)} \qquad \begin{aligned} &\min q(x) := c^T x + \frac{1}{2} x^T Q x \\ &Ax = b, x \geq 0 \end{aligned} \qquad 2.9$$

Here Q is a symmetric, positive semi definite, nxn matrix. A is an mxn matrix, $b \in R^n, c \in R^n$, where x is the variable in which the minimization is carried out. The dual formulation of QP is

$$\begin{aligned} &\max d(x, y) := b^T x - \frac{1}{2} x^T Q x \\ &A^T y - Qx + s = c, s \geq 0, y \in R^m \end{aligned} \qquad 2.10$$

It is well known that for all x feasible for (QP) and $(\bar{x}, y)$ feasible for (QD).

We have

$$d(\bar{x}, y) \leq z^* \leq q(x);$$

Where $z^*$ denote the optimal objective value of (QP). Optimality holds if and only of the complimentary slackness relation $x^T s = 0$ is satisfied and then q(x)=d$(\bar{x}, y)$.

We assume that (QP) has a strictly feasible solution such that x>0, and that the feasible set of (QP) is bounded.

Consider the Logarithmic Barrier function:

$$\phi_\beta(y,\mu)=\frac{c^T x+\frac{1}{2}x^T Qx}{\mu}-\sum_{i=1}^{n}\ln x_i \qquad \text{.............. } 2.11$$

Where $\mu$ is a positive parameter. The first and second derivatives of $\phi_\beta$ will be denoted by $g=g(x,\mu)$ and $H=H(x,\mu)$ respectively.

So

$$g=\nabla\phi_\beta(x,\mu)=\frac{c+Qx}{\mu}-x^{-1}e$$

$$H=\nabla^2\phi_\beta(x,\mu)=\frac{1}{\mu}Q+x^{-2},$$

Where $x^{-1}=\begin{bmatrix} x_1^{-1} & & \\ & x_2^{-1} & \\ & & x_n^{-1}\end{bmatrix}$ and $x^{-2}=\begin{bmatrix} x_1^{-2} & & \\ & x_2^{-2} & \\ & & x_3^{-1}\end{bmatrix}$

It can be verified that under the assumptions made the logarithmic barrier function has a unique minimum. The necessary and sufficient first order optimality Karush-Kuhn-Tucker conditions for this unique minimum of $\phi_\beta(x,\mu)$ are:

$$\begin{cases} A^T y-Qx+s=c, s\geq 0 \\ Ax=b \\ Xs=\mu e \end{cases} \qquad \text{Where } y\in R^m, s\in R^n$$

the unique solution of this system is again denoted by $(x(\mu),y(\mu),s(\mu))$. The following proposition states that the primal objective decreases along the primal path and the dual objective increases along the dual path.

**Proposition 2.1.1** The objective $q(x(\mu))$ of the primal problem (QP) is monotonically decreasing and the objective $d(x(\mu),y(\mu))$ of the dual problem (QD) is monotonically increasing if $\mu$ increases.

Analogous to linear programming , we use $\delta(x,\mu)$ to measure the distance to the central path of non centered points.

$$\delta(x,\mu) := \min_{y} \left\| \frac{x}{\mu}(c + Qx - A^T y) - c \right\|$$

It can be verified that $x = x(\mu)$ if and only if $\delta(x,\mu)=0$

And moreover $\delta(x,\mu)=0$ if and only of $y(x,\mu) = y(\mu)$.

In the algorithm approximate line searches along projected Newton direction are carried out to minimize $\phi_\beta$ for fixed $\mu$ i.e the directions corresponding to exact minimization of the quadratic approximation $\phi_\beta$ on the affine subspace Ax=b. This means that the

Newton direction $p = p(x,\mu)$ is determined by solving

$g + H_p = A^T \bar{y}, Ap = 0$ as a linear system of equations in p and $\bar{y}$.

**Example:** Lets consider the following simple convex quadratic programming problem. (m=1,n=1)

$$(QP) \qquad \begin{array}{l} \max - y_1 - y_2 \\ y_1^2 + y_2^2 \le 1 \end{array}$$

The logarithmic barrier function associated with (QP) is

$$\phi_\beta(y,\mu) = \frac{y_1 + y_2}{\mu} - \ln(1 - y_1^2 - 2y_2^2)$$

This function is minimized by setting the derivatives (with respect to $y_1$ and $y_2$) equal to zero. Solving this equations we obtain

$$y_1(\mu) = \frac{2}{3}\mu - \frac{1}{3}\sqrt{4\mu^2 + 6}$$

$$y_2(\mu) = \frac{1}{3}\mu - \frac{1}{6}\sqrt{4\mu^2 + 6}$$

Note that the central path really ends in the optimal solution $(\frac{-2}{\sqrt{6}}, \frac{-1}{\sqrt{6}})$

Since $\lim_{\mu \to 0} y_1(\mu) = \frac{-2}{\sqrt{6}}$ and $\lim_{\mu \to 0} y_2(\mu) = \frac{-1}{\sqrt{6}}$

More over, This central path starts in (0,0), the analytic center of the feasible region ; $\lim_{\mu\to\infty} y_1(\mu) = \lim_{\mu\to\infty} y_2(\mu) = 0$.

### 2.1.3 Logarithmic barrier method for convex optimization.

Lets recall the primal convex programming problem

$$(CP) \qquad \begin{array}{l} \max f_o(y) \\ g_i(y) \le 0 \quad \text{and} \\ y \in R^m \end{array}$$

the lagrange dual formulation;

$$(CD) \qquad \begin{array}{l} \min f_o(y) - \sum_{i=1}^{n} x_i f_i(y) \\ \sum_{i=1}^{n} x_i \nabla f_i(y) = \nabla f_o(y) \\ x_i \ge 0 \end{array}$$

We will make the following assumptions $-f_o(y)$ and $f_i(y), i = 1,2,..,n$ are convex and continuously differentiable and that the feasible region of (CP) is bounded and has a non-empty interior.

The logarithmic barrier function associated with (CP):

$$\phi_\beta(y,\mu) = \frac{-f_o(y}{\mu} - \sum_{i=1}^{n} \ln(-f_i(y)$$

And the logarithmic barrier function ($\phi_\beta^d$ )associated with (CD):

$$\phi_{\beta}^{d}(y,\mu)=\frac{-f_o(y}{\mu}+\frac{1}{\mu}\sum_{i=1}^{n}x_i f_i(y)+\sum_{i=1}^{n}\ln x_i+n(1-\ln\mu) \qquad 2.13$$

**Lemma 2.1.6** We have that $\phi_{\beta}(\bar{y},\mu)\geq\phi_{\beta}^{d}(x,y,\mu)$ for all primal feasible $\bar{y}$ and dual feasible (x,y). Moreover $\phi_{\beta}(\bar{y},\mu)=\phi_{\beta}^{d}(x(\mu),y(\mu),\mu)$.

**Theorem 2.1.2** The objective function $f_o(y(\mu))$ of the primal problem (CP) is monotonically increasing. and the objective function of the dual problem (CD), $f_o(y(\mu)-\sum_{i=1}^{n}x_i(\mu)f_i(y(\mu))$
is monotonically decreasing if $\mu$ decreases, where $x(\mu)$ and $y(\mu)$ are defined by (2.5).

We now introduce the logarithmic penalty function $L(y,\mu)$ for the primal problem of convex programming which has the same property with logarithmic barrier function.

Let $g:=(f_1,f_2,...,f_n)^T$ and $L(y,\mu):=f(y)-\mu\sum_{i=1}^{n}\ln f_i(y)=f(y)+\mu I(g)...$

where $I(g)=-\sum_{i=1}^{n}\ln f_i$

Let us now see the basic convergence Theorem for convex programming.

**Theorem 2.1.3 (convergence to solution of primal convex programming problem by logarithmic penalty function method).**

For any strictly positive sequence $\{g^i\}$, $\lim_{l\to\infty} I(g^l)=\infty$, if $\lim_{i\to\infty} f_i^l=0$ for some i,j

And $\{\{\mu_k\}$ is a strictly decreasing null . sequence. Then

(i) The interior pint unconstrained minimization function $U(x,\mu_k)=f(x)+\mu_k I(g(x))$.

has a finite unconstrained minimum $x(\mu_k)\in S^o,\mu_k>0$

(ii) $U(x,\mu_k)$. is a convex function in $S^o$

(iii) Every local unconstrained minimum of $U(x,\mu_k)$. in $S^{o.}$ is a global unconstrained minimum of $U(x,\mu_k)$ in $S^o$

(iv) $\lim_{n\to\infty} f(x(\mu_k)) = \lim_{n\to\infty} U[x(\mu_k),\mu_k] = v^* = \min_S f(x)$

(v) $\{f(x,\mu_k)\}$ is monotonically decreasing sequence, and

(vi) $\{I[x(\mu_k)]$ is monotonically increasing

We will prove v and vii for convergence case

**Proof:** (The proof of v and vi)

Let $f(x(\mu_k),\mu_k)=f^{k,}$ Because each $x^k$ is a global minimum of U in $S^o$,

$$f^k + \mu_k I^k \le f^{k+1} + \mu_k I^{k+1} \quad \text{.... (2.14)}$$

$$f^{k+1} + \mu_{k+1} I^{k+1} \le f^{k+1} + \mu_{k+1} I^k$$

Multiplying (2.14) by $\frac{\mu_{k+1}}{\mu_k}$, adding to the second inequality, and rearranging yields

$$(1-\frac{\mu_{k+1}}{\mu_k})f^{k+1} \le (1-\frac{\mu_{k+1}}{\mu_k})f^k$$

Since $\mu_{k+1} < \mu_k$. Part (vii) and (2.14), (vii) follows. By (vii) and (2.14),(vii) follows.

**Theorem 2.3 (Dual convergence-logarithmic Penalty function method)**

If I(g) is separable (i.e. I(g)= $\sum_{i=1}^{n} I_i(f_i)$ and each $I_i(f_i)$ is differentiable in $f_i$ where $f_i>0$,

then $[x(\mu_k),u(\mu_k)]$ is a point feasible to the constraints of the dual problem (CD) where:

$$u_i(\mu_k) = -\mu_k \frac{\partial I_i\{f_i[x(\mu_k)]\}}{\partial f_i}, i = 1,2....n \quad 2.15$$

All limit points of his sequence are finite solutions of (CD)and at least one limit point exits.

**Example** . Consider the convex programming problem;

minimize[ f(x)=-min($x_1$,$x_2$)]

s.t. $f_1(x_1,x_2) = -x_1^2 - x_2^2 + 1 \geq 0$

For our unconstrained minimization function we choose

$$L(x,\mu) = -\min(x_1, x_2) - \mu \ln(-x_1^2 - x_2^2 + 1)$$

It can be verified from the geometric depiction that at the minimum of L, $\forall \mu > 0, x_1(\mu) = x_2(\mu)$. substituting this in L yields

$$L(x,\mu) = x_1 - \mu \ln(-2x_1^2 + 1)$$

differentiating with respect to $x_1$, setting that equal to zero, and solving $x_1(\mu)$ gives

$$x_1(\mu) = \frac{-2\mu \pm \sqrt{4\mu^2 + 2}}{2}$$

Since $x_1(\mu) = x_2(\mu)$, and the minimizing point must be interior, the positive root must be taken. the dual variable given by the above Theorem is

$$u_1(\mu) = (-4\mu + \sqrt{16\mu^2 + 8})^{-1}$$

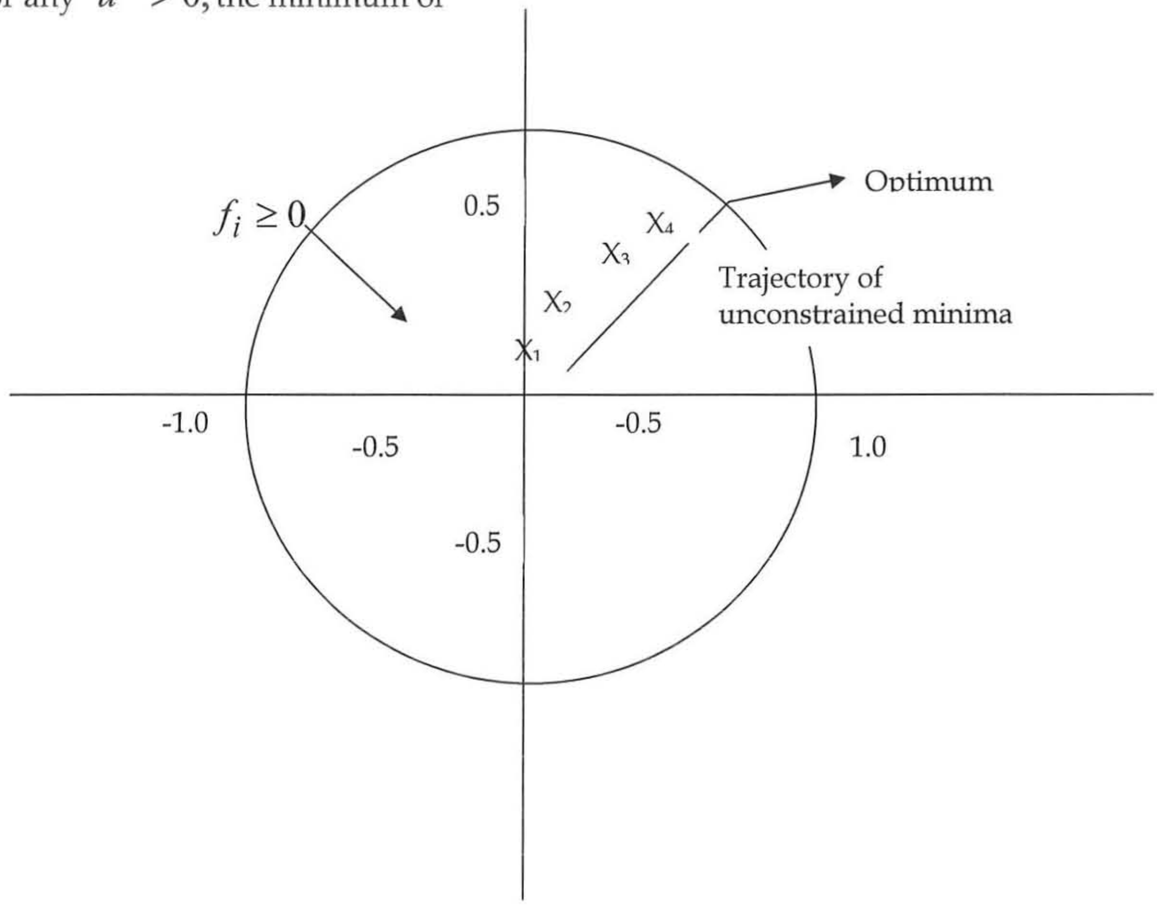
According to Theorem 2.3. $[x(\mu_k), u(\mu_k)]$ minimizes

$$L(x,u(\mu)) = -\min(x_1, x_2) - (-4\mu + \sqrt{16\mu^2 + 8})^{-1}(-x_1^2 - x_2^2 + 1)$$

and $L(x,u(\mu)) \leq v^* = -\dfrac{\sqrt{2}}{2}$, To verify this we note that

$$-\min(x_1, x_2) = \frac{-x_1 - x_2 + |x_1 - x_2|}{2} \qquad (*)$$

Then for any $u^o > 0$, the minimum of




$$\frac{-x_1 - x_2 + |x_1 - x_2|}{2} - u^o(-x_1^2 - x_2^2 + 1)$$

Fig 2.5

(*) must be minimized at a finite point since it is strictly convex and approaches $+\infty$ as $|x_1|, |x_2| \to +\infty$, and since it is symmetric in $x_1$ and $x_2$. And the minimum of (*) is achieved at $x_1(\mu), x_2(\mu)$, and $u_1(\mu)$ satisfy this equality

Furthermore $-(\frac{1}{4u^o}) - u^o[-2(\frac{1}{4u^o})^2 + 1] = \frac{-1}{8u^o} - u^o$

This is a concave function in $u^o$ when >0 nad has a maximum at $u^{o} = \sqrt{\frac{1}{8}}$

where it has the value $-\sqrt{\frac{1}{2}}$

## 2.2. THE CENTER METHOD

The center method was introduced and studied by Huard. as the logarithmic barrier function , this method also follows the central path to the optimum. The central ideas in the analysis of both logarithmic barrier and center method for linear, quadratic and convex programming are the same. First, the "Hessian norm" of the Newton search direction is always used to measure the distance to the current reference point on the central path. Second the quadratic convergence result in the vicinity of the central path enables us to many other properties, which are needed for the complexity analysis. The only difference from the logarithmic barrier method is the parameterization. Due to similarities of the two methods, many results can be followed from the previous sections for example, the quadratic convergence property for linear programming is an immediate consequence of the Lemma 2.1.1. The analysis of this method can be nicely extended to convex programming.
Recall that the convex programming problem (CP) as given in 2.1.

$$
\text{(CP)} \qquad \begin{aligned} &\max f_o(y) \\ &f_i(y) \le 0, i = 1,2,\ldots n \\ &y \in R^m \end{aligned}
$$

and its Wolfe's[152] dual problem

$$
\text{(CD)} \qquad \begin{aligned} &\min f_o(y) - \sum_{i=1}^{n} x_i f_i(y) \\ &\sum_{i=1}^{n} x_i \nabla f_i(y) = \nabla f_o(y) \\ &x_i \ge 0 \end{aligned}
$$

Let $\phi_D(y,z)$ be the distance function of variables y and z. We now make the same assumptions as in (2.1.). The distance function associated with (CP) is

$$\phi_D(y,z) = -q\ln(f_o(y)-z) - \sum_{i=1}^{n}\ln(-f_i(y) \qquad 2.16$$

Where z is a lower bound for the optimal value $z^*$, and q is a given positive integer. We have

$\phi_D(y,z)$ is strictly convex on its domain S°. Hence the distance function achieves the minimal value in its domain (for fixed z) at a unique point, which is denoted by y(z). the necessary and sufficient Karush -Kuhn -Tucker condition for this minimum are ;

$$\begin{aligned} & f_i(y) \le 0, 1 \le i \le n \\ & \sum_{i=1}^{n} x_i \nabla f_i(y) = \nabla f_o(y), x \ge 0 \\ & -f_i(y)x_i = \frac{f_o(y)-z}{q}, 1 \le i \le n \end{aligned} \qquad 2.17$$

Comparing the system of files in (2.50 it can be verified that y(z) on the central path of the problem i.e., $y(z) = y(\mu)$ for $\mu = \dfrac{f_o(y(z))-z}{q}$. Consequently the logarithmic barrier function and the distance function yield two different parameterization of the same central path.

We will use a measure $\| \ \|_H$ to measure distance between points especially from y(z) We define $\| \ \|_H$ as $\|v\|_H = \sqrt{v^T H v}$, where H is the Hessian of the distance function. i.e.

$$H(y,\mu) := \nabla^2\phi_D(y,\mu) = \sum_{i=1}^{n+q}[\frac{\nabla^2 f_i(y)}{-f_i(y)} + \frac{\nabla f_i(y)\nabla f_i(y)^T}{f_i(y)^2}]$$

To find an (iterate close to) the analytic center, which is in fact equivalent to minimizing $\phi_D(y,z)$, we use again Newton's method with (approximate) line search procedures. Note that the Newton direction is given by;

$p(y,z) = -H(y,z)^{-1} g(y,z) = -H^{-1} g$ where g(y,z) is the first derivative of $\phi_D(y,z)$

$$g(y,\mu) := \nabla \phi_D(y,z) = \sum_{i=1}^{n+q} \frac{\nabla f_i(y)}{-f_i(y)} \qquad 2.19$$

The proximity criterion for this minimizing process is a gian the Hessian Norm of the search direction. More precisely we stop if $\|p\|_H=0$ then y=y(z). if the proximity criterion is satisfied, we update the lower bound z as follows: $z = z + \theta(f_o(y) - z)$, for some $0 < \theta < 1$

We now describe the algorithm to find an $\varepsilon$ – optimal solution

**Center algorithm**

Input:

$\varepsilon$ : is the accuracy parameter

$\tau$ : is the proximity tolerance

$\theta$: is the updating criteria , $0 < \theta < 1$

$z^o < f_o(y^o)$ is the lower bound for the optimal value.

$y^o$: is a given interior feasible point such that

$$\|p(y^o, z^o)\|_{H(y^o, z^o)} \leq \tau$$

Begin

$y:=y^o$; $z=z^o$; $\Delta := 4(1+\frac{n}{q})$;

while $f_o(y)-z>\frac{\varepsilon}{\Delta}$ do;

begin(outer step)

$z:=z+\theta(f_o(y)-z$;

while$\|p\|_H>\tau$ do;

begin(inner step)

$$\hat{\alpha} := \arg\min_{\alpha>0}\{\phi_D(y+\alpha p, z) : y+\alpha p \in S^o;$$

$$y := y + \hat{\alpha}\, p$$

end (inner step)

end(outer step)

We stop the minimizing procedure if tolerance $\tau > \|p\|_H$.

### 2.2.1 Central Path

Definition 2.2.1 the analytic center of the bounded convex region $S_z$ is the point which maximizes

$$\prod_{i=1}^{n+q}(-f_i(y)), \text{ Product of slack values}$$

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

since y(z) maximizes

It easily follows that y(z0 is the analytic center of $S_z$.

Given a lower bound z for the optimal value , we try to reach the vicinity of y(z), the analytic center of the current feasible region. Then we increase the lower bound, which means that q objective constrains are shifted, and we try to reach the vicinity of the new center.

**Example 1.** Consider the problem

$$\begin{cases} \max 2y_1 + y_2) \\ y_1 \le 1 \\ y_2 \le 1 \end{cases}$$

Take q=n=2. Then

$$\phi_D(y,z) = -2\ln(2y_1 + y_2 - z) - \ln(1 - y_1) - \ln(1 - y_2)$$

The function is minimized by setting derivatives( with respect to $y_1$ and $y_2$)equal to zero

$$-\frac{4}{2y_1 + y_2 - z} + \frac{1}{1 - y_1} = 0$$

and

$$-\frac{2}{2y_1 + y_2 + -z} + \frac{1}{1 - y_2} = 0$$

This implies that $y_1(z) = \frac{5}{8} + \frac{1}{8}z$ and $y_2(z) = \frac{1}{4} + \frac{1}{4}z$

Hence the optimal value of the problem can be calculated from

$$z = f_o(y(z)) \Rightarrow z = 2y_1(z) + y_2(z) = 3$$

Consequently the optimal value of this problem is $(y_1(3),y_2(3))=(1,1)$.

Hence the path of centers is a line ending in the optimal solution (1,1).

**Example 2**. The observation made I example 2 of section 2.1. is true also in a more general setting :

If the feasible region of a linear program is a cone then the central path is a line, ending in the top of the cone. Again suppose that the constraints of the linear problem describe a cone in $R^m$. More over , let the objective function be such that the top of the cone is an optimal solution. Let the constraint matrix be $A^T$, which is $m \times n$ matrix and let $b \in R^m$ so the problem is a as follows;

$$\begin{cases} \max b^T y \\ A^T y \le c \end{cases}$$

Note that $A^T c=t$ is the top of the come. Te distance function for this problem is

$$\phi_D(y,z) = -q\ln(b^T y - z) - \sum_{i=1}^{m} \ln(c_i - a_i{}^T y)$$

Where $a_i$ is the $i^{th}$ row of $A^T$. It can easily be verified the Karush-kuhn-Tucker condition for the Dual problem (LD):

$$\begin{cases} A^T y + s = c, s \geq 0 \\ Ax = b, x \geq 0 \\ Xs = \dfrac{b^T y - z}{q} e \end{cases}$$

We can calculate y(z) in a similar way as we calculated $y(\mu)$

$$y(z) = A^{-T}(c - \frac{b^T y - z}{q} v), \text{ Where } \quad v_i = \frac{1}{(A^{-1}b)_i}$$

Equivalently, we have $(I + A^{-T}\dfrac{vb^T}{q})y = A^{-T}c + \dfrac{z}{q}A^{-T}v.$

Consequently the path of centers is a line. It can be verified that this line ends in the optimal point, namely the top of the cone ($A^Tc$=t).

### 2.2.1. <u>Center Method for Linear Programming</u>

Consider the dual linear programming problem LD making the same assumption as in (2.7) the distance function for this problem is given by:

$$\phi_D(y,z) = -q\ln(b^T y - z) - \sum_{i=1}^{m} \ln(cs_i) \qquad 2.20$$

Where q is a positive integer. The Hessian of this function:

$$\frac{q}{(b^T y - z)^2} bb^T + As^{-2}A^T \quad \text{where} \quad s^{-2} = \begin{bmatrix} s_1^{-2} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & s_n^{-2} \end{bmatrix}$$

is positive definite since we assumed that A has full rank. Hence the distance function $\phi_D(y,z)$ achieves a minimum value at a unique point, denoted as y(z). the necessary and sufficient Karush Kuhn Tucker conditions for y(z) are

$$\begin{cases} A^T y + s = c, s \geq 0 \\ Ax = b, x \geq 0 \\ Xs = \dfrac{b^T y - z}{q} e \end{cases}$$

where $x \in R^n$. The unique solution of this system is denoted by (x(z),y(z)s(z)). y(z) lies on the central trajectory of the problem (LD). More precisely we have $y(\mu) = y(z)$ for $\mu = \dfrac{b^T y - z}{q}$

The minimizing point y(z) can be considered as the analytic center of $S_z$, which is defined as

$$\{y : A^T y \leq c, -b^T y \leq -z, \ldots, -b^T y \leq -z\}$$

q-times

Note that $F_Z$ is bounded, since we assumed the optimal set is bounded.

Now the matrix $\bar{A} := (a_1, a_2, \ldots, a_{n+1} \ldots a_{n+q})$ where $a_i$=-b for $n+1 \leq i \leq n+q$.

The components of the diagonal matrix $\bar{S}$ are defined as $S_i$. for some $1 \leq i \leq n$

and

$b^T y - z$ for $n+1 \leq i \leq n+q$. The components of the vector $\bar{c}$ are defined as $c_i$ for $1 \leq i \leq n$ and -Z for $n+1 \leq i \leq n+q$. Consequently

$$\phi_D(y,z) = -\sum_{i=1}^{n+q} \ln \bar{s}_i \,.$$

We observe that y(z0 is the analytic center of $F_z = \{y : \bar{A}^T y \le \bar{c}\}$

And the equivalent Karush-Kuhn-Tucker conditions are:

$$\begin{cases} \bar{A}^T y + \bar{s} = \bar{c}, s \ge 0 \\ \bar{A}\bar{x} = 0, x \ge 0 \\ \bar{X}\bar{s} = e \end{cases} \qquad 2.22$$

The unique solution if this system is denoted by $(\bar{x}(z), y(z), \bar{s}(z))$. The following Lemma states that the dual objective $b^T y(z)$ increases, the primal objective $c^T x(z)$ and

$b^T y(z) - z$ are monotonically decreasing if z increases.

**Proposition 2.2.1** The dual objective $b^T y(z)$ is monotonically increasing the primal objective $c^T x(z)$ and $b^T y(z) - z$ are monotonically decreasing if z increases.

We can get the corresponding measure for the distance of the interior feasible point to the center of $S_z$

$$\sigma(y,z) = \left\| \bar{s}\bar{x}(y,z) - e \right\|$$

The Newton direction p(y,z) is given by

$$p(y,z) = -H^{-1}g = -(\bar{A}\,s^{-2}\bar{A}^T)^{-1}\,\bar{A}\bar{s}^{-1}e$$

Most of the propositions stated below are direct consequences of Lemmas stated in section 2.1

**Proposition 2.2.2**

For a given y and z we have

$$\sigma(y,z) = \|p(y,z)\|_H = \left\|\bar{s}^{-1}\bar{A}^{-1} p(y,z)\right\| \geq \left\|s^{-1}A^T p(y,z)\right\|$$

**Proposition 2.2.3** If $\sigma(y,z) < 1$, then y+=y+p is strictly feasible point of $S_z$. more over $\sigma(y^+, z) < \sigma(y,z)^2$

**Proposition 2.2.3 If** $\sigma := \sigma(y,z) < 1$, then $\phi_D(y,z) - \phi_D(y(z),z) \leq \dfrac{\sigma^2}{1-\sigma^2}$

**Proof:** As $\phi_D(y,z)$ is a logarithmic barrier function for the following problem. $\max\{0 : y \in S_z\}$, the proposition 2.2.4 is an immediate consequence of Lemma 2.1.3

### 2.2.2. Center method for convex Programming

In this section we will try to see the monotonocity of the primal and dual objective values along the path defined by the distance function. For the analysis, we again assume that the functions -$f_o$(y), $0 \leq i \leq n$, are convex with first and second order derivatives in S. we will prove that the objective value $f_o$(y(z)) increases and the dual objective $f_o(y(z)) - \sum_{i=1}^{n+q} x_i(z) f_i(y(z))$ and $f_o$(y(z))-z decreases, if are z decreases

**Theorem 2.2.1.**

The primal objective fo(y(z)) is monotonically increasing , and the dual objective $f_o(y(z)) - \sum_{i=1}^{n+q} x_i(z) f_i(y(z))$ and $f_o$(y(z))-z are monotonically decreasing if z increases.

**Proof**: we first prove the last part of the Theorem .Suppose $\bar{z} > z$ such that $f_o(y(z) > \bar{z}$. Consider the function:

$$\phi_D^d(\bar{x},y,z) = \sum_{i=1}^{n+q} \bar{x_i} f_i(y) + \sum_{i=1}^{n+q} \ln \bar{x_i} + (n+q)$$

Note that while $\phi_D(y,z)$ can be interpreted (up to a constant) as logarithmic barrier function for the problem

$\max\{0 : f_i(y) \leq 0, 1 \leq i \leq n+q\}$,

We have that $\phi_D^d(\bar{x},y,z)$ can be interpreted(up to constant)as the logarithmic barrier function for its dual problem:

$$\min\{-\sum_{i=1}^{n+q} \bar{x_i} f_i(y) : \sum_{i=1}^{n+q} \bar{x_i} \nabla f_i(y) = 0, \bar{x} \geq 0\}$$

It is easy to verify that a maximizing point for $\phi_D^d(\bar{x},y,z)$ has to satisfy(2.17), the last of coordinates of $\bar{x}$ are equal to $\bar{x}_{n+1}$. The first n components of $\bar{x}$ are denoted as the vector $\bar{x}$.

Now since $(\bar{x}(\bar{z}), y(\bar{z}))$ maximizes $\phi_D^d(\bar{x},y,\bar{z})$.. We have

$\phi_D^d(\bar{x}(z), y(z), z) \geq \phi_D^d(\bar{x}(\bar{z}), y(\bar{z}), z)$ and

$\phi_D^d(\bar{x}(\bar{z}), y(\bar{z}), \bar{z}) \geq \phi_D^d(\bar{x}(\bar{z}), y(z), \bar{z})$ or equivalently

$$\sum_{i=1}^{n+q} \bar{x_i}(z) f_i(y(z)) + q[z - f_o(y(z))]\bar{x}_{n+1}(z) + \sum_{i=1}^{n+q} \ln \bar{x_i}(z) + (n+q)$$

$$\geq \sum_{i=1}^{n+q} \bar{x_i}(\bar{z}) f_i(y(\bar{z})) + q[z - f_o(y(\bar{z}))]\bar{x}_{n+1}(\bar{z}) + \sum_{i=1}^{n+q} \ln \bar{x_i}(\bar{z}) + (n+q)$$

$$\geq \sum_{i=1}^{n+q} \bar{x}_i(\bar{z}) f_i(y(\bar{z})) + q[\bar{z} - f_o(y(\bar{z}))]\bar{x}_{n+1}(\bar{z}) + \sum_{i=1}^{n+q} \ln \bar{x}_i(\bar{z}) + (n+q)$$

$$\geq \sum_{i=1}^{n+q} \bar{x}_i(z) f_i(y(z)) + q[\bar{z} - f_o(y(z))]\bar{x}_{n+1}(z) + \sum_{i=1}^{n+q} \ln \bar{x}_i(z) + (n+q)$$

Adding the two inequalities gives:

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

This meanse that

$\bar{x}_{n+1}(\bar{z})] \geq \bar{x}_{n+1}(z)$

Now from (2.17) we have that

$$\bar{x}_{n+1}(z) = \frac{1}{-f_{n+1}(y(z))} = \frac{1}{f_o(y(z)) - z}$$

As the similar expression for $\bar{x}_{n+1}(\bar{z})$. Substituting this in to (2.14) gives $f_o(y(z))$-z Monotonically decreases if z decreases.

Finally,note that $y(z)=y(\mu)$, for $\mu = \dfrac{f_o(y(z)) - z}{q}$, and $y(\bar{z}) = y(\bar{\mu})$ for $\bar{\mu} = \dfrac{f_o(y(\bar{z})) - \bar{z}}{q}$, since $\bar{\mu} \leq \mu$, it follows from Theorem 2.1 that

$f_o(y(z)) - \sum_{i=1}^{n} x_i(z) f_i(y(z))$ monotonically decreases and $f_o(y(z))$ monotonically increases if z increases.

There is also a direct proof for the monotonocity of $f_o(y(z))$ since $y(z)$ minimizes $\phi_D(y,z)$

and $y(\bar{z})$ minimizes $\phi_D(y,\bar{z})$ we obviously have

$$\phi_D(y(z),z) \leq \phi_D(y(\bar{z}),z)$$

$$\phi_D(y(\bar{z}),\bar{z}) \leq \phi_D(y(z),\bar{z})$$ or equivalently

$$-q\ln(f_o(y(z)-z)-\sum_{i=1}^{n}\ln(-f_i(y(z)) \leq -q\ln(f_o(y(z)-\bar{z})-\sum_{i=1}^{n}\ln(-f_i(y(z))$$

Adding the two inequalities gives

$$-q\ln(f_o(y(z)-z)-q\ln(f_o(y(\bar{z})-\bar{z}) \leq -q\ln(f_o(y(\bar{z})-z)-q\ln(f_o(y(z))-\bar{z}),$$

or equivalently, by taking the exponential we reduce to

$$(z-\bar{z})(f_o(y(\bar{z}))-f_o(y(z)) \leq 0 \qquad //$$

## 3.0 DISCUSSION ON NON PATH FOLLOWING METHODS

Path following methods based on the classical logarithmic barrier function method and the center method are treated in the previous sections. Most interor pont methods follow or are related to the central path. It is often observed that shortly before convergence interior point methods drastically accelerate. A distinctive nature of interior point methods is that a stronger quadratic convergence results holds whenever the reference point $\bar{x}$ is an interior feasible solution. In that case, It turns out that the proximity measure is in effect computable. More over, a quadratic convergence neighborhood $\bar{x}$ is defined a simple inequality on the proximity measure. One obtains in this way a local convergence result. This result can be further be made global if one can provide a bound on the jumber of iterations required to reach the nieghbourhood . This property makes it to move in a controlled way the trarget $\bar{x}$ to approach to an optimal point $\bar{x}$. This is essentially the philosophy of the so called path following methods. The other alternative is the non path following methods like the potential
Reduction method, Karmarar's method, Affine scaling method, scalen steepest method, etc.

The affine scaling metod was probably the first interior point method defined for linear programming problems. The most prominenet of all interior point metods is Karmarkar's algorithm in 1984. There are two reasons for this : the thoretical reason for this is hat Karmarkar was able to establish a polynomial bound on the number of iterations as afunction of the lengeth of the input data. The important additional feature of Karmarkar's method was that this algorithm in contrast to the ellipsoid method-performs very well on large scale practical problems and this was the first real competitor of the simplex method. We will only see the main ideas deriving

Karmarkar's algorithm. In practical terms we will not go into the details of a complexity analysis for the algorithm.

Basically there are four classes,path following mathods ,affine scaling methods , projective potential reduction method and affine potentail reduction methods. Both affine scaling method and potential reduction methods ust the central path implicitly. The basic idea behind tis methods is the same as for path following methdods: define a reference point on the central path ,try to come to its vicinity,if current itertate is too close then update the parameter(lower or upper bound,barrier parameter,dulaity gap)which meanse that the reference point is shifted. The diference is that the progress of the potentail reduction methods

I measured by a potential function. It appeared also that the search directions used in interior point methods are all linear combinations of two characterstic vectors: the affine scaling and the centering direction.

References:


1. D. den hertog ,1994 Interior Point approach to linear,quadratic,and Convex programming
2. Semu Mitiku, On minimal pairs of compact convex sets and of convex functions, 2002
3. ANTHONY V. FIACCO,1968 Non linear programming sequential Unconstrained Minimization Techniques
4. Clovisc Gonzaga, papers on path following methods for Linear programming
5. R.Deumlich, Optimization theory I, AAU.Department of mathematics
6. J.E. Beasley,1996 Advances in linear and integer programming
7. E.Polak,1971, Computational methods in optimization
8. Springer-Verlag ,Linear optimization and approximation 1983.