

# UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE CIENCIAS

Escuela Profesional de Matemática



TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE  
LICENCIATURA; ESPECIALIDAD: MATEMÁTICA

TITULADA:

GENERALIZACION DE ALGORITMOS DE SEGUIMIENTO  
DE TRAYECTORIA PARA LA PROGRAMACIÓN LINEAL

PRESENTADA POR:

JUAN JAVIER VILLALOBOS SOLANO

Lima - Perú  
2001

# RESUMEN

En este trabajo de tesis, desarrollamos un algoritmo de seguimiento de trayectoria con una función de barrera generalizada, las cuales son usadas para caracterizar una familia de trayectorias interiores que conducen a la solución óptima del problema de la programación lineal. Manteniéndonos en una vecindad de la trayectoria interior y usando direcciones de movimiento adecuadas, proponemos un algoritmo general de seguimiento de trayectoria para una función de barrera generalizada dada. Estudiamos además condiciones suficientes para la convergencia polinomial de estos algoritmos.

# INTRODUCCION

Desde que Karmarkar, publicó su algoritmo proyectivo para resolver problemas de programación lineal en tiempo polinomial, muchas variantes de los llamados algoritmos de punto interior han sido propuestos. En términos generales estos algoritmos pueden ser clasificados en tres clases: Los algoritmos de escalado afin, los algoritmos de reducción de potencial y los algoritmos de seguimiento de trayectoria. En este trabajo de tesis estamos interesados en los algoritmos de seguimiento de trayectoria, estos algoritmos han sido construidos usando la función de barrera logarítmica y para ellos, se han probado resultados de convergencia en tiempo polinomial. El objetivo de esta tesis es desarrollar un algoritmo de seguimiento de trayectoria con una función de barrera generalizada y a partir de allí, estudiar cuales son las condiciones suficientes que una función de barrera debe satisfacer para obtener un algoritmo en tiempo polinomial y desde esta perspectiva general estudiar porque otras funciones de barrera (función de barrera entrópica, función de barrera inversa) no generan algoritmos polinomiales

En el capítulo 1 de esta tesis, hacemos una revisión breve de algunos resultados fundamentales de la programación lineal. En el capítulo 2, estudiamos el uso de la función de barrera logarítmica aplicado al problema de la programación lineal, aquí estudiamos en detalle dicha aproximación. En el capítulo 3, desarrollamos los algoritmos de seguimiento de trayectoria con barrera logarítmica poniendo énfasis en los llamados algoritmos de paso corto, ya que ellos nos darán una base para su generalización. La parte central de esta tesis es el capítulo 4, presentamos los algoritmos de seguimiento de trayectoria con una función de barrera generalizada, y derivamos condiciones suficientes para la convergencia polinomial, en este capítulo presentamos también algunas conclusiones que se desprenden de esta aproximación, los cuales nos permiten ver porque otras funciones de barrera fallan en generar algoritmos con cotas de convergencia polinomial.

# INDICE

## CAPITULO 1

### PROGRAMACIÓN LINEAL Y MÉTODOS DE PUNTO INTERIOR

1.1	PRELIMINARES MATEMATICOS.....	1
1.1.1	Notaciones.....	1
1.1.2	Espacio nulo y espacio imagen.....	3
1.1.3	Conjuntos convexos.....	4
1.2	FORMA ESTANDAR DEL PROBLEMA DE PROGRAMACION LINEAL.....	6
1.3	DUALIDAD, CONDICIONES DE OPTIMALIDAD Y CONJUNTOS SOLUCION.....	8
1.4	LA PARTICION $B \cup N$ Y LA COMPLEMENTARIEDAD ESTRICTA.....	12
1.5	RANGO DE LA MATRIZ A.....	15
1.6	DIRECCIONES ADMISIBLES Y TEST DE LA RAZÓN.....	16
1.7	TAMAÑO DE UN PROBLEMA DE PROGRAMACIÓN LINEAL.....	18
1.8	SOLUCIÓN DE UN PROBLEMA DE PROGRAMACIÓN LINEAL.....	23
1.8.1	Solución por el método simplex.....	23
1.8.2	Solución por un método interior.....	24

## CAPITULO 2

### EL METODO DE BARRERA LOGARITMICA PARA LA PROGRAMACIÓN LINEAL

2.1	INTRODUCCIÓN.....	28
2.2	LA FUNCIÓN DE BARRERA LOGARÍTMICA.....	30
2.2.1	Función de barrera dual.....	33
2.3	RELAJACIÓN DE LA HIPÓTESIS DE PUNTO INTERIOR.....	46

## **CAPITULO 3**

### **ALGORITMO DE SEGUIMIENTO DE TRAYECTORIA**

3.1	CONSTRUCCION DE UN ALGORITMO DE SEGUIMIENTO DE TRAYECTORIA .....	48
3.1.1	Medida de Proximidad .....	57
3.1.2	Selección de los parámetros de barrera.....	59
3.1.3	Algoritmo de seguimiento de trayectoria para el problema primal .....	60
3.2	ANALISIS DEL ALGORITMO .....	63

## **CAPITULO 4**

### **GENERALIZACIÓN DE ALGORITMOS DE SEGUIMIENTO DE TRAYECTORIA PARA EL PROBLEMA DE LA PROGRAMACIÓN LINEAL**

4.1	FUNCIONES DE BARRERA GENERALIZADA PARA EL PROBLEMA DE LA PROGRAMACIÓN LINEAL .....	70
4.2	CALCULO DE LAS DIRECCIONES DE BÚSQUEDA .....	77
4.3	CRITERIO DE CERCANÍA .....	79
4.4	ALGORITMO GENÉRICO DE SEGMENTO DE TRAYECTORIA .....	83
4.5	CONDICIONES SUFICIENTES PARA CONVERGENCIA Y POLINOMIALIDAD .....	84
4.6	EJEMPLOS DE APLICACIÓN .....	91
4.7	CONCLUSIONES .....	94
	BIBLIOGRAFÍA.....	97

# CAPITULO 1

## PROGRAMACIÓN LINEAL Y MÉTODOS DE PUNTO INTERIOR

En este capítulo presentamos el problema de la programación lineal en su forma estándar y discutimos algunos aspectos de la teoría de dualidad que han tenido una influencia directa en los métodos de punto interior, presentamos también el teorema de Goldman-Tucker, el cual establece la existencia de soluciones estrictamente complementarias. Asimismo tratamos el problema del cálculo de las direcciones admisibles, caracterizando a éstas como vectores en el núcleo de la matriz de restricciones. Además presentamos algunos resultados relacionados con la complejidad algorítmica del problema de la programación lineal.

Por otro lado para los resultados sobre la complejidad del problema de la programación lineal, hemos usado el libro de Padberg [14], y para lo concerniente a resultados específicos de la teoría de dualidad y tópicos relacionados el libro de Wright [19]

### 1.1 PRELIMINARES MATEMATICOS

#### 1.1.1 Notaciones

Trabajaremos en el espacio euclidiano  $\mathbb{R}^n$ . Los puntos (vectores) de  $\mathbb{R}^n$  serán representados por vectores columnas  $x = (x_1, \dots, x_n)^T$ . La transpuesta de una matriz  $A$  será denotada por  $A^T$  y el producto escalar de los vectores  $x, y \in \mathbb{R}^n$  por  $x^T y$ . Los subíndices normalmente serán usados para denotar las componentes  $x_1, \dots, x_n$  de un vector y los superíndices denotarán vectores en  $\mathbb{R}^n$ . A un vector  $x \in \mathbb{R}^n$  se le puede asociar una infinidad de normas distintas, las mas comúnmente usadas son:

La norma euclidiana:

$$\|x\| = \sqrt{x^T x} = \left( \sum_{i=1}^n x_i^2 \right)^{1/2},$$

La norma infinita:

$$\|x\|_{\infty} = \max_{i=1, \dots, n} |x_i|,$$

La norma 1:

$$\|x\|_1 = \sum_{i=1}^n |x_i|$$

En general, la norma  $p$  es

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

La siguiente relación entre las normas es importante

$$\|x\|_{\infty} \leq \|x\| \leq \|x\|_1,$$
$$\frac{\|x\|_1}{\sqrt{n}} \leq \|x\| \leq \sqrt{n} \|x\|_{\infty}.$$

Para un vector  $x \in \mathbb{R}^n$ , con  $X$  denotaremos a la matriz diagonal en  $\mathbb{R}^{n \times n}$  cuyos elementos diagonales son las componentes de  $x$ , i.e,

$$X = \text{diag}(x_1, \dots, x_n)$$

$\{x^k\}_0^{\infty}$  denotará a una sucesión ordenada  $x^0, x^1, x^2, \dots, x^k, \dots$ , una sucesión  $\{x^k\}_0^{\infty}$  es convergente a  $\bar{x}$ , lo cual será denotado por  $x^k \rightarrow \bar{x}$ , si

$$\|x^k - \bar{x}\| \rightarrow 0$$

Un punto  $x$  es un punto límite de  $\{x^k\}_0^{\infty}$  si hay una subsucesión de  $\{x^k\}_0^{\infty}$  convergente a  $x$ . Si  $g(x) \geq 0$  es una función real de variable real, la notación  $g(x) = \mathcal{O}(x)$  significa que  $g(x) \leq \bar{c} x$

para alguna constante  $\bar{c}$ ; la notación  $g(x) = \Omega(x)$  significa  $g(x) \geq cx$  para alguna constante  $c$ ; la notación  $g(x) = \theta(x)$  significa que  $c x \leq g(x) \leq \bar{c} x$ .

Un conjunto de funciones reales  $f_1, f_2, \dots, f_m$  definidas en  $\mathbb{R}^n$  puede ser escrita como una función vectorial  $f = (f_1, f_2, \dots, f_m)^T \in \mathbb{R}^m$ . Si  $f$  tiene derivadas parciales continuas de orden  $p$ , escribiremos  $f \in C^p$ . El gradiente de una función  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$  es el vector

$$\nabla f(x) = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)^T.$$

Si  $f \in C^2$ , definimos el Hessiano de  $f$  como la matriz simétrica de orden  $n$

$$\nabla^2 f(x) = [a_{ij}]_{n \times n} \text{ donde } a_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

para  $i = 1, \dots, n$ ,  $j = 1, \dots, n$ .

Si  $f = (f_1, f_2, \dots, f_m)^T \in \mathbb{R}^m$ , la matriz Jacobiana de  $f$  es

$$\nabla f(x) = \begin{pmatrix} \nabla f_1(x) \\ \nabla f_2(x) \\ \vdots \\ \nabla f_m(x) \end{pmatrix}.$$

### 1.1.2 Espacio nulo y espacio imagen

En lo que sigue, consideremos  $A \in \mathbb{R}^{m \times n}$  con  $m \leq n$ , de rango máximo definimos:

**Definición 1.1:** El espacio nulo de  $A$  es denotado por  $N(A)$  y es definido mediante

$$N(A) = \{x \in \mathbb{R}^n : Ax = 0\}.$$

El espacio nulo de una matriz es el conjunto de vectores ortogonales a las filas de la matriz  $A$ .

**Definición 1.2:** El espacio imagen de  $A^T$  es denotado por  $\text{Im}(A^T)$  y es definido mediante

$$\text{Im}(A^T) = \{y \in \mathbb{R}^m : y = A^T x, \text{ por algún } x \in \mathbb{R}^n\}$$

El siguiente Lema tomado de Nash y Sofer [13] da una relación importante entre  $N(A)$  y  $\text{Im}(A^T)$ .



**Lema 1.1:**  $N(A) \perp \text{Im}(A^T)$

**Prueba:** Sea  $y \in \text{Im}(A^T)$ , entonces  $y = A^T x$ , para algún  $x \in \mathbb{R}^n$ , y de este modo para todo vector  $\hat{x} \in N(A)$  tenemos

$$y^T \hat{x} = x^T A \hat{x} = 0 \quad \blacksquare$$

luego si  $x \in \mathbb{R}^n$ ,  $x$  puede escribirse como

$$x = \hat{p} + \hat{q} \text{ donde } \hat{p} \in N(A) \quad \text{y} \quad \hat{q} \in \text{Im}(A^T)$$

Multiplicando esta última expresión por  $A$  obtenemos:

$$\begin{aligned} Ax &= A(\hat{p} + \hat{q}) \\ &= A\hat{q} \\ &= A A^T x' \text{ para algún } x' \in \mathbb{R}^n \end{aligned}$$

de donde obtenemos  $x' = (AA^T)^{-1}Ax$ . Sustituyendo en  $x = \hat{p} + \hat{q}$  y despejando  $\hat{p}$  obtenemos la componente de  $x$  en el espacio nulo.

$$\begin{aligned} \hat{p} &= x - A^T(AA^T)^{-1}Ax \\ &= (I - A^T(AA^T)^{-1}A)x \end{aligned}$$

la matriz  $n \times n$

$$P = I - A^T(AA^T)^{-1}A$$

es llamada la matriz de proyección ortogonal en  $N(A)$ .

### 1.1.3 Conjunto Convexos

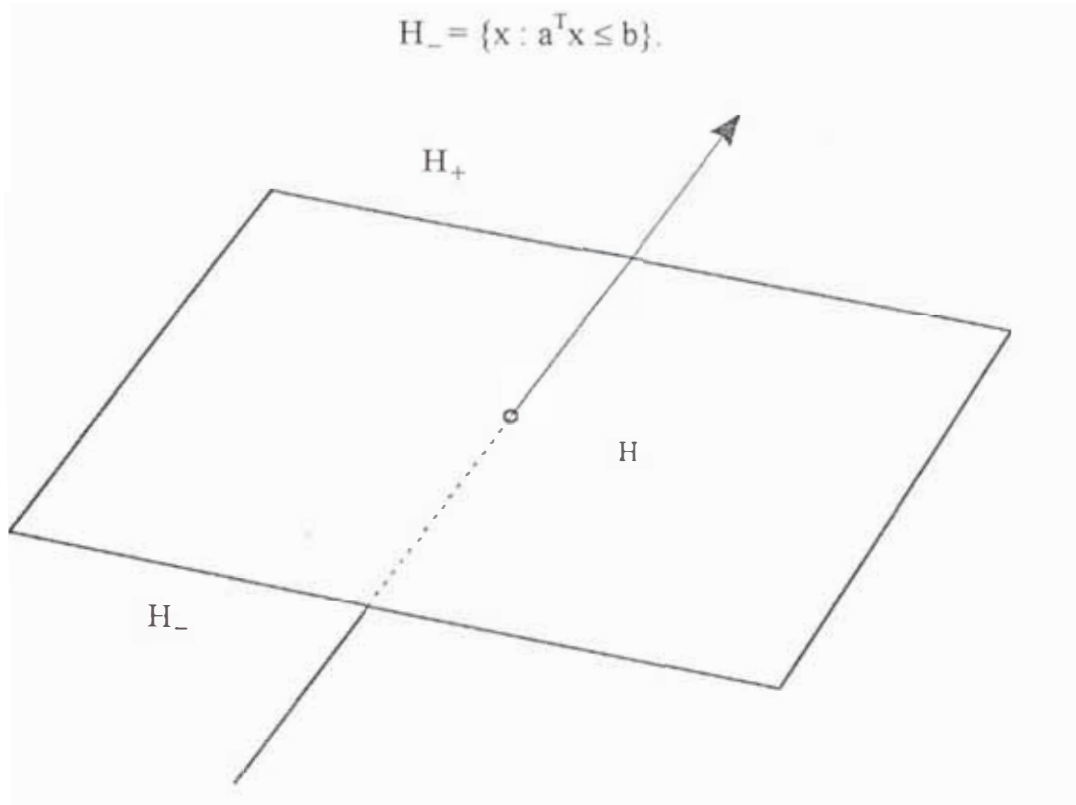
Un conjunto  $C \subset \mathbb{R}^n$  es convexo si para cada par de puntos  $x^1, x^2 \in C$  y para todo  $\alpha \in [0, 1]$ , el punto  $\alpha x^1 + (1 - \alpha)x^2 \in C$ .

El más importante tipo de conjunto convexo es un *hiperplano* y son fundamentales en teoría de optimización. Sea  $a \in \mathbb{R}^n$   $a \neq 0$ , y sea  $b \in \mathbb{R}$ . El conjunto

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

es un hiperplano en  $\mathbb{R}^n$  (figura 1.1) relacionado a un hiperplano, tenemos *los semiespacios positivos y negativos* definido por:

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



**Figura 1.1: Un hiperplano y semiespacios**

Un conjunto que puede ser expresado como la intersección de un número finito de semiespacios cerrados es llamado un *poliedro convexo*.

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

Sea  $P \subset \mathbb{R}^n$  un poliedro y  $F \subseteq \mathbb{R}^n$  un conjunto cualquiera.

- i)  $F$  es una cara de  $P$  si y solo si existe un vector  $(c, f_0) \in \mathbb{R}^{n+1}$  con  $c \neq 0$  tal que

$$F = \{x \in P : c^T x = f_0\} \text{ y } c^T x < f_0 \quad \forall x \in P \setminus F$$

- ii)  $F$  es una faceta de  $P$  si y solo si  $F$  es una cara de  $P$  y  $\dim F = \dim P - 1$

- iii) Una cara de dimensión 0 es un punto extremo de  $P$ . Una cara de dimensión 1 es una arista de  $P$ .

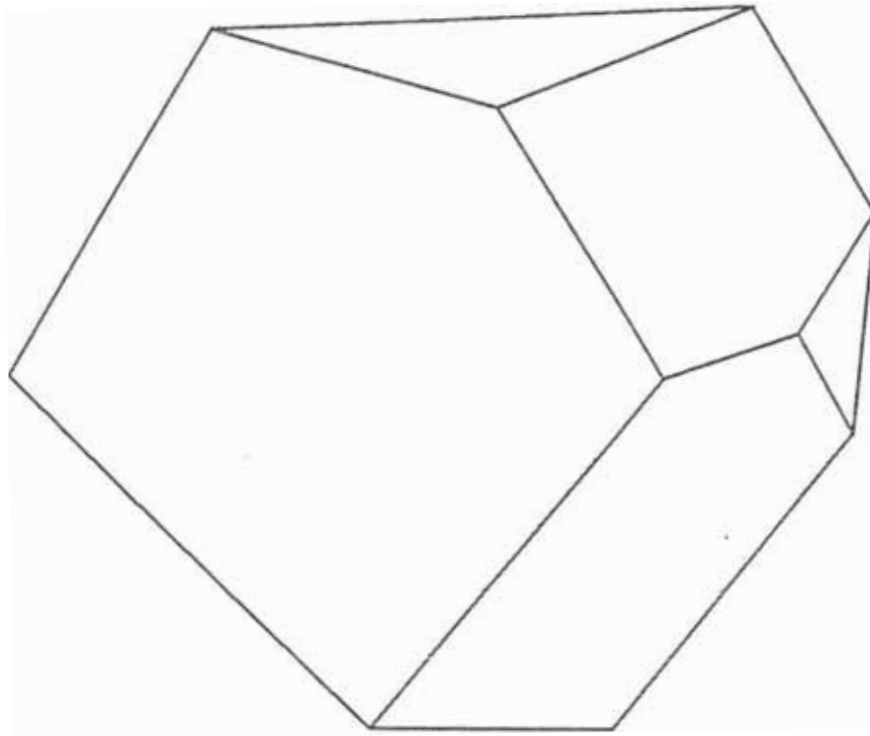


Figura 1.2: Facetas de un poliedro

## 1.2 FORMA ESTANDAR DEL PROBLEMA DE PROGRAMACION LINEAL

En esta sección presentamos la forma de problema de programación lineal que será empleada en el presente trabajo de tesis.

La forma estándar del problema de programación lineal es

$$\begin{aligned} \min \quad & c^T x \\ \text{sujeto a} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{1.1}$$

donde  $c, x \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^m$ ,  $A \in \mathbb{R}^{m \times n}$ .

La función lineal  $c^T x$  es llamada la *función objetivo* y  $x$  es llamado el vector de *variables de decisión*.

El dual de (1.1) es

$$\begin{aligned} & \max \quad b^T y \\ & \text{sujeto a} \quad A^T y + s = c \\ & \quad \quad \quad s \geq 0 \end{aligned} \tag{1.2}$$

donde  $y \in \mathbb{R}^m$ ,  $s \in \mathbb{R}^n$ . Las componentes de  $s$  son llamadas *variables de holgura*, note que  $y \in \mathbb{R}^m$  es un vector 'libre'.

El conjunto

$$\mathcal{P} = \{x : Ax = b, \quad x \geq 0\},$$

es llamado el *conjunto factible* o *región factible de (1.1)*. Un punto  $x \in \mathcal{P}$  es llamado un *punto factible*, y un punto factible  $x^*$  es llamado una *solución óptima* si  $c^T x^* \leq c^T x, \forall x \in \mathcal{P}$ , si existe una sucesión  $\{x^k\}$  tal que  $x^k$  es factible y  $c^T x^k \rightarrow -\infty$  entonces (1.1) es *no acotado*.

De igual forma el conjunto

$$\mathcal{D} = \{(y, s) : A^T y + s = c, \quad s \geq 0\},$$

es llamado el conjunto factible del problema (1.2).

Todo problema de programación lineal puede ser llevado a la forma estándar (1.1) mediante transformaciones simples.

Por ejemplo, una restricción del tipo  $g^T x \leq d$  puede ser transformada introduciendo una variable de holgura  $z$  para obtener

$$g^T x + z = d, \quad z \geq 0,$$

la cual está en la forma estándar. Las restricciones de la forma  $x_i > l_i$  (donde  $x_i$  es la  $i$ -ésima componente de  $x$  y  $l_i$  es un número finito) puede ser transformado a una restricción del tipo  $> 0$ , mediante el cambio

$$x_i \leftarrow x_i - l_i$$

Las componentes  $x_i$  que no tienen restricciones de signo pueden ser transformadas en la diferencia de dos componentes no negativas como

$$x_i = x_i^+ - x_i^-, \quad x_i^+ \geq 0, \quad x_i^- \geq 0$$

Los software para programación lineal realizan estas transformaciones automáticamente y convierten problemas de programación lineal generales en la forma estándar (1.1).

Debido a la equivalencia entre las diferentes formulaciones, es que no perdemos generalidad si consideramos la forma (1.1) y es lo que haremos en este trabajo de tesis.

### 1.3 DUALIDAD, CONDICIONES DE OPTIMALIDAD Y CONJUNTOS SOLUCION

En esta sección, discutimos algunos aspectos importantes de la teoría de dualidad que han tenido una participación directa en el diseño y análisis de métodos de punto interior. No hacemos un tratamiento completo de este importante tópico, debido a ello presentamos algunos resultados sin su respectiva demostración. Por otro lado las condiciones de optimalidad (es decir, condiciones algebraicas que deben satisfacer las soluciones de los problemas de programación lineal) son presentadas. A continuación enunciamos el teorema de dualidad débil según Wright [19], el cual es fundamental dentro del contexto de la teoría de la dualidad.

#### Teorema 1.2: (Teorema de dualidad débil)

Sean los conjuntos factibles no vacíos  $\mathcal{P}$  y  $\mathcal{D}$ . Entonces,  $c^T x \geq b^T y$  donde  $x \in \mathcal{P}$ ,  $(y, s) \in \mathcal{D}$ .

**Prueba:** La prueba es directa, se tiene

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

Por tanto,

$$\begin{aligned} c^T x - b^T y &= c^T x - (Ax)^T y \\ &= (c - A^T y)^T x \\ &= s^T x \geq 0 \end{aligned}$$

la última ecuación se obtiene de  $x^T s = \sum_{j=1}^n x_j s_j$  y  $x, s \geq 0$  ■

La diferencia  $c^T x - b^T y$  es llamada la *brecha dual*. Para ella tenemos importantes resultados.

### Teorema 1.3: (Teorema de dualidad fuerte)

Sean los conjuntos factibles no vacíos  $\mathcal{P}$  y  $\mathcal{D}$ . Entonces,  $x^*$  es óptimo para el problema (1.1) si y solo si las siguientes condiciones se verifican.

- i)  $x^* \in \mathcal{P}$ ,
- ii) Existe  $(y^*, s^*) \in \mathcal{D}$ ,
- iii)  $c^T x^* = b^T y^*$  ■

El teorema de dualidad fuerte es un teorema muy importante en programación lineal porque da un método para verificar optimalidad, en otras palabras, lo que el Teorema 1.3 nos dice es que un punto  $(x^*, y^*, s^*)$  es un par primal-dual óptimo si se verifican las siguientes condiciones.

- *Factibilidad primal* (condición i del teorema 1.3)

$$Ax^* = b, \quad x^* > 0.$$

- *Factibilidad dual* (condición ii del teorema 1.3)

$$A^T y^* + s^* = c, \quad s^* \geq 0$$

- *Optimalidad*

$$c^T x^* - b^T y^* = 0.$$

Una formulación alternativa para las condiciones de optimalidad del teorema 1.3, es decir que un primal-dual  $(x, y, s)$  satisfaciendo

$$\begin{aligned} Ax &= b, & x &> 0, \\ A^T y + s &= c, & s &> 0, \\ x_j s_j &= 0, & j &= 1, \dots, n, \end{aligned} \tag{1.3}$$

es una solución óptima. Estas condiciones contienen tres componentes, la última de ellas es llamada *condición de complementariedad*.

Si asumimos ahora que  $(x, y, s)$  satisfacen las ecuaciones (1.3), entonces la brecha dual es dada por

$$\begin{aligned} c^T x - b^T y &= x^T s \\ &= \sum_{j=1}^n x_j s_j \\ &= 0. \end{aligned}$$

Verificando que  $(x, y, s)$  es una solución óptima.

Usaremos la notación  $\mathcal{P}^*$  y  $\mathcal{D}^*$  para denotar los conjuntos solución de los problemas primal y dual respectivamente.

$$\mathcal{P}^* = \{x^* / x^* \text{ es solución de (1.1)}\}$$

$$\mathcal{D}^* = \{(y^*, s^*) / (y^*, s^*) \text{ es solución de (1.2)}\}$$

El siguiente lema según Wright [19] proporciona una propiedad importante de los conjuntos  $\mathcal{P}^*$  y  $\mathcal{D}^*$

**Lema 1.4:** Los conjuntos solución  $\mathcal{P}^*$  y  $\mathcal{D}^*$  son cerrados.

**Prueba:** Este lema es fácilmente verificado en el caso de  $\mathcal{P}^*$  si usamos la siguiente caracterización equivalente

$$\mathcal{P}^* = \{x^* / Ax^* = b, x^* > 0, c^T x^* = z^*\} \quad (1.4)$$

donde  $z^*$  es el valor óptimo del problema (1.1).

Por otro lado (1.4) muestra que  $\mathcal{P}^*$  es una intersección de subconjuntos cerrados de  $\mathbb{R}^n$  y de este modo  $\mathcal{P}^*$  también un conjunto cerrado. Si análogamente se demuestra para el conjunto  $\mathcal{D}^*$ . ■

Según Wright [19], se define:

**Definición 1.3:** Un vector  $x \in \mathbb{R}^n$  es llamado estrictamente factible para el problema (1.1) si

$$Ax = b, \quad x > 0,$$

Esto es, la condición de no negatividad  $x \geq 0$  es reemplazada por una inecuación estricta.

Similarmente, el vector  $(y, s)$  es estrictamente factible para el problema (1.2) si

$$A^T y + s = c, \quad s > 0$$

El siguiente teorema tomado de Wright [19] da una condición para la existencia y acotamiento de los conjuntos solución  $\mathcal{P}^*$  y  $\mathcal{D}^*$ .

**Teorema 1.5:** Supongamos que los problemas primal y dual son factibles. Entonces, si el problema dual tiene un punto estrictamente factible, el conjunto solución  $\mathcal{P}^*$  es no vacío y acotado. Similarmente, si el problema primal tiene un punto estrictamente factible, el conjunto

$$\{s^* / (y^*, s^*) \in \mathcal{D}^*, \text{ para algún } y^* \in \mathbb{R}^n\}$$

es no vacío y acotado.

**Prueba:** Probamos la primera afirmación, la segunda es similar y no la haremos.

Denotemos el punto dual estrictamente factible por  $(\bar{y}, \bar{s})$ , y sea  $\hat{x}$  cualquier punto factible primal (no necesariamente estricto). Entonces tenemos

$$0 \leq \bar{s}^T \hat{x} = c^T \hat{x} - b^T \bar{y} \tag{1.4}$$

Ahora consideremos el conjunto de puntos  $\mathcal{J}$  definido por

$$\mathcal{J} = \{x : Ax = b, \quad x \geq 0, \quad c^T x \leq c^T \hat{x}\}$$

El conjunto  $\mathcal{J}$  es no vacío (ya que  $\hat{x} \in \mathcal{J}$ ) y es obviamente cerrado.

Para cada  $x \in \mathcal{J}$ , tenemos

$$\sum_{i=1}^n \bar{s}_i x_i = \bar{s}^T x = c^T x - b^T \bar{y} \leq c^T \hat{x} - b^T \bar{y} = \bar{s}^T \hat{x}$$

Como todas los términos en la suma del lado izquierdo son no negativos, tenemos

$$x_i \leq \frac{1}{\bar{s}_i} \bar{s}^T \hat{x} \Rightarrow \|x\|_\infty \leq \left( \max_{i=1,2,\dots,n} \frac{1}{\bar{s}_i} \right) \bar{s}^T \hat{x}$$



Como  $x$  es un elemento arbitrario de  $\mathcal{F}$ , concluimos de esta inecuación que  $\mathcal{F}$  está acotado, así como también es no vacío y cerrado. Luego la función  $c^T x$  alcanza su mínimo en  $\mathcal{F}$ . Esto es, existe un punto  $x^*$  tal que

$$x^* \in \mathcal{F}, \quad c^T x^* \leq c^T x \quad \forall x \in \mathcal{F}$$

Obviamente, el conjunto de puntos  $x^*$  con esta propiedad coincide con  $\mathcal{P}^*$ . Luego  $\mathcal{P}^*$  es no vacío y acotado, desde que es un subconjunto del conjunto acotado  $\mathcal{F}$ . ■

Como un comentario, diremos que el recíproco del teorema 1.5 es también verdadero: Si el conjunto solución  $\mathcal{P}^*$  es no vacío y acotado, el problema dual tiene un punto estrictamente factible (similarmente para  $\mathcal{D}^*$  y el conjunto factible primal). La prueba hace uso de un 'teorema de alternativa' (ver por ejemplo, Mangasarian [capítulo 2]), omitimos aquí la prueba.

#### 1.4 LA PARTICION $B \cup N$ Y LA COMPLEMENTARIEDAD ESTRICTA

En esta sección damos un teorema que nos muestra la existencia de una solución estrictamente complementaria, es decir de una solución  $(x^*, y^*, s^*)$  que cumple

$$x_j^* s_j^* = 0 \quad \text{y} \quad x_j^* + s_j^* > 0.$$

Este resultado importante para los métodos de punto interior es conocido como el **Teorema de Goldman-Tucker**.

Para esto definamos dos conjuntos de índices  $B$  y  $N$  como sigue

$$B = \{j \in \{1, 2, \dots, n\} / x_j^* \neq 0 \text{ para algún } x^* \in \mathcal{P}^*\}, \quad (1.5)$$

$$N = \{j \in \{1, 2, \dots, n\} / s_j^* \neq 0 \text{ para algún } (y^*, s^*) \in \mathcal{D}^*\}, \quad (1.6)$$

El siguiente lema tomado de Wright prueba que estos conjuntos son disjuntos.

**Lema 1.6:** Sean  $B$  y  $N$  definidos como en (1.5) y (1.6) respectivamente. Entonces,  $B \cap N = \emptyset$ .

**Prueba:** Suponga que existiese un índice  $j$  que pertenece a ambos conjuntos entonces existiría una solución primal  $x^*$  y una solución dual  $(y^*, s^*)$  tal que  $x_j^* > 0$  y  $s_j^* > 0$ . Pero entonces  $x_j^* s_j^* > 0$  contradiciendo la condición de complementariedad. Luego  $B \cap N = \emptyset$ . ■

Lo que deseamos es probar que  $B \cup N$  forma una partición del conjunto  $\{1, 2, \dots, n\}$  para lo cual solo nos falta probar que  $B \cup N = \{1, 2, \dots, n\}$  y este resultado es conocido como *teorema de Goldman-Tucker*.

Para probar este teorema, necesitamos el conocido lema de Farkas el cual lo presentamos a continuación sin prueba y a sido tomado de Wright [19].

**Lema 1.7 (Lema de Farkas)**

Para cada matriz  $G \in \mathbb{R}^{p \times n}$  y para cada vector  $g \in \mathbb{R}^n$ , se cumple I) o II) pero nunca ambos

I)  $Gd > 0$ ,  $g^T d < 0$ , tiene una solución  $d \in \mathbb{R}^n$

II)  $G^T h = g$ ,  $h > 0$ , tiene una solución  $h \in \mathbb{R}^p$ . ■

Geoméricamente, este resultado muestra que si definimos un cono tomando una combinación convexa de las filas de  $G$ , entonces  $g$  está en el cono (caso II) o hay un plano definido por  $d$  que separa  $g$  del cono (Caso I).

Obtenemos un simple corolario del lema de Farkas introduciendo algunas igualdades.

**Corolario 1.8:** Para cada par de matrices  $G \in \mathbb{R}^{p \times n}$  y  $H \in \mathbb{R}^{q \times n}$  y cada vector  $g \in \mathbb{R}^n$ , se cumple I) o II) pero nunca ambos.

I)  $Gd > 0$ ,  $Hd = 0$ ,  $g^T d < 0$ , tiene una solución  $d \in \mathbb{R}^n$

II)  $G^T h + H^T \bar{h} = g$ ,  $h > 0$ , tiene una solución  $h \in \mathbb{R}^p$ ,  $\bar{h} \in \mathbb{R}^q$ . ■

Este corolario es usado para probar el teorema de Goldman-Tucker el cual presentamos a continuación.

**Teorema 1.9:**  $B \cup N = \{1, 2, \dots, n\}$ . Esto es, existe al menos una solución primal  $x^* \in \mathcal{P}^*$  y una solución dual  $(y^*, s^*) \in \mathcal{D}^*$  tal que  $x^* + s^* > 0$ .

**Prueba:** Sea  $J$  el conjunto de índices en  $\{1, 2, \dots, n\}$  que no están en  $B$  o  $N$ . Probaremos el resultado mostrando que  $J$  es vacío. Hemos probado en el lema 1.6 que  $B \cap N = \emptyset$  luego  $B \cup N \cup J$  es una partición de  $\{1, 2, \dots, n\}$ .

Si denotamos por  $A_B$  y  $A_J$  las submatrices formadas por las columnas de  $A$  que corresponden a  $B$  y  $J$ , respectivamente

$$\begin{aligned} A_{\cdot i}^T w &< 0, \\ -A_{\cdot j}^T w &\geq 0, \quad \text{para } j \in J \setminus \{i\}, \\ A_B^T w &= 0, \end{aligned} \tag{1.7}$$

donde  $A_{\cdot i}$  denota la  $i$ -ésima columna de  $A$ .

Suponga primero que  $w$  satisfaciendo (1.7) existe.

Sea  $(x^*, y^*, s^*)$  una solución primal-dual para la cual  $s_N^* > 0$  y definamos el vector  $(\bar{y}, \bar{s})$  como

$$\bar{y} = y^* + \varepsilon w, \quad \bar{s} = c - A^T \bar{y} = s^* - \varepsilon A^T w,$$

escogiendo  $\varepsilon > 0$  suficientemente pequeño tal que

$$\begin{aligned} \bar{s}_i &= s_i^* - \varepsilon A_{\cdot i}^T w > 0, \\ \bar{s}_j &= s_j^* - \varepsilon A_{\cdot j}^T w \geq 0, \quad j \in J \setminus \{i\}, \\ \bar{s}_B &= s_B^* = 0, \\ \bar{s}_N &= s_N^* - \varepsilon A_N^T w > 0. \end{aligned}$$

Se sigue de estas relaciones que  $(\bar{y}, \bar{s})$  es factible para el problema dual. En efecto, esta es también optimal, desde que todo vector solución primal  $x^*$  debe tener  $x_N^* = 0$  y de este modo  $\bar{s}^T x^* = 0$ . De este modo por la definición (1.6), debemos tener  $i \in N$ . Suponga, alternativamente, que el vector  $w$  no satisface (1.7). Aplicando el corolario 1.8, deducimos que el siguiente sistema debe tener una solución:

$$\sum_{j \in J \setminus \{i\}} h_j A_j + A_B \bar{h} = A_i, \quad h_j > 0 \quad \forall j \in J \setminus \{i\}$$

Ahora definiendo un vector  $v \in \mathbb{R}^{(j)}$  como

$$v_i = 1, \quad v_j = h_j \quad \forall j \in J \setminus \{i\}, \quad (1.8)$$

podemos reescribir (1.8) como

$$A_j v = A_B \bar{h}, \quad v \geq 0, \quad v_i > 0 \quad (1.9)$$

Ahora, sea  $x^*$  una solución primal para la cual  $x_B^* > 0$ , y definimos  $\bar{x}$  por

$$\bar{x}_B = x_B^* - \varepsilon \bar{h}, \quad \bar{x}_J = \varepsilon v, \quad \bar{x}_N = 0.$$

Sustituyendo en (1.9), tenemos que  $A \bar{x} = b$ , y para  $\varepsilon > 0$  suficientemente pequeño, también tenemos que  $\bar{x} \geq 0$ . Así  $\bar{x}$  es factible y, en efecto, optimal debido a  $\bar{x}_N = 0$ . Desde que  $v_i = 1$ , también tenemos que  $\bar{x}_i = \varepsilon > 0$ ; aquí,  $i \in B$  por (1.5)

Hemos probado que el índice  $i \in J$  también, de esta forma, por la definición de  $J$ , tenemos  $J = \emptyset$  ■

## 1.5 RANGO DE LA MATRIZ A

En el desarrollo de los algoritmos de punto interior es a menudo conveniente asumir que la matriz  $A \in \mathbb{R}^{m \times n}$  en (1.1) tiene rango  $m$ . Existen problemas de programación lineal que no satisfacen esta condición, pero tales problemas pueden siempre ser transformados a problemas equivalentes que si satisfacen esta asunción. Por ejemplo, el problema trivial

$$\begin{aligned} & \min \quad x_1 + x_2 \\ & \text{sujeto a} \quad \begin{bmatrix} 1 & 2 \\ -2 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -6 \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geq 0, \end{aligned}$$

tiene la solución (primal)  $x^* = (0, 3/2)$  pero la matriz de restricciones  $2 \times 2$  no tiene rango 2 ( $m = 2$ ). La segunda restricción es obviamente redundante, podemos eliminar esta y considerar el problema equivalente

$$\begin{aligned} & \min \quad x_1 + x_2 \\ & \text{sujeto a} \quad \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3, \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geq 0, \end{aligned}$$

En general, la matriz  $A$  puede ser reducida a una matriz de rango  $m$ , vía la factorización QR. La factorización QR aplicados a un sistema consistente  $Ax = b$ , obtiene una matriz ortogonal  $Q_{m \times n}$  tal que

$$QA = \begin{bmatrix} \bar{A} \\ 0 \end{bmatrix}, \quad Qb = \begin{bmatrix} \bar{b} \\ 0 \end{bmatrix}$$

Donde  $\bar{A}$  y  $\bar{b}$  tienen el mismo número de filas y  $\bar{A}$  tiene rango  $m$ . Los sistemas  $Ax = b$  y  $\bar{A}x = \bar{b}$  son equivalentes, esto es, todo vector  $x$  que satisface una de estas ecuaciones también satisface la otra. El algoritmo de la factorización QR se encuentra en textos especializados de matemática numérica.

## 1.6 DIRECCIONES ADMISIBLES Y TEST DE LA RAZÓN

La palabra dirección es sinónimo de vector. Un vector  $p \in \mathbb{R}^n$  es llamado un vector de dirección, cuando se está interesado en estudiar puntos  $x + \alpha p$  para  $\alpha \geq 0$ . Estos son los puntos obtenidos desplazándose sobre la dirección  $p$  a partir del punto  $x$ . Esto es ilustrado en la siguiente figura

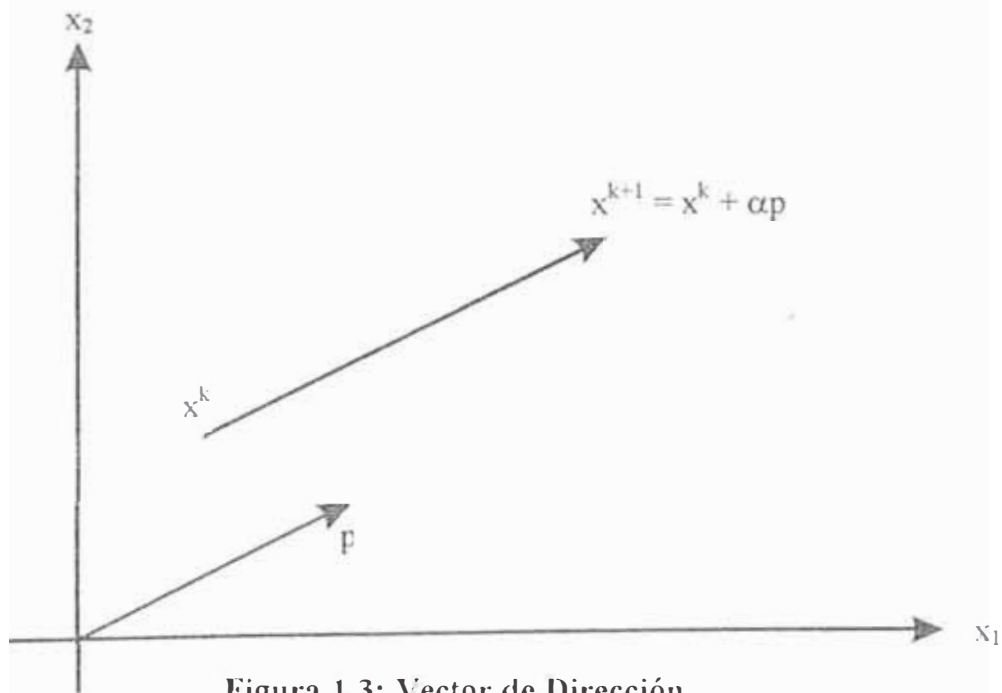


Figura 1.3: Vector de Dirección

**Definición 1.4:** Un vector  $p \in \mathbb{R}^n$  es llamado una dirección admisible a partir de  $x \in \mathcal{P}$  ( $\mathcal{P}$  es el conjunto admisible del problema (1.1)) si y solo si existe  $\alpha > 0$  tal que para cualquier  $\alpha \in [0, \bar{\alpha}]$ ,  $x + \alpha p \in \mathcal{P}$ .

**Lema 1.10:** Sea  $x$  un punto interior de  $\mathcal{P}$ . Entonces,  $p \in \mathbb{R}^n$  es una dirección admisible a partir de  $x$  si y solo si  $p \in N(A)$ .

**Prueba:** Como  $x + \alpha p \in \mathcal{P}$

$$\begin{aligned} \Leftrightarrow A(x + \alpha p) = b &\Leftrightarrow Ap = 0 \\ &\Leftrightarrow p \in N(A) \quad \blacksquare \end{aligned}$$

Como siempre trabajamos con puntos interiores, las direcciones admisibles serán normalmente confundidas con los vectores de  $N(A)$ . Será siempre necesario saber cual es la mayor longitud de paso que se puede tomar sobre la dirección admisible, de modo de preservar factibilidad. La mayor longitud de paso está determinada por el test de la razón, presentado en el siguiente lema.

**Lema 1.11:** Consideremos un punto  $x \in \mathcal{P}$  y una dirección  $p \in \mathbb{R}^n$ , admisible a partir de  $x$ .

Entonces:

Si  $p \geq 0$ ,  $x + \alpha p$  es admisible para cualquier  $\alpha > 0$ .

Si no, el mayor valor de  $\alpha$ , tal que  $x + \alpha p \in \mathcal{P}$  es dado por

$$(1.10) \quad \alpha = \min \left\{ \frac{-x_i}{p_i} \mid p_i < 0, \quad i = 1, \dots, n \right\}$$

**Prueba:** Como  $p \in N(A)$  entonces para cualquier  $\alpha \in \mathbb{R}$  se tiene que  $A(x + \alpha p) = b$ . Basta examinar la restricción de no negatividad. Si  $p > 0$ , entonces para cualquier  $\alpha > 0$ ,  $x + \alpha p > 0$ , probando la primera parte.

Si alguna componente de  $p$  es negativa, entonces la máxima longitud de paso va a ser determinado por la primera entre esas variables a anularse. A cada componente  $i = 1, \dots, n$  asóciase un valor

$$\bar{\alpha}_i = \sup \{ \alpha / x_i + \alpha p_i \geq 0 \}$$

de lo anterior si,  $p_i \geq 0$  entonces  $\bar{\alpha}_i = +\infty$ . Si la  $p_i < 0$ , entonces  $x_i + \bar{\alpha}_i p_i = 0$  y luego  $\bar{\alpha}_i = -\frac{x_i}{p_i}$ . El mayor paso que garantiza la factibilidad es ahora

$$\min_{i=1, \dots, n} \bar{\alpha}_i,$$

el cual es calculado en la expresión (1.10) ■

## 1.7 TAMAÑO DE UN PROBLEMA DE PROGRAMACIÓN LINEAL

En esta sección tratamos el problema de la determinación del número de iteraciones que requiere un algoritmo para resolver el problema de programación lineal (1.1) hasta la precisión deseada. Para estudiar este problema, necesitamos definir una "medida" del tamaño del problema lineal (1.1), como también lo que debemos entender por una instancia del problema de programación lineal, así como también el concepto de tiempo de ejecución de un algoritmo y el concepto de algoritmo polinomial.

Según Padberg [14] se tiene la siguiente definición:

**Definición 1.5:** Por una instancia del problema de programación lineal 1.1 entenderemos un conjunto de valores para la matriz  $A$ , para el vector  $c$  y para el vector  $b$  de restricciones.

**Definición 1.6:** Sea  $\mathcal{A}$  un algoritmo para resolver un determinado problema de optimización combinatoria. El tiempo de ejecución  $t(n)$  de  $\mathcal{A}$  es el máximo número de operaciones elementales en el computador que requiere  $\mathcal{A}$  para resolver una determinada instancia del problema de tamaño  $n$  (aquí el máximo está tomado sobre todas las instancias de tamaño  $n$ ).

**Definición 1.7:** El algoritmo  $\mathcal{A}$  es *polinomial* si y solo si existe un polinomio  $p(n)$  tal que:

$$t(n) = \mathcal{O}(p(n)).$$

En caso contrario, se dirá que  $\mathcal{A}$  es *exponencial*.

Estamos usando aquí el concepto de algoritmo en términos bastante amplios como cualquier método computacional capaz de generar una solución para todas las instancias del problema.

Sabemos que los datos descriptivos del problema lineal (1.1) son los siguientes:

- $n$  el número de variables
- $m$  el número de restricciones
- $c_j$  los coeficientes de la función objetivo
- $b_i$  los elementos del vector del lado derecho
- $a_{ij}$  los coeficientes tecnológicos

La primera medida del tamaño del problema (1.1) la llamaremos *tamaño digital*, pero introduciremos una nueva medida  $L$  de (1.1) con la cual será más fácil de trabajar. Además probaremos que

$$L < \text{tamaño digital de (1.1)},$$

así que todo algoritmo polinomial en  $L$  también será un algoritmo en tiempo polinomial en el tamaño digital de (1.1)

Asumiremos, sin pérdida de generalidad, que todos los elementos de la matriz de coeficientes  $A$ , así como también los elementos de  $b$  y  $c$  son números enteros, ya que todo problema de programación lineal con coeficientes racionales puede ser transformado fácilmente en uno con coeficientes enteros (multiplicando por el M.C.M.).

Según Padberg [14] para cada número entero  $n$ , se define su tamaño como sigue:

**Definición 1.8:** Sea  $n \in \mathbb{Z}$ , el tamaño de  $n$  es denotado por  $\langle n \rangle$  y definido por

$$\langle n \rangle = 1 + \lceil \log_2(1 + |n|) \rceil$$

$\langle n \rangle$  representa el número de bits necesarios para escribir  $n$  en binario y donde el primer 1 es necesario para almacenar el signo de  $n$ . Aquí  $\lceil \cdot \rceil$ , denotará el máximo entero.

Análogamente, definimos el tamaño de un vector  $p \times 1$ , y de una matriz  $p \times l$  como sigue.



**Definición 1.9:** Sea  $v \in \mathbb{R}^p$ , y  $M \in \mathbb{R}^{p \times l}$  el tamaño de  $v$  y  $M$  se denota  $\langle v \rangle$ ,  $\langle M \rangle$  y se define

por:

$$\langle v \rangle = \sum_{i=1}^p \langle v_i \rangle,$$

$$\langle M \rangle = \sum_{i=1}^p \sum_{j=1}^l \langle m_{ij} \rangle,$$

Estamos ahora en condiciones de definir el tamaño digital de un problema de programación lineal.

**Definición 1.10:** Consideremos el problema de programación lineal (1.1), entonces, el tamaño digital de este problema es denotado por  $\langle LP \rangle$  y definido por

$$\langle LP \rangle = \langle A \rangle + \langle b \rangle + \langle c \rangle$$

Una definición más conveniente del tamaño de un problema de programación lineal es la siguiente:

**Definición 1.11:** El tamaño de un problema de programación lineal es dado por

$$L = \langle \det_{\max} \rangle + \langle b_{\max} \rangle + \langle c_{\max} \rangle + m + n$$

donde

$$\det_{\max} = \max_{A'} (|\det(A')|),$$

$$b_{\max} = \max_i (|b_i|),$$

$$c_{\max} = \max_j (|c_j|),$$

y  $A'$  es cualquier submatriz cuadrada de  $A$ .

El siguiente lema adaptado de Padberg [14], se usará para probar que  $L < \langle LP \rangle$

**Lema 1.12:**

1) Si  $n \in \mathbb{Z}$ , entonces,  $|n| < 2^{(n)-1} - 1$ .

2) Si  $v \in Z^n$ , entonces,  $\|v\| \leq \|v\|_1 \leq 2^{\langle v \rangle - n} - 1$ .

3) Si  $A \in Z^{n \times n}$ , entonces,  $|\det(A)| \leq 2^{\langle A \rangle - n^2} - 1$ .

**Prueba:**

1) Por definición

$$\begin{aligned} 2) \quad 1 + \|v\| &\leq 1 + \|v\|_1 = 1 + \sum_{i=1}^n |v_i| \leq \prod_{i=1}^n (1 + |v_i|) \\ &\leq \prod_{i=1}^n 2^{\langle v_i \rangle - 1} = 2^{\langle v \rangle - n} \end{aligned}$$

3) Si  $a_1, \dots, a_n$  las columnas de  $A$ .

Usando desigualdad de Hadamard

$$|\det A| \leq \prod_{i=1}^n \|a_i\|,$$

y lo probado en 2), tenemos

$$\begin{aligned} 1 + |\det(A)| &\leq 1 + \prod_{i=1}^n \|a_i\| \leq \prod_{i=1}^n (1 + \|a_i\|) \\ &\leq \prod_{i=1}^n 2^{\langle a_i \rangle - n} = 2^{\langle A \rangle - n^2} \end{aligned}$$

de donde

$$|\det(A)| \leq 2^{\langle A \rangle - n^2} - 1 \quad \blacksquare$$

El siguiente lema adaptado de Padberg [14] prueba que  $L$  es menor que el tamaño digital de (1.1).

**Lema 1.13:**

$$L < \langle LP \rangle, \quad \forall A, b, c.$$

**Prueba:** Si  $B$  es una submatriz cuadrada de  $A$  entonces, por definición,  $\langle B \rangle < \langle A \rangle$ .

Además por el lema anterior, se tiene

$$1 + |\det(B)| \leq 2^{\langle B \rangle} - 1.$$

Por tanto

$$1) \quad \begin{aligned} \lceil \log(1 + |\det(B)|) \rceil &\leq \langle B \rangle - 1 \\ &< \langle B \rangle \leq \langle A \rangle \end{aligned}$$

Sea  $v \in Z^p$ . Entonces

$$\begin{aligned} \langle v \rangle &\geq \left\langle \max_j |v_j| \right\rangle + p - 1 \\ &= \left\lceil \log(1 + \max_j |v_j|) \right\rceil + p \end{aligned}$$

Por tanto:

$$2) \quad \begin{aligned} \langle b \rangle + \langle c \rangle &\geq \left\lceil \log(1 + \max_j |c_j|) \right\rceil \\ &\quad + \left\lceil \log(1 + \max_i |b_i|) \right\rceil + m + n \end{aligned}$$

Finalmente combinando las ecuaciones 1) y 2) obtenemos el resultado deseado. ■

En lo que sigue trabajaremos con  $L$  como el tamaño de la entrada para un algoritmo que resuelve el problema lineal (1.1).

Sabemos que si el problema de programación lineal (1.1) es factible, hay al menos un vértice el cual es una solución óptima.

Entonces, cuando buscamos una solución óptima para (1.1) podemos restringir nuestra atención solo a los vértices. El siguiente teorema tomado de Padberg [14], nos asegura que tales vértices tienen una representación compacta.

**Teorema 1.14:** Sea  $x$  un vértice poliedro definido por  $Ax = b$ ,  $x \geq 0$ . Entonces,

$$x^T = \left( \frac{p_1}{q}, \frac{p_2}{q}, \dots, \frac{p_n}{q} \right),$$

donde

$$p_i, \quad q \in \mathbb{N}, \quad i = 1, \dots, n$$

y

$$0 \leq p_i < 2^L,$$
$$1 \leq q < 2^L.$$

**Prueba:** Como  $x$  es una solución básica factible, existe una base  $B$  tal que  $x_B = A_B^{-1}b$  y  $x_N = 0$ . Entonces,  $p_j = 0 \quad \forall j \in N$ .

Luego solo nos interesarán los  $p_j$  con  $j \in B$ , por otro lado sabemos del álgebra lineal que

$$x_B = A_B^{-1}b = \frac{1}{\det(A_B)} (\text{cof}(A_B))^T b$$

donde  $\text{cof}(A_B)$  es la matriz de cofactores de  $A_B$ . Cada entrada de  $A_B$  consiste de un determinante de alguna submatriz de  $A$ . Sea  $q = |\det(A_B)|$ , entonces  $q$  es un entero ya que  $A_B$  tiene componentes enteros,  $q \geq 1$  ya que  $A_B$  es invertible, y  $q \leq \det_{\max} < 2^L$ . Finalmente desde

$$p_i = q x_{B_i} = |\text{cof}(A_B)_{i,j}| |b_j|$$
$$\leq m \det_{\max} b_{\max} < 2^L. \blacksquare$$

## 1.8 SOLUCIÓN DE UN PROBLEMA DE PROGRAMACIÓN LINEAL

Sabemos que el método gráfico es útil para resolver un problema de programación lineal con dos variables, pero en el caso de más variables es indispensable contar con una metodología algorítmica para resolver dichos problemas.

### 1.8.1 Solución por el método simplex

En 1947, Dantzing elaboró un método de solución para el problema de la programación lineal, y en esa oportunidad propuso un algoritmo, el cual (hablando en términos muy generales) evolucionaba a través de los vértices del poliedro determinado por las restricciones.

La ventaja de esta aproximación es que los cálculos para pasar de un vértice a otro nos obliga a efectuar a lo más  $\mathcal{O}(m^2 + mn)$  operaciones elementales, donde  $m$  es el número de restricciones y  $n$  el número de variables.

Pero sus inconvenientes son:

- 1) Dificultades en cuanto a la estabilidad numérica sobre determinados problemas de programación lineal.
- 2) El algoritmo puede ejecutar en ciertos casos un número exponencial de iteraciones del orden  $2^n$ , lo que en la práctica no es admitido; es decir, el método simplex no es polinomial.
- 3) Para determinar un vértice de partida, debe resolverse un problema de programación lineal del mismo valor de dificultad que el mismo problema inicial.

### **1.8.2 Solución por un método interior**

En la misma época de Dantzing, otros investigadores como Hoffman, Frish, Carroll, describieron una metodología distinta a la de Dantzing, llamada de punto interior.

Este método contrariamente al método simplex que progresa de vértice en vértice, evoluciona en el interior de la región factible y genera una sucesión  $\{x^k\}$  en el interior de la región factible, con la propiedad que para  $k$  suficientemente grande la solución aproximada  $x^k$  puede transformarse en una solución óptima  $x^*$ . El resultado más importante de estos métodos es la de estimar polinomialmente el número de operaciones elementales necesarias para obtener una solución óptima.

En la página siguiente presentamos un esquema gráfico que muestra la manera de operar de ambas aproximaciones, la del método simplex y la de punto interior.

A continuación se presenta un teorema debido a Khachiyan[9], el cual sirve para la demostración del corolario que nos mostrará que la cantidad  $2^{-2L}$  puede ser usada para la identificación de una solución óptima.

**Teorema 1.15:** Sean  $x_1, x_2$  vértices de

$$\begin{aligned} Ax &= b, \\ x &\geq 0. \end{aligned}$$

Si  $c^T x_1 \neq c^T x_2$ , entonces,  $|c^T x_1 - c^T x_2| > 2^{-2L}$ .

Prueba: Por el teorema 1.14 existen  $q_1, q_2$  tal que  $1 \leq q_1 < 2^L$ ,  $1 \leq q_2 < 2^L$  y  $q_1 x_1, q_2 x_2 \in N^n$ .

Además

$$\begin{aligned} |c^T x_1 - c^T x_2| &= \left| \frac{q_1 c^T x_1}{q_1} - \frac{q_2 c^T x_2}{q_2} \right| \\ &\geq \left| \frac{q_1 q_2 (c^T x_1 - c^T x_2)}{q_1 q_2} \right| \\ &\geq \frac{1}{q_1 q_2} \quad (\text{ya que } c^T x_1 - c^T x_2 \neq 0) \\ &\geq \frac{1}{2^L 2^L} = 2^{-2L} \end{aligned}$$

**Colorario 1.16:**

Si  $z = \min \{c^T x : Ax = b, x \geq 0\}$ ,

y si  $x$  es factible para el problema lineal (1.1), y tal que  $c^T x \leq z + 2^{-2L}$ .

Entonces, cada vértice  $x'$  tal que  $c^T x' \leq c^T x$  es una solución óptima del problema de programación lineal (1.1).

**Prueba:** Suponga que  $x'$  no es óptimo. Entonces existe  $x^*$  un vértice óptimo, tal que  $c^T x^* = z$ . Como  $x'$  no es óptimo,  $c^T x' \neq c^T x^*$ , y entonces por el teorema 1.15.

$$\begin{aligned} c^T x' - c^T x^* &> 2^{-2L} \\ c^T x' &> c^T x^* + 2^{-2L} \\ &= z + 2^{-2L} \\ &\geq c^T x && (\text{por definición de } x) \\ &\geq c^T x' && (\text{por definición de } x') \end{aligned}$$

por consiguiente  $c^T x' > c^T x'$  lo cual es una contradicción ■

Lo que este corolario dice es que no necesitamos demasiada precisión cuando escogemos un vértice optimal. Es decir, solo necesitamos completar la función objetivo con un error menor a  $2^{-2L}$ . Si encontramos un vértice que está dentro de este margen de error, entonces éste debe ser optimal. De esta forma, para propósitos teóricos, es común tomar  $\varepsilon = 2^{-2L}$  como criterio de parada de un algoritmo de punto interior.

Los métodos de punto interior para la programación lineal, usan los resultados del corolario en la siguiente forma: Suponga que  $f^*$  es el valor óptimo del problema (1.1), y sea  $\{x^k\}$  la sucesión generada por un método de punto interior tal que para algún  $k$  tenemos

$$c^T x^k - f^* \leq 2^{-2L} \quad (1.11)$$

Como  $c^T x^* - b^T y^* = 0$  y  $b^T y^k \leq b^T y^*$ ,

Tenemos  $(s^k)^T x^k = c^T x^k - b^T y^k = c^T x^k - c^T x^* + b^T y^* - b^T y^k \geq c^T x^k - c^T x^*$

y el test (1.11) se puede reemplazar por

$$c^T x^k - b^T y^k \leq 2^{-2L} \quad \text{ó} \quad (s^k)^T x^k \leq 2^{-2L}$$

Por otro lado un punto  $x^k$  que satisface (1.11) puede transformarse en un punto extremo  $\hat{x}$  con la propiedad  $c^T \hat{x} \leq c^T x^k$ , tales transformaciones son llamadas a menudo procedimientos de redondeo.

La idea del procedimiento de redondeo es iniciar en  $x^k$  y en cada paso generar un punto factible el cual tiene al menos una coordenada nula mas que el punto anterior y que al mismo tiempo no incrementa el valor de la función objetivo.

# CAPITULO 2

## EL METODO DE BARRERA LOGARITMICA PARA LA PROGRAMACIÓN LINEAL

En este capítulo aplicamos el método de barrera logarítmica al problema de la programación lineal, y vemos como esta metodología de punto interior puede ser usada para derivar aspectos centrales de la teoría de la programación lineal. Usando solo resultados básicos del cálculo y del álgebra lineal, probamos que para cada valor del parámetro de barrera, la función de barrera logarítmica, tiene minimizadores únicos y la trayectoria definida por estos minimizadores (trayectoria central) converge a una solución óptima estrictamente complementaria. Esto nos da nuevas pruebas de los resultados fundamentales de la programación lineal, incluyendo la existencia de una solución estrictamente complementaria y el teorema de dualidad fuerte.

Los resultados de este capítulo, en particular la existencia de la trayectoria central son utilizados para generar algoritmos polinomiales para la programación lineal basados en la idea de seguimiento de trayectoria (i.e. algoritmos de seguimiento de trayectoria).

Este capítulo ha sido desarrollado teniendo como base los artículos de Güler [7], Megiddo [9], Kojima [10], y los libros de Wright [19] y Bazaraa - Shatty [2], este último para un resultado del análisis convexo.

### 2.1 INTRODUCCION

En esta sección, damos algunas definiciones y teoremas necesarios para darle fluidez a lo que desarrollaremos posteriormente.

Consideremos un problema de programación lineal en la forma estándar



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

donde  $A \in \mathbb{R}^{m \times n}$ ,  $c, x \in \mathbb{R}^n$ , y  $b \in \mathbb{R}^m$ . El problema lineal (P) tiene asociado el programa dual

$$(D) \quad \max \{b^T y : A^T y + s = c, s \geq 0\},$$

con  $s \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^m$ . Las regiones factibles de (P) y (D) serán denotados como en el capítulo 2 por:

$$\mathcal{P} := \{x : Ax = b, x \geq 0\},$$

$$\mathcal{D} := \{(y, s) : A^T y + s = c, s \geq 0\}$$

Asumiremos que  $A$  tiene rango  $m$ . Esto implica que existe una correspondencia uno a uno, entre  $y$  y  $s$  y los pares  $(y, s) \in \mathcal{D}$ .

Los interiores de  $\mathcal{P}$  y  $\mathcal{D}$  serán denotados con  $\mathcal{P}^0$  y  $\mathcal{D}^0$ , así tenemos:

$$\mathcal{P}^0 := \{x : Ax = b, x > 0\},$$

$$\mathcal{D}^0 := \{(y, s) : A^T y + s = c, s > 0\}.$$

Ahora presentamos un conocido resultado del análisis convexo. Un resultado relacionado puede encontrarse en Bazaraa y Shetty [2] (teorema 3.4.3 y Corolario 1, pp. 101-102), pero antes enunciamos una definición, la cual es también tomada de esta referencia.

**Definición 2.1:** El conjunto  $C \subset \mathbb{R}^k$  es llamado *abierto relativo*, si  $C$  es abierto en el menor subespacio afin de  $\mathbb{R}^k$  que contiene a  $C$ .

**Teorema 2.1:** Sea  $f : D \rightarrow \mathbb{R}$  una función convexa diferenciable, donde  $D \subset \mathbb{R}^k$  es un conjunto abierto y sea  $C$  un subconjunto abierto relativo de  $D$ . Entonces,  $x \in C$  minimiza a  $f$  sobre  $C$  si y solo si

$$\nabla f(x)^T (z - x) = 0, \quad \forall z \in C \tag{2.1}$$

**Prueba:** Como  $f$  es convexa y diferenciable, tenemos que  $\forall z, x \in C$

$$f(z) > f(x) + \nabla f(x)^T (z - x)$$

La suficiencia de la condición (2.1) se sigue inmediatamente.

Para probar la necesidad de (2.1), considere el punto  $z_t := x + t(z - x)$ ,  $0 \leq t \leq 1$ , tenemos que  $z_t \in C$  y entonces

$$f(z_t) \geq f(x) + t \nabla f(x)^T (z - x).$$

Ya que  $f(z_t) > f(x)$ , haciendo  $t \rightarrow 0$  tenemos:

$$\nabla f(x)^T (z - x) = \lim_{t \rightarrow 0} \frac{f(z_t) - f(x)}{t} \geq 0.$$

Ahora, tomando a  $x$  como el mínimo de  $f$ . Si  $z \in C$  y  $w := z - x$ , entonces para algún  $\lambda > 0$  también  $z := x - \lambda w \in C$ . Ahora tenemos  $\nabla f(x)^T w \geq 0$  y  $\nabla f(x)^T (-\lambda w) \geq 0$  luego (2.1) se cumple. ■

En esta sección asumiremos que los conjuntos  $\mathcal{P}^0$  y  $\mathcal{D}^0$  son no vacíos. Nos referimos a este hecho como la *hipótesis de punto interior*, y es generalmente asumida en el estudio de métodos de punto interior.

**Hipótesis 2.1:** Existen  $(x^0, y^0, s^0)$  tal que

$$Ax^0 = b, \quad x^0 > 0, \quad A^T y^0 + s^0 = c, \quad s^0 > 0$$

es decir  $\mathcal{P}^0$  y  $\mathcal{D}^0$  contienen un vector positivo.

Posteriormente, consideraremos el caso donde la hipótesis 2.1 es relajada.

## 2.2 LA FUNCIÓN DE BARRERA LOGARÍTMICA

Definimos a continuación, la función de barrera logarítmica.

**Definición 2.2:** Para cada número  $\mu > 0$ , definimos:

$$\beta : \mathcal{P}^0 \rightarrow \mathbb{R}$$

por 
$$\beta(x, \mu) := c^T x - \mu \sum_{j=1}^n \ln x_j$$

$\beta$  es llamada la *función barrera logarítmica* para el problema (P) con parámetro de barrera  $\mu$ .

El método de barrera logarítmica, usualmente usado en programación no lineal, consiste en estudiar la familia de problemas.

$$(P_\mu) \quad \min \{ \beta(x, \mu) : x \in \mathcal{P}^0 \}$$

y estudiar las propiedades de sus soluciones. La idea básica es resolver el problema  $(P_\mu)$  para diferentes valores del parámetro  $\mu$ , con  $\mu$  decreciendo a cero y obtener como resultado, una sucesión de puntos factibles convergiendo a la solución óptima del problema  $(P)$ .

El siguiente lema, adaptado de Güler [7], establece una equivalencia importante.

**Lema 2.2:** Sea  $\mu > 0$ . Las siguientes afirmaciones son equivalentes:

- i) Existe un único minimizador de  $\beta(x, \mu)$
- ii) Existen,  $x, s \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^m$  tal que

$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0 \\ Xs &= \mu e \end{aligned} \tag{2.2}$$

donde  $X := \text{diag}(x_1, \dots, x_n)$  y  $e = (1, \dots, 1)^T$ .

**Prueba:** Primero nótese que siempre que  $(x, y, s)$  resuelven (2.2), entonces debido a la tercera ecuación  $x$  y  $s$  son ambos estrictamente positivos. De este modo las condiciones de no negatividad para  $x$  y  $s$  en (2.2) pueden ser reemplazados por requerir que  $x$  y  $s$  sean estrictamente positivos.

Ahora, es fácil probar que  $\beta(x, \mu)$  es estrictamente convexa, para esto nótese que

$$\nabla \beta(x, \mu) = c - \mu X^{-1} e, \quad \nabla^2 \beta(x, \mu) = \mu X^{-2}$$

Sea ahora  $v \in \mathbb{R}^n \setminus \{0\}$ ,  $v^T \nabla^2 \beta(x, \mu) v = \mu v^T X^{-2} v$

$$= \mu \sum_{j=1}^n (v_j x_j^{-1})^2 > 0$$

esto prueba que el hessiano de  $\beta(x, \mu)$  es definido positivo, lo cual implica que  $\beta(x, \mu)$  es estrictamente convexa.

(i)  $\Rightarrow$  (ii)

Como  $\beta(x, \mu)$  es estrictamente convexa. Entonces existe un único minimizador de  $\beta(x, \mu)$  sobre el conjunto  $\mathcal{P}^0$ . Desde que el dominio de  $\beta(x, \mu)$  es un abierto relativo, el teorema 2.1 puede aplicarse. Entonces se obtiene que  $\beta(x, \mu)$  tiene a  $x \in \mathcal{P}^0$  como un minimizador si y solo si

$$\begin{aligned} Ax &= b, \quad x > 0, \\ (c - \mu X^{-1}e)^T w &= 0, \quad \forall w \in N(A) \end{aligned}$$

desde  $N(A)$  denota el núcleo de  $A$ . La segunda ecuación es equivalente a afirmar que  $c - \mu X^{-1}e$  pertenece al complemento ortogonal de  $N(A)$ .

Sabemos por un resultado del álgebra lineal que el complemento ortogonal de  $N(A)$  es el espacio fila  $R(A^t)$  de  $A$ . Entonces se sigue que  $x \in \mathcal{P}^0$  minimiza  $\beta(x, \mu)$  si y solo si

$$c - \mu X^{-1}e = A^T y$$

para algún  $y \in R^m$  haciendo  $s := \mu X^{-1}e$  se prueba que (i)  $\Rightarrow$  (ii).

(ii)  $\Rightarrow$  (i) Es trivial, ya que el sistema (2.2) son las condiciones de óptimo para el problema  $(P\mu)$ , el cual al ser un problema convexo, estas condiciones son necesarias y suficientes. ■

Sabemos por lo desarrollado en el capítulo 1 que las condiciones de óptimo para los problemas (P) y (D) son dadas por el sistema no lineal.

$$\begin{aligned} Ax &= b, \quad x > 0, \\ A^t y + s &= c, \quad s \geq 0, \\ Xs &= 0, \\ X &:= \text{diag}(x_1, \dots, x_n) \end{aligned}$$

Comparando este sistema con el sistema (2.2) podemos ver que los métodos de punto interior relajan la condición (no lineal) de complementariedad  $Xs = 0$ .

Por razones que quedarán claras más adelante las ecuaciones en el sistema (2.2) son conocidas como *condiciones de centrado*. Megiddo [9] (1989), Bayer y Lagarias [1] (1989) y Sonnevend [16] (1986) fueron los primeros en darse cuenta de la importancia del sistema

(2.2) y lo analizarán en detalle. Los resultados más importantes los iremos presentando en el desarrollo de esta sección.

**2.2.1 Función de barrera dual:** Sea  $H$  la matriz cuyas filas forman una base para el subespacio  $N(A)$ . Entonces,  $N(A) = R(H^T)$ ,  $R(A^T) = N(H)$ , y  $R^n = R(H^T) \oplus N(H)$  es una descomposición ortogonal de  $R^n$ . Ahora sea  $x^0 \in \mathcal{D}^0$  y  $(y^0, s^0) \in \mathcal{D}^0$ . Entonces la primera ecuación en el sistema (2.2) puede escribirse como  $Ax = Ax^0$ , y, usando la matriz  $H$ , la segunda ecuación como  $Hs = Hs^0$ .

Entonces obtenemos la siguiente forma simétrica de (2.2):

$$\begin{aligned} Ax &= Ax^0, & x &\geq 0, \\ Hs &= Hs^0, & s &\geq 0, \\ Xs &= \mu e. \end{aligned} \tag{2.3}$$

Usando la simetría en  $x$  y  $s$  podemos aplicar el *lema 2.2* a la función

$$h: \{s \mid Hs = Hs^0, s > 0\} \rightarrow \mathbb{R}$$

definida por

$$h(s, \mu) := (x^0)^T s - \mu \sum_{j=1}^n \ln s_j.$$

Esto prueba que  $h$  tiene un único maximizador si y solo si el sistema de ecuaciones (2.2) tiene una solución. Observemos que,

$$\begin{aligned} (x^0)^T s &= (x^0)^T (c - A^T y) \\ &= (x^0)^T c - b^T y \end{aligned} \tag{2.4}$$

Así, definimos la función de barrera logarítmica para el problema (D) por.

**Definición 2.3:** Para cada  $\mu > 0$ , definimos:

$$\begin{aligned} \beta_D &: \mathcal{D}^0 \rightarrow \mathbb{R} \\ \text{por } \beta_D(s, \mu) &:= b^T y + \mu \sum_{j=1}^n \ln s_j, \end{aligned}$$

la cual es llamada función de barrera logarítmica dual.

Nótese que  $\beta_D$  y  $-h$  difieren por una constante,  $\beta_D$  tiene un maximizador  $s = s(\mu)$  si y solo si el sistema (2.2) tiene una solución. La prueba del siguiente teorema, usa este resultado, así como también un lema del cálculo, el cual lo enunciamos sin prueba.

**Lema 2.3:** Sea la función  $f : ]0, +\infty[ \rightarrow \mathbb{R}$  definida por  $f(x) := \sigma x - \mu \ln x$ , donde  $\sigma$  y  $\mu$  son positivos. Entonces  $f(x)$  está acotado inferiormente. Además, para cada  $m \in \mathbb{R}$  existe un intervalo  $[a, b]$ , donde  $0 < a < b$ , tal que  $f(x) \geq m$  siempre que  $x \notin [a, b]$ . ■

El siguiente teorema es adaptado de Güler [7].

**Teorema 2.4:** Sea  $\mu > 0$ . Entonces, las siguientes afirmaciones son equivalentes:

- i)  $\mathcal{P}$  y  $\mathcal{D}$  contienen vectores positivos;
- ii) Existe un único minimizador de  $\beta(x, \mu)$ ;
- iii) Existe un único maximizador de  $\beta_D(s, \mu)$ ;
- iv) El sistema (2.2) tiene una solución.

**Prueba:** La equivalencia de ii) y iv) está contenida en el Lema 2.2, y la equivalencia de ii) y iii) la hemos visto líneas arriba, cuando construimos la función de barrera logarítmica dual.

Entonces, faltaría probar que (i) es equivalente a (ii) - (iv), note que es obvio que (iv) implica (i). Así para la prueba del teorema es suficiente probar que (i)  $\Rightarrow$  (ii).

Asumiendo (i), tenemos  $x^0 \in \mathcal{P}^0$  y  $(y^0, s^0) \in \mathcal{D}^0$  tal que  $x^0 > 0$  que  $x^0 > 0$  y  $s^0 > 0$ .

Primero observe que minimizar la función barrera  $\beta(x, \mu)$  sobre el conjunto  $\mathcal{P}^0$  es equivalente a minimizar la función.

$$K(x, \mu) := (s^0)^T x - \mu \sum_{j=1}^n \ln x_j,$$

sobre el mismo conjunto. Esto es debido a la relación

$$(s^0)^T x = (c - A^T y^0)^T = c^T x - b^T y^0$$

Así la prueba esta completa si probamos que  $K(x, \mu)$  tiene un minimizador en  $\mathcal{P}^0$ . Para este fin escribimos

$$K(x, \mu) = \sum_{j=1}^n (s_j^0 x_j - \mu \ln x_j)$$

Debido al Lema 2.3, tenemos números  $\tau_j$  tal que el  $j$ -ésimo término en la suma dada anteriormente es acotada inferiormente por  $\tau_j$ . Además, definiendo

$$m_j := K(x^0, \mu) - (\tau_1 + \dots + \tau_{j-1} + \tau_{j+1} + \dots + \tau_n), \quad 1 < j \leq n$$

Tenemos  $m_j > \tau_j$ . Usando el lema 2.3 repetidamente, obtenemos la existencia de intervalos  $[a_j, b_j]$ ,  $0 < a_j < b_j$ , tal que el valor del  $j$ -ésimo término excede  $m_j$ , siempre que  $x_j \in [a_j, b_j]$ . Así, si  $x \in \mathcal{P}^0$  y  $x_j \notin [a_j, b_j]$  para algún  $j$ , entonces

$$K(x, \mu) > \tau_1 + \dots + \tau_{j-1} + m_j + \tau_{j+1} + \dots + \tau_n \geq K(x^0, \mu)$$

Concluimos de esto que el conjunto

$$\hat{L} := \{x \in \mathcal{P}^0; K(x, \mu) \leq K(x^0, \mu)\},$$

es un subconjunto del producto cartesiano de los intervalos  $[a_j, b_j]$ ,  $0 \leq j \leq n$ . De aquí note que  $\hat{L}$  está acotado. Note además que  $\hat{L}$  es no vacío, porque  $x^0 \in \hat{L}$ . Como  $K(x, \mu)$  es continua, entonces tiene un minimizador en  $\hat{L}$  y consecuentemente también en  $\mathcal{P}^0$ . Esto completa la prueba. ■

Para cada  $\mu > 0$ , denotamos al minimizador del problema  $(P_\mu)$  como  $x(\mu)$  y al maximizador de  $\beta_D$  como  $(y(\mu), s(\mu))$ .

El siguiente lema, tomado de Wright [19], nos muestra como calcular la *brecha dual*, para cada valor de  $\mu > 0$ .

**Lema 2.5:** La brecha dual  $c^T x(\mu) - b^T y(\mu)$  satisface

$$c^T x(\mu) - b^T y(\mu) = n\mu$$

**Prueba:**

$$\begin{aligned} c^T x(\mu) - b^T y(\mu) &= c^T x(\mu) - (Ax(\mu))^T y(\mu) \\ &= (c - A^T y(\mu))^T x(\mu) \\ &= s^T(\mu) x(\mu) \\ &= n\mu \quad \blacksquare \end{aligned}$$

En conclusión toda solución  $(x(\mu), y(\mu), s(\mu))$  que satisface el sistema (2.2) y en consecuencia de ello, por el *lema 2.2*, es una solución óptima del problema.

$$(P_\mu) \quad \min \{\beta(x, \mu) : x \in \mathcal{F}^0\}$$

define un par primal-dual factible. Además la brecha dual es dada por  $n\mu$  (lema 2.5). Esto implica que una solución óptima del problema  $(P_\mu)$ , para  $\mu$  suficientemente pequeño es una solución aproximado para el problema  $(P)$ , porque usando el teorema de dualidad, tenemos

$$c^T x(\mu) - c^T x^* \leq c^T x(\mu) - b^T y(\mu) \leq n\mu,$$

donde  $x^*$  es la solución óptima del problema  $(P)$ .

Para los minimizadores del problema  $(P_\mu)$  y a los maximizadores de  $\beta_D$  se tienen las siguientes definiciones. Según Wright [19] se define:

**Definición 2.4:** El punto  $x = x(\mu)$  que resuelve el problema  $(P_\mu)$  y donde la función objetivo del problema  $(P)$  tiene una brecha dual igual a  $n\mu$  es llamado un **centro asociado al valor  $\mu$  o el  $\mu$ -centro**.



Gráficamente:

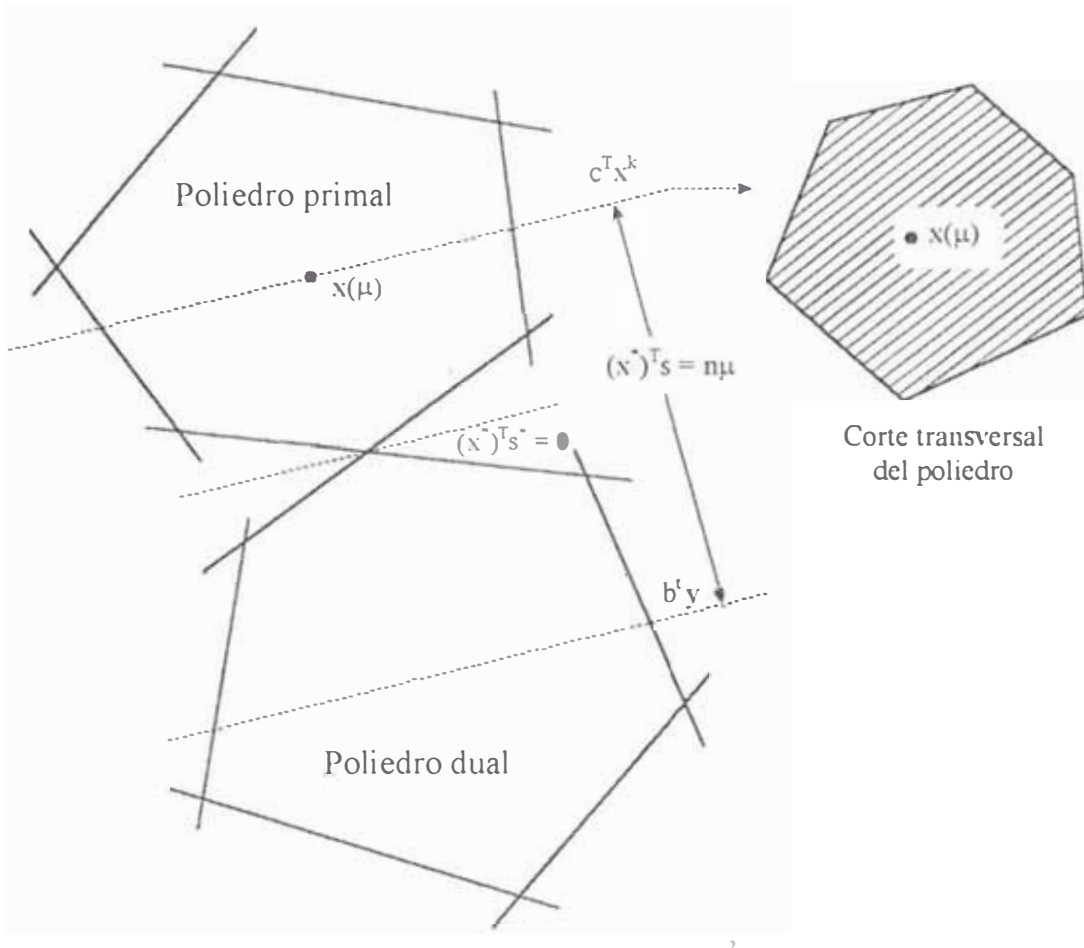


Figura 2.1: Centro asociado a  $\mu$

En otras palabras, a cada valor de  $\mu > 0$  le corresponde un centro  $x = x(\mu)$  y solo uno.

Recíprocamente, un centro  $x(\mu)$  está completamente determinado por el valor de  $\mu$ , es decir

$$\mu \rightarrow x(\mu) \text{ es una biyección.}$$

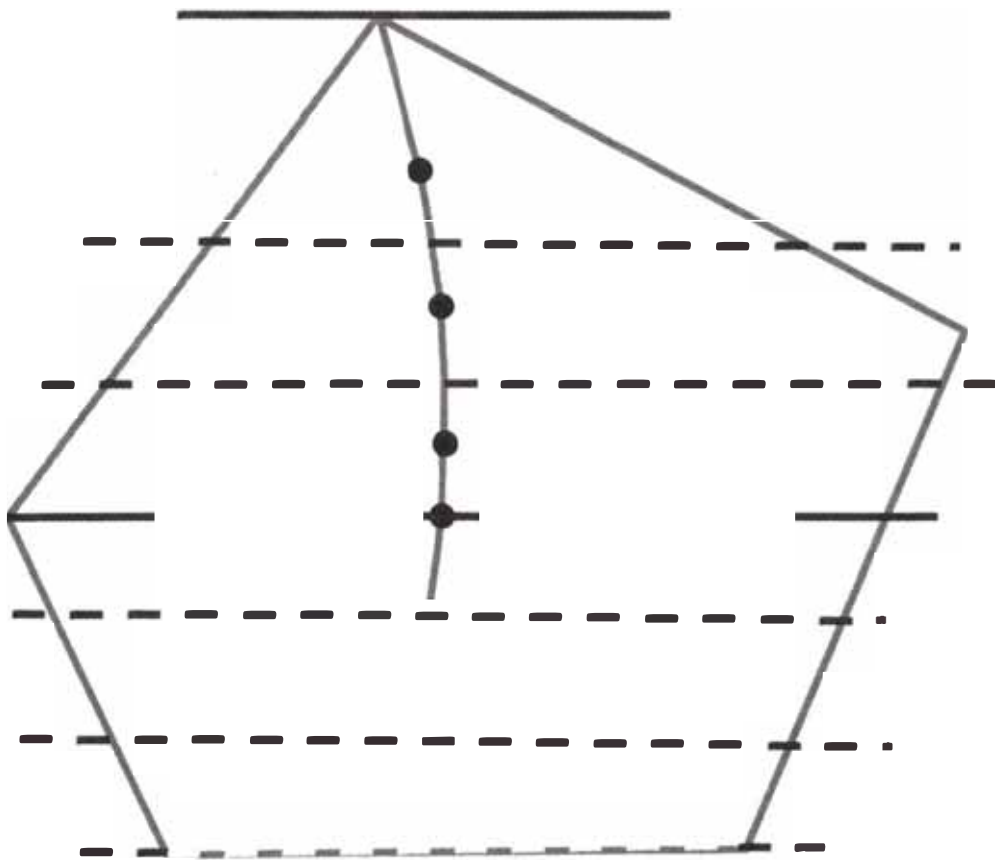
Según Wright [19] se define:

**Definición 2.5:** El conjunto de centros

$$\Gamma = \{x(\mu) / \mu > 0\} \subset \mathcal{P}^0,$$

forman una curva en  $\mathcal{P}^0$ , llamada **trayectoria central**

Gráficamente:



**Fig. 2.2:** Trayectoria central

De igual forma, si denotamos al maximizador de  $\beta_D(s, \mu)$  por  $(y(\mu), s(\mu))$  para cada  $\mu > 0$ . El conjunto  $\{y(\mu) : \mu > 0\}$  o  $\{s(\mu) : \mu > 0\}$  es llamado la trayectoria central de (D).

**Teorema 2.6:** Si ambos problemas primal (P) y dual (D) tienen puntos interiores factibles para un conjunto dado  $(A, b, c)$ . Entonces para cada  $0 < \mu < \infty$ , la trayectoria central  $(x(\mu), y(\mu), s(\mu))$  existe y es único.

**Prueba:**

La prueba de este teorema se obtiene de manera inmediata teniendo en cuenta la convexidad estricta de los problemas de barrera

$$\begin{array}{ll} \min c^T x - \mu \sum_{j=1}^n \ln x_j & \max b^T y + \mu \sum_{j=1}^n \ln s_j \\ \text{sujeto a } Ax = b, x \geq 0 & \text{sujeto a } A^T y + s = c, s > 0 \end{array}$$

Los minimizadores  $x(\mu) \in \mathcal{P}^0$  y  $(y(\mu), s(\mu)) \in \mathcal{D}^0$  de estos problemas existen y son únicos. ■

**Lema 2.7:** Si  $\mu$  decrece entonces el valor objetivo  $c^T x(\mu)$  del problema primal (P) decrece monótonamente y el valor objetivo del problema dual (D) crece monótonamente.

**Prueba:** Sea  $\bar{\mu} < \mu$ , como  $x(\mu)$  minimiza  $\beta(x, \mu)$  y  $x(\bar{\mu})$  minimiza  $\beta(x, \bar{\mu})$  obviamente tenemos

$$\begin{array}{l} \beta(x(\mu), \mu) < \beta(x(\bar{\mu}), \mu), \\ \text{y} \quad \beta(x(\mu), \bar{\mu}) < \beta(x(\mu), \mu), \end{array}$$

lo cual es equivalente a:

$$c^T x(\mu) - \mu \sum_{j=1}^n \ln x_j(\mu) \leq \mu \sum_{j=1}^n \ln x_j(\bar{\mu}),$$

y

$$c^T x(\bar{\mu}) - \bar{\mu} \sum_{j=1}^n \ln x_j(\bar{\mu}) \leq c^T x(\mu) - \bar{\mu} \sum_{j=1}^n \ln x_j(\mu),$$

o equivalente.

$$\frac{c^T x(\mu)}{\mu} - \sum_{j=1}^n \ln x_j(\mu) \leq \frac{c^T x(\bar{\mu})}{\mu} - \sum_{j=1}^n \ln x_j(\bar{\mu})$$

$$\frac{c^T x(\bar{\mu})}{\bar{\mu}} - \sum_{j=1}^n \ln x_j(\bar{\mu}) \leq \frac{c^T x(\bar{\mu})}{\mu} - \sum_{j=1}^n \ln x_j(\mu)$$

sumando estas dos desigualdades tenemos:

$$\frac{c^T x(\mu)}{\mu} + \frac{c^T x(\bar{\mu})}{\bar{\mu}} \leq \frac{c^T x(\bar{\mu})}{\mu} + \frac{c^T x(\bar{\mu})}{\bar{\mu}},$$

reemplazando tenemos:

$$\left( \frac{1}{\mu} - \frac{1}{\bar{\mu}} \right) (c^T x(\mu) - c^T x(\bar{\mu})) \leq 0,$$

y como  $\mu < \bar{\mu}$  entonces  $c^T x(\bar{\mu}) \leq c^T x(\mu)$ .

La segunda parte del lema se puede probar en forma similar, teniendo en cuenta que  $y(\mu)$  es el único punto que maximiza la función de barrera logarítmica dual.

$$\beta_0(y, \mu) = b^T y + \mu \sum_{j=1}^n \ln s_j \quad \blacksquare$$

Recordemos de los *Teoremas 1.2 y 1.3* que si  $(x, s) \in \mathcal{P} \times \mathcal{D}$ , entonces  $x^T s \geq 0$ , y que si  $x^T s = 0$  entonces  $x$  es optimal para  $(\mathcal{P})$  y  $s$  es optimal para  $(\mathcal{D})$ .

Procedemos ahora a probar la existencia de una *solución complementaria*  $(x^*, s^*) \in \mathcal{P} \times \mathcal{D}$ , esto es una solución satisfaciendo  $(x^*)^T s^* = 0$ . Necesitamos el siguiente lema adaptado de Güler [7].

**Lema 2.8:** Sea  $\bar{\mu} > 0$ . El conjunto

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

está acotado.

**Prueba:** Desde que  $x^0 - x(\mu) \in N(A)$  y  $s^0 - s(\mu) \in R(A^t)$ , tenemos  $(x^0 - x(\mu))^t (s^0 - s(\mu)) = 0$ .

Para cada  $j$ ,  $1 < j < n$ , esto implica

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

Esto prueba que  $x_j(\mu) < (n\bar{\mu} + (x^0)^t s^0) / s_j^0$ , así que  $\{x(\mu)\}_{0 < \mu \leq \bar{\mu}}$  está acotado. La prueba para  $\{s(\mu)\}_{0 < \mu \leq \bar{\mu}}$  es similar. ■

El siguiente teorema es tomado de Güler [7].

**Teorema 2.9:** Si  $x$  y  $s$  son factibles para (P) y (D), respectivamente. Entonces los puntos  $x$  y  $s$  son soluciones optimales si y solo si  $x^T s = 0$ , además, ambos problemas tienen soluciones óptimas.

**Prueba:** En virtud del teorema 1.3, es suficiente probar que existe soluciones optimales  $x^*$  para (P) y  $(y^*, s^*)$  para (D) tal que la brecha dual  $(x)^T s$  es cero.

Sea  $\{\mu_k\}$  una sucesión de parámetros de barreras, tales que  $\{\mu_k\} \rightarrow 0$ .

Dado que el conjunto  $\{(x(\mu_k), s(\mu_k))\}$  está acotado, entonces contienen una subsucesión convergente a el punto  $(x^*, s^*)$ , tenemos que  $(x^*, s^*) \in \mathcal{P} \times \mathcal{D}$  y como

$$x(\mu_k)^t s(\mu_k) = n\mu_k \rightarrow 0 \quad (x^*)^T s^* = 0 \quad \blacksquare$$

Un resultado en la teoría de la programación lineal, conocido como el teorema de Goldman - Tucker, es cual lo establecimos en el capítulo 1, la existencia de una *solución optimal primal dual estrictamente complementaria*.

Esto es una solución óptima  $(x^*, s^*)$  satisfaciendo  $x^* + s^* > 0$ . Como describiremos posteriormente, hay una rica estructura asociada con estas soluciones para esto necesitamos introducir algunas notaciones. Denotemos por  $z^*$  el valor óptimo común de los problemas (P) y (D). Los conjuntos de soluciones óptimas para (P) y (D) serán denotadas, como en el capítulo 1, por  $\mathcal{P}^*$  y  $\mathcal{D}^*$  respectivamente, tenemos:

$$\mathcal{P}^* = \{x : Ax = b, c^T x = z^*, x > 0\}$$

$$\mathcal{D}^* = \{(y, s) : A^T y + s = c, b^T y = z^*, y \in \mathbb{R}^m, s \geq 0\}.$$

Según Güler [7] se define:

**Definición 2.6:** Sea  $x \in \mathbb{R}^n$ , el *soporte* de  $x$ , denotado por  $\sigma(x)$ , es el conjunto

$$\sigma(x) = \{i / x_i > 0\}$$

Sea  $(x, y, s) \in \mathcal{P} \times \mathcal{D}$ . Como una consecuencia al teorema de dualidad (teorema 2.9) tenemos  $\sigma(x) \cap \sigma(s) = \emptyset$  si y solo si  $(x, s)$  un par primal-dual óptimo. Nótese que  $(x, s)$  es un par primal-dual óptimo estrictamente complementario si y solo si  $\sigma(x) \cup \sigma(s) = \{1, 2, \dots, n\}$ . Ahora presentamos un teorema que garantiza la existencia de tal par, bajo la *hipótesis 2.1*.

El siguiente teorema es tomado de Güler [7].

**Teorema 2.10:** Existen puntos  $x^* \in \mathcal{P}^*$  y  $s^* \in \mathcal{D}^*$ , tal que  $(x^*, s^*)$  es un par óptimo estrictamente complementario.

**Prueba:** Sea  $x^*$  y  $s^*$  definidas en la prueba del *Teorema 2.9*. Entonces tenemos una sucesión  $\{(x(\mu_k), s(\mu_k))\}$ , con  $\mu_k \rightarrow 0$ , la cual converge a  $(x^*, s^*)$ . Afirmamos que  $(x^*, s^*)$  es un par estrictamente complementario. La prueba se inicia como sigue:

Desde que  $x(\mu_k) - x^* \in N(A)$  y  $s(\mu_k) - s^* \in R(A^T)$ , tenemos:

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

Reordenando los términos de esta ecuación, y notando que  $x(\mu_k)^T s(\mu_k) = n\mu_k$ ,  $(x^*)^T s^* = 0$ , llegamos a:

$$\sum_{j \in \sigma(x^*)} x_j^* s_j(\mu_k) + \sum_{j \in \sigma(s^*)} x_j(\mu_k) s_j^* = n \mu_k.$$

Dividiendo ambos lados por  $\mu_k$  y teniendo en cuenta que  $x_j(\mu_k) s_j(\mu_k) = \mu_k$ , obtenemos:

$$\sum_{j \in \sigma(x^*)} \frac{x_j^*}{x_j(\mu_k)} + \sum_{j \in \sigma(s^*)} \frac{s_j^*}{s_j(\mu_k)} = n.$$

Haciendo  $k \rightarrow \infty$ , la primera suma se hace igual al número de coordenadas no nulas en  $x^*$ . Similarmente, la segunda suma se hace igual al número de coordenadas no nulas en  $s^*$ . Concluimos que el par  $(x^*, s^*)$  es estrictamente complementario. ■

Con  $x$  y  $s$  en el *teorema 2.10*, particionamos al conjunto índice  $\{1, 2, \dots, n\}$  por

$$\{1, 2, \dots, n\} = B \cup N$$

con  $B := \{i : x_i^* > 0\}$ , (2.4)

$$N := \{i : s_i^* > 0\}.$$

El siguiente teorema es consecuencia directa del anterior.

**Teorema 2.11:** La partición dada en (2.4) es única. En otras palabras, cada par optimal estrictamente complementario  $(x, s)$  tiene el mismo soporte:

$$\sigma(x) = B \text{ y } \sigma(s) = N$$

**Prueba:** Como una consecuencia del teorema 2.9,  $x \in \mathcal{P}$  es optimal si y solo si  $x^T s^* = 0$ .

Similarmente,  $s \in \mathcal{D}$  es optimal si y solo si  $(x^*)^T s = 0$ .

De aquí se sigue que

$$x \in \mathcal{P}^* \Leftrightarrow \sigma(x) \subseteq B,$$

$$s \in \mathcal{D}^* \Leftrightarrow \sigma(s) \subseteq N. \blacksquare$$

La partición (2.4) del conjunto de índices  $\{1, 2, \dots, n\}$  será llamado la *partición optimal* del problema (P) y del problema (D).

Usaremos la notación  $x_B$  y  $x_N$  para referirnos a la restricción de un vector  $x \in \mathbb{R}^n$  a los conjuntos de coordenadas B y N respectivamente. Las submatrices  $A_B$  y  $A_N$  de la matriz A son definidas de igual forma.

Como una consecuencia del teorema 2.12, tenemos:

$$\mathcal{P} = \{x \geq 0 : A_B x_B = b, x_N = 0\},$$

$$\mathcal{D}^* = \{(y, s) : A^T y + s = c, y \in \mathbb{R}^m, s \geq 0, s_B = 0\}.$$

El siguiente corolario, se obtiene inmediatamente de las descripciones de  $\mathcal{P}^*$  y  $\mathcal{D}^*$  dadas arriba.

**Corolario 2.12:** El conjunto de soluciones óptimas estrictamente complementarias coincide con el interior relativo de las facetas óptimas  $\mathcal{P}^*$  y  $\mathcal{D}^*$ .

**Prueba:** Consecuencia directa de las descripciones de  $\mathcal{P}^*$  y  $\mathcal{D}^*$  ■

Una observación interesante nos lleva ahora a la noción de solución central de un problema de programación lineal. Cada solución óptima  $x$  de (P) satisface  $x_N = 0$ , y existe al menos una solución óptima tal que  $x_B > 0$ . Debido al acotamiento de la faceta óptima el siguiente problema de maximización.

$$\max \left\{ \sum_{i \in B} \ln x_i : x \in \mathcal{P}^* \right\} \quad (2.5)$$

tiene una solución óptima: Esto es debido a que la función objetivo es estrictamente convexa, luego la solución óptima es única. Este es el llamado centro analítico de  $\mathcal{P}^*$ . Nos referimos a esta solución óptima de (P) como la solución central de (P). La solución central de (D) es definida similarmente, es decir es la única solución del problema de optimización.

$$\max \left\{ \sum_{i \in N} \ln s_i : s \in \mathcal{D}^* \right\} \quad (2.6)$$



Concluimos esta sección probando que las soluciones centrales son los límites de las trayectorias centrales de (P) y (D).

El siguiente lema es debido a Mc linder [11].

**Lema 2.13:** Las trayectorias centrales de (P) y (D) convergen a los centros analíticos de  $\mathcal{P}^*$  y  $\mathcal{D}^*$ .

**Prueba:** En la prueba del teorema 2.10 exhibimos una sucesión decreciente  $\{\mu_k\}_{k=1}^{\infty}$  con la propiedad de que las sucesiones asociadas  $\{x(\mu_k)\}$  y  $\{s(\mu_k)\}$  sobre las trayectorias centrales de (P) y (D) convergen a los puntos límites  $x^*$  en  $\mathcal{P}^*$  y  $s^*$  en  $\mathcal{D}^*$ . Ahora sea  $x^0$  y  $s^0$  puntos arbitrarios en  $\mathcal{P}^*$  y  $\mathcal{D}^*$ . Entonces, usando  $(x^0 - x(\mu_k))^T (s^0 - s(\mu_k)) = 0$ , obtenemos  $(x^0)^T s(\mu_k) + (s^0)^T x(\mu_k) = n\mu_k$ , (ya que  $(s^0)^T x^0 = 0$ ). Usando  $\sigma(x^0) \subseteq B$  y  $\sigma(s^0) \subseteq N$ , tenemos

$$\sum_{i \in B} x_i^0 s_i(\mu_k) + \sum_{i \in N} x_i(\mu_k) s_i^0 = n\mu_k$$

Dividiendo ambos lados por  $\mu_k$  y usando que  $x_i(\mu_k) s_i(\mu_k) = \mu_k$ , obtenemos

$$\sum_{i \in B} \frac{x_i^0}{x_i(\mu_k)} + \sum_{i \in N} \frac{s_i^0}{s_i(\mu_k)} = n$$

Haciendo  $k \rightarrow \infty$ , esto implica

$$\sum_{i \in B} \frac{x_i^0}{x_i^*} + \sum_{i \in N} \frac{s_i^0}{s_i^*} = n$$

si escogemos  $s^0 = s^*$ , vemos que

$$\sum_{i \in B} \frac{x_i^0}{x_i^*} = |B|,$$

donde  $|B|$  es la cardinalidad de  $B$ . Reescribiendo esto, obtenemos

$$c_B^T (X_B^*)^{-1} (x_B^0 - x_B^*) = 0, \forall x^0 \in \mathcal{P}^*$$

Ahora, usando el *Teorema 2.1* observamos que este es la condición necesaria y suficiente para que  $x_B^*$  sea una solución optimal dual problema de optimización (2.5). De este modo  $x_B^*$  debe ser la solución central de (P). Pero el argumento de esta demostración, deja ver claramente que hay solo uno de tales puntos limites, precisamente la solución central de (P). Esto prueba el teorema para el problema primal. Una prueba similar puede ser dada para el problema dual ■.

### 2.3 RELAJACIÓN DE LA HIPÓTESIS DE PUNTO INTERIOR

En esta sección, renovemos la *hipótesis 2.1*, para esto escogamos vectores positivos arbitrarios  $x^0, s^0 \in \mathbb{R}^n$  y un vector arbitrario  $y \in \mathbb{R}^m$ . También escogemos números 'grandes'  $M_p$  y  $M_d$  tal que

$$M_p > (b - A x^0)^T y^0, \quad M_d > (A^T y^0 + s^0 - c)^T x^0$$

Por el momento llamaremos a cada par de parámetros  $(M_p, M_d)$  factible.

Consideremos el par de problemas  $(\tilde{P})$  y  $(\tilde{D})$

$$\begin{aligned}
 (P) \quad & \min \quad c^T x + M_p x_{n+1} \\
 & \text{sujeto a} \quad Ax + (b - Ax^0)x_{n+1} = M_d \\
 & \quad \quad \quad (A^T y^0 + s^0 - c)^T x + x_{n+2} = M_d \\
 & \quad \quad \quad x > 0, x_{n+1} > 0, x_{n+2} > 0,
 \end{aligned}$$

$$\begin{aligned}
 (D) \quad & \min \quad b^T y + M_d x_{m+1} \\
 & \text{sujeto a} \quad A^T y + (A^T y^0 - s^0 - c)y_{m+1} + s = c, \\
 & \quad \quad \quad (b - Ax^0)^T y - s_{n+1} = M_p, \\
 & \quad \quad \quad y_{m+1} - s_{n+2} = 0, \\
 & \quad \quad \quad s > 0, s_{n+1} > 0, s_{n+2} > 0.
 \end{aligned}$$

Sea:

$\tilde{x} = (x, x_{n+1}, x_{n+2})$ ,  $\tilde{s} = (s, s_{n+1}, s_{n+2})$  con  $x, s \in \mathbb{R}^n$  y  $\tilde{y} = (y, y_{m+1})$ ,  $y \in \mathbb{R}^m$  también definamos la matriz  $\tilde{A}$  de orden  $(m+1) \times (n+2)$  y los vectores  $\tilde{b} \in \mathbb{R}^{m+1}$ ,  $\tilde{c} \in \mathbb{R}^{n+2}$

$$\tilde{b} = \begin{pmatrix} b \\ M_d \end{pmatrix}, \quad \tilde{c} = \begin{pmatrix} c \\ M_p \\ 0 \end{pmatrix}, \quad \tilde{A} = \begin{pmatrix} A & b - Ax^0 & 0 \\ (A^T y^0 + s^0 - c)^T & 0 & 1 \end{pmatrix}$$

Podemos escribir  $(\tilde{P})$  y  $(\tilde{D})$  en forma compacta como

$$(\tilde{P}) \quad \min \left\{ \tilde{c}^T \tilde{x} : \tilde{A} \tilde{x} = \tilde{b}, \tilde{x} \geq 0 \right\}$$

$$(\tilde{D}) \quad \max \left\{ \tilde{b}^T \tilde{y} : \tilde{A}^T \tilde{y} + \tilde{s} = \tilde{c}, \tilde{s} \geq 0 \right\}$$

Es fácil ver que el par  $(\tilde{P})$ - $(\tilde{D})$  satisface la *hipótesis 2.1*, considerando el punto  $(\tilde{x}^0, \tilde{y}^0, \tilde{s}^0)$ :

$$\tilde{x}^0 = (x^0, 1, x_{n+2}^0) \in \mathbb{R}^{n+2}$$

$$\tilde{s}^0 = (s^0, s_{n+1}^0, 1) \in \mathbb{R}^{n+2}$$

$$\tilde{y}^0 = (y^0, -1) \in \mathbb{R}^{m+1}$$

Con  $x_{n+2}^0 = M_d - (A^T y^0 + s^0 - c)^T x^0$ , y  $s_{n+1}^0 = M_p - (b - Ax^0)^T y^0$ . Tenemos  $\tilde{A} \tilde{x}^0 = \tilde{b}$  y  $\tilde{A}^T \tilde{y}^0 + \tilde{s}^0 = \tilde{c}$ . Desde que  $\tilde{x}^0$  y  $\tilde{s}^0$  son positivos (debido a  $M_p > (b - Ax^0)^T y^0$ ,  $M_d > (A^T y^0 + s^0 - c)^T x^0$ ) la *hipótesis 2.1* es satisfecha.

El par de programas aumentado  $(\tilde{P})$ - $(\tilde{D})$  ha sido usado en [6, 9, 10] para inicializar los algoritmos de punto interior.

# CAPITULO 3

## ALGORITMO DE SEGUIMIENTO DE TRAYECTORIA

En este capítulo presentamos un algoritmo de seguimiento de trayectoria de paso corto, el cual resulta de la aplicación del método de barrera logarítmica al problema de la programación lineal, una de las razones por la cual nos restringimos a este tipo de algoritmos es debido a que cierta variante de el (el algoritmo primal-dual-predictor-corrector) ha demostrado ser particularmente eficiente.

Iniciamos la presentación de nuestro algoritmo como un método de barrera, para luego mediante una sucesión de lemas llegar a el teorema central que establece su convergencia polinomial.

Este capítulo ha sido escrito siguiendo el artículo de Roos – Vial [15] y el libro de D. den Hertog [4].

### 3.1 CONSTRUCCIÓN DE UN ALGORITMO DE SEGUIMIENTO DE TRAYECTORIA

En este capítulo, consideraremos el problema de la programación lineal en su forma standard:

$$(P) \quad \begin{cases} \min c^T x \\ \text{sujeto a } Ax = b, \\ x \geq 0; \end{cases}$$

con las siguientes hipótesis

H1) El interior de la región factible

$$\mathcal{P}^0 = \{x \in \mathbb{R}^n / Ax = b, x > 0\} \text{ es no vacío}$$

H2) El conjunto factible  $\mathcal{P} = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$  es acotado.

H3) A tiene rango  $m < n$ .

donde  $X^{-1} = \text{diag} \left( \frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_n} \right)$ ,  $e = (1, 1, \dots, 1)^T$ , y el vector de multiplicadores de Lagrange es denotado por  $y$ .

Introduciendo la variable de holgura  $s = \mu X^{-1}e$ , podemos reescribir (3.1) como

$$\begin{cases} A^T y + s - c = 0, \\ Ax - b = 0, \\ \mu e - Xs = 0. \end{cases} \quad (3.2)$$

De la tercera ecuación de (3.2),  $Xs = \mu e$  tenemos que  $x_j s_j = \mu$ . Ya que  $x_j$  es positivo,  $s_j$  es también positivo, y luego  $s > 0$ . Luego, el multiplicador de Lagrange  $y$  corresponde a la solución dual factible  $(y, s)$ . Además, cuando  $\mu$  es pequeño, la condición de holguras complementarias están cerca de ser satisfechas.

Asumamos que  $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)^T > 0$ ,  $\bar{s} = (\bar{s}_1, \bar{s}_2, \dots, \bar{s}_n)^T > 0$  y  $\bar{y} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n)^T$  son dados, con  $A\bar{x} = b$  y  $A^T\bar{y} + \bar{s} = c$ . Nuestro objetivo es resolver (3.2) vía el método de Newton. Si definimos

$$g(x_i, s_i) = \frac{\mu}{x_i} - s_i,$$

vemos que esta es la única expresión no lineal del sistema (3.2), solo necesitamos linealizar esta cuando aplicamos el método de Newton. Para esto consideramos

$$g(x_i, s_i) = 0$$

y entonces

$$0 - g(\bar{x}_i, \bar{s}_i) = \nabla g(\bar{x}_i, \bar{s}_i) \begin{pmatrix} x_i - \bar{x}_i \\ s_i - \bar{s}_i \end{pmatrix},$$

sustituyendo la expresión que define  $g$  obtenemos

$$\bar{s}_i - \frac{\mu}{\bar{x}_i} = \left( -\frac{\mu}{\bar{x}_i^2}, -1 \right) \begin{pmatrix} x_i - \bar{x}_i \\ s_i - \bar{s}_i \end{pmatrix}$$

de aquí

$$s_i = \frac{2\mu}{\bar{x}_i} - \frac{\mu}{\bar{x}_i^2} x_i. \quad (3.3)$$

Haciendo  $X = \text{diag}(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$  notamos que (3.3) se escribe matricialmente como

$$s = 2\mu\bar{X}^{-2}\bar{x} - \mu\bar{X}^{-2}x.$$

Usando  $A^T y + s = c$ , vemos que

$$x = \frac{1}{\mu} \bar{X}^2 \left[ A^T y + 2\mu\bar{X}^{-2}\bar{x} - c \right]. \quad (3.4)$$

Multiplicando ambos lados de (3.4) por la matriz  $A$  obtenemos

$$b = Ax = \frac{1}{\mu} A\bar{X}^2 \left[ A^T y + 2\mu\bar{X}^{-2}\bar{x} - c \right].$$

Consecuentemente,

$$y = \left( A\bar{X}^2 A^T \right)^{-1} \left( A\bar{X}^2 c - \mu b \right), \quad (3.5)$$

colocando (3.5) en (3.4) obtenemos

$$\Delta x = x - \bar{x} = -\frac{1}{\mu} \bar{X} \left[ I - \bar{X} A^T (A\bar{X}^2 A^T)^{-1} A \bar{X} \right] \bar{X} c + \bar{X} \left[ I - \bar{X} A^T (A\bar{X}^2 A^T)^{-1} A \bar{X} \right] e,$$

es decir

$$\Delta x = -\frac{1}{\mu} \bar{X} \left[ I - \bar{X} A^T (A\bar{X}^2 A^T)^{-1} A \bar{X} \right] \bar{X} c + \bar{X} \left[ I - \bar{X} A^T (A\bar{X}^2 A^T)^{-1} A \bar{X} \right] e,$$

la cual es la dirección de búsqueda que se usará en el algoritmo de seguimiento de trayectoria primal que desarrollaremos. Un nuevo punto interior factible es dado por  $\bar{x} + \alpha \Delta x$  con un apropiado  $\alpha \in (0, 1]$ .

### Ejemplo 3.1: (Cálculo de la dirección generada por el algoritmo)

Consideremos el problema lineal

$$\begin{aligned} \text{Min } z &= -x_1 - 2x_2 \\ \text{Sujeto a } &-2x_1 + x_2 \leq 2 \\ &-x_1 + 2x_2 \leq 7 \\ &x_1 \leq 3 \\ &x_1, x_2 \geq 0 \end{aligned}$$

Después de añadir variables de holgura y poner el problema en forma standard, tenemos que la matriz de restricciones y al vector de costos son:

$$A = \begin{pmatrix} -2 & 1 & 1 & 0 & 0 \\ -1 & 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \text{ y } c = (-1 \quad -2 \quad 0 \quad 0 \quad 0)^T$$

Calculamos la dirección generada por el algoritmo con punto inicial

$$x = x_0 = (0.5 \quad 0.5 \quad 2.5 \quad 6.5 \quad 2.5)^T$$

usamos  $\mu = 10$ . Entonces

$$c - \mu X^{-1}e = \begin{pmatrix} -1 \\ -2 \\ 0 \\ 0 \\ 0 \end{pmatrix} - 10 \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.1539 & 0 \\ 0 & 0 & 0 & 0 & 0.4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -21.000 \\ -22.000 \\ -4.000 \\ -1.539 \\ -4.000 \end{pmatrix}$$

ya que

$$AX = \begin{pmatrix} -1 & 0.5 & 2.5 & 0 & 0 \\ -0.5 & 1 & 0 & 6.5 & 0 \\ 0.5 & 0 & 0 & 0 & 2.5 \end{pmatrix} \text{ y } AX^2A^T = \begin{pmatrix} 7.50 & 1.00 & -0.50 \\ 1.00 & 43.50 & -0.25 \\ -0.50 & -0.25 & 6.50 \end{pmatrix},$$

Obtenemos

$$y = (AX^2A^T)^{-1}AX^2(c - \mu X^{-1}e) = \begin{pmatrix} -2.7832 \\ -1.5908 \\ -4.9291 \end{pmatrix}, \quad s = c - A^T y = \begin{pmatrix} -3.2280 \\ 3.9647 \\ 2.7832 \\ 1.5908 \\ 4.9291 \end{pmatrix},$$

y

$$\Delta x = x - \frac{1}{\mu} X^2 s = \begin{pmatrix} 0.5807 \\ 0.4009 \\ 0.7605 \\ -0.2211 \\ -0.5807 \end{pmatrix}.$$

Ahora veamos otra forma de obtener esta misma dirección de búsqueda, que será importante cuando hagamos el estudio de los algoritmos generalizados de seguimiento de trayectoria en el capítulo 4. Antes de hacer esto, damos dos resultados simples del álgebra lineal.

**Lema 3.1:** Si  $A \in \mathbb{R}^{m \times n}$   $m < n$ , con rango máximo, y  $U \in \mathbb{R}^{n \times (n-m)}$  de rango máximo tal que  $AU = 0$ , entonces

$$(I - A^T(AA^T)^{-1}A)x = U(U^T U)^{-1}U^T x, \quad \forall x \in \mathbb{R}^n$$

**Prueba:** Desde que  $A$  tiene rango máximo, cada  $x \in \mathbb{R}^n$  puede ser descompuesto como

$$x = A^T u_1 + U u_2 \quad \text{donde } u_1 \in \mathbb{R}^m, \text{ y } u_2 \in \mathbb{R}^{n-m}$$

Entonces

$$Ax = AA^T u_1 + AU u_2 = AA^T u_1.$$

Luego

$$u_1 = (AA^T)^{-1} Ax$$

Similarmente,

$$U^T x = U^T A^T u_1 + U^T U u_2 = U^T U u_2.$$

De este modo,

$$u_2 = (U^T U)^{-1} U^T x.$$



Por tanto,

$$x = A^T(AA^T)^{-1}Ax + U(U^T U)^{-1}U^T x \text{ para } x \in \mathbb{R}^n$$

En otras palabras,

$$(I - A^T(AA^T)^{-1}A)x = U(U^T U)^{-1}U^T x \quad \forall x \in \mathbb{R}^n \quad \blacksquare$$

Note que el operador  $P = U(U^T U)^{-1}U^T$  satisface  $P^2 = P$  y  $AP = 0$ . De este modo, es una proyección ortogonal sobre el espacio nulo de  $A$ .

Ya que hemos asumido que la matriz  $A$  es de rango máximo, el espacio nulo de  $A$  es un subespacio de  $\mathbb{R}^n$  de dimensión  $n-m$ , el cual es isomórfico a  $\mathbb{R}^{n-m}$ . La matriz  $U$  en el lema juega el rol de isomorfismo entre el espacio nulo de  $A$  y  $\mathbb{R}^{n-m}$ .

**Lema 3.2:** Sea  $Q \in \mathbb{R}^{n \times n}$  una matriz simétrica definida positiva, y sea  $A$  y  $U$  definidas como en el lema 3.1. Entonces

$$Q(I - QA^T(AQ^2A^T)^{-1}AQ)Qx = U(U^T Q^{-2}U)^{-1}U^T x$$

**Prueba:** Sea  $\hat{A} = AQ$  y  $\hat{U} = Q^{-1}U$ .

Como  $\hat{A}\hat{U} = 0$ , el resultado se obtiene cuando reemplazamos  $A$ ,  $U$  y  $x$  en el lema 3.1 por  $\hat{A}$ ,  $\hat{U}$  y  $Qx$  respectivamente.  $\blacksquare$

Denotando la función objetivo  $c^T x - \mu \sum_{i=1}^n \ln x_i$  en el problema  $(P_\mu)$  por  $\beta(x, \mu)$ , notemos que

es estrictamente convexa y continuamente diferenciable en el interior del dominio factible. En particular, para una solución factible interior dada  $\bar{x}$ , tenemos

$$\beta(x, \mu) - \beta(\bar{x}, \mu) = \nabla \beta(\bar{x}, \mu)^T (x - \bar{x}) + o(\|x - \bar{x}\|),$$

donde

$$\nabla \beta(\bar{x}, \mu) = c - \mu \bar{X}^{-1} e.$$

Encontrar la dirección de descenso más rápida en  $x$  es equivalente a minimizar  $\nabla \beta(\bar{x}, \mu)^T (x - \bar{x})$ .

Entonces definimos un subproblema  $(P_s)$  de  $(P_\mu)$  como sigue:

$$(P_s) \begin{cases} \min & \nabla\beta(\bar{x}, \mu)^T (x - \bar{x}) \\ \text{sujeto a} & A(x - \bar{x}) = 0, \\ & \|Q^{-1}(x - \bar{x})\|^2 \leq \beta^2 < 1. \end{cases}$$

Donde  $Q^{-1}$  podrá ser cualquier matriz simétrica definida positiva y  $\beta$  es escogido tal que la superficie del elipsoide  $\|Q^{-1}(x - \bar{x})\| = \beta$  se mantenga inscrita en la región factible de  $(P_\mu)$ . Los ejes principales del elipsoide son los vectores propios de  $Q^{-1}$ . En particular si escogemos  $Q^{-1}$

como  $\bar{X}^{-1} = \text{diag}\left(\frac{1}{\bar{x}_1}, \frac{1}{\bar{x}_2}, \dots, \frac{1}{\bar{x}_n}\right)$ , entonces  $(P_s)$  puede ser establecido como una versión

del espacio nulo, por usar el isomorfismo  $U$  entre el espacio nulo de  $A$  y  $\mathbb{R}^{n-m}$ . Reemplazando  $\Delta x = x - \bar{x}$  por  $h \in \mathbb{R}^{n-m}$ , el subproblema se convierte en:

$$\begin{aligned} \min & \nabla\beta(\bar{x}, \mu)^T U h \\ \text{sujeto a} & \|\bar{X}^{-1} U h\|^2 \leq \beta^2, \\ & h \in \mathbb{R}^{n-m}. \end{aligned}$$

Considerando el lagrangiano  $L_1(h, \lambda) = \nabla\beta(\bar{x}, \mu)^T U h + \lambda (\|\bar{X}^{-1} U h\|^2 - \beta^2)$ , donde  $\lambda \geq 0$  es el vector de multiplicadores de Lagrange.

Calculando  $\frac{\partial L_1}{\partial h} = 0$ , tenemos  $U^T \nabla\beta(\bar{x}, \mu) + 2\lambda (U^T \bar{X}^{-2} U) h = 0$ , y por tanto

$$h = -\frac{1}{2\lambda} (U^T \bar{X}^{-2} U)^{-1} U^T \nabla\beta(\bar{x}, \mu),$$

transformando  $h$  hacia el espacio nulo de  $A$  en  $\mathbb{R}^n$  mediante la aplicación de  $U$  a  $h$  tenemos

$$\Delta x = U h = -U (U^T \bar{X}^{-2} U)^{-1} U^T \nabla\beta(\bar{x}, \mu),$$

donde  $\frac{1}{2\lambda}$  no ha sido considerado, debido a que solo interesa el vector de movimiento.

Sustituyendo  $\nabla\beta(\bar{x},\mu) = c - \mu \bar{X}^{-1}e$  y haciendo uso del lema 3.2, vemos que la solución para  $(P_s)$  nos proporciona la dirección de movimiento

$$\Delta x = -\frac{1}{\mu} \bar{X} \left[ I - \bar{X} A^T (A \bar{X}^2 A^T)^{-1} A \bar{X} \right] \bar{X} c + \bar{X} \left[ I - \bar{X} A^T (A \bar{X}^2 A^T)^{-1} A \bar{X} \right] e,$$

la cual es exactamente igual a la encontrada por el método de Newton aplicado a (3.2). Este último desarrollo será eficaz para generar direcciones de movimiento bajo funciones de barrera generalizada.

Ahora examinando nuevamente la dirección de movimiento  $\Delta x$  y definiendo  $P = I - \bar{X} A^T (A \bar{X}^2 A^T)^{-1} A \bar{X}$ , podemos escribir esta como

$$\begin{aligned} \Delta x &= -\frac{1}{\mu} \bar{X} \left[ I - \bar{X} A^T (A \bar{X}^2 A^T)^{-1} A \bar{X} \right] \bar{X} c + \bar{X} \left[ I - \bar{X} A^T (A \bar{X}^2 A^T)^{-1} A \bar{X} \right] e, \\ &= -\frac{1}{\mu} \bar{X} P \bar{X} c + \bar{X} P e, \\ &= -\frac{1}{\mu} \bar{X} P \bar{X} (c - \mu \bar{X}^{-1} e), \\ &= -\frac{1}{\mu} \left( \bar{X}^2 - \bar{X}^2 A^T (A \bar{X}^2 A^T)^{-1} A \bar{X}^2 \right) (c - \mu \bar{X}^{-1} e), \\ &= -\frac{1}{\mu} \bar{X}^2 (c - A^T y - \mu \bar{X}^{-1} e), \\ &= -\frac{1}{\mu} \bar{X}^2 (s - \mu \bar{X}^{-1} e) = \bar{x} - \frac{1}{\mu} \bar{X}^2 s, \end{aligned} \tag{3.6}$$

es decir

$$\Delta x = \bar{x} - \frac{1}{\mu} \bar{X}^2 s. \tag{3.7}$$

Esta forma simplificada de  $\Delta x$  será útil posteriormente, además probaremos que cuando  $x$  converge a  $x^*$ , el vector  $y$  converge a la solución óptima dual, y el vector  $s = c - A^T y$  converge a el vector óptimo de holguras duales. Por esta razón nos referiremos a  $(y, s)$  como el 'dual estimado' en  $x$ .

### 3.1.1 Medida de Proximidad

Sabemos por lo expuesto en el capítulo 2 que la trayectoria central  $\{x(\mu): \mu > 0\}$  es una curva diferenciable en el interior de la región factible del problema primal (P) y cuando  $\mu \rightarrow 0$  converge a la solución óptima de (P).

Sin embargo, es difícil hallar el verdadero mínimo  $x(\mu)$  del problema  $(P_\mu)$  para cada  $\mu > 0$ . De esta forma si renunciamos a la idea de hallar puntos centrados  $x(\mu)$ , lo que el algoritmo generará son puntos aproximadamente centrados. Así, es necesario tener una medida o criterio de cercanía para determinar cuando un punto  $x$  es 'próximo' a  $x(\mu)$ .

¿Qué es una medida de cercanía a la trayectoria central?

Para un valor dado de  $\mu$ ,  $x$  es "cercano" a la trayectoria central si, para algún vector de holgura dual  $s$ , las componentes de  $x_i s_i$  son aproximadamente iguales a  $\mu$ . Aquí "aproximadamente igual" es definido relativo a  $\mu$ .

Roos y Vial [15] definen la siguiente medida de proximidad para la distancia de un punto interior factible  $x$  a  $x(\mu)$ .

**Definición 3.1:** Dado el punto interior estrictamente factible  $x$  y dado  $\mu > 0$ , la medida de proximidad  $\delta(x, \mu)$  es definida por

$$\delta(x, \mu) = \min_s \left\{ \left\| \frac{Xs}{\mu} - e \right\| : A^T y + s = c, y \in \mathbb{R}^m \right\},$$

donde  $\|\cdot\|$  denota la norma euclidiana.

La única solución de este problema de minimización convexo es denotado por  $s(x,\mu)$ . Así podemos también escribir

$$\delta(x,\mu) = \left\| \frac{Xs(x,\mu)}{\mu} - e \right\|. \quad (3.10)$$

El siguiente Lema establece una propiedad de  $\delta(x,\mu)$ .

**Lema 3.3** Para cada  $\mu > 0$ ,  $\delta(x(\mu), \mu) = 0$ . Recíprocamente si  $\delta(x,\mu) = 0$ , entonces  $x$  debe ser la solución óptima de  $(P_\mu)$ .

**Prueba:** Observe que  $\delta(x,\mu) > 0$ , así 0 es el valor mínimo del problema convexo dado en la definición 3.1. Como  $x(\mu)$  es la solución óptima del problema  $(P_\mu)$  y ya que el par  $(y(\mu), s(\mu))$  asociado a el, satisface

$$A^T y(\mu) + s(\mu) - c = 0, \text{ así como también}$$

$$X(\mu)s(\mu) = \mu e, \text{ vemos que}$$

$$\delta(x(\mu),\mu) = \left\| \frac{X(\mu)s(\mu)}{\mu} - e \right\| = 0$$

luego  $x(\mu)$  da el mínimo valor al problema convexo de la definición 3.1 en otras palabras

$$\delta(x(\mu),\mu) = 0$$

Recíprocamente, como el problema  $(P_\mu)$  tiene una única solución, si  $\delta(x,\mu) = 0$  entonces  $x$  debe ser la solución óptima del problema  $(P_\mu)$ . ■

El siguiente lema debido a Roos y Vial [15] establece una relación entre la medida  $\delta(x,\mu)$  y la norma del vector  $\Delta x$ .

**Lema 3.4:** Dados  $x$  y  $\mu$ , tenemos

$$\delta(x,\mu) = \|X^{-1}\Delta x\|$$

**Prueba:** Sabemos que el vector  $\Delta x$  es determinado de

$$\Delta x = x - \frac{1}{\mu} X^2 s$$

de donde

$$X^{-1}(x-\Delta x) = \frac{1}{\mu} X s,$$

$$e - \bar{X}^{-1} \Delta x = \frac{1}{\mu} X s,$$

Por lo tanto

$$\left\| \frac{Xs}{\mu} - e \right\| = \left\| X^{-1} \Delta x \right\|,$$

$$\delta(x, \mu) = \|X^{-1} \Delta x\| \quad \blacksquare$$

### 3.1.2 Selección de los parámetros de barrera

La trayectoria central está conformado por los minimizadores de una familia de problemas de programación convexos con diferentes parámetros  $\mu$ , esta curva es continua en  $\mu$  y lleva a la solución óptima del problema (P) cuando  $\mu$  tiende a cero. Es obvio que la selección de los parámetros de barrera afecta significativamente la velocidad de convergencia y el número total de iteraciones.

La forma más común de seleccionar los parámetros de barrera es usando una sucesión geométrica del siguiente tipo:  $\mu_{k+1} = (1 - \theta)\mu_k$ .

Cuando  $\theta = \frac{\gamma}{\sqrt{n}}$ ,  $\gamma \in \langle 0, 1 \rangle$ , se ha obtenido que la función de barrera logarítmica trabaja

bien en la práctica y proporciona un análisis de complejidad simple que puede ser fácilmente generalizado.

En particular, si seleccionamos  $\mu_{k+1} = (1 - \theta)\mu_k$  para  $k = 0, 1, 2, \dots$  y terminamos las iteraciones cuando  $\mu_k \leq e^{-L}$  para algún número grande  $L > 0$ , entonces requerimos que

$$(1 - \theta)^k \mu_0 \leq e^{-L}$$

$$k \ln(1 - \theta) + \ln \mu_0 \leq -L.$$

De este modo,

$$\ln \mu_0 + L \leq -k \ln(1 - \theta).$$

Ya que  $-k'\ln(1 - \theta) > k'\theta$ , es suficiente escoger  $k'\theta \geq \ln\mu_0 + L$ .

Ahora, cuando

$$k' \geq \frac{1}{\theta} (\ln\mu_0 + L),$$

podemos terminar las iteraciones con  $\mu_k \leq e^{-L}$ .

Obsérvese que en la sucesión geométrica atrae nuestra atención el estudio de  $\theta$ , el cual realmente controla la razón de convergencia de todo el proceso iterativo. Si deseamos diseñar un algoritmo en tiempo polinomial (aun sin considerar lo complicado que puede ser cada

paso) requerimos que  $\frac{1}{\theta}$  sea un factor polinomial en el tamaño del problema.

### 3.1.3 Algoritmo de seguimiento de trayectoria para el problema primal

Los resultados de las secciones previas sugieren el siguiente algoritmo para la solución de (P).

#### Algoritmo 3.1

Datos:  $\beta \in \langle 0, 1 \rangle$  es el parámetro de proximidad,

$\theta = \frac{1}{6\sqrt{n}}$  es el parámetro de reducción,

$\mu_0$  es un valor inicial del parámetro de barrera,

$x^0$  un punto interior inicial factible de (P) con  $\delta(x^0, \mu_0) \leq \beta$

$L$  es el tamaño del problema

**Paso 0:** Sea  $k := 0$

**Paso 1:** Si  $\mu_k \leq e^{-L}$  entonces parar, sino ir al siguiente paso.

**Paso 2:** Calcular

$$y^k := (AX_k^2 A^T)^{-1} (AX_k^2 (c - \mu_k X_k^{-1} e)),$$

$$s^k := c - A^T y^k,$$

$$\Delta x^k := x^k - \frac{1}{\mu_k} X_k^2 s^k,$$

donde  $X_k := \text{diag}(x_1^k, x_2^k, \dots, x_n^k)$ ,

$$e := (1, 1, \dots, 1)^T,$$

**Paso 3:** Sea  $\delta(x^k, \mu_k) := \left\| X_k^{-1} \Delta x^k \right\|$ .

Si  $(\delta(x^k, \mu_k) < \beta)$  hacer

$$\mu_{k+1} := (1 - \theta)\mu_k,$$

$$x^{k+1} := x^k + \Delta x^k,$$

$$k := k + 1,$$

ir al paso 1.

Según D. den Hertog [4] se define

**Definición 3.2:** Un algoritmo de seguimiento de trayectoria será llamado un:

a) Algoritmo de paso largo: si  $\theta \in (0, 1)$  es constante e independiente de  $n$  y  $\varepsilon$ ;

b) Algoritmo de paso medio: si  $\theta = \frac{\gamma}{\sqrt{n}}$ , donde  $\gamma > 0$  es una constante arbitraria,

posiblemente grande e independiente de  $n$  y  $\varepsilon$ .

c) Algoritmo de paso corto: si  $\theta = \frac{\gamma}{\sqrt{n}}$ ,  $\gamma$  es pequeño (por ejemplo,  $1/6$ ) y tal que, después

de una reducción de  $\mu$ , un paso de Newton de longitud unitaria es suficiente para llevarlo a la vecindad de un nuevo  $\mu$ -centro.

El algoritmo de paso largo parece ser el más adecuado para propósitos prácticos. Sin embargo, es posible obtener un resultado que muestre que la complejidad teórica de estos algoritmos es inferior a la complejidad de los algoritmos de paso medio y corto. Es debido a ésta que solo presentamos un algoritmo de seguimiento de trayectoria de paso corto. La figura 3.1 muestra algunas iteraciones para el algoritmo de paso largo y paso medio para un determinado problema de programación lineal. La figura 3.2 muestra algunas iteraciones del algoritmo de paso corto.



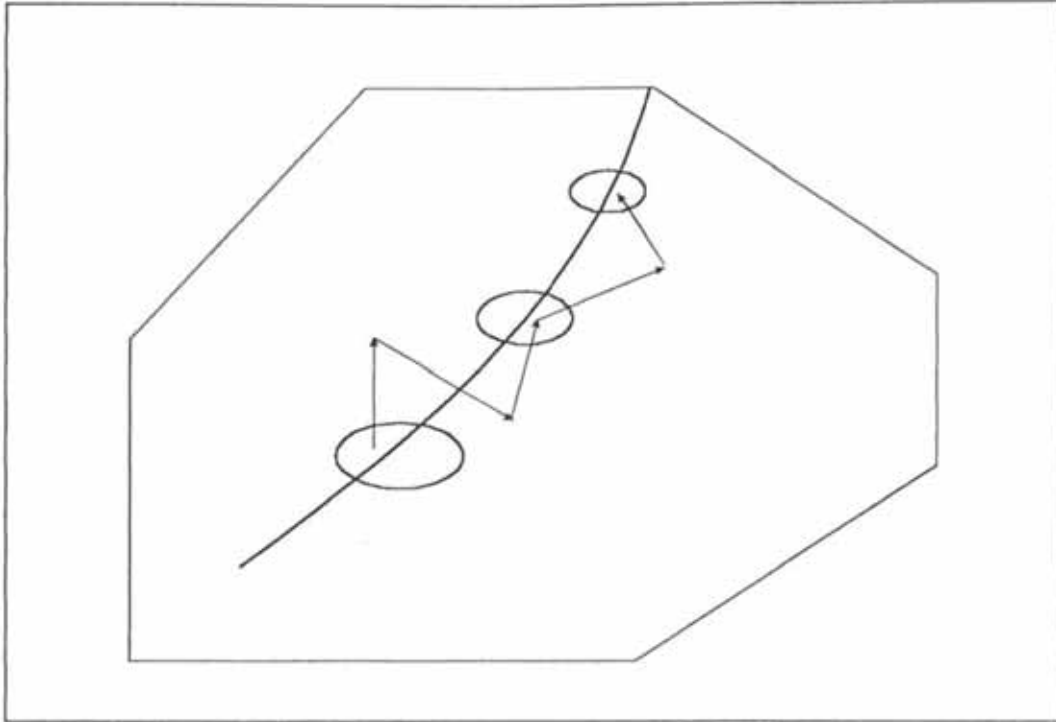


Figura 3.1: Iteraciones del Algoritmo de paso largo (medio)

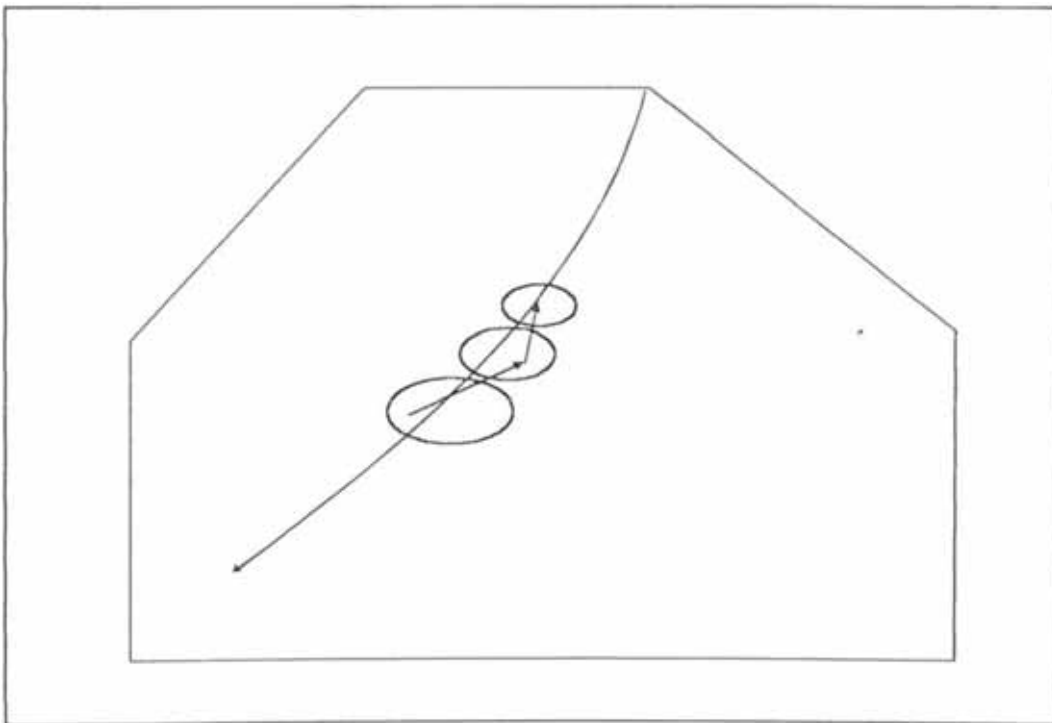


Figura 3.2: Iteraciones del Algoritmo de paso corto

$$\frac{\bar{X}_s}{\mu} = \frac{S\bar{x}}{\mu} = \frac{2Sx}{\mu} - \frac{SX^2_s}{\mu^2} = \frac{2X_s}{\mu} - \frac{(XS)(X_s)}{\mu^2}$$

sea

$$t = \frac{X_s}{\mu} y \quad T = \text{diag}(t) = \frac{XS}{\mu}$$

Entonces

$$\frac{\bar{X}_s}{\mu} - e = 2t - T^2 e - e$$

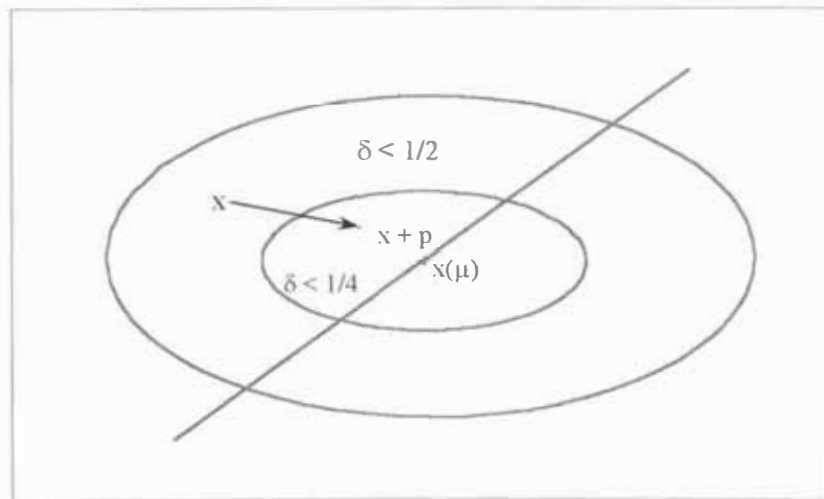
De este modo

$$\delta(\bar{x}, \mu)^2 = \sum_{j=1}^n (2t_j - t_j^2 - 1)^2 = \sum_{j=1}^n (t_j - 1)^4 \leq \left[ \sum_{j=1}^n (t_j - 1)^2 \right]^2 = \delta(x, \mu)^4$$

por tanto

$$\delta(\bar{x}, \mu) \leq \delta(x, \mu)^2 \quad \blacksquare$$

La importancia de este lema es evidente si repetidamente reemplazamos  $x$  por  $\bar{x}$ , obtenemos una sucesión de puntos que converge cuadráticamente a  $x(\mu)$ .



**Fig. 3.3: Convergencia cuadrática en una vecindad de un  $\mu$ -centro**

El siguiente lema, tomado de Roos y Vial [15] da una cota sobre la medida de proximidad de  $x$  con respecto a un nuevo valor (reducido) del parámetro de barrera  $\mu$ .

El siguiente lema, tomado de Roos y Vial [15] da una cota sobre la medida de proximidad de  $x$  con respecto a un nuevo valor (reducido) del parámetro de barrera  $\mu$ .

**Lema 3.6:** Sea  $\bar{\mu} = (1 - \theta)\mu$ , con  $0 < \theta < 1$ , entonces

$$\delta(x, \bar{\mu}) \leq \frac{1}{1 - \theta} (\delta(x, \mu) + \theta \sqrt{n}).$$

**Prueba:** Por la definición de la medida de proximidad tenemos

$$\begin{aligned} \delta(x, \bar{\mu}) &= \min_{\bar{s}} \left\{ \left\| \frac{X\bar{s}}{\bar{\mu}} - e \right\| : A^T \bar{y} + \bar{s} = c \right\} \\ &\leq \left\| \frac{Xs}{\bar{\mu}} - e \right\| = \left\| \frac{Xs}{(1 - \theta)\mu} - e \right\| \\ &= \left\| \frac{t}{1 - \theta} - e \right\| \\ &\leq \frac{1}{1 - \theta} (\|t - e\| + \theta \|e\|). \end{aligned}$$

Ya que  $\|t - e\| = \delta(x, \mu)$  y  $\|e\| = \sqrt{n}$  esto implica el lema ■

Ahora combinamos los resultados de los lemas 3.5 y 3.6, para obtener una cota sobre las medidas de proximidad de un punto obtenido por un paso de Newton, para un nuevo valor del parámetro de barrera.

En particular, si  $\delta(x, \mu) \leq \frac{1}{2}$ , y la reducción en  $\mu$  es suficientemente conservadora (esto es,  $\theta$

es cercano a 1) entonces también  $\delta(\bar{x}, \bar{\mu}) \leq \frac{1}{2}$ . Este es precisamente el contenido del

siguiente lema tomado de Roos y Vial [15].

**Lema 3.7:** Sea  $x$  un punto estrictamente factible para el problema primal y supongamos que

$$\delta(x, \mu) \leq \frac{1}{2}. \text{ Si } \bar{x} = x + \Delta x \text{ y teniendo en cuenta que } \theta = \frac{1}{6\sqrt{n}}, \text{ entonces } \delta(\bar{x}, \bar{\mu}) \leq \frac{1}{2}.$$

**Prueba:** Usando el lema 3.5 y el lema 3.6 sucesivamente, podemos escribir

$$\begin{aligned} \delta(\bar{x}, \bar{\mu}) &\leq \frac{1}{1-\theta} (\delta(\bar{x}, \mu) + \theta\sqrt{n}) \\ &\leq \frac{1}{1-\theta} \left( \delta(x, \mu)^2 + \theta\sqrt{n} \right) \\ &\leq \frac{1}{1-\theta} \left( \frac{1}{4} + \frac{1}{6} \right) = \frac{5}{12(1-\theta)} \leq \frac{1}{2} \end{aligned}$$

y por tanto  $\delta(\bar{x}, \bar{\mu}) \leq \frac{1}{2}$  ■

**Observación:** De los lemas 3.6 y 3.7 podemos concluir que si tenemos un punto estrictamente factible  $x_0$  y un parámetro de barrera satisfaciendo  $\delta(x_0, \mu_0) \leq \frac{1}{2}$ , entonces la sucesión de puntos iterados  $(x_k, \mu_k)$  obtenida por tomar repetidamente un solo paso de Newton

y reducir  $\mu$  por el factor  $\left(1 - \frac{1}{6\sqrt{n}}\right)$ , es estrictamente factible y mantiene una medida de

proximidad

$$\delta(x_k, x_k) \leq \frac{1}{2}.$$

¿Convergerá esta sucesión a una solución óptima cuando  $\mu$  tiende a cero?. El siguiente lema, tomado de Roos y Vial [15], nos ayudará a responder esta pregunta.

Este lema proporciona cotas sobre la brecha dual en un punto  $x$  en términos de  $\mu$  y  $\delta(x, \mu)$

### Lema 3.8: (Brecha dual acotada)

Sea  $x$  un punto estrictamente factible para el problema primal, y sea  $(y, s)$  el dual estimado en  $x$  con respecto a  $\mu$ . Si  $\delta(x, \mu) \leq \delta \leq 1$  entonces  $(y, s)$  es dual factible y

$$\mu(n - \delta\sqrt{n}) \leq c^T x - b^T y \leq \mu(n + \delta\sqrt{n}).$$

**Prueba:** Primero probaremos la factibilidad dual. Por definición  $A^T y + s = c$ , de aquí solo necesitamos probar que  $s \geq 0$ .

Ahora por hipótesis

$$\delta(x, \mu) = \left\| \frac{Xs}{\mu} - e \right\| \leq 1,$$

cada componente de  $\frac{Xs}{\mu} - e$  es a lo más 1 en valor absoluto. Esto implica que  $x_j s_j \geq 0 \forall j$ , y

como  $x > 0$ , se sigue que  $s \geq 0$ .

Debido a que  $x$  es factible para el primal y  $(y, s)$  es factible para el dual,

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

Ahora

$$\Delta x = x - \frac{1}{\mu} X^2 s = X e - \frac{1}{\mu} X^2 s,$$

por tanto

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

Consecuentemente, la brecha dual es

$$\begin{aligned} x^T s &= \mu x^T X^{-1} (e - X^{-1} \Delta x) = \mu e^T (e - X^{-1} \Delta x) \\ &= \mu(n - e^T X^{-1} \Delta x), \end{aligned}$$

así tenemos que

$$\mu(n - \|e\| \|X^{-1} \Delta x\|) \leq x^T s \leq \mu(n + \|e\| \|X^{-1} \Delta x\|),$$

$$-k \ln \left( 1 - \frac{1}{6\sqrt{n}} \right) \geq M$$

Usando la desigualdad

$$-\ln(1 - \alpha) \geq \alpha \quad \forall \alpha : 0 < \alpha < 1,$$

esta inecuación se mantiene válida si  $\frac{k}{6\sqrt{n}} \geq M$ , esto es, si  $k \geq 6\sqrt{n} M$ . Entonces tenemos la cota requerida sobre el número de iteraciones.

# CAPITULO 4

## GENERALIZACIÓN DE ALGORITMOS DE SEGUIMIENTO DE TRAYECTORIA PARA EL PROBLEMA DE LA PROGRAMACIÓN LINEAL

En este capítulo estudiamos un modelo general de algoritmo de seguimiento de trayectoria para el problema de la programación lineal. Definimos funciones de barrera generalizada para este problema, de esta forma generamos una trayectoria interior ideal para seguimiento de trayectoria. Las componentes claves requeridas por el algoritmo genérico, incluyendo las direcciones de movimiento y el criterio de cercanía son también estudiados.

Proponemos un algoritmo genérico de seguimiento de trayectoria y estudiamos las condiciones suficientes para convergencia y polinomialidad. Este capítulo ha sido escrito teniendo como base los artículos de Sheu [17] y Fang [5].

### 4.1 FUNCIONES DE BARRERA GENERALIZADA PARA EL PROBLEMA DE LA PROGRAMACIÓN LINEAL

En esta sección definimos una función de barrera generalizada para el problema de la programación lineal y a partir de ella deducimos propiedades similares a las encontradas para la función de barrera logarítmica.

A lo largo de este capítulo, consideraremos el problema de la programación lineal en la forma estándar (1.1) del capítulo 1, junto con las siguientes hipótesis:

H1) El interior de la región factible

$$\mathcal{P}^0 = \{x : Ax = b, x > 0\} \text{ es no vacío}$$

H2) El conjunto factible  $\mathcal{P} = \{x : Ax = b, x > 0\}$  está acotado.

H3) A tiene rango  $m$  con  $m \leq n$ .

Según Sheu [17] se define

**Definición 4.1:** Una función  $\varphi(x)$  es una función de barrera generalizada o simplemente una función GBLP para el problema de la programación lineal, si

i)  $\varphi : \mathcal{P} \rightarrow \overline{\mathbb{R}}$  es propia, estrictamente convexa y diferenciable, donde  $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ , es la recta real extendida.

ii) Si la sucesión  $\{x^k\} \subset \mathcal{P}^0$  converge a  $x$  con  $x_i = 0$ , para al menos un  $i$ , entonces

$$\lim_{k \rightarrow \infty} (\nabla \varphi(x^k))_i = -\infty,$$

iii)  $\varphi$  toma valores finitos al menos para  $x$  en  $\mathcal{P}^0$ , es decir el dominio efectivo de  $\varphi$  contiene a  $\mathcal{P}^0$ .

Ahora hacemos las siguientes observaciones:

1) Ya que  $\varphi(x)$  es estrictamente convexa y diferenciable, esta debe ser continuamente diferenciable en  $\mathcal{P}$ . En otras palabras,  $\nabla \varphi(x)$  es continua sobre  $\mathcal{P}$ .

2)  $\varphi$  es finita sobre  $\mathcal{P}^0$ , y solo asume el valor de  $+\infty$  cuando  $x_i = 0$  para algún  $i$ .

3) Toda combinación lineal positiva de funciones de barreras generalizadas son nuevamente funciones de barrera generalizada. En otras palabras, función de barreras mixtas pueden ser aplicadas para problemas de programación lineal.

4) Por ejemplo,  $-\sum_{i=1}^n \ln x_i$ ,  $\sum_{i=1}^n x_i \ln x_i$  y  $\sum_{i=1}^n \frac{1}{x_i^r}$  para  $r > 0$ , son todas funciones GBLP.

Ahora para  $\mu > 0$ , definimos un problema primal aumentado  $(P_\mu)$ , asociado con una función GBLP como sigue:



$$(P_\mu) \begin{cases} \min : & c^T x + \mu\varphi(x) \\ \text{sujeto a} & Ax = b, \\ & x \geq 0. \end{cases}$$

Bajo las hipótesis H1 - H2 - H3 se cumple el siguiente teorema, el cual ha sido tomado de Sheu [17]

**Teorema 4.1:**

- i) Para cada  $\mu > 0$ ,  $(P_\mu)$  tiene una única solución optimal  $x(\mu) \in \mathcal{P}^0$ .
- ii) Para  $\mu > 0$ ,  $\{c^T x(\mu)\}$  es una sucesión monótona decreciente en  $\mu$ .

**Prueba:**

- i) Como  $\varphi(x)$  es estrictamente convexa en  $\mathcal{P}$ , la función  $c^T x + \mu\varphi(x)$  es también estrictamente convexa. Además, el dominio factible  $\mathcal{P}$  es convexo y acotado. Esto asegura que  $(P_\mu)$  tiene una única solución óptima  $x(\mu)$  en  $\mathcal{P}$ . Además  $x(\mu)$  satisface la condición de Karush-Kuhn-Tucker (K-K-T) siguiente:

$$\begin{cases} c + \mu \nabla \varphi(x) - A^T y - \lambda = 0, \\ Ax = b, \\ x^T \lambda = 0, \\ x \geq 0, \quad \lambda \geq 0 \end{cases}$$

donde  $y \in \mathbb{R}^m$  y  $\lambda \in \mathbb{R}^n$ , son los vectores langrageanos.

De la definición de la función de barrera generalizada, sabemos que  $(\nabla \varphi(x))_i$  se aproxima a  $-\infty$  si  $x_i$  se aproxima a cero para algún  $i$ . Desde que  $\mu$  es positiva, si  $x_i$  se aproxima a cero para algún  $i$ ,  $\mu(\nabla \varphi(x))_i$  se aproxima a  $-\infty$  y por esto las ecuaciones K-K-T podrían no tienen una solución finita. Por lo tanto,  $x(\mu)$  tiene que estar en  $\mathcal{P}^0$  y las ecuaciones se reducen al siguiente sistema:

$$\begin{cases} c + \mu \nabla \varphi(x) - A^T y = 0 \\ Ax = b, \\ x > 0 \end{cases}$$

ii) Por definición de  $x(\mu_1)$  y  $x(\mu_2)$ , tenemos,  $c^T x(\mu_1) + \mu_1 \varphi(x(\mu_1)) \leq c^T x(\mu_2) + \mu_1 \varphi(x(\mu_2))$  y  $c^T x(\mu_2) + \mu_2 \varphi(x(\mu_2)) < c^T x(\mu_1) + \mu_2 \varphi(x(\mu_1))$  o equivalente

$$-c^T x(\mu_1) - \mu_2 \varphi(x(\mu_1)) \leq -c^T x(\mu_2) - \mu_2 \varphi(x(\mu_2))$$

De estas desigualdades se obtiene:

$$(\mu_1 - \mu_2) \varphi(x(\mu_1)) \leq (\mu_1 - \mu_2) \varphi(x(\mu_2)).$$

si  $\mu_1 > \mu_2 > 0$ , tenemos  $\varphi(x(\mu_1)) < \varphi(x(\mu_2))$ . De este modo, de la segunda inecuación

$$\begin{aligned} c^T x(\mu_2) + \mu_2 \varphi(x(\mu_2)) &< c^T x(\mu_1) + \mu_2 \varphi(x(\mu_1)) \\ &\leq c^T x(\mu_1) + \mu_2 \varphi(x(\mu_2)) \end{aligned}$$

lo cual implica que

$$c^T x(\mu_2) < c^T x(\mu_1) \text{ para } 0 < \mu_2 < \mu_1 \quad \blacksquare$$

Además en los siguientes dos teoremas tomados de Sheu [17], vemos que cuando  $\mu$  tiende a cero, la curva formada por todas los minimizadores  $x(\mu)$  es continua en  $\mu$  y converge a una solución optimal  $x^*$  del problema lineal (1.1).

**Teorema 4.2:** El conjunto  $\{x(\mu): \mu > 0\}$  es una curva continua en  $\mathcal{P}^0$ .

**Prueba:** Por el teorema 4.1 solo debemos probar la parte referente a la continuidad. Suponga que la curva es discontinua en  $\bar{\mu} > 0$ . Entonces, existe un  $\varepsilon > 0$  tal que para cada  $\Delta\mu > 0$ , tenemos  $\|x(\bar{\mu} + \Delta\mu) - x(\bar{\mu})\| > 2\varepsilon$ . Probaremos por contradicción que  $x(\bar{\mu} + \Delta\mu)$  no es el mínimo de  $(P_{\bar{\mu} + \Delta\mu})$ . Como  $c^T x + \bar{\mu} \mu(x)$  es estrictamente convexa y desde que  $x(\bar{\mu})$  es el mínimo de  $(P_{\bar{\mu}})$  para  $x \in \mathcal{P}^0$  y  $\|x - x(\bar{\mu})\| > \varepsilon$ , existe un  $\tau > 0$  tal que

$$c^T x + \bar{\mu} \varphi(x) > c^T x(\bar{\mu}) + \bar{\mu} \varphi(x(\bar{\mu})) + \tau \quad (4.1)$$

En particular,

$$c^T x(\bar{\mu} + \Delta\mu) + \bar{\mu} \varphi(x(\bar{\mu} + \Delta\mu)) > c^T x(\bar{\mu}) + \bar{\mu} \varphi(x(\bar{\mu})) + \tau,$$

no importando cuan pequeño es  $\Delta\mu$ .

### Caso 1:

Primero asumiremos que el dominio efectivo de  $\varphi(x)$  es  $\mathcal{P}$ . En este caso, ya que  $\varphi(x)$  es continua sobre el conjunto compacto  $\mathcal{P}$ , podemos siempre escoger  $\Delta\mu$  suficientemente pequeño tal que

$$0 < |\Delta\mu| \|\varphi(x)\|_{\infty} < \frac{\tau}{2}, \quad (4.2)$$

donde  $\|\cdot\|_{\infty}$  es la norma infinita. De este modo

$$\begin{aligned} c^T x(\bar{\mu}) + (\bar{\mu} + \Delta\mu) \varphi(x(\mu)) & \\ & < c^T x(\bar{\mu}) + \bar{\mu} \varphi(x(\bar{\mu})) + \tau/2 \quad \text{por (4.2)} \\ & < c^T x(\bar{\mu} + \Delta\mu) + \bar{\mu} \varphi(x(\bar{\mu} + \Delta\mu)) - \tau/2 \quad \text{por (4.1)} \\ & < c^T x(\bar{\mu} + \Delta\mu) + (\bar{\mu} + \Delta\mu) \varphi(x(\bar{\mu} + \Delta\mu)) \quad \text{por (4.2)} \end{aligned}$$

Lo cual prueba que  $x(\bar{\mu} + \Delta\mu)$  no es el mínimo de  $(P_{\bar{\mu} + \Delta\mu})$ .

### Caso 2:

Por otro lado, si  $\varphi(x)$  toma el valor  $+\infty$  sobre la frontera, tal que  $\varphi(x(\bar{\mu} + \Delta\mu))$  podría posiblemente estar no acotada superiormente para  $\Delta\mu$  pequeño, entonces

$$\begin{aligned} c^T x(\bar{\mu}) + (\bar{\mu} + \Delta\mu) \varphi(x(\bar{\mu})) & \\ & < c^T x(\bar{\mu} + \Delta\mu) + (\bar{\mu} + \Delta\mu) \varphi(x(\bar{\mu} + \Delta\mu)) \quad (4.3) \end{aligned}$$

es automáticamente verdadero desde que el lado derecho de (4.3) es no acotado ■

Como  $\{x(\mu)_k : \mu > 0\}$  es una curva continua y bien definida en  $\mathcal{P}^0$ , podemos entonces definir

$$x^* = \lim_{k \rightarrow \infty} x(\mu_k),$$

donde  $\mu_k$  tiende a cero. Todo lo que necesitamos hacer para probar esto es que  $x^*$  es verdaderamente el mínimo de  $x(\mu)$ . Esto es precisamente el contenido del siguiente teorema.

**Teorema 4.3:** Dada una sucesión positiva decreciente  $\{\mu_k\}$  tal que  $\lim_{k \rightarrow \infty} \mu_k = 0$ , si  $x^* = \lim_{k \rightarrow \infty} x(\mu_k)$ , entonces  $x^*$  es el mínimo del problema de programación lineal (1.1)

**Prueba:**

**Caso 1:** Suponga que el dominio efectivo de  $\varphi(x)$  es  $\mathcal{P}$ , entonces para cada  $x \in \mathcal{P}$ , tenemos

$$c^T x(\mu_k) + \mu_k \varphi(x(\mu_k)) \leq c^T x + \mu_k \varphi(x)$$

para cada  $k$ .

Como  $\mathcal{P}$  es compacto,  $x^*$  debe estar en  $\mathcal{P}$ . tomando límites a ambos lados de la desigualdad anterior y considerando la continuidad de  $\varphi$ , tenemos

$$c^T x^* < c^T x.$$

Además, como  $x(\mu_k) > 0$ ,  $Ax(\mu_k) = b$ , tenemos

$$\lim_{k \rightarrow \infty} x(\mu_k) = x^* > 0 \quad \text{y} \quad Ax^* = b.$$

Luego  $x^*$  es la única solución del problema lineal (1.1).

**Caso 2:** Si  $\varphi(x)$  toma el valor de  $+\infty$  en la frontera, entonces haremos la prueba por contradicción. Asumamos que  $x^*$  no es el mínimo del problema lineal (1.1), pero  $v^*$  si lo es. Entonces, podemos encontrar  $u \in \mathcal{P}^0$  tal que  $c^T v^* < c^T u < c^T x^*$ .

Como  $\lim_{k \rightarrow \infty} \mu_k = 0$ , podemos escoger  $k$  suficientemente grande y un escalar  $\delta > 0$  tal que

$$c^T \mu + \mu_k \varphi(\mu) < c^T x^* - \delta$$

Suponga que  $x^* \in \mathcal{P}^0$ . Como este punto es el límite de  $\{x(\mu_k)\}$ , cuando  $k$  es suficientemente grande,  $x(\mu_k)$  estará en una pequeña vecindad de  $x^*$  sobre la cual  $\varphi(x(\mu_k))$  es acotada y toma valores finitos. De este modo tenemos

$$c^T x^* - \delta < c^T x(\mu_k) + \mu_k \varphi(x(\mu_k))$$

para  $k$  suficientemente grande. Pero esto contradice el hecho de que  $x(\mu_k)$  es el mínimo de  $(P_{\mu_k})$ .

Por otro lado, si  $x^* \in \mathcal{P} \setminus \mathcal{P}^0$  y  $\varphi(x)$  toma el valor  $+\infty$  en  $x^*$ ,  $\mu_k \varphi(x(\mu_k))$  podrá no tener un límite. Sin embargo, tenemos  $\liminf_{k \rightarrow \infty} \{\mu_k \varphi(x(\mu_k))\} \geq 0$ .

De este modo para  $k$  suficientemente grande tenemos

$$c^T x^* - \frac{\delta}{2} < c^T x(\mu_k) + \liminf_{k \rightarrow \infty} \{\mu_k \varphi(x(\mu_k))\}$$

$$< c^T x(\mu_k) + \mu_k \varphi(x(\mu_k)) + \frac{\delta}{2}$$

consecuentemente,

$$c^T \mu + \mu_k \varphi(\mu) < c^T x^* - \delta < c^T x(\mu_k) + \mu_k \varphi(x(\mu_k))$$

lo cual nuevamente es una contradicción.

De este modo  $x^*$  debe ser la solución óptima del problema lineal (1.1) ■

**Teorema 4.4:** Para cada sucesión positiva decreciente  $\{\mu_k\}$  tal que  $\lim_{k \rightarrow \infty} \mu_k = 0$ ,  $\{x(\mu_k)\}$

debe converger a  $x^*$ , la solución óptima del problema lineal (1.1).

**Prueba:** Como el dominio factible  $\mathcal{P}$  es compacto, la sucesión  $\{x(\mu_k)\}$ , debe tener al menos un punto límite.

Para cada punto límite podemos encontrar una subsucesión que converge a este punto límite. Aplicando el teorema 4.3, podemos concluir que cada punto límite es a su vez una solución óptima del problema lineal (1.1). Pero hemos asumido que este problema tiene una única solución óptima, de este modo  $\{x(\mu_k)\}$  tiene uno y solo un punto límite y entonces la sucesión converge a este único punto límite ■

## 4.2 CALCULO DE LAS DIRECCIONES DE BUSQUEDA

Las direcciones de búsqueda son la esencia de un algoritmo iterativo. Una vez que la trayectoria definida por una función de barrera generalizada una dirección de movimiento es necesaria para pasar a lo largo de la trayectoria para optimalidad. Basados en el esquema desarrollado en el capítulo 3, sección 3.1, tenemos las siguientes fórmulas extendidas dadas por el corolario del siguiente teorema, las cuales son tomadas de Sheu [17].

**Teorema 4.5:** Sea  $\tilde{X}^{-1}$  una matriz simétrica definida positiva arbitraria y sea  $\bar{x}$  una solución factible interior de  $(P_\mu)$ . Entonces, una solución optima de

$$(P'_\mu) \begin{cases} \min (c + \mu \nabla \phi(\bar{x}))^T \Delta x \\ \text{sujeto a } A \Delta x = 0, \\ \|\tilde{X}^{-1} \Delta x\|^2 \leq \beta^2 < 1. \end{cases}$$

define una dirección de búsqueda dada por

$$\Delta x = -\tilde{X} \left[ \tilde{I} - \tilde{X} A^T (A \tilde{X}^2 A^T)^{-1} A \tilde{X} \right] \tilde{X} (c + \mu \nabla \phi(\bar{x}))$$

**Prueba:** Primero observamos que  $(P'_\mu)$ , bajo la asunción de punto interior, es siempre factible, además por la hipótesis H3,  $A$  tiene rango máximo, el espacio nulo de  $A$  es un subespacio de  $\mathbb{R}^n$  de dimensión  $n - m$ , el cual es isomórfico a  $\mathbb{R}^{n-m}$ . Sea  $U$  una matriz base del espacio nulo de  $A$  y hagamos  $\Delta x = Uh$  para  $h \in \mathbb{R}^{n-m}$ . En esta forma, podemos eliminar las restricciones en un problema de minimización sobre  $h$ , es decir,

$$\begin{aligned} & \min (c + \mu \nabla \phi(\bar{x}))^T Uh, \\ & \text{sujeto a } \|\tilde{X}^{-1} Uh\|^2 \leq \beta^2, \\ & h \in \mathbb{R}^{n-m}. \end{aligned}$$

La función Lagrangeana asociada es

$$L(h, \lambda) = (c + \mu \nabla \phi(\bar{x}))^T Uh + \lambda (\|\tilde{X}^{-1} Uh\|^2 - \beta^2),$$

con multiplicador de Lagrange asociado  $\lambda \geq 0$ .

Tomando  $\frac{\partial L}{\partial h} = 0$ , resolvemos para  $h$  y obtenemos

$$h = -\frac{1}{2\lambda} (U^T \tilde{X}^{-2} U^{-1})^{-1} U^T (c - \mu \nabla \varphi(\bar{x}))^T,$$

$$\text{y } \Delta x = Uh = -\frac{1}{2\lambda} (U^T \tilde{X}^{-2} U^{-1})^{-1} U^T (c + \mu \nabla \varphi(\bar{x}))^T,$$

por el Lema 3.1, la siguiente identidad

$$\tilde{X} \left[ I - \tilde{X} A^T (A \tilde{X}^2 A^T)^{-1} A \tilde{X} \right] \tilde{X} x = U (U^T \tilde{X}^{-2} U^{-1})^{-1} U^T x,$$

se mantiene para todo  $x$  en  $\mathbb{R}^n$ , de este modo

$$\Delta x = -\tilde{X} \left[ I - \tilde{X} A^T (A \tilde{X}^2 A^T)^{-1} A \tilde{X} \right] \tilde{X} (c + \mu \nabla \varphi(\bar{x}))^T, \quad (4.4)$$

donde  $\frac{1}{2\lambda}$  es omitido, ya que un escalar no afecta la dirección de movimiento. Esto completa

la prueba ■

#### Corolario 4.6: (Dirección de Newton para $(P_\mu)$ )

Sea una función de barrera generalizada  $\varphi(x)$  dos veces diferenciable en el dominio factible y si escogemos  $\tilde{X}^{-2}$  como la matriz hessiana de  $c^T x + \mu \varphi(x)$  en  $\bar{x}$  y la denotamos por  $H_\varphi^2$ , entonces  $\Delta x$  obtenido en el teorema 4.5 es la dirección de Newton al problema  $(P_\mu)$ .

#### Prueba:

Las condiciones K-K-T de  $(P_\mu)$  son

$$\begin{cases} c + \mu \nabla \varphi(x) - A^T y = 0, \\ Ax = b, \\ x > 0 \end{cases} \quad (4.5)$$

Si definimos  $\mu \nabla \varphi(x) = -s$ , entonces el sistema (4.5) queda en la forma

$$\begin{cases} A^T y + s - c = 0, \\ Ax - b = 0, \\ \mu \nabla \varphi(x) = -s, \\ x > 0 \end{cases} \quad (4.6)$$

Nótese que el problema  $(P_\mu)$  es un programa convexo teniendo una única solución optimal, la cual es completamente caracterizada por (4.6). La aproximación de primer orden de  $\mu \nabla \varphi(x) + s = 0$ , da

$$-\bar{s} - \mu \nabla \varphi(\bar{x}) = \begin{bmatrix} H_\varphi^{-2} & I_n \end{bmatrix} \begin{pmatrix} x - \bar{x} \\ s - \bar{s} \end{pmatrix},$$

donde  $H_\varphi^{-2} = \mu \nabla^2 \varphi(x)$ , simplificando esta expresión, tenemos

$$s = \gamma - H_\varphi^{-2} x \quad (4.7)$$

donde  $\gamma = -\mu \nabla \varphi(\bar{x}) + H_\varphi^{-2} c$ . Reemplazando  $s$  en (4.7) por  $c - A^T y$  y escribiendo  $x$  en términos de  $\gamma$ , obtenemos

$$x = H_\varphi^2 (-c + \gamma + A^T y) \quad (4.8)$$

Aplicando  $A$  a ambos lados de (4.8), tenemos

$$y = (A H_\varphi^2 A^T)^{-1} (b + A H_\varphi^2 c - A H_\varphi^2 \gamma) \quad (4.9)$$

Colocando (4.9) en (4.8), tenemos la dirección de búsqueda

$$\Delta x = -H_\varphi [I - H_\varphi A^T (A H_\varphi^2 A^T)^{-1} A H_\varphi] H_\varphi (c + \mu \nabla \varphi(x)). \quad (4.10)$$

### 4.3 CRITERIO DE CERCANÍA

En la sección 4.1, caracterizamos la trayectoria interior ideal, como una curva continua en  $\mu$ , la cual lleva a la solución óptima del problema lineal (1.1). Sin embargo por lo desarrollado en el capítulo 3, sabemos que calcular el valor exacto  $x(\mu)$  de  $(P_\mu)$  para cada  $\mu > 0$  es muy difícil. De esta forma necesitamos un 'criterio de cercanía', el cual puede ser como una



$$\begin{aligned}
d(\bar{x}, \mu) &= \left\| \bar{X} \left( \frac{1}{\mu} s + \nabla \varphi(\bar{x}) \right) \right\| \\
&= \left\| \left( I - \bar{X} A^T (A \bar{X}^2 A^T)^{-1} A \bar{X} \right) \bar{X} \left( \frac{1}{\mu} c + \nabla \varphi(\bar{x}) \right) \right\| \quad (4.11)
\end{aligned}$$

donde  $s = c - A^T y$  y donde  $y$  es el dual estimado (4.10).

Varias propiedades de  $d(\bar{x}, \mu)$  pueden probarse aquí, en particular los siguientes teoremas tomados de Sheu [17], prueba que  $d(\bar{x}, \mu)$  es la mejor medida en  $\bar{x}$  con menor error cuadrático entre todos los pares duales  $(y, s)$ .

**Teorema 4.7:**  $d(\bar{x}, \mu)$  es la solución óptima del siguiente problema de programación convexo

$$(\text{PC}) \begin{cases} \min_{y, s} \left\| \bar{X} \left( \frac{1}{\mu} s + \nabla \varphi(\bar{x}) \right) \right\| \\ s = c - A^T y, \\ y \in \mathbb{R}^m \end{cases}$$

**Prueba:** Sustituyendo  $s = c - A^T y$  en la función objetivo, tenemos un problema no restringido equivalente

$$\min_y \left\| \bar{X} \left( \frac{1}{\mu} (c - A^T y) + \nabla \varphi(\bar{x}) \right) \right\|$$

Además,

$$\begin{aligned}
& \left\| \bar{X} \left( \frac{1}{\mu} (c - A^T y) + \nabla \varphi(\bar{x}) \right) \right\|^2 \\
&= \left\langle \left[ \frac{1}{\mu} c^T - \frac{1}{\mu} y^T A + \nabla \varphi(\bar{x})^T \right] \bar{X}, \bar{X} \left[ \frac{1}{\mu} (c - A^T y) + \nabla \varphi(\bar{x}) \right] \right\rangle \\
&= \left\| \bar{X} \left( \frac{1}{\mu} c + \nabla(\bar{x}) \right) \right\|^2 + \frac{1}{\mu^2} y^T (A \bar{X}^2 A^T) y - \frac{2}{\mu} y^T A \bar{X}^2 \left[ \frac{1}{\mu} (c + \nabla \varphi(\bar{x})) \right]
\end{aligned}$$

Note que esta última igualdad es una función de  $y$  y entonces, su derivada con respecto a  $y$  debe ser cero, entonces obtenemos

$$y = (A \bar{X}^2 A^T)^{-1} A \bar{X}^2 (c + \mu \nabla \varphi(\bar{x}))$$

el cual es la solución óptima del problema (PC). Pero  $y$  es el dual estimado inducido por  $(\bar{x}, \mu)$ , de este modo, por definición,  $d(\bar{x}, \mu)$  es el mínimo de (PC) ■

El siguiente teorema tomado de Sheu [17], da una propiedad importante de esta función criterio de cercanía.

**Teorema 4.8:** Para cada  $\mu > 0$ ,  $d(x(\mu), \mu) = 0$  recíprocamente, si  $d(\bar{x}, \mu) = 0$ , entonces  $\bar{x}$  debe ser la solución óptima de  $(P_\mu)$ .

**Prueba:** Por el Teorema 4.7 sabemos que el problema de programación convexa

$$(PC): \min_{(y,s)} \left\{ \left\| \bar{X} \begin{pmatrix} 1 \\ \mu \end{pmatrix} s + \nabla \varphi(\bar{X}) \right\| : s = c - A^T y, y \in \mathbb{R}^m \right\}$$

tiene solución única y además observamos que  $d(\bar{x}, \mu) \geq 0$  y este es el valor mínimo de (PC), el cual depende solo de la solución primal factible dada  $\bar{x}$  y  $\mu$ . Ya que  $x(\mu)$  es la solución óptima de  $(P_\mu)$  y desde que su par dual  $(y(\mu), s(\mu))$  satisface  $A^T y(\mu) + s(\mu) - c = 0$  así como

también  $\begin{pmatrix} 1 \\ \mu \end{pmatrix} s(\mu) = -\nabla \varphi(x(\mu))$ , vemos que  $\left\| X(\mu) \left( \begin{pmatrix} 1 \\ \mu \end{pmatrix} s(\mu) + \nabla \varphi(x(\mu)) \right) \right\| = 0$ . Luego, este

debe ser el valor mínimo de (PC) en  $x(\mu)$ . En otras palabras,  $d(x(\mu), \mu) = 0$ . Recíprocamente, ya que  $(P_\mu)$  tiene una solución única, si  $d(\bar{x}, \mu) = 0$ , entonces  $\bar{x}$  debe ser la solución óptima de  $(P_\mu)$  ■

Nótese que la función  $d(\bar{x}, \mu)$  es solo un número el cual muestra cuan cercano está el punto iterado  $\bar{x}$  del minimizador real  $x(\mu)$ . Esta no es una función distancia definida sobre  $(x, \mu)$ , ya que ésta no dice nada acerca de la relación entre las distancias entre dos puntos factibles arbitrarios se tiene que en particular,  $d(x^1, \mu) < d(x^2, \mu)$  no implica que  $x^1$  está más cercano a  $x(\mu)$  que  $x^2$ .

Ya que la d-función es definida usando la norma  $\bar{X}^{-1}$  bajo la cual el espacio ha sido distorsionado según diferentes puntos actuales, y de este modo puede no reflejar la verdadera relación geométrica en el espacio original.

#### 4.4 ALGORITMO GENÉRICO DE SEGUIMIENTO DE TRAYECTORIA

Teniendo todas las componentes esenciales, podemos construir un algoritmo genérico de seguimiento de trayectoria usando una función de barrera generalizada como sigue

##### Algoritmo 4.4.1

##### Paso 1 (Inicialización)

Sea  $\theta > 0$ ,  $\beta < 1$  constantes,  $L$  el tamaño de (1.1),  $\mu_0 > 0$ , y  $x^0$  una solución inicial interior factible de (1.1) tal que  $d(x^0, \mu_0) \leq \beta$ .

Sea  $f_\mu(x) = c^T x + \mu\varphi(x)$  y  $k = 0$

##### Paso 2 (Comprobación de optimalidad)

Si  $\mu_k < e^{-L}$ , entonces parar. En otro caso, ir al segundo paso.

##### Paso 3 (Cálculo de los valores duales)

$$y^k := (A X_k^2 A^T)^{-1} A X_k^2 (c + \mu \nabla \varphi(x^k))$$

$$s^k := c - A^T y^k$$

$$\gamma^k := X_k \left( \frac{1}{\mu_k} s^k + \nabla \varphi(x^k) \right) \quad \text{donde } X_k = \text{diag}(x_1^k, \dots, x_n^k)$$

##### Paso 4 (Comprobar criterio de cercanía)

$$d(x^k, \mu_k) := \|\gamma^k\|$$

si  $(d(x^k, \mu_k) \leq \beta)$ , poner

$$\mu_{k+1} := (1-\theta)\mu_k$$

$$x^{k+1} := x^k - X_k \gamma^k$$

$$k := k+1$$

Ir al paso 2

En otro caso, ir al siguiente paso.

### Paso 5 (Cálculo de una solución cercana a $x(\mu_k)$ )

$$\delta := \operatorname{argmin}_{0 < \alpha \leq 1} \left\{ f_{\mu_k} \left( x^k - \alpha X_k \gamma^k \right) / x^k - \alpha X_k \gamma^k > 0 \right\}$$
$$x^k := x^k - \delta X_k \gamma^k$$

Ir al paso 3, sin adaptar  $k$  y  $\mu_k$ .

Note que asumimos que el algoritmo inicia con un parámetro de barrera  $\mu_0 > 0$  y una solución interior factible  $x^0$  tal que el criterio inicial de cercanía,  $d(x^0, \mu_0) \leq \beta$  es satisfecho. Obtener un punto inicial  $x^0$  requiere un procedimiento de fase 1, en el capítulo 2, presentamos tal procedimiento para  $\varphi(x) = -\sum_{i=1}^n \ln x_i$  como un caso especial.

Para una función de barrera generalizada, se requieren estudios adicionales para diseñar un procedimiento análogo.

## 4.5 CONDICIONES SUFICIENTES PARA CONVERGENCIA Y POLINOMIALIDAD

En esta sección, imponemos algunas restricciones sobre las funciones GBLP para asegurar que el algoritmo 4.4.1 converge a una solución óptima del problema (1.1). Luego estudiaremos condiciones suficientes bajo las cuales el algoritmo converge en tiempo polinomial.

Sea  $\varphi(x)$  una función GBLP y  $X = \operatorname{diag}(x_1, x_2, \dots, x_n)$ ,  $x = (x_1, \dots, x_n) \in \mathcal{P}$ . Consideremos la siguiente condición.

**Condición 1:** Existe un  $M > 0$  tal que  $\|X \nabla \varphi(x)\| < M$ , para cada  $x \in \mathcal{P}$ .

El siguiente teorema, tomado de Sheu [17] estudia el paso 5 del algoritmo 4.4.1.

**Teorema 4.9:** Para cada  $\mu_k > 0$ , si hay una solución primal factible  $x^k$  tal que  $d(x^k, \mu_k) > \beta$ , entonces el paso 5 del algoritmo generará una solución primal factible  $x'$  tal que

$$d(x', \mu_k) < \beta$$

**Prueba:** Recuerde que:

$$\begin{aligned}\Delta x^k &= -X_k \left[ I - X_k A^T (A X_k^2 A^T)^{-1} A X_k \right] X_k \left( \frac{c}{\mu_k} + \nabla \varphi(x^k) \right) \\ &= -X_k \gamma^k\end{aligned}$$

Nótese que  $d(x^k, \mu_k) = \|\gamma^k\| > \beta \neq 0$  y  $d(x^k, \mu_k) = 0$  si y solo si  $\|\gamma^k\| = 0$ . Si consideramos la derivada direccional de  $f_{\mu_k}$  en la dirección  $\Delta x^k$  y definimos

$$g(t) = f_{\mu_k}(x^k + t \Delta x^k),$$

Entonces

$$g'(0) = \left. \frac{d}{dt} f_{\mu_k}(x^k + t \Delta x^k) \right|_{t=0} = -\mu_k \|\Delta x^k\|^2 < 0$$

Ya que  $f_{\mu_k}$  es continuamente diferenciable,  $g'(t)$  es continua en  $t = 0$ , luego debe existir un intervalo  $[-\alpha, \alpha]$  con  $\alpha > 0$ , tal que  $g'(t) < 0 \forall t \in [-\alpha, \alpha]$ .

Esto implica que  $g(t)$  es monótonamente decreciente en  $[-\alpha, \alpha]$ . En otras palabras, cuando iniciamos en  $x^k$ , moviéndonos a lo largo de la dirección  $\Delta x^k$  con longitud de paso  $\alpha$ , el punto resultante dará un valor menor de la función objetivo con relación al valor en  $x^k$ . Si

$$\delta := \arg \min_{0 < \alpha \leq 1} \left\{ f_{\mu_k}(x^k + \alpha \Delta x^k) / x^k + \alpha \Delta x^k > 0 \right\};$$

$$x^k := x^k + \delta \Delta x^k$$

y repetimos el mismo procedimiento hasta que  $\|\Delta x^k\| = 0$ . En esta forma, el algoritmo genera una sucesión de soluciones primales factibles con valor objetivo monótonamente decreciente. Y debido a que  $f_{\mu_k}$  tiene una única solución  $x(\mu_k)$  con  $d(x(\mu_k), \mu_k) = 0$ , la sucesión debe converger a  $x(\mu_k)$ . Luego, el algoritmo generará una solución primal factible  $x'$  con  $d(x', \mu_k) < \beta$ , después de repetir varias veces el paso 5 ■.

El siguiente teorema, contenido en Sheu [17] continua con el análisis del algoritmo.

**Teorema 4.10:** Bajo la condición 1, si  $\{x^k\}$  es la sucesión generada por el algoritmo 4.4.1 satisfaciendo  $d(x^k, \mu_k) < \beta$ , es decir,

$$\left\| X_k \left( \frac{s^k}{\mu_k} + \nabla \phi(x^k) \right) \right\| \leq \beta, \quad k = 1, 2, \dots$$

entonces

$$\lim_{k \rightarrow \infty} X_k s^k = 0$$

**Prueba:** En el teorema 4.9 se prueba que tal sucesión existe, también,

$$\left\| X_k \frac{s^k}{\mu_k} \right\| - \left\| X_k \nabla \phi(x^k) \right\| \leq \left\| X_k \left( \frac{s^k}{\mu_k} + \nabla \phi(x^k) \right) \right\| \leq \beta$$

la condición 1 implica que  $\left\| X_k s^k \right\| \leq \mu_k (\beta + M)$ .

Cuando  $\mu_k \rightarrow 0$ , observamos que, cuando  $k$  tiende a infinito

$$\sum_{i=1}^n (x_i^k s_i^k)^2 \rightarrow 0.$$

Esto a su vez implica que  $X_i^k \rightarrow 0$  ó  $s_i^k \rightarrow 0$  para cada componente  $i$ . De este modo

$$\lim_{k \rightarrow \infty} X_k s^k = 0 \quad \blacksquare$$

El siguiente teorema, tomando de Sheu [17], muestra la convergencia de la sucesión, generada por el algoritmo 4.4.1.

**Teorema 4.11:** Si la sucesión  $\{x^k\}$  en el teorema 4.10 converge, esta converge a la solución óptima del problema lineal (1.1).

**Prueba:** Supongamos que  $\{x^k\}$  converge al punto  $x^*$ . Ya que  $\{x^k\}$  es una sucesión de soluciones primales factibles,  $Ax^* = b$  y  $x^* \geq 0$ . Si podemos probar que la sucesión correspondiente  $\{s^k\}$  converge a la solución dual factible  $s^*$ , entonces el teorema 4.10 generará la condición de holguras complementarias para  $s^*$  y  $x^*$ . De este modo  $x^*$  será una solución óptima para el problema lineal (1.1).

De acuerdo al algoritmo 4.4.1,  $s^k$  es definida por

$$s^k = c - A^T (AX_k^2 A^T)^{-1} AX_k^2 (c + \mu_k \nabla \phi(x^k)),$$

bajo la hipótesis de no degeneración primal-dual,  $AX_k$ , siempre tiene rango máximo para  $x^k \in \mathcal{P}$ . Luego, la matriz  $(AX_k^2 A^T)$  es invertible.

Por otro lado,  $X_k \nabla \varphi(x^k)$  está uniformemente acotada (condición 1), y de aquí  $s^k$  es una función continua bien definida en  $x$  y  $\mu_k$ , tomando límites, tenemos

$$\lim_{m \rightarrow \infty} s^k = s^* = c - A^T (A(X^*)^2 A^T)^{-1} A(X^*)^2 c$$

Suponga que  $x^*$  no es dual factible, es decir, algún componente de  $s^*$  es negativo,  $s_j^* < 0$ , entonces podemos encontrar  $N_1 > 0$  tal que  $s_j^k \leq \zeta < 0$  para  $k \geq N_1$ , además, sabemos que

$$x_j^{k+1} = x_j^k - \delta(x_j^k)^2 \left( \frac{s_j^k}{\mu_k} + (\nabla \varphi(x^k))_j \right)$$

Sin embargo, del teorema 4.10,  $x_j^*$  tiende a ser cero y esto fuerza a la  $j$ -ésima componente de  $\nabla \varphi(x^*)$  a ser  $-\infty$  (por la definición de función de barrera generalizada). Esto indica que existe  $N_2 > 0$  tal que

$$(\nabla \varphi(x^k))_j < 0 \text{ para } k \geq N_2$$

Tomando  $N = \max\{N_1, N_2\}$ , vemos que

$$x_j^{k+1} = x_j^k - \delta(x_j^k)^2 \left( \frac{s_j^k}{\mu_k} + (\nabla \varphi(x^k))_j \right) > x_j^k,$$

para  $k > N$ . De este modo  $\lim_{k \rightarrow \infty} x_j^k > x_j^N > 0$ . Esto contradice  $x_j^* = 0$ . Luego  $s^*$  debe ser dual

factible ■

**Teorema 4.12:** Bajo la condición 1, el algoritmo 4.4.1 converge a la solución óptima del problema lineal (1.1)

**Prueba:** Para cada  $\mu_k$  iniciando desde  $\mu^0$ , el algoritmo 4.4.1 genera una sucesión  $\{x^k\}$  con  $d(x^k, \mu_k) \leq \beta$ . Desde que el dominio factible es compacto, hay una subsucesión  $\{x^{k_m}\}$  que

converge a una solución óptima del problema lineal (1.1) (por los teoremas 4.10 y 4.11. Pero el problema lineal (1.1) es dual no degenerado (hipótesis) y entonces tienen una única solución óptima. Luego  $\{x^k\}$  debe converger a la misma solución óptima del problema lineal (1.1) ■

Nótese que, para una función de barrera generalizada particular, si podemos encontrar un  $\theta > 0$  y  $1/\theta \in \mathcal{O}(n^s)$  con  $s > 0$ , tal que solo un número polinomial de iteraciones del paso 5 son requeridas para generar un  $x^k$  satisfaciendo  $d(x^k, \mu_k) \leq \beta$  para cada  $k$ , entonces el algoritmo 4.4.1 obviamente tiene una cota en tiempo polinomial. En particular, consideramos la siguiente condición

**Condición 2:** Existe una constante  $0 < \theta' < 1$  con  $\frac{1}{\theta'} \in \mathcal{O}(n^s)$  para  $s > 0$  y otra constante  $0 < \gamma < 1$  tal que, para todo  $\theta \in (0, \theta']$  y  $\mu_{k+1} = (1 - \theta)\mu_k$ , tenemos  $d(x^{k+1}, \mu_{k+1}) < \gamma d(x^k, \mu_{k+1})$ , para  $k = 0, 1, 2, \dots$  donde  $x^{k+1} = x^k + \Delta x$  y

$$\Delta x = \bar{X} \left[ I - \bar{X} A^T (A \bar{X}^2 A^T)^{-1} A \bar{X} \right] \bar{X} \begin{pmatrix} c \\ \mu \end{pmatrix} + \nabla \varphi(\bar{x})$$

los siguientes teoremas tomados de Sheu [17], completan nuestro análisis

**Teorema 4.13:** Bajo las condiciones 1 y 2, podemos escoger  $\theta = \min \left\{ \theta', \frac{\beta(1 - \gamma)}{2\gamma M + \beta} \right\}$  tal que

$$d(x^k, \mu_k) \leq \frac{\beta}{2} \text{ implica } d(x^k, \mu_{k+1}) \leq \frac{\beta}{2}, \text{ para } k = 0, 1, 2, \dots$$

**Prueba:** Recordemos del teorema 4.7

$$d(\bar{x}, \mu) = \min \left\{ \left\| \bar{X} \begin{pmatrix} 1 \\ \mu \end{pmatrix} s + \nabla \varphi(\bar{x}) \right\| : s = c - A^T y, y \in \mathbb{R}^m \right\}$$

Denotemos  $S(\bar{x}, \mu)$  como el mínimo del problema anterior. Entonces,



$$d(\bar{x}, \mu) = \left\| \bar{X} \left( \frac{1}{\mu} s(\bar{x}, \mu) + \nabla \varphi(\bar{x}) \right) \right\|$$

y

$$d(x^k, \mu_{k+1}) = \left\| X_k \left[ \frac{1}{\mu_{k+1}} x(x^k, \mu_{k+1}) + \nabla \varphi(x^k) \right] \right\| \quad (4.12)$$

para todo  $0 < \theta < 1$  y  $\mu_{k+1} = (1 - \theta)\mu_k$ . Note que (4.12) es verdadero debido a que tanto  $s(x^k, \mu_{k+1})$  y  $s(x^k, \mu_k)$  Satisface  $s = c - A^T y$  para algún  $y \in \mathbb{R}^m$  y  $s(x^k, \mu_{k+1})$  logra el valor optimal de  $d(x^k, \mu_{k+1})$ .

Por esto tenemos

$$\begin{aligned} & \left\| X_k \left[ \frac{1}{\mu_{k+1}} s(x^k, \mu_k) + \nabla \varphi(x^k) \right] \right\| \\ &= \left\| X_k \left[ \frac{1}{(1-\theta)\mu_k} s(x^k, \mu_k) + \frac{1}{1-\theta} \nabla \varphi(x^k) - \frac{1}{1-\theta} \nabla \varphi(x^k) + \nabla \varphi(x^k) \right] \right\| \\ &= \left\| \frac{1}{1-\theta} X_k \left( \frac{1}{\mu_k} s(x^k, \mu_k) + \nabla \varphi(x^k) \right) - X_k \left( \frac{\theta}{1-\theta} \nabla \varphi(x^k) \right) \right\| \\ &\leq \frac{1}{1-\theta} d(x^k, \mu_k) + \left( \frac{\theta}{1-\theta} \right) \left\| X_k \nabla \varphi(x^k) \right\| \\ &\leq \frac{1}{1-\theta} d(x^k, \mu_k) + \left( \frac{\theta}{1-\theta} \right) M \\ &\leq \left( \frac{1}{1-\theta} \right) \frac{\beta}{2} + \left( \frac{\theta}{1-\theta} \right) M. \end{aligned}$$

Por la condición 2, existe  $\theta'$  y  $\gamma$  tal que, para todo  $\theta \in (0, \theta']$ ,

$$d(x^{k+1}, \mu_{k+1}) \leq \gamma d(x^k, \mu_{k+1}) \leq \frac{\gamma\beta}{2(1-\theta)} + \left( \frac{\gamma\theta}{1-\theta} \right) M$$

ya que deseamos  $\frac{\gamma\beta}{2(1-\theta)} + \left(\frac{\gamma\theta}{1-\theta}\right)M \leq \frac{\beta}{2}$  ó equivalentemente,  $\gamma\beta + 2\gamma\theta M \leq \beta - \beta\theta$ ,

necesitamos  $\theta \leq \frac{\beta(1-\gamma)}{2\gamma M + \beta}$ .

Por esta razón, si  $\theta := \min \left\{ \theta', \frac{\beta(1-\gamma)}{2\gamma M + \beta} \right\}$ , entonces  $d(x^{k+1}, \mu_{k+1}) \leq \frac{\beta}{2}$ .

**Teorema 4.14:** Bajo las condiciones 1 y 2 con  $\mu_{k+1} = (1-\theta)\mu_k$  y  $\theta = \min \left\{ \theta', \frac{\beta(1-\gamma)}{2\gamma M + \beta} \right\}$ , si

el algoritmo 4.4.1 inicia desde  $\mu_0$  con  $d(x^0, \mu_0) < \beta$ , este algoritmo termina en tiempo acotado polinomialmente.

**Prueba:** Del teorema 4.13, vemos que si las condiciones 1 y 2 son satisfechas, el algoritmo 4.4.1, encuentra un  $x$  tal que  $d(x, \mu) \leq \beta$  cuando  $\mu$  cambia. Debido a esto, el número total de iteraciones requeridas para alcanzar la solución óptima es determinada solo por las iteraciones, que hacen decrecer  $\mu$  desde  $\mu_0$  hasta ser suficientemente pequeño. Suponga que

$\frac{\beta(1-\gamma)}{2\gamma M + \beta}$  es menor que  $\theta'$  y  $\gamma, \beta \in \langle 0, 1 \rangle$ , el único factor que queda para el análisis de la

complejidad es  $M$ , el cual depende de la dimensión del problema.

Además,  $M$  es la cota uniforme de  $\|X\nabla\phi(x)\|$  y cada componente de  $X\nabla\phi(x)$  está acotada

para  $x \in \mathcal{P}$ . De esta forma  $M \in O(\sqrt{n})$  y  $\theta \in O\left(\frac{1}{\sqrt{n}}\right)$ .

Sea  $c_0$  una constante tal que

$$\theta = \frac{1}{c_0\sqrt{n}} \leq \frac{\beta(1-\gamma)}{2\gamma M + \beta}.$$

De acuerdo a la regla de parada del algoritmo, éste termina cuando  $\mu_k \leq e^{-L}$ , para algún  $k$ . En

otras palabras  $\left(1 - \frac{1}{c_0 \sqrt{n}}\right)^k \mu_0 \leq e^{-L}$  o equivalentemente

$$k \ln \left(1 - \frac{1}{c_0 \sqrt{n}}\right) + \ln \mu_0 \leq -L,$$

$$\ln \mu_0 + L \leq -k \ln \left(1 - \frac{1}{c_0 \sqrt{n}}\right) \quad (4.13)$$

Ya que  $-k \ln \left(1 - \frac{1}{c_0 \sqrt{n}}\right) \geq \frac{k}{c_0 \sqrt{n}}$  como conocemos que  $\frac{k}{c_0 \sqrt{n}} \geq \ln \mu_0 + L$ , entonces

(4.13) se mantiene. En otras palabras, cuando  $k \geq c_0 \sqrt{n} (\ln \mu_0 + L)$ , tenemos  $\mu_k \leq e^{-L}$ .

Por otro lado si  $\theta' < \frac{\beta(1-\gamma)}{2nM + \beta}$ , ya que  $\frac{1}{\theta'} = O(n^s)$  con  $s > 0$ , con el mismo argumento podemos

probar que el algoritmo termina en a lo más  $O(n^s(\ln(\mu_0 + L)))$  pasos ■

#### 4.6 EJEMPLOS DE APLICACIÓN

Estudiemos ahora, en términos de las condiciones 1 y 2, los algoritmos de seguimiento de trayectoria con algunas funciones de barrera bien conocidas.

Para la función de barrera logarítmica, tenemos los siguientes resultados.

**Lema 4.15:** Si  $\varphi(x) = -\sum_{i=1}^n \ln x_i$ , entonces la condición 1 se mantiene.

**Prueba:** Como  $\nabla \varphi(x) = \left(-\frac{1}{x_1}, -\frac{1}{x_2}, \dots, -\frac{1}{x_n}\right)$  y además  $X \nabla \varphi(x) = (-1, -1, \dots, -1)$ , tenemos

que  $\|X \nabla \varphi(x)\| = \sqrt{n}$ . Tomando  $M = \sqrt{n}$ , tenemos  $\|X \nabla \varphi(x)\| \leq M, \forall x \in \mathcal{P}$  ■

**Lema 4.16:** Si  $\varphi(x) = - \sum_{i=1}^n \ln x_i$  y  $x = \bar{x} + \Delta x$ , entonces la desigualdad  $d(\bar{x}, \mu) \leq \beta < 1$  implica

que  $d(x, \mu) \leq d(\bar{x}, \mu)^2$

**Prueba:** Sea

$$\bar{t} = \bar{X} \left[ \frac{s(\bar{x}, \mu)}{\mu} - \bar{X}^{-1} e \right] \quad y$$

$\bar{T} = \text{diag}(\bar{t}_1, \bar{t}_2, \dots, \bar{t}_n)$ , entonces

$d(\bar{x}, \mu) = \|\bar{t}\|$  y  $x = \bar{x} - \bar{X}\bar{t}$  también,

$$\begin{aligned} d(\bar{x}, \mu) &= \left\| \bar{X} \left[ \frac{s(x, \mu)}{\mu} - \bar{X}^{-1} e \right] \right\| \\ &\leq \left\| \bar{X} \left[ \frac{s(\bar{x}, \mu)}{\mu} - \bar{X}^{-1} e \right] \right\| \\ &= \left\| \bar{X} (I - \bar{T}) \frac{s(\bar{x}, \mu)}{\mu} - e \right\| \\ &= \left\| \bar{X} \frac{s(\bar{x}, \mu)}{\mu} - e - \bar{X} \bar{T} \frac{s(\bar{x}, \mu)}{\mu} \right\| \\ &= \left\| \bar{t} - \bar{T} \bar{X} \frac{s(\bar{x}, \mu)}{\mu} \right\| \\ &= \left\| \bar{T} \left( e - \bar{X} \frac{s(\bar{x}, \mu)}{\mu} \right) \right\| \\ &= \left\| \bar{T}(-\bar{t}) \right\| \\ &= \left\| \bar{t}^t \bar{t} \right\| \leq \left\| \bar{t} \right\|^2 = d(\bar{x}, \mu)^2 \quad \blacksquare \end{aligned}$$

**Lema 4.17:** Si:  $d(x^k, \mu_k) \leq \frac{\beta}{2}$ , escogiendo  $\theta' = \frac{\beta}{2(\beta + M)}$  y  $\gamma = \beta$ , entonces

$\varphi(x) = -\sum_{i=1}^n \ln x_i$  satisface la condición 2.

**Prueba:** De la prueba del teorema 4.13, para cada  $0 < \theta < 1$ ,

$$d(x^k, \mu_{k+1}) \leq \left( \frac{1}{1-\theta} \right) \frac{\beta}{2} + \left( \frac{\theta}{1-\theta} \right) M$$

También del lema 4.16 tenemos

$$d(x^{k+1}, \mu_{k+1}) \leq d(x^k, \mu_{k+1})^2.$$

Si podemos probar que  $\left( \frac{1}{1-\theta} \right) \frac{\beta}{2} + \left( \frac{\theta}{1-\theta} \right) M \leq \beta$ , entonces, tomando  $\gamma = \beta$ , tenemos

$d(x^{k+1}, \mu_{k+1}) \leq d(x^k, \mu_{k+1})^2 \leq \gamma d(x^k, \mu_{k+1})$ . De aquí se tiene que  $\varphi(x) = -\sum_{i=1}^n \ln x_i$  satisface la

condición 2. Consecuentemente, para activar

$$\left( \frac{1}{1-\theta} \right) \frac{\beta}{2} + \left( \frac{\theta}{1-\theta} \right) M \leq \beta$$

necesitamos  $\theta \leq \frac{\beta}{2(\beta + M)}$ . Si escogemos  $\theta' = \frac{\beta}{2(\beta + M)}$ , entonces  $\frac{1}{\theta'} \in O(\sqrt{n})$  y

$$d(x^{k+1}, \mu_{k+1}) \leq \beta \quad \forall \theta \in (0, \theta'] \quad \blacksquare$$

En resumen los tres lemas anteriores, prueban que el algoritmo de seguimiento de trayectoria con función de barrera logarítmica converge a la solución óptima del problema lineal (1.1) en un tiempo acotado polinomialmente.

Para el algoritmo de seguimiento de trayectoria con función de barrera inversa

$$\varphi(x) = \frac{1}{r} \sum_{i=1}^n \frac{1}{x_i^r} \quad \text{para } r > 0.$$

Den Hertog [3] prueba la convergencia, pero la cota superior del número total de iteraciones necesarias a obtener una  $\varepsilon$ -solución optimal es no polinomial note que al gradiente de

$$\frac{1}{r} \sum_{i=1}^n \frac{1}{x_i^r} \text{ es } -X_e^{-r-1} \text{ y su matriz hessiana } \blacksquare$$

$H_\varphi$  es  $(r+1) X^{-r-2}$ . Escogiendo  $H_\varphi^{-1/2} = \frac{1}{\sqrt{r+1}} X^{\frac{r}{2}+1}$  como una matriz de escalado, vemos

que la condición 1 requiere que

$$\|H_\varphi^{-1/2} \nabla \varphi(x)\| = \left\| \frac{1}{\sqrt{r+1}} X^{-\frac{r}{2}} e \right\|$$

sea acotada. Obviamente, esto no es verdadero para todo  $r > 0$ , aún si escogemos  $X$  como la matriz de escalado,  $\|X \nabla \varphi(x)\| = \|X^{-r} e\|$  es no acotado. De este modo, la función de barrera

inversa falla en satisfacer la condición 1. Además,  $\frac{\beta(1-\gamma)}{2\gamma M + \beta}$  y  $\frac{\beta}{2(\beta + M)}$  requiere que  $M$

aparezca en el denominador, esto es no hace posible que un apropiado  $\theta > 0$  puede ser encontrado para una prueba de polinomialidad.

Para la función de barrera entrópica  $\varphi(x) = \sum_{i=1}^n x_i \ln x_i$ , si asumimos que el dominio factible

es acotado, puede verificarse fácilmente que la condición 1 es satisfecha.

De este modo al algoritmo 4.4.1 converge para todo  $\theta < \theta < 1$ . Desafortunadamente, la condición 2 no es satisfecha, así una prueba de polinomialidad no se puede dar.

## 4.7 CONCLUSIONES

El objetivo de esta tesis ha sido el de tratar de unificar la existencia de los algoritmos de seguimiento de trayectoria para la programación lineal (en inglés: path-following algorithms) mediante una estructura matemática fundamental.

Hemos caracterizado las trayectorias que se obtienen por introducir una clase de funciones de barrera generalizadas y entonces discutimos todas las componentes requeridas para diseñar algoritmos de seguimiento de trayectoria, hemos propuesto además un algoritmo genérico de seguimiento de trayectoria además, dos condiciones suficientes para convergencia y polinomialidad han sido establecidas.

La condición 1 requiere que  $\|X\nabla\varphi(x)\|$  este acotado y esto es crucial en establecer las propiedades de convergencia así como también las de polinomialidad. Nótese que, para un problema aumentado  $(P_\mu)$ , las condiciones K-K-T son

$$\begin{cases} A^T y + s - c = 0, \\ Ax - b = 0, \\ \frac{1}{\mu} s = -\nabla\varphi(x), \\ x > 0, \end{cases}$$

la cual puede ser interpretada como al problema de programación lineal original (1.1) con una 'brecha dual estimada'  $\|Xs\| = \|\mu X\nabla\varphi(x)\|$ . Para ser mas preciso, cada vez que escogemos una función GBLP, estamos introduciendo una brecha dual  $\|X\nabla\varphi(x)\|$  en el problema de programación lineal (1.1).

Desde que los algoritmos de seguimiento de trayectoria disminuyen la brecha dual hasta que tienda a cero, esto hace sensato escoger una función de barrera generalizada con una 'brecha dual estimada' acotada. La razón por la cual Den Hertog [3] falla en encontrar una cota polinomial por usar la función de barrera inversa, es debido al hecho de que esta función incrementa la brecha dual a infinito de modo tal que esta no puede ser reducida a cero eficientemente, de este modo la condición 1 no se cumple.

Los algoritmos de seguimiento de trayectoria con una función de barrera logarítmica toma  $\mathcal{O}(\sqrt{n}L)$  iteraciones para converger. Esta es probablemente la mejor cota para convergencia