

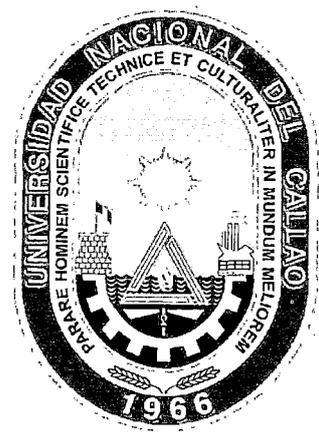
74

JUN 2014



UNIVERSIDAD NACIONAL DEL CALLAO
 FACULTAD DE INGENIERÍA QUÍMICA
 INSTITUTO DE INVESTIGACIÓN DE LA FACULTAD DE
 INGENIERÍA QUÍMICA

UNIVERSIDAD NACIONAL DEL CALLAO
 VICE-RECTORADO DE INVESTIGACIÓN
 14 MAY 2014
 16:15



1119
 Almey
 05-105-2014
 10-30h.
 171.
 UNIVERSIDAD NACIONAL DEL CALLAO
 VICE-RECTORADO DE INVESTIGACIÓN
RECIBIDO
 216
 14 MAYO 2014
 CENTRO DE DOCUMENTACION
 CIENTIFICA Y TRADUCCIONES

INFORME FINAL DEL PROYECTO DE INVESTIGACIÓN
 “BARRERAS AUTO CONCORDANTES Y SU
 APLICACIÓN EN EL MÉTODO DE PUNTO
 INTERIOR: PRIMAL DUAL”

AUTOR: SANTOS PANTALEÓN RODRÍGUEZ CHUQUIMANGO

PERIODO DE EJECUCIÓN: Del 01 de junio del 2012 al 31 de mayo del 2014.

Resolución de aprobación N° 502-2012-R

Callao, 2014

INDICE

Resumen	1
Abstract	2
CAPITULO I: INTRODUCCIÓN	
1.1. Situación problemática	3
1.2. Formulación del problema	5
1.3. Importancia y justificación de la investigación	5
CAPÍTULO II: MARCO TEÓRICO	
2.1. Tópicos de programación lineal	6
2.1.1 Función convexa	
2.1.2 Condición de primer orden	
2.1.3 Condición de segundo orden	
2.1.1 Conjuntos subnivel	
2.1.2 Operaciones que preservan convexidad	
2.1.3 Problema de optimización	
2.1.4 Teorema fundamental de dualidad	
2.1.5 Optimización convexa	
2.2. Funciones auto concordantes y el método de Newton	9
2.2.1. Funciones auto concordantes en R	10
2.2.2. Funciones auto concordantes en R^n	12
2.2.3. Operaciones con funciones auto concordantes	13
a) Escalamiento	
b) Suma	
c) Composición con una función afín	
d) Norma	
2.2.4. El decremento de Newton para funciones auto concordantes	15
2.2.5. Cota superior e inferior sobre la segunda derivada	16
2.2.6. Cota sobre sub optimalidad	17
2.2.7. Análisis del método de Newton para funciones auto concordantes	20
a) Fase Newton amortiguado	



b) Fase convergencia cuadrática	
c) Cota final de complejidad	
2.3. Métodos de punto interior	26
2.3.1. Problema convexo	27
2.3.2. Curva central	30
2.3.3. Puntos duales de la curva central	32
2.4. El método barrera	34
2.4.1. Algoritmo	35
2.4.2. Exactitud de centrado	35
2.4.3. Análisis de complejidad	36
2.5. Métodos de punto interior primal dual	41
2.5.1. Dirección de búsqueda primal dual	41
2.5.2. Comparación con direcciones de búsqueda del método barrera	43
2.5.3. La dualidad gap sustituta	46
2.5.4. Algoritmo del método de punto interior primal dual	46
2.5.5. Línea de búsqueda	47
2.5.6. Implementación	48
a) Problemas espase	
b) Programación lineal en forma estándar	
CAPÍTULO III : MATERIALES Y MÉTODOS	51
CAPÍTULO IV: RESULTADOS	52
CAPÍTULO V: DISCUSION	54
REFERENCIALES	56
APÉNDICE	57
ANEXO	60



GRÁFICOS

Gráfico 2.1.	Comparación con la función cuadrática	19
Gráfico 2.2.	Función estrictamente decreciente	23
Gráfico 2.3.	Comparación con la función cúbica	25
Gráfico 2.4.	Función indicatriz	29
Gráfico 2.5	Comportamiento cuadrático y lineal	40
Gráfico 5.1.	Comportamiento lineal y cuadrático	55

A handwritten signature in black ink, located in the bottom right corner of the page. The signature is stylized and appears to read "Sant'Anna".

“A ti, oh Dios de mis padres te doy
gracias y te alabo, porque me has
dado sabiduría y fuerza...”

Profeta Daniel

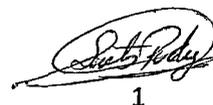
A handwritten signature in black ink, appearing to read 'Profeta Daniel', written in a cursive style.

RESUMEN

En el presente trabajo de investigación haremos una presentación de la teoría de funciones auto concordantes y su aplicación en el análisis de la convergencia del método de punto interior: primal dual. Los métodos de punto interior se caracterizan porque hacen uso del método de Newton para minimizar un problema de programación lineal irrestricto, el cual tiene complejidad polinomial. El análisis clásico de convergencia del método de Newton depende de la convexidad fuerte de la función objetivo que incluye constantes desconocidas en la mayor parte de problemas y de la constante de Lipchitz para la Hessiana de la función objetivo. En el trabajo se ha usado la teoría de funciones auto concordantes para el análisis de la convergencia del método de Newton y consecuentemente del método de punto interior primal dual.

En el marco teórico se presenta la teoría básica y fundamental de las funciones o barreras auto concordantes comenzando con funciones de una sola variable y generalizando la definición para una función de varias variables. Se ha demostrado la convergencia y la complejidad polinomial del método de Newton que no depende de la función objetivo ni de constantes desconocidas como es el caso del análisis de convergencia tradicional.

Se presenta el método de punto interior primal dual y se establece su convergencia y complejidad polinomial en base a la convergencia y complejidad polinomial del método de Newton y el método barrera.



Abstract

In the present research we introduced the theory of self concordant functions and their application in the analysis of convergence of the interior point method: primal dual. The interior point methods are characterized because they make use of Newton's method to minimize a problem of unrestricted linear programming, which has polynomial complexity. The classical convergence analysis of Newton's method depends on the strong convexity of the objective function that includes the unknown constants in most problems and constant Lipschitz for the Hessian of the objective function. At work we have used the self concordant barriers to convergence analysis of Newton's method and consequently the method of primal dual interior point theory.

Basic and fundamental theory of functions and barriers self concordant starting with a single variable functions and generalizing the definition for a function of several variables is presented in the theoretical framework. It has been shown convergence and complexity polynomial Newton method which is not dependent on the objective function or unknown constants such as traditional convergence analysis.

The method of primal dual interior point is presented and its convergence and polynomial complexity are established based on the convergence and polynomial complexity of the method and the Newton barrier method.

Key Word: interior point Method, self concordant Barrier.



2

CAPITULO I

INTRODUCCION

1.1. Situación problemática

En muchos problemas de la ciencia, la técnica y la economía se tratan de asignar o compartir determinados recursos, disponibles sólo en cantidades limitadas, los mismos que se requiere optimizar. La Programación Lineal es una parte de la programación matemática, que busca encontrar aquella solución que mejor identifica un determinado criterio lineal atendiendo a diversas condiciones también lineales, entre las muchas posibles de un problema.

Desde la aparición del método Simplex en 1947 para resolver problemas lineales de grandes dimensiones, ha habido un gran desarrollo de métodos para solucionar no solamente problemas lineales sino también los no lineales. A pesar que el método simplex ha demostrado ser altamente eficaz para resolver todo tipo de problemas de programación lineal, el hecho de que en determinadas circunstancias su complejidad computacional sea exponencial ha motivado en los últimos años un elevado número de intentos, tanto teóricos como prácticos, de obtener otros algoritmos con un mejor comportamiento computacional, algunos de estos son los métodos de punto interior.

Los métodos de punto interior tratan de encontrar la solución del problema de optimización a través de puntos en el interior de la región factible, a diferencia del método simplex que lo hace en los vértices de la región. Los procedimientos numéricos de puntos interiores para resolver problemas de programación lineal basan su estrategia en conseguir solucionar las tres cuestiones siguientes:

- i) Encontrar un punto de partida factible en el interior de la región factible del problema.
- ii) Definir una dirección de movimiento tal que, conservando la factibilidad del problema, moviéndose a lo largo de ella, se avance hasta un punto en el que se reduzca el valor de la función objetivo (si se está minimizando).
- iii) Obtener un criterio de parada, es decir, cuantas veces se debe realizar la operación descrita en el punto (ii) y cómo identificar que se ha alcanzado el óptimo del problema.

Uno de los aspectos importantes en la elaboración de algoritmos para solucionar problemas de optimización es la convergencia y la complejidad computacional del algoritmo. Que un algoritmo sea convergente significa que a través de la iteración propuesta, este converge al óptimo del problema. La complejidad se refiere al costo computacional o el número de pasos que se requiere para llegar a la solución del problema.

Los métodos de punto interior hacen uso del método de Newton para un problema irrestricto. Los resultados tradicionales para el método de Newton establecen su convergencia cuadrática, asumiendo razonables supuestos de suavidad y no degeneración. Estos resultados, poseen un marco genérico conceptual: la descripción cuantitativa de la región de convergencia cuadrática, así mismo, la convergencia es dada en términos del número de condición de la Hessiana de la función objetivo en el minimizador y la constante de Lipschitz de esta Hessiana. Estas cantidades son “marco-dependientes”: Están definidas no solamente por la función objetivo misma, sino también por la estructura euclidiana en el espacio de las variables. Cuando cambiamos la estructura euclidiana, el gradiente y la Hessiana están sujetos a cierta transformación que no mantienen invariantes las cantidades. Como resultado, la descripción tradicional del comportamiento del método depende no solo de la función objetivo sino también de la estructura



euclidiana usada en la descripción; esto contradice la naturaleza afín invariante del método. Para superar este inconveniente se usa de la teoría de funciones o barreras autoconcordantes.

1.2. Formulación del problema

¿Cómo las propiedades básicas de barreras auto concordantes conducen a demostrar la convergencia y la complejidad polinómica del algoritmo provisto por el método de punto interior: primal dual?

El objetivo en el trabajo es utilizar las propiedades de barreras autoconcordantes para demostrar la convergencia del método de punto interior: primal dual y la complejidad polinómica del mismo.

1.3. Importancia y justificación de la investigación

El estudio de este tema es importante desde el punto de vista teórico porque nos permite conocer las propiedades de barreras auto concordantes y su aplicación a la optimización convexa.

Desde el punto vista práctico nos permite demostrar la convergencia del método de punto interior: primal dual y su complejidad polinomial; nos permite en suma corroborar la utilidad práctica de los conceptos y teorías matemáticas en las diferentes aplicaciones, especialmente en la optimización y el diseño de algoritmos. De hecho la mayor parte de investigadores en este tema están dedicados a mejorar la complejidad de los algoritmos mas que en la implementación de los métodos.



CAPÍTULO II

MARCO TEORICO

2.1. Tópicos de programación lineal

2.1.1 Función convexa

Una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa si $\text{dom}f$ es un conjunto convexo y si para todo $x, y \in \mathbb{R}^n, 0 \leq t \leq 1$ se tiene

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

Decimos que la función es estrictamente convexa si se cumple la desigualdad estricta, $x \neq y$, y $0 < t < 1$.

2.1.2 Condición de primer orden

Supongamos que f es diferenciable (es decir, el ∇f existe en cada punto del $\text{dom}f$ que es abierto), entonces f es convexa si y solamente si $\text{dom}f$ es convexo y

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

La función afín de y dada por $f(x) + \nabla f(x)^T(y - x)$ es de hecho la aproximación de Taylor de primer orden de f cerca de x . La desigualdad establece que para una función convexa, la aproximación de Taylor de primer orden es un estimador global de la función y recíprocamente. En otras palabras, de información local acerca de una función, podemos deducir información global. Así, por ejemplo, si $\nabla f(x) = 0$, entonces para todo y en el $\text{dom}f$, $f(y) \geq f(x)$, es decir x es un minimizador global de la función f .

2.1.3 Condición de segundo orden

Asumiendo que f es dos veces continuamente diferenciable, es decir la matriz hessiana de f existe en cada punto del $\text{dom}f$, que es un conjunto



abierto, entonces f es convexa si y solamente si f es convexa y su matriz hessiana es semidefinida positiva, es decir $\nabla^2 f(x) \succeq 0$.

2.1.4 Conjuntos subnivel

El subconjunto de nivel de una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ está definido como

$$C_\alpha = \{x \in \text{dom}f / f(x) \leq \alpha\}$$

Los conjuntos sub nivel de funciones convexas son convexas, para cualquier valor de α . El recíproco no es verdad. Una función puede tener todos sus subconjuntos sub nivel convexas pero la función no es convexa.

2.1.5 Operaciones que preservan convexidad

- a) Si f es convexa y $\alpha \geq 0$, entonces la función αf es convexa.
- b) Si f_1, \dots, f_m son funciones convexas, w_1, \dots, w_m son números reales, entonces $w_1 f_1 + \dots + w_m f_m$ es una función convexa.
- c) Si $f: \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa y $Ax + b$ es una aplicación afin, entonces $f(Ax + b)$ es convexa.
- d) Si f_1, f_2 son funciones convexas, entonces $\max \{f_1(x), f_2(x)\}$ es convexa.

2.1.6 Problema de optimización

- a) Problema primal lineal: Es un problema de la forma
Minimizar $c^T x$
Sujeto a: $Ax = b, x \succeq 0$



Donde $c, x \in \mathbb{R}^n, b \in \mathbb{R}^m, A$ es una matriz de orden $m \times n$.

Si x satisface la restricción $Ax = b, x \geq 0$, llamamos a x un punto factible. El conjunto de todos los puntos factibles se llama conjunto factible.

Asociado con cualquier programa lineal hay otro programa lineal llamado el dual, que consiste de los mismos datos reordenados en una diferente forma. (Ver Stephen J. Wright(5), capítulo 1, página 3)

b) Problema dual

Maximizar $b^T \lambda$

Sujeto a : $A^T \lambda + s = c, s \geq 0$

Donde $\lambda \in \mathbb{R}^m, s \in \mathbb{R}^n$. Llamamos a las componentes de λ las variables duales, mientras que s es el vector dual de variables libres. El problema dual puede ser establecido mas compactamente al eliminar la variable s , reescribiendo las restricciones como $A^T \lambda \leq c$.

2.1.7 Teorema fundamental de dualidad

Un problema primal tiene solución si y solo si el problema dual lo tiene. Además si existe una solución, entonces, las funciones objetivo de los problemas primal y dual alcanzan el mismo valor óptimo

2.1.8 Optimización convexa

a) Considere el problema de optimización

$$P: \begin{cases} \text{optimizar } f(x) \\ \text{sujeto a : } x \in X \subset \mathbb{R}^n \end{cases}$$

b) Si P es un problema convexo, entonces el conjunto de mínimos globales de notado por X^* es convexo.



- c) Si P es un problema convexo y f es estrictamente convexa, entonces, si el mínimo existe, es único
- d) Considere el problema convexo

$$M: \begin{cases} \text{minimizar } f(x) \\ \text{sujeto a : } x \in X \subset \mathbb{R}^n \end{cases}$$

Donde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ es diferenciable en todo \mathbb{R}^n

- e) x^* es un mínimo sí y solamente sí $\nabla f(x^*)(x - x^*) \geq 0$

2.2. Funciones auto concordantes y el método de Newton

Hay dos principales deficiencias del análisis clásico de convergencia del método de Newton. El primero es práctico: La complejidad resultante envuelve las tres constantes m , M , L que casi nunca se conocen en la práctica. Como resultado la cota sobre el número de iteraciones requerida para el paso de Newton es desconocida. De hecho el análisis de la convergencia y complejidad es aun conceptualmente útil.

La segunda desventaja es que mientras, el paso de Newton es afin invariante, el análisis clásico de convergencia del método de Newton depende del sistema de coordenadas usado. Si cambiamos las coordenadas las constantes m , M , L también cambian. Debemos ver, aunque sea por una razón estética, un análisis del método que sea independiente del cambio de coordenadas, en otras palabras veremos una alternativa a las condiciones

$$ml \leq \nabla^2 f(x) \leq MI, \quad \|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2$$

que sea independiente o afin al cambio de coordenadas y también nos conduzca al análisis del método de Newton.



Una simple y elegante forma que cumple este propósito fue descubierto por Nesterov y Nemirovskii, quienes dieron el nombre de auto concordancia a sus condiciones. Funciones auto concordantes son importantes por varias razones

- i) Ellas incluyen muchas de las barreras logarítmicas que juegan un rol importante en métodos de punto interior para resolver problemas de optimización convexa.
- ii) El análisis del método de Newton para barreras auto concordantes no depende de constantes desconocidas.
- iii) Auto concordancia es una propiedad afín invariante, es decir, si aplicamos una transformación lineal de variables a funciones auto concordantes, obtenemos funciones auto concordantes. Por lo tanto la complejidad estimada que obtenemos para el método de Newton aplicado a funciones auto concordantes es independiente de cambio de coordenadas.

2.2.1. Funciones auto concordantes en R

Una función convexa $f: R \rightarrow R$ es auto concordante si $\forall x \in \text{dom } f$

$$|f'''(x)| \leq 2f''(x)^{3/2}$$

Ejemplo 1.

Como para funciones cuadráticas y lineales la tercera derivada es cero, es evidente que son auto concordantes.

Ejemplo 2.

Las siguientes funciones son auto concordantes

a) $f(x) = -\log x,$

b) $f(x) = x \log x - \log x$

Probemos el ejemplo en b). Haciendo los cálculos se tiene:

$$f''(x) = \frac{x+1}{x^2}, \quad f'''(x) = -\frac{x+2}{x^3}$$

Luego,

$$\frac{|f'''(x)|}{2f''(x)^{3/2}} = \frac{x+2}{2(x+1)^{3/2}} = g(x)$$

La función g en el lado derecho alcanza su máximo en $x = 0$ y además $g(0) = 1$, por lo tanto la propiedad de auto concordancia se cumple.

Debemos hacer dos importantes observaciones acerca de la definición de función autoconcordante. La primera es concerniente a la misteriosa constante 2 que aparece en la definición. Esta constante se escoge por conveniencia, a fin de simplificar las fórmulas más tarde; cualquier otra constante puede usarse.

Supongamos por ejemplo, que la función convexa $f: R \rightarrow R$ satisface

$$|f'''(x)| \leq kf''(x)^{3/2}$$

donde k es una constante positiva. Entonces la función,

$$\tilde{f}(x) = \left(\frac{k^2}{4}\right) f(x)$$

satisface

$$\begin{aligned} |\tilde{f}'''(x)| &= \left(\frac{k^2}{4}\right) |f'''(x)| \\ &\leq \left(\frac{k^3}{4}\right) f''(x)^{3/2} \\ &= \left(\frac{k^3}{4}\right) \left((4/k^2)\tilde{f}''(x)\right)^{3/2} \\ &= 2\tilde{f}''(x)^{3/2} \end{aligned}$$

y por lo tanto es auto concordante. Así, lo importante es que la tercera derivada de la función sea acotada por algún múltiplo de la $3/2$ potencia



de la segunda derivada, con un escalamiento apropiado, podemos cambiar el múltiplo por la constante 2.

El segundo comentario es un simple cálculo que demuestra porque auto concordancia es tan importante: esta es afín invariante. Supongamos que definimos la función $g(y) = f(ay + b)$, $a \neq 0$, entonces g es auto concordante si y solo si f lo es.

En efecto,

$$g''(y) = a^2 f''(x), \quad g'''(y) = a^3 f'''(x)$$

Asumiendo que f es auto concordante se tiene

$$\begin{aligned} |g'''(y)| &= a^3 |f'''(x)| \\ &\leq a^3 2 f''(x)^{3/2} \\ &= a^3 2 \left(\frac{1}{a^2} g''(y) \right)^{3/2} \\ &= 2 g''(y)^{3/2} \end{aligned}$$

Estrictamente hablando, la condición de auto concordancia es una forma de limitar la tercera derivada de una función de tal forma que sea independiente de cambio afín de coordenadas.

2.2.2. Funciones auto concordantes en R^n

Consideraremos ahora funciones sobre R^n . Decimos que una función $f: R^n \rightarrow R$ es auto concordante, si es auto concordante en cada línea de su dominio, es decir, si la función $\tilde{f}(t) = f(x + tv)$ es un función autoconcordante de t para todo $x \in \text{dom } f$ y para todo $v \in R^n$.

Se tiene una definición equivalente (Ver Arcadii Nemirovskii (3), Capítulo 2, página 21). Sea $\emptyset \neq Q \subset R^n$ un conjunto convexo, abierto y



$\tilde{F} \in C^3(Q)$ una función convexa suave definida sobre Q . F es llamada auto concordante sobre Q si posee las siguientes dos propiedades:

- 1) Propiedad de barrera: $F(x_i) \rightarrow \infty$ a través de cada sucesión $\{x_i \in Q\}$ convergente a un punto frontera de Q , cuando $i \rightarrow \infty$.
- 2) Desigualdad de auto concordancia: F satisface la desigualdad diferencial

$$|D^3F(x)[h, h, h]| \leq 2(D^2F(x)[h, h])^{3/2} \quad (2.1)$$

para todo $x \in Q$ y todo $h \in R^n$. Donde

$$D^kF(x)[h_1, \dots, h_k] \equiv \left. \frac{\partial^k}{\partial t_1 \dots \partial t_k} \right|_{t_1=\dots=t_k=0} F(x + t_1h_1 + \dots + t_kh_k)$$

denota la diferencial de F de orden k en x en las direcciones h_1, \dots, h_k .

2.2.3. Operaciones con funciones auto concordantes

a) Escalamiento

Auto concordancia es preservada por escalamiento por un factor mayor que uno: si f es auto concordante y $a \geq 1$, entonces af es auto concordante.

b) Suma.

Auto concordancia es preservada por la adición: Si f_1, f_2 son auto concordantes, entonces $f_1 + f_2$ es auto concordante.

En efecto, para demostrar esto es suficiente considerar dos funciones $f_1, f_2: R \rightarrow R$ auto concordantes. Usando la desigualdad triangular y la auto concordancia se tiene

$$\begin{aligned} |f_1'''(x) + f_2'''(x)| &\leq |f_1'''(x)| + |f_2'''(x)| \\ &\leq 2(f_1''(x)^{3/2} + f_2''(x)^{3/2}) \end{aligned}$$



$$\leq 2(f_1''(x) + f_2''(x))^{3/2}$$

Para el último paso se usó la desigualdad

$$(u^{3/2} + v^{3/2})^{2/3} \leq u + v, \quad \forall u, v \geq 0$$

Verifiquemos esta propiedad, elevando al cuadrado en

$$u^{3/2} + v^{3/2} \leq (u + v)^{3/2}$$

$$u^3 + v^3 + 2u^{3/2}v^{3/2} \leq u^3 + v^3 + 3u^2v + 3uv^2$$

$$2u^{3/2}v^{3/2} \leq 3u^2v + 3uv^2$$

$$2u^{1/2}v^{1/2} \leq 3(u + v)$$

Elevando al cuadrado en la última desigualdad se tiene

$$4uv \leq 9u^2 + 9v^2 + 18uv$$

$$0 \leq 9u^2 + 9v^2 + 14uv$$

Aplicando discriminante

$$\Delta = -14^2v^2 - 4(81)v^2 < 0$$

lo que prueba la identidad.

c) Composición con una función afín

Si $f: R^n \rightarrow R$ es autoconcordante, y $A \in R^{n \times m}$, $b \in R^n$, entonces,

$F^+(x) = f(Ax + b)$ es autoconcordante.

La prueba es directa de la definición

$$DF^+(x)[h] = Df(Ax + b)A$$

$$D^2F^+(x)[h, h] = D^2f(Ax + b)A^2$$

$$D^3F^+(x)[h, h, h] = D^3f(Ax + b)A^3$$



$$\begin{aligned}
|D^3F^+(x)[h, h, h]| &= |D^3f(Ax + b)A^3| \\
&\leq 2(D^2f(Ax + b))^{3/2}A^3 \\
&= 2(D^2F^+(x)[h, h]A^{-2})^{3/2}A^3 \\
&= 2(D^2F^+(x)[h, h])^{3/2}
\end{aligned}$$

Ejemplo 3. Barrera logarítmica para desigualdades lineales.

La función

$$f(x) = -\sum_{i=1}^n \log(b_i - a_i^T x)$$

Con $\text{dom } f = \{x: a_i^T x < b_i, i = 1, 2, \dots, m\}$ es auto concordante. Cada término $-\log(b_i - a_i^T x)$ es la composición de $-\log y$ con una función afin $b_i - a_i^T x$ y por lo tanto auto concordante, y por ende la suma también es auto concordante. Esta función es llamada barrera logarítmica para desigualdades lineales.

d) Norma

Sea Q un conjunto convexo abierto en R^n y F auto concordante sobre Q . Para $x \in Q, h, g \in E$, definimos el producto escalar y la norma por

$$\langle g, h \rangle_x = D^2F(x)[g, h], \quad |h|_x = \langle h, h \rangle_x^{1/2} = \sqrt{D^2F(x)[h, h]}$$

De manera general, $|\cdot|_x$ es una seminorma euclidiana sobre E , es una norma si y solo si $D^2F(x)$ es no singular y definida positiva.

2.2.4. El decremento de Newton para funciones auto concordantes

En el análisis clásico de convergencia se hace uso de convexidad fuerte para deducir cotas de sub optimalidad en un punto x en términos de la



norma del gradiente en x . Para funciones auto concordantes estrictamente convexas, podemos obtener similares cotas en términos del decremento de Newton

Para una función auto concordante el decremento de newton se define por

$$\lambda(x) = (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2}$$

(Se puede demostrar que la Hessiana de una función auto concordante estrictamente convexa es definida positiva). Al igual que las cotas basadas en la norma del gradiente, las cotas basadas en el decremento de Newton no son afectadas por cambio afín de coordenadas.

Notemos que el decremento de Newton puede también darse como

$$\lambda(x) = \max_{v \neq 0} \frac{-v^T \nabla f(x)}{(v^T \nabla^2 f(x) v)^{1/2}}$$

En otras palabras tenemos que para cualquier $v \neq 0$

$$\lambda(x) \geq \frac{-v^T \nabla f(x)}{(v^T \nabla^2 f(x) v)^{1/2}} \quad (2.2)$$

La igualdad se cumple cuando $v = N = -\nabla^2 f(x)^{-1} \nabla f(x)$ (el paso de Newton)

2.2.5. Cota superior e inferior sobre la segunda derivada

Supongamos que $f: R \rightarrow R$ es una función auto concordante estrictamente convexa. Podemos escribir la desigualdad de auto concordancia como (ver Stephen Boyd (5) capítulo 9, página 501)

$$\left| \frac{d}{dt} (f''(t)^{-1/2}) \right| \leq 1, \quad \text{para todo } t \in \text{dom } f.$$



Integrando entre 0 y t obtenemos

$$-t \leq \int_0^t \frac{d}{ds} (f''(s)^{-1/2}) ds \leq t$$

Es decir

$$-t \leq f''(t)^{-1/2} - f''(0)^{-1/2} \leq t$$

De esto tenemos cotas superior e inferior para $f''(t)$

$$\frac{f''(0)}{\left(1 + tf''(0)^{\frac{1}{2}}\right)^2} \leq f''(t) \leq \frac{f''(0)}{\left(1 - tf''(0)^{\frac{1}{2}}\right)^2} \quad (2.3)$$

La cota inferior es válida para todo $t \in \text{dom } f$ no negativo; la cota superior es válida si $0 \leq t < f''(0)^{-1/2}$.

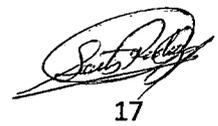
2.2.6. Cota sobre suboptimalidad

Sea $f: R^n \rightarrow R$ una función auto concordante estrictamente convexa, y sea v una dirección descendente (es decir, cualquier dirección satisfaciendo $v^T \nabla f(x) < 0$, no necesariamente la dirección de Newton). Definamos $\tilde{f}: R \rightarrow R$ como $\tilde{f}(t) = f(x + tv)$. Por definición, la función \tilde{f} es auto concordante.

Integrando la cota inferior en (2.3) tenemos una cota para $\tilde{f}'(t)$

$$\tilde{f}'(t) \geq \tilde{f}'(0) + \tilde{f}''(0)^{\frac{1}{2}} - \frac{\tilde{f}''(0)^{1/2}}{1 + t\tilde{f}''(0)^{\frac{1}{2}}} \quad (2.4)$$

Integrando de nuevo tenemos una cota para $\tilde{f}(t)$:



$$\tilde{f}(t) \geq \tilde{f}(0) + t\tilde{f}'(0) + t\tilde{f}''(0)^{\frac{1}{2}} - \log\left(1 + t\tilde{f}'''(0)^{\frac{1}{2}}\right) \quad (2.5)$$

El lado derecho de (2.5) alcanza su mínimo en

$$\tilde{t} = \frac{-\tilde{f}'(0)}{\tilde{f}''(0) + \tilde{f}''(0)^{1/2}\tilde{f}'(0)}$$

Evaluando en \tilde{t} provee una cota inferior para $\tilde{f}(t)$:

$$\begin{aligned} \inf_{t \geq 0} \tilde{f}(t) &\geq \tilde{f}(0) + \frac{-\tilde{f}'(0)\tilde{f}'(0)}{\tilde{f}''(0) + \tilde{f}''(0)^{1/2}\tilde{f}'(0)} + \frac{-\tilde{f}'(0)\tilde{f}''(0)^{\frac{1}{2}}}{\tilde{f}''(0) + \tilde{f}''(0)^{1/2}\tilde{f}'(0)} \\ &\quad - \log\left(1 + \frac{-\tilde{f}'(0)\tilde{f}''(0)^{\frac{1}{2}}}{\tilde{f}''(0) + \tilde{f}''(0)^{1/2}\tilde{f}'(0)}\right) \\ &= \tilde{f}(0) - \tilde{f}'(0)\tilde{f}''(0)^{-\frac{1}{2}} + \log\left(1 + \tilde{f}'(0)\tilde{f}''(0)^{-\frac{1}{2}}\right) \end{aligned}$$

La desigualdad (2.2) puede ser expresada como

$$\lambda(x) \geq -\tilde{f}'(0)\tilde{f}''(0)^{-\frac{1}{2}}$$

(con igualdad cuando $v = N$), puesto que se tiene

$$\tilde{f}'(0) = v^T \nabla f(x), \quad \tilde{f}''(0) = v^T \nabla^2 f(x) v.$$

Ahora, usando el hecho que $u + \log(1 - u)$ es monótona decreciente, tenemos

$$\inf_{t \geq 0} \tilde{f}(t) \geq \tilde{f}(0) + \lambda(x) + \log(1 - \lambda(x))$$

Esta desigualdad se cumple para cualquier dirección descendente v . Por lo tanto



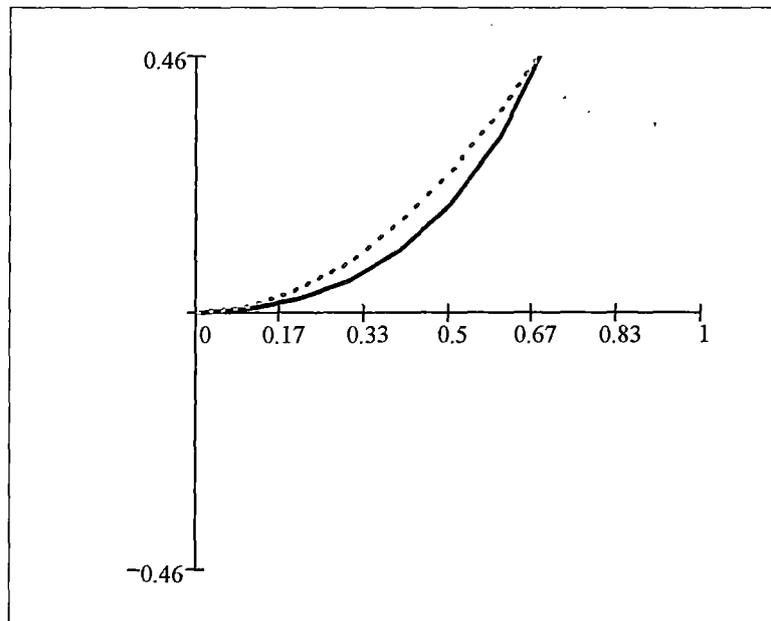
$$p^* \geq \tilde{f}(0) + \lambda(x) + \log(1 - \lambda(x)), \quad \lambda(x) < 1$$

Graficando $-\lambda - \log(1 - \lambda)$ se observa que

$$-\lambda - \log(1 - \lambda) \approx \frac{\lambda^2}{2},$$

Gráfico 2.1.

Comparación con la función cuadrática



Gráfica de $-\lambda - \log(1 - \lambda)$ en línea sólida y λ^2 en línea punteada. Gráfica elaborada con el software Mathcad.

Para λ pequeño, y la cota

$$-\lambda - \log(1 - \lambda) \leq \lambda^2$$

con $\lambda \leq 0.68$, tenemos la cota para sub optimalidad

$$p^* \geq f(x) - \lambda(x)^2 \quad (2.6)$$

válida para $\lambda(x) \leq 0.68$.

Recalcando que $\frac{\lambda^2}{2}$ es el estimado de $f(x) - p^*$ basado sobre el modelo cuadrático en x ; la desigualdad (2.6) demuestra que para funciones auto concordantes, duplicando esta estimación nos da una probable cota. En particular esto demuestra que para funciones auto concordantes, podemos usar el criterio de parada

$$\lambda(x)^2 \leq \epsilon$$

(donde $\epsilon < 0.68^2$), y garantiza que $f(x) - p^* < \epsilon$. (Ver Stephen Boyd (5) capítulo 9, página 502)

2.2.7. Análisis del método de Newton para funciones auto concordantes

Haremos el análisis del método de Newton usando una línea de búsqueda con retroceso, cuando lo aplicamos a una función auto concordante estrictamente convexa. Asumiremos que se conoce el punto de inicio $x^{(0)}$, y que el conjunto subnivel

$$S = \{x: f(x) \leq f(x^{(0)})\}$$

es cerrado. También asumiremos que f es acotada inferiormente. (esto implica que tiene un mínimo en x^*).

El análisis es muy similar al análisis clásico, excepto que usamos auto concordancia como la condición básica en vez de convexidad fuerte y la condición de Lipschitz de la Hessiana, el decremento de Newton juega el rol de la norma del gradiente. Demostraremos que existen números



$\gamma, \eta > 0$, con $0 < \eta \leq 1/4$ que dependen solamente de los parámetros de la línea de búsqueda α, β tal que, lo siguiente se cumple:

- Si $\lambda(x^{(k)}) > \eta$, entonces

$$f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma \quad (2.7)$$

- Si $\lambda(x^{(k)}) < \eta$, entonces la línea de búsqueda con retroceso selecciona $t = 1$, y

$$2\lambda(x^{(k+1)}) \leq (2\lambda(x^{(k)}))^2 \quad (2.8)$$

La segunda condición puede ser aplicada recursivamente, concluyendo que para todo $l \geq k$, se tiene $\lambda(x^{(l)}) \leq \eta$, y

$$2\lambda(x^{(l)}) \leq (2\lambda(x^{(k)}))^{2^{l-k}} \leq (2\eta)^{2^{l-k}} \leq \left(\frac{1}{2}\right)^{2^{l-k}}$$

Como consecuencia, para todo $l \geq k$,

$$f(x^{(l)}) - p^* \leq \lambda(x^{(l)})^2 \leq \frac{1}{4} \left(\frac{1}{2}\right)^{2^{l-k+1}} \leq \left(\frac{1}{2}\right)^{2^{l-k+1}}$$

Y entonces $f(x^{(l)}) - p^* \leq \epsilon$, si $l - k \geq \log_2 \log_2(1/\epsilon)$.

La primera desigualdad implica que la fase amortiguada no requiere mas que $\frac{f(x^{(0)}) - p^*}{\gamma}$ pasos. El número total de iteraciones requeridas para obtener una exactitud $f(x) - p^* \leq \epsilon$, comenzando en el punto $x^{(0)}$ es acotada por

$$\frac{f(x^{(0)}) - p^*}{\gamma} + \log_2 \log_2(1/\epsilon) \quad (2.9)$$



a) Fase Newton amortiguado

Sea $\tilde{f}(t) = f(x + tN)$, entonces se tiene

$$\tilde{f}'(0) = -\lambda(x)^2, \quad \tilde{f}''(0) = \lambda(x)^2$$

Si integramos la cota superior en (2.3) dos veces, obtenemos una cota superior para $\tilde{f}(t)$:

$$\begin{aligned} \tilde{f}(t) &\leq \tilde{f}(0) + t\tilde{f}'(0) - t\tilde{f}''(0)^{\frac{1}{2}} - \log\left(1 - t\tilde{f}''(0)^{\frac{1}{2}}\right) \\ &= \tilde{f}(0) - t\lambda(x)^2 - t\lambda(x) - \log(1 - t\lambda(x)) \end{aligned} \quad (2.10)$$

Esta cota es válida para $0 \leq t < \frac{1}{\lambda(x)}$.

Podemos usar esta cota para demostrar que la línea de búsqueda con retroceso resulta en un tamaño de paso $t \geq \frac{\beta}{1+\lambda(x)}$. Para probar esto note que el punto $\tilde{t} = \frac{1}{1+\lambda(x)}$ satisface la condición de salida de la línea de búsqueda:

$$\begin{aligned} \tilde{f}(t) &\leq \tilde{f}(0) - \tilde{t}\lambda(x)^2 - \tilde{t}\lambda(x) - \log(1 - \tilde{t}\lambda(x)) \\ &= \tilde{f}(0) - \lambda(x) + \log(1 + \lambda(x)) \\ &\leq \tilde{f}(0) - \alpha \frac{\lambda(x)^2}{1 + \lambda(x)} \\ &= \tilde{f}(0) - \alpha\lambda(x)^2\tilde{t} \end{aligned}$$

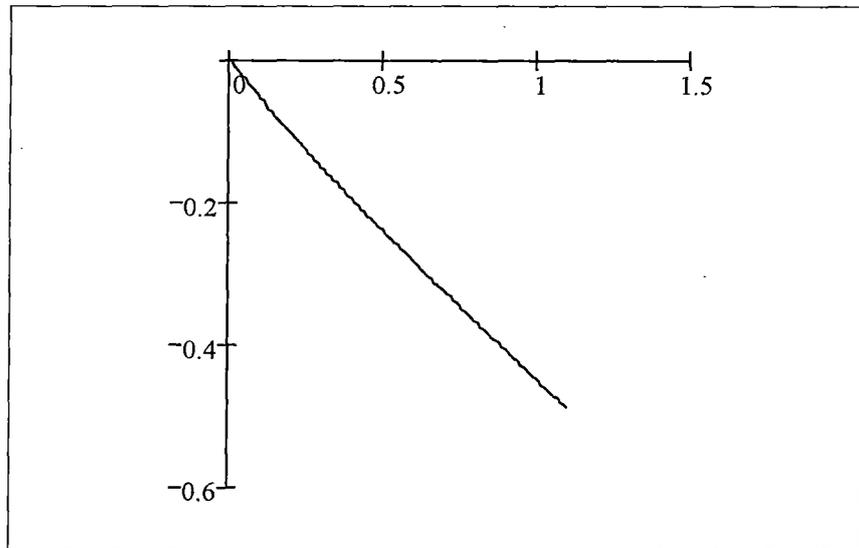
La segunda desigualdad se cumple del hecho que para $x \geq 0$



$$-x + \log(1+x) + \frac{x^2}{2(1+x)} \leq 0$$

Gráfico 2.2

Función estrictamente decreciente



Gráfica de $-x + \log(1+x) + \frac{x^2}{2(1+x)} \leq 0$, en donde se nota que es estrictamente decreciente. Gráfica elaborada con el software Mathcad.

Puesto que $t \geq \frac{\beta}{1+\lambda(x)}$, se tiene

$$\tilde{f}(t) - \tilde{f}(0) \leq -\alpha\beta \frac{\lambda(x)^2}{1+\lambda(x)}$$

De modo que (2.7) se cumple con

$$\gamma = \alpha \frac{\eta^2}{1+\eta}$$

b) Fase convergencia cuadrática

Demostraremos que podemos tomar

$$\eta = (1 - 2\alpha)/4$$

(que satisface $0 < \eta < 1/4$, puesto que $0 < \alpha < 1/2$), es decir, si $\lambda(x^{(k)}) \leq (1 - 2\alpha)/4$, entonces la línea de búsqueda con retroceso acepta el paso unitario y (2.8) se cumple.

Primeramente notemos que la cota superior en (2.10) implica que un paso unitario $t = 1$ nos lleva a un punto en el dominio de f si $\lambda(x) < 1$. Además, si $\lambda(x) \leq (1 - 2\alpha)/2$, tenemos, usando (2.10),

$$\begin{aligned}\tilde{f}(1) &\leq \tilde{f}(0) - \lambda(x)^2 - \lambda(x) - \log(1 - \lambda(x)) \\ &\leq \tilde{f}(0) - \frac{1}{2}\lambda(x)^2 + \lambda(x)^3 \\ &\leq \tilde{f}(0) - \alpha\lambda(x)^2\end{aligned}$$

De modo que el paso unitario satisface la condición suficiente de decrecimiento. La segunda línea se sigue del hecho que

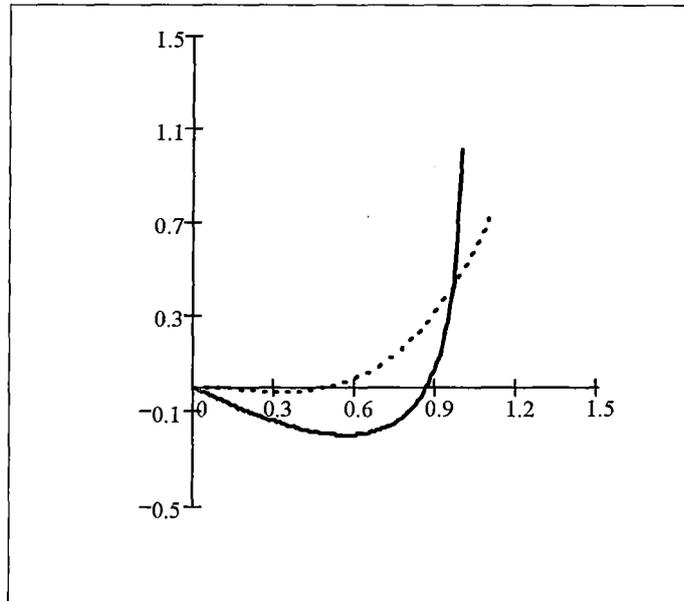
$$-x - \log(1 - x) \leq -\frac{1}{2}x^2 + x^3 \text{ para } 0 \leq x \leq 0.81$$

(Véase el gráfico N° 2.3)



Gráfico 2.3

Comparación con la función cúbica



Gráfica de $-x - \log(1-x)$ en línea sólida y $-\frac{1}{2}x^2 + x^3$ en línea punteada. Gráfica elaborada con el software Mathcad.

La desigualdad (2.8) se sigue del hecho que, si $\lambda(x) < 1$, y

$$x^+ = x - \nabla^2 f(x)^{-1} \nabla f(x)$$

Entonces,

$$\lambda(x^+) \leq \frac{\lambda(x)^2}{(1 - \lambda(x))^2} \quad (2.11)$$

En particular, si $\lambda(x) < 1/4$,

$$\lambda(x^+) \leq 2\lambda(x)^2$$

Lo que demuestra que (2.8) se cumple cuando $\lambda(x^{(k)}) \leq \eta$

c) Cota final de complejidad

Juntando las cotas halladas, la cota (2.9) sobre el número de iteraciones de Newton

$$\frac{f(x^{(0)}) - p^*}{\gamma} + \log_2 \log_2 \left(\frac{1}{\epsilon} \right) = \frac{20 - 8\alpha}{\alpha\beta(1 - 2\alpha)^2} (f(x^{(0)}) - p^*) + \log_2 \log_2(1/\epsilon)$$

Esta cota depende solamente de los parámetros de la línea de búsqueda α, β , y la exactitud final ϵ . Además el término que involucra ϵ puede ser reemplazada por la constante 6, de modo que la cota depende solo de α y β . Para típicos valores de α y β las constantes que escalan $f(x^{(0)}) - p^*$ están sobre el orden de varios cientos. Por ejemplo, $\alpha = 0.1$ y $\beta = 0.8$ el factor que escala es 375. Con tolerancia $\epsilon = 10^{-10}$, obtenemos la cota

$$375(f(x^{(0)}) - p^*) + 6$$

Veremos que esta cota es bastante conservativa. Un análisis mucho mas refinado, como el que originalmente dieron Nesterov y Nemirovskii, da una cota similar, con una sustancial constante que escala $f(x^{(0)}) - p^*$. (Ver Stephen Boyd(5), capítulo 9, página 505).

2.3. Métodos de punto interior

En este ítem discutiremos métodos de punto interior para resolver problemas de optimización convexa que incluyen restricciones de desigualdad.



2.3.1. Problema convexo

Es un problema de la forma

$$\begin{aligned} & \text{minimizar} && f_0(x) \\ & \text{Sujeto a:} && f_i(x) \leq 0, i = 1, \dots, m \\ & && Ax = b \end{aligned} \tag{2.12}$$

donde f_0, \dots, f_m son funciones de $\mathbb{R}^n \rightarrow \mathbb{R}$, convexas, dos veces continuamente diferenciables, $A \in \mathbb{R}^{p \times n}$, con rango de $A = p < n$.

Asumiendo que el problema es factible, existe x^* óptimo, $f_0(x^*) = p^*$ es su valor. Asumamos también que el problema es estrictamente factible, es decir, existe $x \in D$ que satisface $Ax = b$, $f_i(x) < 0$, $\forall i = 1, \dots, m$.

Esto significa que las condiciones de Slater se cumplen, de modo que existen $\lambda^* \in R^m$, $v^* \in R^p$ óptimos duales, que juntamente con x^* satisfacen las condiciones de KKT

$$\begin{aligned} Ax^* &= b, \quad f_i(x^*) \leq 0, \quad i = 1, \dots, m \\ \lambda^* &\geq 0 \\ \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + A^T v^* &= 0 \\ \lambda_i^* f_i(x^*) &= 0, \quad i = 1, \dots, m \end{aligned} \tag{2.13}$$

Métodos de punto interior resuelven el problema (2.12) o (2.13) aplicando el método de Newton a una sucesión de problemas restringidos. Nuestro objetivo es reformular (2.12) con ecuaciones restricción para aplicar el método de Newton.



1° El primer paso es reescribir (2.12) agregando la inecuación restricción implícita en la función objetivo

$$\text{minimizar } f_0(x) + \sum_{i=1}^m I_-(f_i(x)) \quad (2.14)$$

$$\text{sujeto a : } Ax = b$$

donde $I_- : \mathbb{R} \rightarrow \mathbb{R}$ es la función indicatriz

$$I_-(u) = \begin{cases} 0, & u \leq 0 \\ \infty, & u > 0 \end{cases}$$

Luego

$$I_-(f_i(x)) = \begin{cases} 0, & f_i(x) \leq 0 \\ \infty, & f_i(x) > 0 \end{cases}$$

El problema (2.14) no tiene inecuaciones restricción, pero la función objetivo no es en general diferenciable, así, el método de Newton no puede aplicarse.

2° La idea básica del método es aproximar la función indicatriz I_- por la función.

$$\hat{I}_-(u) = -\frac{1}{t} \log(-u), \quad \text{dom } \hat{I}_- = -\mathbb{R}_{++}$$

El problema (2.14) puede reformularse haciendo

$$\text{minimizar } f_0(x) + \sum_{i=1}^m -\frac{1}{t} \log(-f_i(x)) \quad (2.15)$$

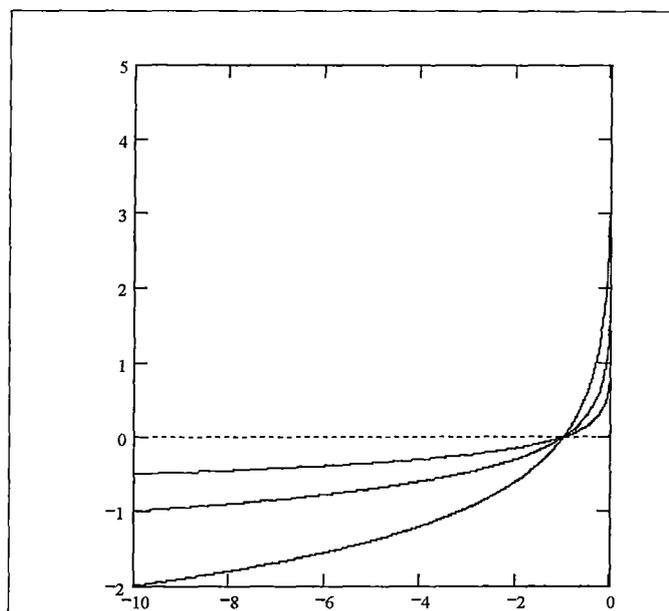
$$\text{sujeto a : } Ax = b$$



La función objetivo es ahora convexa y diferenciable, puesto que $-\frac{1}{t}\log(-u)$ es convexa y decreciente en u , y diferenciable. Asumiendo que una condición apropiada de cercanía se cumple, el método de Newton puede usarse para resolver este problema.

Gráfica 2.4

Función indicatriz



Gráfica de la función indicatriz para $t=2$, $t=1$, $t=1/2$. Gráfica elaborada con el software Mathcad.

La función

$$\phi(x) = - \sum_{i=1}^m \log(-f_i(x)),$$

$$\text{dom } \phi = \{x \in \mathbb{R}^n: f_i(x) < 0, i = 1, \dots, m\}$$

es llamada función barrera logarítmica para el problema (2.12).

3º Un problema equivalente a (2.15) es el siguiente

$$\text{minimizar } t f_0(x) + \phi(x) \quad (2.16)$$

$$\text{sujeto a : } Ax = b$$

El cual tiene los mismos minimizadores que (2.15). Asumimos ahora que (2.16) puede ser resuelto vía el método de Newton y en particular, que este tiene una única solución, para cada $t > 0$.

2.3.2. Curva central

Para cada $t > 0$ definimos $x^*(t)$ como la solución de (2.16), la curva central asociada con el problema (2.12) se define como el conjunto de puntos $x^*(t)$, $t > 0$. Los puntos sobre la curva central están caracterizados por la siguiente condición necesaria y suficiente: $x^*(t)$ es estrictamente factible, es decir, satisface

$$Ax^*(t) = b, \quad f_i(x^*(t)) < 0, \quad i = 1, \dots, m$$

y existe un $\hat{v} \in R^p$ tal que

$$t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{v} = 0$$

$$t\nabla f_0(x^*(t)) + \sum_{i=1}^m \frac{1}{-f_i(x^*(t))} \nabla f_i(x^*(t)) + A^T\hat{v} = 0 \quad (2.17)$$

Sea el programa lineal

$$\text{Minimice } c^T x$$

$$\text{Sujeto a: } Ax \leq b$$



La barrera logarítmica está dada por

$$F(x) = - \sum_{i=1}^n \log(b_i - a_i^T x), \quad \text{dom}F = \{x: Ax < b\},$$

donde a_1^T, \dots, a_m^T son las filas de la matriz A . El gradiente y la Hessiana de la barrera son

$$\nabla F(x) = \sum_{i=1}^m \frac{1}{b_i - a_i^T x} a_i, \quad \nabla^2 F(x) = \sum_{i=1}^m \frac{1}{(b_i - a_i^T x)^2} a_i a_i^T$$

O, mas compactamente

$$\nabla F(x) = A^T d, \quad \nabla^2 F(x) = A^T \text{diag}(d)^2 A,$$

donde los elementos de $d \in \mathbb{R}^m$ están dados por $d_i = \frac{1}{b_i - a_i^T x}$ y puesto que x es estrictamente factible se tiene que $d > 0$, de modo que la Hessiana de F es no singular sí y solamente sí A tiene rango n .

La condición de centralidad es

$$tc + \sum_{i=1}^m \frac{1}{b_i - a_i^T x} a_i = tc + A^T d = 0.$$

Podemos dar una interpretación geométrica de la condición de centralidad. En un punto $x^*(t)$ sobre la curva central el gradiente $\nabla F(x^*(t))$, que es normal a los conjuntos de nivel de F que pasan por $x^*(t)$, deberían ser paralelos a $-c$. En otras palabras, el hiperplano $c^T x = c^T x^*(t)$ es tangente al conjunto de nivel de F que pasa por $x^*(t)$.



2.3.3. Puntos duales de la curva central

De (2.17) podemos deducir una propiedad importante de la curva central. Cada punto central nos lleva a un punto dual factible, y entonces, a una cota inferior sobre el valor optimal p^* . Mas específicamente, definimos

$$\lambda_i^*(t) = \frac{1}{-t f_i(x^*(t))}, \quad i = 1, \dots, m, \quad v^*(t) = \frac{\hat{v}}{t} \quad (2.18)$$

Afirmamos que el par $\lambda^*(t), v^*(t)$ es dual factible.

Primeramente, es claro que $\lambda^*(t) > 0$ puesto que $f_i(x^*(t)) < 0$, $i = 1, \dots, m$. Expresando (2.17) como

$$t \nabla f_0(x^*(t)) + \sum_{i=1}^m \lambda_i^*(t) \nabla f_i(x^*(t)) + A^T v^*(t) = 0$$

Vemos que $x^*(t)$ minimiza el Lagrangiano

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + v^T (Ax - b)$$

Para $\lambda = \lambda^*(t), v = v^*(t)$, lo cual significa que $\lambda^*(t), v^*(t)$ es un par factible dual. Por lo tanto, la función dual $g(\lambda^*(t), v^*(t))$ es finita, y

$$\begin{aligned} g(\lambda^*(t), v^*(t)) &= f_0(x^*(t)) + \sum_{i=1}^m \lambda_i^*(t) f_i(x^*(t)) + \\ &+ v^*(t)^T (A(x^*(t)) - b) \\ &= f_0(x^*(t)) - \frac{m}{t}. \end{aligned}$$



En particular, la dualidad gap asociada con $x^*(t)$ y el par factible dual $\lambda^*(t), v^*(t)$ es simplemente m/t . Como una importante consecuencia tenemos

$$f_0(x^*(t)) - p^* \leq \frac{m}{t},$$

es decir, $x^*(t)$ no es más que m/t -sub óptimo. Esto confirma la idea intuitiva que $x^*(t)$ converge a un punto óptimo cuando $t \rightarrow \infty$.

El dual del programa lineal es

$$\begin{aligned} & \text{Maximice } -b^T z \\ & \text{Sujeto a: } A^T z + c = 0, \quad z \geq 0 \end{aligned}$$

De las condiciones de optimalidad tenemos

$$z_i^*(t) = \frac{1}{t(b_i - a_i^T x^*(t))}, \quad i = 1, \dots, m,$$

Es dual factible, con valor objetivo dual

$$-b^T z^*(t) = c^T x^*(t) + (Ax^*(t) - b)^T z^*(t) = c^T x^*(t) - \frac{m}{t}$$

Podemos también interpretar las condiciones de curva central (2.17) como una deformación continua de las condiciones de optimalidad (2.13). Un punto x es igual a $x^*(t)$ si y solamente si, existe λ, v tales que

$$\begin{aligned} Ax &= b, \quad f_i(x) \leq 0, \quad i = 1, \dots, m \\ \lambda &\geq 0 \end{aligned}$$



$$\nabla f_0(x) + \sum_{i=1}^n \lambda_i \nabla f_i(x) + A^T v = 0 \quad (2.19)$$

$$-\lambda_i f_i(x) = 1/t, \quad i = 1, \dots, m$$

La única diferencia con (2.13) es que la condición de complementaridad ahora es,

$$-\lambda_i f_i(x) = 1/t.$$

En particular, para t grande, $x^*(t)$ y el punto dual asociado $\lambda^*(t), v^*(t)$ satisfacen las condiciones de optimalidad para el problema (2.12).

2.4. El método barrera

Hemos visto que el punto $x^*(t)$ es m/t -sub optimal, y que una certificación de esta exactitud está provista por el par factible dual $\lambda^*(t), v^*(t)$. Esto sugiere un método directo para resolver el problema original (2.12) con una exactitud especificada ϵ : Simplemente tomamos $t = m/\epsilon$ y resolvemos el problema restringido

$$\begin{aligned} &\text{minimizar } (m/\epsilon) f_0(x) + \phi(x) \\ &\text{sujeto a : } \quad Ax = b \end{aligned}$$

usando el método de Newton. Este método podría ser llamado método de minimización irrestricta, puesto que nos lleva a resolver el problema (2.12) con una exactitud garantizada por resolver un problema irrestricto. Aunque este método puede trabajar bien para problemas pequeños, buenos puntos de inicio, y modesta exactitud (es decir, ϵ no tan pequeño), este no trabaja bien en otros casos, como resultado es poco usado.



2.4.1. Algoritmo

Dado x estrictamente factible, $t = t^{(0)} > 0$, $\mu > 1$, tolerancia $\epsilon > 0$.

Repetir

1. Paso de centrado: compute $x^*(t)$ minimizando $tf_0 + \phi$ sujeto a $Ax = b$, comenzando en x .
2. Actualizar $x = x^*(t)$.
3. Criterio de parada: salir si $\frac{m}{t} < \epsilon$.
4. Incrementar t , haciendo $t = \mu t$

En cada iteración (excepto el primero) computamos $x^*(t)$ comenzando del punto central anterior, y luego incrementamos t por un factor $\mu > 1$. El algoritmo puede también retornar el par ϵ -optimal dual $\lambda^*(t), v^*(t)$ o certificación para x .

Nos referimos a cada ejecución del paso 1 como un paso de centrado (puesto que un punto central está siendo calculado) o una iteración de salida, y al primer paso de centrado (el cálculo de $x^*(t^{(0)})$) como el paso central inicial. Aunque cualquier método para minimización irrestricta puede ser usado en el paso 1, asumiremos que se usa el método de Newton. Nos referiremos a las iteraciones o pasos de Newton ejecutados durante el paso de centrado como iteraciones internas. En cada paso interno, se tiene un punto factible primal; tenemos un punto factible dual, sin embargo, solamente al final de cada paso de salida.

2.4.2. Exactitud de centrado

Computar $x^*(t)$ exactamente no es necesario, puesto que la curva central no tiene significancia mas que el hecho que nos lleva a una



solución del problema original cuando $t \rightarrow \infty$, centrado inexacto, aun nos llevará a una sucesión de puntos $x^{(k)}$ que converge a un punto optimal. Centrado inexacto, sin embargo, significa, que los puntos $\lambda^*(t), v^*(t)$ computados de (2.19) no son exactamente dual factibles.

Por otro lado el costo de computar un minimizador extremadamente exacto de $tf_0 + \phi$, comparado con el costo de computar un buen minimizador es solamente unos pocos mas pasos de Newton, por esta razón no es necesario asumir centrado exacto.

2.4.3. Análisis de complejidad

Usando el análisis de complejidad del método de newton para funciones auto concordantes, daremos un análisis de complejidad del método barrera. El análisis aplica a muchos problemas comunes, y nos lleva a interesantes conclusiones: esto da una cota rigurosa sobre el número total de pasos de Newton requeridos para resolver un problema usando el método barrera, y justifica nuestra observación que problemas de centrado no son más dificultosos cuando t se incrementa.

Estableceremos los siguientes requerimientos

- La función $tf_0 + \phi$ es cerrada y auto concordante para todo $t \geq t^{(0)}$
- Los conjuntos subnivel de (2.12) son acotados

El segundo requerimiento implica que el problema de centrado tiene conjuntos subnivel acotados, y por lo tanto, el problema de centrado tiene solución. La acotación de los conjuntos subnivel también implica que la Hessiana de $tf_0 + \phi$ es definida positiva en todo su dominio. Mientras la condición de auto concordancia restringe el análisis de



complejidad a una clase particular de problemas; es importante enfatizar que el método barrera trabaja bien en general, cuando la condición de auto concordancia se cumple o no.

La condición de auto concordancia se cumple para una gran variedad de problemas, incluyendo todos los problemas lineales y cuadráticos. Si las funciones f_i son lineales o cuadráticas, entonces

$$tf_0 - \sum_{i=1}^m \log(-f_i)$$

Es auto concordante para todo $t \geq 0$.

La teoría de complejidad del método de Newton para funciones auto concordantes, demuestra que el número de iteraciones requerida para minimizar una función cerrada f auto concordante estrictamente convexa es acotada por

$$\frac{f(x) - p^*}{\gamma} + c \quad (2.20)$$

donde x es el punto de inicio, $p^* = \inf_x f(x)$ es el valor óptimo. La constante γ depende solamente de los parámetros de la línea de búsqueda con retroceso α, β y está dada por

$$\frac{1}{\gamma} = \frac{20 - 8\alpha}{\alpha\beta(1 - 2\alpha)^2}$$

La constante c depende solamente de la tolerancia ϵ_{nt} ,

$$c = \log_2 \log_2(1/\epsilon_{nt})$$



y puede ser aproximado como $c = 6$. La expresión en (2.20) es una cota conservativa sobre el número de pasos de Newton requeridos, pero nuestro interés en esta sección es solamente establecer una cota de complejidad, concentrándonos en como esta se incrementa con el tamaño del problema y parámetros del algoritmo.

Usaremos este resultado para deducir una cota sobre el número de iteraciones de Newton requeridos por una iteración de salida del método barrera, es decir, para computar $x^*(\mu t)$ a partir de $x^*(t)$. Para aligerar la notación usaremos x para denotar $x^*(t)$, el iterado actual, y usamos x^+ para denotar $x^*(\mu t)$, el siguiente iterado. Usamos λ , ν para denotar $\lambda^*(t)$, $\nu^*(t)$ respectivamente.

La condición de auto concordancia implica que

$$\frac{\mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+)}{\gamma} + c \quad (2.21)$$

Es una cota superior sobre el número de pasos que se requiere para computar x^+ , comenzando en x . Desafortunadamente, no conocemos x^+ , y entonces la cota (2.21) hasta que efectivamente computemos x^+ , es decir, el iterado del algoritmo de Newton.

Podemos, sin embargo, obtener una cota para (2.21) como sigue

$$\begin{aligned} & \mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+) \\ &= \mu t f_0(x) - \mu t f_0(x^+) + \sum_{i=1}^m \log(-\mu t \lambda_i f_i(x^+)) - m \log \mu \\ &\leq \mu t f_0(x) - \mu t f_0(x^+) - \mu t \sum_{i=1}^m \lambda_i f_i(x^+) - m - m \log \mu \end{aligned}$$



$$\begin{aligned}
&= \mu t f_0(x) - \mu t \left(f_0(x^+) - \sum_{i=1}^m \lambda_i f_i(x^+) + v^T(Ax^+ - b) \right) - m \\
&\quad - m \log \mu \\
&\leq \mu t f_0(x) - \mu t g(\lambda, v) - m - m \log \mu \\
&= m(\mu - 1 - \log \mu)
\end{aligned}$$

Para obtener la segunda línea, usamos que $\lambda_i = 1/tf_i$, en la segunda desigualdad hemos usado el hecho que $\log a \leq a - 1$ para todo $a > 0$. Para obtener la cuarta línea de la tercera, hemos usado que $Ax^+ = b$, de modo que el término $v^T(Ax^+ - b) = 0$. La segunda desigualdad se sigue de la definición de la función dual:

$$\begin{aligned}
g(\lambda, v) &= \min_z \left(f_0(z) + \sum_{i=1}^m \lambda_i f_i(z) + v^T(Az - b) \right) \\
&\leq f_0(x^+) + \sum_{i=1}^m \lambda_i f_i(x^+) + v^T(Ax^+ - b)
\end{aligned}$$

La última línea se sigue del hecho que $g(\lambda, v) = f_0(x) - m/t$.

La conclusión es que

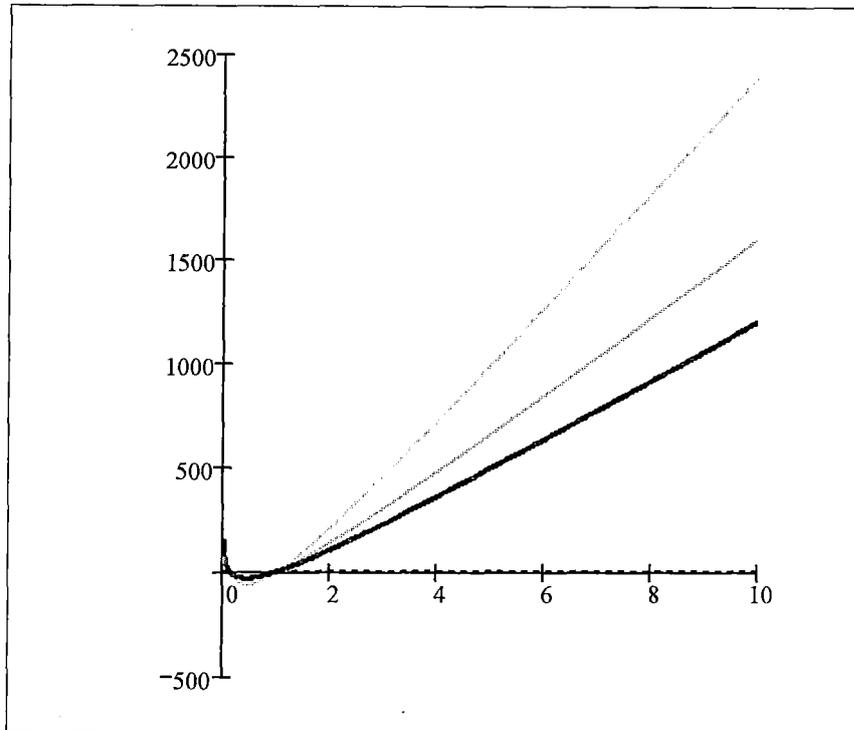
$$\frac{m(\mu - 1 - \log \mu)}{\gamma} + c \tag{2.22}$$

es una cota de (2.21), y por lo tanto una cota sobre el número de iteraciones requeridas para una iteración de salida del método barrera. La función $m(\mu - 1 - \log \mu)$ se muestra en el gráfico N° 2.5. Para μ pequeño es aproximadamente cuadrático, para μ grande crece aproximadamente linealmente.



Gráfica 2.5

Comportamiento cuadrático y lineal



Gráfica de $m(\mu - 1 - \log \mu)$ para $m = 150, 200, 300$. Gráfica elaborada con el software Mathcad.

La cota en (2.22) demuestra que el número de iteraciones de Newton requeridas en cada paso de centrado es acotada por una cantidad que depende solamente de μ , el factor por el que t es actualizado en cada iteración de salida del método barrera, y m , el número de restricciones del problema. Esta cota también depende, débilmente, de los parámetros α , β usada en la línea de búsqueda para las iteraciones internas, y en una forma muy débil de la tolerancia usada para terminar las iteraciones internas. Es interesante notar que las cotas no dependen de n , la dimensión de la variable, o p , el número de ecuaciones restricción, o de la función objetivo y las funciones restricción. Finalmente, notemos que esta no depende de t ; en particular, cuando

$t \rightarrow \infty$, una cota uniforme sobre el número de pasos de Newton por iteración de salida se cumple.

2.5. Métodos de punto interior primal dual

Métodos de punto interior primal dual son muy similares al método barrera, con algunas diferencias.

1. Existe solamente una iteración o bucle, es decir, no hay distinción entre iteraciones internas e iteraciones de salida, como en el método Barrera. En cada iteración, ambas variables primal y dual son actualizadas.
2. Las direcciones de búsqueda en un método de punto interior primal dual son obtenidos del método de newton, aplicado a las ecuaciones de KKT modificadas (es decir, las condiciones de optimalidad para el problema de centrado en la barrera logarítmica). Las direcciones de búsqueda primal dual son similares, pero no las mismas, que las direcciones de búsqueda en el método barrera.
3. En un método de punto interior primal dual, los iterados primal dual no son necesariamente factibles.

Métodos de punto interior primal dual son a menudo más eficientes, que el método barrera, especialmente cuando se requiere alta precisión, puesto que estos exhiben mejor convergencia que la lineal. Otra ventaja de algoritmos primal dual sobre el método barrera es que ellos pueden trabajar cuando el problema es factible, pero no estrictamente factible. \blacktriangledown

2.5.1. Dirección de búsqueda primal dual

Como en el método barrera, comenzaremos con las condiciones modificadas de KKT expresadas como $r_t(x, \lambda, \nu) = 0$, donde definimos



$$r_t(x, \lambda, \nu) = \begin{bmatrix} \nabla f_0(x) + Df(x)^T + A^T \nu \\ -diag(\lambda)f(x) - \left(\frac{1}{t}\right) 1 \\ Ax - b \end{bmatrix}$$

para $t > 0$. Aquí $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ y sus derivadas están dadas por

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}, \quad Df(x) = \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix}$$

Si x, λ, ν satisfacen $r_t(x, \lambda, \nu) = 0$ ($f_i(x) < 0$), entonces $x = x^*(t)$, $\lambda = \lambda^*(t)$, $\nu = \nu^*(t)$, en particular x es primal factible, λ y ν son dual factibles, con dualidad gap m/t .

El primer bloque componente de r_t

$$r_{dual} = \nabla f_0(x) + Df(x)^T + A^T \nu$$

es llamado dual residual, y el último bloque componente,

$$r_{pri} = Ax - b$$

es llamado el residual primal. El bloque del medio

$$r_{cent} = -diag(\lambda)f(x) - \left(\frac{1}{t}\right) 1$$

es el residual central, es decir, el residual para las condiciones de complementariedad modificadas.

Ahora consideremos el paso de Newton para resolver las ecuaciones no lineales $r_t(x, \lambda, \nu) = 0$ para t fijo, en un punto (x, λ, ν) que satisface $f(x) < 0, \lambda > 0$, denotaremos el punto actualizado y el paso de Newton como

$$y = (x, \lambda, \nu), \quad \Delta y = (\Delta x, \Delta \lambda, \Delta \nu)$$



respectivamente. El paso de Newton está caracterizado por la ecuación lineal

$$r_t(y + \Delta y) \approx r_t(y) + Dr_t(y)\Delta y = 0$$

es decir, $\Delta y = -Dr_t(y)^{-1}r_t(y)$.

En términos de x, λ y v , se tiene

$$\begin{bmatrix} \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\ -diag(\lambda)Df(x) & -diag f(x) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta v \end{bmatrix} = \begin{bmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{bmatrix} \quad (2.23)$$

La dirección de búsqueda primal dual $\Delta y_{pd} = (\Delta x_{pd}, \Delta \lambda_{pd}, \Delta v_{pd})$ está definida por la solución del sistema (2.23).

Las direcciones de búsqueda primal dual están unidas, ambas a través de la matriz de los coeficientes y los residuales. Por ejemplo, la dirección de búsqueda primal Δx_{pd} depende del valor actualizado de las variables duales λ, v , como también x . Notemos también que si x satisface $Ax = b$, es decir, el residual primal de factibilidad r_{pri} es cero, entonces $A\Delta x_{pd} = 0$, de modo que Δx_{pd} define una dirección factible (primal): para cualesquier $s, x + s\Delta x_{pd}$ se cumple $A(x + s\Delta x_{pd}) = b$.

2.5.2. Comparación con direcciones de búsqueda del método barrera

Las direcciones de búsqueda primal-dual están relacionadas con las direcciones de búsqueda usadas por el método barrera, pero no son las mismas.



En el sistema (2.23) que definen las direcciones de búsqueda primal dual, eliminamos la variable $\Delta\lambda_{pd}$ usando

$$\Delta\lambda_{pd} = -\text{diag}(f(x))^{-1} \text{diag}(\lambda) Df(x) \Delta x_{pd} + \text{diag}(f(x))^{-1} r_{cent}$$

que viene del segundo bloque de ecuaciones. Sustituyendo esto en el primer bloque de ecuaciones resulta

$$\begin{aligned} \begin{bmatrix} H_{pd} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{pd} \\ \Delta v_{pd} \end{bmatrix} &= - \begin{bmatrix} r_{dual} + Df(x)^T \text{diag}(f(x))^{-1} r_{cent} \\ r_{pri} \end{bmatrix} \\ &= - \begin{bmatrix} \nabla f_0(x) + \left(\frac{1}{t}\right) \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + A^T v \\ r_{pri} \end{bmatrix} \end{aligned} \quad (2.24)$$

Donde

$$H_{pd} = \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) + \sum_{i=1}^m \frac{\lambda_i}{-f_i(x)} \nabla f_i(x) \nabla f_i(x)^T \quad (2.25)$$

Podemos comparar (2.24) con las ecuaciones que definen el paso de Newton para el problema de centrado en el método barrera con parámetro t . Esta ecuación puede escribirse como

$$\begin{aligned} \begin{bmatrix} H_{bar} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{bar} \\ v_{bar} \end{bmatrix} &= - \begin{bmatrix} t \nabla f_0(x) + \nabla \varphi(x) \\ r_{pri} \end{bmatrix} \\ &= - \begin{bmatrix} t \nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) \\ r_{pri} \end{bmatrix} \end{aligned} \quad (2.26)$$

donde

$$H_{bar} = t \nabla^2 f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla^2 f_i(x) + \sum_{i=1}^m \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T \quad (2.27)$$



(Aquí hemos dado la expresión general para el paso de Newton factible; si el x actualizado es factible, es decir $r_{pri} = 0$, entonces Δx_{bar} coincide con el paso de Newton factible Δx_{nt})

Nuestra primera observación es que los dos sistemas de ecuaciones (2.24) y (2.26) son muy similares. La matriz de los coeficientes en ambos tiene la misma estructura, en verdad las matrices H_{pd} y H_{bar} son ambas combinaciones lineales positivas de las matrices

$$\nabla^2 f_0(x), \nabla^2 f_1(x), \dots, \nabla^2 f_m(x), \nabla f_1(x) \nabla f_1(x)^T, \dots, \nabla f_m(x) \nabla f_m(x)^T$$

Esto significa que el mismo método puede ser usado para computar las direcciones de búsqueda primal dual y el paso de Newton del método barrera.

Podemos decir más acerca de la relación entre las ecuaciones (2.24) y (2.26). Supongamos que dividimos el primer bloque de ecuaciones en (2.26) por t , y definimos la variable $\Delta v_{bar} = (1/t)v_{bar} - v$ (donde v es arbitrario).

Entonces obtenemos

$$\begin{bmatrix} (1/t)H_{bar} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{bar} \\ v_{bar} \end{bmatrix} = - \begin{bmatrix} \nabla f_0(x) + (1/t) \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + A^T v \\ r_{pri} \end{bmatrix}$$

En esta forma, el lado derecho es idéntico al lado derecho de las ecuaciones primal dual.

La matriz de los coeficientes difiere solamente en el 1, 1 bloque:

$$H_{pd} = \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) + \sum_{i=1}^m \frac{\lambda_i}{-f_i(x)} \nabla f_i(x) \nabla f_i(x)^T$$



$$(1/t)H_{bar} = t\nabla^2 f_0(x) + \sum_{i=1}^m \frac{\nabla^2 f_i(x)}{-tf_i(x)} + \sum_{i=1}^m \frac{1}{tf_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T$$

Cuando x y λ satisfacen $-f_i(x)\lambda_i = 1/t$, entonces la matriz de los coeficientes, y por lo tanto también las direcciones de búsqueda, coinciden.

2.5.3. La dualidad gap sustituta

En el método punto interior primal dual los iterados $x^{(k)}, \lambda^{(k)}, y v^{(k)}$ no son necesariamente factibles, excepto en el límite cuando el algoritmo converge. Esto significa que no podemos fácilmente evaluar una dualidad gap $\eta^{(k)}$ asociado con el paso k del algoritmo, como se hizo en el método barrera. Definimos la dualidad gap sustituta, para cualquier x que satisfice $f(x) < 0, \lambda \geq 0$, como

$$\hat{\eta}(x, \lambda) = -f(x)^T \lambda$$

El gap sustituto debería ser la dualidad gap, si x es primal factible y λ, v son dual factible, es decir, si $r_{pri} = 0$ y $r_{dual} = 0$. Note que el valor del parámetro t que corresponde a la dualidad gap sustituta es $m/\hat{\eta}$.

2.5.4. Algoritmo del método de punto interior primal dual

Dado un x tal que $f_1(x) < 0, \dots, f_m(x) < 0, \lambda > 0, \mu > 1, \epsilon_{fcas} > 0,$

$\epsilon > 0$

Repetir

1) Determine t . Hacer $t := \mu m / \hat{\eta}$



- 2) Compute la dirección de búsqueda primal dual Δy_{pd}
- 3) Línea de búsqueda y actualización

Determine la longitud de paso $s > 0$ y hacer $y = y + s\Delta y_{pd}$

Hasta que $\|r_{pri}\|_2 \leq \epsilon_{fcas}$, y $\|r_{dual}\|_2 \leq \epsilon_{fcas}$ y $\hat{\eta} \leq \epsilon$.

En el paso 1, el parámetro es μ veces $m/\hat{\eta}$, que es el valor de t asociado con la dualidad sustituta actualizada $\hat{\eta}$.

El método termina cuando x es primal factible y λ, μ son dual factibles (dentro de la tolerancia ϵ_{fcas}) y la tolerancia gap sustituta es mas pequeña que la tolerancia ϵ .

2.5.5. Línea de búsqueda

La línea de búsqueda en el método de punto interior primal dual es una línea de búsqueda con retroceso estándar, basada sobre la norma del residual y modificada para asegurar que $\lambda > 0$ y $f(x) < 0$. Denotamos el iterado actualizado como x, λ y v y el siguiente iterado como x^+, λ^+ y v^+ , es decir

$$x^+ = x + s\Delta x_{pd}, \quad \lambda^+ = \lambda + s\Delta \lambda_{pd}, \quad v^+ = v + s\Delta v_{pd}$$

El residual, evaluado en y^+ , será denotado por r^+ .

Calculamos primeramente la longitud de paso positivo mayor, sin exceder 1, esto da $\lambda^+ \geq 0$, es decir,

$$s^{max} = \sup\{s \in [0,1]: \lambda + s\Delta \lambda \geq 0\}$$

$$= \min\{1, \min\{-\lambda_i/\Delta \lambda_i < 0\}\}$$



Comenzamos el retroceso con $s = 0.99s^{max}$, y multiplicamos s por $\beta \in (0,1)$ mientras tenemos $f(x^+) < 0$. Continuamos multiplicando s por β hasta que tenemos

$$\|r_t(x^+, \lambda^+, v^+)\|_2 \leq (1 - \alpha s) \|r_t(x, \lambda, v)\|_2$$

Elecciones comunes para los parámetros de retroceso α, β son los mismos como los del método de Newton: α es típicamente escogido en el rango de 0.01 hasta 0.1, y β es escogido entre los valores de 0.3 a 0.8.

Una iteración del algoritmo de punto interior primal dual es el mismo que el de infactibilidad del método de Newton, aplicado al resolver $r_t(x, \lambda, v) = 0$ pero modificado para asegurar $\lambda > 0$ y $f(x) < 0$. Los mismos argumentos usados en la prueba de convergencia del método de Newton demuestran que la línea de búsqueda para el método primal dual siempre termina en un número finito de pasos.

2.5.6. Implementación

El principal reto en el método barrera es computar el paso de Newton para el problema de centrado, que consiste en resolver un conjunto de ecuaciones lineales de la forma

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ v_{nt} \end{bmatrix} = - \begin{bmatrix} g \\ 0 \end{bmatrix}$$

Donde

$$H = t \nabla^2 f_0(x) + \sum_{i=1}^m \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$



$$g = t\nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x)$$

Las ecuaciones de Newton para el método primal dual tienen exactamente la misma estructura, así que nuestras observaciones en esta sección aplican al método primal dual también.

a) Problemas esparsos

Si el problema original es esparsos, lo cual significa que la función objetivo y cada función restricción depende solamente de un número modesto de variables, entonces las matrices gradiente y hessiana son también esparsas.

b) Programación lineal en forma estándar

Discutiremos la implementación del método barrera para la forma estándar de un problema de programación lineal

Minimice $c^T x$

Sujeto a: $Ax = b, x \geq 0$

Con $A \in \mathbb{R}^{m \times n}$. Las ecuaciones de Newton para el problema de centrado

Minimice $tc^T x - \sum_{i=1}^n \log x_i$

Sujeto a: $Ax = b, x \geq 0$

están dados por

$$\begin{bmatrix} \text{diag}(x)^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ v_{nt} \end{bmatrix} = \begin{bmatrix} -tc + \text{diag}(x)^{-1} \mathbf{1} \\ 0 \end{bmatrix}$$



Estas ecuaciones son usualmente resueltas por eliminación del bloque Δx_{nt} . De la primera ecuación

$$\begin{aligned}\Delta x_{nt} &= \text{diag}(x)^2 (-tc + \text{diag}(x)^{-1} 1 - A^T v_{nt}) \\ &= -t \text{diag}(x)^2 c + x - \text{diag}(x)^2 A^T v_{nt}\end{aligned}$$

Sustituyendo en la segunda ecuación se tiene

$$A \text{diag}(x)^2 A^T v_{nt} = -tA \text{diag}(x)^2 c + b$$

La matriz de los coeficientes es definida positiva puesto que por hipótesis $\text{rango } A = m$. Además si A es esparsa, entonces usualmente $A \text{diag}(x)^2 A^T$ es esparsa.

CAPÍTULO III

MATERIALES Y MÉTODOS

Como se trata de una investigación básica aplicada, se ha usado el método inductivo deductivo.

- 3.1. Se ha presentado la teoría de funciones o barreras auto concordantes comenzando con funciones de una sola variable para luego generalizar para funciones de varias variables.
- 3.2. Se ha mostrado la convergencia y complejidad del método primal dual que está basado en la convergencia y complejidad del método de Newton y del método barrera.
- 3.3. En las demostraciones hemos usado propiedades que se pueden corroborar con la gráfica de las funciones involucradas, haciendo uso del software Mathcad. Lo cual nos permite visualizar las propiedades enunciadas, que sustituye la demostración analítica.

CAPÍTULO IV

RESULTADOS

En el presente trabajo se han tenido los siguientes resultados

4.1. El análisis del método de Newton para barreras auto concordantes no depende de constantes desconocidas. Auto concordancia es una propiedad afin invariante, es decir, si aplicamos una transformación lineal de variables a funciones auto concordantes, obtenemos funciones auto concordantes. Por lo tanto la complejidad estimada que obtenemos para el método de Newton aplicado a funciones auto concordantes es independiente de cambio de coordenadas.

4.2. Iteraciones de Newton por paso de centrado

La teoría de complejidad del método de Newton para funciones auto concordantes, demuestra que el número de iteraciones requerida para minimizar una función cerrada f auto concordante estrictamente convexa es acotada por

$$\frac{f(x) - p^*}{\gamma} + c$$

donde x es el punto de inicio, $p^* = \inf_x f(x)$ es el valor óptimo. La constante γ depende solamente de los parámetros de la línea de búsqueda con retroceso α, β y está dada por

$$\frac{1}{\gamma} = \frac{20 - 8\alpha}{\alpha\beta(1 - 2\alpha)^2}$$

La constante c depende solamente de la tolerancia ϵ_{nt} ,

$$c = \log_2 \log_2(1/\epsilon_{nt})$$

Se demostró que



$$\frac{m(\mu - 1 - \log \mu)}{\gamma} + c$$

es una cota de (2.21), y por lo tanto una cota sobre el número de iteraciones requeridas para una iteración de salida del método barrera.

- 4.3. Se ha demostrado en 2.5 que una iteración del algoritmo de punto interior primal dual es el mismo que el de infactibilidad del método de Newton, aplicado al resolver $r_t(x, \lambda, \nu) = 0$ pero modificado para asegurar $\lambda > 0$ y $f(x) < 0$. Los mismos argumentos usados en la prueba de convergencia del método de Newton demuestran que la línea de búsqueda para el método primal dual siempre termina en un número finito de pasos.



CAPÍTULO V

DISCUSIÓN DE RESULTADOS

5.1. La cota encontrada para el número de iteraciones del paso de Newton

$$\frac{m(\mu - 1 - \log \mu)}{\gamma} + c$$

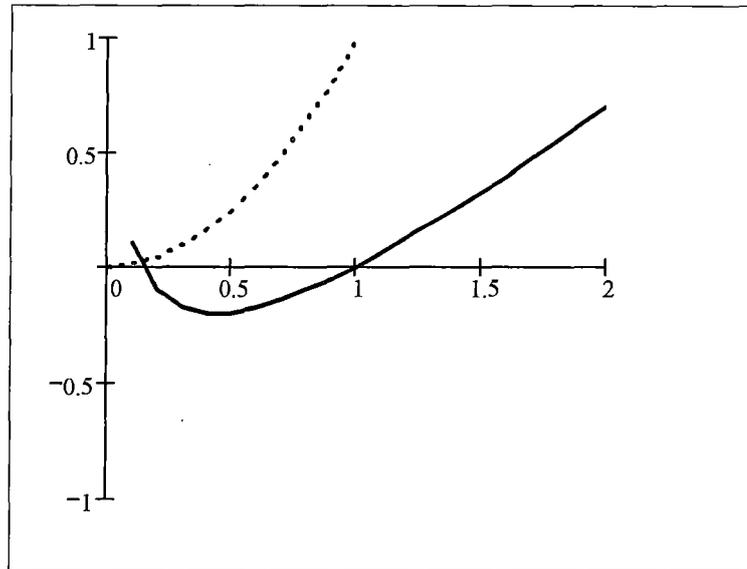
en cada paso de centrado es acotada por una cantidad que depende del factor por el cual t es actualizado en cada iteración de salida del método barrera, y m , el número de restricciones en el problema. También depende débilmente de los parámetros para la línea de búsqueda usados en la línea de búsqueda para las iteraciones internas. Es interesante notar que la cota no depende de n , la dimensión de la variable, o p , el número de ecuaciones restricción, o el valor particular de la data del problema, es decir, la función objetivo y funciones restricción (probado que la condición de auto concordancia se cumple).

La función $m(\mu - 1 - \log \mu)$ se muestra en el gráfico N° 2.5. Para μ pequeño es aproximadamente cuadrático, para μ grande crece aproximadamente linealmente.

En el gráfico N° 5.1. se compara la función $(\mu - 1 - \log \mu)$ con una cuadrática y se puede observar el comportamiento cuadrático para valores pequeños, el comportamiento lineal para valores de μ grandes.

Grafico 5.1

Comportamiento lineal y cuadrático



Gráfica de x^2 en línea punteada y $(\mu - 1 - \log \mu)$ lo cual muestra que para μ pequeño tiene un comportamiento cuadrático y para μ grande tiene un comportamiento lineal.

- 5.2. Métodos de punto interior primal dual son mas eficientes que el método barrera, especialmente cuando se requiere de alta exactitud, puesto que estos exhiben mejor convergencia que la lineal. Para problemas de optimización no convexa, métodos de punto interior primal dual son un tópico de activa investigación.

REFERENCIALES

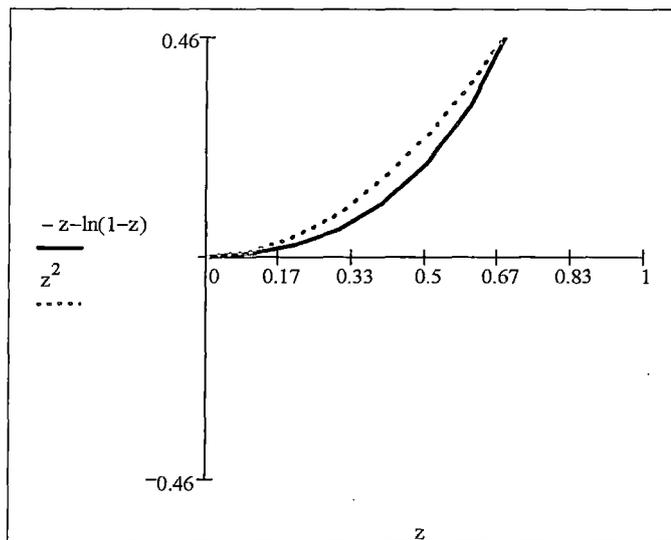
- (1) ARKADII NEMIROVSKII, **Interior point polynomial time methods in convex programming**, Georgia Institute of Technology School of Industrial and Systems Engineering, 2004.
- (2) ARKADII S. NEMIROVSKII AND M. J. TODD, **Interior-point methods for optimization**, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA, Cambridge University Press, 2008.
- (3) YURII NESTEROV AND ARKADII NEMIROVSKII, **Interior- Point Polynomial Algorithms in Convex Programming**, USA SIAM, 1994.
- (4) MAGNUS R. HESTENES, **Conjugate direction methods in optimization**, New York Inc, Springer-Verlag, 1980.
- (5) STEPHEN J. WRIGHT, **Primal dual interior point methods**, Philadelphia, Society for industrial and applied Mathematics, 1997.
- (6) STEPHEN BOYD, **Convex Optimization**, New York, Cambridge University Press, 2009.



APENDICE

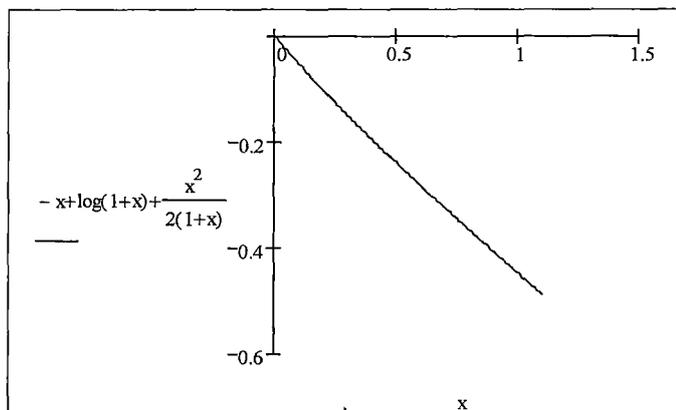
Formato de los gráficos elaborados con el software Mathcad

Gráfica 2.1



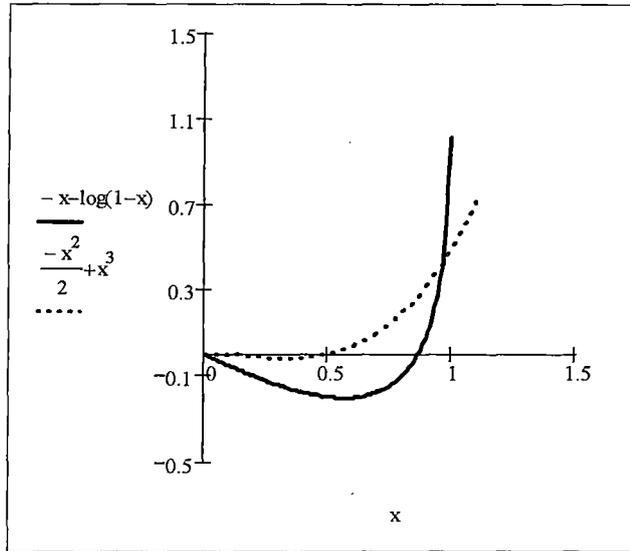
Gráfica de $-z - \ln(1-z)$ en línea sólida y z^2 en línea punteada. Gráfica elaborada con el software Mathcad.

Gráfica 2.2



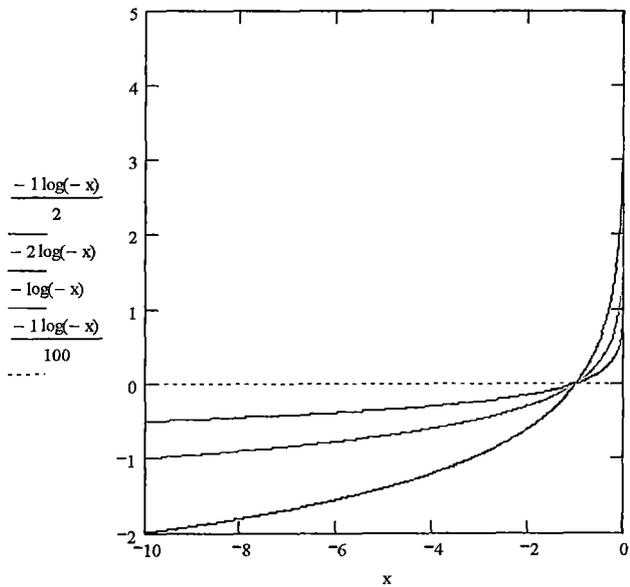
Gráfica de $-x + \log(1+x) + \frac{x^2}{2(1+x)} \leq 0$, en donde se nota que es estrictamente decreciente. Gráfica elaborada con el software Mathcad.

Gráfica 2.3



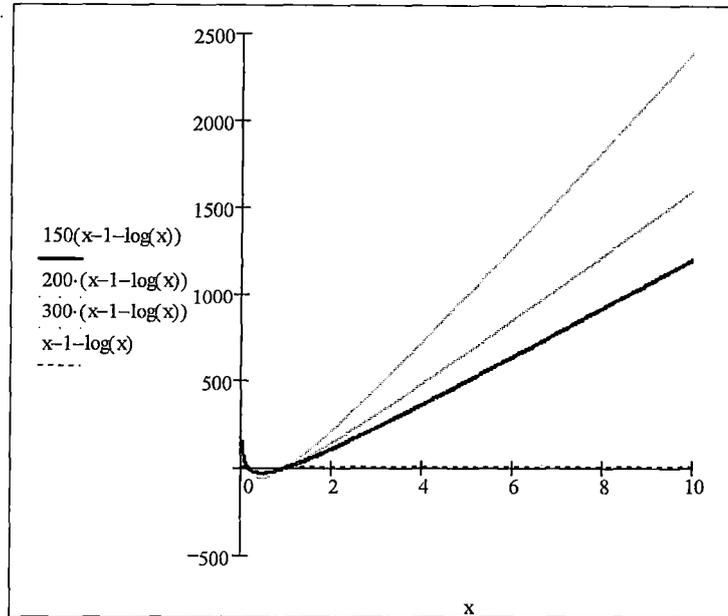
Gráfica de $-x - \log(1-x)$ en línea sólida y $-\frac{1}{2}x^2 + x^3$ en línea punteada para $0 \leq x \leq 0.81$ Gráfica elaborada con el software Mathcad.

Gráfica 2.4



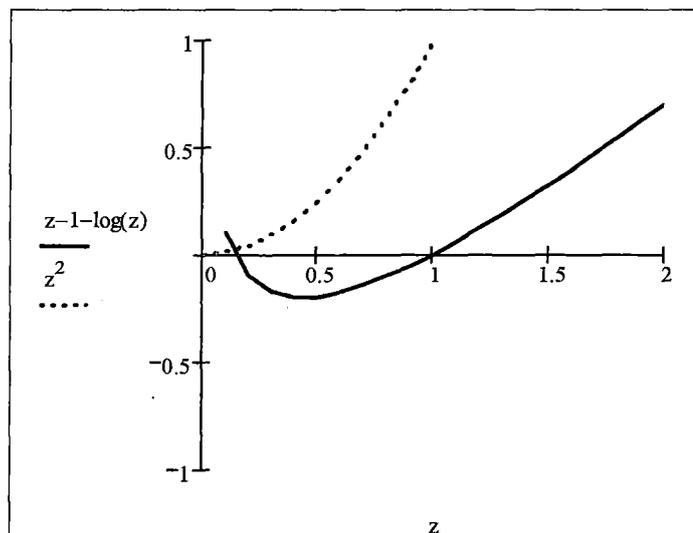
Gráfica de la función indicatriz para $t=2$, $t=1$, $t=1/2$. Gráfica elaborada con el software Mathcad.

Gráfica 2.5



Gráfica de $m(\mu - 1 - \log \mu)$ para $m = 150, 200, 300$. Gráfica elaborada con el software Mathcad

Gráfica 5.1



Gráfica de z^2 en línea punteada y $(\mu - 1 - \log \mu)$ lo cual muestra que para μ pequeño tiene un comportamiento cuadrático y para μ grande tiene un comportamiento lineal.

ANEXOS

Matriz de consistencia

PROBLEMA	objetivo	Hipótesis
<p>¿Cómo las propiedades básicas de barreras auto concordantes conducen a demostrar la convergencia y la complejidad polinómica del algoritmo provisto por el método de punto interior: primal dual?</p>	<p>El objetivo en el trabajo es utilizar las propiedades de barreras auto concordantes para demostrar la convergencia del método de punto interior: primal dual y la complejidad polinómica del mismo</p>	<p>Las propiedades de las barreras auto concordantes nos permiten demostrar la convergencia y complejidad polinomial del método de punto interior primal dual.</p>

