

**UNIVERSIDAD CENTROCCIDENTAL
"LISANDRO ALVARADO"**

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

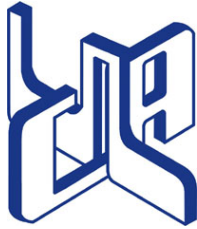


**"NUEVO METODO DE BUSQUEDA LINEAL INEXACTA
PARA OPTIMIZACION SIN RESTRICCIONES.."**

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

José Quiroz

COMO REQUISITO FINAL
PARA OBTENER EL TÍTULO DE LICENCIADA
EN CIENCIAS MATEMÁTICAS
ÁREA DE CONOCIMIENTO: **OPTIMIZACIÓN**
TUTOR: PROF. JOSÉ MEZA



Universidad Centroccidental
 "Lisandro Alvarado"
 Decanato de Ciencias y Tecnología
 Licenciatura en Ciencias Matemáticas



ACTA
 TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designados por el Jefe del Departamento de Matemáticas del Decanato de Ciencias y Tecnología de la Universidad Centroccidental "Lisandro Alvarado", para examinar y dictar el veredicto sobre el Trabajo Especial de Grado titulado:

"NUEVO METODO DE BUSQUEDA LINEAL INEXACTA PARA OPTIMIZACION SIN RESTRICCIONES.."

Presentado por la ciudadana JOSÉ QUIROZ titular de la Cédula de Identidad N° 20.044.814. Con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciada en Ciencias Matemáticas.

Luego de realizada la Defensa y en los términos que imponen los Lineamientos para el Trabajo Especial de Grado de la Licenciatura en Ciencias Matemáticas, se procedió a discutirlo con el interesado habiéndose emitido el veredicto que a continuación se expresa:

1 _____

Con una calificación de _____ puntos.

En fe de lo expuesto firmamos la presente Acta en la Ciudad de Barquisimeto a los _____ días del mes de _____ de _____.

TUTOR

FIRMA

JURADO

FIRMA

JURADO

FIRMA

OBSERVACIONES:

¹ Aprobado ó Reprobado

Universidad Centroccidental

“Lisandro Alvarado”

**“NUEVO METODO DE BUSQUEDA LINEAL INEXACTA
PARA OPTIMIZACION SIN RESTRICCIONES..”**

José Quiroz

Tutor: Prof. José Meza

RESUMEN

Un método de descenso eficiente para resolver problemas de optimización sin restricciones es el método de búsqueda lineal en el cual se quiere elegir el tamaño de paso, después de determinar una dirección de descenso. Hay muchas formas de elegir el tamaño de paso, como son la búsqueda lineal exacta ó la búsqueda lineal inexacta: Armijo, Goldstein, Wolfe, entre otras.

En este trabajo se desarrolla una nueva búsqueda lineal inexacta para problemas de métodos de descenso en general propuesto por ZHEN JUN y JIE SHEN (Ver: [10]), el cual es una modificación del método de Armijo, se establecen algunas propiedades de convergencia global. Se muestran algunos resultados numéricos donde se evidencia que esta nueva búsqueda lineal es mucho mas eficiente que la de Armijo.

El presente trabajo es una prueba de "la perseverancia de quien quiso llegar a ser lo que alguna vez se propuso", del Apoyo y el Amor que me brindan mis padres y hermanos , del estímulo de mis amigos, y de mi Profesor tutor José Meza

AGRADECIMIENTOS

A la fuerza motivadora de que todo lo puede: DIOS, por haberme guiado a través de las sendas de este mundo tan complejo y distinto a mí que es LA MATEMÁTICA, por darme el conocimiento, la fortaleza y las herramientas necesarias para seguir adelante y vencer el desánimo que en muchas situaciones me embargaban.

Ser hijo de Luisa Ortiz y José Isaias Quiroz es mi mayor premio en la vida; gracias a su educación, su apoyo, a la perseverancia, Mis dos hermanos Pedro luis y Luisannys Carolina que me dieron pues los motivos por el cual jamás me rendí. Sé lo mucho que me aman familia, y ese amor es recíproco .

Sra. Aura, Usted ha sido mi segunda madre, ha estado pendiente de mi y de mis hermanos desde niños y también le agradezco este logro.

La amistad duplica las alegrías y divide las tristezas, esa alegría la pude conseguir en ustedes Javier Montes, Betzabet Tovar, Francisco López, Kissy Álvarez, Dannimar Gonzales, Alejandro Gil, Gustavo Castillo, Hernan Terán , Yuribel Agüero, Ana López, Joelviz Vizcalla, Jesus Valadares. Dejaron ser mis compañeros y se convirtieron en mis buenos amigos.

Y eternamente agradecido a quien fortaleció mis conocimientos, vio potencial en Mí para seguir en esta carrera y me dio la oportunidad de trabajar con ella reforzando mis ideas con su conocimiento: A mi tutor Prof. Jose Meza.

ÍNDICE

Agradecimientos	i
Introducción	1
Capítulo 1. Preliminares	3
1.1. Definiciones Básicas	3
1.2. Métodos de direcciones de descensos	5
1.3. Tipos de Convergencia	8
Capítulo 2. Consideración de la nueva Búsqueda lineal Inexacta	10
2.1. Definición de la Nueva Búsqueda lineal Inexacta	10
Capítulo 3. Radio de Convergencia Con la Nueva Búsqueda lineal inexacta	14
3.1. Análisis de Convergencia Global	14
3.2. Radio de Convergencia Lineal	20
Capítulo 4. Resultados Numéricos	24
4.1. Estimación de Parámetros L_k	24
4.2. Resultados Numéricos	25
Conclusiones	28
Referencias	29

INTRODUCCIÓN

Los métodos de descenso son una técnica general para resolver problemas de optimización y la mayor parte de éstos son de búsqueda lineal que requieren un procedimiento después de determinar una dirección de búsqueda en cada iteración. Es un hecho conocido que la búsqueda lineal exacta es ideal pero difícil o incluso imposible para aplicar en el cálculo práctico. Como resultado, muchas personas han estudiado varias búsquedas lineales inexactas como la de Armijo, la de Goldstein y la de Wolfe, entre otras.

Aparentemente, la búsqueda lineal de Armijo es más fácil de aplicar y más útil en los cálculos prácticos que la búsqueda lineal exacta. La ventaja más importante de la búsqueda lineal de Armijo es que nos permite estimar un tamaño de paso inicial de pruebas, el cual llamaremos s . Una buena estimación de s nos puede hacer disminuir las evaluaciones de las funciones objetivos en cada iteración.

Algunos resultados importantes de convergencia global de varios métodos de la antes mencionada búsqueda lineal se han presentado, véase, por ejemplo Refs. [1] - [7] Los métodos No-monótonos búsqueda lineal también han sido investigados por muchos autores, ver por ejemplo Refs [13] - [15] De hecho, el método Barzilai-Borwein (véanse las referencias [16] - [19]) es una forma de descenso no monótona, que es un algoritmo eficiente para la solución de algunos problemas especiales.

En este trabajo, se desarrolla una nueva búsqueda lineal que es similar a la regla Armijo, presentado por ZHEN JUN y JIE SHEN (Ver: [1]). Esta nueva búsqueda lineal permite elegir el tamaño de paso más grande en cada iteración y además reducir el número de evaluaciones de función objetivo en cada paso. En algunos casos especiales, el nuevo método de descenso puede reducir al método de Barzilai y Borwein Ref. [17] - [19] que es considerado como un algoritmo eficiente. Los resultados numéricos muestran que estos nuevos métodos de búsqueda lineal son eficientes para resolver problemas de optimización sin restricciones.

El resto de este trabajo se organiza de la siguiente manera:

Capítulo I: Preliminares.

En este capítulo se enuncian algunas definiciones básicas, teoremas y métodos fundamentales en la teoría de optimización, los cuales servirán de base a los posteriores capítulos.

Capítulo II: Consideración de una Nueva Búsqueda Lineal Inexacta.

Este capítulo describe el objetivo central de este trabajo, que es presentar la nueva búsqueda lineal inexacta y establecer el algoritmo relacionado con esta nueva búsqueda.

Capítulo III: Radio de Convergencia con la Nueva Búsqueda Lineal Inexacta.

Este capítulo determina el radio de convergencia de esta nueva búsqueda, tal como: El radio de convergencia Lineal, Para ello se plantean antes algunos teoremas y lemas indispensables a la hora de determinar el radio de convergencia.

Capítulo IV: Experimentos Numéricos.

Para finalizar, este capítulo muestra algunos reportes numéricos de esta nueva búsqueda lineal inexacta comparada con la búsqueda lineal inexacta de Armijo.

Conclusiones y algunas referencias Bibliográficas.

CAPÍTULO 1

PRELIMINARES

§1.1. DEFINICIONES BÁSICAS

A continuación desarrollaremos y trataremos algunas definiciones, métodos y teoremas esenciales para la realización de este trabajo.

DEFINICIÓN 1.1.1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continuamente diferenciable y sea $x^* \in \mathbb{R}^n$.

- Diremos que x^* es un MÍNIMO LOCAL de f si existe $\epsilon > 0$ y una vecindad V_ϵ de x^* tal que: $f(x^*) \leq f(x)$, para todo $x \in V_\epsilon$.
- Diremos que x^* es un MÍNIMO LOCAL ESTRICTO de f si existe $\epsilon > 0$ y una vecindad V_ϵ de x^* tal que: $f(x^*) < f(x)$, para todo $x \in V_\epsilon$.
- Diremos que x^* es un MÍNIMO LOCAL AISLADO de f si existe $\epsilon > 0$ y una vecindad V_ϵ de x^* tal que x^* es el único mínimo local en V_ϵ .
- Diremos que x^* es un MÍNIMO GLOBAL de f si, y sólo si $f(x^*) \leq f(x)$, para todo $x \in \mathbb{R}^n$.

DEFINICIÓN 1.1.2. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continuamente diferenciable y sea $x^* \in \mathbb{R}^n$.

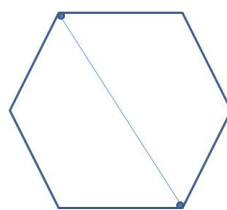
- Diremos que x^* es un MÁXIMO LOCAL de f si, y sólo si x^* es un MÍNIMO LOCAL de $-f$.
- Diremos que x^* es un MÁXIMO GLOBAL de f si, y sólo si x^* es un MÍNIMO GLOBAL de $-f$.

DEFINICIÓN 1.1.3. Una función $f : M \rightarrow N$ entre espacios métricos M y N se dice que es Lipschitz continua si existe una constante $K > 0$ tal que $d(f(x), f(y)) \leq Kd(x, y)$ para todo x e y en M . En tal caso, K es llamada la constante Lipschitz de la función.

DEFINICIÓN 1.1.4. (Conjunto Convexo). Sea C un subconjunto de \mathbb{R}^n , C se dice convexo si y sólo si $\forall x, y \in C$, se tiene que:

$$ax + (1 - \alpha)y \in C, \quad \forall \alpha \in (0, 1).$$

Es decir, un conjunto es convexo, cuando el segmento de recta que une a cualquier par de puntos del conjunto está totalmente contenida en éste.



Conjunto Convexo

FIGURA 1.1.1: Conjunto convexo

DEFINICIÓN 1.1.5. Una función f definida en un conjunto convexo $\Omega \subset \mathbb{R}^n$, es convexa si para todo $x, \tilde{x} \in \Omega$ y todo $\alpha, 0 \leq \alpha \leq 1$, se cumple:

$$f(\alpha x + (1 - \alpha)\tilde{x}) \leq \alpha f(x) + (1 - \alpha)f(\tilde{x})$$

o equivalentemente,

$$f(\alpha x + (1 - \alpha)\tilde{x}) \leq f(\tilde{x}) + \alpha(f(x) - f(\tilde{x}))$$

Si, además para todo $\alpha, 0 < \alpha < 1$ y $x \neq \tilde{x}$, se cumple:

$$\begin{aligned} f(\alpha x + (1 - \alpha)\tilde{x}) &< \alpha f(x) + (1 - \alpha)f(\tilde{x}) \\ &= f(\tilde{x}) + \alpha(f(x) - f(\tilde{x})) \end{aligned}$$

Entonces, se dice que f es estrictamente convexa.

TEOREMA 1.1.1. Sea $f \in C^1$. Entonces, f es convexa en un conjunto convexo $\Omega \subset \mathbb{R}^n$ si, y sólo si

$$f(\tilde{x}) \geq f(x) + \nabla f(x)(\tilde{x} - x), \text{ para todo } x, \tilde{x} \in \Omega$$

DEFINICIÓN 1.1.6. Sea A una matriz simétrica de orden $n \times n$ y P una matriz de orden $n \times 1$.

- Diremos que A es DEFINIDA POSITIVA si para todo $P \neq$ vector nulo, se cumple:

$$P^t A P > 0$$

- Diremos que A SEMIDEFINIDA POSITIVA si para todo P , se cumple:

$$P^t A P \geq 0$$

TEOREMA 1.1.2. Sea $\Omega \subset \mathbb{R}^n$ un conjunto convexo y sea $f : \Omega \rightarrow \mathbb{R}$ de clase C^2 .

- Si la matriz hessiana $\nabla^2 f(x)$ es semidefinida (definida) positiva para todo $x \in \Omega$, entonces f es convexa (estrictamente).

- Si $\Omega = \mathbb{R}^n$ y f es convexa (estrictamente), entonces $\nabla^2 f(x)$ es semidefinida (definida) positiva.
- Si $f(x) = x^t Q x$, $Q \in \mathbb{M}_{n \times n}$, f es convexa si y sólo si Q es semidefinida positiva.

DEFINICIÓN 1.1.7. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable, $x^* \in \mathbb{R}^n$. Diremos que x^* es un punto estacionario de f si, y sólo si $\nabla f(x^*) = 0$.

TEOREMA 1.1.3. Sea $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa en un conjunto convexo Ω .

- Cualquier mínimo local $x^* \in \mathbb{R}^n$, es un mínimo global de f . Además, si f es diferenciable, entonces todo punto estacionario x^* es un mínimo global de f .
- Si f es estrictamente convexa, existe a lo sumo un mínimo.

TEOREMA 1.1.4. Sea $f \in C^2$, $x^* \in \mathbb{R}^n$. Sea x^* un punto estacionario y $\nabla^2 f(x^*)$ es indefinido. Entonces, x^* es un punto silla.

DEFINICIÓN 1.1.8. Sea f una función diferenciable, $x^* \in \mathbb{R}^n$.

- Si x^* satisface $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es definida positiva, entonces x^* es un mínimo local no singular.
- Si x^* satisface $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es semidefinida positiva, entonces x^* es un mínimo local singular.

TEOREMA 1.1.5. Todo mínimo local no singular es localmente único.

§1.2. MÉTODOS DE DIRECCIONES DE DESCENSOS

Consideremos el problema:

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

donde f es una función diferenciable en todo $x \in \mathbb{R}^n$.

Este problema se puede resolver aplicando los métodos de descenso. Estos métodos generan una secuencia de puntos que van reduciendo el valor de la función objetivo, ó más formalmente generan una sucesión $\{x_k\}_{k \geq 1}$ que cumple $f(x_1) > f(x_2) > \dots > f(x_k) > \dots$. Dichos procedimientos terminan cuando se satisface algún criterio de parada.

La familia más importante de estos métodos son los llamados métodos de dirección de descenso, los cuales constan de dos pasos fundamentales. En el primero se selecciona una dirección de descenso

y en el segundo se calcula el desplazamiento a lo largo de esa dirección. Estos métodos están definidos por el algoritmo iterativo:

$$x_{k+1} = x_k + \alpha_k d_k$$

Donde α_k es un escalar no-negativo que minimiza a $f(x_k + \alpha d_k)$, $\alpha \geq 0$, y d_k es una dirección de descenso.

Los métodos de dirección de descenso convergen a puntos candidatos a ser mínimos locales, y paran cuando el punto actual está lo suficientemente próximo a uno de estos candidatos. La regla de parada más utilizada es $\|\nabla f(x_k)\| \leq \epsilon$, está basada en la condición necesaria de optimalidad $\nabla f(x) = 0$ y en la continuidad de la función $\nabla f(x)$ que garantiza que la norma del gradiente es pequeña en un entorno del mínimo.

DEFINICIÓN 1.2.1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $\bar{x} \in \mathbb{R}^n$. Diremos que $d \in \mathbb{R}^n$ es una dirección de descenso de f en \bar{x} si y sólo si existe $\bar{\alpha} > 0$ tal que $f(\bar{x} + \alpha d) < f(\bar{x})$ para todo $\alpha \in (0, \bar{\alpha})$.

TEOREMA 1.2.1 (Caracterización de direcciones de descenso mediante el vector gradiente).
Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de clase C^1 en $\bar{x} \in \mathbb{R}^n$. Sea d un vector de \mathbb{R}^n . Diremos que d es una dirección de descenso de f en \bar{x} si y sólo si

$$\nabla f(\bar{x})^t d < 0$$

Demostración:

Supongamos que d es dirección de descenso de f en \bar{x} .

Así, por definición, existe $\alpha > 0$ tal que $f(\bar{x} + \alpha d) < f(\bar{x})$.

Luego, utilizando expansión de Taylor de orden 1, tenemos

$$f(\bar{x} + \alpha d) = f(\bar{x}) + \alpha \nabla f(\bar{x}) d + o(\alpha)$$

Entonces;

$$0 > f(\bar{x} + \alpha d) - f(\bar{x}) = \alpha (\nabla f(\bar{x}) d + \frac{o(\alpha)}{\alpha}) \implies (\nabla f(\bar{x}) d + \frac{o(\alpha)}{\alpha}) < 0$$

Pero, $\frac{o(\alpha)}{\alpha} \rightarrow 0$ cuando $\alpha \rightarrow 0$. Es decir, para un ϵ dado, existe un α tal que $\frac{o(\alpha)}{\alpha} \rightarrow 0$ puede hacerse muy pequeño.

Por tanto,

$$\nabla f(\bar{x}) d < 0$$

Recíprocamente, supongamos que $\nabla f(\bar{x}) d < 0$.

Igualmente por expansión de Taylor,

$$f(\bar{x} + \alpha d) - f(\bar{x}) = \alpha (\nabla f(\bar{x}) d + \frac{o(\alpha)}{\alpha})$$

Como $\nabla f(\bar{x})d < 0$ y por argumentación anterior $\frac{o(\alpha)}{\alpha}$ puede hacerse muy pequeño.

Podemos decir que, existe α tal que $f(\bar{x} + \alpha d) < f(\bar{x})$

Por tanto, d es dirección de descenso de f en \bar{x} .

Hay una estructura fundamental subyacente para casi todos los algoritmos descendentes que se analizan. Se comienza en un punto inicial, se determina, de acuerdo a una regla fija, una dirección de movimiento y después se sigue esa dirección hacia un mínimo (relativo) de la función objetivo de esa recta. En el punto nuevo se determina una nueva dirección, y se repite el proceso. Las diferencias entre algoritmos radican en la regla mediante la cual se seleccionan las direcciones sucesivas del movimiento. Una vez que se ha hecho la selección, todos los algoritmos exigen movimiento hacia el punto mínimo de la recta correspondiente. Este proceso de determinar el punto mínimo de una recta dada se denomina **Búsqueda lineal**.

Hay varias alternativas de reglas de búsqueda lineal para elegir α_k a lo largo de la recta $S = x_k + \alpha d_k | \alpha > 0$. A saber:

(A) Regla de Minimización. En cada iteración, α_k se selecciona de manera que

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k) \quad (1.2.1)$$

(B) la regla de minimización aproximada. En cada iteración, una k se selecciona de manera que

$$\alpha_k = \min\{\alpha | g(x_k + \alpha d_k)^T d_k = 0, \alpha > 0\} \quad (1.2.2)$$

(C) la regla de Armijo. Sean los escalares $s > 0$, $\beta \in (0, 1)$, y $\sigma \in (0, \frac{1}{2})$. Definiremos α_k como el más grande α en el conjunto $\{s, \beta s_k, \beta^2 s, \dots\}$ de tal manera que

$$f_k - f(x_k + \alpha d_k) \geq -\sigma \alpha g_k^T d_k. \quad (1.2.3)$$

Algoritmo 1.2.1: Regla de Armijo

Paso 1: Elegir $s > 0$, $\beta \in (0, 1)$ y $\sigma \in (0, \frac{1}{2})$. Se toma $\alpha = s$.

Paso 2: Si $f(x_k + \alpha d_k) \leq f(x_k) + \sigma \alpha g_k^T d_k$, entonces se para y se toma $\alpha_k = \alpha$. En caso contrario se va al paso 3.

Paso 3: Hacer $\alpha = \alpha \beta$ y se va al paso 2.

(D) Regla de Minimización limitada. Sea $s_k = -\frac{g_k^T d_k}{L \|d_k\|^2}$ donde α_k es definido por

$$f(x_k + \alpha_k d_k) = \min_{\alpha \in [0, s_k]} f(x_k + \alpha d_k) \quad (1.2.4)$$

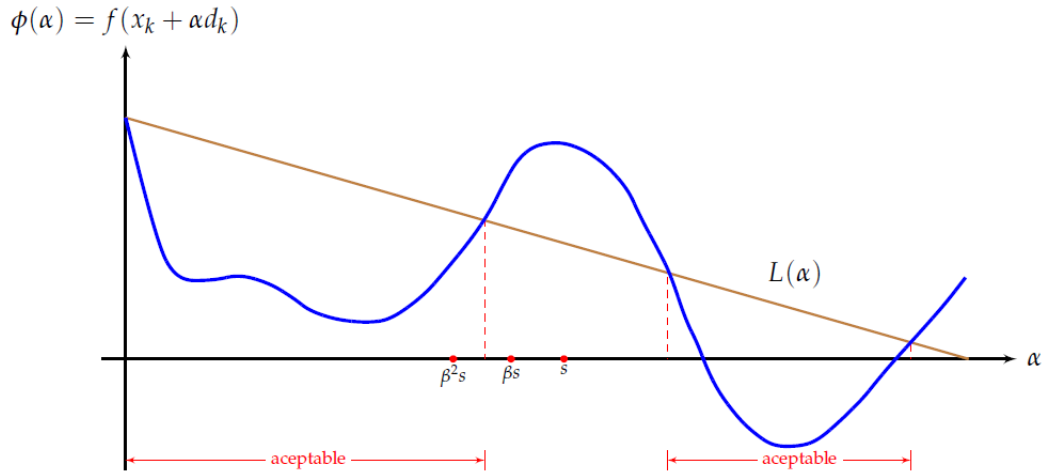


FIGURA 1.2.1: Regla de Armijo

y $L > 0$ es una constante dada.

(E) **Regla de Goldstein.** Un escalar fijo $\sigma \in (0, \frac{1}{2})$ es seleccionado, y α_k se elige con el fin de satisfacer

$$\sigma \leq \frac{f(x_k + \alpha_k d_k) - f_k}{\alpha_k g_k^T d_k} \leq 1 - \sigma \tag{1.2.5}$$

(F) **Regla Fuerte Wolfe.** En la k -ésima iteración, α_k satisface al mismo tiempo

$$f_k - f(x_k + \alpha_k d_k) \geq -\sigma \alpha_k g_k^T d_k \tag{1.2.6}$$

y

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\beta g_k^T d_k, \tag{1.2.7}$$

donde $\sigma \in (0, \frac{1}{2})$ y $\beta \in (\sigma, 1)$

(G) **Regla de Wolfe.** En la k -ésima iteración, α_k satisface (12) y

$$g(x_k + \alpha_k d_k)^T d_k \geq \beta g_k^T d_k \tag{1.2.8}$$

§1.3. TIPOS DE CONVERGENCIA

DEFINICIÓN 1.3.1 (Orden de Convergencia). Sea la sucesión $\{x_k\} \in \mathbb{R}^n$ convergente a x^* . Se define el orden de convergencia de dicha sucesión como el máximo de los números no-negativos p que

satisfagan :

$$0 \leq \limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} < \infty$$

DEFINICIÓN 1.3.2 (Tipos de Convergencia). Si la sucesión $\{x_k\} \in \mathbb{R}^n$ converge a x^* para todo k . Para $\alpha \geq 1, \alpha \in \mathbb{R}^n$ de manera que:

- Decimos que la sucesión converge a x^* por lo menos con orden α o que la convergencia es por lo menos de orden α , si existen $C > 0$ y un $N \in \mathbb{N}$ tales que

$$\|x_{k+1} - x^*\| \leq C \|x_k - x^*\|^\alpha \quad \text{para } k \geq N$$

siendo $C < 1$ cuando $\alpha = 1$; en este último caso se dice que la convergencia es por lo menos lineal o que la sucesión converge a x^* por lo menos linealmente.

- Se tiene que la sucesión converge Q–linealmente si existe una constante $r \in (0, 1)$ tal que $\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq r$, para toda k suficientemente grande.
- La sucesión converge R–linealmente si existe una secuencia de escalares no–negativos tal que $\|x_k - x^*\| \leq V_k$ para toda k , y $\{V_k\}$ convergiendo Q–linealmente a cero.
- $\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \beta < 1$, la sucesión converge linealmente a x^* con tasa de convergencia β .
- $\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$, la sucesión converge superlinealmente.
- $\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = c < \infty$, la sucesión converge cuadráticamente.

CAPÍTULO 2

CONSIDERACIÓN DE LA NUEVA BÚSQUEDA LINEAL INEXACTA

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continuamente diferenciable. Un método de búsqueda lineal que resuelve el problema:

$$\min f(x); \quad x \in \mathbb{R}^n. \quad (2.0.1)$$

esta definida por la ecuación

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, 3, \dots \quad (2.0.2)$$

donde $x_0 \in \mathbb{R}^n$ es un punto inicial, d_k es una dirección de descenso de $f(x)$ en x_k , y α_k es el tamaño de paso. Denotamos el gradiente $\nabla f(x_k)$ por g_k y el valor de la función $f(x_k)$ por f_k .

La elección de la dirección búsqueda d_k y la determinación del tamaño de paso α_k a lo largo de la dirección en cada iteración son las principales tareas del método de búsqueda lineal. La dirección d_k requiere generalmente satisfacer que

$$g_k^T d_k < 0, \quad (2.0.3)$$

esto garantiza que d_k es una dirección de descenso de $f(x)$ en x_k . Para garantizar la convergencia global, se requiere a veces que d_k satisfaga condición de descenso suficiente

$$g_k^T d_k \leq -c \|g_k\|^2 \quad (2.0.4)$$

donde $c > 0$ es una constante.

Después que se obtiene d_k , la siguiente tarea es encontrar un tamaño de paso α_k a lo largo de la dirección de búsqueda para satisfacer la propiedad de descenso:

$$f(x_k + \alpha_k d_k) < f(x_k), \quad k = 0, 1, 2, \dots,$$

§2.1. DEFINICIÓN DE LA NUEVA BÚSQUEDA LINEAL INEXACTA

A lo largo del trabajo se suponen las siguientes hipótesis

(H1) La función $f(x)$ tiene un límite inferior en el conjunto de nivel

$$L_0 = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}, \text{ donde } x_0 \text{ es dado}$$

(H2) El gradiente $g(x)$ de $f(x)$ es Lipschitz continua en un conjunto abierto convexo \mathbf{B} que contiene

a L_0 ; i.e, Existe L tal que $\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \mathbf{B}$.

Definición de la búsqueda de Armijo modificada:

Dados s_k, β, L_k, μ , y σ con $s_k = \frac{-g_k^T d_k}{L_k \|d_k\|^2}$, $\beta \in (0, 1)$, $L_k > 0$, $\mu \in [0, 2)$, y $\sigma \in (0, \frac{1}{2})$. Sea α_k el valor mas grande α en el conjunto $\{s_k, \beta s_k, \beta^2 s_k, \dots\}$ tal que satisface:

$$f(x_k + \alpha d_k) - f(x_k) \leq \sigma \alpha [g_k^T d_k + \frac{1}{2} \alpha \mu L_k \|d_k\|^2]. \quad (2.1.1)$$

El algoritmo que permite obtener el tamaño de paso α_k es el siguiente:

Algoritmo 2.1.1: Nueva búsqueda lineal inexacta

Paso 1: Elegir $\beta \in (0, 1)$, $L_k > 0$, $\mu \in [0, 2)$, $\sigma \in (0, \frac{1}{2})$ y $s_k = \frac{-g_k^T d_k}{L_k \|d_k\|^2}$. Se toma $\alpha = s_k$.

Paso 2: Si $f(x_k + \alpha d_k) - f(x_k) \leq \sigma \alpha [g_k^T d_k + \frac{1}{2} \alpha \mu L_k \|d_k\|^2]$, entonces se para y se toma $\alpha_k = \alpha$. En caso contrario se va al paso 3.

Paso 3: Hacer $\alpha = \alpha \beta$ y se va al paso 2.

En esta sección, no hablamos de cómo elegir d_k en cada iteración, sino investigar cómo elegir el tamaño de paso α_k . Algunas de las reglas de búsqueda lineal que han sido mencionada en la sección anterior son útiles. Se describe aquí una nuevo regla de búsqueda lineal inexacta que contiene la regla de Armijo como caso especial. El tamaño de paso definido en esta nueva regla es mayor que la definida en la norma original de la regla de búsqueda lineal de Armijo . En otras palabras, el tamaño de paso definido por la nueva búsqueda lineal es más fácil de encontrar que con la definida por la regla de Armijo.

Observación 1:

1.1 De (2.1.1) tenemos:

$$\begin{aligned} f(x_k + \alpha d_k) &\leq f(x_k) + \sigma \alpha [g_k^T d_k + \frac{1}{2} \alpha \mu L_k \|d_k\|^2] \\ f(x_k + \alpha d_k) &= f(x_k) + \alpha \sigma g_k^T d_k + \frac{\alpha^2}{2} \sigma \mu L_k \|d_k\|^2 \end{aligned}$$

Definamos $C(\alpha) = f(x_k) + \sigma \alpha [g_k^T d_k + \frac{1}{2} \alpha \mu L_k \|d_k\|^2]$; y $L(\alpha) = f(x_k) + \alpha \sigma g_k^T d_k$

Claramente $C(\alpha) \leq L(\alpha)$ ya que $\frac{\alpha^2}{2} \sigma \mu L_k \|d_k\|^2 > 0$

Derivando

$$\begin{aligned} C'(\alpha) &= \sigma g_k^T d_k + \alpha \mu L_k \|d_k\|^2 = 0 \Rightarrow \bar{\alpha} = -\frac{g_k^T d_k}{\mu L_k \|d_k\|^2} \\ C''(\alpha) &= \mu \sigma L_k \|d_k\|^2 > 0 \end{aligned}$$

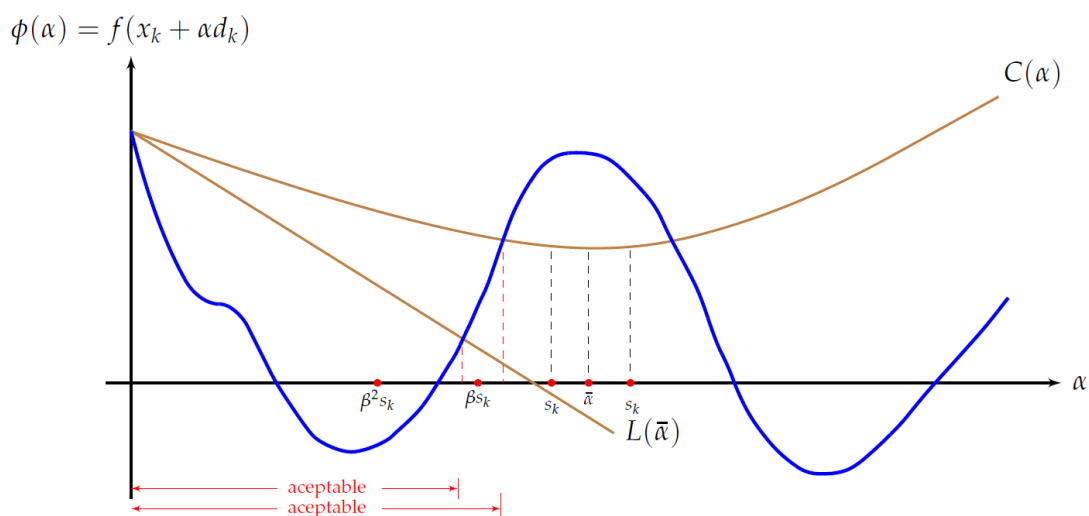


FIGURA 2.1.1: Nueva Búsqueda Lineal Vs. Armijo

Por lo tanto $\bar{\alpha}$ es un mínimo absoluto

En la nueva Búsqueda lineal el backtracking comienza con $s_k = -\frac{g_k^T d_k}{L_k \|d_k\|^2}$

si $\mu \in (0, 1)$ entonces $s_k < \bar{\alpha}$

si $\mu \in [1, 2)$ entonces $\bar{\alpha} < s_k$

1.2 Supongamos que α_k se define por el método de Armijo y que α'_k se define por la nueva búsqueda lineal inexacta, entonces, $\alpha'_k \geq \alpha_k$. En efecto,

Sea α'_k que satisface:

$$\phi(\alpha'_k) \leq C(\alpha'_k)$$

Observemos que si tomamos $s = s_k$ en la búsqueda lineal de Armijo y α_k satisface:

$\phi(\alpha_k) \leq L(\alpha_k)$ entonces se tiene que:

$$L(\alpha_k) = f(x_k) + \alpha_k \sigma g_k^T d_k \leq f(x_k) + \alpha_k \sigma g_k^T d_k + \frac{1}{2} \alpha^2 \sigma \mu L_k \|d_k\|^2$$

Es decir, α_k también satisface:

$$\phi(\alpha_k) \leq C(\alpha_k)$$

Por lo tanto,

$$\alpha_k \leq \alpha'_k$$

En otras palabras, A^c denota el conjunto del tamaño de paso de Armijo y sea $A^{c'}$ el conjunto de tamaños de pasos nuevos, entonces; tenemos $A^c \subseteq A^{c'}$. De hecho, si $L_k \equiv L$ y existe α_k satisfaciendo (1.2.3), entonces α_k es cierto para satisfacer (2.1.1)

Por otra parte, si $\mu = 0$, entonces la nueva de búsqueda lineal se reduce a la regla de Armijo (c).

En lo que sigue, se analiza la convergencia global de la nuevo método de búsqueda lineal.

CAPÍTULO 3

RADIO DE CONVERGENCIA CON LA NUEVA BÚSQUEDA LINEAL INEXACTA

§3.1. ANÁLISIS DE CONVERGENCIA GLOBAL

Teorema 3.1 Supongamos que (H1) y (H2) se cumplen, la búsqueda de la dirección d_k satisface (2.0.3), y α_k se determina por la Nueva Búsqueda lineal. Entonces el Algoritmo genera una secuencia infinita $\{x_k\}$ y

$$0 < L_k \leq m_k L, \quad (3.1.1)$$

Donde m_k es un entero positivo y $m_k \leq M_0 < +\infty$, con M_0 una constante positiva grande. Entonces

$$\sum_{k=1}^{\infty} (g_k^T d_k / \|d_k\|)^2 < +\infty \quad (3.1.2)$$

Demostración: Sean

$$K_1 = \{k | \alpha_k = s_k\}, \quad K_2 = \{k | \alpha_k < s_k\}.$$

Si $k \in K_1$, Entonces

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &\leq \sigma \alpha_k [g_k^T d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2] \\ &= \frac{-\sigma g_k^T d_k}{L_k \|d_k\|^2} [g_k^T d_k + \frac{1}{2} \frac{-g_k^T d_k}{L_k \|d_k\|^2} \mu L_k \|d_k\|^2] \quad (\text{Sustituyendos}_k) \\ &= \frac{-\sigma (g_k^T d_k)^2}{L_k \|d_k\|^2} - \frac{1}{2} \frac{\sigma (g_k^T d_k)^2}{L_k \|d_k\|^2} \mu \\ &= -\left[\frac{\sigma(1 - (1/2)\mu)}{L_k} \right] \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 \end{aligned}$$

Así,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\left[\frac{\sigma(1 - (1/2)\mu)}{L_k} \right] \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2, \quad k \in \mathbf{K}_1 \quad (3.1.3)$$

Sea

$$\eta_k = -\left(\frac{\sigma(1 - (1/2)\mu)}{L_k} \right), \quad k \in \mathbf{K}_1;$$

Pero $L_k \leq m_k L \leq M_0 L$, Entonces

$$\frac{1}{M_0 L} \leq \frac{1}{L_k}$$

Así

$$\eta_k \leq -\left(\frac{\sigma(1-(1/2)\mu)}{M_0 L}\right)$$

Sea

$$\eta' = \left(\frac{\sigma(1-(1/2)\mu)}{M_0 L}\right)$$

Esto implica que $\eta_k \leq -\eta'$ y

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\eta' \left(\frac{g_k^T d_k}{\|d_k\|}\right)^2 \quad k \in \mathbf{K}_1 \quad (3.1.4)$$

Si $k \in \mathbf{K}_2$, entonces $\alpha_k \leq s_k$.

$$f(x_k + \alpha_k d_k) - f_k \leq \sigma \alpha_k [g_k^T d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2]$$

Ya que α_k satisface la nueva búsqueda y claramente se observa que $\leq s_k$ no la satisface, de ser así entonces $\alpha_k \geq s_k$ lo cual sería una contradicción. Por tanto $\alpha_k \leq \beta s_k$

Afirmamos que $\alpha = \frac{\alpha_k}{\beta}$ no puede satisfacer la nueva búsqueda, por tanto

$$f(x_k + \alpha_k d_k / \beta) - f_k > \sigma \alpha_k / \beta [g_k^T d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2 / \beta]$$

Usando el teorema del valor medio en el lado izquierdo de la desigualdad anterior, vemos que $\exists \theta_k \in [0, 1]$ tal que

$$f(x_k + \frac{\alpha_k}{\beta} d_k) - f(x_k) = g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k)^T d_k \frac{\alpha_k}{\beta}$$

$$g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k)^T d_k \frac{\alpha_k}{\beta} > \sigma g_k^T d_k + \frac{\sigma}{2} \mu L_k \frac{\|d_k\|^2}{\beta};$$

Por lo tanto,

$$[g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k]^T d_k > (\sigma - 1) g_k^T d_k + \frac{\sigma}{2} \mu L_k \frac{\|d_k\|^2}{\beta} \quad (3.1.5)$$

Por (H2), la desigualdad de Cauchy-Schwarz, y (3.1.5), obtenemos

$$\begin{aligned} [g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k]^T d_k &\leq \|g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k\| \|d_k\| \\ &\leq L \frac{\theta_k}{\beta} \alpha_k \|d_k\|^2 \\ &\leq \frac{L}{\beta} \alpha_k \|d_k\|^2 \quad 0 \leq \theta_k \leq 1 \end{aligned}$$

Así,

$$L \frac{\theta_k}{\beta} \alpha_k \|d_k\|^2 > -(1 - \sigma) g_k^T d_k + \frac{\sigma}{2} \mu \alpha_k L_k \|d_k\|^2$$

Por tanto

$$L \frac{\theta_k}{\beta} \alpha_k \|d_k\|^2 > -(1 - \sigma) g_k^T d_k,$$

Esto implica que

$$\alpha_k > -\left[\frac{\beta(1 - \sigma)}{L}\right] \frac{g_k^T}{\|d_k\|^2}, \quad k \in \mathbf{K}_2. \quad (3.1.6)$$

Siendo

$$s'_k = -\left[\frac{\beta(1 - \sigma)}{L}\right] \frac{g_k^T d_k}{\|d_k\|^2}, \quad k \in \mathbf{K}_2,$$

Se tiene

$$s_k > \alpha_k > s'_k \quad k \in \mathbf{K}_2, \quad (3.1.7)$$

Por (2.1.1) y (3.1.7), se tiene

$$f_{k+1} - f_k \leq \sigma \alpha_k (g_k^T d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2)$$

Llamemos $C(\alpha_k) = (g_k^T d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2)$

Como $s_k = \frac{-g_k^T d_k}{L_k \|d_k\|^2}$ y por observación 1 se tiene:

$$C(\alpha_k) \leq C(s_k)$$

Pero $C(\alpha)$ es decreciente en $[s'_k, s_k]$, entonces

$$C(\alpha_k) \leq C(s'_k)$$

Por tanto

$$f_{k+1} - f_k \leq \sigma s'_k (g_k^T d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2)$$

Sustituyendo s'_k nos queda

$$f_{k+1} - f_k \leq -[\sigma \beta (1 - \sigma) \frac{(1 - (1/2)\mu)}{L}] \left(\frac{g_k^T d_k}{\|d_k\|}\right)^2$$

Tomando

$$\eta'' = \sigma\beta(1 - \sigma) \frac{(1 - (1/2)\mu)}{L} \quad (3.1.8)$$

Se tiene

$$f_{k+1} - f_k \leq \eta'' \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2, \quad k \in \mathbf{K}_2 \quad (3.1.9)$$

Sea

$\eta_0 = \min(\eta', \eta'')$; Por (3.1.9) y (3.1.9), se tiene

$$f_{k+1} - f_k \leq \eta_0 \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 \leq 0, \quad \forall k \quad (3.1.10)$$

Esto implica que

$$f_{k+1} \leq f_k$$

$\therefore \{f_k\}$ es decreciente

Por otro lado $f_k \leq f_0 = f(x_0)$, es acotada para todo k

$$\{f_k\} \subset L_0$$

Por tanto $\{f_k\}$ es decreciente y está acotada inferiormente, entonces $\{f_k\}$ converge digamos a f

Así por (3.1.10) se tiene que

$$\left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 \leq \frac{1}{n'_0} (f_k - f_{k+1}), \quad \forall k$$

$$\text{Pero } \sum_{k=0}^{+\infty} \frac{1}{n'_0} (f_k - f_{k+1}) = \frac{1}{n'_0} (f_0 - f) < \infty$$

$$\therefore \sum_{k=0}^{+\infty} \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 < \infty$$

Corolario 3.1 si la condición en el Teorema 3.1 se cumple, entonces

$$\lim_{k \rightarrow \infty} \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 = 0. \quad (3.1.11)$$

en efecto, Asumir que (H2) puede ser sustituida por la siguiente hipótesis más débil:

(H2') El gradiente $g(x)$ de $f(x)$ es uniformemente continua en un conjunto abierto convexo B que contiene L_0 .

Teorema 3.2. Supongamos que (H1) y (H2') se cumplen, la búsqueda de la dirección d_k satisface (2.0.3), y α_k se determina por la Nueva Búsqueda lineal. Entonces el Algoritmo genera una secuencia infinita $\{x_k\}$ y

$$0 < L_k \leq M'_0, \quad (3.1.12)$$

Donde M'_0 es una constante positiva grande. Entonces

$$\lim_{k \rightarrow \infty} \left(\frac{-g_k^T d_k}{\|d_k\|} \right) = 0 \quad (3.1.13)$$

Demostración:

como en la demostración del teorema anterior, si $k \in \mathbf{K}_1$, se probó que

$$\lim_{k \in \mathbf{K}_1, k \rightarrow \infty} \left(\frac{-g_k^T d_k}{\|d_k\|} \right)^2 = 0 \quad (3.1.14)$$

En efecto, por (3.1.4) se tiene

$$\begin{aligned} f_{k+1} - f_k &\leq -\eta' \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 \quad k \in \mathbf{K}_1 \\ \Rightarrow \eta' \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 &\leq f_k - f_{k+1} \\ \Rightarrow 0 &\leq \frac{g_k^T d_k}{\|d_k\|} \leq \sqrt{f_k - f_{k+1}} \\ \Rightarrow \lim_{k \in \mathbf{K}_1, k \rightarrow \infty} 0 &\leq \lim_{k \in \mathbf{K}_1, k \rightarrow \infty} \frac{g_k^T d_k}{\|d_k\|} \leq \lim_{k \in \mathbf{K}_1, k \rightarrow \infty} \sqrt{f_k - f_{k+1}} \end{aligned}$$

Por el teorema del Emparedado se obtiene

$$\lim_{k \in \mathbf{K}_1, k \rightarrow \infty} \frac{g_k^T d_k}{\|d_k\|} = 0 \quad (3.1.15)$$

Ya que $\lim_{k \in \mathbf{K}_1, k \rightarrow \infty} \sqrt{f_k - f_{k+1}} = 0$ por ser f_k decreciente y acotada (H1) y por lo tanto converge

En el caso en que $k \in \mathbf{K}_2$, por la regla de Armijo modificada se tiene

$$\begin{aligned} f_{x_k + \alpha_k d_k} - f_k &\leq \sigma \alpha_k [g_k^T d_k + \frac{1}{2} \alpha_k \mu L_k \|d_k\|^2] \\ &\leq \sigma \alpha_k [g_k^T d_k + \frac{1}{2} s_k \mu L_k \|d_k\|^2] \\ &= \sigma \alpha_k [g_k^T d_k + \frac{1}{2} \left(-\frac{g_k^T d_k}{L_k \|d_k\|^2} \right) \mu L_k \|d_k\|^2] \\ &= \sigma \alpha_k (g_k^T d_k - \frac{1}{2} \mu g_k^T d_k) \end{aligned}$$

Así

$$f_{x_k + \alpha_k d_k} - f_k \leq \sigma \alpha_k (g_k^T d_k - \frac{1}{2} \mu g_k^T d_k)$$

de igual forma como en (3.1.15) se prueba

$$\lim_{k \in \mathbf{K}_2, k \rightarrow \infty} (-\alpha_k g_k^T d_k) = 0 \quad (3.1.16)$$

Queremos probar que

$$\lim_{k \in \mathbf{K}_2, k \rightarrow \infty} \left(\frac{g_k^T d_k}{\|d_k\|} \right) = 0$$

Supongamos por absurdo que $\lim_{k \in \mathbf{K}_2, k \rightarrow \infty} \left(\frac{g_k^T d_k}{\|d_k\|} \right) \neq 0$ esto es, existe un $\epsilon_0 > 0$ y un subconjunto infinito $\mathbf{K}_3 \subseteq \mathbf{K}_2$ tal que

$$\frac{-g_k^T d_k}{\|d_k\|} \geq \epsilon_0, \quad \forall k \in \mathbf{K}_3, \quad (3.1.17)$$

De aquí se observa

$$\begin{aligned} -g_k^T d_k &\geq \epsilon_0 \|d_k\| \\ \frac{-\alpha_k g_k^T d_k}{\epsilon_0} &\geq \alpha_k \|d_k\| > 0 \\ \frac{1}{\epsilon_0} \lim_{k \in \mathbf{K}_3, k \rightarrow \infty} -g_k^T d_k &\geq \lim_{k \in \mathbf{K}_3, k \rightarrow \infty} \alpha_k \|d_k\| > \lim_{k \in \mathbf{K}_3, k \rightarrow \infty} 0 \end{aligned}$$

Por teorema del emparedado

$$\lim_{k \in \mathbf{K}_3, k \rightarrow \infty} \alpha_k \|d_k\| = 0 \quad (3.1.18)$$

Por otro lado, por (3.1.5) tenemos

$$[g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k]^T d_k > (\sigma - 1) g_k^T d_k + \frac{\sigma}{2} \mu L_k \frac{\|d_k\|^2}{\beta}$$

con $\theta_k \in [0, 1]$ definido como en la demostración del Teorema 3.1. Por tanto

$$[g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k]^T d_k > (\sigma - 1) g_k^T d_k \quad k \in \mathbf{K}_3 \quad (3.1.19)$$

Por la desigualdad de Cauchy - Schawars y (3.1.19) se tiene

$$\begin{aligned} \|g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k\| &= \|g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k\| \frac{\|d_k\|}{\|d_k\|} \\ &\geq [g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k]^T \frac{d_k}{\|d_k\|} \\ &\geq -(1 - \sigma) g_k^T d_k \frac{d_k}{\|d_k\|} \quad \text{por(3,1,19)} \end{aligned}$$

Por tanto

$$\|g(x_k + \theta_k \frac{\alpha_k}{\beta} d_k) - g_k\| \geq -(1 - \sigma) g_k^T d_k \frac{d_k}{\|d_k\|}$$

Ahora por (H2') como g es uniformemente continua, se tiene como en (3.1.15), se prueba que

$$\lim_{k \in \mathbf{K}_3, k \rightarrow \infty} \left(\frac{g_k^T d_k}{\|d_k\|} \right) = 0$$

Lo que contradice (3.1.17). y Así

$$\lim_{k \in \mathbf{K}_3, k \rightarrow \infty} \left(\frac{g_k^T d_k}{\|d_k\|} \right) = 0 \text{ es lo que queríamos demostrar.} \quad \square$$

§3.2. RADIO DE CONVERGENCIA LINEAL

Con el fin de analizar la velocidad o el tipo de convergencia, asumiremos que la sucesión x_k generada por el nuevo algoritmo converge a un nuevo x^* . Para ello, es necesario asumir la siguiente hipótesis.

(H3) $\nabla^2 f(x^*)$ es una matriz simétrica definida positiva y $f(x)$ es dos veces continuamente diferenciable en una vecindad $N_0(x^*, \epsilon_0)$ de x^*

Lema 4.1. Supongamos que (H3) se cumple. Entonces, existe $\epsilon > 0$ y $0 < m' \leq M'$. tal que (H1) y (H2) se cumplen para $x_0 \in N(x^*, \epsilon) \subseteq N_0(x^*, \epsilon_0)$ y

$$m' \|y\|^2 \leq y^T \nabla^2 f(x) y \leq M' \|y\|^2, \quad \forall x, y \in N(x^*, \epsilon); \quad (3.2.1)$$

$$\frac{1}{2} m' \|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2} M' \|x - x^*\|^2, \quad \forall x \in N(x^*, \epsilon); \quad (3.2.2)$$

$$M' \|x - y\|^2 \geq (g(x) - g(y))^T (x - y) \geq m' \|x - y\|^2, \quad \forall x, y \in N(x^*, \epsilon); \quad (3.2.3)$$

y así

$$M' \|x - x^*\|^2 \geq g(x)^T (x - x^*) \geq m' \|x - x^*\|^2, \quad \forall x \in N(x^*, \epsilon_0). \quad (3.2.4)$$

Por (3.2.3) y (3.2.4) también podemos obtener, a partir de la desigualdad de Cauchy-Schwartz, que

$$M' \|x - x^*\| \geq \|g(x)\| \geq m' \|x - x^*\|, \quad \forall x \in N(x^*, \epsilon), \quad (3.2.5)$$

$$\|g(x) - g(y)\| \leq M' \|x - y\|, \quad \forall x, y \in N(x^*, \epsilon) \quad (3.2.6)$$

Su prueba se puede ver a partir de Refs. [7] - [8] o Refs. [21] - [22]

Lema 4.2. Si (H1) y (H2) se cumplen, la búsqueda dirección d_k satisface la siguiente propiedad en cada iteración

$$\cos \langle -g_k, d_k \rangle = \frac{-g_k^T d_k}{(\|g_k\| \cdot \|d_k\|)} \geq \eta_0, \quad (3.2.7)$$

donde $\eta_0 \in [0, 1]$ es una constante y $\langle g_k, d_k \rangle$ denota el ángulo entre los vectores $-g_k$ y d_k . El nuevo algoritmo genera una secuencia infinita $\{x_k\}$, entonces existe $\eta > 0$ tal que

$$f_k - f_{k+1} \geq \eta \|g_k\|^2, \quad \forall k. \quad (3.2.8)$$

Demostración: por el Teorema 3.1 tenemos lo siguiente

$$\begin{aligned} f_{k+1} - f_k &\leq -\eta'_0 \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 \\ &= -\eta'_0 \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 \frac{\|g_k\|^2}{\|g_k\|^2} \\ &= -\eta'_0 \left(\frac{g_k^T d_k}{\|d_k\| \|g_k\|} \right)^2 \cdot \|g_k\|^2 \end{aligned} \quad (I)$$

Como $\frac{-g_k^T d_k}{(\|g_k\| \cdot \|d_k\|)} \geq \eta_0$ Entonces

$$\begin{aligned} \frac{-g_k^T d_k}{(\|g_k\| \cdot \|d_k\|)} &\leq -\eta_0 \\ \left(\frac{g_k^T d_k}{(\|g_k\| \cdot \|d_k\|)} \right)^2 &\leq (-\eta_0)^2 \\ \frac{g_k^T d_k}{(\|g_k\| \cdot \|d_k\|)} &\leq \eta_0^2 \quad \text{ya que } \eta_0 \in [0, 1] \end{aligned}$$

volviendo a (I) nos queda

$$f_{k+1} - f_k \leq -\eta'_0 \eta_0^2 \cdot \|g_k\|^2$$

Escogiendo $\eta = \eta'_0 \eta_0^2$ se tiene

$$f_{k+1} - f_k \leq -\eta \|g_k\|^2$$

Así

$$f_k - f_{k+1} \leq \eta \|g_k\|^2$$

Teorema 4.1 Si (H3) se cumplen, la búsqueda dirección d_k satisface la siguiente propiedad en cada iteración

$$\cos \langle -g_k, d_k \rangle = \frac{-g_k^T d_k}{(\|g_k\| \cdot \|d_k\|)} \geq \eta_0,$$

donde $\eta_0 \in [0, 1]$ es una constante y $\langle g_k, d_k \rangle$ denota el ángulo entre los vectores $-g_k$ y d_k . El nuevo algoritmo genera una secuencia infinita $\{x_k\}$, y $x_k \in N(x^*, \epsilon)$ para un k suficientemente grande. Entonces $\{x_k\}$ converge a x^* por lo menos linealmente.

Demostración: Sabemos que $\{x_k\}$ converge a x^* , esto es: $\lim_{k \rightarrow \infty} x_k = x^*$ por lema 4.2 se tiene:

$$\begin{aligned} f_k - f_{k+1} &\geq -\eta \|g_k^T\|^2 \\ &\geq \eta m'^2 \|x - x^*\|^2 && \text{por(3,2,5)} \\ &\geq \left(\frac{2\eta m'^2}{M'}\right)(f_k - f^*) && \text{por(3,2,2)} \end{aligned} \quad (I)$$

Sea $\theta = m' \sqrt{\frac{2\eta}{M'}}$, Entonces

$$\begin{aligned} \theta^2 &= \frac{2m'^2\eta}{M'} \\ &\leq \frac{2m'^2\eta_0^2\eta'_0}{M'} && \text{sustituyendo } \eta = \eta_0^2\eta'_0 \\ &\leq \frac{2m'^2\eta'_0}{M'} \\ &\leq \frac{2m'^2\eta'}{M'} && \text{ya que } \eta'_0 = \min(\eta', \eta'') \\ &\leq \left(\frac{\sigma(1 - (1/2)\mu)}{M_0L}\right) \frac{2m'^2}{M'} && \text{por teorema 3,1 } \eta' = \left(\frac{\sigma(1 - (1/2)\mu)}{M_0L}\right) \\ &\leq \left(\frac{2\sigma(1 - (1/2)\mu)}{M_0L}\right) \frac{m'^2}{M'} && (L < M') \\ &\leq \left(\frac{2\sigma(1 - (1/2)\mu)}{M_0L}\right) \frac{M'}{M'} && (m' < M') \\ &\leq \frac{2\sigma}{M_0} < 1 && \text{ya que } \sigma \in (0, 1/2), \mu \in [0, 2) \end{aligned}$$

Por tanto $\theta < 1$

Sea $\omega = \sqrt{1 - \theta^2}$

obviamente $\omega < 1$. regresando a (I) tenemos:

$$\begin{aligned} f_k - f_{k+1} &\geq \theta^2(f_k - f^*) \\ &= \theta^2 f_k - \theta^2 f^* \\ f_{k+1} &\leq f_k - \theta^2 f_k + \theta^2 f^* \\ f_{k+1} - f^* &\leq (1 - \theta^2)f_k - f^* + \theta^2 f^* \\ &= (1 - \theta^2)f_k - (1 - \theta^2)f^* \\ \therefore f_{k+1} - f^* &\leq (1 - \theta^2)(f_k - f^*) \end{aligned}$$

Pero $\omega^2 = \sqrt{1 - \theta^2}$, Entonces:

$$\begin{aligned} f_{k+1} - f^* &\leq \omega^2(f_k - f^*) \\ &\leq \omega^2 \omega^2(f_{k-1} - f^*) \\ &\leq \omega^{2 \times 2} \omega^2(f_{k-2} - f^*) \\ &\leq \omega^{3 \times 2} \omega^2(f_{k-3} - f^*) \\ &\leq \omega^{5 \times 2} \omega^2(f_{k-4} - f^*) \\ &\leq \dots \\ &\leq \omega^{2(k-1)}(f_{k-(k-2)} - f^*) \\ f_{k+1} - f^* &\leq \omega^{2(k-1)}(f_2 - f^*) \\ \therefore f_{k+1} - f^* &\leq \omega^{2k}(f_1 - f^*) \quad (II) \end{aligned}$$

Por lema 4.1

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \left(\frac{2}{m'}\right)(f_{k+1} - f^*) \\ &\leq \omega^{2k} \left[\frac{2(f_1 - f^*)}{m'}\right] \quad \text{por(II)} \end{aligned}$$

Así

$$\|x_{k+1} - x^*\|^2 \leq \omega^k \sqrt{\frac{2(f_1 - f^*)}{m'}} \quad (3.2.9)$$

Por tanto:

$$\begin{aligned} \lim_{k \rightarrow \infty} \|x_k - x^*\|^{1/k} &= \lim_{k \rightarrow \infty} [\omega^{k-1} \sqrt{\frac{2(f_1 - f^*)}{m'}}]^{1/k} \\ &= \omega \lim_{k \rightarrow \infty} [\sqrt{\frac{2(f_1 - f^*)}{m'}}]^{1/k} \\ &= \omega < 1 \end{aligned}$$

Lo que demuestra que $\{x_k\}$ converge por lo menos linealmente a x^* □

CAPÍTULO 4

RESULTADOS NUMÉRICOS

En esta sección, se discute la aplicación del nuevo algoritmo. La técnica de elección de los parámetros es razonable y eficaz para resolver los problemas prácticos de la teoría y los aspectos numéricos.

§4.1. ESTIMACIÓN DE PARÁMETROS L_k

En la nueva de búsqueda lineal (modificación de la regla Armijo), hay un parámetro de L_k que debe ser estimado en cada iteración. Ahora bien, L_k debería aproximar a la constante Lipschitz M' del gradiente $g(x)$ de la función objetivo $f(x)$. Si M' es conocido entonces, deberíamos tomar $L_k = M'$. Sin embargo, M' es desconocida en muchas situaciones por lo tanto, L_k necesita ser estimada.

Definamos para $k = 1, 2, 3, 4, \dots$ a las siguientes expresiones:

$$\delta_{k+1} = x_k - x_{k-1}, \quad y_{k-1} = g_k - g_{k-1}.$$

Se puede utilizar como aproximaciones del parámetro L_k , a las ecuaciones:

$$L_k = \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|} \quad (4.1.1)$$

o

$$L_k = \max\left\{\frac{\|y_{k-i}\|}{\|\delta_{k-i}\|} \quad / i = 1, 2 \dots M\right\} \quad (4.1.2)$$

siempre que $k = M + 1$, donde M es un entero positivo.

A continuación, el método Bazzilai y Borwein. Ref. [16] - [19] nos motiva también a encontrar una manera de estimar M' Aquí, Resolviendo la minimización

$$\min \|L_k \delta_{k-1} - y_{k-1}\|$$

Obtenemos

$$L_k = \frac{\delta_k^T y_{k-1}}{\|\delta_{k-1}\|^2} \quad (4.1.3)$$

Obviamente, si $k \geq M + 1$, también podemos tomar

$$L_k = \max \left\{ \frac{\delta_k^T y_{k-1}}{\|\delta_{k-1}\|^2} \quad / i = 1, 2 \dots M \right\} \quad (4.1.4)$$

Por otro lado, podemos tomar

$$L_k = \frac{\|y_{k-1}\|^2}{\delta_k^T y_{k-1}} \quad (4.1.5)$$

O

$$L_k = \max \left\{ \frac{\|y_{k-i}\|^2}{\delta_k^T y_{k-1}} \quad / i = 1, 2 \dots M \right\} \quad (4.1.6)$$

Siempre que $k \geq M + 1$.

Hay muchas otras técnicas de estimación de la constante de Lipschitz M' ; ver Ref. [8] Vamos a utilizar (4.1.1) - (4.1.6) para estimar M' .

§4.2. RESULTADOS NUMÉRICOS

En lo que sigue, vamos a discutir los resultados numéricos de la nueva búsqueda lineal. Los problemas de prueba se obtuvieron de Ref [20] y el algoritmo utilizado aplicando la nueva búsqueda lineal es el siguiente:

Algoritmo 4.2.1: Algoritmo de descenso con la nueva búsqueda lineal inexacta

Paso 0: Teniendo en cuenta algunos parámetros $\sigma \in (0, \frac{1}{2})$, $\beta \in (0, 1)$, $\mu \in [0, 2)$, L_0 y $x_0 \in \mathbb{R}^n$.

Para $k = 0, 1, \dots$ hasta obtener convergencia.

Paso 1: si $\|g_k\| = 0$, Entonces se detiene, de otra manera vaya al paso 3:

Paso 2: Hacer $d_k = -g_k$ y $\alpha = -\frac{g_k^T d_k}{L_k \|d_k\|^2}$.

Paso 3: Si $f(x_k + \alpha d_k) - f(x_k) \leq \sigma \alpha [g_k^T d_k + \frac{1}{2} \alpha \mu L_k \|d_k\|^2]$, entonces $\alpha_k = \alpha$ ir al paso 5.

Paso 4: Hacer $\alpha = \alpha \beta$ e ir al paso 3.

Paso 5: Hacer $x_{k+1} = x_k + \alpha_k d_k$.

Paso 6: Hacer $\delta_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$, actualizar L_{k+1} utilizando algunas de las ecuaciones dadas por (4.1.1) - (4.1.6).

Paso 7: Hacer $k = k + 1$; ir al Paso 1.

A continuación se muestran los resultados numéricos obtenidos aplicando la nueva búsqueda lineal inexacta los cuales son comparados con la búsqueda lineal de Armijo. Para ello se programaron los algoritmos en MATLAB 7.6.0(R2008a) y se ejecutaron en una portátil con un procesador Pentium(R) Dual–Core i3 CPU T4400 @ 2.20 GHz 2.20 GHz.

Denotaremos por $New(4,1,1)$, $New(4,1,3)$ y $New(4,1,5)$ al método de descenso con la nueva búsqueda lineal inexacta utilizando los valores de L_k (4.1.1),(4.1.3) y (4.1.5) respectivamente.

Los problemas seleccionados para obtener los resultados son: (P14) (función de Wood), (P16) (función Brown and Dennis), (21)(Función de Rosenbrock extendida), (23) (Función de Penalidad) tales funciones de obtienen del artículo [20]. Se utilizo como criterio de parada en todas las corridas, $\|g_k\| \leq 10^{-6}$.

En la Tablas donde se indican los resultados, comprende: en la primera columna el problema (P) utilizado, en la columna 2 la dimensión del problema (n) y en columnas 3-5 se indica el número de iteraciones el número de veces que se evalúa la función y el total de veces que se activo la búsqueda lineal hasta obtener la solución (Backtracking) aplicados a los métodos de Armijo y al método de descenso con la nueva búsqueda lineal inexacta utilizando los valores de L_k (4.1.1),(4.1.3) y (4.1.5) respectivamente. Para obtener las tablas 1,2 y 3 se utilizaron los parametros $\sigma = 0,38$, $\beta = 0,87$ y $L_0 = s = 1$.

Tabla 1 (Iteraciones, Nro de evaluación de la función, Backtracking) $\mu = 0,5$

P	n	Armijo	New(4.1.1)	New(4.1.3)	New(4.1.5)
14	4	6921/26922/24668	283/1237/953	1807/16070/14262	no converge
16	4	235/2443/2207	136/626/489	109/640/530	44/184/139
21	8	6936/26937/23455	2079/3104/1024	1874/4347/2472	no converge
21	100	4991/15128/10136	10815/36567/25751	4060/9250/5189	no converge
21	1000	63036/240845/177808	52206/174444/122237	29710/64283/34572	no converge
23	8	9901/38888/28986	542/3421/2878	68/915/856	85/352/267
23	100	3852/15227/11374	1150/6315/5164	76/829/752	83/345/261
23	1000	1659/6460/4800	458/ 2358/1899	69/536/466	80/348/267

Tabla 2 (Iteraciones, Backtracking, Nro de evaluación de la función,) $\mu = 1,$

P	n	Armijo	New(4.1.1)	New(4.1.3)	New(4.1.5)
14	4	6921/26922/24668	2055/8633/6577	1261/9339/8077	no converge
16	4	235/2443/2207	73/372/298	122/727/604	68/1050/981
21	8	6936/26937/23455	2441/3645/1203	1010/2300/1289	no Converge
21	100	4991/15128/10136	8064/25306/17241	3262/19267/16004	no converge
21	1000	63036/240845/177808	43152/133136/89983	29503/63739/34235	no converge
23	8	9901/38888/28986	542/3421/2878	64/825/760	85/341/255
23	100	3852/15227/11374	160/998/837	80/867/786	83/344/260
23	1000	1659/6460/4800	70/372/301	69/536/466	75/307/231

Tabla 3 (Iteraciones, Backtracking, Nro de evaluación de la función,) $\mu = 1,5$

P	n	Armijo	New(4.1.1)	New(4.1.3)	New(4.1.5)
14	4	6921/26922/24668	1617/6246/4628	1293/10264/8970	no converge
16	4	235/2443/2207	131/2801/2669	116/657/540	59/415/355
21	8	6936/26937/23455	2010/2996/985	1399/3326/1926	no converge
21	100	4991/15128/10136	6874/21045/14170	4050/9146/5095	no converge
21	1000	63036/240845/177808	37201/110578/73376	28946/62656/33709	no converge
23	8	9901/38888/28986	400/2634/2233	64/811/746	80/307/226
23	100	3852/15227/11374	190/1077/886	76/802/725	78/307/228
23	1000	1659/6460/4800	176/976/799	69/536/466	75/290/214

Ajustando los parámetros, $\sigma = 0,04$, $\beta = 0,7$ y $L_0 = s = 1$, obtenemos lo siguiente tabla:

Tabla 4 (Iteraciones, Nro de evaluación de la función, Backtracking)

P	n	μ	New(4.1.1)	New(4.1.3)	New(4.1.5)
14	4	0.5	204/448/243	1339/5016/3676	163/293/129
14	4	1	342/773/430	1295/4816/3520	163/293/129
14	4	1.5	342/773/430	1260/4596/3335	153/261/107
23	100	0.5	192/572/379	no converge	no converge
23	100	1	192/572/379	no converge	no converge
23	100	1.5	192/572/379	no converge	no converge

En estos resultados, se tomo como la dirección de descenso, $d_k = -g_k$ en cada caso.

CONCLUSIONES

De los resultados numéricos obtenidos se observa:

- La nueva búsqueda lineal propuesta realiza menos iteraciones para obtener la solución óptima que la búsqueda lineal de Armijo. Además en la mayoría de los casos el número de backtracking para esta nueva búsqueda, es menor que Armijo, es decir, el tamaño de paso se obtiene más rápido con la nueva búsqueda lineal que con Armijo para obtener la solución.
- De las actualizaciones (4.1.1), (4.1.3) y (4.1.5) para L_k , utilizadas en la obtención de los resultados se observa que con (4.1.1) todos los resultados convergen, donde más fallo fue con (4.1.5) y en los casos donde hubo convergencia, (4.1.3) obtenía la solución en menos iteraciones.
- μ es un parámetro muy importante porque el nuevo método funciona mejor cuando se aumenta en $[0,2)$. Los resultados numéricos de la Tabla 2 y 3 lo muestran.
- En la tabla 4 se aprecia que al ajustar los parámetros se puede mejorar la rapidez del algoritmo para obtener la solución para algunas funciones, mientras que para otras funciones puede que no se obtiene convergencia. Esto significa que el nuevo algoritmo depende del ajuste de los parámetros involucrados para poder obtener la solución.

Un trabajo de investigación a futuro, puede ser considerar más funciones y utilizar dirección Cuasi-Newton y compararlo con otras búsquedas lineales inexactas. Además, analizar que opciones en esta nueva búsqueda lineal inexacta, se puede aplicar para encontrar de hacer mejores ajustes de los parámetros involucrados.

REFERENCIAS

- [1] L. ARMIJO, *Minimization of function having Lipschitz continuous first partial derivatives*, Pacific J. Math., 16 (1966), pp. 1-13.
- [2] J.E. DENNIS, R.B. SCHANBLE, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, New Jersey, (1983)
- [3] J. NOCEDAL, *Theory of Algorithms for Unconstrained Optimization*. Acta Numerica, vol.1 pp. 199-242.(1992)
- [4] M.J.D POWELL, *Direct Search Algorithms for Optimization Calculations*. Acta Numerica, Vol. 7, pp.287-336 (1998)
- [5] P.WOLFE, *Convergence Conditions for Ascent Methods*.Vol.11, pp. 226-235.(1969)
- [6] J.Z SHI, *Convergence of Line-Search Methods for Unconstrained Optimization*. Applied Mathematics and Computation, Vol.157, pp. 393-405. (2004)
- [7] J.Z SHI, *A Supermemory Gradient Method for Unconstrained Optimization Problems* Chinese Journal of Engineering Mathematics, Vol.17, pp. 99-104, (2000).
- [8] M.J TODD, *On Convergence Properties of Algorithms for Unconstrained Minimization*. Journal of Numerical Analysis, Vol.9, pp. 435-441
- [9] SUN, W. Y., HAN J. Y. AND SUN, J., *Global Convergence of Nonmonotone Descent Methods for Unconstrained Optimization Problems* Journal of Computational and Applied Mathematics, Vo.146, pp. 89-98 (2002).
- [10] Z.J. SHI AND J. SHEN, *New Inexact line Search Method for Unconstrained Optimization*. Journal of optimization theory and applications, Vol. 127, pp 425-446. (2005)
- [11] Z.L. WEI AND H. JIANG *Some Convergence Properties of Descent Methods*, Journal of Optimization Theory and Applications, Vol. 95, pp. 177-188. (1997)
- [12] M.N. VRAHATIS, G. S. ANDROULAKIS AND G. MANOUSSAKIS *A New Unconstrained Optimization Method for Imprecise Function and Gradient Values*, Journal of Mathematical Analysis and Applications, Vol. 197, pp. 586-607, (1996).
- [13] L. GRIPPO, F. LAMPARIELLO AND S. LUCIDI *A Class of Nonmonotone Stability Methods in Unconstrained Optimization*, Numerische Mathematik Vol. 62 pp. 779-805, (1991).

-
- [14] Y. H. DAI *On the Nonmonotone Line Search*, *Journal of Optimization Theory and Applications* vol. 112, pp. 315-330, (2002)
- [15] W. Y. SUN, J. Y. HAN, AND J. SUN *Global Convergence of Nonmonotone Descent Methods for Unconstrained Optimization Problems*, *Journal of Computational and Applied Mathematics* Vol. 146, pp. 89-98, (2002)
- [16] J. BARZILAI, AND J. M. BORWEIN *Two-Point Stepsize Gradient Methods*, *IMA Journal of Numerical Analysis* Vol. 8, pp. 141-148, (1988).
- [17] Y. H. DAI, AND L.Z. LIAO *R-Linear Convergence of the Barzilai and Borwein Gradient Method*, *IMA Journal of Numerical Analysis* Vol. 22, pp. 1-10, (2002).
- [18] M. RAYDAN *The Barzilai and Borwein Method for the Large-Scale Unconstrained Minimization Problem*, *SIAM Journal on Optimization* Vol. 7 pp: 26-33, (1997)
- [19] M. RAYDAN *On the Barzilai-Borwein Gradient Choice of Steplength for the Gradient Method*, *IMA Journal of Numerical Analysis* Vol. 13, pp. 321-326. (1993)
- [20] J. MORE *Testing Unconstrained Optimization Software*, *ACM Transactions on Mathematical Software* Vol. 7, pp. 17-41. (1981)
- [21] J. NOCEDAL AND S. J. WRIGHT *Numerical Optimization*, Springer Verlag (1999)
- [22] E. POLAK *Optimization: Algorithms and Consistent Approximations*, Springer Verlag (1997)