

UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas



“SOBRE UNA NUEVA BÚSQUEDA LINEAL TIPO ARMIJO Y
SU RELACIÓN CON EL MÉTODO DE REGION DE
CONFIANZA”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

BR. RANGHELY DEL C. HERNÁNDEZ CASTAÑEDA

COMO REQUISITO FINAL
PARA OBTENER EL TÍTULO DE LICENCIADO
EN CIENCIAS MATEMÁTICAS
ÁREA DE CONOCIMIENTO: OPTIMIZACIÓN.
TUTOR: DR. RÓMULO CASTILLO

*Dedicado A mis Padres Carmen y Rafael,
A mis Hermanos, Raidicar, Wilfredo,
Rodolfo, A mis Abuelos, Lucila, Rafael, A
mi novio Miguelangel.*

AGRADECIMIENTOS

A **DIOS** primeramente por darme la oportunidad de vivir y regalarme una familia maravillosa, por que a pesar de los obstáculos me llenaste de fuerza y voluntad para poder culminar mi carrera.

A mis **padres**, por haberme dado la vida por su apoyo y esfuerzo, por la confianza depositada en mi, por estar a mi lado en todo momento a quienes debo este triunfo, por incentivar-me a seguir adelante y por todos esos sacrificios que hicieron por mi para lograr que sea una profesional.

A mi **hermano**, por sus palabras y consejos que nunca faltaban siempre las tuve presente, por su espíritu luchador y emprendedor eres mi ejemplo a seguir.

A mi **hermanos**, Raidicar, Wilfredo y Rodolfo, por que no sólo son mis hermanos, son mis mejores amigos siempre aconsejándome para bien, ayudándome siempre en todo lo que necesite gracias por nunca faltarme su apoyo tanto financiero como personal y por confiar en mi.

A mis tias (Nelly, Alida), primas (Aliana, Albany) y primos (Keiber, David) por sus consejos, por el apoyo brindado.

A mis amigos (Maria G., Elena, Beomay, Williams, Orangel,), por compartir conmigo buenos y malos momentos, por las enseñanzas que me dejaron cada uno de ellos.

A mis compañeros de clase (Mary Ines, Dayana, Maria E.), por su disposición para ayudarme cuando los necesitaba.

A Miguelangel, por tu apoyo incondicional que me permitió mantenerme constante y no decaer, por motivarme a luchar por lo que quiero.

A mi tutor Rómulo Castillo por ser un gran Mentor y amigo.

Y a todos aquellos que de una u otra forma colaboraron conmigo.

“SOBRE UNA NUEVA BÚSQUEDA LINEAL TIPO ARMIJO Y
SU RELACIÓN CON EL MÉTODO DE REGION DE
CONFIANZA”

RESUMEN

El propósito central de este trabajo es relacionar los Método búsqueda lineal con el Método de Región de Confianza, los cuales son dos importantes clases de técnicas para resolver problemas de optimización irrestrictas y tienen sus ventajas respectivamente. En este trabajo se utiliza la Regla de Armijo la cual es un Método de búsqueda lineal inexacto, y se propone una Nueva Búsqueda Lineal de problemas de optimización sin restricciones. Convergencia global y la velocidad de convergencia del nuevo método se analizan en leves condiciones. Además, la nueva estrategia tipo Armijo de búsqueda lineal se muestra como equivalente a una aproximación de un método de región de confianza, entonces tiene tanto ventajas de la estrategia de búsqueda en línea y la estrategia de región de confianza.

Introducción

Al resolver problemas de optimización irrestricta nos vemos en la necesidad de estudiar principalmente dos tipos de metodologías que generan sucesiones, las cuales se esperan de alguna forma converger a un minimizador de la función objetivo. Por una parte tenemos los métodos basados en búsquedas lineales, en los cuales, una vez obtenida una dirección de avance, realizamos una búsqueda lineal en esa dirección; tales búsquedas pueden ser exactas o inexactas. Entre estas últimas tenemos las búsquedas de Armijo, Wolfe y Goldstein. Por otra parte los métodos de región de confianza constituyen otra metodología para el mismo fin; en este último caso primero se preocupa uno del tamaño del paso y luego de la dirección resultante por medio de una aproximación cuadrática local de la función objetivo en cada iteración. En el presente trabajo realizaremos un estudio general de las búsquedas lineales y el método de región de confianza, pero nos detendremos en el análisis de una particular búsqueda tipo Armijo propuesta recientemente que relaciona el iterado obtenido mediante esta nueva búsqueda con la solución aproximada del subproblema correspondiente al método de región de confianza.

ÍNDICE

Agradecimientos	i
Resumen	ii
1. PRELIMINARES	1
1.0.1. Dirección de Descenso	6
1.0.2. Dirección de Newton	8
1.0.3. Otras Direcciones de Descenso	9
1.0.4. Condición de Suficiente Descenso para d_k	10
1.0.5. MÉTODO DE BÚSQUEDA LINEAL	10
1.0.6. Método de Región de Confianza	15
1.1. Razón de Convergencia	21
Referencias	24
2. SOBRE UNA BÚSQUEDA LINEAL TIPO ARMIJO Y SU RELACION CON EL MÉTODO DE REGIÓN DE CONFIANZA	25
2.1. Un novedoso uso de la regla de Armijo	25
2.2. Convergencia Global	27
2.3. Razón de Convergencia	31
2.3.1. Convergencia Lineal	31
2.4. Convergencia Superlineal	33
2.5. Convergencia Cuadrática	36
2.6. Relación con el Método de Región de Confianza	37

2.7. Conclusión	40
Referencias bibliográficas.	41

Índice de figuras

1.1. Gráfica de la función con corte de planos	4
1.2. Conjunto de nivel	5
1.3. Dirección de Máximo Descenso	7
1.4. La dificultad es que el dirección de máximo descenso es casi ortogonal a la dirección que conduce al mínimo cuando las superficies de costo de f son alargada. Con lo que resulta, que el método vaya en zig-zag sin hacer avance rápido.	7
1.5. Dirección de Descenso	7
1.6. Método de Búsqueda Lineal	11
1.7. Gráfica de $\phi(\alpha)$	12
1.8. Método de Bisección	13
1.9. Regla de Armijo	14
1.10. Regla de Goldstein: lustra el conjunto de tamaños de paso que son aceptables en la regla de Goldstein.	15
1.11. Region de Confianza	17
1.12. Punto de Cauchy.	19
1.13. Punto de Pata de Perro.	20

Capítulo 1

PRELIMINARES

En este capítulo se presentan los fundamentos de la optimización sin restricciones.

En la optimización sin restricciones, se minimiza una función objetivo que depende de variables reales, sin restricción alguna en los valores de estas variables. Consideremos el problema de minimización sin restricción.

$$\min f(x), \quad x \in \mathbb{R}^n. \quad (1.1)$$

Donde \mathbb{R}^n denota un espacio Euclideo n-dimensional y $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función continuamente diferenciable. Para la minimización de dicha función se procede por medio de varios métodos, entre los cuales nos enfocaremos en el Método de Búsqueda Lineal y el Método de Región de Confianza.

Definición 1.0.1. Un punto $x^* \in \Omega$ se dice que es un punto mínimo relativo o punto mínimo local de f sobre Ω si existe un $\varepsilon > 0$ tal que $f(x^*) \leq f(x)$ para todo $x \in \Omega$, dentro de una distancia ε de x^* (Esto es, $x \in \Omega$ y $|x - x^*| < \varepsilon$). Si $f(x^*) < f(x)$ para todo $x \in \Omega$ y $|x - x^*| < \varepsilon$, $x \neq x^*$, entonces x^* es llamado punto mínimo local estricto de f sobre Ω .

Definición 1.0.2. Un punto $x^* \in \Omega$ se dice punto mínimo global de f sobre Ω , si $f(x^*) \leq f(x)$ para todo $x \in \Omega$. Si $f(x^*) < f(x)$ para todo $x \in \Omega$ y $x \neq x^*$, entonces x^* es un punto mínimo global estricto de f sobre Ω .

Definición 1.0.3. (Función Continua)

Sea $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es continua si:

$$\forall \varepsilon > 0, \exists \delta > 0; x \in X, \|x - a\| < \delta \Rightarrow \|f(x) - f(a)\| < \varepsilon \quad (1.2)$$

Donde $\|\cdot\|$ es la norma euclidea.

Definición 1.0.4. (Función Uniformemente Continua)

Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $n \geq 1$ es uniformemente continua en un intervalo $A \subset \mathbb{R}^n$ si para todo $\varepsilon > 0$ existe algún $\delta > 0$ tal que para todo $x, y \in A$ se cumple que si $\|x - y\| < \delta$, entonces $\|f(x) - f(y)\| < \varepsilon$

Teorema 1.0.5. (Teorema del Valor Medio, de Lagrange)

Sea A un abierto en \mathbb{R}^n , sea $f : A \rightarrow \mathbb{R}$ diferenciable sobre A . Si A contiene el segmento de línea con extremos a y $a + h$, entonces existe un punto $c = a + \theta h$, con $\theta \in [0, 1]$, tal que:

$$f(a + h) - f(a) = g(a + \theta h)h \tag{1.3}$$

La prueba se encuentra en cualquier libro de análisis ver

Teorema 1.0.6. (Teorema Fundamental del Cálculo) Sea A un abierto en \mathbb{R}^n , sea $f : A \rightarrow \mathbb{R}$. Si f es continua sobre A , entonces

$$\int_x^a \nabla f = f(x) - f(a) \tag{1.4}$$

Prueba ver

Definición 1.0.7. (Lipschitz continua)

Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, se dice que es Lipschitz continua o simplemente Lipschitziana, si existe una constante $M > 0$, tal que:

$$\|f(x) - f(y)\| \leq M\|x - y\|, \forall x, y \in \mathbb{R}^n. \tag{1.5}$$

Donde M es llamada la constante Lipschitz de la función.

Igualmente es Lipschitz continua en un conjunto $X \subset \mathbb{R}^n$ si existe una constante $M > 0$, tal que:

$$\|f(x) - f(y)\| \leq M\|x - y\|, \forall x, y \in X. \tag{1.6}$$

Si M es la constante Lipschitz de f , entonces $\|\nabla f(x)\| \leq M$ toda vez que el gradiente de la función exista. Contrariamente, si $f : I \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es una función diferenciable con derivada acotada, $\|\nabla f(x)\| \leq L$ para toda x en I , entonces f es Lipschitz continua con constante Lipschitz $K = L$, una consecuencia del teorema del valor medio.

Proposición 1.1 Sea $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ Lipschitziana, entonces es continuamente uniforme

Demostración.

Sea f Lipschitziana, luego existe un $c > 0$, tal que $\|f(x) - f(y)\| \leq c\|x - y\|, \forall x, y \in X$. Así para cualquier $\varepsilon > 0$, basta tomar $\delta = \varepsilon/c$. Por Lipschitziana tenemos que

$$\|x - y\| < \delta \Rightarrow \|f(x) - f(y)\| < \varepsilon$$

Entonces f es continuamente uniforme

Definición 1.0.8. (Función Convexa)

El término "convexo" se puede aplicar tanto a conjuntos como funciones. Un conjunto $S \in \mathbb{R}^n$ es convexo si el segmento de recta que une dos puntos cualesquiera de S se encuentra totalmente dentro de S . Formalmente, para dos puntos cualesquiera $x, y \in S$, tenemos que

$$\alpha x + (1 - \alpha)y \in S, \forall \alpha \in [0, 1]. \tag{1.7}$$

La función f es una función convexa si su dominio S es un conjunto convexo y si para cualquier par de puntos $x, y \in S$, se satisface la siguiente propiedad:

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y), \forall \alpha \in [0, 1]. \tag{1.8}$$

Ejemplos de funciones convexas incluyen la función lineal $f(x) = c^T x + v$, para cualquier vector constante $c \in \mathbb{R}^n$ y escalar v , y la función cuadrática convexa $f(x) = xHx$, donde H es una matriz simétrica semidefinida positiva.

Decimos que f es estrictamente convexa si la desigualdad (1.8) es estricta siempre que $x \neq y$ y α está en el intervalo abierto $(0, 1)$. Una función f se dice que es cóncava si f es convexa.

Si la función objetivo es convexa en todo su dominio, entonces cualquier solución local del problema es, de hecho, una solución global.

Definición 1.0.9. (Función Pseudoconvexa)

Sea $f : S \rightarrow \mathbb{R}$ una función diferenciable sobre S , f es pseudoconvexa si para cada $x_1, x_2 \in S$ se cumple que,

$$\text{Si } f(x_2) < f(x_1), \text{ entonces } \nabla f(x_1)^T (x_2 - x_1) < 0 \tag{1.9}$$

o de manera equivalente,

$$\text{Si } \nabla f(x_1)^T (x_2 - x_1) \geq 0, \text{ entonces } f(x_2) \geq f(x_1) \tag{1.10}$$

Y tenemos como consecuencia que,

$$\text{Si } \nabla f(\bar{x}) = 0, \text{ entonces } \bar{x} \text{ es un minimo global.} \tag{1.11}$$

Ya que si $\nabla f(\bar{x}) = 0$, entonces $\nabla f(\bar{x})^T(x_2 - \bar{x}) \geq 0, \forall x_2 \in S$ se cumple $f(x_2) \geq f(\bar{x}) \forall x_2 \in S$

Definición 1.0.10. (Conjunto de Nivel)

Sea H un conjunto y $f : H \rightarrow \mathbb{R}$. El conjunto de nivel C_k para la función f es el subconjunto de puntos x en H para los cuales $f(x) = k$.

En Simbolo:

$$C_k = \{x \in H : f(x) = k\} \tag{1.12}$$

- Si $H = \mathbb{R}^2$, los conjuntos de nivel son en general curvas y se las llama curvas de nivel.
- Si $H = \mathbb{R}^3$, los conjuntos de nivel suelen ser superficies y se les llama superficies de nivel.

Cuando $H = \mathbb{R}^2$, $f(x) = z$ con $x \in H$, la gráfica de dicha función corresponde al conjunto $gr(f) = \{(x_1, x_2, f(x_1, x_2)) : (x_1, x_2) \in H\}$. Al ubicar los puntos en el espacio \mathbb{R}^3 , obtenemos una superficie en dicho espacio. Las curvas de nivel se obtienen cortando la superficie con planos horizontales situados a distintas alturas. Si cortamos la gráfica con varios planos horizontales obtenemos una serie de curvas situadas sobre la gráfica:

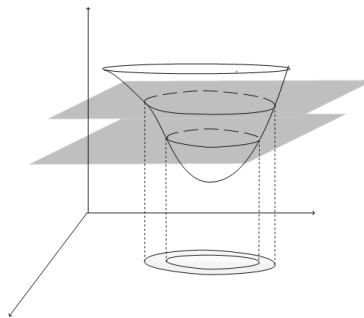


FIGURA 1.1: GRÁFICA DE LA FUNCIÓN CON CORTE DE PLANOS

Y si ahora proyectamos esas curvas sobre el plano xy (lo cual equivale a mirar la gráfica, desde arriba, a vista de pajar) vemos una familia de curvas planas, que son el conjunto de las curvas de nivel de esta gráfica:

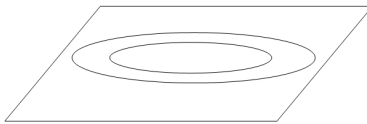


FIGURA 1.2: CONJUNTO DE NIVEL

Conjuntos de nivel y gradientes

Si el conjunto H coincide con \mathbb{R}^n y el campo escalar es de clase C^1 entonces los vectores gradiente del campo escalar son ortogonales a los conjuntos de nivel en el siguiente sentido: Sea C_k un conjunto de nivel y $c : I \subset \mathbb{R} \rightarrow C : k$ una curva diferenciable. Los vectores gradiente del campo f sobre la curva, son ortogonales a los vectores velocidad de la curva. En efecto, para todo t en I , $f(c(t)) = k$. Derivando respecto de t se obtiene (usando la derivada de una composición de funciones), $\nabla f(c(t))c'(t) = 0$

En particular, las curvas integrales asociadas al campo vectorial generado por el gradiente de f son ortogonales a los conjuntos de nivel asociadas a dicha función.

Teorema 1.0.11. (Teorema de Taylor)

Supongamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continuamente diferenciable y $p \in \mathbb{R}^n$, entonces tenemos que:

$$f(x + d) = f(x) + \nabla f(x + td)^T d \tag{1.13}$$

Para algún $t \in (0, 1)$. Además, si f es dos veces continuamente diferenciable, tenemos:

$$\nabla f(x + d) = \nabla f(x) + \int_0^1 \nabla^2 f(x + td) dtd \tag{1.14}$$

y que,

$$f(x + d) = f(x) + d^T \nabla f(x) + \frac{1}{2} d^T \nabla^2 f(x + td) d \tag{1.15}$$

Para algún $t \in (0, 1)$.

Condición necesaria para optimalidad es la derivada para asumir que x^* es un mínimo local y entonces probar los hechos sobre $\nabla f(x^*)$ y $\nabla^2 f(x^*)$

Teorema 1.0.12. (Condición necesaria de primer orden)

Si x^* es un mínimo local de f y f es continuamente diferenciable en un vecindad abierta de x^* , entonces $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es semidefinida positiva.

Teorema 1.0.13. (Condición necesaria de segundo orden)

Si x^* es un mínimo local de f y f es dos veces continuamente diferenciable en un vecindad abierta de x^* , entonces $\nabla f(x^*) = 0$.

Teorema 1.0.14. (Condición suficiente de segundo orden)

Supongamos que $\nabla^2 f(x^*)$ es continua en una vecindad abierta de x^* y que $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es definida positiva, entonces x^* es un mínimo local estricto de f .

§1.0.1. Dirección de Descenso

Para deducir las propiedades de la dirección de descenso, hacemos llamativo al teorema de Taylor (1.0.11), que nos dice que para cualquier dirección d y tamaño de paso α , tenemos que:

$$f(x_k + \alpha d) = f(x_k) + \alpha d^T \nabla f(x_k) + \frac{1}{2} \alpha^2 d^T \nabla^2 f(x_k + td) d \quad (1.16)$$

Para algún $t \in (0, \alpha)$.

Por conveniencia denotaremos $\nabla f(x_k)$ por ∇f_k o en otros casos por g_k , $f(x_k)$ por f_k , $\nabla^2 f(x_k)$ por $\nabla^2 f_k$ o por G_k y $f(x^*)$ por f^* , respectivamente.

El factor reductor de cambio en f a lo largo de la dirección d desde x_k es simplemente el coeficiente de α , a saber, $d^T \nabla f(x_k)$. Por lo tanto, la única dirección d de más rápido decrecimiento es la solución del problema.

$$\min_d d^T \nabla f_k, \text{ sujeto a } \|d\| = 1 \quad (1.17)$$

Donde $d^T \nabla f_k = \|d\| \|\nabla f_k\| \cos \theta = \|\nabla f_k\| \cos \theta$, y θ es el ángulo entre d y ∇f_k , es fácil ver que el reductor se alcanza cuando $\cos \theta = -1$, esto ocurre cuando $\theta = \pi$ y

$$d = \frac{-\nabla f_k}{\|\nabla f_k\|} \quad (1.18)$$

Esta dirección es llamada la dirección de máximo descenso, la figura (1.3) ilustra la dirección ortogonal en el contorno de la función.

La dirección de máximo descenso es usada en método de búsqueda lineal que se mueve a lo largo de $d_k = -\nabla f_k$ en cada paso. Se puede elegir la longitud α_k en una variedad de formas que se discute mas adelante. Una ventaja de la dirección de máximo descenso es que requiere el cálculo de ∇f_k pero no de la segunda derivada. Sin embargo puede ser muy lento en los problemas difíciles.

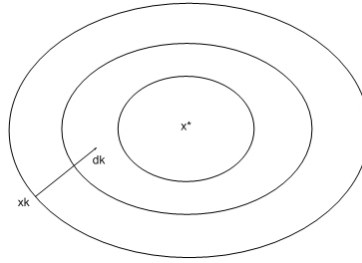


FIGURA 1.3: DIRECCIÓN DE MÁXIMO DESCENSO

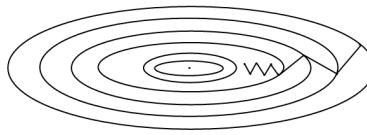


FIGURA 1.4: LA DIFICULTAD ES QUE EL DIRECCIÓN DE MÁXIMO DESCENSO ES CASI ORTOGONAL A LA DIRECCIÓN QUE CONDUCE AL MÍNIMO CUANDO LAS SUPERFICIES DE COSTO DE f SON ALARGADA. CON LO QUE RESULTA, QUE EL MÉTODO VAYA EN ZIG-ZAG SIN HACER AVANCE RÁPIDO.

Métodos de Búsqueda lineal pueden usar una dirección que no sea la dirección de máximo descenso. En general, como hemos visto que el máximo descenso ocurre en $\theta = \pi$ cualquier dirección que forme un ángulo estrictamente mayor que $\frac{\pi}{2}$ con ∇f_k , o análogamente que forme un ángulo estrictamente menor que $\frac{\pi}{2}$ con $-\nabla f_k$ garantiza una disminución en f , siempre que la longitud del paso sea suficiente pequeña, ver figuras (1.5).

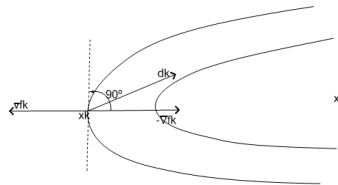


FIGURA 1.5: DIRECCIÓN DE DESCENSO

Podemos probar esta afirmación usando el teorema de Taylor

$$f(x_k + \varepsilon d_k) = f_k + \varepsilon d^T \nabla f_k + O(\varepsilon_k) \tag{1.19}$$

Donde d^T es una dirección de descenso, el ángulo θ_k entre p_k y f_k tiene $\cos \theta_k < 0$,

de modo que

$$(d_k)^T \nabla f_k = \|d_k\| \|\nabla f_k\| \cos \theta_k < 0 \quad (1.20)$$

Si tomamos en cuenta el ángulo θ_k , como el ángulo entre d_k y $-g_k$, podemos verlo de la siguiente forma

$$-(d_k)^T g_k = -\|d_k\| \|g_k\| \cos \theta_k > 0 \quad (1.21)$$

O lo que es lo mismo

$$\cos \langle -g_k, d_k \rangle = -\frac{(g_k)^T d_k}{\|g_k\| \|d_k\|} > 0 \quad (1.22)$$

Y por completitud en \mathbb{R}

$$\cos \langle -g_k, d_k \rangle = -\frac{(g_k)^T d_k}{\|g_k\| \|d_k\|} \geq \tau, \quad (1.23)$$

Donde $0 < \tau \leq 1$.

De esto se deduce que $f(x_k + \varepsilon d_k) < f(x_k)$ para todo valor positivo pero suficiente pequeño de ε .

§1.0.2. Dirección de Newton

Otra importante búsqueda de dirección, tal vez la más importante de todas es la dirección de Newton. Esta dirección se deriva de la serie de Taylor de segundo orden aproximando $f(x_k + d)$, que es

$$f(x_k + d) \approx f_k + d^T \nabla f_k + \frac{1}{2} d^T \nabla^2 f_k d = m_k(d) \quad (1.24)$$

Asumamos por el momento que $G_k = \nabla^2 f(x_k)$ es definida positiva, obtenemos la dirección de Newton por la búsqueda del vector d que minimiza $m_k(d)$. Por la simple igualación de la derivada de la función $m_k(d)$ a 0, obtenemos la siguiente fórmula explícita:

$$g_k + d^T G_k = 0 \Rightarrow (d_k)^N = -(G_k)^{-1} g_k \quad (1.25)$$

La Dirección de Newton es confiable cuando la diferencia entre la función verdadera $f(x_k + d)$ y su modelo cuadrático $m_k(d)$ no es tan grande. Por comparación (1.16) y

(1.24), vemos que la única diferencia entre estas funciones es que la matriz $\nabla^2 f(x_k + d)$ en el tercer termino de la expansión ha sido reemplazado por $\nabla^2 f_k$. Si $\nabla^2 f$ es suficientemente suave, esta diferencia introduce una perturbación de solo $O\|d\|^3$ en la expansión, que por que lo que cuando $\|d\|$ es pequeña la aproximación $f(x_k + d) \approx m_k(d)$ es bastante exacta.

La dirección de Newton puede usarse en métodos de búsqueda lineal cuando $\nabla^2 f_k$ es definida positiva. En este caso sustituyendo (1.25) y multiplicando por $\nabla f(x_k)$ tenemos

$$\nabla(f_k)^T(d_k)^N = -\nabla(f_k)^T(\nabla^2 f_k)^{-1}\nabla f_k \tag{1.26}$$

y

$$-\nabla(f_k)^T(\nabla^2 f_k)^{-1}\nabla f_k \leq -\sigma_k\|\nabla f_k\| \tag{1.27}$$

Para algún $\sigma_k > 0$. A menos que el ∇f_k (y por lo tanto el paso $(d_k)^N$) es igual a cero, tenemos que $\nabla(f_k)^T(d_k)^N < 0$, por lo que la Dirección de Newton es una dirección de descenso.

Cuando $\nabla^2 f_k$ no es definida positiva, la dirección de Newton ni siquiera se puede definir, ya que $(\nabla^2 f_k)^{-1}$ puede no existir. Incluso cuando se puede definir, puede que no cumpla con la propiedad de descenso $\nabla(f_k)^T(d_k)^N < 0$, en cuyo caso no es adecuada como una dirección de búsqueda.

Los métodos que utilizan la dirección de Newton tiene una rápida tasa de convergencia local, por lo general cuadrática. Después de una vecindad de la solución se alcanza, la convergencia con alta precisión a menudo se produce en tan sólo unas pocas iteraciones. El principal inconveniente de la dirección de Newton es el la necesidad de calcular la Hessiana $\nabla^2 f(x)$. El cálculo explícito de esta matriz de segundas derivadas a veces puede ser un proceso engorroso, propenso a errores, y caro.

§1.0.3. Otras Direcciones de Descenso

Además la dirección de descenso a menudo tiene la forma

$$d_k = -(D_k)^{-1}\nabla f_k \tag{1.28}$$

donde D_k es una matriz simétrica y no singular. En el método de la dirección de máximo descenso, D_k es simplemente la matriz identidad I_n , mientras que en el

método de Newton, D_k es la exacta Hessiana $\nabla^2 f_k$. También podemos mencionar los métodos Cuasi-Newton, D_k es una aproximación a la de Hessiana, que se actualiza en cada iteración por medio de una fórmula de bajo rango.

§1.0.4. Condición de Suficiente Descenso para d_k

Para asegurar que algún método no se quede pegado se considera las condiciones técnicas, que pueden ser fácilmente aplicadas a la mayoría de los algoritmos de interés. Para el caso en que $d_k = -(D_k)^{-1}\nabla f_k$, y D_k es definida positiva los autovalores de la matriz simétrica D_k son acotados superiormente y lejos de cero, es decir, para algunos escalares positivos c_1 y c_2 , tenemos

$$c_1\|z\|^2 \leq z^T D_k z \leq c_2\|z\|^2 \quad (1.29)$$

Se puede apreciar entonces que tomando solo la desigualdad izquierda que,

$$|\nabla f(x_k)^T d_k| = |\nabla f(x_k)^T D_k \nabla f(x_k)| \geq c_1 \|\nabla f(x_k)\|^2 \quad (1.30)$$

Luego esto implica que

$$\nabla f(x_k)^T D_k \nabla f(x_k) \geq c_1 \|\nabla f(x_k)\|^2 \text{ o } \nabla f(x_k)^T D_k \nabla f(x_k) \leq -c_1 \|\nabla f(x_k)\|^2 \quad (1.31)$$

En orden de garantizar la convergencia global, nosotros sometemos el requerimiento que d_k satisfasga

$$\nabla f(x_k)^T d_k \leq -c_1 \|\nabla f(x_k)\|^2 \quad (1.32)$$

Para detalles de la prueba ver referencia

§1.0.5. MÉTODO DE BÚSQUEDA LINEAL

Cada iteración de un método de búsqueda lineal, calcula una búsqueda de dirección y luego decide cuan lejos se moverá a lo largo de esa dirección. La iteración esta dada por

$$x_{k+1} = x_k + \alpha_k d_k \quad (1.33)$$

Donde el escalar positivo α_k es llamado el tamaño de paso. El éxito de un método de búsqueda lineal depende de las decisiones eficaces, tanto de la dirección d_k como el tamaño de paso α_k .

La mayoría de los algoritmo de búsqueda lineal requieren que d_k sea una dirección de descenso, esto es que se cumpla $(d_k)^T \nabla f_k < 0$, porque esta característica garantiza que la función f puede reducirse a lo largo esta dirección, como se discutió anteriormente.

Una vez que la dirección de descenso d_k se determina se debe buscar un tamaño de paso a lo largo la dirección de descenso y completa una iteración. Ahora consideremos cuidadosamente la elección del tamaño de paso, parámetro α_k .

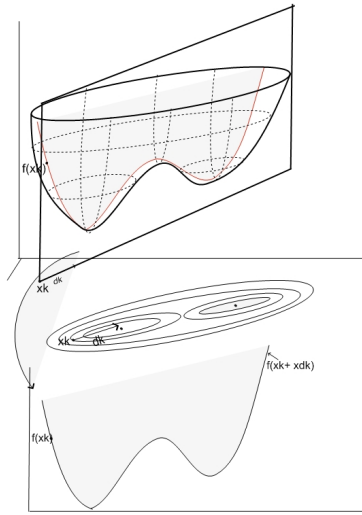


FIGURA 1.6: MÉTODO DE BÚSQUEDA LINEAL

Tamaño de paso

En el tamaño de paso α_k , se encuentra una disyuntiva. Nos gustaría elegir α_k para dar una reducción sustancial de la f , pero al mismo tiempo, no queremos gastar mucho tiempo de costo en hacer la elección. La opción ideal sería el minimizador global de la función univaluada $\phi(\Delta)$ definida por

$$\phi(\alpha) = f(x_k + \alpha d_k), \quad \alpha > 0, \tag{1.34}$$

Cuando la búsqueda del tamaño de paso se realiza resolviendo la minimization de (1.35), entonces se trata de una "Búsqueda lineal exacta". Entra las cuales mencionaremos el Método de Bisección.

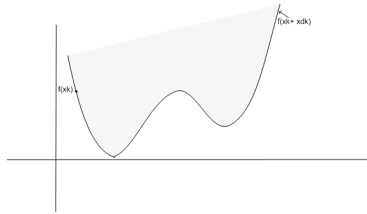


FIGURA 1.7: GRÁFICA DE $\phi(\alpha)$

Método de Bisección

Consideremos una función $\phi(\alpha) = f(x_k + \alpha d_k)$ Pseudoconvexa sobre un intervalo $[a_1, b_1]$. Supongamos que $[a_k, b_k]$ es el intervalo de incertidumbre en una iteración k . Supongamos que $c_k = \frac{b_k - a_k}{2} \in [a_k, b_k]$ y $\phi'(c_k)$ existe. Luego pueden suceder uno de los tres casos,

1. $\phi'(c_k) = 0$, entonces c_k es un mínimo.
2. $\phi'(c_k) > 0$, entonces para $c > c_k$ $\phi'(c_k)(c - c_k) > 0$, entonces $\phi(c) > \phi(c_k)$, luego

$$[a_k + 1, b_k + 1] = [a_k, c_k]$$

3. $\phi'(c_k) < 0$, entonces para $c < c_k$, $\phi'(c_k)(c - c_k) > 0$, entonces $\phi(c) > \phi(c_k)$. Así,

$$[a_k + 1, b_k + 1] = [c_k, b_k]$$

Esto nos da un método que se describe en el siguiente algoritmo

(Algoritmo (1))

Entrada x_0 punto inicial, a_0, b_0 puntos extremos del intervalo de entrada, Tol tolerancia $k=0$; Paso 1: Calcula $c_k = \frac{b_k - a_k}{2}$ Paso 2: Calcular $\phi'(c_k)$

Paso 3: Si $\phi'(c_k) = 0$ o $\frac{b_k - a_k}{2} < \text{Tol}$ Salida ($\alpha c = c_k$) PARAR;

Si no; Paso 4: Si

$\phi'(c_k) > 0$ $[a_k + 1, b_k + 1] = [a_k, c_k]$, y $k = k + 1$; Volver al paso 1 Si no $[a_k + 1, b_k + 1] = [c_k, b_k]$, $k = k + 1$; Volver al paso 1

Para ilustrar el método de Bisección mostramos la siguiente figura.

$$\phi(\alpha) = f(x_k + \alpha d_k), \quad \alpha > 0, \tag{1.35}$$

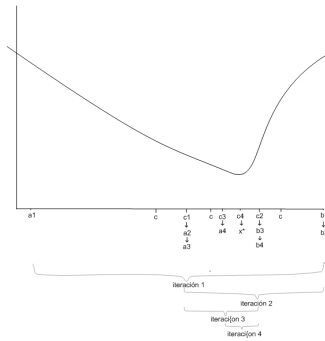


FIGURA 1.8: MÉTODO DE BISECCIÓN

Pero, en general, es demasiado caro identificar el valor mínimo de la función $\phi(\alpha)$. Para encontrar incluso un minimizador local de f con una precisión moderada generalmente requiere demasiadas evaluaciones de la función objetivo f , y posiblemente el gradiente ∇f . Estrategias más prácticas realizan una búsqueda lineal para identificar a un tamaño de paso que logre reducciones que son necesarias en f a un costo mínimo a lo cual se le llama "Búsqueda lineal inexacta", las más usadas son la regla de Armijo, Regla de Goldstein y Regla de Wolfe, .

Regla de Armijo

Un práctico criterio popular para determinar una búsqueda lineal es la regla de Armijo. La idea esencial es que la regla debe primero garantizar que la selección de α no demasiado grande.

En la regla más simple de este tipo un tamaño de paso inicial s se elide, y si el vector correspondiente $x_k + sd_k$ no produce una mejora del valor de f , es decir, $f(x_k + sd_k) \geq f(x_k)$, el tamaño del paso se reduce, quizás varias veces, por un determinado factor, hasta que el valor de f se mejore.

Aquí, escalares fijos $s > 0$, ρ , y μ , con $0 < \rho < 1$, y $0 < \mu < 1$ se eligen, y establecemos $\alpha_k = s\rho^{m_k}$, donde m_k es el primer entero no negativo tal que

$$f(x_k) - f(x_k + s\rho^m d_k) \geq -\mu s \rho^m \nabla f(x_k)^T d_k \tag{1.36}$$

Así el tamaño de paso $\alpha_k = s\rho^m$, con $m = 0, 1, \dots$, se tratan sucesivamente hasta que la desigualdad anterior se cumpla, En otras palabras, se elide α_k el mas grande

α en $\{s, s\rho, s\rho^2, \dots\}$ tal que

$$f(x_k) - f(x_k + \alpha d_k) \geq -\alpha \mu \nabla f(x_k)^T d_k \quad (1.37)$$

Por lo tanto, el tamaño de paso no debe ser sólo positivo, sino que debe ser lo suficientemente pequeño para que se cumpla (1.37). En función de ver esta regla geoméricamente podemos reescribir (1.37) como

$$f(x_k + \alpha d_k) \leq f(x_k) + \alpha \mu \nabla f(x_k)^T d_k \quad (1.38)$$

la figura siguiente ilustra esta regla.

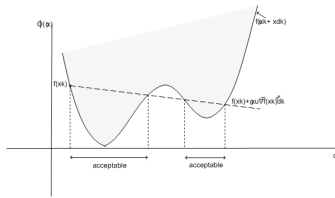


FIGURA 1.9: REGLA DE ARMIJO

La búsqueda lineal por la regla de Armijo. Comenzamos con el estudio del tamaño de paso s y continuar con $s\rho, s\rho^2, \dots$ hasta la primera vez que $s\rho^m$ caiga dentro del conjunto de tamaños de paso que satisface la desigualdad, donde la línea punteada representa la función $f(x_k) + \alpha \mu \nabla f(x_k)^T d_k$ en la Figura (1.10).

Por lo general, μ se elige cerca del cero, por ejemplo, $\mu \in [10^{-5}, 10^{-1}]$. El factor de reducción β se elige generalmente de $1/2$ a $1/10$ en función de la confianza que tenemos del paso inicial s . Siempre podemos tomar $s = 1$ y multiplicar la dirección d_k por un factor de escala.

Regla De Goldstein

Otra regla de búsqueda lineal de precisión que es usada frecuentemente es la regla de Goldstein. Aquí, un escalar fijo $\mu \in (0, 1/2)$ es seleccionado, y α_k se elige para satisfacer

$$\mu \leq \frac{f(x_k + \alpha d_k) - f(x_k)}{\alpha \nabla f(x_k)^T d_k} \leq 1 - \mu \quad (1.39)$$

Reescribiendo esta condición tenemos, que un valor α es aceptable si

$$\mu \alpha \nabla f(x_k)^T d_k + f(x_k) \leq f(x_k + \alpha d_k) \leq (1 - \mu) \alpha \nabla f(x_k)^T d_k + f(x_k) \quad (1.40)$$

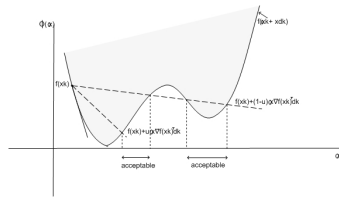


FIGURA 1.10: REGLA DE GOLDSTEIN: LUSTRA EL CONJUNTO DE TAMAÑOS DE PASO QUE SON ACEPTABLES EN LA REGLA DE GOLDSTEIN.

la figura siguiente ilustra esta regla.

Hay algoritmos muy simples para encontrar el tamaño de paso, pero no lo haremos entrar en detalles, ya que en la práctica, la simple regla de Armijo parece ser universalmente preferida. La regla de Goldstein se incluye aquí debido principalmente a de su importancia histórica: fue la primer regla para generar una búsqueda lineal que no se basara en la minimización exacta de f a lo largo de la dirección de descenso, y fue la idea fundamental para posteriormente proponer la regla era Armijo.

Regla de Wolfe

Si derivamos de la función objetivo, sus valores pueden ser evaluados con relativa facilidad, entonces la regla de wolfe, que es una variación de la anterior, es también preferida. En este caso c_1 es seleccionado con $0 < c_1 < 1/2$ y α requiere satisfacer

$$\phi'(\alpha) \geq (1 - c_1)\phi'(0) \tag{1.41}$$

Esto se ilustra

§1.0.6. Método de Región de Confianza

Método de Región de confianza define una región alrededor de la iteración actual en la que se confía que el modelo sea una representación adecuada de la función objetivo, y luego elegir el paso a ser el reductor aproximado del modelo en esta región. Así, el método elige la dirección y la longitud del paso de forma simultánea.

La función del modelo m_k que se utiliza en cada x_k iteración es cuadrática. Además, m_k se basa en el desarrollo de la serie de Taylor de f alrededor de x_k que es

$$f(x_k + p) = f_k + g_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p \tag{1.42}$$

Donde $f_k = f(x_k)$, $g_k = \nabla f(x_k)$ y t es cierto escalar en el intervalo $(0, 1)$. Mediante el uso de una aproximación B_k a la Hessiana $\nabla^2 f(x_k) = G_k$ en el término de segundo orden, m_k se define como sigue:

$$m_k = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad (1.43)$$

Donde B_k es una matriz simétrica. La diferencia entre $f(x+p)$ y m_k es $O(\|p\|^2)$ que es pequeña cuando p es pequeño. Cuando $B_k = G_k$ se dice que el método de Región de Confianza es tipo Newton.

Para obtener cada paso, buscamos una solución del subproblema

$$\min_{p \in R^n} m_k = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad s.t. \|p\| \leq \Delta_k \quad (1.44)$$

Donde $\Delta_k > 0$ es el radio de la región de confianza, el cual determina la longitud del paso de x_k a x_{k+1} . En la mayor parte de nuestras discusiones, se define $\|\cdot\|$ para ser la norma euclídea, de modo que la solución p^* de (1.11) es el reductor de m_k en la bola del radio Δ_k .

Uno de los ingredientes clave en un algoritmo de Región de Confianza es la estrategia para la elección del radio Δ_k en cada iteración p_k se define la relación de reducción

$$r_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)} = \frac{\text{actualreduccion}}{\text{previstareduccion}} \quad (1.45)$$

Si la relación de reducción es grande por ejemplo $\rho_k > 3/4$, el tamaño de la región de confianza aumenta en la siguiente iteración.

Si la relación de reducción es pequeña, por ejemplo $\rho_k < 1/4$, el tamaño de la región de confianza es reducido en la siguiente iteración.

Además, el paso p_k solo se aceptara si le relación de reducción no es demasiado pequeña. Lo que nos lleva al siguiente algoritmo

Algoritmo (2)

Dado $\hat{\Delta} > 0$, $\Delta_o \in (0, \hat{\Delta})$ y $\mu \in [0, 1/4)$ Para $k = 0, 1, 2, \dots$, hasta que x_k es optimal
Obtener la solución aproximada para tentativa del paso p_k

$$\min_{p \in R^n} m_k = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad s.t. \|p\| \leq \Delta_k \quad (1.46)$$

Calcular la relación de reducción

$$r_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$$

Para p_k Actualizar el punto actual

$$x_{k+1} = \begin{cases} x_k + p_k & \text{si } p_k > \mu \\ x_k & \text{Enotro caso} \end{cases}$$

Actualizar el radio de la región de confianza

$$\Delta_{k+1} = \begin{cases} \frac{1}{4}\Delta_k & \text{si } r_k < \frac{1}{4} \\ \min(2\Delta_k, \hat{\Delta}) & \text{si } r_k > \frac{3}{4} \text{ y } \|p_k\| = \Delta_k \\ \Delta_k & \text{Enotro caso.} \end{cases}$$

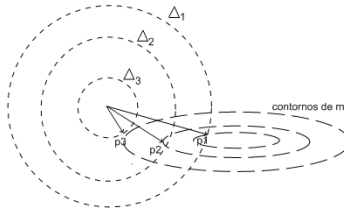


FIGURA 1.11: REGION DE CONFIANZA

A continuación se describen dos estrategias para encontrar soluciones aproximadas del subproblema (1.43)

El Punto de Cauchy

El punto de Cauchy, no es más que el reductor de m_k a lo largo de la dirección de máximo descenso $-g_k$. Sujeto a la confianza. Como hemos visto, los métodos de búsqueda lineal pueden ser convergente, incluso cuando la longitud del paso óptimo no se utiliza en cada iteración. De hecho, la longitud del paso α_k sólo necesita satisfacer criterios. Aunque en principio buscamos la solución óptima del subproblema (1.43), es suficiente para fines de convergencia global encontrar una solución aproximada que se encuentra dentro de la región la confianza y da una reducción suficiente en el modelo. La reducción suficiente puede ser cuantificado en términos del punto de Cauchy, que denotamos por p_k^c y definir en términos del simple procedimiento siguiente.

Algoritmo(3) Encontrar el vector p_k^s que resuelva una version lineal de (1.43), que es,

$$p_k^s = \operatorname{argmin}_{p \in \mathbb{R}^n} f_k + g^T p \quad \text{s.t. } \|p\| \leq \Delta_k \quad (1.47)$$

Calcular $\tau_k > 0$ que minimice $m_k(\tau p_k^s)$ sujeto a satisfacer el conjunto de región de confianza, esto es,

$$\tau_k = \operatorname{argmin}_{\tau > 0} m_k(\tau p_k^s) \quad \text{s.t. } \|\tau p_k^s\| \leq \Delta_k \quad (1.48)$$

Establecer $p_k^c = \tau_k p_k^s$

Es fácil deducir que la solución de (1.47) es simplemente

$$p_k^s = \frac{-\Delta_k \cdot g_k}{\|g_k\|} \quad (1.49)$$

Esto lo podemos comprobar del mismo modo que lo hicimos para deducir la dirección de descenso, sabiendo que el coeficiente de cambio en (1.47) es $g_k^T p$ y ver que la máxima reducción sucede s.t $\|p\| = \Delta_k$.

Para obtener τ_k explícitamente, consideramos el caso que $g_k^T B_k g_k \leq 0$ y $g_k^T B_k g_k > 0$ por separado.

Para el primer caso, la función $m_k(\tau p_k^s)$ decrece monotonamente con τ cuando $g_k \neq 0$, Así $\tau_k = 1$ el valor mas grande que se satisface en la región de confianza.

Para el caso $g_k^T B_k g_k > 0$, $m_k(\tau p_k^s)$ es cuadrada convexa en τ , Así τ_k se obtiene de la minimización no restringida de esta cuadrática, esto es, derivando con respecto a τ a $m_k(\tau p_k^s)$ e igualando a 0, obtenemos:

$$\begin{aligned} p_k^s g_k + \tau (p_k^s)^T B_k p_k^s &= 0 \\ \tau &= \frac{-p_k^s g_k}{(p_k^s)^T B_k p_k^s} \end{aligned}$$

Sustituyendo $p_k^s = \frac{-\Delta_k g_k}{\|g_k\|}$

$$\begin{aligned} \tau &= \frac{\frac{\Delta_k g_k}{\|g_k\|} \cdot g_k}{\frac{-\Delta_k g_k^T}{\|g_k\|} B_k \frac{-\Delta_k g_k}{\|g_k\|}} \\ &= \frac{\frac{\Delta_k}{\|g_k\|} \|g_k\|^2}{\frac{\Delta_k^2 g_k^T B_k g_k}{\|g_k\|^2}} \\ &= \frac{\|g_k\|^3}{\Delta_k g_k^T B_k g_k} \end{aligned}$$

Así que en el caso que $g_k^T B_k g_k > 0$, $\tau = \|g_k\|^3 / (\Delta_k g_k^T B_k g_k)$ o es 1, cualquiera que ocurra primero.

En resumen tenemos,

$$p_k^c = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k$$

Donde

$$\tau_k = \begin{cases} 1 & sig_k^T B_k g_k \leq 0 \\ \min\left(\frac{\|g_k\|^3}{\Delta_k g_k^T B_k g_k}, 1\right) & \text{Enotro caso} \end{cases}$$

La figura.3 ilustra el Punto de Cauchy para en que B_k es definida positiva

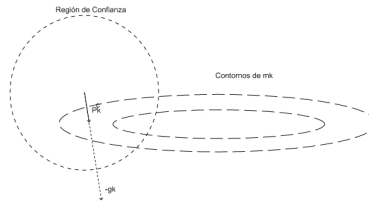


FIGURA 1.12: PUNTO DE CAUCHY.

El Método de Dogled

El enfoque de este método va con el título descriptivo del método de pata de perro. El cual puede ser usado solo cuando B_k es definida positiva. El subproblema de la región de confianza.

$$\min_p m_k = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad s.t. \|p\| \leq \Delta_k$$

Es un problema difícil. La solución no restringida de m_k es $p^B = -B^{-1}g$, cuando este punto es factible para (1.46) es $\min (1.43)$, es evidente una solución por lo que tenemos

$$p^*(\Delta) = p^B, \text{ cuando } \Delta \geq \|p^B\| \tag{1.50}$$

Cuando Δ es pequeño en relación a p^B , la restricción $\|p\| \leq \Delta$ asegura que el término de segundo grado en m tiene poco efecto en la solución "4.5. Así podemos obtener la solución por

$$p^*(\Delta) \approx -\Delta \frac{g}{\|g\|}, \text{ cuando } \Delta \text{ es pequeño} \tag{1.51}$$

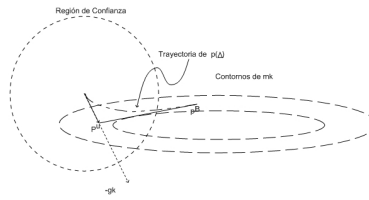


FIGURA 1.13: PUNTO DE PATA DE PERRO.

Para valores intermedios de Δ , la solución $p^*(\Delta)$, típicamente sigue una trayectoria curva como la figura (1.13)

El método de pata de perro aproxima la trayectoria curva de $p^*(\Delta)$, mediante un ruta de acceso que consta de dos líneas.

Formalmente definimos la trayectoria por $\tilde{p}(\tau)$ para $\tau \in [0, 2]$, como:

$$\tau_k = \begin{cases} \tau p & \text{si } 0 \leq \tau \leq 1 \\ p + (\tau - 1)(p^B - p^u) & \text{si } 1 \leq \tau \leq 2 \end{cases}$$

El punto p^u es el de Cauchy, es decir, el minimizador de m a lo largo de la dirección de máximo descenso;

$$p^u = -\frac{g^T g}{g^T B g} g \tag{1.52}$$

El método de pata de perro elige p para minimizar el modelo m a lo largo de este camino, con sujeción a la región confianza. El siguiente lema muestra que el mínimo lo largo de la ruta de acceso pata de perro se pueden encontrar fácilmente.

Lema 1.0.15. Sea B definida positiva. Entonces

- $\|\tilde{p}(\tau)\|$ es una función creciente de τ , y
- $m(\tilde{p}(\tau))$ es una función decreciente de τ .

La demostración de este lema es muy fácil de probar y se encuentra en cualquier libro de optimización ver [?]libro Nocedal). Este lema garantiza que que el camino $\tilde{p}(\tau)$ se cruza con la región de confianza $\|p\| = \Delta$ exactamente en un punto si $\|p^B\| \geq \Delta$, Y en ninguna otra parte. Como m es decreciente a lo largo del camino, el valor

elegido de p estará en p^B si $\|p^B\| \leq \Delta$. En este caso, se calcula el valor apropiado de τ por la solución de la ecuación cuadrática escalar siguiente:

$$\|p^u + (\tau - 1)(p^B - p^u)\| = \Delta^2 \tag{1.53}$$

Consideremos ahora el caso en el que $B = G_k$, cuando G_k es definida positiva, podemos simplemente tomar el conjunto $B = G_k$ (es decir, $p^B = G_k^{-1}gk$) y aplicar el procedimiento anterior para encontrar el paso Newton de pata de perro. El método de Newton-pata de perro es el más apropiado cuando la función objetivo es convexa (es decir, G_k es siempre semidefinida positiva).

Definición 1.0.16. (Notación o , O) Si g es una función de variable real a valores reales, la notación $g(x) = O(x)$, significa que $g(x)$ tiende a cero por lo menos tan rápidamente como lo hace x , esto es, que la hay una $k \geq 0$, tal que:

$$\frac{|g(x)|}{|x|} \leq k, \text{ cuando } x \rightarrow 0 \tag{1.54}$$

La notación $g(x) = o(x)$, significa que $g(x)$ tiende a cero más rápido que lo hace x , es equivalente a que el k anterior es cero.

§1.1. Razón de Convergencia

Considere una sucesión de números reales $\{x_i\}_{i=0}^{\infty}$ convergente a el límite x^* . Definiremos la noción de velocidad de convergencia de la sucesión.

Definición 1.1.1. Sea la sucesión $\{x_i\}_{i=0}^{\infty}$ convergente a x^* . El orden de convergencia de $\{x_i\}$ es definido como el número supremo no negativo p satisfaciendo

$$o \leq \lim_{i \rightarrow \infty} \frac{|x_{i+1} - x^*|}{|x_i - x^*|^p} < \infty \tag{1.55}$$

Si la sucesión tiene orden p el límite

$$\lim_{i \rightarrow \infty} \frac{|x_{i+1} - x^*|}{|x_i - x^*|^p} = \beta \quad \text{Existe} \tag{1.56}$$

Entonces asintóticamente tenemos

$$|x_{i+1} - x^*| = \beta |x_i - x^*|^p$$

Donde β se denomina la tasa de convergencia o radio de convergencia.

Valores mas grandes del orden p , implica en sentido de distancia para el límite x^* , más rápida convergencia. En efecto, si la sucesión tiene orden p , podemos ver la siguiente comparación, como $\{x_i\}_{i=0}^{\infty}$ convergente a x^* , luego $|x_i - x^*| < 1$ para i suficientemente grande, y vemos que $|x_i - x^*|^p$ se hará mas pequeño mientras p sea mas grande, así el siguiente iterado estará a una menor distancia de x^* , sin ignorar a β con lo que tenemos que a menor radio más rápida convergencia.

Es decir, que a valores mas grandes de p y valores mas pequeños de β tendremos aun mayor rapidez de convergencia.

El estudio de la velocidad de convergencia además de estudiar la eficacia relativa de los algoritmos de los métodos es un análisis de aproximación local el cual a tenido considerable éxito en predicción de comportamientos de distintos métodos donde la función de costo puede ser bien aproximada por una cuadrática. Sin embargo, el enfoque de análisis local también tiene algunos inconvenientes importantes, el más importante de las cuales es que no se tiene en cuenta el ritmo de progreso en las primeras iteraciones. No obstante, en muchas situaciones prácticas, no es una omisión grave porque el progreso es rápido en las iteraciones iniciales y con menor crecimiento sólo en el límite (Las razones de esto parecen difíciles de entender, estos son problemas dependientes). Además, a menudo en la práctica, los puntos de partida que están cerca de una solución son fácilmente obtenibles por una combinación de heurística y experiencia, en cuyo caso el análisis local es más significativo.

El análisis local no es muy útil para problemas que involucran ya sea singularidades o mínimos locales, que son difíciles de encontrar en el sentido de que los principales métodos le toma muchas iteraciones para llegar cerca de su solución en la que se aplica el análisis local.

Si los métodos a comparar tienen igual orden de convergencia la comparación se basara en los correspondientes radios de convergencia, con menor valor del radio mayor sera la rapidez de la convergencia.

Definición 1.1.2. sea la sucesión $\{x_k\}$ convergente a x^* , si existe un $\beta \in (0, 1)$ y tal que

$$\lim_{i \rightarrow \infty} \frac{|x_{i+1} - x^*|}{|x_i - x^*|} = \beta \tag{1.57}$$

La sucesión se dice convergente linealmente a x^* con radio de convergencia β .

Una sucesión convergente linealmente con radio de convergencia β , puede decirse tener una convergencia por lo menos tan rápido como la secuencia geométrica $c\beta^k$

para alguna constante c .

En el caso cuando $\beta = 0$ es denominada convergencia superlineal.

Ahora decimos que la sucesión converge con orden p en x^* para $p > 1$, si

$$\lim_{i \rightarrow \infty} \frac{|x_{i+1} - x^*|}{|x_i - x^*|^p} = \beta \quad (1.58)$$

En particular, convergencia con orden:

- 2 se denomina convergencia cuadrática.
- 3 se denomina convergencia cubica.
- etc.

Definición 1.1.3. (Definición ampliada)

El inconveniente de las definiciones anteriores es que éstas no perciben algunas sucesiones que todavía convergen razonablemente rápido, pero cuya "velocidad" es variable. Por lo tanto, la definición de orden de convergencia a veces se extiende como sigue. Según la nueva definición, la sucesión $\{x_k\}$ converge con al menos orden q a x^* , si existe una sucesión $\{\varepsilon_k\}$ tal que

$$\|x_k - x^*\| < \varepsilon_k^q \quad \forall k \quad (1.59)$$

y la sucesión $\{\varepsilon_k\}$ converge a cero con orden p de acuerdo a la definición anterior. Para distinguir esta definición, se le llama convergencia R-lineal, convergencia R-cuadrática, etc (Con el q permanente). Es decir, si $q = 1$ la sucesión converge R-lineal, así sucesivamente.

Definición 1.1.4. Una sucesión $\{x_n\}$, se llama uniformemente acotada, si existe una constante $L > 0$, tal que $\|x_n\| \leq L \forall n$. El número L se llama cota uniforme de $\{x_n\}$.

Definición 1.1.5. (Convergencia Global)

Un método interactivo se dice que converge globalmente cuando la sucesión de puntos producidos por los iterados del método converge a la solución donde el punto inicial es arbitrario.

REFERENCIAS

Capítulo 2

SOBRE UNA BÚSQUEDA LINEAL TIPO ARMIJO Y SU RELACIÓN CON EL MÉTODO DE REGIÓN DE CONFIANZA

En este capítulo presentamos un uso novedoso de la regla de Armijo y desarrollamos un método de búsqueda de línea. En las secciones 2 y 3 se analiza la convergencia global y la tasa de convergencia, respectivamente del nuevo método. En la Sección 4 se pone de manifiesto algunas relaciones entre el nuevo método de búsqueda lineal y el Método de Región Confianza. Conclusiones se dan en la sección 5.

§2.1. Un novedoso uso de la regla de Armijo

Primero asumamos que:

(H1). La función objetivo $f(x)$ es continuamente diferenciable y es acotada inferiormente en \mathbb{R}^n .

(H2). El gradiente $g(x)$ de $f(x)$ es continuo uniformemente en un conjunto convexo abierto B que contiene el conjunto de nivel $L_0 = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$, donde x_0 es dado.

(H2)'. El gradiente $g(x)$ de $f(x)$ es Lipschitz continua en un conjunto convexo abierto

B que contiene el conjunto de nivel L_0 , es decir, existe M' tal que

$$\|g(x) - g(y)\| \leq M' \|x - y\|, \forall x, y \in B.$$

Es evidente que (H2) implica (H2) ver preliminares.

Definimos un nuevo uso de la Regla de Armijo o simplemente lo llamamos una Nueva Búsqueda de Armijo.

Nueva Búsqueda de Armijo Dado $\mu \in (0, \frac{1}{2})$ y $\rho \in (0, 1)$, B_k es una aproximación de $G_k = \nabla^2 f(x_k)$ y \widehat{B}_k es definida de la siguiente procuración: tomar i el entero mas pequeño tal que $d_k^T B_k d_k + i \|d_k\|^2 > 0$. Establecer $s_k = -\frac{g_k^T d_k}{d_k^T \widehat{B}_k d_k}$ y α_k es el mas grande α en $\{s_k, s_k \rho, s_k \rho^2, \dots\}$ tal que

$$f_k - f(x_k + \alpha d_k) \geq -\alpha \mu [g_k^T d_k + \frac{1}{2} \alpha d_k^T B_k d_k]$$

Algoritmo (A)

Paso 0. Elegir $x_0 \in \mathbb{R}^n$ y establecer $k := 0$.

Paso 1. Si $\|g_k\| = 0$ entonces parar; si no ir al paso 2;

Paso 2. Establecer $x_{k+1} = x_k + \alpha_k d_k$, donde d_k es una dirección de descenso de $f(x)$ en x_k y α_k es seleccionado por la Nueva Búsqueda de Armijo;

Paso 3. Establecer $k := k + 1$.

Lema 2.1.1. Si (H1) se tiene y $g_k^T < 0$, entonces la Nueva Búsqueda de Armijo esta bien definida.

Demostración.

Por (H1), tenemos

$$\begin{aligned} \lim_{\alpha \rightarrow 0^+} \left[\frac{f(x_k + \alpha_k d_k) - f_k - \frac{1}{2} \mu \alpha^2 d_k^T B_k d_k}{\alpha} \right] &= \lim_{\alpha \rightarrow 0^+} \left[\frac{f(x_k + \alpha_k d_k) - f_k}{\alpha} \right] - \frac{1}{2} \lim_{\alpha \rightarrow 0^+} [\mu \alpha d_k^T B_k d_k] \\ &= g_k^T d_k - 0, \\ &< \mu g_k^T d_k. \end{aligned}$$

Por lo tanto, existe un $\bar{\alpha}_k < 0$, tal que

$$\frac{f(x_k + \alpha_k d_k) - f_k - \frac{1}{2} \mu \alpha^2 d_k^T B_k d_k}{\alpha} \leq \mu g_k^T d_k, \forall \alpha \in [0, \bar{\alpha}_k].$$

Luego,

$$f(x_k + \alpha_k d_k) - f_k \leq \alpha \mu g_k^T d_k + \frac{1}{2} \mu \alpha^2 d_k^T B_k d_k, \forall \alpha \in [0, \bar{\alpha}_k].$$

Entonces

$$f_k - f(x_k + \alpha_k d_k) \geq -\alpha \mu [g_k^T d_k + \frac{1}{2} \alpha d_k^T B_k d_k], \forall \alpha \in [0, \bar{\alpha}_k].$$

Lo que implica que la Nueva Búsqueda de Armijo esta bien definida.

§2.2. Convergencia Global

Teorema 2.2.1. Si (H1) y (H2) se cumplen, d_k satisface (??)Y α_k es definida por la Nueva Búsqueda de Armijo. Algoritmo (A) genera una secuencia infinita $\{x_k\}$, con una secuencia acotada $\{B_k\}$, esto es, hay un β tal que $\|B_k\| \leq \beta, \forall k$. Entonces

$$\lim_{k \rightarrow \infty} \left(-\frac{g_k^T d_k}{\|d_k\|} \right) = 0 \quad (2.1)$$

Demostración.

Por contradicción, Supongamos que existe un subconjunto infinito $K \subseteq \{0, 1, 2, 3, \dots\}$ y un $\varepsilon > 0$, tal que

$$-\frac{g_k^T d_k}{\|d_k\|} \geq \varepsilon, k \in K \quad (2.2)$$

Entonces

$$-g_k^T d_k \geq \varepsilon \|d_k\|, k \in K \quad (2.3)$$

Por la Nueva Búsqueda de Armijo, en el caso de $d_k^T B_k d_k \leq 0$ ($k \in K$), tenemos:

$$\begin{aligned} f_k - f_{k+1} &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2} \alpha_k d_k^T B_k d_k] \\ &\geq -\alpha_k \mu g_k^T d_k \quad (\text{Ya que } d_k^T B_k d_k \leq 0) \\ &\geq \alpha_k \mu \varepsilon \|d_k\| \quad (\text{Por (2.3)}); \end{aligned}$$

Y en el caso de $d_k^T B_k d_k > 0$ ($k \in K$), tenemos:

$$\begin{aligned} f_k - f_{k+1} &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2} \alpha_k d_k^T B_k d_k] \\ &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2} s_k d_k^T B_k d_k] \quad (\text{Desde que } \alpha_k \leq s_k) \\ &= -\frac{1}{2} \alpha_k \mu g_k^T d_k \quad (\text{Sustituyendo } s_k \text{ y restando}) \\ &\geq \frac{\mu \varepsilon}{2} \alpha_k \|d_k\|, \forall k \in K \quad (\text{Por (2.3)}). \end{aligned}$$

De aquí vemos que en ambos casos llegamos a que $f_k - f_{k+1} \geq 0$, ya que $\alpha_k, \mu, \varepsilon$, son escalares positivos, así $f_k \geq f_{k+1}, \forall k \in K$, luego la función es monótona decreciente y acotada inferiormente por (H1), con lo que tenemos que

$$\lim_{k \rightarrow \infty} f_k - f_{k+1} = 0, \forall k \in K.$$

Esto, implica que

$$\alpha_k \|d_k\| \leq f_k - f_{k+1} \rightarrow 0, (k \in K, k \rightarrow \infty) \quad (2.4)$$

En el paper se encuentra un pequeño error debido a que hacen la siguiente comparación de un vector con un escalar; $\alpha_k d_k \leq \alpha_k \|d_k\|$

También por la Nueva Búsqueda de Armijo, $\|B_k\| \leq \beta$ implica que

$$\|\widehat{B}_k\| \leq 2\beta + 1, \forall k. \quad (2.5)$$

Sea $K_1 = \{k \in K \mid \alpha_k = s_k\}$, $K_2 = \{k \in K : \alpha_k < s_k\}$,

Podemos probar que K_1 es un subconjunto finito. En efecto, si K_1 es un subconjunto infinito, vemos que:

$$\alpha_k \|d_k\| = -\frac{g_k^T d_k}{d_k^T \widehat{B}_k d_k} \|d_k\| \quad (\text{Sustituyendo } \alpha_k) \quad (2.6)$$

Por (2.5) tenemos

$$d_k^T \widehat{B}_k d_k \leq (2\beta + 1) \|d_k\|^2 \Rightarrow \frac{1}{d_k^T \widehat{B}_k d_k} \geq \frac{1}{(2\beta + 1) \|d_k\|^2}$$

Así en ((2.6)), tendríamos

$$-\frac{g_k^T d_k \|d_k\|}{(2\beta + 1) \|d_k\|^2} \leq -\frac{g_k^T d_k}{d_k^T \widehat{B}_k d_k} \|d_k\| = \alpha_k \|d_k\| \rightarrow 0 (k \in K_1, k \rightarrow \infty), \quad (\text{Por (2.5)})$$

Que contradice (2.2). En consecuencia K_2 debe ser un subconjunto infinito y $\alpha_k/\rho \leq s_k$, $\forall k \in K_2$. Además, tenemos que α_k es el mas grande α tal que la desigualdad de la Nueva Búsqueda de Armijo se cumple, con lo que $\alpha = \alpha_k/\rho$ hará que la desigualdad en la Nueva Búsqueda de Armijo falle, es decir,

$$f_k - f(x_k + (\alpha/\rho)d_k) < -(\alpha_k/\rho)\mu[g_k^T d_k + \frac{1}{2}d_k^T B_k d_k]$$

Por lo tanto

$$\begin{aligned} f(x_k + (\alpha/\rho)d_k) - f_k &\geq (\alpha_k/\rho)\mu[g_k^T d_k + \frac{1}{2}(\alpha_k/\rho)d_k^T B_k d_k] \\ &\geq (\alpha_k/\rho)\mu[g_k^T d_k - \frac{1}{2}(\alpha_k/\rho)d_k^T \widehat{B}_k d_k] \\ &\geq (\alpha_k/\rho)\mu[g_k^T d_k - \frac{1}{2}s_k d_k^T \widehat{B}_k d_k] \quad (\alpha_k/\rho \leq s_k) \\ &= \frac{3}{2}(\alpha_k/\rho)\mu g_k^T d_k, k \in K_2. \quad (\text{Sustituyendo } s_k = -\frac{g_k^T d_k}{d_k^T \widehat{B}_k d_k}) \end{aligned}$$

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

$$f(x_k + (\alpha/\rho)d_k) - f_k = (\alpha_k/\rho)g(x_k + \theta_k(\alpha_k/\rho)d_k)^T d_k,$$

Luego

$$(\alpha_k/\rho)g(x_k + \theta_k(\alpha_k/\rho)d_k)^T d_k \geq \frac{3}{2}(\alpha_k/\rho)\mu g_k^T d_k, k \in K_2.$$

De aquí,

$$g(x_k + \theta_k(\alpha_k/\rho)d_k)^T d_k \geq \frac{3}{2}\mu g_k^T d_k, k \in K_2. \quad (2.7)$$

Restando a ambos lados $\frac{g_k^T d_k}{\|d_k\|}$ y acomodando tenemos:

$$\begin{aligned} -(1 - \frac{3}{2}\mu) \frac{g_k^T d_k}{\|d_k\|} &\leq \frac{[g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k]^T d_k}{\|d_k\|} \\ &\leq \frac{\|g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k\| \|d_k\|}{\|d_k\|} \quad (\text{Por la desigualdad de Cauchy - Schwartz}) \\ &= \|g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k\| \rightarrow 0 (k \in K_2, k \rightarrow \infty) \quad (\text{Por (2.4) y por (H2)}) \end{aligned}$$

Lo que contradice (2.2) y por tanto, la conclusión es válida.

Teorema 2.2.2. Si (H1) y (H2) se cumplen, d_k satisface (??) y α_k es definida por la Nueva Búsqueda de Armijo. Algoritmo (A) genera una secuencia infinita $\{x_k\}$ y B_k es uniformemente acotada, esto es, existe un $\beta > 0$ tal que $\|B_k\| \leq \beta, \forall k$. Entonces

$$\sum_{k=0}^{\infty} \left(-\frac{g_k^T d_k}{\|d_k\|} \right)^2 < \infty \quad (2.8)$$

Demostración.

Dado que (H2) implica (H2) ver en preliminares, la conclusion del Teorema (2.2.1) se cumple.

Sea

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

Por la Nueva Búsqueda de Armijo, en el caso de que $k \in K_1$ tenemos:

$$\begin{aligned} f_k - f_{k+1} &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2} \alpha_k d_k^T B_k d_k] \\ &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2} \alpha_k d_k^T \widehat{B}_k d_k] \quad (d_k^T \widehat{B}_k d_k \geq d_k^T B_k d_k) \\ &= -\frac{1}{2} \alpha_k \mu g_k^T d_k \quad (\text{Evaluando } \alpha_k = s_k \text{ en el segundo termino}) \\ &= \frac{\mu}{2} \cdot \frac{(g_k^T d_k)^2}{d_k^T \widehat{B}_k d_k} \quad (\text{Sustituyendo } \alpha_k \text{ por } s_k) \\ &\geq \frac{\mu}{2(2\beta+1)} \left(-\frac{g_k^T d_k}{\|d_k\|} \right)^2, k \in K_1. \quad (\text{Por (2.5)}) \end{aligned}$$

Así

$$f_k - f_{k+1} \geq \frac{\mu}{2(2\beta+1)} \left(-\frac{g_k^T d_k}{\|d_k\|} \right)^2, k \in K_1. \quad (2.9)$$

En el caso de $k \in K_2$ tenemos que $\alpha_k \leq s_k$, podemos probar similar a (2.7) que

$$g(x_k + \theta_k(\alpha_k/\rho)d_k)^T d_k \geq \frac{3}{2}\mu g_k^T d_k, k \in K_2, \quad (2.10)$$

Y por la desigualdad de Cauchy Schwartz, (2.5), (2.4) y restando a ambos lados por $\frac{g_k^T d_k}{\|d_k\|}$, obtuvimos:

$$-(1 - \frac{3}{2}\mu) \frac{g_k^T d_k}{\|d_k\|} \leq \|g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k\|, k \in K_2$$

Por lo que (H2) implica que existe un M tal que:

$$\begin{aligned} -(1 - \frac{3}{2}\mu) \frac{g_k^T d_k}{\|d_k\|} &\leq \|g(x_k + \theta_k(\alpha_k/\rho)d_k) - g_k\| \\ &\leq M' \alpha_k / \rho \|d_k\|, \quad (M' = \theta_k M) (k \in K_2) \end{aligned}$$

Y que despejando a α_k , obtenemos:

$$\alpha_k \geq \frac{-\rho(1 - \frac{3}{2}\mu) g_k^T d_k}{M' \|d_k\|^2}, k \in K_2. \quad (2.11)$$

Por consiguiente,

$$\begin{aligned} f_k * f_{k+1} &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2} \alpha_k d_k^T B_k d_k] \\ &\geq -\alpha_k \mu [g_k^T d_k + \frac{1}{2} s_k d_k^T \widehat{B}_k d_k] \quad (\alpha_k d_k^T B_k d_k \leq s_k d_k^T \widehat{B}_k d_k) \\ &= -\alpha_k \mu [g_k^T d_k - \frac{1}{2} g_k^T d_k] \quad (Sustituyendo s_k) \\ &= -\frac{1}{2} \alpha_k \mu g_k^T d_k \\ &\geq \frac{\mu \rho (1 - \frac{3}{2}\mu)}{2M'} \left(-\frac{g_k^T d_k}{\|d_k\|}\right)^2, k \in K_2. \quad (Por (2.10)) \end{aligned}$$

Por la inecuación (2.9) y tomando

$$\eta' = \frac{\mu}{2} \min\left(\frac{1}{2\beta + 1}, \frac{\rho(1 - \frac{3}{2}\mu)}{M'}\right),$$

Se tiene que:

$$f_f - f_{k+1} \geq \eta' \left(-\frac{g_k^T d_k}{\|d_k\|}\right)^2, \forall k. \quad (2.12)$$

Por (H1) f es continuamente diferenciable y acotada por debajo y además la función es decreciente, así

$$\eta' \sum_{k=0}^{\infty} \left(-\frac{g_k^T d_k}{\|d_k\|}\right)^2 \leq \sum_{k=0}^{\infty} \infty (f_k - f > k + 1) < +\infty.$$

Esta desigualdad implica que (2.8) se cumple.

Corolario 2.2.3. Si (H1) y (H2) se cumplen, d_k satiface (??) y α_k es definida por la Nueva Búsqueda de Armijo. Algoritmo (A) genera una secuencia infinita $\{x_k\}$ y $\|B_k\| \leq \beta, \forall k$. Entonces

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad (2.13)$$

Demostración.

Por Teorema (2.2.1), se tiene:

$$\begin{aligned} \tau \|g_k\| &\leq -\frac{g_k^T d_k}{\|g_k\| \|d_k\|} \\ &= -\frac{g_k^T d_k}{\|d_k\|} \rightarrow 0, (k \rightarrow \infty). \end{aligned}$$

La prueba termina.

§2.3. Razón de Convergencia

En orden de analizar la razón de convergencia, además asumimos que:

(H3). $x_k \rightarrow x^*$ cuando $k \rightarrow \infty$, $\nabla^2 f(x^*) \neq 0$ y $f(x)$ es dos veces continuamente diferenciable en $N(x^*, \varepsilon_0) = \{x \mid \|x - x^*\| < \varepsilon_0\}$.

Lema 2.3.1. Asumamos que (H3) se cumple. Entonces existe $0 < m' \leq M'$ y $\varepsilon \leq \varepsilon_0$ tal que:

$$m' \|y\|^2 \leq y^T \nabla^2 f(x) y \leq M' \|y\|^2, \forall x, y \in N(x^*, \varepsilon); \tag{2.14}$$

$$\frac{1}{2} m' \|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2} M' \|x - x^*\|^2, \forall x \in N(x^*, \varepsilon); \tag{2.15}$$

$$M' \|x - y\|^2 \geq (g(x) - g(y))^T (x - y) \geq m' \|x - y\|^2, \forall x, y \in N(x^*, \varepsilon); \tag{2.16}$$

Y así,

$$M' \|x - x^*\|^2 \geq g(x)^T (x - x^*) \geq m' \|x - x^*\|^2, \forall x \in N(x^*, \varepsilon); \tag{2.17}$$

Por (2.17) y (2.16) podemos obtener, de la desigualdad de Cauchy Schwartz, que

$$M' \|x - x^*\|^2 \geq \|g(x)\| \geq m' \|x - x^*\|^2, \forall x \in N(x^*, \varepsilon); \tag{2.18}$$

Y

$$\|g(x) - g(y)\| \leq M' \|x - y\|^2, \forall x, y \in N(x^*, \varepsilon); \tag{2.19}$$

La prueba se puede encontrar en la literatura de

§2.3.1. Convergencia Lineal

Teorema 2.3.2. Asumir que (H3) se cumple, d_k satisface (??) y α_k es definida por la Nueva Búsqueda de Armijo y que $\|B_k\| \leq \beta, \forall k$. Si el Algoritmo (A) genera un secuencia infinita $\{x_k\}$, entonces $\{x_k\}$ converge a x^* por lo menos R-lineal.

Demostración.

Si (H3) se cumple entonces existe un k' tal que x_k in $N(x^*, \varepsilon_0)$, $\forall k \geq k'$ y (H1) y (H2)'se cumplen si x_0 in $N(x^*, \varepsilon_0)$. Por Teorema (2.2.2) y (??) tenemos:

$$\begin{aligned} f - f_{k+1} &\geq \eta' \left(-\frac{g_k^T d_k}{\|d_k\|} \right)^2 \\ &\geq \eta' \tau^2 \|g_k\|, \quad k \geq k' \quad (\text{Por}(??)) \end{aligned}$$

Por la desigualdad anterior y lema ??, Haciendo $\eta = \eta' \tau^2$, se obtiene

$$\begin{aligned} f - f_{k+1} &\geq \eta \|g_k\| \\ &\geq \eta m'^2 \|x_k - x^*\|^2 \quad (\text{Por}(2.18)) \\ &\geq \frac{2\eta m'^2}{M'} (f_k - f^*). \quad (\text{Por}(2.15)) \end{aligned}$$

Sea

$$\theta = m' \sqrt{\frac{2\eta}{M'}} > 0$$

Probemos que $\theta < 1$. En efecto, por la definición de η y η' en la prueba del Teorema (2.2.2), se obtiene:

$$\begin{aligned} \theta^2 &= \frac{2m'^2\eta}{M'} = \frac{2m'^2\tau^2\eta'}{M'} \\ &\leq \frac{2m'^2\tau^2}{M'} \frac{\mu\rho(1-\frac{3}{2}\mu)}{2M'} \quad (\text{Por } \eta') \\ &\leq \tau^2\mu\rho(1-\frac{3}{2}\mu). \quad (0 < m' \leq M') \\ &\leq \mu\rho(1-\frac{3}{2}\mu). \quad (0 < \tau \leq 1) \\ &< 1. \quad (0 < \mu \leq 1/2)(0 < \rho < 1) \end{aligned}$$

Definamos

$$0 < w = \sqrt{1 - \theta^2} < 1$$

Retomando la desigualdad de arriba que

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

Luego,

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

Esto y por (2.15) de el Lema (2.3.1) se tiene

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \frac{2}{m'}(f_{k+1} - f^*) \\ &\leq w^{2(k-k')} \frac{2(f_{k'+1} - f^*)}{m'} \end{aligned}$$

Así

$$\|x_{k+1} - x^*\| \leq w^{k-k'} \sqrt{\frac{2(f_{k'+1} - f^*)}{m'}}$$

Es decir, que si cambiamos el subíndice k por $k - 1$, tenemos

$$\begin{aligned} \|x_k - x^*\| &\leq w^{k-1-k'} \sqrt{\frac{2(f_{k'+1} - f^*)}{m'}} \\ &= \frac{w^k}{w^{k'+1}} \sqrt{\frac{2(f_{k'+1} - f^*)}{m'}} \\ &= w^k \sqrt{\frac{2(f_{k'+1} - f^*)}{m'w^{2(k'+1)}}}. \end{aligned}$$

Notemos que $\sqrt{\frac{2(f_{k'+1} - f^*)}{m'w^{2(k'+1)}}$ es una constante, y aplicando la raíz k -ésima y limite a ambos lados, tenemos finalmente que:

$$R_1\{x_k\} = \lim_{k \rightarrow \infty} \|x_k - x^*\|^{1/k} \leq \lim_{k \rightarrow \infty} w \left(\frac{2(f_{k'+1} - f^*)}{m'w^{2(k'+1)}} \right)^{1/2k} = w < 1$$

Que muestra que $\{x_k\}$ converge a x^* al menos R-Lineal.

§2.4. Convergencia Superlineal

Supongamos además que,

(H4). $\{BB_k\}$ es una secuencia de matrices definida positiva y $\|B_k\| \leq \beta, \forall k$. Algoritmo (A) con $d_k = -B_k^{-1}g_k$ satisface la siguiente condición

$$\lim_{k \rightarrow \infty} \frac{\|[B_k - \nabla^2 f(x^*)]d_k\|}{\|d_k\|} = 0 \tag{2.20}$$

Lema 2.4.1. Si (H3) y (H4) se cumplen. Algoritmo (A) genera una secuencia infinita $\{x_k\}$. Entonces existe k' tal que

$$\alpha_k = 1, \forall k \geq k'. \tag{2.21}$$

Demostración.

Por corolario (2.2.3) y (H4)

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \Rightarrow \lim_{k \rightarrow \infty} \|-B_k^{-1}g_k\| = 0$$

Esto y por (H3), implica que

$$\lim_{k \rightarrow \infty} x_k = x^*, \quad \lim_{k \rightarrow \infty} \|d_k\| = 0, \quad (2.22)$$

Y así

$$\lim_{k \rightarrow \infty} (x_k + td_k - x^*) = 0, \quad (2.23)$$

Donde $t \in [0, 1]$. Además, (H4) implica que

$$\begin{aligned} 0 &\leq \lim_{k \rightarrow \infty} \frac{\|d_k^T [B_k - \nabla^2 f(x^*)] d_k\|}{\|d_k\|^2} \\ &\leq \lim_{k \rightarrow \infty} \frac{\|d_k\| \| [B_k - \nabla^2 f(x^*)] d_k\|}{\|d_k\|^2} \\ &= \lim_{k \rightarrow \infty} \frac{\| [B_k - \nabla^2 f(x^*)] d_k\|}{\|d_k\|^2} \end{aligned}$$

Luego

$$d_k^T [B_k - \nabla^2 f(x^*)] d_k = o(\|d_k\|^2). \quad (2.24)$$

En el paper presenta un detalle al declarar que es el teorema del valor medio, pero la validez de la siguiente igualdad es por el teorema fundamental de calculo:

$$f(x_k + d_k) - f_k = g_k^T d_k + \int_0^1 (1-t) d_k^T \nabla^2 f(x_k + td_k) d_k dt$$

En efecto,

$$\begin{aligned} g_k^T d_k + \int_0^1 (1-t) d_k^T \nabla^2 f(x_k + td_k) d_k dt &= g_k^T d_k + \int_0^1 d_k^T \nabla^2 f(x_k + td_k) d_k dt \\ &\quad - \int_0^1 t d_k^T \nabla^2 f(x_k + td_k) d_k dt \\ &= g_k^T d_k + \nabla f(x_k + d_k)^T d_k - g_k^T d_k \\ &\quad - \int_0^1 t d_k^T \nabla^2 f(x_k + td_k) d_k dt \\ &= \nabla f(x_k + d_k)^T d_k - \int_0^1 t d_k^T \nabla^2 f(x_k + td_k) d_k dt, \end{aligned}$$

Integrando por parte, con $u = t$ y $dv = d_k^T \nabla^2 f(x_k + td_k) d_k dt$,

$$\begin{aligned} g_k^T d_k + \int_0^1 (1-t) d_k^T \nabla^2 f(x_k + td_k) d_k dt &= \nabla f(x_k + d_k)^T d_k - \nabla f(x_k + d_k)^T d_k \\ &\quad + \int_0^1 \nabla f(x_k + td_k) d_k dt \\ &= \int_0^1 \nabla f(x_k + td_k) d_k dt \\ &= f(x_k + d_k) - f_k \end{aligned}$$

si para un k suficiente grande, tenemos:

$$\begin{aligned} f(x_k + d_k) - f_k &= g_k^T d_k + \int_0^1 (1-t) d_k^T \nabla^2 f(x_k + td_k) d_k dt \\ &= [g_k^T d_k + \frac{1}{2} d_k^T B_k d_k] + \int_0^1 (1-t) d_k^T [\nabla^2 f(x_k + td_k) - \nabla^2 f(x^*)] d_k dt \\ &\quad + \frac{1}{2} d_k^T [\nabla^2 f(x^*) - B_k] d_k \\ &= [g_k^T d_k + \frac{1}{2} d_k^T B_k d_k] + o(\|d_k\|^2). \quad (\text{Por (2.23) y (2.24)}) \end{aligned}$$

Donde $d_k = -B_k^{-1}g_k$, sustituimos en la desigualdad derecha anterior vemos que,

$$\begin{aligned} g_k^T d_k + \frac{1}{2}d_k^T B_k d_k &= g_k^T d_k - \frac{1}{2}g_k^T d_k. \quad (\text{Porque } g_k^T d_k \in \mathbb{R}). \\ &= \frac{1}{2}g_k^T d_k \\ &< 0. \quad (\text{Por}(??)) \end{aligned}$$

Por esto y (2.22), se tiene en la igualdad de arriba que:

$$f(x_k + d_k) - f_k = [g_k^T d_k + \frac{1}{2}d_k^T B_k d_k] + o(\|d_k\|^2) \leq \mu[g_k^T d_k + \frac{1}{2}d_k^T B_k d_k]$$

Lo que implica que existe un k' para que el (2.21) es valida.

Teorema 2.4.2. Si (H3) y (H4) se cumplen. Algoritmo (A) genera una sucesión infinita $\{x_k\}$. Entonces $\{x_k\}$ converge a x^* superlinealmente.

Demostración.

Por Corolario (2.2.3) y Lema (2.3.1) sabemos que $\{x_k\} \rightarrow x^*$. Por Lema (2.4.1), existe k' tal que (2.21) se cumple y tenemos:

$$x_{k+1} = x_k + d_k, k \geq k' \tag{2.25}$$

Donde $d_k = -B_k^{-1}g_k$. Por el teorema fundamental del calculo se sigue que:

$$\begin{aligned} g_{k+1} - g_k &= \int_0^1 \nabla^2 f(x_k + t(x_{k+1} - x_k))(x_{k+1} - x_k) dt \\ &= \int_0^1 \nabla^2 f(x_k + td_k) d_k dt. \quad (\text{Por (2.25)}) \\ &= \nabla^2 f(x^*) d_k + \int_0^1 [\nabla^2 f(x_k + td_k) - \nabla^2 f(x^*)] d_k dt \end{aligned}$$

Veamos que:

$$\lim_{k \rightarrow \infty} \frac{\|\nabla^2 f(x_k + td_k) - \nabla^2 f(x^*)\| \|d_k\|}{\|d_k\|} \leq \lim_{k \rightarrow \infty} \frac{\|\nabla^2 f(x_k + td_k) - \nabla^2 f(x^*)\| \|d_k\|}{\|d_k\|} = 0$$

Asi $\int_0^1 [\nabla^2 f(x_k + td_k) - \nabla^2 f(x^*)] d_k dt = o(\|d_k\|)$ y en la igualdad de arriba obtendríamos:

$$\begin{aligned} g_{k+1} &= g_k + \nabla^2 f(x^*) d_k + o(\|d_k\|) \\ &= -B_k d_k + \nabla^2 f(x^*) d_k + o(\|d_k\|) \\ &= -[B_k - \nabla^2 f(x^*)] d_k + o(\|d_k\|) \\ &= o(\|d_k\|). \quad (\text{Por (2.20)}) \end{aligned}$$

Entonces

$$\lim_{k \rightarrow \infty} \frac{\|g_{k+1}\|}{\|d_k\|} = 0 \tag{2.26}$$

De (2.18) y (2.26) se deduce que:

$$\begin{aligned}
 \frac{\|g_{k+1}\|}{\|d_k\|} &\geq \frac{m'\|x_{k+1}-x^*\|}{\|d_k\|} \\
 &= \frac{m'\|x_{k+1}-x^*\|}{\|x_{k+1}-x_k\|} \quad (\text{Por (2.25)}) \\
 &\geq \frac{m'\|x_{k+1}x^*\|}{\|x_{k+1}-x^*\|+\|x_k-x^*\|} \\
 &= m' \frac{\|x_k-x^*\|}{1+\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}}
 \end{aligned}$$

Y así

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|} = 0$$

Lo que implica que $\{x_k\}$ converge a x^* Superlinealmente.

§2.5. Convergencia Cuadrática

Si tomamos $B_k = \nabla^2 f(x_k)$ en el Algoritmo (A), entonces (H4) se cumple. y tenemos el siguiente resultado.

Teorema 2.5.1. Si (H3) se cumple, $B_k = \nabla^2 f(x_k)$ para k suficiente grande. Algoritmo (A) genera una sucesión infinita $\{x_k\}$. Entonces $\{x_k\}$ converge a x^* al menos Superlinealmente.

Demostración.

En este caso, (H4) se cumple automáticamente, en consecuencia el resultado del Teorema (2.4.2) se cumple.

Teorema 2.5.2. Si (H3) se cumple, $B_k = \nabla^2 f(x_k)$ para k suficiente grande. Además, existe una epsilon vecindad $N(x^*, \varepsilon) = \{x \in \mathbb{R}^n \mid \|x - x^*\| < \varepsilon\}$ de x^* con $\varepsilon < \varepsilon_0$ tal que $\nabla^2 f(x)$ es Lipschitz continua en $N(x^*, \varepsilon)$, es decir, existe $L(\varepsilon)$ tal que

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L(\varepsilon)\|x - y\|, \quad \forall x, y \in N(x^*, \varepsilon) \quad (2.27)$$

Algoritmo (A) genera una secuencia infinita $\{x_k\}$. Entonces $\{x_k\}$ converge a x^* Cuadráticamente.

Demostración.

Por Corolario (2.2.3), Lema (2.3.1) y (2.4.1), se sigue que $\{x_k\}$ converge a x^* y existe k' tal que para todo $k \geq k'$, $x_k \in N(x^*, \varepsilon)$, $B_k = \nabla^2 f(x_k)$, y $\alpha_k = 1$. Sea $\varepsilon_k = x_k - x^*$. Por teorema fundamental del calculo tenemos:

$$\begin{aligned}
 \varepsilon_{k+1} &= x_{k+1} - x^* \\
 &= x_k - x^* + d_k \quad (\text{Por (2.25)}) \\
 &= \varepsilon_k - \nabla^2 f(x_k)^{-1} g_k \\
 &= \varepsilon_k - \nabla^2 f(x_k)^{-1} (g_k - g^*) \\
 &= \varepsilon_k - \nabla^2 f(x_k)^{-1} \int_0^1 \nabla^2 f(x^* + t\varepsilon_k) \varepsilon_k dt \\
 &= \nabla^2 f(x_k)^{-1} [\nabla^2 f(x_k) - \int_0^1 \nabla^2 f(x^* + t\varepsilon_k) \varepsilon_k dt] \\
 &= \nabla^2 f(x_k)^{-1} \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + t\varepsilon_k)] \varepsilon_k dt
 \end{aligned}$$

Esto y (2.27) implica que:

$$\begin{aligned}
 \|\varepsilon_{k+1}\| &= \|\nabla^2 f(x_k)^{-1} \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + t\varepsilon_k)] \varepsilon_k dt\| \\
 &\leq \|\nabla^2 f(x_k)^{-1}\| \int_0^1 \|\nabla^2 f(x_k) - \nabla^2 f(x^* + t\varepsilon_k)\| dt \|\varepsilon_k\| \\
 &\leq |\nabla^2 f(x_k)^{-1}| L(\varepsilon) \int_0^1 \|x_k + t\varepsilon_k - x_k\| dt \|\varepsilon_k\| \\
 &= |\nabla^2 f(x_k)^{-1}| L(\varepsilon) \|\varepsilon_k\|^2 \int_0^1 t dt \\
 &= \frac{1}{2} \|\nabla^2 f(x_k)^{-1}\| L(\varepsilon) \|\varepsilon_k\|^2.
 \end{aligned}$$

Por lo tanto,

$$\lim_{k \rightarrow \infty} \frac{\|\varepsilon_{k+1}\|}{\|\varepsilon_k\|}$$

Con lo que $\{x_k\}$ converge a x^* Cuadráticamente.

§2.6. Relación con el Método de Región de Confianza

La relación entre el método de línea nueva búsqueda y el método de región de confianza se dará a conocer en esta sección.

Ya vimos en el Capítulo 1 El Método de Región de Confianza, el cual se basa en buscar una solución al subproblema

$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p, \quad s.t. \|p\| \leq \Delta_k, \quad (2.28)$$

Dónde Δ_k es de un radio de la Región de Confianza. Definimos $\|\cdot\|$ por la norma euclidiana, de modo que la solución p_k^* de (2.28) es el reductor de $m_k(p)$ en la bola de radio Δ_k . Por lo tanto, el Método de Región de Confianza nos obliga a resolver una secuencia del subproblemas (2.28) en el que la función objetivo y restricciones (Que se puede escribir como $p^T p \leq \Delta_k^2$) son ambos cuadráticos.

El primer problema que surge en la definición del Método de Región de Confianza es la estrategia para la elección del radio de la región la confianza Δ_k en cada

iteración. Con base a esta elección en la relación entre el modelo m_k y la función objetivo f en la anterior iteraciones, se define la relación de reducción

$$r_k = \frac{f_k - f(x_k + p_k)}{m_k(0) - m_k(p_k)}, \quad (2.29)$$

Y la influencia de esta relación de reducción en la elección de dicho radio Δ_k la vimos en el **Algoritmo(2)** del Capítulo 1.

La convergencia global del Método de Región de Confianza depende de las reducciones obtenidas en la solución del modelo cuadrático m_k . Pero, no requiere resolver el subproblema (2.28) del Método de la Región de confianza con exactitud, podemos encontrar un p_k que satisfaga:

$$m_k(0) - m_k(p_k) \geq c_1 \|g_k\| \min(\Delta_k, \frac{\|g_k\|}{\|B_k\|}), \quad (2.30)$$

Y

$$\|p_k\| \leq \gamma \Delta_k, \quad (2.31)$$

Para $\gamma \geq 1$ y $c_1 \in (0, 1]$. Lo cual garantiza la convergencia del método.

En efecto, la solución exacta P_k^* de (2.28) satisface (2.30) y (2.31) ([3]). Basta ver que $m_k(0) - m_k(p_k^*) \geq m_k(0) - m_k(p_k), \forall p_k$ y $\forall k$.

Lema 2.6.1. Sea $\mu = 0$ en el Algoritmo(2). Supongamos que $\|B_k\| \leq \beta$ para alguna constante β , que f es Lipschitz continua diferenciable y acotada inferiormente en el conjunto de nivel

$$L_0 = \{x \in \mathbb{R} | f(x) \leq f(x_0)\},$$

Y que la aproximación de la solución de (2.28) satisface (2.30) y (2.31) para constantes positivas c_1 y γ . Entonces se tiene

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad (2.32)$$

Lema 2.6.2. Sea $\mu \in (0, \frac{1}{4})$ en el **Algoritmo(2)**. Supongamos que $\|B_k\| \leq \beta$ para alguna constante β , que f es Lipschitz continua diferenciable y acotada inferiormente en el conjunto de nivel L_0 , Y que la aproximación de la solución de (2.28) satisface (2.30) y (2.31) para constantes positivas c_1 y γ . Entonces se tiene

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad (2.33)$$

En la Nueva Búsqueda de Armijo, si el conjunto $p = \alpha_k d_k$ entonces

$$f_k - f(x_k + p_k) \geq \mu[-g_k^T p_k - \frac{1}{2} p_k^T B_k p_k]$$

Y $x_{k+1} = x_k + p_k$ es solo una aceptable reducción en el Método Región de Confianza y sabemos que $m_k(0) - m_k(p_k) = f_k - f_k - g_k^T p_k - \frac{1}{2} p_k^T B_k p_k = -g_k^T p_k - \frac{1}{2} p_k^T B_k p_k > 0$, tenemos

$$r_k = \frac{f_k - f_{k+1}}{m_k(0) - m_k(p_k)} \geq \mu, \tag{2.34}$$

Es obvio que (2.34) coincide con la condición de aceptación en (??).

Este r_k es justamente la relación de la reducción actual y la reducción predicha. Si la relación anterior se cumple entonces el nuevo punto $\bar{x}_k = x_k + p_k$ es aceptado tanto por la Nuevo Método de Búsqueda lineal como por el Método de Región de Confianza. De lo contrario, debemos ajustar el tamaño de paso en el Nuevo Método de Búsqueda lineal o ajustar el radio de la región de confianza en el Método de Región de Confianza. De hecho, si el nuevo punto $x_k = x_k + p_k$ es rechazado, debemos reducir el tamaño del paso en el Nuevo Método de Búsqueda lineal o reducir el radio de la región de confianza en el Método de Región de Confianza. Desde este punto de vista, el Método de Región de Confianza y el Método de Búsqueda Lineal se pueden unificar en una forma general.

§2.7. Conclusión

En esta Investigación se utiliza La Búsqueda Lineal Tipo Armijo en una forma novedosa y proponen un Nuevo Método de Búsqueda Lineal para problemas de optimización sin restricciones. La convergencia global y razón de convergencia del Nuevo Método se analizan en leves condiciones. Además, cada iteración generada por La Nueva Búsqueda Lineal tipo Armijo se demuestra que es una solución aproximada del subproblema de correspondiente al Método de Región de Confianza, lo que revela la relación entre el Método Búsqueda lineal y el Método de Región de Confianza, en cierto sentido. Para ponerlo en detalle, si tomamos $p_k = \alpha_k d_k$ en el propuesto Método de Búsqueda Lineal entonces, se tiene la condición de aceptación

$$r_k = -\frac{f_k - f_{k+1}}{g_k^T d_k + \frac{1}{2} p_k^T B_k p_k} \geq \mu$$

Que es exactamente la condición de avance en el Método de Región de Confianza.

Por otra parte, cabe destacar que el paper presenta algunos errores, que fueron arreglados en la desarrollo de la Investigación, y que se mencionan a continuación:

- La ecuación (2.4), era originalmente $\alpha_k d_k \leq \alpha_k \|d_k\|$, pero esta fue cambiada debido a que no cabe comparar un vector con un número real.
- En el Lema (2.6.1), falta la premisa de que la función f sea Lipschitz Continua, ya que es una hipótesis fundamental para la demostración del Lema (ver [1]). Además, la definición de L_0 (2.6.1) en el mismo Lema, $f(x_0)$ en el paper es $f(x_1)$ lo cual no concuerda con la definición de conjunto de nivel.
- En el Lema (2.6.2) el paper menciona el Algoritmo (2.1), el cual además de no encontrarse en el paper es una referencia equivocada, realmente se quería referir al Algoritmo (2).

REFERENCIAS BIBLIOGRÁFICAS.

- [1] P.-A. Absil, C. G. Baker, and K. A. Gallivan, Trust-region methods on Riemannian manifold with applications in numerical linear algebra, Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), Leuven, Belgium, 5-9 July 2004, 2004.
- [2] Zhen-Jun Shi y Xiang-Sun Zhang. From Line Search Method to Trust Region Method. Operations Research and Its Applications, Lecture Notes in Operations Research, vol.5, pp.156-170, World Publishing Corporation, 2005.
- [3] J. Nocedal and J. S. Wright, Numerical Optimization, Springer-Verlag New York, Inc.(1999).
- [4] Bertsekas D. Nonlinear Programming, Athena Scientific, 2000.