

UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas



“SOBRE EL MÉTODO DEL GRADIENTE ESPECTRAL
PROYECTADO NO-MONÓTONO EN CONJUNTOS
CONVEXOS.”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

JESÚS FREÍTEZ

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



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



ACTA
 TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designado 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:

“SOBRE EL MÉTODO DEL GRADIENTE ESPECTRAL PROYECTADO NO-MONÓTONO EN CONJUNTOS CONVEXOS.”

Presentado por el ciudadano JESÚS FREÍTEZ titular de la Cédula de Identidad N° 18.135.907. Con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciado 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

 PRINCIPAL

 FIRMA

 PRINCIPAL

 FIRMA

OBSERVACIONES:

¹ Aprobado ó Reprobado

*Dedico este trabajo a DIOS
TODOPODEROSO y a mis PADRES
quienes con su amor y su apoyo me han
ayudado a cumplir todas mis metas.*

AGRADECIMIENTOS

Agradezco primeramente a **DIOS TODOPODEROSO** quien me dio la energía positiva e iluminó mi camino con perseverancia, sabiduría y fortaleza.

A mis PADRES Miriam Rebeca Falcón de Freítez y Evodio Rafael Freítez pilares fundamentales de mi educación quienes con su amor, afecto, atención y apoyo incondicional llenan mi vida de seguridad para lograr los éxitos. A ustedes mil gracias!!!... Esto se lo debo a ustedes!!!... LOS AMO!!!...

A mi hermana Daniela que con su inocencia, picardía y travesuras alivia mi estrés, para ti este humilde ejemplo.

A mi novia Angelica Rodriguez persona muy especial para mi, quien ha sido mi amiga y mi compañera por mucho tiempo, y me ha brindado motivación y ayuda. GRACIAS!!!...

A mi inmensa familia que siempre me han motivado con su alegría a seguir adelante. En especial a mi abuela Maria Honoria Flores, mi tía Gloria Falcón y mi tío Otilio Falcón que siempre estuvieron pendiente de mis estudios.

A el Dr. Rómulo Castillo, tutor de esta investigación por su empeño, experiencia y dedicación en mi formación ética. Y a otros profesores como Javier Hernández, Eibar Hernández, Hugo Lara, entre otros que de una u otra forma me ayudaron a la realización de este trabajo.

A mis compañeros de clases y amigos (Moisés, Willennys, Alexander, Luis, Liseth, entre otros.) por los momentos vividos, los conocimientos adquiridos y las ayudas prestadas. Muchas Gracias!!!...

“SOBRE EL MÉTODO DEL GRADIENTE ESPECTRAL
PROYECTADO NO-MONÓTONO EN CONJUNTOS
CONVEXOS.”

RESUMEN

En el presente trabajo se analiza principalmente el artículo científico *Nonmonotone Spectral Projected Gradient Methods on Convex Sets* cuyos autores son Ernesto G. Birgin, José Mario Martínez y Marcos Raydan (ver [3]). Para lo cual se han consultado las referencias mostradas al final del mismo. El trabajo ha consistido en lograr una buena comprensión de los métodos y teoremas allí presentados, como novedad podría considerarse la interpretación geométrica de dicho método, presente en la monografía que corroboran la comprensión del mismo.

En el artículo principal mencionado describen las técnicas del método del gradiente proyectado espectral no monótono considerado para la minimización de funciones diferenciables sobre conjuntos convexos. El diseño del gradiente proyectado clásico es extendido para incluir una estrategia de tamaño de paso que es basada sobre la búsqueda lineal no monótona desarrollada por Grippo, Lampariello y Lucidi (ver [7]). En la referencia principal (ver [3]) la estrategia no monótona es combinada con la elección del tamaño de paso espectral introducido por Barzilai y Borwein (ver [1]) y analizado por Raydan (ver [12]), para acelerar la convergencia del proceso. En adición a la ruta no lineal del gradiente proyectado clásico. La comprensión y el análisis de los algoritmos, los teoremas de convergencia y un resumen de los resultados de la aplicación de los algoritmos obtenidos en el artículo [3] serán presentados en la siguiente monografía.

ÍNDICE

Agradecimientos	i
Resumen	ii
Introducción	1
1. Preliminares	5
1.1. Nociones Básicas	5
1.1.1. Interpolación Cuadrática	6
1.2. Proyección Sobre Conjuntos Convexos	7
1.3. Regla de Armijo	11
1.4. Algunos Métodos de Optimización	12
1.4.1. Método de Máximo Descenso	12
1.4.2. Método de Newton	14
1.4.3. Método Quasi-Newton	16
1.4.4. Método de la Secante	17
1.4.5. Método del Gradiente Projectado	18
1.5. Regla de Armijo a lo largo del arco de proyección	19
1.6. El Método Gradiente con tamaño de paso de dos puntos	21
1.7. Sobre la elección del tamaño de paso de Barzilai y Borwein para el método gradiente	24
1.8. Una técnica de Búsqueda lineal no monótona para el método de Newton	25
2. Método del Gradiente Projectado Espectral	29
2.1. Algoritmos	30
2.2. Teoremas de Convergencia	36
2.3. Análisis de Resultados Numéricos	50
3. Conclusión	53

Referencias

55

INTRODUCCIÓN

Estudiaremos el problema de programación no lineal dado por.

$$\begin{aligned} &\text{Minimizar} && f(x) \\ &\text{sujeto a} && x \in \Omega, \end{aligned} \tag{1}$$

donde $\Omega \subset \mathbb{R}^n$ es un conjunto convexo y cerrado y $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continua diferenciable. El objetivo principal de la siguiente monografía es el de analizar el método del gradiente proyectado espectral (ver [3]) para resolver el problema (1) y lograr una comprensión general de la metodología allí involucrada. Para ello realizamos un estudio preliminar de diferentes aspectos teóricos que están presentes en dicho método y que requieren de una comprensión básica inicial para profundizar posteriormente en dicha metodología. Estas consideraciones son abordadas con detalles en el capítulo 2.

No se pretende obtener resultados originales puesto que escapa a los objetivos del trabajo, pero se aborda con detalles los contenidos presentes para lograr un buen entendimiento de los mismos.

Para resolver el problema (1) existen diferentes metodologías entre ellas el método del gradiente proyectado el cual está basado en el método de proyecciones, para una comprensión del mismo (ver [2]). Por otra parte las búsquedas unidimensionales como por ejemplo la regla de Armijo (ver [2]) han sido consideradas en diferentes métodos de minimización irrestricta, en la mayoría de las veces a lo largo de direcciones de descenso obteniendo métodos en donde la disminución de los valores de la función objetivo es monótona. En [7] se presenta un algoritmo basado en una búsqueda lineal no monótona que involucra el conocimiento de los valores objetivos de iteraciones previas. Un caso particular de una búsqueda lineal que contiene un tamaño de paso específico es dado en [12], el cual está relacionado con la condición secante de los métodos Quasi-Newton (ver [5]) y con la definición del cociente de Rayleigh (ver [10]).

El método del gradiente proyectado espectral no monótono cuyo análisis representa la base principal de esta monografía combina en el método del gradiente proyectado la estrategia de búsqueda lineal no monótona antes citada (ver [7]) con el tamaño de paso espectral mostrado en [1] y [12]. Dos variantes de este método y su completo desarrollo serán analizados más profundamente en el capítulo 3, no obstante, a manera de introducción seguiremos con algunas consideraciones generales importantes.

Este método es una estrategia no monótona de implementación eficiente en cuanto a tiempos de ejecución y es una de las muchas variaciones del método del gradiente proyectado en los últimos años. El método tiene la propiedad común de mantener la factibilidad de los iterados por la frecuencia de los pasos de prueba proyectados sobre el conjunto convexo factible. El proceso de proyección es la parte más costosa del método del gradiente proyectado. En algunos casos donde la proyección no es costosa, como en el caso de las restricciones de caja, la velocidad de convergencia del método del gradiente proyectado es muy lenta y en el caso en que las curvas de nivel de f parecen largos valles, la sucesión de iteraciones generadas por el método de máximo descenso expone una trayectoria de zig-zag y el método es considerablemente lento. Esta situación motiva a combinar el método del gradiente proyectado con dos nuevos proyectos desarrollados en la optimización ya mencionados. Esta combinación requiere menos trabajo computacional y ayuda a acelerar notablemente la velocidad de convergencia, y por lo tanto se asegura la convergencia global del método.

Barzilai y Borwein descubren que para alguna elección del tamaño de paso α_k , el método de Cauchy converge superlinealmente a la solución, si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función cuadrática convexa publicado en 1986 (ver [1]), luego en 1991 Raydan prueba la convergencia del método de Barzilai y Borwein para funciones cuadráticas estrictamente convexas arbitrarias (ver [12]). Raydan muestra que este método es más efectivo que el método de máximo descenso y el tamaño de paso viene dado por.

$$\alpha_k = \frac{s_{k-1}y_{k-1}}{s_{k-1}s_{k-1}}.$$

En el método de Barzilai y Borwein se estudia el método de máximo descenso, donde la elección del tamaño de paso a lo largo del gradiente negativo es derivado

de una aproximación de dos puntos de la ecuación secante fundamental del método Quasi Newton (ver [5]). La idea principal de la elección del tamaño de paso espectral es acelerar la convergencia del método de máximo descenso.

Una estrategia de globalización que acepte el presente tamaño de paso desarrollado por Barzilai y Borwein con la mayor frecuencia posible y conserve las propiedades locales del método es basado sobre una regla de búsqueda lineal no monótona para el gradiente proyectado desarrollada por Grippo, Lampariello y Lucidi en 1985 (ver [7]) y ésta, a la vez está basada en una condición tipo Armijo a lo largo del arco de proyección (ver [2]) como sigue.

$$f(x_k + \sigma^h a d_k) \leq \max_{0 \leq j \leq m(k)} \{f(x_{k-j})\} + \gamma \sigma^h a g_k^T d_k,$$

donde $m(0) = 0$, $0 \leq m(k) \leq \min\{m(k-1) + 1, M\}$, $k \geq 1$, M es un entero no negativo, $\sigma \in (0, 1)$, $\gamma \in (0, 1)$, $a > 0$

En este trabajo presentaremos algunos métodos y condiciones de descenso monótonas y no monótonas, bases para poder estudiar el método central, luego definiremos el método del gradiente proyectado espectral, explicaremos sus algoritmos, desarrollaremos sus teoremas de convergencia y se realizara un análisis de los resultados numéricos que se obtuvieron en la referencia principal (ver [3]).

CAPÍTULO 1

PRELIMINARES

§1.1. Nociones Básicas

A continuación desarrollaremos y trataremos algunas definiciones, métodos y teoremas necesarios para una mejor comprensión de este trabajo, los cuales han sido extraídos de las referencias citadas y serán usados durante el desarrollo del trabajo.

DEFINICIÓN 1.1 (Conjunto Convexo). (ver [2]) Sea C un subconjunto de \mathbb{R}^n , Diremos que C es un conjunto convexo si

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

DEFINICIÓN 1.2 (Autovalor). (ver [2]) Sea A una matriz $n \times n$. Diremos que $\lambda \in K$ ($K = \mathbb{C}$ ó $K = \mathbb{R}$) es un autovalor o valor propio de A si existe un vector $v \in \mathbb{R}^n$ (ó \mathbb{C}^n), $v \neq 0$ tal que

$$Av = \lambda v$$

DEFINICIÓN 1.3 (Condición de una Matriz). (ver [5]) El número de condición de una matriz esta definido por.

$$K(A) = \|A\| \|A^{-1}\|,$$

donde A es una matriz $n \times n$.

Una matriz A es bien condicionada si $K(A)$ está cerca a 1 y es mal condicionada si $K(A)$ es significativamente mayor que 1.

DEFINICIÓN 1.4 (Radio Espectral). (ver [2]) El radio espectral $\rho(A)$ de una matriz A está definido por.

$$\rho(A) = \max |\lambda|,$$

donde λ es un valor característico de A .

DEFINICIÓN 1.5. (ver [2]) Sea A una matriz simétrica de orden $n \times n$, A es llamada definida positiva si $x^T Ax > 0$ para todo $x \in \mathbb{R}^n$, $x \neq 0$. Ésta es llamada definida no negativa o semidefinida positiva si $x^T Ax \geq 0$ para todo $x \in \mathbb{R}^n$.

DEFINICIÓN 1.6 (Cociente de Rayleigh). (ver [10])

Sea x una aproximación de un autovector v de la matriz A asociado a un autovalor λ ($Av = \lambda v$) que queremos calcular. El cociente $\frac{x^T Ax}{x^T x}$ se le denomina Cociente de Rayleigh del vector x respecto de la matriz A .

El sistema $\lambda x = Ax$ de orden $n \times 1$, resulta en general incompatible por no ser x exactamente un autovector sino una aproximación. Por lo que la solución que mejor se ajusta es la pseudosolución del sistema, que nos dará la mejor aproximación del valor λ .

$$\begin{aligned}\lambda x^T x &= x^T Ax. \\ \lambda &= \frac{x^T Ax}{x^T x}.\end{aligned}$$

El Cociente de Rayleigh puede utilizarse para acelerar la convergencia de los métodos, pues en una iteración k dicho cociente da una mejor aproximación al valor propio.

DEFINICIÓN 1.7 (Punto Estacionario). (ver [2]) Sea Ω un conjunto convexo, diremos que \bar{x} es un punto estacionario si.

$$\nabla f(\bar{x})^T (x - \bar{x}) \geq 0, \quad \forall x \in \Omega.$$

PROPOSICIÓN 1.1 (Desigualdad de Schwartz). (ver [2]) Para cualquier par de vectores x e y , tenemos que.

$$|x^T y| \leq \|x\| \cdot \|y\|.$$

La igualdad se tiene si y solo si $x = \alpha y$ para algún escalar α .

§1.1.1. Interpolación Cuadrática

La interpolación cuadrática (ver [2]) consiste en obtener la parábola que determinan tres puntos del plano. Si los puntos tienen abscisas distintas su ecuación es

un polinomio de segundo grado.

Este método usa tres puntos a , b y c tal que $a < b < c$, $g(a) > g(b)$ y $g(b) < g(c)$. Tal conjunto de puntos es referido como un modelo de tres puntos, allí se puede ver que un mínimo local de g debe estar entre los puntos extremos a y c de los tres puntos modelos a , b y c . En cada iteración el método fija un polinomio cuadrático a los tres valores $g(a)$, $g(b)$ y $g(c)$, y uno sustituye uno de los puntos a , b y c por el punto minimizador de este polinomio cuadrático.

Así tenemos un único mínimo $\bar{\alpha} \in (a, c)$ dado por.

$$\bar{\alpha} = \frac{1}{2} \frac{g(a)(c^2 - b^2) + g(b)(a^2 - c^2) + g(c)(b^2 - a^2)}{g(a)(c - b) + g(b)(a - c) + g(c)(b - a)}. \quad (1.1)$$

Luego formamos un nuevo modelo de tres puntos como sigue.

- si $\bar{\alpha} > b$, reemplazamos a ó c por $\bar{\alpha}$ dependiendo si $g(\bar{\alpha}) < g(b)$ o $g(\bar{\alpha}) > g(b)$ respectivamente.
- si $\bar{\alpha} < b$, reemplazamos c ó a por $\bar{\alpha}$ dependiendo si $g(\bar{\alpha}) < g(b)$ o $g(\bar{\alpha}) > g(b)$ respectivamente.

§1.2. Proyección Sobre Conjuntos Convexos

El siguiente desarrollo ha sido estudiado y extraído de la referencia [8].

Sea C un conjunto convexo, cerrado y no vacío de \mathbb{R}^n . Fijamos un $x \in \mathbb{R}^n$ y consideramos el siguiente problema.

$$\inf \left\{ \frac{1}{2} \|y - x\|^2 : y \in C \right\}. \quad (1.2)$$

Nos interesan los puntos que pertenecen a C que están más cerca de x . Así definimos la siguiente función. $f_x : \mathbb{R}^n \rightarrow \mathbb{R}$, dada por:

$$f_x(y) = \frac{1}{2} \|y - x\|^2. \quad (1.3)$$

Tomamos en cuenta el siguiente conjunto $S = \{y \in \mathbb{R}^n : f_x(y) \leq f_x(c), c \in C\}$. Así (1.2) es equivalente a.

$$\inf\{f_x(y) : y \in C \cap S\}.$$

El cual tiene una solución ya que f_x es continua y por lo tanto S y $S \cap C$ son compactos.

Así definimos un operador proyección, el cual es la función $x \mapsto P_c(x)$, $x \in \mathbb{R}^n$ donde P_c denota la proyección ortogonal sobre un subespacio $C \in \mathbb{R}^n$, y viene dado por.

$$P_c(x) = \operatorname{argmin} \left\{ \frac{1}{2} \|y - x\|^2 : y \in C \right\}.$$

TEOREMA 1.1. *Un punto $y_k \in C$ es la proyección $P_c(x)$ si y solo si.*

$$\langle x - y_k, y - y_k \rangle \leq 0, \quad \forall y \in C. \quad (1.4)$$

Demostración. Sea y_x la solución (1.4), sea $y \in C$, tal que $y_k + \alpha(y - y_k) \in C$ para algún $\alpha \in (0, 1)$ entonces usando la notación de (1.3) obtenemos lo siguiente.

$$f_x(y_x) \leq f_x(y_k + \alpha(y - y_k)) = \frac{1}{2} \|y_x - x + \alpha(y - y_x)\|^2. \quad (1.5)$$

Dado que y_x es solución de (1.4), luego.

$$\frac{1}{2} \|y_x - x\|^2 \leq \frac{1}{2} (\|y_x - x\|^2 + 2\alpha \langle y_x - x, y - y_x \rangle + \alpha^2 \|y - y_x\|^2).$$

$$\frac{1}{2} \|y_x - x\|^2 \leq \frac{1}{2} \|y_x - x\|^2 + \alpha \langle y_x - x, y - y_x \rangle + \frac{1}{2} \alpha^2 \|y - y_x\|^2.$$

Así.

$$0 \leq \alpha \langle y_x - x, y - y_x \rangle + \alpha^2 \|y - y_x\|^2.$$

Dividiendo por $\alpha > 0$ y sea $\alpha \downarrow 0$ tenemos que.

$$0 \leq -\langle x - y_x, y - y_x \rangle.$$

Por lo tanto.

$$0 \geq \langle x - y_x, y - y_x \rangle.$$

Ahora supongamos que $y_x \in C$ y.

$$\langle x - y_x, y - y_x \rangle \leq 0 \quad \forall y \in C.$$

Si $y_x = x$ tenemos que y_x resuelve (1.2).

Ahora analicemos en el caso en que $y_x \neq x$ con $y_x \in C$.

Por hipótesis tenemos que.

$$\begin{aligned} 0 &\geq \langle x - y_x, y - y_x \rangle \\ &= \langle x - y_x, y - x + x - y_x \rangle \\ &= \|x - y_x\|^2 + \langle x - y_x, y - x \rangle \\ &\geq \|x - y_x\|^2 - \|x - y_x\| \|x - y\|. \end{aligned}$$

Aquí usamos la desigualdad de Cauchy Schwartz, luego dividiendo por $\|x - y_x\| > 0$, ($y_x \neq x$).

$$\begin{aligned} 0 &\geq \|x - y_x\| - \|x - y\|, \\ \|x - y\| &\geq \|x - y_x\|, \\ \frac{1}{2}\|x - y\|^2 &\geq \frac{1}{2}\|x - y_x\|^2, \\ f_x(y) &\geq f_x(y_x) \quad \text{para todo } y \in C. \end{aligned}$$

Así y_x satisface (1.2).

Por otra parte, esta demostración prueba al mismo tiempo que la desigualdad (1.4) tiene una única solución. □

Observación 1. La desigualdad de Cauchy - Schwartz define el ángulo $\theta \in [0, \pi]$ de dos vectores distintos de ceros x e y por.

$$\cos(\theta) = \frac{\langle x, y \rangle}{\|x\| \|y\|} \in [-1, 1].$$

Así la expresión (1.4) nos dice que el ángulo entre $y - y_x$ y $x - y_x$ es obtuso, para todo $y \in C$.

PROPOSICIÓN 1.2. *La función $f : \mathbb{R}^n \rightarrow C$ definida por $f(x) = P_C(x)$ es continua y no expansiva, o lo que es lo mismo.*

$$\|P_C(x) - P_C(y)\| \leq \|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

Demostración. Escribimos (1.4) con $x = x_1$ y $y = P_c(x_2) \in C$.

$$\langle P_c(x_2) - P_c(x_1), x_1 - P_c(x_1) \rangle \leq 0.$$

Asimismo.

$$\langle P_c(x_1) - P_c(x_2), x_2 - P_c(x_2) \rangle \leq 0.$$

Y concluimos por adición.

$$\langle P_c(x_2) - P_c(x_1), x_1 - P_c(x_1) \rangle + \langle P_c(x_1) - P_c(x_2), x_2 - P_c(x_2) \rangle \leq 0.$$

Esto implica.

$$\langle P_c(x_1) - P_c(x_2), x_2 - x_1 + P_c(x_1) - P_c(x_2) \rangle \leq 0.$$

Así.

$$\langle P_c(x_1) - P_c(x_2), x_2 - x_1 \rangle + \|P_c(x_1) - P_c(x_2)\|^2 \leq 0.$$

Luego.

$$\|P_c(x_1) - P_c(x_2)\|^2 \leq \langle P_c(x_1) - P_c(x_2), x_1 - x_2 \rangle.$$

Por lo tanto.

$$\|P_c(x_1) - P_c(x_2)\|^2 \leq \|P_c(x_1) - P_c(x_2)\| \|x_1 - x_2\|.$$

En consecuencia.

$$\|P_c(x_1) - P_c(x_2)\| \leq \|x_1 - x_2\|.$$

□

TEOREMA 1.2. *(ver [13]) Supongamos que f es una aplicación continua de un espacio métrico compacto X en un espacio métrico Y , entonces $f(X)$ es compacto.*

TEOREMA 1.3 (Teorema del Valor Medio). (ver [2]) Si $f : \mathbb{R} \rightarrow \mathbb{R}$ es continuamente diferenciable sobre un intervalo abierto I , entonces para todo $x, y \in I$ existe algún $\xi \in [x, y]$ tal que.

$$f(y) - f(x) = \nabla f(\xi)(y - x).$$

Nota 1. Estos teoremas serán utilizados en las demostraciones de los lemas y los teoremas de convergencia.

§1.3. Regla de Armijo

La regla de Armijo forma una parte fundamental en el desarrollo de la metodología propuesta en la referencia principal estudiada (ver [3]), aquí presentamos un estudio basado en la referencia [2].

La regla de Armijo es una regla de reducción sucesiva en la cual se selecciona un tamaño de paso inicial s , y si el vector $x^k + sd^k$ no produce un valor mejorado de f ; es decir, $f(x^k + sd^k) \geq f(x^k)$, el tamaño de paso s es reducido por un cierto factor hasta que el valor de f es mejorado. Este método es teóricamente inestable ya que los valores mejorados en cada iteración pueden no ser suficientes para garantizar convergencia a un mínimo.

Para eliminar la dificultad de la convergencia teórica, fijamos los siguientes escalares $s > 0$, $\beta \in (0, 1)$, $\sigma \in (0, 1)$ y hacemos $\alpha^k = \beta^{m_k} s$, donde m_k es el primer entero no negativo m para el cual se cumple la siguiente desigualdad.

$$f(x^k) - f(x^k + \beta^m s d^k) \geq -\sigma \beta^m s \nabla f(x^k)^T d^k,$$

o lo que es equivalente.

$$f(x^k + \beta^m s d^k) \leq f(x^k) + \sigma \beta^m s \nabla f(x^k)^T d^k. \quad (1.6)$$

La búsqueda lineal por la regla de Armijo, comienza con un tamaño de paso de prueba s y continua con $\beta s, \beta^2 s, \dots$, hasta la primera vez que $\beta^m s$ se inscribe en el conjunto de tamaño de pasos α satisfaciendo la desigualdad (1.6).

Para ilustrar la desigualdad (1.6), consideramos.

$$g(\alpha) = f(x^k + \alpha d^k),$$

vista como una función del tamaño de paso α . Luego para $\alpha = 0$ tenemos.

$$g(0) = f(x^k).$$

Y por otra parte.

$$g'(\alpha) = \nabla f(x^k + \alpha d^k)^T d^k.$$

Y así.

$$g'(0) = \nabla f(x^k) d^k.$$

Por lo tanto podemos escribir la desigualdad (1.6) como.

$$g(\alpha) \leq g(0) + \sigma \alpha g'(0) d^k.$$

Así observamos que la expresión $L : g(0) + \sigma \alpha g'(0) d^k$ es una recta donde su pendiente es negativa ya que d^k es una dirección de descenso.

Entonces el tamaño de paso a elegir son los que hacen que la recta $L(\alpha)$ estén por encima del gráfico $g(\alpha)$ (ver figura (1.1)).

§1.4. Algunos Métodos de Optimización

A continuación desarrollamos concisamente varios métodos de optimización irrestricta: método de máximo descenso (método de Cauchy), método de Newton, método Quasi Newton y el método de la secante, cuyo desarrollo nos permite una mejor comprensión de los temas estudiados.

§1.4.1. Método de Máximo Descenso

El método clásico de máximo descenso (ver [2]) para la minimización sin restricciones de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es de la forma.

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

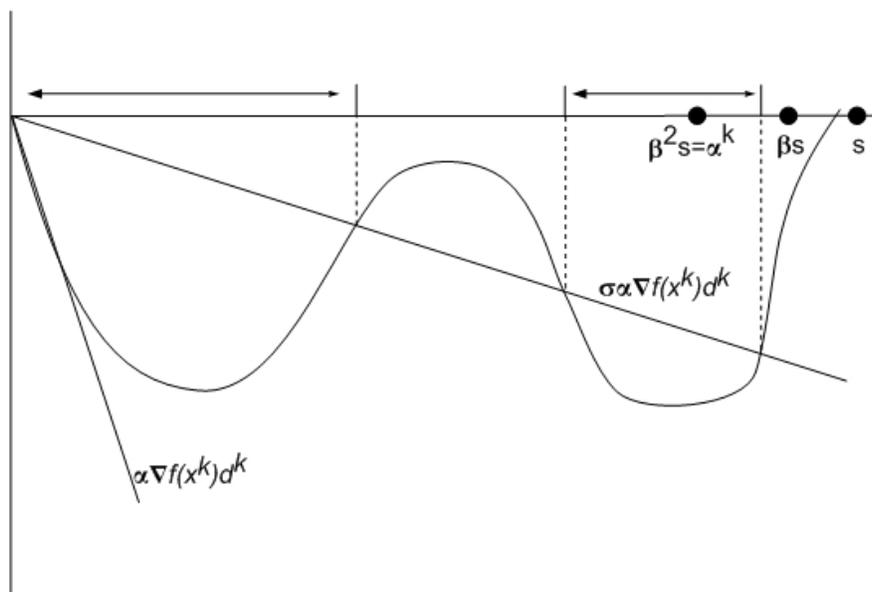


FIGURA 1.1: REGLA DE ARMIJO

La dirección de búsqueda $d_k \in \mathbb{R}^n$ es escogida como el gradiente negativo de f en x_k .

$$d_k = -\nabla f(x_k).$$

Y el tamaño de paso α_k es dado por.

$$\alpha_k = \operatorname{argmin}_{\alpha} f(x_k + \alpha d_k).$$

Si d es una dirección de descenso entonces d y $\nabla f(x)$ forman un ángulo mayor que $\pi/2$.

Una de las ventajas de la dirección de máximo descenso es que sólo se requiere de la evaluación del gradiente de la función.

TEOREMA 1.4. (ver [9]) Para algún $x_0 \in E^n$ el método de máximo descenso converge al único punto mínimo x^* de f . Es mas, con.

$$E(x) = \frac{1}{2}(x - x^*)^T Q(x - x^*). \tag{1.7}$$

se tiene que para todo paso k .

$$E(x_{k+1}) \leq \left(\frac{A - a}{A + a} \right)^2 E(x_k),$$

donde $Q = \nabla^2 f(x)$ es una matriz simétrica definida positiva y A y a son el más grande y el más pequeño autovalor de Q .

Observación 2. En el teorema E^n denota a \mathbb{R}^n con la norma euclídea. El teorema nos dice que la tasa de convergencia del método de máximo descenso es retardado como los contornos de f se hacen más excéntricos. En el caso en el cual $A = a$, corresponde a contornos circulares, y así la convergencia ocurre en un solo paso. Sin embargo, si $n - 1$ de los autovalores son iguales y uno es diferente de estos, la convergencia será muy lenta. Y así un simple autovalor distinto puede destruir la efectividad del tamaño de paso del método de máximo descenso.

§1.4.2. Método de Newton

El método de Newton (ver [5]) es conocido como una técnica para encontrar raíces de una función real. Sea $\phi(t) : \mathbb{R} \rightarrow \mathbb{R}$. Consideremos la ecuación $\phi(x^*) = 0$.

El método de Newton es basado en una aproximación lineal. Supongamos que tenemos un x_i muy cercano a x^* .

Trazamos la recta tangente a la curva en el punto $(x_i, f(x_i))$ ésta cruza al eje x en un punto x_{i+1} que será nuestra siguiente aproximación a la raíz x^* ver la figura (1.2).

Para calcular el punto x_{i+1} , calculamos primero la ecuación de la recta tangente. Sabemos que tiene pendiente $m = f'(x_i)$.

Y por lo tanto la ecuación de la recta tangente es.

$$y - f(x_i) = f'(x_i)(x - x_i).$$

Como vemos en la figura (1.2).

Hacemos $y = 0$ y despejamos x .

$$x = x_i - \frac{f(x_i)}{f'(x_i)}.$$

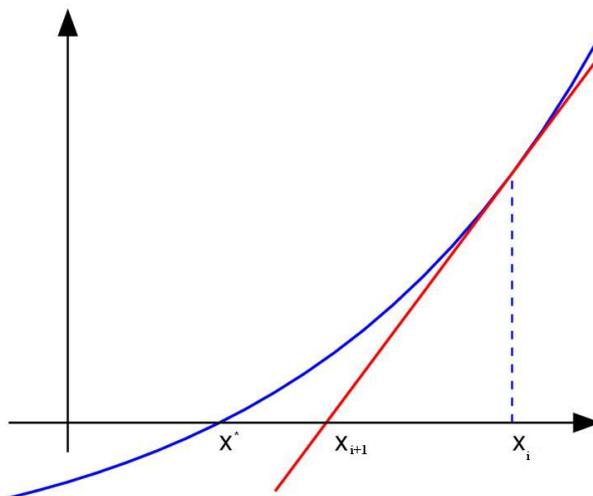


FIGURA 1.2: MÉTODO DE NEWTON

Así obtenemos la siguiente fórmula iterativa de Newton.

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}, \quad \text{si } f'(x) \neq 0. \quad (1.8)$$

Note que el método de Newton no trabaja con intervalos donde nos asegure que encontraremos la raíz, y de hecho no tenemos ninguna garantía de que nos aproximaremos a dicha raíz. Desde luego, existen ejemplos donde este método no converge a la raíz, en cuyo caso se dice que el método diverge. Sin embargo, en los casos donde sí converge a la raíz lo hace con una rapidez impresionante, por lo cual es uno de los métodos preferidos por excelencia.

También observe que hay casos donde el método no se puede aplicar, por ejemplo, cuando la derivada de f se anula en un punto, no hay definición del siguiente iterado (ver 1.8).

La dirección de Newton es obtenida después de considerar una aproximación cuadrática de f en una vecindad de x^k .

$$f(x^k + d) \approx f(x^k) + d^T \nabla f(x^k) + \frac{1}{2} d^T \nabla^2 f(x^k) d = m_k(d).$$

Si $\nabla^2 f(x)$ es definida positiva, entonces el vector que minimiza a m_k es

$$d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k).$$

La ventaja es que obtenemos una tasa de convergencia cuadrática cuando estamos cerca del punto óptimo, pero en cada iteración es necesario el cálculo del Hessiano, cosa que puede ser muy costosa.

§1.4.3. Método Quasi-Newton

El método Quasi-Newton (ver [11]) es un procedimiento que se basa en el método de Newton para encontrar el punto estacionario de una función, la idea fundamental del método es usar una aproximación de la inversa de la Hessiana en lugar de la verdadera inversa que es requerida en el método de Newton. Así que en la aproximación cuadrática, en lugar de $\nabla^2 f(x^k)$ se utiliza una aproximación B_k , la cual es actualizada en cada iteración.

El método Quasi-Newton es un proceso iterativo muy básico para resolver los problemas.

$$\text{minimizar } f(x),$$

y su iteración viene dada por.

$$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k)^T.$$

La matriz B_k es una matriz simétrica $n \times n$ y α_k es elegido para minimizar $f(x_{k+1})$. Observamos que si B_k es la Hessiana de f obtenemos el método de Newton, si $B_k = I$ tenemos el método del paso de máximo descenso.

La dirección Quasi-Newton está dada por.

$$d_k = -B_k^{-1} \nabla f(x_k).$$

La actualización de B_k en cada iteración, debe garantizar una propiedad del Hessiano. Llamada la ecuación secante.

$$B_{k+1} s_k = y_k,$$

donde $s_k = x_{k+1} - x_k$ y $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, además, B_k debe ser simétrica definida positiva.

Por otra parte s_k y y_k deben satisfacer la condición de curvatura.

$$s_k^T y_k > 0.$$

Una de las actualizaciones más usadas es la formula BFGS cuyo nombre es derivado por las iniciales de sus inventores Broyden, Fletcher, Goldfar y Shanno.

$$B_{k+1} = B_k - \frac{B_k S_k S_k^T B_k}{S_k^T B_k S_k} + \frac{Y_k Y_k^T}{S_k^T Y_k}.$$

Entre otras actualizaciones esta la formula DFP y SR1.

Al igual que en el método de Newton, este método utiliza una aproximación de segundo orden para encontrar el mínimo de una función $f(x)$.

§1.4.4. Método de la Secante

El método de la secante (ver [5]) es un procedimiento que se emplea para encontrar los ceros de una función, en este método en vez de calcular la derivada de la función en el punto de estudio, como en el método de Newton, teniendo en mente la definición de derivada, se aproxima la pendiente a la recta que una la función evaluada en el punto de estudio y en el punto de la iteración anterior. Esta técnica es de especial interés cuando el coste computacional de derivar la función de estudio y evaluarla es demasiado elevado, por lo que el método de Newton no resulta atractivo.

El método se define por la relación de recurrencia.

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n).$$

Para aplicar este método se necesitará dos aproximaciones iniciales de la raíz para poder inducir una pendiente inicial.

El método se basa en obtener la ecuación de la recta que pasa por los puntos $(x_{k-1}, f(x_{k-1}))$ y $(x_k, f(x_k))$, luego se escoge como siguiente elemento x_{k+1} la intersección de la recta secante con el eje X y su imagen; es decir, el punto $(x_{k+1}, f(x_{k+1}))$. A dicha recta se le conoce como recta secante por cortar la gráfica de la función.

Por otra parte para el caso donde el método de la secante encuentra los ceros del gradiente de la función ($\nabla f(x)$) obtenemos la siguiente iteración.

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{\nabla f(x_k) - \nabla f(x_{k-1})} \nabla f(x_k). \quad (1.9)$$

§1.4.5. Método del Gradiente Projectado

Presentamos un breve resumen del método del gradiente proyectado y la regla de Armijo a lo largo del arco de proyección (ver [2]) en lo cual se basa fuertemente la referencia principal desarrollada [ver [3]].

El método del gradiente proyectado es un método de dirección factible de la forma.

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k), \quad (1.10)$$

donde.

$$\bar{x}_k = [x_k - s_k \nabla f(x_k)]^+. \quad (1.11)$$

Así, $[\cdot]^+$ denota la proyección sobre el conjunto X , $\alpha_k \in (0, 1]$, es un tamaño de paso, y s_k es un escalar positivo. Así, para obtener el vector \bar{x}_k , tomamos un paso $-s_k \nabla f(x_k)$ a lo largo del gradiente negativo. Como en el paso de máximo descenso. Entonces proyectamos el resultado $x_k - s_k \nabla f(x_k)$ sobre X , por ello mantiene la factibilidad.

Podemos también ver el escalar s_k como tamaño de paso, esto es evidente cuando seleccionamos $\alpha_k = 1$ para todo k , el cual caso $x_{k+1} = \bar{x}_k$ y el método hace.

$$x_{k+1} = [x_k - s_k \nabla f(x_k)]^+. \quad (1.12)$$

Si $x_k - s_k \nabla f(x_k)$ es factible, la iteración del gradiente proyectado se convierte en una iteración del método de máximo descenso sin restricciones en la forma (1.10) o la forma (1.12) en este caso (1.10) toma la forma $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$ donde (1.12) toma la forma $x_{k+1} = x_k - s_k \nabla f(x_k)$, note que $x^* = [x^* - s \nabla f(x^*)]^+$ para todo $s > 0$ si y solo si x^* es estacionario, de manera que el método se detiene en ese punto.

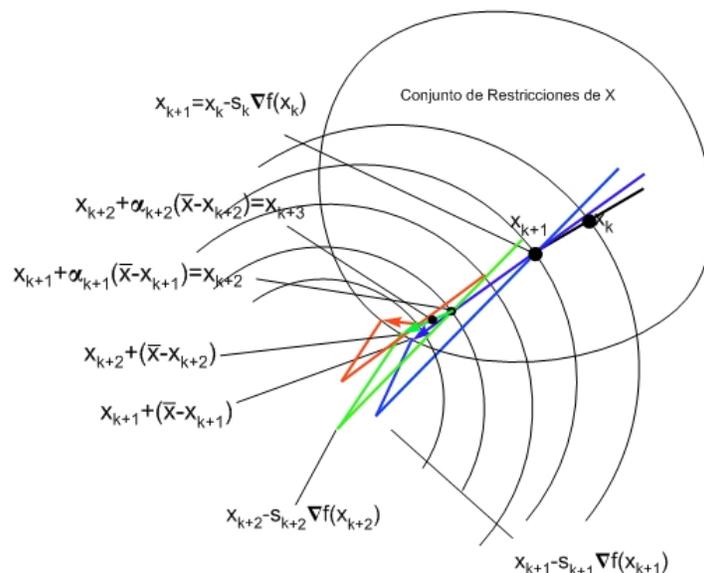


FIGURA 1.3: ITERACIONES DEL MÉTODO DEL GRADIENTE PROYECTADO.

En la figura (1.3) observamos algunas iteraciones del método del gradiente proyectado.

A fin de que el método para hacer sentido práctico, es necesario que la proyección de funcionamiento sea bastante simple. Esto será si X tiene una estructura relativamente simple. Por ejemplo, cuando existen cotas sobre las variables.

$$X = \{x / \alpha_i \leq x_i \leq \beta_i, i = 1, \dots, n\}.$$

La i -ésima coordenada de la proyección de un vector x es dada por.

$$[x]_i^+ = \begin{cases} \alpha_i, & \text{si } x_i \leq \alpha_i; \\ \beta_i, & \text{si } x_i \geq \beta_i; \\ x_i, & \text{e. o. c.} \end{cases}$$

§1.5. Regla de Armijo a lo largo del arco de proyección

Existen varios procedimientos para la selección del tamaño de paso en el método del gradiente proyectado entre ellos esta la regla de Armijo a lo largo del arco de

proyección (ver [2]).

En esta regla el tamaño de paso α^k es dado por.

$$\alpha^k = 1, \quad k = 0, 1, \dots$$

Y el tamaño de paso s^k es determinado por una reducción sucesiva hasta que una desigualdad parecida a Armijo es satisfecha; Es decir que, x_{k+1} es determinado por una búsqueda parecida a Armijo sobre el arco de proyección.

$$\{x^k(s) | s > 0\},$$

donde, para todo $s > 0$, $x^k(s)$ es definida por.

$$x^k(s) = [x^k - s\nabla f(x^k)]^+. \quad (1.13)$$

Fijando escalares \bar{s} , β , y $\sigma > 0$, con $\bar{s} > 0$, $\beta \in (0, 1)$, y $\sigma \in (0, 1)$ son elegidos, y hacemos $s^k = \beta^{m_k}$, donde m_k es el primer entero no negativo m para el cual se cumple.

$$f(x^k) - f(x^k(\beta^m \bar{s})) \geq \sigma \nabla f(x^k)^T (x^k - x^k(\beta^m \bar{s})). \quad (1.14)$$

Esta desigualdad es llamada la regla de Armijo a lo largo del arco de proyección.

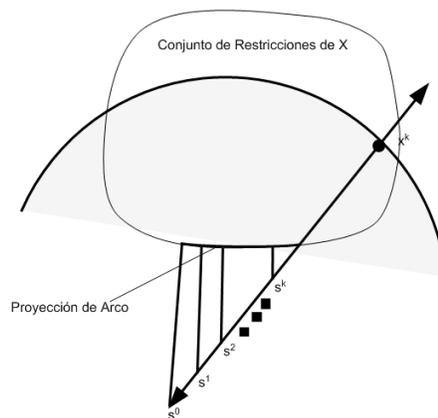


FIGURA 1.4: ARCO DE PROYECCIÓN

Podemos observar geoméricamente el arco de proyección en la figura (1.4).

Ejemplo 1.5.1. Encontrar los autovalores de la siguiente función $f(x, y) = \frac{x^2}{4} + \frac{y^2}{9}$.

Tenemos que la representación matricial de $f(x, y)$ viene dada por.

$$f(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{9} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (1.15)$$

Sea

$$A = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{9} \end{pmatrix}$$

Así el polinomio característico es el siguiente.

$$\begin{aligned} P(\lambda) &= \det(A - \lambda I) \\ &= \begin{vmatrix} \frac{1}{4} - \lambda & 0 \\ 0 & \frac{1}{9} - \lambda \end{vmatrix} \\ &= \left(\frac{1}{4} - \lambda\right) \left(\frac{1}{9} - \lambda\right). \end{aligned}$$

Luego para conseguir los autovalores hacemos $P(\lambda) = 0$, por lo tanto.

$$\left(\frac{1}{4} - \lambda\right) \left(\frac{1}{9} - \lambda\right) = 0$$

Esto implica que $\lambda = \frac{1}{4}$ ó $\lambda = \frac{1}{9}$.

§1.6. El Método Gradiente con tamaño de paso de dos puntos

La siguiente síntesis de la referencia [1] nos permite apreciar la obtención del tamaño de paso de Barzilai y Borwein, lo que es de gran importancia para llevar a cabo el método central de este trabajo.

El método de Cauchy (método de máximo descenso) a pesar de su simplicidad y que su tamaño de paso viene dado por $\alpha_k = \arg \min f(x_k + \alpha d_k)$, al ser aplicada a funciones de forma cuadrática ($f(x) = \frac{1}{2}x^T Qx - x^T b$, donde Q es una matriz simétrica y positiva semidefinida) la rata de convergencia depende de el radio de las

longitudes de los ejes de los contornos elípticos de la función, el cual la longitud del eje corresponde, al i -ésimo autovector de Q que es proporcional a $\frac{1}{\lambda_i}$, donde λ_i es un autovalor de Q para $i = 1, 2, \dots, n$ (ver ejemplo (1.5.1)). Por esto, en algunos casos el método de Cauchy funciona mal, este es muy afectado por el mal condicionamiento.

El método gradiente con tamaño de dos puntos produce dos nuevos tamaños de pasos que pueden usarse independientemente en conjunción con la dirección negativa del gradiente, estos tamaños de paso requieren menos cálculo computacional y son menos afectados por el mal condicionamiento. Este método genera la iteración.

$$x_{k+1} = x_k - S_k g_k,$$

donde $g_k = \nabla f(x_k)$, $S_k = \alpha_k I$ y α_k minimiza a $\|\Delta x - \alpha \Delta g\|^2$ con $\Delta x = x_k - x_{k-1}$ y $\Delta g = g_k - g_{k-1}$.

Esta elección se debe a que este proporciona una aproximación de dos puntos a la ecuación de la secante subyacente del método Quasi-Newton. Esto produce la iteración.

$$x_{k+1} = x_k - \alpha_k g_k, \tag{1.16}$$

donde α_k es dado por.

$$\alpha_k = \frac{\langle \Delta x, \Delta g \rangle}{\Delta g, \Delta g}. \tag{1.17}$$

En efecto.

Definimos $f(\alpha) = \|\Delta x - \alpha \Delta g\|^2$.

Tenemos que.

$$\begin{aligned} f(\alpha_k) &= (\langle \Delta x - \alpha \Delta g, \Delta x - \alpha \Delta g \rangle) \\ f(\alpha_k) &= (\langle \Delta x, \Delta x \rangle - \alpha \langle \Delta x, \Delta g \rangle - \alpha \langle \Delta g, \Delta x \rangle + |\alpha|^2 \langle \Delta g, \Delta g \rangle). \end{aligned}$$

Así.

$$f'(\alpha_k) = -\langle \Delta x, \Delta g \rangle - \langle \Delta g, \Delta x \rangle + 2\alpha \langle \Delta g, \Delta g \rangle.$$

Dado que α_k minimiza a $\|\Delta x - \alpha \Delta g\|^2$, $f'(\alpha_k) = 0$, así.

$$-\langle \Delta x, \Delta g \rangle - \langle \Delta g, \Delta x \rangle + 2\alpha \langle \Delta g, \Delta g \rangle = 0. \quad (1.18)$$

Por lo tanto despejando α .

$$\alpha = \frac{\langle \Delta x, \Delta g \rangle}{\langle \Delta g, \Delta g \rangle}.$$

Por simetría, podemos minimizar $\|\Delta x - \alpha \Delta g\|^2$ con respecto a α y nos resulta.

$$\alpha_k = \frac{\langle \Delta x, \Delta x \rangle}{\langle \Delta x, \Delta g \rangle}. \quad (1.19)$$

En el caso unidimensional observamos que usando (1.16) con (1.17) ó (1.19) obtenemos la iteración del método de la secante para el caso donde el método encuentra los ceros del gradiente de la función (la ecuación (1.9)). En efecto.

En este caso tenemos que α_k minimiza $|\Delta x - \alpha \Delta g|^2$ o lo que es equivalente $(\Delta x - \alpha \Delta g)^2$. Consideremos la función.

$$g(\alpha) = (\Delta x - \alpha \Delta g)^2.$$

Así.

$$g'(\alpha_k) = 0.$$

$$2(\Delta x - \alpha \Delta g) \Delta g = 0.$$

Dado que $\Delta g \neq 0$, entonces.

$$\begin{aligned} 2(\Delta x - \alpha \Delta g) &= 0 \\ \Delta x &= \alpha \Delta g \\ \alpha &= \frac{\Delta x}{\Delta g}. \end{aligned}$$

Esto implica que.

$$\alpha = \frac{x_k - x_{k-1}}{\nabla f(x_k) - \nabla f(x_{k-1})}.$$

Luego.

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{\nabla f(x_k) - \nabla f(x_{k+1})} \nabla f(x_k).$$

Como los algoritmos de dos puntos no son algoritmos de descenso, tienen una ventaja con respecto a los algoritmos de descenso debido a que las restricciones de los algoritmos de descenso muchas veces resultan en pequeños tamaños de pasos para problemas mal condicionados.

§1.7. Sobre la elección del tamaño de paso de Barzilai y Borwein para el método gradiente

Presentamos un breve estudio de la referencia [12], la cual nos permite una mejor comprensión del tamaño de paso de Barzilai y Borwein que da origen al término espectral.

El método de Barzilai y Borwein para la minimización sin restricciones de una función diferenciable $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es definida por.

$$x_{k+1} = x_k - \frac{1}{\alpha_k} g_k. \quad (1.20)$$

donde g_k es el vector gradiente de f en x_k y el escalar α_k es dado por.

$$\alpha_k = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}. \quad (1.21)$$

o

$$\alpha_k = \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}, \quad (1.22)$$

donde $s_{k-1} = x_k - x_{k-1}$ y $y_{k-1} = g_k - g_{k-1}$.

Cuando $f(x) = \frac{1}{2}x^T Ax - b^T x + c$ es una función cuadrática y A es una matriz definida positiva, entonces α_k en (1.21) y (1.22) se convierte en.

$$\alpha_k = \frac{s_{k-1}^T A s_{k-1}}{s_{k-1}^T s_{k-1}}. \quad (1.23)$$

$$\alpha_k = \frac{s_{k-1}^T A^2 s_{k-1}}{s_{k-1}^T A s_{k-1}}. \quad (1.24)$$

Respectivamente. En efecto.

Tenemos que.

$$\begin{aligned} y_{k-1} &= g_k - g_{k-1} = (Ax_k - b) - (Ax_{k-1} - b) \\ &= A(x_k - x_{k-1}) \\ &= As_{k-1}. \end{aligned}$$

Por lo tanto.

$$\alpha_k = \frac{s_{k-1}^T As_{k-1}}{s_{k-1}^T s_{k-1}}$$

ó

$$\alpha_k = \frac{s_{k-1}^T A^2 s_{k-1}}{s_{k-1}^T As_{k-1}}.$$

En este caso, α_k es el cociente de Rayleigh de A en el vector s_{k-1} o en el vector $\sqrt{A}s_{k-1}$ por la definición (1.6). Como A es simétrica definida positiva, entonces.

$$0 < \lambda_{\min} \leq \alpha_k \leq \lambda_{\max}, \quad \text{para todo } k,$$

donde λ_{\min} y λ_{\max} son respectivamente los autovalores mas pequeño y mas grande de A . Por lo tanto, no se divide por cero en (1.20).

Las formulas (1.21) y (1.22) son usadas para generar los coeficientes α_k con tal de que ellos estén cercanos a 0 y que ellos nos sean muy grandes; es decir. El método usa protección $0 < \alpha_{\min} < \alpha_{\max} < \infty$ y define, para cada iteración.

$$\alpha_{k+1} = \min \left\{ \alpha_{max}, \max \left\{ \alpha_{min}, \frac{s_k^T y_k}{s_k^T s_k} \right\} \right\}.$$

§1.8. Una técnica de Búsqueda lineal no monótona para el método de Newton

En esta sección mostraremos un análisis del trabajo realizado por Grippo, Lampariello y Lucidi (ver[7]). Quienes exponen una técnica de búsqueda lineal no monótona para el método del gradiente proyectado (ver 1.4.5). Este trabajo es relevante en el desarrollo del método principal (ver [1]).

Se sabe que el método de Newton en su forma pura tiene serias desventajas, ya que requiere el uso de una técnica de búsqueda lineal el cual garantice un decrecimiento monótono de la función objetivo y esto hace que el método tenga poca

rapidez en la presencia de curvas y valles estrictos.

La técnica de búsqueda lineal no monótona para el método de Newton puede ser vista como una generalización de la regla de Armijo, en el sentido que este permita un aumento en la función sin afectar la propiedad de convergencia.

Sea (x_k) una sucesión definida por.

$$x_{k+1} = x_k + \alpha_k d_k \quad \text{con} \quad d_k \neq 0.$$

Para llevar a cabo dicha técnica se deben cumplir las siguientes condiciones.

Condición 1 i $g_k^T d_k \leq -C_1 \|g_k\|^2$, C_1 positivo.

ii $\|d_k\| \leq C_2 \|g_k\|$, C_2 positivo.

Condición 2 $\alpha_k = \sigma^{h_k} a$, donde h_k es el primer entero no negativo h para el cual.

$$f(x_k + \sigma^h a d_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \gamma \sigma^h a g_k^T d_k \quad (1.25)$$

donde $m(0) = 0$, $0 \leq m(k) \leq \min[m(k-1) + 1, M]$, $k \geq 1$, M es un entero no negativo, $\sigma \in (0, 1)$, $\gamma \in (0, 1)$, $a > 0$.

Si la condición 1 es satisfecha se establece que d_k es “gradient related” (ver [2]), la cual básicamente significa que d_k no es “muy grande” ni “muy pequeña” relativa al gradiente de $f(x_k)$ y que el ángulo entre d_k y $\nabla f(x_k)$ no está “muy cerca” a 90° . La condición 2 representa un criterio de reducción suficiente sobre f , el cual acota el tamaño de paso de una manera que, en caso de convergencia, tenemos que $\alpha_k \|d_k\| \rightarrow 0$.

Ahora describimos el procedimiento de la técnica de búsqueda lineal no monótona para el método de Newton.

Algoritmo

Datos: x_0 , entero $M \geq 0$, $c_1 > 0$, $c_2 > 0$, $\gamma \in (0, 1)$, $\sigma \in (0, 1)$.

Paso 1 Sea $k = 0$, $m(0) = 0$ y calcular $f_0 \triangleq f(x_0)$.

Paso 2 Calcular g_k ; si $g_k = 0$ para, sino se calcula H_k ; si H_k es singular, hacemos

$d_k = -g_k$, $m(k) = 0$ y vamos al paso 5.

Paso 3 Calcular $d_k = -H_k^{-1}g_k$. Si $|g_k^T d_k| < c_1 \|g_k\|^2$ o $\|d_k\| > c_2 \|g_k\|$, hacemos $d_k = -g_k$, $m(k) = 0$ y vamos al paso 5.

Paso 4 Si $g^T d_k > 0$, hacemos $d_k = -d_k$.

Paso 5 Sea $\alpha = 1$.

Paso 6 Calcular $f_\alpha = f(x_k + \alpha d_k)$. Si

$$f_\alpha \leq \max_{0 \leq j \leq m(k)} \{f_{k-j}\} + \gamma \alpha g_k^T d_k,$$

entonces se define $f_{k+1} = f_\alpha$, $x_{k+1} = x_k + \alpha d_k$, $k = k+1$, $m(k) \leq \min\{m(k-1)+1, M\}$ y vamos al paso 2.

Paso 7 Sea $\alpha = \sigma \alpha$ y vamos al paso 6.

CAPÍTULO 2

MÉTODO DEL GRADIENTE PROYECTADO ESPECTRAL

En este capítulo estudiamos y presentamos un resumen de los principales aspectos desarrollados en la referencia principal [3]. Mostramos en forma detallada las demostraciones de los teoremas allí considerados y de dos lemas cuyas demostraciones fueron analizadas de la referencia [2]. Un aspecto que podría considerarse como una novedad es la inclusión de gráficos que nos permite una mejor comprensión en el tema estudiado.

El método del gradiente proyectado espectral no monótono es aplicado a problemas de la forma.

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeto a} && x \in \Omega, \end{aligned} \tag{2.1}$$

donde Ω es un conjunto convexo y cerrado en \mathbb{R}^n el cual es el conjunto de restricciones factibles y $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función continua que admite derivadas parciales continuas en un abierto que contiene Ω y es llamada función objetivo.

En esta sección describiremos los algoritmos de dicho método y demostraremos la buena definición de los algoritmos, y que cada punto de acumulación de la sucesión generada de estos algoritmos es un punto estacionario.

El método del gradiente proyectado espectral es una estrategia no monótona de implementación eficiente en cuanto a tiempos de ejecución (ver [3]). En cada iteración, el método combina el método del gradiente proyectado con el esquema de búsqueda lineal no monótona desarrollada por Grippo, Lampariello y Lucidi para el método de Newton (ver [7]) y esta variante combinada con el tamaño de paso espectral introducido por Raydan y analizado por Barzilai y Borwein (ver [12] y [1]).

Esta combinación nos ayuda a acelerar notablemente la velocidad de convergencia, requiere menos trabajo computacional y esto nos asegura la convergencia global del método.

El método del gradiente proyectado espectral para el problema (2.1) viene dado por la iteración.

$$x_{k+1} = x_k + \lambda d_k \quad \text{ó} \quad x_{k+1} = P(x_k - \alpha_k \nabla f(x_k)),$$

donde $\lambda = \alpha_k$ y la dirección viene dada por.

$$d = P(x_k - \alpha_k \nabla f(x_k)) - x_k.$$

Tenemos que $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$ y $\alpha_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}$ el cociente de Rayleigh inverso.

El uso de x_{k+1} depende del algoritmo que se utilizara; es decir, si usaremos el algoritmo SPG1 o el algoritmo SPG2 que se presentan en la siguiente sección.

§2.1. Algoritmos

A continuación describiremos los algoritmos del método del gradiente proyectado espectral, los cuales son llamados SPG1 y SPG2 (ver [1]).

Los algoritmos comienzan con un punto inicial $x_0 \in \Omega$ y usan un entero $M \geq 1$, un parámetro pequeño $\alpha_{\min} > 0$, un parámetro grande $\alpha_{\max} > \alpha_{\min}$, un parámetro de suficiente reducción $\gamma \in (0, 1)$, los parámetros de protección $0 < \sigma_1 < \sigma_2 < 1$ y denotamos $g(x) = \nabla f(x)$.

Dado $x_k \in \Omega$ y $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$, tomamos inicialmente $\alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$ arbitrario, los algoritmos SPG1 y SPG2 describen como obtener x_{k+1} , α_{k+1} y cuando termina el proceso.

Algoritmo SPG1.

Paso 1 Detectar si el punto actualizado es estacionario.

Si $\|P(x_k - \alpha_k g(x_k)) - x_k\| = 0$, para, declarando que x_k es estacionario.

Paso 2 Backtracking.

Paso 2.1 Sea $\lambda \leftarrow \alpha_k$.

Paso 2.2 Sea $x_+ = P(x_k - \lambda g(x_k))$.

Paso 2.3 Si.

$$f(x_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \gamma \langle x_+ - x_k, g(x_k) \rangle, \quad (2.2)$$

entonces se define $\lambda_k = \lambda$, $x_{k+1} = x_+$, $s_k = x_{k+1} - x_k$, $y_k = g(x_{k+1}) - g(x_k)$, y vamos al Paso 3.

Si (2.2) no se cumple, definimos.

$$\lambda_{nuevo} \in [\sigma_1 \lambda, \sigma_2 \lambda]. \quad (2.3)$$

Sea $\lambda \leftarrow \lambda_{nuevo}$, y vamos al Paso 2.2.

Paso 3

Calcular $b_k = \langle s_k, y_k \rangle$.

Si $b_k \leq 0$, sea $\alpha_{k+1} = \alpha_{\max}$; sino, calcular $a_k = \langle s_k, s_k \rangle$ y.

$$\alpha_{k+1} = \min\{\alpha_{\max}, \max\{\alpha_{\min}, a_k/b_k\}\}.$$

El algoritmo SPG1 toma puntos de la forma $P(x_k - \lambda g(x_k))$ para $\lambda \in (0, \alpha_k]$ durante el proceso de búsqueda unidimensional, el cual en general forma una ruta curvilínea. Por lo tanto, el producto escalar $\langle x_+ - x_k, g(x_k) \rangle$ en la condición (2.2) será calculado para cada punto de prueba x_+ . Mas aún, en la formulación del algoritmo SPG1 la distancia entre dos puntos consecutivos podría ser muy pequeña o siempre nula en la vecindad de puntos esquinas del conjunto Ω . De hecho, la distancia entre las proyecciones de dos puntos sobre el conjunto factible pueden ser pequeñas, siempre y cuando los puntos son distantes entre sí. Claramente, al evaluar la función objetivo sobre dos puntos cercanos representa un mal uso de la información.

Ahora veamos gráficamente el caso cuando λ no cumple la condición (2.2) (ver figura (2.1)) y algunas iteraciones del SPG1 (ver figura (2.2)).

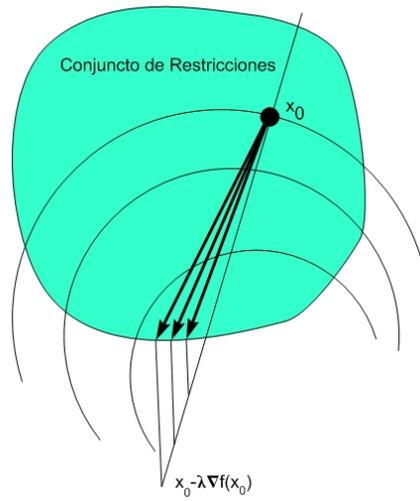


FIGURA 2.1: CASO DONDE λ NO CUMPLE LA CONDICIÓN (2.2)

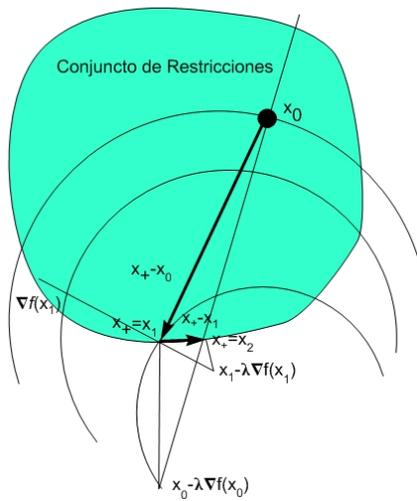


FIGURA 2.2: ITERACIONES DEL SPG1

Debido a la deficiente información dada por el algoritmos SPG1 al evaluar la función en dos puntos cercanos, los autores de la referencia [3] se motivaron a definir el algoritmo SPG2. Este algoritmo también está basado sobre la dirección del gradiente proyectado espectral $P(x_k - \alpha_k g(x_k)) - x_k$, con α_k como el cociente de Rayleigh inverso.

$$\alpha_k = \frac{\langle s_{k-1}, s_{k-1} \rangle}{\langle s_{k-1}, y_{k-1} \rangle}.$$

Observamos que $\frac{\langle s_{k-1}, y_{k-1} \rangle}{\langle s_{k-1}, s_{k-1} \rangle}$ es el cociente de Rayleigh correspondiente al promedio de la matriz Hessiana $\int_0^1 \nabla^2 f(x_{k-1} + ts_{k-1}) dt$. En efecto.

El método Quasi Newton para optimización sin restricciones obedece la formula recursiva.

$$x_{k+1} = x_k + \lambda_k B_k^{-1} \nabla f(x_k).$$

La sucesión de matrices $\{B_k\}$ satisface la ecuación secante.

$$B_{k+1} s_k = y_k. \tag{2.4}$$

Deseamos una matriz con una simple estructura que cumpla (2.4), mas precisamente, deseamos.

$$B_{k+1} = \alpha_{k+1} I,$$

donde $\alpha_{k+1} \in \mathbb{R}$ y así la ecuación (2.4) nos queda.

$$\alpha_{k+1} s_k = y_k.$$

Aceptando el estudio realizado en (1.6) de que α_k minimiza $\|\alpha s_k - y_k\|^2$ obtenemos que.

$$\alpha_{k+1} = \frac{s_k^T y_k}{s_k^T s_k}. \tag{2.5}$$

Ahora observamos que $y_k = \int_0^1 \nabla^2 f(x_k + ts_k) dt$.

haciendo un cambio de variable.

$$u = x_k + ts_k.$$

Tenemos que.

$$du = s_k dt.$$

Así.

$$\begin{aligned}
\left(\int_0^1 \nabla^2 f(x_k + ts_k) dt \right) s_k &= \int_{x_k}^{x_k+s_k} \nabla^2 f(u) du \\
&= \nabla f(x_k + s_k) - \nabla f(x_k) \\
&= \nabla f(x_{k+1}) - \nabla f(x_k) \\
&= y_k.
\end{aligned}$$

Por lo tanto la ecuación (2.5) nos queda.

$$\alpha_{k+1} = \frac{s_k \left(\int_0^1 \nabla^2 f(x_k + ts_k) dt \right) s_k}{s_k^T s_k},$$

y por la definición (1.6) se tiene lo deseado.

En el algoritmo SPG2 en el caso de rechazar el primer punto de prueba, el siguiente punto sera calculado a lo largo de la misma dirección. Como consecuencia, $\langle x_+ - x_k, g(x_k) \rangle$ deberá ser calculado en la primera prueba y la operación proyección deberá ser realizada solo una vez por iteración, este algoritmo coincide con SPG1 excepto en el paso backtracking.

Algoritmo SPG2.

Paso 1. Detectar si el punto actualizado es estacionario.

Si $\|P(x_k - \alpha_k g(x_k)) - x_k\| = 0$, para, declarando que x_k es estacionario.

Paso 2 Backtracking.

Paso 2.1 Calcular $d_k = P(x_k - \alpha_k g(x_k)) - x_k$. Sea $\lambda \leftarrow 1$.

Paso 2.2 Sea $x_+ = x_k + \lambda d_k$.

Paso 2.3 Si.

$$f(x_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \gamma \lambda \langle d_k, g(x_k) \rangle, \quad (2.6)$$

entonces se define $\lambda_k = \lambda$, $x_{k+1} = x_+$, $s_k = x_{k+1} - x_k$, $y_k = g(x_{k+1}) - g(x_k)$, y vamos al Paso 3.

Si (2.6) no se cumple, definimos.

$$\lambda_{nuevo} \in [\sigma_1 \lambda, \sigma_2 \lambda]. \quad (2.7)$$

Sea $\lambda \leftarrow \lambda_{nuevo}$, y vamos al Paso 2.2.

Paso 3

Calcular $b_k = \langle s_k, y_k \rangle$.

Si $b_k \leq 0$, sea $\alpha_{k+1} = \alpha_{max}$; sino, calcular $a_k = \langle s_k, s_k \rangle$ y.

$$\alpha_{k+1} = \min\{\alpha_{\max}, \max\{\alpha_{\min}, a_k/b_k\}\}.$$

En las figuras (2.3) y (2.4) observaremos geoméricamente el desarrollo del algoritmo SPG2, en el caso en que el λ no cumple (2.6) y el caso cuando λ cumple con la condición (2.6).

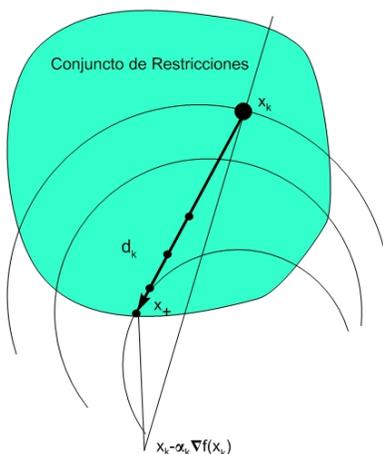


FIGURA 2.3: CASO DONDE λ NO CUMPLE LA CONDICIÓN (2.6)

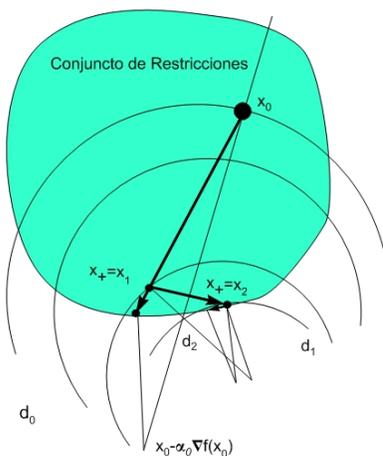


FIGURA 2.4: ITERACIONES DEL SPG2

§2.2. Teoremas de Convergencia

Ahora desarrollaremos los teoremas de convergencia de los algoritmos SPG1 y SPG2, estos teoremas fueron estudiados y analizados de la referencia principal (ver [3]). Antes de llevar a cabo el desarrollo de los teoremas, mostraremos dos lemas los cuales fueron obtenidos de la referencia [2], dichos lemas son utilizados en la evolución de las demostraciones de los teoremas de convergencia.

Definamos la siguiente función.

$$g_t(x) = P(x - tg(x)) - x \quad \text{para todo } x \in \Omega, \quad t > 0, \quad (2.8)$$

donde Ω es un conjunto convexo y cerrado en \mathbb{R}^n .

LEMA 2.1. *Para todo $x \in \Omega$, $t \in (0, \alpha_{max}]$.*

$$(i) \quad \langle g(x), g_t(x) \rangle \leq -\frac{1}{t} \|g_t(x)\|_2^2 \leq -\frac{1}{\alpha_{max}} \|g_t(x)\|_2^2.$$

(ii) *El vector $g_t(\bar{x}) = 0$ si y solo si \bar{x} es un punto estacionario restringido.*

Demostración. (i)

Por el teorema (1.1) tenemos lo siguiente.

$$\langle x - tg(x) - P(x - tg(x)), x - P(x - tg(x)) \rangle \leq 0 \quad \text{para todo } x \in \Omega.$$

Y de aquí obtenemos lo siguiente.

$$\begin{aligned} \langle x - tg(x) - P(x - tg(x)), x - P(x - tg(x)) \rangle &\leq 0, \\ \langle P(x - tg(x)) - x + tg(x), P(x - tg(x)) - x \rangle &\leq 0, \\ \langle g_t(x) + tg(x), g_t(x) \rangle &\leq 0, \\ \|g_t(x)\|^2 + t \langle g(x), g_t(x) \rangle &\leq 0, \\ t \langle g(x), g_t(x) \rangle &\leq -\|g_t(x)\|^2, \\ \langle g(x), g_t(x) \rangle &\leq -\frac{1}{t} \|g_t(x)\|^2. \end{aligned}$$

Por otra parte tenemos que $t \in (0, \alpha_{max}]$, así.

$$\begin{aligned}
t &\leq \alpha_{max}, \\
\frac{1}{t} &\geq \frac{1}{\alpha_{max}}, \\
-\frac{1}{t} &\leq -\frac{1}{\alpha_{max}}, \\
-\frac{1}{t}\|g_t(x)\|_2^2 &\leq -\frac{1}{\alpha_{max}}\|g_t(x)\|_2^2.
\end{aligned}$$

Así obtenemos.

$$\langle g(x), g_t(x) \rangle \leq -\frac{1}{t}\|g_t(x)\|^2 \leq -\frac{1}{\alpha_{max}}\|g_t(x)\|_2^2.$$

(ii) Tenemos que $g_t(\bar{x})$ viene dado por.

$$g_t(\bar{x}) = P(\bar{x} - tg(\bar{x})) - \bar{x}, \quad \bar{x} \in \Omega, \quad t > 0.$$

Supongamos que $g_t(\bar{x}) = 0$, Así $P(\bar{x} - tg(\bar{x})) = \bar{x}$.

Como \bar{x} es la proyección de $\bar{x} - tg(\bar{x})$ por el teorema (1.1) obtenemos que.

$$(\bar{x} - tg(\bar{x}) - \bar{x})^T(x - \bar{x}) \leq 0, \quad \text{para todo } x \in \Omega.$$

Esto es equivalente a.

$$-tg(\bar{x})^T(x - \bar{x}) \leq 0 \quad \text{para todo } x \in \Omega.$$

Si y solo si.

$$g(\bar{x})^T(x - \bar{x}) \geq 0 \quad \text{para todo } x \in \Omega.$$

Por definición (1.7) tenemos que \bar{x} es un punto estacionario. \square

LEMA 2.2. (i) Para todo $x \in \Omega$ y $z \in \mathbb{R}^n$, la función $h : [0, \infty) \rightarrow \mathbb{R}$ dada por.

$$h(s) = \frac{\|P(x + sz) - x\|}{s}, \quad \text{para todo } s > 0. \quad (2.9)$$

es monotonicamente decreciente.

(ii) Para todo $x \in \Omega$ existe $s_x > 0$ tal que para todo $s \in [0, s_x]$ se tiene que.

$$f(x) - f(P(x - sg(x))) \geq \sigma \nabla f(x)^T(x - P(x - sg(x))). \quad (2.10)$$

Demostración. (i) Fijemos $x \in \Omega$, $z \in \mathbb{R}^n$ y $\gamma > 1$.

Luego denotamos.

$$a = x + z, \quad b = x + \gamma z,$$

y sean \bar{a} y \bar{b} la proyección sobre Ω de a y b respectivamente.

Queremos probar que $h(\gamma) \leq h(1)$, tenemos que.

$$h(\gamma) = \frac{\|P(x + \gamma z) - x\|}{\gamma} = \frac{\|\bar{b} - x\|}{\gamma}.$$

$$h(1) = \|\bar{a} - x\|.$$

Así que mostrar que $h(\gamma) \leq h(1)$ es equivalente a mostrar que.

$$\|\bar{b} - x\| \leq \|\bar{a} - x\|\gamma. \quad (2.11)$$

Ahora estudiemos los siguientes casos.

Caso I. Si $\bar{a} = x$ entonces $\bar{b} = x$.

Así obtenemos lo siguiente.

$$\|\bar{b} - x\| = \|x - x\| = 0$$

y

$$\|\bar{a} - x\| = \|x - x\| = 0.$$

Por lo tanto se cumple la ecuación (2.11).

Caso II. Si $a \in \Omega$, entonces $\bar{a} = a$, por lo tanto.

$$\begin{aligned} \gamma\|\bar{a} - x\| &= \gamma\|x + z - x\| \\ &= \gamma\|z\| \\ &= \|\bar{b} - x\|. \end{aligned}$$

Dado que la proyección es una aplicación no expansiva [ver proposición (1.2)] tenemos que.

$$\|\bar{b} - \bar{x}\| \leq \|b - x\|,$$

donde \bar{x} es la proyección de x en Ω , pero como $x \in \Omega$, $\bar{x} = x$, entonces.

$$\|\bar{b} - \bar{x}\| \leq \|\bar{b} - x\|.$$

Con esto vemos que se cumple la ecuación (2.11).

Caso III. En el caso en que $\bar{a} = \bar{b}$ se tiene (2.11)

Caso IV. Veamos el caso en que $\bar{a} \neq \bar{b}$, $\bar{a} \neq x$, $\bar{b} \neq x$ y a no pertenece a Ω .

Sean H_a y H_b dos hiperplanos ortogonales a $\bar{b} - \bar{a}$ y pasan por \bar{a} y \bar{b} respectivamente. Como observamos en la figura (2.5).

Como.

$$(\bar{b} - \bar{a})^T (b - \bar{b}) \leq 0$$

$$(\bar{b} - \bar{a})^T (a - \bar{a}) \leq 0.$$

Esto nos dice que ni a ni b están estrictamente entre los hiperplanos H_a y H_b . Además x esta en el mismo lado de H_a como a , así x no pertenece a H_a .

Denotamos la intersección de la línea $\{x + \alpha(b - x)/\alpha \in \mathbb{R}\}$ con H_a y H_b por S_a y S_b respectivamente.

Por otra parte denotamos la intersección de la línea $\{x + \alpha(\bar{a} - x)/\alpha \in \mathbb{R}\}$ con H_b por w .

Además tenemos que.

$$\geq \frac{\|\bar{b} - x\|}{\|\bar{a} - x\|}.$$

Donde la sexta igualdad la obtenemos por similitud de triángulos, la siguiente y la última desigualdad la obtenemos por la relación de ortogonalidad $(w - \bar{b})^T(\bar{b} - \bar{a}) = 0$, y las últimas desigualdades siguen de la desigualdad triangular, así.

$$\|\bar{b} - x\| \leq \gamma \|\bar{a} - x\|.$$

(ii)

Sea $x(s) = P(x - s\nabla f(x))$.

Por el teorema (1.1) tenemos lo siguiente.

$$(x - x(s))^T(x - s\nabla f(x) - x(s)) \leq 0 \quad \text{para todo } x \in \Omega, \quad s > 0.$$

Así obtenemos.

$$(x - x(s))^T(x - x(s)) - (x - x(s))^T s \nabla f(x) \leq 0.$$

Luego.

$$\frac{\|x - x(s)\|^2}{s} \leq \nabla f(x)^T(x - x(s)). \quad (2.12)$$

Por el teorema del valor medio (1.3), tenemos que para todo $x \in \Omega$ y $s \geq 0$.

$$\begin{aligned} f(x) - f(x(s)) &= \nabla f(\xi_s)^T(x - x(s)) \\ &= \nabla f(\xi_s)^T(x - x(s)) - \nabla f(x)^T x + \nabla f(x)^T x + \nabla f(x)^T x(s) - \nabla f(x)^T x(s) \\ &= \nabla f(x)^T(x - x(s)) + (\nabla f(\xi_s) - \nabla f(x))^T(x - x(s)). \end{aligned}$$

Donde ξ_s está en el segmento de línea que une a x y $x(s)$.

Si x es estacionario se tiene que.

$$x - x(s) = 0.$$

Así.

$$f(x) - f(x(s)) = 0.$$

Luego la desigualdad (2.10) se cumple.

Por lo tanto la conclusión se tiene con s_x algún escalar positivo.

Supongamos que x no es un punto estacionario, Así.

$$\|x - x(s)\| \neq 0, \quad \text{para todo } s > 0.$$

Usando nuevamente el teorema del valor medio (1.3), tenemos que para todo $x \in X$ y $s \geq 0$.

$$f(x) - f(x(s)) = \nabla f(x)^T(x - x(s)) + (\nabla f(\xi_s) - \nabla f(x))^T(x - x(s)).$$

Luego la desigualdad.

$$f(x) - f(x(s)) \geq \sigma \nabla f(x)^T(x - x(s)).$$

Puede escribirse como.

$$\nabla f(x)^T(x - x(s)) + (\nabla f(\xi_s) - \nabla f(x))^T(x - x(s)) \geq \sigma \nabla f(x)^T(x - x(s)).$$

Entonces.

$$(1 - \sigma) \nabla f(x)^T(x - x(s)) \geq (\nabla f(\xi_s) - \nabla f(x))^T(x - x(s)). \quad (2.13)$$

Por la desigualdad (2.12) y la parte (i), tenemos que para todo $s \in (0, 1]$.

$$\nabla f(x)^T(x - x(s)) \geq \frac{\|x - x(s)\|^2}{s} \geq \|x - x(1)\| \|x - x(s)\|.$$

Por lo tanto, la desigualdad (2.13) es satisfecha para todo $s \in (0, 1]$ tal que.

$$(1 - \sigma) \|x - x(1)\| \geq (\nabla f(x) - \nabla f(\xi))^T \frac{x - x(s)}{\|x - x(s)\|}.$$

Así vemos que existe $s_x > 0$ tal que la desigualdad anterior, y por lo tanto las ecuaciones (2.13) y (2.10), son satisfecha para $s \in (0, s_x]$. \square

TEOREMA 2.1. *El algoritmo SPG1 está bien definido, y todo punto de acumulación de la sucesión $\{x_k\}$ que éste genera es un punto estacionario restringido.*

Demostración. Del lema (2.2(ii)), tenemos que para todo $\lambda \in [0, \min\{s_{x_k}, \alpha_{\min}\}]$ se cumple.

$$\begin{aligned} f(P(x_k - \lambda g(x_k))) - \max_{0 \leq j \leq M-1} f(x_{k-j}) &\leq f(P(x_k - \lambda g(x_k))) - f(x_k) \\ &\leq \gamma \langle g(x_k), g_\lambda(x_k) \rangle. \end{aligned}$$

Por lo tanto, un tamaño de paso que satisfaga (2.2) será encontrado después de un número finito de pruebas, y así el algoritmo SPG1 está bien definido.

Sea $\bar{x} \in \Omega$ un punto de acumulación de $\{x_k\}$, y reetiquetamos $\{x_k\}$ como una subsucesión convergiendo a \bar{x} . Consideraremos dos casos.

- **Caso 1.** Si $\inf \lambda_k = 0$, entonces existe una subsucesión $\{x_k\}_{K \in \mathbb{N}}$ tal que.

$$\lim_{k \in K} \lambda_k = 0.$$

En ese caso, de la forma en que λ_k es escogido en (2.2), existe un índice \bar{k} tal que para todo $k \geq \bar{k}$, $k \in K$, existe ρ_k , $0 < \sigma_1 \leq \rho_k \leq \sigma_2$, para el cual $\psi_k = \frac{\lambda_k}{\rho_k} > 0$ no satisface la condición (2.2), es decir.

$$\begin{aligned} f(P(x_k - \psi_k g(x_k))) &> \max_{0 \leq j \leq M-1} f(x_{k-j}) + \gamma \langle g(x_k), P(x_k - \psi_k g(x_k)) - x_k \rangle \\ &\geq f(x_k) + \gamma \langle g(x_k), P(x_k - \psi_k g(x_k)) - x_k \rangle. \end{aligned}$$

Por lo tanto, esto sigue que.

$$f(P(x_k - \psi_k g(x_k))) - f(x_k) > \gamma \langle g(x_k), g_{\psi_k}(x_k) \rangle. \quad (2.14)$$

Por el teorema del valor medio (1.3) obtenemos.

$$\begin{aligned} f(P(x_k - \psi_k g(x_k))) - f(x_k) &= g(\xi_k)^T (P(x_k - \psi_k g(x_k)) - x_k), \\ &= g(\xi_k)^T g_{\psi_k}(x_k), \\ &= g(\xi_k)^T g_{\psi_k}(x_k) + g(x_k)^T P(x_k - \psi_k g(x_k)), \\ &\quad - g(x_k)^T P(x_k - \psi_k g(x_k)) + g(x_k)^T x_k - g(x_k)^T x_k, \\ &= g^T(x_k) g_{\psi_k}(x_k) + (g(\xi_k) - g(x_k))^T g_{\psi_k}(x_k), \\ &= \langle g(x_k), g_{\psi_k}(x_k) \rangle + \langle g(\xi_k) - g(x_k), g_{\psi_k}(x_k) \rangle. \end{aligned}$$

En consecuencia tenemos que

$$f(P(x_k - \psi_k g(x_k))) - f(x_k) = \langle g(x_k), g_{\psi_k}(x_k) \rangle + \langle g(\xi_k) - g(x_k), g_{\psi_k}(x_k) \rangle. \quad (2.15)$$

Donde ξ_k pertenece al segmento de recta que conecta a x_k y $P(x_k - \psi_k g(x_k))$.

Combinando (2.14) y (2.15) obtenemos que para todo $k \in K$ suficientemente grande.

$$(1 - \gamma) \langle g(x_k), g_{\psi_k}(x_k) \rangle > \langle g(x_k) - g(\xi_k), g_{\psi_k}(x_k) \rangle. \quad (2.16)$$

Usando los lemas (2.1) y (2.2), tenemos.

$$\langle g(x_k), g_{\psi_k}(x_k) \rangle \leq \frac{-1}{\psi_k} \|g_{\psi_k}(x_k)\|_2^2 \leq \frac{-1}{\alpha_k} \|g_{\alpha_k}(x_k)\|_2 \|g_{\psi_k}(x_k)\|_2, \quad (2.17)$$

donde α_k es el tamaño de paso inicial en la iteración k . Combinando (2.16) y (2.17) y usando la desigualdad de Schwartz (1.1), obtenemos para $k \in K$ suficientemente grande.

$$\begin{aligned} \frac{(1 - \gamma)}{\alpha_k} \|g_{\alpha_k}(x_k)\|_2 \|g_{\psi_k}(x_k)\|_2 &< \langle g(\xi_k) - g(x_k), g_{\psi_k}(x_k) \rangle \\ &\leq \|g(\xi_k) - g(x_k)\|_2 \|g_{\psi_k}(x_k)\|_2. \end{aligned}$$

Usando que $\|g_{\psi_k}(x_k)\|_2 \neq 0$, tenemos.

$$\frac{(1 - \gamma)}{\alpha_k} \|g_{\alpha_k}(x_k)\|_2 < \|g(\xi_k) - g(x_k)\|_2. \quad (2.18)$$

Como $\psi_k \rightarrow 0$ y $x_k \rightarrow \bar{x}$ cuando $k \rightarrow \infty$, $k \in K$, entonces $\xi_k \rightarrow \bar{x}$ cuando $k \rightarrow \infty$, $k \in K$.

Tomando una subsucesión conveniente $\bar{K} \subseteq K$ tal que $\{\alpha_k\}$ es convergente a $\bar{\alpha} \in [\alpha_{\min}, \alpha_{\max}]$, y tomando límites en (2.18) cuando $k \rightarrow \infty$, $k \in \bar{K}$, deducimos que.

$$\|g_{\bar{\alpha}}(\bar{x})\|_2 \leq 0$$

Por lo tanto, $g_{\bar{\alpha}}(\bar{x}) = 0$, y \bar{x} es un punto estacionario estricto.

- **Caso 2.** Asumimos que $\inf \lambda_k \geq \rho > 0$. Supongamos por contradicción que \bar{x} no es un punto estacionario restringido. Por lo tanto $\|g_{\lambda}(\bar{x})\| > 0$ para todo $\lambda \in (0, \alpha_{\text{máx}}]$.

Por otra parte, sea $X = [\rho, \alpha_{\text{máx}}]$ el cual es un conjunto compacto.

Tenemos que $\|g_{\lambda}(\bar{x})\|$ es continua en X . Así por el teorema (1.2) tenemos que para cada λ , $\|g_{\lambda}(\bar{x})\| \in [a, b]$ para algún $a, b \in \mathbb{R}$

Pero como $\|g_{\lambda}(\bar{x})\| > 0$, entonces.

$$0 < a \leq \|g_{\lambda}(\bar{x})\| \leq b.$$

Haciendo $a = \delta$ obtenemos que,

$$\|g_{\lambda}(\bar{x})\| \geq \delta > 0 \quad \text{para todo } \lambda \in [\rho, \alpha_{\text{máx}}].$$

Ahora definamos la siguiente sucesión $\{f(x_{l(k)})\}$ dada por.

$$f(x_{l(k)}) = \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}),$$

donde $l(k)$ es un entero tal que $k - \min\{k, M-1\} \leq l(k) \leq k$, la cual es una sucesión monótonicamente no creciente (ver [7]). Por (2.2) sigue que, para $k > M-1$.

$$f(x_{l(k)}) \leq f(x_{l(l(k)-1)}) + \gamma \langle g(x_{l(k)-1}), g_{\lambda_{l(k)-1}}(x_{l(k)-1}) \rangle.$$

Por continuidad, para $k \geq \bar{k}$ suficientemente grande, tenemos que $\|g_\lambda(\bar{x}_k)\| \geq \frac{\delta}{2}$. En efecto.

Tomemos una subsucesión $\{\bar{x}_k\}$ que converge a \bar{x} y tenemos que la función $\|g_\lambda\|$ es continua en \bar{x} , esto implica que.

$$\lim_{k \rightarrow \infty} \|g_\lambda(\bar{x}_k)\| = \|g_\lambda(\bar{x})\|.$$

Esto quiere decir que, para todo $\epsilon > 0$ existe un $\bar{k} \in \mathbb{N}$ tal que para $k > \bar{k}$, $|\|g_\lambda(\bar{x}_k)\| - \|g_\lambda(\bar{x})\|| < \epsilon$.

Luego.

$$-\epsilon < \|g_\lambda(\bar{x}_k)\| - \|g_\lambda(\bar{x})\| < \epsilon.$$

Entonces.

$$-\epsilon + \|g_\lambda(\bar{x})\| < \|g_\lambda(\bar{x}_k)\| < \epsilon + \|g_\lambda(\bar{x})\|.$$

Haciendo $\epsilon = \frac{\|g_\lambda(\bar{x})\|}{2} > 0$. Obtenemos que.

$$\|g_\lambda(\bar{x}_k)\| > \frac{\|g_\lambda(\bar{x})\|}{2} \geq \frac{\delta}{2}.$$

Así, usando el lema (2.1), obtenemos.

$$f(x_{l(k)}) \leq f(x_{l(l(k)-1)}) - \frac{\gamma}{\alpha_{\text{máx}}} \|g_{\lambda_{l(k)-1}}(x_{l(k)-1})\|_2^2 \leq f(x_{l(l(k)-1)}) - \frac{\gamma\delta^2}{4\alpha_{\text{máx}}}.$$

Cuando $k \rightarrow \infty$, claramente $f(x_{l(k)}) \rightarrow -\infty$, lo cual es una contradicción. De hecho, la función objetivo f es una función continua y así $f(x_k)$ converge a $f(\bar{x})$.

□

TEOREMA 2.2. *El algoritmo SPG2 está bien definido, y todo punto de acumulación de la sucesión $\{x_k\}$ que esté genera es un punto estacionario restringido.*

Demostración. Supongamos que x_k no es un punto estacionario restringido, entonces por el lema (2.1).

$$\langle g(x_k), d_k \rangle = \langle g(x_k), g_{\alpha_k}(x_k) \rangle \leq -\frac{1}{\alpha_{\text{máx}}} \|g_{\alpha_{\text{máx}}}(x_k)\|_2^2 < 0,$$

y la dirección de búsqueda es una dirección de descenso. Así, un tamaño de paso satisfaciendo (2.6) será encontrado después de un número finito de pruebas, por lo tanto el algoritmo SPG2 esta bien definido.

Sea $\bar{x} \in \Omega$ un punto de acumulación de $\{x_k\}$. Y reetiquetamos $\{x_k\}$ una subsecuencia convergiendo a \bar{x} . Consideraremos dos casos.

- **Caso 1.** Asumimos que $\inf \lambda_k = 0$. Supongamos por reducción al absurdo, que \bar{x} no es estacionario. Por lo tanto por la definición (1.7) tenemos que.

$$\langle g(\bar{x}), P(\bar{x} - \alpha g(\bar{x})) - \bar{x} \rangle < 0.$$

Así.

$$\left\langle g(\bar{x}), \frac{P(\bar{x} - \alpha g(\bar{x})) - \bar{x}}{\|P(\bar{x} - \alpha g(\bar{x})) - \bar{x}\|} \right\rangle < 0.$$

Definamos.

$$h(\alpha) = \left\langle g(\bar{x}), \frac{P(\bar{x} - \alpha g(\bar{x})) - \bar{x}}{\|P(\bar{x} - \alpha g(\bar{x})) - \bar{x}\|} \right\rangle,$$

y $Y = [\alpha_{\text{mín}}, \alpha_{\text{máx}}]$ el cual es un conjunto compacto.

Tenemos que $h(\alpha)$ es continua en Y , luego por el teorema (1.2) $h(\alpha) \in [c, d]$ para $c, d \in \mathbb{R}$. Como $h(\alpha) < 0$, esto implica que.

$$c \leq h(\alpha) \leq d < 0.$$

Haciendo $d = -\delta$ con $\delta > 0$ tenemos que.

$$h(\alpha) \leq -\delta < 0 \quad \text{para todo } \alpha \in [\alpha_{\text{mín}}, \alpha_{\text{máx}}],$$

esto equivale a que.

$$\left\langle g(\bar{x}), \frac{P(\bar{x} - \alpha g(\bar{x})) - \bar{x}}{\|P(\bar{x} - \alpha g(\bar{x})) - \bar{x}\|} \right\rangle < -\delta \quad \text{para todo } \alpha \in [\alpha_{\min}, \alpha_{\max}].$$

Lo que implica que.

$$\left\langle g(x_k), \frac{P(x_k - \alpha g(x_k)) - x_k}{\|P(x_k - \alpha g(x_k)) - x_k\|} \right\rangle < -\frac{\delta}{2} \quad \text{para todo } \alpha \in [\alpha_{\min}, \alpha_{\max}]. \quad (2.19)$$

En efecto. tomemos una subsucesión $\{x_k\}$ que converge a \bar{x} y definamos.

$$w(x_k) = \left\langle g(x_k), \frac{P(x_k - \alpha g(x_k)) - x_k}{\|P(x_k - \alpha g(x_k)) - x_k\|} \right\rangle.$$

Dado que w es continua en \bar{x} , esto implica que.

$$\lim_{k \rightarrow \infty} w(x_k) = w(\bar{x})$$

Esto quiere decir que, para todo $\epsilon > 0$ existe $\bar{k} \in \mathbb{N}$ tal que para $k > \bar{k}$, $|w(x_k) - w(\bar{x})| < \epsilon$

Así

$$-\epsilon < w(x_k) - w(\bar{x}) < \epsilon$$

Luego

$$w(\bar{x}) - \epsilon < w(x_k) < w(\bar{x}) + \epsilon$$

Dado que $w(\bar{x}) = h(\alpha)$, entonces $w(\bar{x}) < -\delta < 0$.

Tomando $\epsilon = \frac{-w(\bar{x})}{2} > 0$ obtenemos que.

$$w(x_k) < \frac{w(\bar{x})}{2} < \frac{-\delta}{2} \quad \text{para todo } \alpha \in [\alpha_{\min}, \alpha_{\max}].$$

O lo que es equivalente.

$$\left\langle g(x_k), \frac{P(x_k - \alpha g(x_k)) - x_k}{\|P(x_k - \alpha g(x_k)) - x_k\|} \right\rangle < -\frac{\delta}{2} \quad \text{para todo } \alpha \in [\alpha_{\min}, \alpha_{\max}]. \quad (2.20)$$

Como $\inf \lambda_k = 0$, existe una subsucesión $\{x_k\}_K$ tal que.

$$\lim_{k \in K} \lambda_k = 0.$$

En ese caso, de la forma en que λ_k es escogido en (2.6), existe un índice \bar{k} tal que para todo $k \geq \bar{k}$, $k \in K$, existe ρ_k , $0 < \sigma_1 \leq \rho_k \leq \sigma_2$, para el cual $\lambda_k/\rho_k > 0$ no satisface la condición (2.6); es decir.

$$f\left(x_k + \frac{\lambda_k}{\rho_k} d_k\right) > \max_{0 \leq j \leq M-1} f(x_{k-j}) + \gamma \frac{\lambda_k}{\rho_k} \langle g(x_k), d_k \rangle \geq f(x_k) + \gamma \frac{\lambda_k}{\rho_k} \langle g(x_k), d_k \rangle.$$

Así

$$\frac{f\left(x_k + \frac{\lambda_k}{\rho_k} d_k\right) - f(x_k)}{\lambda_k/\rho_k} > \gamma \langle g(x_k), d_k \rangle.$$

Por el teorema del valor medio (1.3), esta relación puede escribirse como

$$\langle g(x_k + t_k d_k), d_k \rangle > \gamma \langle g(x_k), d_k \rangle \quad \text{para todo } k \in K, \quad k \geq \bar{k}, \quad (2.21)$$

donde t_k es un escalar en el intervalo $[0, \lambda_k/\rho_k]$ el cual va a cero cuando $k \in K$ va al infinito.

Tomando una subsucesión conveniente tal que $d_k/\|d_k\|$ es convergente a d , y tomando límites en (2.21), deducimos que $(1 - \gamma)\langle g(\bar{x}), d \rangle \geq 0$. (en efecto, observe que $\{d_k\}_K$ es acotada y así $t_k\|d_k\| \rightarrow 0$). Como $(1 - \gamma) > 0$ y $\langle g(x_k), d_k \rangle < 0$ para todo k , entonces $\langle g(\bar{x}), d \rangle = 0$.

Por continuidad y la definición de d_k , implica que para k suficientemente grande en esa subsucesión tenemos que.

$$\left\langle g(x_k), \frac{P(x_k - \alpha_k g(x_k)) - x_k}{\|P(x_k - \alpha_k g(x_k)) - x_k\|} \right\rangle > -\delta/2$$

Lo que contradice (2.20). Por lo tanto \bar{x} es un punto estacionario restringido.

- **Caso 2.** Asumimos que $\inf \lambda_k \geq \rho > 0$. Supongamos por reducción al absurdo que \bar{x} no es punto estacionario restringido. Por lo tanto $\|g_{\lambda_k}(\bar{x})\| > 0$ para todo

$\lambda \in (0, \alpha_{\text{máx}}]$. Por continuidad y compacidad (ver teorema 1.2), existe $\delta > 0$ tal que $\|g_\lambda(\bar{x})\| \geq \delta > 0$ para todo $\lambda \in [\rho, \alpha_{\text{máx}}]$, (se usa el mismo mecanismo del caso (ii) del teorema anterior).

Como en la prueba del segundo caso del teorema (2.1).

$$f(x_{l(k)}) = \max_{0 \leq j \leq \min\{k, m-1\}} f(x_{k-j}),$$

es una sucesión monotonicamente no creciente. De la desigualdad (2.6) sigue que, para $k > M - 1$.

$$f(x_{l(k)}) \leq f(x_{l(l(k)-1)}) + \gamma \lambda_{l(k)-1} \langle g(x_{l(k)-1}), g_{\alpha_{l(k)-1}}(x_{l(k)-1}) \rangle.$$

Por continuidad, para $k \geq \bar{k}$ suficientemente grande, usando el mismo proceso que en el teorema anterior tenemos que, $\|g_{\alpha_k}(\bar{x}_k)\| \geq \delta/2$. Así, usando el lema (2.1), obtenemos.

$$f(x_{l(k)}) \leq f(x_{l(l(k)-1)}) - \frac{\gamma \rho}{\alpha_{\text{máx}}} \|g_{\alpha_{l(k)-1}}(x_{l(k)-1})\|_2^2 \leq f(x_{l(l(k)-1)}) - \frac{\gamma \delta^2 \rho}{4\alpha_{\text{máx}}}$$

Cuando $k \rightarrow \infty$, claramente $f(x_{l(k)}) \rightarrow -\infty$ que es una contradicción. De hecho, f es una función continua y así $f(x_k)$ converge a $f(\bar{x})$.

□

§2.3. Análisis de Resultados Numéricos

En esta sección hacemos un resumen de las principales conclusiones obtenidas referente a los resultados numéricos correspondiente a la aplicación de los algoritmos SPG1 y SPG2 en comparación con el paquete LANCELOT y una aplicación del método del gradiente proyectado clásico, todo desarrollado y mostrado de la referencia principal [3].

Notemos que los algoritmos SPG1 y SPG2 calculan al menos una proyección sobre el conjunto factible Ω por iteración. Así, estos algoritmos son especialmente interesantes en el caso en cual esta proyección es fácil de calcular como es el caso

donde el conjunto Ω es una caja n-dimensional, posiblemente con algunas cotas infinitas. Como por ejemplo, los buenos algoritmos para la minimización restringida de caja son la herramienta para el desarrollo de los métodos eficaces del lagrangiano aumentado para programación no lineal en general. A partir de esto, implementamos los algoritmos del gradiente proyectado espectral para el cual Ω es descrito por cotas sobre las variables.

Para evaluar la formalidad de los algoritmos del método del gradiente proyectado espectral. En el trabajo desarrollado por Birgin, Martinez y Raydan [3] se analizaron resultados numéricos en el cual comparan los algoritmos de método principal con el buen conocido paquete LANCELOT (ver [6]) usando los problemas restringidos acotados con mas de 50 variables de la colección CUTE (ver [4]).

Para los algoritmos SPG1 y SPG2 usaron $\gamma = 10^{-4}$, $\alpha_{\text{mín}} = 10^{-30}$, $\alpha_{\text{máx}} = 10^{30}$, $\sigma_1 = 0,1$, $\sigma_2 = 0,9$, y $\alpha_0 = \frac{1}{\|g_1(x_0)\|}$, luego corrieron una serie de problemas con $M \in \{5, 10, 15\}$, decidieron usar $M = 10$ ya que las pruebas no mostraron diferencias significativas. Decidieron que los algoritmos pararan cuando $\|g_1(x)\|_\infty \leq 10^{-5}$ declarando convergencia. También la ejecución de los algoritmos para cuando 50000 iteraciones y 200000 evaluaciones de la función fueran completadas sin conseguir la convergencia.

Para completar la visión numérica del comportamiento de los algoritmos, también corrieron el algoritmo del gradiente proyectado (PGA), el cual es idéntico al SPG1, con la elección inicial del tamaño de paso $\alpha_k = 1$.

En este caso implementaron ambos, la versión monótona y no monótona del PGA, el cual corresponde a $M = 1$, y la única no monótona con $M = 10$.

La convergencia de la versión no monótona es un caso particular del teorema (2.1).

Comparando los resultados de los algoritmos del método principal con LANCELOT y PGA en una cantidad de 50 problemas (ver [3]) notaron que cuatro de esos problemas pueden ser considerados fallas de los métodos SPG. Ya que no convergen a

un punto estacionario después de 50000 iteraciones.

En 40 problemas, LANCELOT, SPG1 y SPG2 encuentran una misma solución. En términos de tiempo de calculo, SPG1 fue mas rápido que LANCELOT en 29 problemas y SPG2 supera a LANCELOT también en 29 problemas. No existen diferencias significativas entre la ejecución de los métodos SPG1 y SPG2.

Con la exclusión de los problemas donde la diferencia en tiempos de CPU era inferior a 10 %, el algoritmo SPG1 venció al paquete LANCELOT 28-9 y el SPG2 venció al LANCELOT 28-11. según los resultados numéricos extraídos de la referencia principal [3].

Ahora excluyendo, de los 40 problemas, aquellos donde los 3 algoritmos convergieron en menos de 1 segundo, nos quedamos con 31 problemas. Considerando este conjunto de problemas, el algoritmo SPG1 venció al LANCELOT 20-11 (o 19-9 si excluimos, de nuevo, las diferencias de menos de 10 %) y SPG2 venció a LANCELOT 20-11 (O 19-11).

En la referencia principal ([3]) también se implemento el algoritmo del gradiente proyectado PGA, usando la misma estructura de los algoritmos SPG1 y SPG2 en términos de sistemas de interpolación, ambos con la estrategia monótona y no monótona. La ejecución de ambas alternativas es deficiente, en comparación con los algoritmos SPG1 y SPG2 y otros minimizadores de restricciones de caja. Esto confirma que la elección espectral del tamaño de paso es la característica que coloca eficiencia en la metodología del gradiente proyectado.

CAPÍTULO 3

CONCLUSIÓN

En esta monografía se presento un estudio sobre el método del gradiente proyectado espectral no monótono, el cual es considerado como la combinación del esquema de búsqueda lineal no monótona desarrollada por Grippo, Lampariello y Lucidi (ver [7]) con el tamaño de paso espectral introducido por Barzilai y Borwein (ver [1]) y analizado por Raydan (ver [12]). No hay originalidad en cuanto a nuevos resultados, sin embargo se presento detalladamente las demostraciones de los teoremas citados y una serie de gráficos que ayudan en la comprensión de la metodología estudiada.

El método del gradiente proyectado espectral es aplicado para resolver problemas de programación no lineal en conjuntos convexos restringidos, este método ayuda a acelerar la velocidad de convergencia del método del gradiente proyectado y requiere menos trabajo computacional. Además se estudiaron dos versiones de la metodología llamados SPG1 y SPG2, los cuales son algoritmos del método del gradiente proyectado espectral (ver [3]). Estos algoritmos son más eficaces en el caso en que la proyección es fácil de calcular debido a que estos calculan al menos una proyección ortogonal por iteración.

En la referencia principal (ver[3]) realizan un estudio de los resultados numéricos de la aplicación de los algoritmos SPG1 y SPG2, donde comparan los algoritmos del método del gradiente proyectado espectral con el paquete LANCELOT (ver[6]) y con una aplicación en el método del gradiente proyectado (ver [2]), ellos son comparados en un número de 50 problemas de la colección CUTE (ver [4]).

Según estos resultados observamos que en una minoría de problemas los algoritmos SPG1 y SPG2 divergen, y en más de la mitad de los problemas los algoritmos del método del gradiente proyectado espectral (ver [3]) son mas eficaces que el paque-

te LANCELOT (ver [6]). Por otra parte en comparación con método del gradiente proyectado (ver [2]), los algoritmos SPG1 y SPG2 tienen mayor efectividad. Con esto se verifica que la elección del tamaño de paso espectral (ver [12] y [1]) es la herramienta fundamental que ayuda a mejorar la convergencia del método del gradiente proyectado.

REFERENCIAS

- [1] J. BARZILAI & J. BORWEIN *Two-Point Step Size Gradient Methods*, 1986.
- [2] DIMITRI P. BERTSEKAS *Nonlinear Programming* Athena Scientific, Belmont, Massachusetts, 1995.
- [3] ERNESTO G. BIRGIN, JOSÉ MARTÍNEZ & MARCOS RAYDAN *Nonmonotone Spectral Projected Gradient Methods On Convex Sets*. 1997.
- [4] I. BONGARTZ, A. R. CONN, NICK GOULD & PH. L. TOINT *CUTE: CONSTRAINED AND UNCONSTRAINED TESTING ENVIROMENT* Octubre 1993.
- [5] R. L. BURDEN & J. D. FAIRES *Análisis numérico* septima edición, Grupo Editorial Iberoamérica, Traduccion de Numerical Analysis, 1985.
- [6] A. R. CONN, N. I. M. GOULD & PH. L. TOINT *LANCELOT: a Fortran package for large - scale nonlinear optimization (Release A)* Springer Series in Computational Mathematics, vol 17. Springer Verlag (Heidelberg, New York), 1992b.
- [7] L. GRIPPO, F. LAMPARIELLO & S. LUCIDI *A Nonmonotone line search technique For Newton's Method*, 1985.
- [8] J. B. HIRIART - URRUTY & C. LEMAREDAL *Conver Analysis and Minimization algorithm I* Springer Verlag, New York, 1996.
- [9] DAVID G. LUENBERGER *Linear and Nonlinear Programming* Stanford University, Stanford California, 1984.
- [10] GEORGE NAKOS & DAVID JOYNER *Algebra Lineal con Aplicaciones* International Thomson Editores, 1999.
- [11] JORGE NOCEDAL & STEPHEN J. WRIGHT *Numerical Optimization* Springer-Verlag New York, 1999.

- [12] MARCOS RAYDAN *On the Barzilai and Borwein choice of steplength for the gradient method*, 1991.

- [13] WALTER RUDIN *Principios de Análisis Matemático* segunda edición, Libros McGRAW-HILL de Mexico, 1996.