BENEMÉRITA UNIVERSIDAD
AUTÓNOMA DE PUEBLA

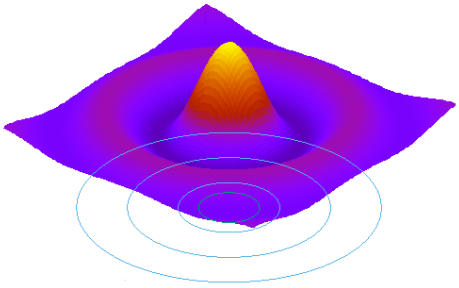
FACULTAD DE CIENCIAS FÍSICO-MATEMÁTICAS

TESIS
PARA OBTENER EL TÍTULO DE:
LICENCIADO EN MATEMÁTICAS APLICADAS

Implementación De Los Métodos Cuasi-Newton

Autor:
Acevedo Vázquez Julio Andrés

Director de tesis:
Dr. Guillermo López Mayo



PUEBLA PUE., DICIEMBRE 2019

Dedicado a
mi familia
amigos y
profesores.

Agradecimientos

Quiero agradecer a todas las personas que siempre estuvieron apoyándome, no solo para la realización de este trabajo, también a lo largo de mi carrera y mi vida, dándome consejos y apoyo cuando más lo necesitaba.

Hago mención especial a mis padres Roman Acevedo Pichardo y Elizabeth Vázquez Ortiz, quienes siempre han hecho todo lo posible y hasta lo imposible para que tanto mis hermanos como yo podamos salir adelante, de todo corazón, ¡gracias!. A mis hermanos Alejandro y Andrea, quienes siempre han estado a mi lado dándome consejos y sacándome una sonrisa.

A mis amigos de la facultad Gustavo, Juan, Roque y Edgar con quienes compartí muy gratos momentos durante la universidad y siempre estuvieron dispuestos a apoyarme. También a mis demás compañeros, quienes aportaron su granito de arena para mi crecimiento tanto académico como personal.

A mi novia América, quien siempre ha estado para mi cuando la necesito y me ha motivado para seguir adelante, ¡Muchas gracias por todo amor! No sé cuales serán los caminos que tomaremos en el futuro, pero siempre tendrás un lugar muy especial en mi corazón.

A mi asesor de tesis, el Dr. Guillermo López Mayo, quien siempre tuvo la mejor disposición y paciencia para explicarme todo lo necesario para la realización de esta tesis.

A mis sinodales, el Dr. José Jacobo Oliveros Oliveros quien siempre me apoyó con la mejor disposición en los temas que no entendía y cuyas platicas son muy amenas, el M.C. Sergio Adán Juárez, por ayudarme en la implementación de los métodos y quien siempre que le pedía ayuda me apoyaba de la manera más atenta y a la Dra, Lourdes Sandoval quien sin conocerme me ayudó y brindó su conocimiento para que el trabajo se concluyera de forma exitosa.

A todos los profesores de la facultad. De cada uno me llevo conocimiento muy valioso que me servirá por el resto de mi vida. En especial a la Dra. Hortensia Josefina Reyes Cervantes, quien no solo fue mi tutora académica, también me apoyo en el difícil proceso de adaptación a la facultad. Al Dr. Carlos Alberto López Andrade quien de manera desinteresada me ayudó en la comprensión de algunas demostraciones que en su momento para mi eran bastante difíciles de entender.

Índice general

1. Conceptos previos **3**
- 1.1. Máximos y Mínimos . . . 3
- 1.2. Métodos de Búsqueda en la Línea . . . 6
- 1.3. Condiciones de Wolfe . . . 7
- 1.4. Tasas de Convergencia . . . 8
- 1.5. Método de Newton . . . 8

2. El método DFP **11**
- 2.1. Construcción del método . . . 11
- 2.2. Propiedades del método . . . 15

3. El método BFGS **19**
- 3.1. Construcción del método . . . 19
- 3.2. Algoritmo BFGS . . . 23

4. Análisis de Convergencia **25**
- 4.1. Convergencia Global del Método BFGS . . . 25
- 4.2. Convergencia Superlineal del Método BFGS . . . 28

5. Resultados de la implementación **33**

6. Conclusiones **37**

A. Conceptos de Cálculo **39**

B. Conceptos de Álgebra **41**

C. Convexidad **45**

D. La notación de la O grande. **49**

Introducción

En optimización, el método de Newton es un método iterativo de búsqueda en la línea, es decir $x_{k+1} = x_k + \alpha_k p_k$, donde $\alpha_k \in \mathbb{R}$ y $p_k \in \mathbb{R}^n$ son llamados longitud de paso y dirección de búsqueda, respectivamente. Dada una función $f : \mathbb{R}^n \to \mathbb{R}$, este método consiste en encontrar un punto x_* donde la función f alcanza su valor mínimo, lo cual se logra mediante la aproximación cuadrática en el punto x_k,

$$m(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p,$$

donde $\nabla f(x_k)$ es el gradiente y $\nabla^2 f(x_k)$ la matriz hessiana evaluados en el punto x_k. Minimizando la función m, se obtiene que el punto p_k donde alcanza el valor mínimo esta dado por

$$p_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k).$$

El p_k obtenido, es utilizado en la actualización de x_{k+1}.

Lo que caracteriza a los distintos métodos de búsqueda en la línea es la elección de la dirección de paso p_k, por ejemplo, en el *método de descenso más rápido (también llamado el método de descenso más pronunciado)* se considera $p_k = -\nabla f(x_k)$. Además, la velocidad con la que los métodos de búsqueda en la línea convergen a la solución, depende de la dirección y la longitud de paso.

Otra clase de métodos a considerar es en los cuales la dirección es considerada como $p_k = -H_k \nabla f(x_k)$, donde H_k es una matriz definida positiva que aproxima a la inversa de la matriz hessiana. Ésta clase es a menudo llamada *métodos Cuasi-Newton.* Uno de los primeros métodos de minimización de una función no lineal usando éste enfoque, es el de Davidon en 1959, el cual fue simplificado y reformulado por Fletcher y Powell en 1963 y es referido como el *método de variable métrica.* Una generalización muy útil del método de *Davidon-Fletcher-Powell (DFP)* fue el propuesto por Broyden en 1967. Esencialmente, Broyden introduce un grado de libertad en la actualización de la matriz H_k . Una elección particular de éste grado de libertad fue propuesto por Broyden, Fletcher, Goldfarb y Shanno en 1970. Esto dio lugar al bien conocido método $BFGS$. Gill et al, en 1972, demostraron que ésta modificación se ejecuta más eficientemente que el método original en una mayor cantidad de problemas.

En 1972, Powell demostró que el método DFP converge a una solución óptima si la función objetivo es convexa y tiene segundas derivadas continuas, usando una línea de búsqueda exacta. Bajo suposiciones más fuertes en 1971 el mismo Powell ya había demostrado que el método converge superlinealmente. En 1973, Broyden et al., dieron resultados de convergencia donde el tamaño de paso es fijado como 1 y bajo ciertas condiciones se provee

convergencia superlineal. Bajo suposiciones adecuadas, Powell en 1976 demostró que una versión del método de variable métrica con líneas de búsqueda inexactas converge a una solución óptima si la función es convexa. Además, demostró que si la matriz Hessiana es definida positiva en el punto solución, la tasa de convergencia es superlineal.

Hoy en día los métodos Cuasi-Newton son de gran importancia para la resolución de problemas de programación no lineal, dado que el costo computacional comparado con el método de Newton, el cual tiene una tasa de convergencia cuadrática, es mucho menor en problemas donde la matriz es muy grande. El presente trabajo se enfocó en el método DFP y el BFGS, ya que el primero fue el primer método Cuasi-Newton obtenido y el segundo es el más robusto.

Los métodos de variable métrica discutidos anteriormente actualizan la matriz H_k sumándole dos matrices de rango 1, por lo que ésta clase es también llamada como *procedimiento de corrección de rango 2*.

El trabajo tiene como objetivo comprobar si los métodos Cuasi-Newton efectivamente convergen a la solución del problema de tal forma que el número de iteraciones no sea mucho mayor que el que alcanza el método de Newton, así como ver cual método de los que se explicarán obtiene una mejor aproximación de la solución. Para esto, el trabajo se dividió en 6 capítulos, los cuales contienen la siguiente información:

1. **Conceptos Previos.** Se presentan las definiciones y teoremas básicos para una mayor comprensión de los métodos, así como una explicación del método de Newton, el cual es de los más rápidos conocidos hasta el momento.

2. **El método DFP.** Se explica como se hizo la construcción del método así como algunas propiedades que cumple, tanto para funciones cuadráticas con línea de búsqueda exacta como para funciones en general.

3. **El método BFGS.** Se presentan la construcción del método y el algoritmo del mismo para su implementación computacional.

4. **Análisis de Convergencia.** Daremos los teoremas que aseguran la convergencia superlineal y global del método BFGS. Nos enfocamos solamente en este método ya que hasta el momento es el método Cuasi-Newton más robusto que existe.

5. **Implementación.** Se muestran los resultados de aplicar los métodos BFGS, DFP y Newton a dos funciones particulares, donde se muestra que en algunos casos es mejor el método BFGS y en otros el método de Newton.

6. **Conclusiones.** Se mencionan las ventajas de los métodos Cuasi-Newton, en especial el BFGS, sobre el método de Newton. Por tanto, se dan razones por las cuales en ocasiones es conveniente hacer uso de los métodos Cuasi-Newton.

CAPÍTULO 1

CONCEPTOS PREVIOS

El objetivo de la optimización es encontrar los puntos máximos o mínimos de una función dada con valores reales. Para ello, es necesario dar una definición precisa de lo que es un punto mínimo y un punto máximo.

1.1. Máximos y Mínimos

Definición 1.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función.*

1. *Un punto x_* es llamado un mínimo global si $f(x_*) \leq f(x)$, para todo $x \in \mathbb{R}^n$.*
2. *Un punto x_* es llamado un mínimo local si existe una vecindad U de x_* tal que $f(x_*) \leq f(x)$, para todo $x \in U$.*
3. *Un punto x_* es llamado un mínimo local estricto (también llamado un mínimo local fuerte) si existe una vecindad U de x_* tal que $f(x_*) < f(x)$, para todo $x \in U$, con $x \neq x_*$.*
4. *Un punto x_* es llamado un mínimo local aislado si existe una vecindad U de x_* tal que x_* es el único mínimo local en U.*
5. *Un punto x^* es llamado un máximo global si $f(x^*) \geq f(x)$, para toda $x \in \mathbb{R}^n$.*
6. *Un punto x^* es llamado un máximo local si existe una vecindad U de x^* tal que $f(x^*) \geq f(x)$, para todo $x \in U$.*
7. *Un punto x^* es llamado un máximo local estricto (también llamado un máximo local fuerte) si existe una vecindad U de x^* tal que $f(x^*) > f(x)$, para todo $x \in U$, con $x \neq x^*$.*
8. *Un vector p es llamado una dirección de descenso de la función $f : \mathbb{R}^n \to \mathbb{R}$ en el punto x_k si se cumple que $p^T \nabla f(x_k) < 0$.*

Matemáticamente hablando, la optimización es la minimización o maximización de una función, sujeto a restricciones en sus variables. Es decir, tenemos que resolver el siguiente problema:

$$\min_{x \in \mathbb{R}^n} f(x); \tag{1.1a}$$

$$\begin{aligned} sujeto\ a \quad & c_i(x) = 0, \quad i \in \mathscr{E} \\ & c_i(x) \geq 0, \quad i \in \mathscr{I}, \end{aligned}$$

donde $\mathscr{I}$ y $\mathscr{E}$ son conjuntos de índices de desigualdad e igualdad, respectivamente, y decimos que:

- x es el vector de variables, también llamado *incógnitas* o *parámetros*.
- f es la función objetivo, una función escalar de x, la cual se desea minimizar o maximizar.
- c_i son funciones restricción, las cuales son funciones escalares de x, que definen ciertas ecuaciones o desigualdades que la incógnita x debe satisfacer.

Los problemas con la forma general (1.1a) pueden ser clasificados de acuerdo a la naturaleza de la función objetivo y restricciones (lineal, no lineal, convexa,...), el número de variables, la suavidad de las funciones (diferenciable o no diferenciable), etcétera. Una importante distinción es entre los problemas que tiene restricciones y aquellos que no los tienen.

- Problemas de *optimización sin restricciones*, para los cuales $\mathscr{E} = \mathscr{I} = \phi$ en (1.1a), surgen directamente en muchas aplicaciones prácticas.
- Problemas de *optimización con restricciones* surgen de modelos en los cuales las restricciones juegan un papel fundamental. Estas restricciones pueden ser simples cotas, tales como $0 \leq x_1 \leq 100$, restricciones lineales más generales tales como $\sum_i x_i \leq 1$, o desigualdades no lineales que representan relaciones complejas entre las variables.

Cuando la función objetivo y todas las restricciones son funciones lineales de x, decimos que (1.1a) es un problema de *programación lineal*. Los problemas de *programación no lineal*, en los cuales al menos una restricción o la función objetivo son funciones no lineales, tienden a surgir naturalmente en las ciencias físicas e ingenierías.

Durante el presente trabajo se hará uso de la definición de *espacio normado*, ya que como se verá más adelante, las matrices cumplen la siguiente definición.

Definición 1.2. *Sea V un espacio vectorial sobre el campo $\mathbb{R}$, una **norma** es una transformación $\|\cdot\| : V \to \mathbb{R}$ tal que cumple lo siguiente:*

1. $\|x\| \geq 0$, *para todo* $x \in V$.
2. $\|x\| = 0$ *si, y sólo si,* $x = 0$.
3. *Si* $\alpha \in \mathbb{R}$ *y* $x \in V$, *entonces* $\|\alpha x\| = |\alpha| \|v\|$.
4. $\|x + y\| \leq \|x\| + \|y\|$, *para todo* $x, y \in V$.

Si existe tal función, a V se le llamará espacio normado.

También es común que el problema consista en maximizar la función objetivo. Transformar el problema de minimización a esta clase de problemas es tarea sencilla, si se toma en cuenta que

$$\max f(x) = -\min[-f(x)],$$

$$\text{mín } f(x) = -\text{máx}[-f(x)].$$

Para verificar que un punto es un mínimo de la función dada, se tienen los siguientes teoremas:

Teorema 1.1 (**Condiciones Necesarias de Primer Orden**). *Si x_* es un mínimo local y f es continuamente diferenciable en una vecindad abierta de x_*, entonces $\nabla f(x_*) = 0$.*

Demostración. Supongamos que $\nabla f(x_*) \neq 0$. Definimos el vector $p = -\nabla f(x_*)$ y notemos que $p^T \nabla f(x_*) = \|\nabla f(x_*)\|^2 < 0$. Dado que ∇f es continua en una vecindad de x_*, existe un escalar $T > 0$ tal que

$$p^T \nabla f(x_* + tp) < 0, \quad para\ todo\ t \in [0, T].$$

Por el teorema de Taylor, para cualquier $\bar{t} \in (0, T]$ tenemos que

$$f(x_* + \bar{t}p) = f(x_*) + \bar{t} p^T \nabla f(x_* + tp), \quad para\ algún\ t \in (0, \bar{t}).$$

Por lo tanto, $f(x_* + \bar{t}p) < f(x_*)$ para todo $\bar{t} \in (0, T]$, así x_* no es un minimizador de f, lo cual es una contradicción. □

Teorema 1.2 (**Condiciones Necesarias de Segundo Orden**). *Si x_* es un mínimo local y $\nabla^2 f$ es continua en una vecindad abierta de x_*, entonces $\nabla f(x_*) = 0$ y $\nabla^2 f(x_*)$ es semidefinida positiva.*

Demostración. Por las condiciones necesarias de primer orden sabemos que $\nabla f(x_*) = 0$. Supongamos por contradicción que $\nabla^2 f(x_*)$ no es definida positiva, así podemos escoger un vector p tal que $p^T \nabla^2 f(x_*) p < 0$ y dado que $\nabla^2 f$ es continua en una vecindad de x_*, entonces existe un escalar $T > 0$ tal que $p^T \nabla^2 f(x_* + tp) p < 0$ para todo $t \in [0, T]$.

Consideramos la expansión en serie de Taylor alrededor de x_*, tenemos que para todo $\bar{t} \in (0, T]$ y algún $t \in (0, \bar{t})$ que

$$f(x_* + \bar{t}p) = f(x_*) + \bar{t} p^T \nabla f(x_*) + \frac{1}{2} \bar{t}^2 p^T \nabla^2 f(x_* + tp) p < f(x_*).$$

Por lo que x_* no es un minimizador de f, lo cual es una contradicción. □

Teorema 1.3 (**Condiciones Suficientes de Segundo Orden**). *Suponga que $\nabla^2 f$ es continua en una vecindad abierta de x_*, que $\nabla f(x_*) = 0$ y que $\nabla^2 f(x_*)$ es definida positiva. Entonces x_* es un mínimo local estricto de f.*

Demostración. Dado que $\nabla^2 f$ es definida positiva en x_*, podemos escoger un radio $r > 0$ tal que $\nabla^2 f$ es defina positiva para todo x en la bola abierta $\mathscr{D} = \{z \in \mathbb{R}^n : \|z - x_*\| < r\}$. Consideremos un vector $p \neq 0$ tal que $\|p\| < r$, entonces tenemos que $(x_* + p) \in \mathscr{D}$ y así

$$\begin{aligned} f(x_* + p) &= f(x_*) + p^T \nabla f(x_*) + \frac{1}{2} p^T \nabla^2 f(z) p \\ &= f(x_*) + \frac{1}{2} p^T \nabla^2 f(z) p, \end{aligned}$$

donde $z = x_* + tp$ para algún $t \in (0, 1)$. Dado que $z \in \mathscr{D}$, tenemos que $p^T \nabla^2 f(z) p > 0$ y por lo tanto $f(x_* + p) > f(x_*)$, de donde se obtiene que x_* es un mínimo local estricto. □

Teorema 1.4. *Si f es una función convexa, entonces cualquier mínimo local x_* es un mínimo global de f. Sí además f es diferenciable, entonces todo punto estacionario es un mínimo global de f.*

Demostración. Supongamos que x_* es un minimizador local pero no global. Entonces podemos encontrar un $z \in \mathbb{R}^n$ tal que $f(z) < f(x_*)$. Consideremos el segmento de línea que une a los puntos z y x_*, esto es

$$x = \lambda z + (1-\lambda)x_*, \quad para\ algún\ \lambda \in (0.1]. \tag{1.3}$$

Dado que la función f es convexa, se obtiene que

$$f(x) \leq \lambda f(z) + (1-\lambda) f(x_*) < f(x_*). \tag{1.4}$$

Cualquier vecindad $\mathcal{N}$ de x_* contiene una pieza del segmento de linea (1.3), así que siempre habrá puntos $x \in \mathcal{N}$ para los cuales (1.4) se satisface. Por lo tanto, x_* no es un mínimo local, lo cual es una contradicción.

Para la segunda parte del teorema, supongamos que x_* no es un minimizador global y escogemos z como en la parte de arriba. Entonces, por la convexidad de f, tenemos que

$$\begin{aligned}
\nabla f(x_*)^T (z - x_*) &= \frac{d}{d\lambda} f(x_* + \lambda(z - x_*))|_{\lambda=0} \\
&= \lim_{\lambda \to 0} \frac{f(x_* + \lambda(z - x_*)) - f(x_*)}{\lambda} \\
&\leq \lim_{\lambda \to 0} \frac{\lambda f(z) + (1-\lambda) f(x_*) - f(x_*)}{\lambda} \\
&= f(z) - f(x_*) < 0.
\end{aligned}$$

Por lo tanto, $\nabla f(x_*) \neq 0$ y así x_* no es un punto estacionario. □

Los métodos empleados en la optimización son iterativos, es decir, se da una aproximación de las solución y se genera una sucesión de estimaciones hasta que se llega a una solución.

1.2. Métodos de Búsqueda en la Línea

Cada iteración de un método de búsqueda en la línea calcula una dirección de búsqueda p_k y decide que tan lejos se moverá a lo largo de esta dirección. La iteración está dada por $x_{k+1} = x_k + \alpha_k p_k$, donde el escalar positivo α_k es llamado *longitud de paso*. El éxito de los métodos de búsqueda en la línea depende de la efectividad de la elección, tanto de p_k como de la longitud α_k.

La mayoría de los métodos de búsqueda en la línea requieren que p_k sea una dirección de descenso, dado que esta propiedad garantiza que la función f puede ser reducida a lo largo de ésta dirección. Además, la dirección de búsqueda con frecuencia tiene la forma

$$p_k = -B_k^{-1} \nabla f(x_k), \tag{1.5}$$

donde B_k es una matriz simétrica y no-singular. Cuando B_k es una matriz definida positiva, tenemos que

$$p_k^T \nabla f(x_k) = -\nabla f(x_k)^T B_k^{-1} \nabla f(x_k) < 0.$$

Por lo tanto p_k es una dirección de descenso. En el método de descenso más rápido, la matriz $B_k = I$, para todo $k \in \mathbb{N}$, mientras que en el método de Newton $B_k = \nabla^2 f(x_k)$. En los métodos Cuasi-Newton, B_k es una aproximación de Hessiano que es actualizada en cada iteración.

1.3. Condiciones de Wolfe

Lo que se busca para la longitud de paso en cada iteración de un método iterativo de búsqueda en la línea, $x_{k+1} = x_k + \alpha_k p_k$, es que proporcione una reducción substancial de la función objetivo f, pero a la vez, que el tiempo para encontrar tal longitud no sea muy grande. La elección ideal sería la solución del subproblema

$$\min_{\alpha>0} \phi(\alpha) = f(x_k + \alpha p_k). \tag{1.6}$$

En general es muy costoso identificar la solución, por lo cuál, usualmente se usan líneas de búsqueda inexactas para identificar una longitud de paso que alcance una máxima reducción en f con un costo mínimo. Una estrategia que nos permite hacer lo anterior, es utilizar las **condiciones de Wolfe**, las cuales se explican brevemente. Una condición que nos asegura un descenso suficiente de la función objetivo f, está dada por la siguiente desigualdad, la cual es conocida como la *condición de Armijo*:

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k, \tag{1.7}$$

para algún $c_1 \in (0,1)$. Pero la condición de Armijo no es suficiente para asegurar que el algoritmo haga un progreso suficiente, ya que la desigualdad se hará válida para valores de α suficientemente pequeños, por lo que es necesaria una segunda desigualdad, que evite pasos inaceptablemente cortos, la cual es conocida como la *condición de curvatura*:

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k, \tag{1.8}$$

para alguna constante $c_2 \in (c_1, 1)$, donde c_1 es la constante de (1.7).

Las condiciones de Armijo y de curvatura, en conjunto, son conocidas como las **condiciones de Wolfe**. Así, las condiciones de Wolfe se escriben como,

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k, \tag{1.9a}$$

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k, \tag{1.9b}$$

con $0 < c_1 < c_2 < 1$.

Una longitud de paso puede satisfacer las condiciones de Wolfe sin estar cerca de la solución de (1.6), sin embargo se puede modificar la condición de curvatura para forzar a que α_k pertenezca a una vecindad del minimizador o un punto estacionario de ϕ. Las **condiciones fuertes de Wolfe** requieren que α_k satisfaga

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k, \tag{1.10a}$$

$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \leq c_2 |\nabla f_k^T p_k|, \tag{1.10b}$$

con $0 < c_1 < c_2 < 1$.

La única diferencia con las condiciones de Wolfe, es que no permitimos que $\phi'(\alpha_k)$ sea muy grande. Por lo tanto, excluimos los puntos que están lejos de los puntos estacionarios de ϕ. En [6] se demuestra que para cada función suave y acotada inferiormente, existen longitudes de paso que satisfacen las condiciones de Wolfe.

1.4. Tasas de Convergencia

La tasa de convergencia de una sucesión $\{x_k\}$ de puntos en $\mathbb{R}^n$, que convergen a un punto x_*, es la velocidad con la que la sucesión se acerca al punto. Las principales tasas de convergencia en los métodos de optimización son los siguientes.

Definición 1.3. *Sea $\{x_k\}$ una sucesión en $\mathbb{R}^n$ que converge a x_*.*

1. *Decimos que la convergencia es $Q-lineal$, si existen $r \in (0,1)$ y $N \in \mathbb{N}$ tales que*

$$\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \leq r, \ para\ todo\ k \geq N.$$

2. *Decimos que la convergencia es $Q-superlineal$, si*

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0.$$

3. *Decimos que la convergencia es $Q-cuadrática$, si existen $M > 0$ y $N \in \mathbb{N}$ tales que*

$$\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|^2} \leq M, \ para\ todo\ k \geq N.$$

1.5. Método de Newton

Consideremos ahora la iteración de Newton, donde la dirección de búsqueda está dada por

$$p_k^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k). \tag{1.11}$$

Aquí discutiremos solo las propiedades de la tasa de convergencia local del método de Newton. Sabemos que para todo x en una vecindad de un punto solución x_* tal que $\nabla^2 f(x_*)$ es definida positiva, la matriz hessiana $\nabla^2 f(x)$ será también definida positiva. El método de Newton estará bien definido en esta vecindad y convergerá cuadráticamente, siempre que las longitudes de paso α_k sean eventualmente siempre 1.

Teorema 1.5. *Supongamos que f es dos veces continuamente diferenciable y que la matriz Hessiana $\nabla^2 f(x)$ es Lipschitz continua en una vecindad de una solución x_*, en la cual se cumplen las condiciones suficientes del Teorema 1.3. Considere la iteración $x_{k+1} = x_k + p_k$, donde p_k está dado por (1.11). Entonces*

1. *si el punto inicial x_0 es suficientemente cercano a x_*, la sucesión de iteraciones converge a x_*;*
2. *la tasa de convergencia de $\{x_k\}$ es cuadrática; y*
3. *La sucesión de normas del gradiente $\|\nabla f(x_k)\|$ converge cuadráticamente a cero.*

Demostración. De la definición del paso de Newton y la condición de optimalidad $\nabla f(x^*) = 0$ tenemos que

$$\begin{aligned} x_k + p_k^N - x^* &= x_k - x^* - \nabla^2 f(x_k)^{-1} \nabla f(x_k) \\ &= \nabla^2 f(x_k)^{-1} \left[\nabla^2 f(x_k)(x_k - x^*) - (\nabla f(x_k) - \nabla f(x_*)) \right]. \end{aligned} \tag{1.12}$$

Dado que

$$\nabla f_k - \nabla f_* = \int_0^1 \nabla^2 f(x_k + t(x^* - x_k))(x_k - x^*)\, dt,$$

tenemos que

$$\begin{aligned} \|\nabla^2 f(x_k)(x_k - x^*) - (\nabla f_k - \nabla f_*)\| &= \left\| \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x_k + t(x^* - x_k))](x_k - x^*)\, dt \right\| \\ &\leq \int_0^1 \|\nabla^2 f(x_k) - \nabla^2 f(x_k + t(x^* - x_k))\| \|x_k - x^*\|\, dt \\ &\leq \|x_k - x^*\|^2 \int_0^1 Lt\, dt \\ &= \frac{1}{2} L \|x_k - x^*\|^2, \end{aligned} \tag{1.13}$$

donde L es la constante de Lipschitz de $\nabla^2 f(x)$ para x cercano a x^*. Dado que $\nabla^2 f(x^*)$ es no-singular, y como $\nabla^2 f_k \to \nabla^2 f(x^*)$, tenemos que $\|\nabla^2 f_k^{-1}\| \leq 2\|\nabla^2 f(x^*)^{-1}\|$ para toda k suficientemente grande. Sustituyendo en (1.12) y (1.13), obtenemos

$$\|x_k + p_k^N - x^*\| \leq L\|\nabla^2 f(x^*)^{-1}\| \ \|x_k - x^*\|^2 = \tilde{L}\|x_k - x^*\|^2,$$

donde $\tilde{L} = L\|\nabla^2 f(x^*)^{-1}\|$. Usando esta desigualdad inductivamente deducimos que si el punto inicial es suficientemente cercano a x^*, entonces la sucesión converge a x^*, y la tasa de convergencia es cuadrática.

Usando las relaciones $x_{k+1} - x_k = p_k^N$ y $\nabla f_k + \nabla^2 f_k p_k^N = 0$, obtenemos que

$$\begin{aligned} \|\nabla f(x_{k+1})\| &= \|\nabla f(x_{k+1}) - \nabla f_k - \nabla^2 f(x_k) p_k^N\| \\ &= \left\| \int_0^1 \nabla^2 f(x_k + t p_k^N)(x_{k+1} - x_k)\, dt - \nabla^2 f(x_k) p_k^N \right\| \\ &\leq \int_0^1 \|\nabla^2 f(x_k + t p_k^N) - \nabla^2 f(x_k)\| \|p_k^N\|\, dt \\ &\leq \frac{1}{2} L \|p_k^N\| \\ &\leq \frac{1}{2} L \|\nabla^2 f(x_k)^{-1}\|^2 \|\nabla f_k\|^2 \\ &\leq 2L \|\nabla^2 f(x^*)^{-1}\|^2 \|\nabla f_k\|^2, \end{aligned}$$

probando que la norma de los gradientes converge a cero cuadráticamente. □

CAPÍTULO 2

EL MÉTODO DFP

De aquí en adelante, f será una función cuyo dominio es $\mathbb{R}^n$ y rango es $\mathbb{R}$, además escribiremos f_k para referirnos a la evaluación de f en la $k-ésima$ iteración, es decir, $f_k = f(x_k)$. Del mismo modo denotamos $\nabla f_k = \nabla f(x_k)$ y $\nabla^2 f_k = \nabla^2 f(x_k)$. El objetivo es trabajar en el problema de optimización sin restricciones, es decir, tenemos que resolver

$$\begin{aligned} &\text{mín } f(x); \\ &sujeto\ a\ \ x \in \mathbb{R}^n. \end{aligned}$$

2.1. Construcción del método

El primer método Cuasi-Newton fue creado por Davidon, más tarde Fletcher y Powell demostraron que el algoritmo creado por Davidon era mucho más rápido y confiable que los ya existentes. Dicho método consiste en considerar una aproximación cuadrática de la función objetivo en la iteración actual x_k. Tal aproximación está dada por

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p, \tag{2.1}$$

donde B_k es una matriz simétrica y definida positiva de tamaño nxn, la cual se actualizará en cada iteración. La matriz B_k no es la matriz Hessiana de la función objetivo, pero es una aproximación de ésta.

Notamos que $m_k(0) = f_k$ y $\nabla m_k(p) = \nabla f_k + B_k p$, así $\nabla m_k(0) = \nabla f_k$. Además, como m_k es una función cuadrática y la matriz B_k es definida positiva, se sigue que m_k es una función convexa, por lo cual su mínimo lo alcanza cuando $\nabla m_k(p_k) = 0$, es decir, $\nabla f_k + B_k p_k = 0$, por lo que

$$p_k = -B_k^{-1} \nabla f_k, \tag{2.2}$$

el cual usaremos en la nueva iteración

$$x_{k+1} = x_k + \alpha_k p_k, \tag{2.3}$$

donde α_k se escoge de tal forma que cumpla las condiciones de Wolf (1.9).

Ahora supongamos que hemos generado la nueva iteración x_{k+1} y construimos un nuevo modelo cuadrático, de la forma

$$m_{k+1}(p) = f_{k+1} + \nabla f_{k+1}^T p + \frac{1}{2} p^T B_{k+1} p.$$

Un requerimiento que parece razonable imponer en B_{k+1} es que el gradiente de m_{k+1} debe coincidir con el gradiente de la función objetivo f en al menos dos iteraciones x_k y x_{k+1}. Notemos que

$$\nabla m_{k+1}(p) = \nabla f_{k+1} + B_{k+1}p \Rightarrow \nabla m_{k+1}(0) = \nabla f_{k+1},$$

por lo que la segunda de estas condiciones se satisface. La primera condición puede ser escrita como

$$\nabla m_{k+1}(x_k - x_{k+1}) = \nabla f_k \Rightarrow \nabla f_{k+1} + B_{k+1}(x_k - x_{k+1}) = \nabla f_k,$$

de donde

$$B_{k+1}(x_{k+1} - x_k) = \nabla f_{k+1} - f_k. \tag{2.4}$$

Para simplificar la ecuación anterior, definimos los vectores

$$s_k = x_{k+1} - x_k = \alpha_k p_k; \tag{2.5a}$$

$$y_k = \nabla f_{k+1} - \nabla f_k. \tag{2.5b}$$

Por lo que la ecuación (2.4) se reescribe como

$$B_{k+1}s_k = y_k. \tag{2.6}$$

La cual es conocida como la ***ecuación secante***.

A partir del hecho de que la matriz B_{k+1} es definida positiva y por la ecuación secante, se tiene que

$$s_k^T B_{k+1} s_k = s_k^T y_k > 0. \tag{2.7}$$

La desigualdad anterior es conocida como la ***condición de curvatura***.

Cuando la función f es fuertemente convexa, la desigualdad (2.7) se cumplirá para cualesquiera dos puntos x_{k+1} y x_k, en efecto, como f es fuertemente convexa, se sigue que f es estrictamente convexa, y por el Teorema C.2 se tiene que :

$$(x_{k+1} - x_k)^T(\nabla f_{k+1} - \nabla f_k) = s_k^T y_k > 0.$$

La condición de curvatura no siempre será verdadera cuando la función es no convexa, en este caso tenemos que forzar a que la desigualdad (2.7) sea verdadera, lo cual se logra imponiendo las condiciones de Wolfe o las condiciones fuertes de Wolfe en α_k. Para verificar esto, notamos de (2.3), (2.5a) y la segunda condición de Wolfe (1.9b) que

$$\begin{aligned}
\nabla f_{k+1}^T p_k \geq c_2 \nabla f_k^T p_k &\Rightarrow \nabla f_{k+1}^T \frac{1}{\alpha_k}(x_{k+1} - x_k) \geq c_2 \nabla f_k^T \frac{1}{\alpha_k}(x_{k+1} - x_k) \\
&\Rightarrow \nabla f_{k+1}^T (x_{k+1} - x_k) \geq c_2 \nabla f_k^T (x_{k+1} - x_k) \\
&\Rightarrow \nabla f_{k+1}^T s_k \geq c_2 \nabla f_k^T s_k \\
&\Rightarrow \nabla f_{k+1}^T s_k - \nabla f_k^T s_k \geq c_2 \nabla f_k^T s_k - \nabla f_k^T s_k \\
&\Rightarrow (\nabla f_{k+1}^T - f_k^T) s_k \geq (c_2 - 1) \nabla f_k^T s_k.
\end{aligned}$$

De donde se tiene que

$$y_k^T s_k \geq (c_2 - 1)\alpha_k \nabla f_k^T p_k. \tag{2.8}$$

Ya que $c_2 < 1$ se sigue que $c_2 - 1 < 0$, además $\alpha_k > 0$ y p_k es una dirección de descenso, por lo que $\nabla f_k^T p_k < 0$, de esta manera el lado derecho de la desigualdad (2.8) es positivo, por lo que se cumple la desigualdad (2.7).

Cuando la condición de curvatura se cumple, la ecuación secante (2.6) siempre tiene solución, de hecho tiene un número infinito de soluciones, ya que hay $\frac{n(n+1)}{2}$ grados de libertad en una matriz simétrica, y la ecuación secante representa solo n condiciones, adicionalmente la condición de que la matriz B_k sea definida positiva aumenta n condiciones más (todos los menores principales deben ser positivos), pero estas condiciones no absorben los restantes grados de libertad.

Para determinar B_{k+1} de forma única, debemos de pedir la condición de que entre todas las matrices simétricas que cumplen la ecuación secante, B_{k+1} debe ser, de alguna forma, la más cercana a la matriz actual B_k, es decir, tenemos que resolver el problema

$$\min_B \|B - B_k\|; \tag{2.9a}$$

$$sujeto\ a \quad B = B^T, \quad Bs_k = y_k. \tag{2.9b}$$

Donde s_k y y_k satisfacen (2.7) y B_k es simétrica y definida positiva. Muchas normas matriciales pueden utilizarse en (2.9a) y cada norma da lugar a un método Cuasi-Newton distinto. Una norma que permite una fácil solución del problema de minimización (2.9) es la norma pesada de Frobenius

$$\|A\|_W \equiv \|W^{\frac{1}{2}} A W^{\frac{1}{2}}\|_F, \tag{2.10}$$

donde $\|\cdot\|_F$ está definida por

$$\|C\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}^2.$$

Notemos que la norma $\|\cdot\|_W$ puede también ser calculada mediante

$$\|A\|_W^2 = traza\left(WA^T WA\right). \tag{2.11}$$

La matriz de peso W puede ser escogida como cualquier matriz que satisfaga la relación $Wy_k = s_k$. Asumiremos que $W = \bar{G}_k^{-1}$, donde $\bar{G}_k$ es el Hessiano promedio definido por

$$\bar{G}_k = \frac{\int_0^1 \nabla^2 f(x_k + \tau\alpha_k p_k) d\tau}{\mathscr{L}(0,1)} = \int_0^1 \nabla^2 f(x_k + \tau\alpha_k p_k) d\tau, \tag{2.12}$$

donde $\mathscr{L}(0,1)$ es la longitud del intervalo $(0,1)$, la cual es 1. Así la integral anterior nos mide el promedio de los Hessianos entre $\nabla^2 f_k$ y $\nabla^2 f_{k+1}$. Veamos que $\bar{G}_k s_k = y_k$. En efecto, de (2.5a) y (2.3) tenemos que $s_k = \alpha_k p_k$ y por el Teorema Fundamental del Cálculo tenemos que:

$$\begin{aligned}
\bar{G}_k s_k &= \bar{G}_k \alpha_k p_k \\
&= \left[\int_0^1 \nabla^2 f(x_k + \tau \alpha_k p_k) d\tau \right] \alpha_k p_k \\
&= \int_0^1 \nabla^2 f(x_k + \tau \alpha_k p_k) \alpha_k p_k d\tau \\
&= \nabla f(x_k + \alpha_k p_k) - \nabla f(x_k) \\
&= \nabla f_{k+1} - \nabla f_k \\
&= y_k.
\end{aligned}$$

De donde se tiene que

$$\bar{G}_k s_k = y_k. \tag{2.13}$$

Dado que $\bar{G}_k$ es el promedio de matrices definidas positivas, se sigue que debe de ser definida positiva, por lo cual es invertible y así ${\bar{G}_k}^{-1} y_k = s_k$.

Con la norma y matriz de peso descritas arriba, la solución única de (2.9) es

$$B_{k+1}^{DFP} = (I - \gamma_k y_k s_k^T) B_k (I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T, \tag{2.14}$$

con

$$\gamma_k = \frac{1}{y_k^T s_k}.$$

Ésta fórmula es llamada la fórmula de actualización DFP, dado que es la propuesta por Davidon en 1959 y después estudiada, implementada y popularizada por Fletcher y Powell.

Ahora denotamos $H_k := B_k^{-1}$, la cual es útil, ya que nos permite calcular la dirección de búsqueda (2.2) por una multiplicación simple de matriz-vector. Usando la fórmula de Sherman-Morrison-Woodbury (B.2) y la ecuación (2.14) se obtiene que

$$H_{k+1}^{DFP} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}. \tag{2.15}$$

Notemos que los dos últimos términos del lado derecho de (2.15) son matrices de rango uno, así que H_k será una modificación de rango 2. Ésta es la idea fundamental de la actualización Quasi-Newton, en lugar de recalcular las matrices desde cero, aplicamos una simple modificación que combina la más reciente información observada acerca de la función objetivo con el conocimiento existente incrustado en nuestra aproximación del Hessiano actual.

2.2. Propiedades del método

El método DFP, definido por (2.15), tiene un número importante de propiedades, algunas de ellas son las siguientes:

1. *Para funciones cuadráticas* (con líneas de búsqueda exactas)
   - (i) termina en a lo más $n+1$ iteraciones, con $H_{n+1} = (\nabla^2 f)^{-1}$;
   - (ii) la ecuación secante se preserva en las iteraciones siguientes;
   - (iii) genera direcciones conjugadas y gradientes conjugados cuando $H_0 = I$.
2. *Para funciones generales*
   - (iv) preserva la condición de que H_k es definida positiva, para todo k, (por lo tanto la propiedad de descenso es verdadera);
   - (v) requiere $3n^2 + O(n)$ multiplicaciones por iteración;
   - (vi) orden de convergencia superlineal;
   - (vii) convergencia global para funciones estrictamente convexas (con líneas de búsqueda exactas).

Para las demostraciones de (i) y (iii) se hará uso de los siguientes resultados.

Teorema 2.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función cuadrática, es decir, $f(x) = c^T x + \frac{1}{2} x^T G x$, donde G es una matriz simétrica de tamaño $n \times n$ y $c^T \in \mathbb{R}^n$. Sean $p_1, \ldots, p_n$ vectores $G-conjugados$ y x_0 un punto inicial arbitrario. Para $k = 1, \ldots, n$ sea α_k una solución óptima del problema de minimizar $f(x_k + \alpha p_k)$ sujeto a que $\alpha \in \mathbb{R}$ y sea $x_{k+1} = x_k + \alpha_k p_k$. Entonces para $k = 1, \ldots, n$ se tiene que:*

1. $\nabla f_{k+1}^T p_j = 0$, *para* $j = 1, \ldots, k$.
2. $\nabla f_0^T p_k = \nabla f_k p_k$.
3. x_{k+1} *es una solución óptima del problema de minimizar* $f(x)$ *sujeto a que* $x - x_0 \in L(p_1, \ldots, p_k)$, *donde* $L(p_1, \ldots, p_k)$ *es el subespacio lineal formado por* $p_1, \ldots, p_k$; *esto es,*

$$L(p_1, \ldots, p_k) = \left\{ \sum_{j=1}^{k} \mu_j p_j : \mu_j \in \mathbb{R}, \ para\ cada\ j \right\}.$$

En particular, x_{n+1} es un punto minimizador de f sobre $\mathbb{R}^n$.

Lema 2.2.1. *Sean $x_0 \in \mathbb{R}^n$ y H_0 una matriz inicial simétrica y definida positiva. Para $j = 1, \ldots, n$ sea $x_{j+1} = x_j + \alpha_j p_j$, donde $p_j = -H_j \nabla f_j$ y α_j resuelve el problema de minimizar $f(x_j + \alpha p_j)$, sujeto a que $\alpha > 0$. Además, H_{j+1} está dado por (2.15). Si $\nabla f_j \neq 0$ para $j = 1, \ldots, n$, entonces $H_0, \ldots, H_n$ son simétricas y definidas positivas y así, $p_1, \ldots, p_n$ son direcciones de descenso.*

Las demostraciones de ambos resultados se encuentran en [1].

Teorema 2.2. *Sea G una matriz simétrica y defina positiva de tamaño $n \times n$ y considere el problema de minimizar $f(x) = c^T x + \frac{1}{2} x^T G x$, sujeto a que $x \in \mathbb{R}^n$. Suponga que el problema es resuelto por el método DFP, comenzando en algún punto inicial x_0 y una matriz simétrica y definida positiva H_0. Además, para $j \in \{1, \dots n\}$ sean α_j la solución óptima del problema de minimizar $f(x_j + \alpha p_j)$, sujeto a que $\alpha \geq 0$ y $x_{j+1} = x_j + \alpha_j p_j$, donde $p_j = -H_j \nabla f_j$ y H_j está determinado por (2.15). Si $\nabla f_j \neq 0$ para cada j, entonces las direcciones $p_1, \dots, p_n$ son $G-conjugadas$ y $H_n = G^{-1}$. Además, x_n es una solución óptima del problema.*

Demostración. Primero demostraremos que para cualquier $j \in \{1, \dots, n\}$ las siguientes proposiciones son verdaderas:

1. $\{p_1, \dots, p_j\}$ es un conjunto linealmente independiente.
2. $p_i^T G p_k = 0$, para $i \neq k$, $i, k \leq j$.
3. $H_{j+1} G p_k = p_k$ para $1 \leq k \leq j$, o equivalentemente $H_{j+1} G s_k = s_k$, donde $s_k = \alpha_k p_k$.

La demostración de éste resultado se realizará por inducción sobre j.

Para $j = 1$, las partes 1 y 2 son obvias. Para demostrar la parte 3, primero notemos que para cualquier k tenemos que $\nabla f_k = c + G x_k$, de donde

$$G s_k = G(\alpha_k p_k) = G(x_{k+1} - x_k) = \nabla f_{k+1} - \nabla f_k = y_k. \tag{2.16}$$

En particular, $G s_1 = y_1$. Así,

$$\begin{aligned} H_2 G s_1 = H_2 y_1 &= \left(H_1 + \frac{s_1 s_1^T}{s_1^T y_1} - \frac{H_1 y_1 y_1^T H_1}{y_1^T H_1 y_1} \right) y_1 \\ &= H_1 y_1 + \frac{s_1 (s_1^T y_1)}{(s_1^T y_1)} - \frac{H_1 y_1 (y_1^T H_1 y_1)}{(y_1^T H_1 y_1)} \\ &= H_1 y_1 + s_1 - H_1 y_1 \\ &= s_1. \end{aligned}$$

Por lo que, la parte 3 es verdadera para $j = 1$.

Ahora supóngase que las partes 1, 2 y 3 son verdaderas para $j \leq n - 1$. Para demostrar que también son verdaderas para $j + 1$, primero notemos que por la parte 1 del Teorema 2.1 se tiene que $p_i^T \nabla f_{j+1} = 0$ para $i \leq j$. Por la hipótesis de inducción de la parte 3 se sigue que $p_i^T = p_i^T G H_{j+1}$, para $i \leq j$. Por lo tanto, observando que $p_{j+1} = -H_{j+1} \nabla f_{j+1}$, para $i \leq j$ se obtiene

$$\begin{aligned} 0 &= p_i^T \nabla f_{j+1} \\ &= p_i^T G H_{j+1} \nabla f_{j+1} \\ &= p_i^T G(-p_{j+1}) \\ &= -p_i^T G p_{j+1}. \end{aligned}$$

En vista de la hipótesis de inducción de la parte 2, la ecuación anterior muestra que la parte 2 es verdadera para $j + 1$. Ahora demostraremos que 3 es verdadera para $j + 1$. Si $k \leq j + 1$, entonces

$$H_{j+2}Gs_k = \left(H_{j+1} + \frac{s_{j+1}s_{j+1}^T}{s_{j+1}^T y_{j+1}} - \frac{H_{j+1}y_{j+1}y_{j+1}^T H_{j+1}}{y_{j+1}^T H_{j+1} y_{j+1}} \right) Gs_k, \tag{2.17}$$

Notando (2.16) y haciendo $k = j+1$ en (2.17), se sigue que

$$\begin{aligned} H_{j+2}Gs_k &= H_{j+2}Gs_{j+1} \\ &= \left(H_{j+1} + \frac{s_{j+1}s_{j+1}^T}{s_{j+1}^T y_{j+1}} - \frac{H_{j+1}y_{j+1}y_{j+1}^T H_{j+1}}{y_{j+1}^T H_{j+1} y_{j+1}} \right) y_{j+1} \\ &= H_{j+1}y_{j+1} + \frac{s_{j+1}(s_{j+1}^T y_{j+1})}{(s_{j+1}^T y_{j+1})} - \frac{H_{j+1}y_{j+1}(y_{j+1}^T H_{j+1} y_{j+1})}{(y_{j+1}^T H_{j+1} y_{j+1})} \\ &= H_{j+1}y_{j+1} + s_{j+1} - H_{j+1}y_{j+1} \\ &= s_{j+1}. \end{aligned}$$

Ahora sea $k \leq j$, dado que 2 es verdadero para $j+1$, se tiene que

$$s_{j+1}^T Gs_k = \alpha_k \alpha_{j+1} p_{j+1} Gp_k = 0. \tag{2.18}$$

Notando la hipótesis de inducción 3, la ecuación (2.16) y el hecho de que 2 es verdadera para $j+1$, tenemos

$$y_{j+1}^T H_{j+1} Gs_k = y_{j+1}^T s_k = s_{j+1}^T Gs_k = \alpha_{j+1}\alpha_k p_{j+1}^T Gp_k = 0. \tag{2.19}$$

Sustituyendo (2.18) y (2.19) en (2.17) y notando la hipótesis de inducción de 3, se obtiene que

$$\begin{aligned} H_{j+2}Gs_k &= \left(H_{j+1} + \frac{s_{j+1}s_{j+1}^T}{s_{j+1}^T y_{j+1}} - \frac{H_{j+1}y_{j+1}y_{j+1}^T H_{j+1}}{y_{j+1}^T H_{j+1} y_{j+1}} \right) Gs_k \\ &= H_{j+1}Gs_k + \frac{s_{j+1}(s_{j+1}^T Gs_k)}{s_{j+1}^T y_{j+1}} - \frac{H_{j+1}y_{j+1}(y_{j+1}^T H_{j+1} Gs_k)}{y_{j+1}^T H_{j+1} y_{j+1}} \\ &= H_{j+1}Gs_k \\ &= s_k. \end{aligned}$$

Por lo tanto, 3 es verdadera para $j+1$.

Para completar el argumento de inducción, solo debemos probar que 1 es cierto para $j+1$. Supóngase que $\sum_{i=1}^{j+1} a_i p_i = 0$. Multiplicando por $p_{j+1}G$ y notando que 2 es cierta para $j+1$, se sigue que $a_{j+1}p_{j+1}^T Gp_{j+1} = 0$. Por la hipótesis de que $\nabla f_{j+1} \neq 0$ y por el Lema 2.2.1, H_{j+1} es definida positiva, así que $p_{j+1} = -H_{j+1}\nabla f_{j+1} \neq 0$. Como G es definida positiva $p_{j+1}^T Gp_{j+1} \neq 0$ y por tanto $a_{j+1} = 0$. Esto implica que $\sum_{i=1}^{j} a_i p_i = 0$ y dado que $p_1, \ldots, p_j$ son linealmente independientes por la hipótesis inductiva, se tiene que $a_i = 0$ para $i = 1, \ldots, j$. Por lo tanto

$p_1,\ldots,p_{j+1}$ son linealmente independientes y 1 es verdadera para $j+1$. De esta manera 1, 2 y 3 son verdaderos. En particular la conjugación de $p_1,\ldots,p_n$ se sigue de 1 y 2 haciendo $j=n$.

Ahora sea $j=n$ en 3, entonces $H_{n+1}Gp_k=p_k$, para $k=1,\ldots,n$. Sea P la matriz cuyas columnas son $p_1,\ldots,p_n$, entonces $H_{n+1}GP=P$. Observe que P es invertible, ya que sus columnas son linealmente independientes y así su determinante es distinto de 0, de este modo $H_{n+1}G=I_n$, lo cuál es posible si $H_{n+1}=G^{-1}$. Finalmente, x_{n+1} es una solución óptima por el Teorema 2.1.

□

Teorema 2.3. *Si $y_k^T s_k>0$ para todo k, entonces la fórmula DFP preserva las matrices definidas positivas H_k.*

Demostración. La demostración se hace por inducción y demostraremos que $z^T H_k z>0$ para todo $z\neq 0$. El resultado es verdadero para $k=0$ por elección. Asumiremos que el resultado es verdadero para algún $k\geq 1$. Escribimos $H_k=LL^T$, ya que la factorización de Cholesky existe y si $a=L^T z$ y $b=L^T y_k$, entonces

$$z^T\left(H_k-\frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right)z=z^T(LL^T)z-\frac{z^T(LL^T)y_k y_k^T(LL^T)z}{y_k^T(LL^T)y_k}=a^T a-\frac{(a^T b)(b^T a)}{b^T b}\geq 0.$$

La última desigualdad se cumple por la desigualdad de Cauchy. Además, dado que se cumple la condición de curvatura $s_k^T y_k>0$, se tiene que

$$z^T\left(\frac{s_k s_k^T}{s_k^T y_k}\right)z=\frac{\left(z^T s_k\right)\left(s_k^T z\right)}{s_k^T y_k}\geq 0.$$

Por lo tanto, $z^T H_{k+1}z\geq 0$. □

CAPÍTULO 3

EL MÉTODO BFGS

La fórmula de actualización DFP es bastante efectiva, pero pronto fue superada por la fórmula BFGS, la cual es actualmente considerada la más eficiente de todas las fórmula de actualización Quasi-Newton.

3.1. Construcción del método

La fórmula de actualización BFGS puede ser derivada haciendo un simple cambio en el argumento que nos lleva a (2.14). En lugar de imponer condiciones en las aproximaciones del Hessiano B_k, imponemos condiciones similares en sus inversos H_k. La aproximación actualizada H_{k+1} debe ser simétrica, definida positiva y debe satisfacer la ecuación secante (2.6), ahora escrita como

$$H_{k+1}y_k = s_k.$$

La condición de la cercanía a H_k es ahora especificada de forma análoga a (2.9)

$$\min_H \|H - H_k\|; \tag{3.1a}$$

$$sujeto\ a \quad H = H^T, \quad Hy_k = s_k. \tag{3.1b}$$

La norma considera es de nuevo la norma pesada de Frobenius descrita en el capítulo anterior, donde la matriz de peso W es ahora cualquier matriz que satisfaga $Ws_k = y_k$. (Por concreción, asumimos de nuevo que W está dado por el Hessiano promedio $\bar{G}_k$ definida en (2.12)). La única solución H_{k+1} a (3.1) está dada por

$$H_{k+1}^{BFGS} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T, \tag{3.2}$$

donde

$$\rho_k = \frac{1}{y_k^T s_k}. \tag{3.3}$$

Otra forma en la que podemos actualizar la matriz, es desarrollando las multiplicaciones que aparecen el la fórmula $BFGS$, con lo cuál H_{k+1} está dada por;

$$H_{k+1}^{BFGS} = H_k + \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right)\frac{s_k s_k^T}{s_k^T y_k} - \left(\frac{s_k y_k^T H_k + H_k y_k s_k^T}{s_k^T y_k}\right). \tag{3.4}$$

Teorema 3.1. *Si la fórmula de actualización BFGS (3.4), es ahora escrita como* $H_{k+1} = H_k + E$, *entonces E resuelve el problema variacional*

$$\min_E \|E\|_W; \tag{3.5a}$$

$$sujeto\ a \quad E = E^T, \tag{3.5b}$$

$$Ey_k = \eta, \tag{3.5c}$$

donde $\eta = s_k - H_k y_k$ *y W satisface que* $Ws_k = y_k$.

Demostración. Primero veamos que el problema (3.5) es un problema convexo. Al ser la función objetivo una norma, es una función convexa, probemos que el conjunto de restricciones es un conjunto convexo. Dadas dos matrices simétricas, mostraremos que la combinación convexa de dos matrices simétricas es una matriz simétrica. Sean A, B dos matrices simétricas de tamaño $n \times n$ con coeficientes reales, así para todo $\alpha \in [0,1]$ se tiene que

$$\begin{aligned} \alpha A + (1-\alpha)B &= \alpha A^T + (1-\alpha)B^T \\ &= (\alpha A)^T + ((1-\alpha)B)^T \\ &= (\alpha A + (1-\alpha)B)^T. \end{aligned}$$

Así, la combinación convexa de matrices simétricas es simétrica.

Veamos ahora que el conjunto de matrices que cumplen la relación $Ey_k = \eta$ es convexo. Sean A y B dos matrices de tamaño $n \times n$ con coeficientes reales que cumplen la relación anterior, es decir $Ay_k = \eta$ y $By_k = \eta$, así para todo $\alpha \in [0,1]$ se tiene que

$$\begin{aligned} (\alpha A + (1-\alpha)B)y_k &= \alpha Ay_k + (1-\alpha)By_k \\ &= \alpha\eta + (1-\alpha)\eta \\ &= \alpha\eta + \eta - \alpha\eta \\ &= \eta. \end{aligned}$$

Por lo tanto, la combinación convexa de matrices que cumplen la igualdad, también la cumple. En consecuencia el problema dado es un problema convexo, por lo que es suficiente encontrar E que satisfaga la condiciones de primer orden. Después de cuadrar la norma, una adecuada función lagrangiana es

$$\mathscr{L} = \frac{1}{4} traza(WE^T WE) + traza(\Lambda^T(E^T - E)) - \lambda^T W(Ey_k - \eta),$$

donde Λ es una matriz de multiplicadores de Lagrange para la restricción $E^T = E$ y $\lambda^T W$ es un vector de multiplicadores de Lagrange para la restricción $Ey_k = \eta$. $\mathscr{L}$ debe ser estacionaria con respecto a E, Λ y λ. Igualando a cero las derivadas respecto a Λ y λ, nos da justamente las restricciones de (3.5). Obsérvese que $\frac{\partial E}{\partial E_{ij}} = e_i e_j^T$ por lo tanto, derivando $\mathscr{L}$ respecto a E_{ij} e igualando a 0, se obtiene que

$$\frac{1}{4}\left[traza\left(We_j e_i^T WE\right) + traza\left(WE^T We_i e_j^T\right)\right] + traza\Lambda\left(e_j e_i^T - e_i e_j^T\right) - \lambda^T We_i e_j^T y_k = 0,$$

o usando la simetría e invarianza de la traza de permutaciones cíclicas,

$$\frac{1}{2}[WEW]_{ij} + \Lambda_{ij} - \Lambda_{ji} = \left[W\lambda y_k^T\right]_{ij}.$$

Trasponiendo y sumando se elimina Λ para obtener

$$WEW = W\lambda y_k^T + y_k\lambda^T W.$$

Usando que $Ws_k = y_k$ y que W es no-singular, se sigue que

$$E = \lambda s_k^T + s_k\lambda^T. \tag{3.6}$$

De la ecuación anterior observamos que la actualización es de rango dos. Ahora es posible resolver para λ usando las restricciones de (3.5). Sustituyendo (3.6) en $Ey_k = \eta$ obtenemos

$$\begin{aligned} Ey_k = \eta &\Rightarrow (\lambda s_k^T + s_k\lambda^T)y_k = \eta \\ &\Rightarrow \lambda s_k^T y_k + s_k\lambda^T y_k = \eta \\ &\Rightarrow \lambda s_k^T y_k = \eta - s_k\lambda^T y_k. \end{aligned}$$

De donde

$$\lambda = \frac{\eta - s_k\lambda^T y_k}{s_k^T y_k}. \tag{3.7}$$

Trasponiendo la igualdad anterior y post-multiplicándola por y_k, se consigue

$$\begin{aligned} \lambda = \frac{\eta - s_k\lambda^T y_k}{s_k^T y_k} &\Rightarrow \lambda^T y_k = \left(\frac{\eta - s_k\lambda^T y_k}{s_k^T y_k}\right)^T y_k \\ &\Rightarrow \lambda^T y_k = \frac{\eta^T y_k - y_k^T\lambda s_k^T y_k}{s_k^T y_k} \\ &\Rightarrow \lambda^T y_k = \frac{\eta^T y_k}{s_k^T y_k} - \frac{y_k^T\lambda(s_k^T y_k)}{s_k^T y_k} \\ &\Rightarrow \lambda^T y_k = \frac{\eta^T y_k}{s_k^T y_k} - y_k^T\lambda \\ &\Rightarrow \lambda^T y_k + y_k^T\lambda = \frac{\eta^T y_k}{s_k^T y_k} \\ &\Rightarrow 2\lambda^T y_k = \frac{\eta^T y_k}{s_k^T y_k} \\ &\Rightarrow \lambda^T y_k = \frac{\eta^T y_k}{2s_k^T y_k}. \end{aligned}$$

Sustituyendo en (3.7) nos da

$$\begin{aligned}\lambda &= \frac{\eta - s_k\left(\frac{\eta^T y_k}{2s_k^T y_k}\right)}{s_k^T y_k} \\ &= \frac{\eta}{s_k^T y_k} - \frac{s_k \eta^T y_k}{2\left(s_k^T y_k\right)^2} \\ &= \frac{s_k - H_k y_k}{s_k^T y_k} - \frac{s_k (s_k - H_k y_k)^T y_k}{2\left(s_k^T y_k\right)^2}.\end{aligned}$$

Sustituyendo en (3.6).

$$\begin{aligned}E &= \lambda s_k^T + s_k \lambda^T \\ &= \left(\frac{s_k - H_k y_k}{s_k^T y_k} - \frac{s_k (s_k - H_k y_k)^T y_k}{2\left(s_k^T y_k\right)^2}\right) s_k^T + s_k \left(\frac{s_k - H_k y_k}{s_k^T y_k} - \frac{s_k (s_k - H_k y_k)^T y_k}{2\left(s_k^T y_k\right)^2}\right)^T \\ &= \frac{s_k s_k^T - H_k y_k s_k^T}{s_k^T y_k} - \frac{s_k (s_k - H_k y_k)^T y_k s_k^T}{2\left(s_k^T y_k\right)^2} + \frac{s_k s_k^T - s_k y_k^T H_k}{s_k^T y_k} - \frac{s_k y_k^T (s_k - H_k y_k) s_k^T}{2(s_k^T y_k)^2} \\ &= \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k s_k^T}{s_k^T y_k} - \frac{s_k (s_k^T y_k) s_k^T}{2\left(s_k^T y_k\right)^2} + \frac{s_k y_k^T H_k y_k s_k^T}{2\left(s_k^T y_k\right)^2} \\ &+ \frac{s_k s_k^T}{s_k^T y_k} - \frac{s_k y_k^T H_k}{s_k^T y_k} - \frac{s_k (y_k^T s_k) s_k^T}{2(s_k^T y_k)^2} + \frac{s_k y_k^T H_k y_k s_k^T}{2(s_k^T y_k)^2} \\ &= \frac{s_k s_k^T}{s_k^T y_k} - \frac{1}{2}\frac{s_k s_k^T}{s_k^T y_k} + \frac{s_k s_k^T}{s_k^T y_k} - \frac{1}{2}\frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \frac{s_k y_k^T H_k y_k s_k^T + s_k y_k^T H_k y_k s_k^T}{2\left(s_k^T y_k\right)^2} \\ &= \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \frac{2 s_k y_k^T H_k y_k s_k^T}{2\left(s_k^T y_k\right)^2} \\ &= \frac{s_k s_k^T}{s_k^T y_k} + \left(\frac{y_k^T H_k y_k}{s_k^T y_k}\right)\frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} \\ &= \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right)\frac{s_k s_k^T}{s_k^T y_k} - \frac{s_k y_k^T H_k + H_k y_k s_k^T}{s_k^T y_k}.\end{aligned}$$

Lo cual al sustituirlo en $H_{k+1} = H_k + E$ nos da la actualización $BFGS$. □

Es fácil ver que H_{k+1} es definida positiva, siempre que H_k sea definida positiva. En efecto, como $y_k^T s_k$ es positivo, la fórmula (3.2) está bien definida. Para cada vector $z \neq 0$, tenemos que

$$z^T H_{k+1} z = w^T H_k w + \rho_k (z^T s_k)^2 \geq 0,$$

donde $w = z - \rho_k y_k (s_k^T z)$. El lado derecho de la igualdad anterior puede ser cero, solo si $s_k^T z = 0$, pero en este caso $w = z \neq 0$, lo cual implica que el primer término es mayor que 0. Por lo tanto, H_{k+1} es definida positiva.

3.2. Algoritmo BFGS

Sólo un problema tiene que ser resuelto antes de que podamos definir un algoritmo BFGS completo. ¿ Cómo debemos escoger la aproximación inicial H_0? Desafortunadamente, no hay una fórmula que funcione bien en todos los casos. Podemos usar información especifica acerca del problema, por ejemplo, configurándolo como el inverso de una aproximación del Hessiano calculado por diferencias finitas en x_0. De otra manera, podemos simplemente configurarla para que sea la matriz identidad, o un múltiplo de la matriz identidad, donde el múltiplo es escogido para reflejar el escalamiento de las variables.

Algoritmo 1 Algoritmo BFGS

Entrada: Punto inicial x_0, tolerancia ε, aproximación del inverso de el Hessiano H_0.
Salida: H_n
$k \leftarrow 0$;
mientras $\|\nabla f_k\| > \varepsilon$ **hacer**
Calcular la dirección de búsqueda

$$p_k = -H_k \nabla f_k; \tag{3.8}$$

Configurar $x_{k+1} = x_k + \alpha_k p_k$, donde α_k es calculado mediante un procedimiento de búsqueda en la línea que satisfaga las condiciones de Wolf;
Definir $s_k = x_{k+1} - x_k$ y $y_k = \nabla f_{k+1} - \nabla f_k$;
Calcule H_{k+1} por medio de (3.2);
$k \leftarrow k + 1$;
fin mientras

Cada iteración puede ser ejecutado con un costo de $O(n^2)$ operaciones aritméticas (más el costo de las evaluaciones en la función y el gradiente); no hay operaciones de $O(n^3)$, tales como solución de un sistema de ecuaciones lineal u operaciones matriz-matriz. El algoritmo es robusto, y su rapidez de convergencia es superlineal, lo cual es suficientemente rápido para la mayoría de los propósitos prácticos. Incluso aunque el método de Newton converge más rápido a la solución (es decir, cuadráticamente), su costo por iteración es mayor, ya que requiere la solución de un sistema lineal. Una ventaja muy importante del método BFGS es, por supuesto, que no requiere el cálculo de segundas derivadas.

Podemos obtener una versión del algoritmo BFGS que funcione con la aproximación del Hessiano B_k en lugar de H_k. La fórmula de actualización para B_k es obtenida simplemente aplicando la fórmula de Sherman-Morrison-Woodbury (B.2) para obtener:

$$B_{k+1}^{BFGS} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}. \tag{3.9}$$

Una sencilla implementación de esta variante no es eficiente para minimización sin restricciones, pues requiere que el sistema $B_k p_k = -\nabla f_k$ sea resuelto para el paso p_k, y de este modo incrementando el costo del cálculo de paso a $O(n^3)$. Con el fin de obtener una fórmula de actualización Cuasi-Newton que sea invariante a cambios en las variables, es necesario que que (2.9a) y (3.1a) sean también invariantes. La elección de la matriz de peso W usada

para definir las norma en (2.9a) y (3.1a) asegura que esta condición sea verdadera. Muchas otras matrices W pueden ser escogidas, cada una de ellas da lugar a distintas fórmulas de actualización. Sin embargo, después de intensas búsquedas, no se encontró una fórmula más efectiva que la actualización $BFGS$.

Es razonable preguntarse: ¿hay situaciones en las que la fórmula de actualización (3.2) puede producir malos resultados?, ¿hay alguna iteración en la cual H_k es una mala aproximación del inverso del hessiano?, si es así, ¿existirá alguna forma de corregirlo? Si, por ejemplo, el producto interno $y_k^T s_k$ es pequeño, pero positivo, entonces se sigue de (3.2) y (3.3) que H_{k+1} será muy grande, ¿es este comportamiento razonable? Una pregunta relacionada al error de redondeo que ocurre en una implementación de precisión finita es, ¿pueden estos errores crecer a tal punto que se pierda información útil en la matriz de aproximación Cuasi-Newton?

Estas preguntas han sido estudiadas analítica y experimentalmente. Ahora se sabe que la fórmula $BFGS$ tiene muy efectivas *propiedades de auto-corrección*. Es decir, que si la matriz H_k no estima correctamente la curvatura de la función objetivo y si esta mala estimación ralentiza la iteración, la aproximación del hessiano tenderá a corregirse en pocos pasos. Es también sabido que la fórmula DFP es menos efectiva en corregir malas aproximaciones; se cree que esta propiedad es la razón por la cual tiene un menor desempeño en la práctica. La propiedad de auto-corrección de la fórmula $BFGS$ es válida cuando se usa línea de búsqueda adecuada. En particular, las condiciones de Wolfe aseguran que los gradientes son muestreados en puntos que permiten al modelo (2.1) capturar una apropiada información de la curvatura.

CAPÍTULO 4

ANÁLISIS DE CONVERGENCIA

En este capítulo presentaremos resultados de convergencia global y local para implementaciones del método BFGS. A pesar de que el método BFGS es conocido por ser bastante robusto en la práctica, no somos capaces de establecer resultados de convergencia global para cualquier función objetivo no-lineal. Es decir, no podemos probar que las iteraciones de éste método Cuasi-Newton converge a un punto estacionario del problema, partiendo de cualquier punto inicial y aproximación de la matriz Hessiana. Por otro lado, existe resultados que aseguran convergencia local y superlineal que son verdaderas bajo ciertas suposiciones. A lo largo la presente sección usaremos $\|\cdot\|$ para referirnos a norma Euclidiana, tanto de vectores como de matrices y denotaremos $G(x) = \nabla^2 f(x)$.

4.1. Convergencia Global del Método BFGS

Para obtener la convergencia global, no solo debemos escoger buenas longitudes de paso, también debemos elegir adecuadamente la dirección de búsqueda p_k. Definimos el ángulo θ_k entre p_k y la dirección de descenso más rápido $-\nabla f_k$ mediante

$$\cos\theta_k = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\|\|p_k\|}. \tag{4.1}$$

El siguiente teorema, debido a Zoutendijk, tiene consecuencias de largo alcance. Muestra, por ejemplo que el método de descenso más rápido es globalmente convergente. Para otros algoritmos describe que tan alejado puede estar p_k de la dirección de descenso más rápido para que aún sea un método globalmente convergente.

Teorema 4.1. *Consideremos cualquier iteración de la forma* (2.3) *, donde p_k es una dirección de descenso y α_k satisface las condiciones de Wolfe. Supongamos que f es acotada inferiormente en $\mathbb{R}^n$ y que f es continuamente diferenciable en un conjunto abierto $\mathcal{N}$ que contiene al conjunto de nivel $\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$, donde x_0 es el punto inicial de la iteración. Supongamos además, que el gradiente ∇f es Lipschitz continua en $\mathcal{N}$, esto es, existe una constante $L > 0$ tal que*

$$\|\nabla f(x) - \nabla f(\bar{x})\| \leq L\|x - \bar{x}\|, \quad para\ todo\ x, \bar{x} \in \mathcal{N}. \tag{4.2}$$

Entonces

$$\sum_{k\geq 0} \cos^2\theta_k \|\nabla f_k\|^2 < \infty. \tag{4.3}$$

Demostración. De las condiciones de Wolfe y de (2.3) tenemos que

$$(\nabla f_{k+1} - \nabla f_k)^T \geq (c_2 - 1)\nabla f_k^T p_k,$$

mientras que la condición de Lipschitz (4.2) implica que

$$(\nabla f_{k+1} - \nabla f_k)^T p_k \leq \alpha_k L \| p_k \|^2.$$

Combinando estas dos relaciones, tenemos que

$$\alpha_k \geq \frac{c_2 - 1}{L} \frac{\nabla f_k^T p_k}{\| p_k \|^2}.$$

Sustituyendo esta desigualdad en la primera condición de Wolfe (1.9a) obtenemos que

$$f_{k+1} \leq f_k - c_1 \frac{(1 - c_2)(\nabla f_k^T p_k)^2}{L \| p_k \|^2}.$$

De la definición (4.1), podemos escribir esta relación como

$$f_{k+1} \leq f_k - \cos^2 \theta_k \| \nabla f_k \|^2,$$

donde $c = c_1(1 - c_2)/L$. Sumando esta expresión sobre todos los índices menores o iguales que k, tenemos que

$$f_{k+1} \leq f_0 - c \sum_{j=0}^{k} \cos^2 \theta_j \| \nabla f_j \|^2. \tag{4.4}$$

Dado que f es acotada inferiormente, tenemos que $f_0 - f_{k+1}$ es menor que alguna constante positiva, para toda k. Por lo tanto, tomando límites en (4.4), obtenemos

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \| \nabla f_k \|^2 < \infty.$$

Lo cual concluye la prueba. □

Consideremos ahora las siguiente suposiciones:
Suposición 4.1

1. La función objetivo f es dos veces continuamente diferenciable.
2. El conjunto de nivel $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ es convexo, y existe constantes positivas m y M tales que
$$m \| z \|^2 \leq z^T G(x) z \leq M \| z \|^2, \tag{4.5}$$
para todo $z \in \mathbb{R}^n$ y $x \in \Omega$.

La segunda parte de esta suposición implica que $G(x)$ es definida positiva en Ω y que f tiene un único minimizador en Ω. Usando (2.13) y (4.5) obtenemos que

$$\frac{y_k^T s_k}{s_k^T s_k} = \frac{s_k^T \bar{G}_k s_k}{s_k^T s_k} \geq m. \tag{4.6}$$

Donde $\bar{G}_k$ es la matriz promedio definida en (2.12). La suposición (4.1) implica que $\bar{G}_k$ es definida positiva, por lo que su raíz cuadrada está bien definida. Por lo tanto, si definimos $z_k = \bar{G}_k^{\frac{1}{2}} s_k$ tenemos que

$$\frac{y_k^T y_k}{y_k^T s_k} = \frac{z_k^T \bar{G}_k z_k}{z_k^T z_k} \leq M. \tag{4.7}$$

Ahora estamos listos para presentar el resultado de convergencia global del método BFGS.

Teorema 4.2. *Sea B_0 una matriz inicial, simétrica y definida positiva, además x_0 un punto inicial para la cual la supocisión (4.1) se satisface. Entonces la sucesión $\{x_k\}$ generada por el Algoritmo 1 converge a el minimizador x_* de f.*

Demostración. Definimos

$$m_k = \frac{y_k^T s_k}{s_k^T s_k}, \quad M_k = \frac{y_k^T y_k}{y_k^T s_k}, \tag{4.8}$$

notemos de (4.6) y (4.7) que

$$m_k \geq m, \quad M_k \leq M. \tag{4.9}$$

Calculando la traza de la aproximación BFGS (3.9), obtenemos que

$$traza(B_{k+1}) = traza(B_k) - \frac{\|B_k s_k\|}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k}. \tag{4.10}$$

Además se tiene que

$$det(B_{k+1}) = det(B_k) \frac{y_k^T s_k}{s_k^T B_k s_k}. \tag{4.11}$$

Definimos ahora

$$\cos\theta_k = \frac{s_k^T B_k s_k}{\|s_k\| \|B_k s_k\|}, \quad q_k = \frac{s_k^T B_k s_k}{s_k^T s_k}, \tag{4.12}$$

así que θ_k es el ángulo entre s_k y $B_k s_k$. Obtenemos entonces que

$$\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} = \frac{\|B_k s_k\|^2 \|s_k\|^2}{(s_k^T B_k s_k)^2} \frac{s_k^T B_k s_k}{\|s_k\|^2} = \frac{q_k}{\cos^2\theta_k}. \tag{4.13}$$

Además, de (4.8) tenemos que

$$det(B_{k+1}) = de(B_k) \frac{y_k^T s_k}{s_k^T s_k} \frac{s_k^T s_k}{s_k^T B_k s_k} = det(B_k) \frac{m_k}{q_k}. \tag{4.14}$$

Ahora combinamos la traza y el determinante introduciendo la siguiente función de una matriz definida positiva B:

$$\psi(B) = traza(B) - \ln(det(B)), \tag{4.15}$$

donde $\ln(\cdot)$ denota el logaritmo natural. Puede observarse que $\psi(B) > 0$. Usando (4.8) y (4.10)-(4.15) obtenemos que

$$\begin{aligned} \psi(B_{k+1}) &= \psi(B_k) + M_k - \frac{q_k}{\cos^2(\theta_k)} - \ln(m_k) + \ln(q_k) \\ &= \psi(B_k) + (M_k - \ln m_k - 1) + \left[1 - \frac{q_k}{\cos^2\theta_k} + \ln\frac{q_k}{\cos^2\theta_k}\right] + \ln\cos^2\theta_k. \end{aligned} \tag{4.16}$$

Ahora, dado que la función $h(t) = 1 - t + \ln(t) \leq 0$ es no positiva para toda $t > 0$, el término dento de los corchetes cuadrados es no positivo, y por lo tanto de (4.9) y (4.16) tenemos

$$0 < \psi(B_{k+1}) \leq \psi(B_1) + ck + \sum_{j=1}^{k} \ln \cos^2 \theta_j, \tag{4.17}$$

donde podemos asumir, sin pérdida de generalidad, que la constante $c = M - \ln m - 1$ es positiva.

Notemos que, a partir del hecho de que $s_k = -\alpha_k B_k^{-1} \nabla f(x_k)$ de la iteración Cuasi-Newton, se tiene que $\cos\theta_k$ definido por (4.12) es el ángulo entre la dirección de descenso más rápido y la dirección de búsqueda, lo cual juega un papel importante en la convergencia global.

Supongamos que $\cos\theta_j \to 0$. Entonces existe $k_1 > 0$ tal que para todo $j > k_1$ tenemos

$$\ln \cos^2 \theta_j < -2c,$$

donde c es la constante definida arriba. Usando esta desigualdad en (4.17) tenemos que las siguientes relaciones son verdaderas para todo $k > k_1$:

$$\begin{aligned} 0 &< \psi(B_1) + ck + \sum_{j=1}^{k_1} \ln \cos^2 \theta_j + \sum_{j=k_1+1}^{k} (-2c) \\ &= \psi(B_1) + ck + \sum_{j=1}^{k_1} \ln \cos^2 \theta_j + 2ck_1 - ck. \end{aligned}$$

Sin embargo, el lado derecho es negativo para k grandes, dando una contradicción. Por lo tanto, existe una subsucesión de índices $\{j_k\}$ tales que $\{\cos\theta_{j_k}\} \geq \delta > 0$. Por el resultado de Zoutendijk (4.3), esto implica que el $\liminf \|\nabla f_k\| \to 0$. Dado que el problema es fuertemente convexo, el último límite es suficiente para probar que $x_k \to x_*$. □

Una extensión del análisis que se acaba de dar, muestra que la tasa de convergencia de las iteraciones es lineal. En particular, podemos probar que la sucesión $\|x_k - x_*\|$ converge a cero lo suficientemente rápido, de modo que

$$\sum_{i=1}^{\infty} \|x_k - x_*\| < \infty, \tag{4.18}$$

No probaremos este hecho, pero estableceremos que si (4.18) es válida, entonces la tasa de convergencia es superlineal.

4.2. Convergencia Superlineal del Método BFGS

El análisis de ésta sección hace uso de la caracterización de Dennis y Moré (4.20) de convergencia superlineal.

Teorema 4.3. *Supongamos que $f : \mathbb{R}^n \to \mathbb{R}$ es tres veces continuamente diferenciable. Considere la iteración $x_{k+1} = x_k + \alpha_k p_k$, donde p_k es una dirección de descenso, y α_k cumple las condiciones de Wolfe (1.9) con $c_1 \leq \frac{1}{2}$. Si la sucesión $\{x_k\}$ converge a un punto x_* tal que $\nabla f(x_*) = 0$ y $\nabla^2 f(x_*)$ es definida positiva, además si la dirección de búsqueda satisface*

$$\lim_{k\to\infty} \frac{\|\nabla f(x_k) + G(x_k) p_k\|}{\|p_k\|} = 0, \tag{4.19}$$

entonces

- *la longitud de paso* $\alpha_k = 1$ *es admisible para toda* $k > k_0$, *para algún* $k_0 \in \mathbb{N}$; *y*
- *si* $\alpha_k = 1$ *para toda* $k > k_0$, *entonces* $\{x_k\}$ *converge superlinealmente a* x_*.

Si p_k *es una dirección de búsqueda de un método Cuasi-Newton de la forma* (1.5), *entonces* (4.19) *es equivalente a*

$$\lim_{k\to\infty} \frac{\|(B_k - G(x_*))p_k\|}{\|p_k\|} = 0. \tag{4.20}$$

Por lo tanto, tenemos el sorprendente resultado de que la tasa de convergencia superlineal puede ser alcanzada, incluso si la sucesión de matrices B_k no converge a $\nabla^2 f(x_*)$; es suficiente que B_K se convierta en una aproximación cada vez más precisa de $\nabla^2 f(x_*)$ a lo largo de las direcciones de búsqueda p_k.

Una importante observación es que la condición (4.20) es necesaria y suficiente para la convergencia superlineal de los métodos Cuasi-Newton.

Teorema 4.4. *Supongamos que* $f : \mathbb{R}^n \to \mathbb{R}$ *es tres veces continuamente diferenciable. Considere la iteración* $x_{k+1} = x_k + p_k$ *(esto es,* $\alpha_k = 1$, *para todo k) y que* p_k *está dado por* (1.5). *Además supongamos que* $\{x_k\}$ *converge a un punto* x_* *tal que* $\nabla f(x_*) = 0$ *y* $G(x_*)$ *es definida positiva. Entonces* $\{x_k\}$ *converge superlinealmente a* x_* *si, y solo si, se cumple* (4.20).

Demostración. Primero demostraremos que (4.20) es equivalente a

$$p_K - p_k^N = o(\|p_k\|), \tag{4.21}$$

donde $p_k^N = -G(x_k)^{-1}\nabla f(x_k)$ es el paso de Newton. Asumiendo que (4.20) es válido, tenemos que

$$\begin{aligned} p_k - p_k^N &= G(x_k)^{-1}\left(G(x_k)p_k + \nabla f(x_k)\right) \\ &= G(x_k)^{-1}\left(G(x_k)p_k - B_k\right)p_k \\ &= O(\|(G(x_k) - B_k)\,p_k\|) \\ &= o(\|p_k\|), \end{aligned}$$

donde hemos usado el hecho de que $\|G(x_k)^{-1}\|$ es acotada superiormente para x_k suficientemente cercanos a x_*, dado que la matriz hessiana $G(x_*)$ es definida positiva. El inverso se sigue fácilmente si multiplicamos ambos lados de (4.21) por $G(x_k)$ y recordamos (1.5).

Además se tiene que

$$\begin{aligned} \|x_k + p_k - x_*\| &\le \|x_k + p_k^N - x_*\| + \|p_k - p_k^N\| \\ &= O(\|x_k - x_*\|^2) + o(\|p_k\|). \end{aligned}$$

Una sencilla manipulación de esta desigualdad revela que $\|p_k\| = O(\|x_k - x_*\|)$, así que se obtiene

$$\|x_k + p_k - x_*\| \le o(\|x_k - x_*\|),$$

dando el resultado de la convergencia superlineal. □

Para el resultado del método BFGS, haremos uso de la siguiente suposición.

Suposición 4.2

La matriz Hessiana es Lipschitz continua en x_*, esto es,

$$\|G(x) - G(x_*)\| \leq L\|x - x_*\|, \tag{4.22}$$

para todo x cercano a x_*, donde L es una constante positiva.

Comenzamos introduciendo las siguientes cantidades

$$\bar{s_k} = G_*^{\frac{1}{2}} s_k, \quad \bar{y_k} = G_*^{\frac{1}{2}} y_k, \quad \bar{B_k} = G_*^{\frac{1}{2}} B_k G_*^{\frac{1}{2}},$$

donde $G_* = G(x*)$ y x_* es un minimizador de f. Como en (4.12), definimos

$$\cos\bar{\theta_k} = \frac{\bar{s_k}^T \bar{B_k} \bar{s_k}}{\|\bar{s_k}\|\|\bar{B_k}\bar{s_k}\|}, \quad \bar{q_k} = \frac{\bar{s_k}^T \bar{B_k} \bar{s_k}}{\|\bar{s_k}\|^2},$$

similarmente definimos

$$\bar{M_k} = \frac{\|\bar{y_k}\|^2}{\bar{y_k}^T \bar{s_k}}, \quad \bar{m_k} = \frac{\bar{y_k}^T \bar{s_k}}{\bar{s_k}^T \bar{s_k}}.$$

Si a la fórmula de actualización BFGS (3.9) la pre y postmultiplicamos por $G_*^{-\frac{1}{2}}$ y agrupando términos apropiadamente, obtenemos

$$\bar{B}_{k+1} = \bar{B_k} - \frac{\bar{B_k}\bar{s_k}\bar{s_k}^T\bar{B_k}}{\bar{s_k}^T\bar{B_k}\bar{s_k}} + \frac{\bar{y_k}\bar{y_k}^T}{\bar{y_k}^T\bar{s_k}}.$$

Dado que esta expresión tiene la misma forma de la fórmula BFGS (3.9), se sigue del argumento que nos llevó a (4.16) que

$$\begin{aligned}\psi(\bar{B}_{k+1}) = \psi(\bar{B_k}) &+ (\bar{M_k} - \ln\bar{m_k} - 1) \\ &+ \left[1 - \frac{\bar{q_k}}{\cos^2\bar{\theta_k}} + \ln\frac{\bar{q_k}}{\cos^2\bar{\theta_k}}\right] \\ &+ \ln\cos^2\bar{\theta_k}.\end{aligned} \tag{4.23}$$

Recordando la ecuación (2.13), tenemos que

$$y_k - G_* s_k = (\bar{G_k} - G_*)s_k,$$

y por lo tanto

$$\bar{y_k} - \bar{s_k} = G_*^{\frac{-1}{2}}(\bar{G_k} - G_*)G_*^{\frac{-1}{2}}\bar{s_k}.$$

Por la suposición 4.2 y recordando la definición (2.12), tenemos

$$\|\bar{y_k} - \bar{s_k}\| \leq \|G_*^{-\frac{1}{2}}\|^2\|\bar{s_k}\|\|\bar{G_k} - G_*\| \leq \|G_*^{-\frac{1}{2}}\|^2\|\bar{s_k}\|L\epsilon_k,$$

donde ϵ_k esta definido por

$$\epsilon_k = \text{máx}\{\|x_{k+1} - x_*\|, \|x_k - x_*\|\}.$$

Por lo tanto hemos demostrado que

$$\frac{\|\bar{y_k} - \bar{s_k}\|}{\|\bar{s_k}\|} \leq \bar{c}\epsilon_k, \tag{4.24}$$

para alguna constante positiva $\bar{c}$. Esta desigualdad y (4.18) juegan un papel importante en la convergencia superlineal, como se verá a continuación.

Teorema 4.5. *Supongamos que f es dos veces continuamente diferenciable y que las iteraciones generadas por el algoritmo BFGS 1 converge a un minimizador x_*, para el cual la suposición 4.2 es válida. Suponga además que se cumple (4.18). Entonces x_k converge a x_* de manera superlineal.*

Demostración. De (4.24), tenemos de la desigualdad del triángulo que

$$\|\bar{y}_k\| - \|\bar{s}_k\| \leq \bar{c}\epsilon_k\|\bar{s}_k\|, \quad \|\bar{s}_k\| - \|\bar{y}_k\| \leq \bar{c}\epsilon_k\|\bar{s}_k\|,$$

así que

$$(1-\bar{c}\epsilon_k)\|\bar{s}_k\| \leq \|\bar{y}_k\| \leq (1+\bar{c}\epsilon_k). \tag{4.25}$$

Cuadrando (4.24) y usando (4.25), obtenemos

$$(1-\bar{c}\epsilon_k)^2\|\bar{s}_k\|^2 - 2\bar{y}_k{}^T\bar{s}_k + \|\bar{s}_k\|^2 \leq \|\bar{y}_k\|^2 - 2\bar{y}_k{}^T\bar{s}_k + \|\bar{s}_k\|^2 \leq \bar{c}^2\epsilon_k^2\|\bar{s}_k\|^2,$$

y por tanto

$$2\bar{y}_k{}^T\bar{s}_k \geq (1-2\bar{c}\epsilon_k + \bar{c}^2\epsilon_k^2 + 1 - \bar{c}^2\epsilon_k^2)\|\bar{s}_k\|^2 = 2(1-\bar{c}\epsilon_k)\|\bar{s}_k\|^2.$$

Se sigue de la definición de $\bar{m}_k$ que

$$\bar{m}_k = \frac{\bar{y}_k{}^T\bar{s}_k}{\|\bar{s}_k\|^2} \geq 1-\bar{c}\epsilon_k. \tag{4.26}$$

Combinando (4.25)y (4.26), obtenemos además que

$$\bar{M}_k = \frac{\|\bar{y}_k\|^2}{\bar{y}_k{}^T\bar{s}_k} \leq \frac{1+\bar{c}\epsilon_k}{1-\bar{c}\epsilon_k}. \tag{4.27}$$

Dado que $x_k \to x_*$, tenemos que $\epsilon_k \to 0$ y por lo tanto por (4.27) existe una constante positiva $c > \bar{c}$ tal que las siguientes desigualdades son verdaderas para k suficientemente grande:

$$\bar{M}_k \leq 1 + \frac{2\bar{c}}{1-\bar{c}\epsilon_k}\epsilon_k \leq 1 + c\epsilon_k, \tag{4.28}$$

Nuevamente hacemos uso de que la función $h(t) = 1-t+\ln t$ es no positiva. Por consiguiente, tenemos

$$\frac{-x}{1-x} - \ln 1 - x = h\left(\frac{1}{1-x}\right) \leq 0.$$

Ahora, para k suficientemente grande podemos asumir que $\bar{c}\epsilon_k < \frac{1}{2}$ y por lo tanto

$$\ln(1-\bar{c}\epsilon_k) \geq \frac{-\bar{c}\epsilon_k}{1-\bar{c}\epsilon_k} \geq -2\bar{c}\epsilon_k.$$

Esta relación y (4.26) implican que para k suficientemente grande, tenemos que

$$\ln \bar{m}_k \geq \ln(1-\bar{c}\epsilon_k) \geq -2\bar{c}\epsilon_k > -2c\epsilon_k. \tag{4.29}$$

Podemos deducir de (4.23), (4.28) y (4.29) que

$$0 < \psi(\bar{B}_{k+1}) \leq \psi(\bar{B}_k) + 3c\epsilon_k + \ln\cos^2\bar{\theta}_k + \left[1 - \frac{\bar{q}_k}{\cos^2\bar{\theta}_k} + \ln\frac{\bar{q}_k}{\cos^2\bar{\theta}_k}\right]. \tag{4.30}$$

Suponiendo esta expresión y haciendo uso de (4.18) tenemos que

$$\sum_{j=0}^{\infty}\left(\ln\frac{1}{\cos^2\bar{\theta}_j}-\left[1-\frac{\bar{q}_j}{\cos^2\bar{\theta}_k}+\ln\frac{\bar{q}_j}{\cos^2\bar{\theta}_k}\right]\right)\leq\psi(\bar{B}_0)+3c\sum_{j=0}^{\infty}\epsilon_j<+\infty.$$

Dado que el término en los corchetes cuadrados es no positivo, y como $\left(1/\cos^2\bar{\theta}_j\right)\geq 0$ para toda j, obtenemos los dos límites

$$\lim_{j\to\infty}\ln\frac{1}{\cos^2\bar{\theta}_j}=0,\quad \lim_{j\to\infty}\left(1-\frac{\bar{q}_j}{\cos^2\bar{\theta}_k}+\ln\frac{\bar{q}_j}{\cos^2\bar{\theta}_k}\right)=0,$$

lo cual implica que

$$\lim_{j\to\infty}\cos^2\bar{\theta}_j=1,\quad \lim_{j\to\infty}\bar{q}_j=1. \tag{4.31}$$

La parte esencial del resultado ha sido probado; solo tenemos que interpretar estos límites en términos de la caracterización de Dennis-Moré de convergencia superlineal.

Recordando (4.13) tenemos

$$\begin{aligned}\frac{\|G_*^{-1/2}(B_k-G_*)s_k\|^2}{\|G_*^{1/2}s_k\|^2}&=\frac{\|(\bar{B}_k-I)\bar{s_k}\|^2}{\|\bar{s_k}\|^2}\\&=\frac{\|\bar{B}_k\bar{s_k}\|^2-2\bar{s_k}^T\bar{B}_k\bar{s_k}+\bar{s_k}^T\bar{s_k}}{\bar{s_k}^T\bar{s_k}}\\&=\frac{\bar{q_k}^2}{\cos\bar{\theta_k}^2}-2\bar{q_k}+1.\end{aligned}$$

Ya que por (4.31) el lado derecho converge a 0, concluimos que

$$\lim_{k\to\infty}\frac{\|(B_k-G_*)s_k\|}{\|s_k\|}=0.$$

El límite (4.20) y el Teorema 4.3 implican que la longitud de paso unitaria $\alpha_k=1$ siempre va a satisfacer la condiciones de Wolfe cerca de la solución y por lo tanto la tasa de convergencia es superlineal. □

CAPÍTULO 5

RESULTADOS DE LA IMPLEMENTACIÓN

Se programó el Algoritmo 1 en lenguaje C, con el fin de minimizar la función de Rosenbrock de $\mathbb{R}^2$ a $\mathbb{R}$, la cual está dada por,

$$\begin{aligned} f &: \mathbb{R}^2 \to \mathbb{R} \\ (x_1, x_2) &\longmapsto 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2. \end{aligned}$$

Se hizo uso de esta función ya que el mínimo global está dentro de un valle plano, largo, estrecho y de forma parabólica. Encontrar el valle es trivial. Sin embargo, converger al mínimo global es difícil. Puede observarse fácilmente que el punto donde alcanza el valor mínimo es $x_* = (1,1)^T$, además $f(x_*) = 0$.

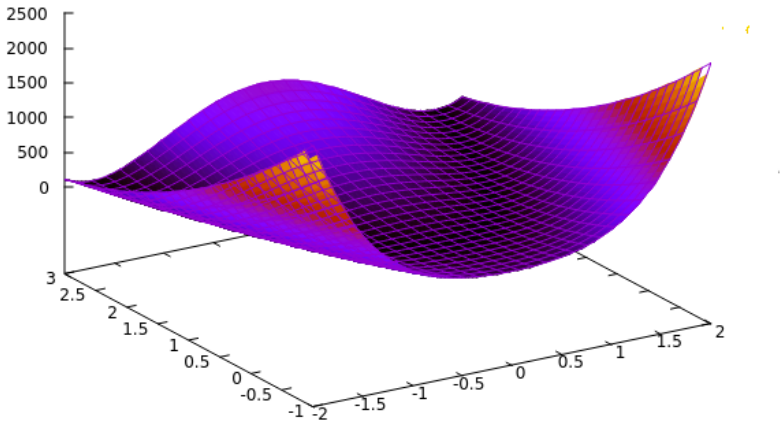


Figura 5.1: Función de Rosenbrock.

A continuación, se presentan los resultados al aplicar los métodos BFGS, DFP y Newton a la función anterior, donde la matriz inicial H_0 se consideró como la matriz identidad, el punto inicial como $x_0 = (-1.2,1)^T$ y una tolerancia de 10^{-5}. Puede observarse que el método DFP alcanzó el mínimo en un total de 36 iteraciones, mientras que el BFGS lo hizo en 35 iteraciones. El método de Newton lo hizo en tan solo 6 iteraciones, sin embargo, el método que obtuvo una mayor aproximación al valor mínimo fue el método BFGS.

```
La solucion con el metodo BFGS es: (1, 1)
f(1, 1) = 2.74564e-017.
Se usaron 35 iteraciones.

La solucion con el metodo DFP es: (1, 1)
f(1, 1) = 1.20714e-013.
Se usaron 36 iteraciones

La solucion con el metodo de Newton es: (0.999996, 0.999991)
f(0.999996, 0.999991) = 1.85274e-011
Se usaron 6 iteraciones

--------------------------------
Process exited after 0.1988 seconds with return value 122
Presione una tecla para continuar . . .
```

Figura 5.2: Resultados de los métodos para la función de Rosenbrock.

Además, el programa fue utilizado para minimizar la siguiente función,

$$\begin{aligned} f &: \mathbb{R}^2 \to \mathbb{R} \\ (x_1, x_2) &\longmapsto x_1^4 + (x_1 + x_2)^2 + \left(e^{x_2} - 1\right)^2. \end{aligned}$$

No es complicado observar que el punto donde alcanza su valor mínimo es $x_* = (0,0)^T$ y que $f(x_*) = 0$. Se presentan los resultados de aplicar el algoritmo a la función dada, donde nuestra matriz inicial nuevamente se consideró como la matriz identidad y la tolerancia nuevamente se consideró como 10^{-5}, pero en este caso se dieron dos puntos iniciales. En la primera imagen se observan los resultados al dar el punto inicial $x_0 = (1,1)^T$ y en la segunda imagen se presentan los resultados cuando $x_0 = (-1,3)^T$.

```
La solucion con el metodo BFGS es: (4.63846e-007, -8.40617e-008)
f(4.63846e-007, -8.40617e-008) = 1.51303e-013.
Se usaron 9 iteraciones.

La solucion con el metodo DFP es: (-1.02514e-007, 3.2446e-007)
f(-1.02514e-007, 3.2446e-007) = 1.54535e-013.
Se usaron 9 iteraciones

La solucion con el metodo de Newton es: (-1.49201e-006, 1.48791e-006)
f(-1.49201e-006, 1.48791e-006) = 2.21391e-012
Se usaron 6 iteraciones

--------------------------------
Process exited after 0.09524 seconds with return value 140
Presione una tecla para continuar . . .
```

Figura 5.3: Resultados de los métodos para el punto inicial $x_0 = (1,1)^T$.

Se observa en la imagen anterior que el método BFGS fue nuevamente el que mejor aproximó la solución, pero eso no es siempre ocurre. Con la misma función y matriz inicial, si ahora consideramos $x_0 = (-1,3)^T$ podemos observar que el método que mejor aproxima la solución es el de Newton.

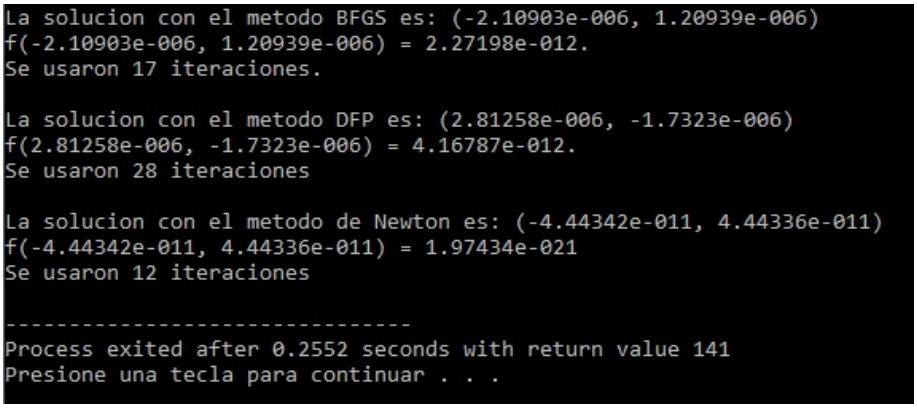


Figura 5.4: Resultados de los métodos para el punto inicial $x_0 = (-1,3)^T$.

Como una última observación, se ejecutó el programa para esta misma función con la misma matriz inicial y la tolerancia de 10^{-5} como se hizo anteriormente, sin embargo se dio el punto inicial $x_0 = (-10,17)^T$ y el único método que obtuvo la solución fue el BFGS, el cual lo hizo en 54 iteraciones, como se muestra a continuación.

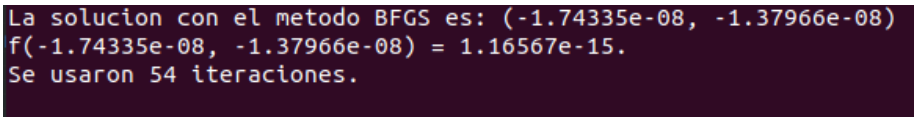


Figura 5.5: Resultados de los métodos para el punto inicial $x_0 = (-10,17)^T$.

CAPÍTULO 6

CONCLUSIONES

Aunque el método de Newton goza de una tasa de convergencia cuadrática, no siempre es la mejor opción al momento de resolver un problema de optimización sin restricciones. Tal como se pudo observar a lo largo del presente trabajo, existen ocasiones en las que el método BFGS proporciona una mejor aproximación de la solución y el número de iteraciones no es grande, además no es necesario hacer demasiadas evaluaciones como en el método de Newton, ya que en este se tiene evaluar el punto en la función, el gradiente y en la matriz hessiana, en cambio los métodos Cuasi-Newton solo hacen evaluaciones en la función y en el gradiente y en lugar de hacer evaluaciones en la matriz hessiana, simplemente actualiza una matriz simétrica y definida positiva mediante multiplicaciones de matrices, vectores y escalares, lo cual tiene un menor costo computacional. Otro de los inconvenientes que tiene el método de Newton es que nada garantiza que en todas las iteraciones la matriz hessiana es definida positiva, por lo cual puede que el método no converja, en cambio los métodos Cuasi-Newton como matriz inicial requieren una que sea definida positiva y aseguran, mediante las actualizaciones, que las consecutivas también lo serán. Por último, el método de Newton no garantiza convergencia si el punto inicial no es cercano a la solución, como pudo observarse en el último ejemplo presentado en los resultados de la implementación, donde se consideró un punto más alejado del minimizador. El único método que logró obtener la solución fue el BFGS, ya que, como se planteó anteriormente, tiene una convergencia global.

Podemos concluir el trabajo realizado haciendo las siguientes observaciones.

- El método BFGS logra obtener la solución aunque el punto inicial no sea cercano a la solución.
- Aunque la tasa de convergencia del método de Newton es cuadrática, la de los métodos BFGS y DFP es superlineal, proporcionando así un método que converge a la solución en un número bastante aceptable de iteraciones.
- En problemas donde la matriz hessiana llega a ser muy grande, el método de Newton requiere de un número bastante grande de evaluaciones por iteración. Por ejemplo, si consideramos una función de $\mathbb{R}^{100}$ a $\mathbb{R}$, la matriz hessiana será de tamaño 100×100, por lo cual requerirá un total de 10,000 evaluaciones en la matriz más las evaluaciones de la función objetivo además de 100 evaluaciones del gradiente y las correspondientes multiplicaciones entre matrices, vectores y escalares en una sola iteración, haciendo el método bastante costoso computacionalmente. En cambio, los métodos Cuasi-Newton estudiados en este trabajo no requieren las evaluaciones en la matriz hessiana y hace multiplicaciones matriz-vector y sumas de matrices, las cuales son sencillas de realizar.

- El método de Newton no asegura que la dirección obtenida sea de descenso y por tanto no podemos asegurar que el método converge a la solución, en cambio los métodos Cuasi-Newton requieren de entrada una matriz definida positiva, asegurando que las consecutivas también lo serán y por tanto las direcciones generadas serán de descenso.

Todo lo anterior hace de los métodos Cuasi-Newton una importante herramienta en la resolución de problemas de optimización, incluso puede llegar a dar mejores resultados que el método de Newton.

APÉNDICE A

CONCEPTOS DE CÁLCULO

Definición A.1. *Sean $D \subset \mathbb{R}^n$ y $f : D \to \mathbb{R}^m$. Decimos que el **límite** de f cuando x tiende a x_0 es L, lo cual escribimos como*

$$\lim_{x \to x_0} f(x) = L,$$

si para todo $\varepsilon > 0$, existe $\delta > 0$ tal que si $x \in D$ y $\|x - x_0\| < \delta$, entonces $\|f(x) - L\| < \varepsilon$.

Decimos que f es continua en x_0, si $x_0 \in D$ y lo anterior es válido para $L = f(x_0)$.

Decimos que f es continua en su dominio D, si f es continua para todo $x \in D$.

Definición A.2. *Sean $D \subset \mathbb{R}^n$ y $f : D \to \mathbb{R}^m$. Decimos que f es $Lipschitz\ continua\ en\ D$ si existe una constante $M > 0$ tal que para cualesquiera dos puntos $x_1, x_2 \in D$ tenemos que:*

$$\|f(x_1) - f(x_2)\| \leq M\|x_1 - x_2\|.$$

Definición A.3. *Sea $\phi : \mathbb{R} \to \mathbb{R}$ una función de variable real. La primera derivada $\phi'(\alpha)$ está definida como*

$$\frac{d\phi}{d\alpha} = \lim_{\varepsilon \to 0} \frac{\phi(\alpha + \varepsilon) - \phi(\alpha)}{\varepsilon}.$$

De forma análoga, la segunda derivada de ϕ está dada por

$$\frac{d^2\phi}{d\alpha^2} = \lim_{\varepsilon \to 0} \frac{\phi'(\alpha + \varepsilon) - \phi'(\alpha)}{\varepsilon}.$$

Decimos que ϕ es una función suave si estos límites existen.

Definición A.4. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función. Si $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, al vector de las primeras derivadas de f, al cual llamaremos el gradiente de f, está definido por:*

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right),$$

donde $\frac{\partial f}{\partial x_i}(x) = \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon e_i) - f(x)}{\varepsilon}$, con $e_i \in \mathbb{R}^n$ es el vector cuya entrada $i - ésima$ es igual a 1 y las demás entradas son 0.

La matriz de las segundas derivadas parciales de f es conocida como la matriz Hessiana, y está definida como

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}.$$

Decimos que f es diferenciable si todas las derivadas parciales de f existen, y continuamente diferenciable si además éstas derivadas son funciones continuas de x. Similarmente, f es dos veces diferenciable si todas las segundas derivadas parciales existen, y dos veces continuamente diferenciable si además las segundas derivadas parciales son funciones continuas.

Nótese que si f es dos veces continuamente diferenciable, entonces la matriz Hessiana es simétrica, dado que para todas $i, j = 1, \ldots, n$ se tiene que

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x).$$

Definición A.5. *Sea F un subconjunto de $\mathbb{R}^n$ y supóngase que $\{x_k\}_{k\in\mathbb{N}}$ es una sucesión de puntos que pertenecen a F, decimos que la sucesión $\{x_k\}$* ***converge*** *a un punto x, lo cuál se escribe como $\lim_{k\to\infty} x_k = x$, si para todo $\varepsilon > 0$ existe algún $K \in \mathbb{N}$ tal que*

$$\|x_k - x\| < \varepsilon,$$

para todo $k \geq K$.

Definición A.6. *Sea F un subconjunto de $\mathbb{R}^n$, decimos que F es un conjunto* ***acotado*** *si existe $M > 0$ tal que $\|x\| \leq M$, para todo $x \in F$.*

APÉNDICE B

CONCEPTOS DE ÁLGEBRA

Teorema B.1. *Dado V espacio vectorial sobre el campo $\mathbb{K}$, con $dim(V) = n$, y un conjunto de vectores l.i.; $X = \{v_1, \ldots, v_r\}$, con $r < n$, entonces existen vectores $v_{r+1}, \ldots, v_n$, tales que el conjunto $\{v1, \ldots, vn\}$ es base de V.*

La demostración se encuentra en [7].

A lo largo del presente trabajo, hacemos uso constante de vectores y matrices sobre el campo $\mathbb{R}$. Los vectores son usualmente denotados por letras minúsculas y las matrices por letras mayúsculas. El espacio de los vectores cuyas entradas son números reales los denotamos por $\mathbb{R}^n$, mientras que el espacio de las matrices de tamaño $m \times n$ con coeficientes en el campo $\mathbb{R}$, lo denotamos por $M_{m\times n}(\mathbb{R})$. Si $A \in A \in M_{n\times n}(\mathbb{R})$, al elemento que se encuentra en la i-ésima fila y en la j-ésima columna lo denotaremos por a_{ij}.

Definición B.1. *A la matriz $A \in M_{n\times n}(\mathbb{R})$ que cumple*

$$a_{ij} = \begin{cases} 1 & \text{si } i = j; \\ 0 & \text{si } i \neq j, \end{cases}$$

le llamaremos la matriz identidad de tamaño n y la denotamos por $A = I_n$.

Definición B.2. *Si $A \in M_{n\times n}(\mathbb{R})$ y existe una matriz $B \in M_{n\times n}(\mathbb{R})$ tal que*

$$AB = BA = I_n,$$

decimos que A es invertible (o no-singular) y denotamos $B = A^{-1}$. Si no existe tal matriz, decimos que A es singular.

Definición B.3. *Si $A \in M_{m\times n}(\mathbb{R})$, definimos la traspuesta de A como la matriz $B \in M_{n\times m}(\mathbb{R})$, cuya entrada ij es igual a la entrada ji de la matriz A, es decir,*

$$b_{ij} = a_{ji}.$$

Lo cual denotamos por $B = A^T$. Además decimos que una matriz $A \in M_{n\times n}(\mathbb{R})$ es simétrica si $A^T = A$.

Definición B.4. *Una matriz $A \in M_{n\times n}(\mathbb{R})$ será semidefinida-positiva si para todo $x \in \mathbb{R}^n$ se tiene que*

$$x^T Ax \geq 0.$$

Si la desigualdad anterior se cumple estrictamente, es decir que $x^T Ax > 0$, decimos que la matriz A es definida positiva.

Definición B.5. *Un número real λ es un valor propio de una matriz $A \in M_{n\times n}(\mathbb{R})$, si existe un vector $x \neq 0$ tal que*

$$Ax = \lambda x.$$

Al vector x le llamaremos el vector propio asociado a λ.

Teorema B.2. *Una matriz $A \in M_{n\times n}(\mathbb{R})$ es no-singular si ninguno de sus valores propios es 0.*

Teorema B.3. *Los valores propios de una matriz simétrica, son números reales, mientras que las matrices que no son simétricas pueden tener valores propios complejos.*

Lema B.0.1. *Si una matriz A es simétrica y definida positiva, sus valores propios serán todos números reales positivos.*

Teorema B.4. *Si A es una matriz simétrica, cuyos valores propios son $\lambda_1, \lambda_2, \ldots, \lambda_n$ y sus respectivos vectores propios son $q_1, q_2, \ldots, q_n$, entonces A tiene la siguiente descomposición espectral*

$$A = \sum_{i=1}^{n} \lambda_i q_i q_i^T.$$

Definición B.6. *Definimos la traza de una matriz $A \in M_{n\times n}(\mathbb{R})$, como*

$$traza(A) = \sum_{i=1}^{n} a_{ii}.$$

Teorema B.5. *Si los valores propios de una matriz $A \in M_{n\times n}(\mathbb{R})$ son denotados por $\lambda_1, \lambda_2, \ldots, \lambda_n$, entonces*

$$traza(A) = \sum_{i=1}^{n} \lambda_i.$$

Definición B.7. *Sea una matriz $A \in M_{n\times n}(\mathbb{R})$, cuyos valores propios son $\lambda_1, \lambda_2, \ldots, \lambda_n$, definimos el determinante de A como*

$$det(A) = \prod_{i=1}^{n} \lambda_i.$$

El determinante de una matriz tiene importantes propiedades:

- $det(A) = 0$ si, y solo si, A es singular;
- $det(AB) = det(A)det(B)$;
- $det(A^{-1}) = \frac{1}{det(A)}$.

Teorema B.6. *Sean $x, y, u, v \in \mathbb{R}^n$, se tiene que*

$$det(I_n + xy^T) = 1 + y^T x; \tag{B.1a}$$

$$det(I_n + xy^T + uv^T) = (1 + y^T x)(1 + u^T v) - (x^T v)(y^T u). \tag{B.1b}$$

Demostración. Para la demostración de (B.1a) supongamos que $x \neq 0$, por el teorema de completitud de bases podemos encontrar $w_1, w_2, \ldots, w_{n-1}$ tales que el conjunto $\{x, w_1, w_2, \ldots, w_{n-1}\}$ es linealmente independiente, por lo que la matiz $Q = [x, w_1, w_2, \ldots, w_{n-1}]$ es invertible, dado que el determinante de Q es distinto de 0. Además se sigue que $x = Qe_1$, donde $e_1 = (1,0,0,\ldots,0)^T \in \mathbb{R}^n$. Si definimos $y^T Q = (z_1, z_2, \ldots, z_n)$, entonces $z_1 = (y^T Q)e_1 = y^T(Qe_1) = y^T x$. Finalmente tenemos que

$$\begin{aligned} det(I_n + xy^T) &= 1 \cdot det(I_n + xy^T) \\ &= det(Q^{-1}Q) det(I_n + xy^T) \\ &= det\left(Q^{-1}(I_n + xy^T)Q\right) \\ &= det(I_n + Q^{-1}xy^T Q) \\ &= det(I_n + e_1 y^T Q) \\ &= 1 + z_1 \\ &= 1 + x^T y. \end{aligned}$$

La demostración de (B.1b) se hace de forma análoga a la de (B.1a). □

Teorema B.7 (Factorización de Cholesky)**.** *Si $A \in M_{n\times n}(\mathbb{R})$ simétrica definida positiva, entonces existe al menos una matriz $B \in M_{n\times n}(\mathbb{R})$ triangular inferior (es decir, $b_{ij} = 0$ si $j > i$), tal que $A = BB^T$. Además, se puede imponer que $b_{ii} > 0$ para todo $i = 1, \ldots, n$, y en tal caso la factorización anterior es única.*

El recíproco del Teorema de factorización de Cholesky es también cierto.

Teorema B.8 (Fórmula de Sherman-Morrison-Woodbury)**.** *Si A es una matriz de tamaño $n \times n$ no singular y tenemos que*

$$\bar{A} = A + ab^T,$$

donde $a, b \in \mathbb{R}^n$, entonces $\bar{A}$ es no singular y además

$$\bar{A}^{-1} = A^{-1} - \frac{A^{-1}ab^T A^{-1}}{1 + b^T A^{-1} a}. \tag{B.2}$$

APÉNDICE C

CONVEXIDAD

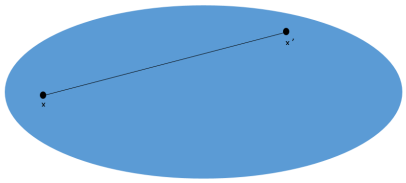
La convexidad, tanto de funciones como de conjuntos, juegan un papel importantísimo en la optimización, que aunque parezcan dos conceptos completamente ajenos, juntos dan lugar a grandes resultados para la existencia del máximo (o mínimo) de una función con valores reales.

Definición C.1. *El conjunto $C \subset \mathbb{R}^n$ se dice que es convexo si $\alpha x + (1-\alpha)x'$ pertenece a C, para cualesquiera x, x' en C y $\alpha \in [0,1]$.*

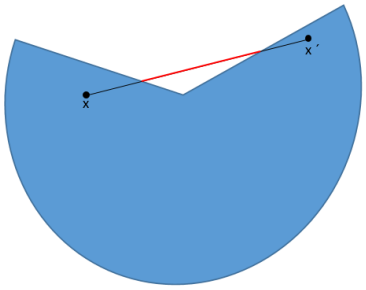
Geométricamente, esto significa que el segmento de linea

$$[x, x'] := \{\alpha x + (1-\alpha)x' : \ \alpha \in [0,1]\},$$

está completamente contenida en C, siempre que $x, x' \in C$.



(a) Conjunto Convexo.



(b) Conjunto No Convexo.

Definición C.2. *Sea C un conjunto convexo no vacío en $\mathbb{R}^n$. Una función $f : C \to \mathbb{R}$ se dice que es convexa en C cuando para todos los pares $(x, x') \in C \times C$ y para toda $\alpha \in [0,1]$, se tiene que*

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha) f(x'). \tag{C.1}$$

Geométricamente esto significa que la función aplicada al segmento de recta $[x, x']$ *está por debajo de el segmento de recta* $[f(x), f(x')]$. *Por ejemplo, la función* $sin(\cdot)$ *es no convexa en el intervalo* $(0, \pi)$ *y es convexa en el intervalo* $[\pi, 2\pi]$.

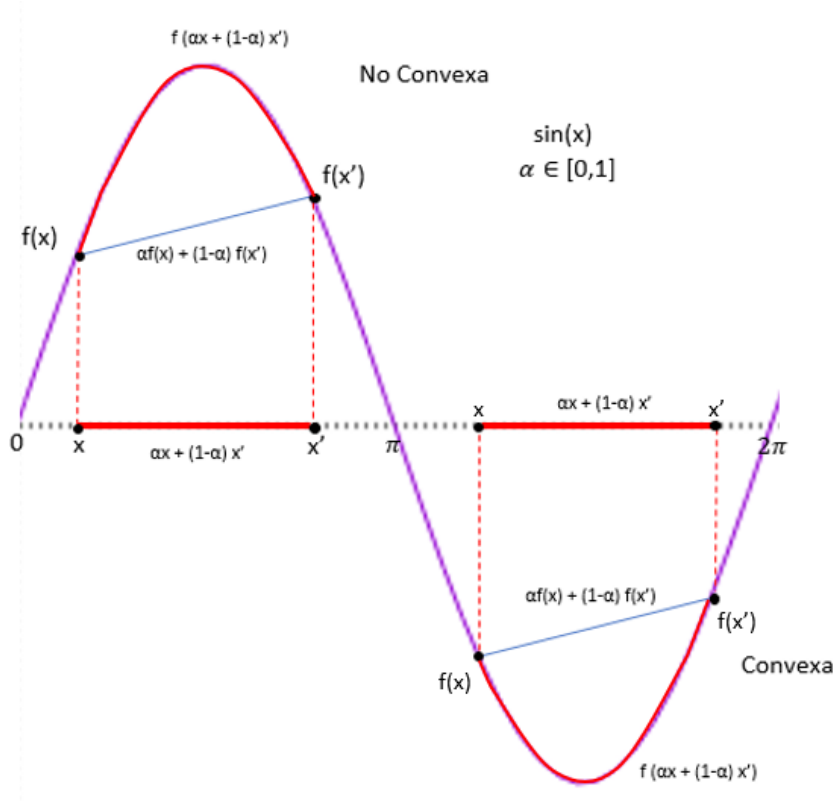


Figura C.1: Función no convexa en el intervalo $(0, \pi)$ y convexa en el intervalo $[\pi, 2\pi]$.

Decimos que f *es estrictamente convexa cuando la desigualdad anterior se cumple como desigualdad estricta. Una propiedad aún más fuerte es cuando existe* $c > 0$ *tal que*

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha) f(x') - \frac{1}{2} c\alpha(1-\alpha)\|x - x'\|^2, \tag{C.2}$$

para todo $(x, x') \in C \times C$ *y todo* $\alpha \in (0,1)$. *En este caso decimos que* f *es fuertemente convexa en* C.

Definición C.3. *Una función* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *no idénticamente* $+\infty$, *se dice convexa cuando para todo* $(x, x') \in \mathbb{R}^n \times \mathbb{R}^n$ *y para todo* $\alpha \in [0,1]$ *se tiene que*

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha) f(x'),$$

considerada como una desigualdad en $\mathbb{R} \cup \{+\infty\}$.

El conjunto de tales funciones es denotada por $Conv\, \mathbb{R}^n$.

Definición C.4. *Dado $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, no idénticamente $+\infty$, la epigráfica de f es el conjunto no vacío*

$$epi(f) := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : \ r \geq f(x)\}.$$

Como un primer resultado que relaciona las funciones convexas con los conjuntos convexos se tiene el siguiente teorema.

Teorema C.1. *Sea $f : \mathbb{R}^n \to \mathbb{R}$. La función f es convexa si, y solo si, $epi(f)$ es un conjunto convexo en el espacio $\mathbb{R}^n \times \mathbb{R}$.*

Teorema C.2. *Sean S un subconjunto convexo, no vacío, abierto de $\mathbb{R}^n$ y $f : S \to \mathbb{R}$ diferenciable en S. Entonces f es una función convexa si, y solo si, para cada $x_1, x_2 \in S$, tenemos que*

$$\left(\nabla f(x_2) - \nabla f(x_1)\right)^T (x_2 - x_1) \geq 0.$$

Similarmente, f es estrictamente convexa si, y sólo si, para cada $x_1, x_2 \in S$, com $x_1 \neq x_2$, tenemos que

$$\left(\nabla f(x_2) - \nabla f(x_1)\right)^T (x_2 - x_1) \geq 0.$$

APÉNDICE D

LA NOTACIÓN DE LA O GRANDE.

La notación "de la O grande" es usada en ciencias de la computación para describir el funcionamiento o complejidad del algoritmo. La O grande describe específicamente el peor escenario posible y puede ser usado para describir el tiempo de ejecución requerido o el espacio usado (por ejemplo en memoria o en el disco) por el algoritmo.

$O(1)$ Describe un algoritmo que siempre se ejecutará en el mismo tiempo (o espacio) independientemente de el tamaño del conjunto de datos de entrada.

$O(n)$ Describe un algoritmo cuyo funcionamiento crecerá linealmente y en proporción directa a el tamaño del conjunto de datos de entrada.

$O(n^2)$ Representa un algoritmo cuyo funcionamiento es directamente proporcional al cuadrado del tamaño de el conjunto de datos de entrada. Éste es común con algoritmos que involucren iteraciones anidadas sobre el conjunto de datos. Iteraciones anidadas más profundas resultarán en $O(n^3)$, $O(n^4)$, etc.

$O(2^n)$ Denota un algoritmo cuyo crecimiento se duplica con cada adición al conjunto de datos de entrada. La curva de crecimiento de una función $O(2^n)$ es exponencial, empezando muy lento, y después creciendo muy rápidamente.

Algunos ejemplos de cada tipo de algoritmo se encuentran en [2].

Aunque la "O grande" es la notación más usada para comparar algoritmos, existen otras notaciones.

- **Omega grande** Ω Representa la cota inferior, así que no es de mucha utilidad.

  $\Omega(n)$ significa que el algoritmo se ejecutará en al menos n pasos.

- **Theta grande** Θ Representa la cota inferior y la cota superior de un algoritmo

  $\Theta(n)$ significa que el algoritmo se ejecutará en al menos n y en a lo más n pasos.

- **O grande** O Representa la cota superior. Es la notación más usada ya que representa el peor escenario posible. Podemos garantizar que el algoritmo se detendrá en un tiempo determinado, incluso puede parar antes pero nunca se detendrá después.

  $O(n)$ significa que el algoritmo ejecutará en lo más n pasos.

Referencias

[1] Bazaraa, M., Sherali H., & Shetty, C. (2006). *Nonlinear Programming. Theory and Algorithms*. New Jersey: Wiley-Interscience.

[2] Bell, R.(2009). *beginner's guide to Big O notation*. Recuperado de https://rob-bell.net/2009/06/a-beginners-guide-to-big-o-notation/.

[3] El Gabry, O. (2016). *Big Scary O Notation*. Recuperado de https://medium.com/omarelgabrys-blog/the-big-scary-o-notation-ce9352d827ce.

[4] Fletcher, R. (1987).*Practical Methods of Optimization*. England: John Wiley & Sons.

[5] López, G. (2005). *Curso de programación lineal*. Puebla: Benemérita Universidad Autónoma de Puebla.

[6] Nocedal J. & Wright S. (1999). *Numerical Optimization*. New York: Springer-Verlag.

[7] *Semana 7: Espacios Vectoriales*. (s.f.). Departamento de Ingeniería Matemática. Universidad de Chile. Recuperado de https://docencia.dim.uchile.cl/algebra_lineal/material/tutoria_semana/Semana07_alglin.pdf

[8] *Práctica 1bis: Recordatorio de descomposiciones matriciales LU y Cholesky*. (s.f.). Recuperado de: https://personal.us.es/pmr/images/pdfs/1314gm-cnii-practica1bis.pdf