Tarea de Modelos de Optimización.

Nombre: Ronald Cabrera Martínez.
Grupo: 311.

Función objetivo a minimizar:

$$f(x,y) = -\, tan^2(0.5sen(x) + 0.5sen(y)).$$

Estamos en presencia de un problema de optimización no lineal (uso de funciones trigonométricas) sin restricciones y con variables continuas.

Analicemos ahora la existencia o no de solución:

Recordando…..

Por el **teorema de Weierstrass**: Si f es **continua** y el conjunto factible S es **no vacío, cerrado y acotado**, entonces existe al menos un mínimo global.

El dominio de nuestro problema es R^2 y R^2 no es acotado por lo que Weierstrass no funciona pero notemos que podemos hacer una composición de funciones:

$$f(x,y) = -\, tan^2(0.5sen(x) + 0.5sen(y)) := g(h(x,y))$$

donde $h(x,y) = 0.5sen(x) + 0.5sen(y) \quad y \quad g(u) = -\, tan^2(u).$

Ahora para cualquier x,y ∈ R se tiene que $sen(x), sen(y)$ ∈ [-1,1] por tanto h(x,y)=[-1,1] y con esto tenemos que el conjunto [-1,1] es no vacío,

cerrado y acotado, además $g(u) = \ - tan^2(u)$ es continua en [-1,1] ya que ($u <= 1 < \pi/2$) por lo que se cumple Weierstrass para g y podemos garantizar la existencia de solución.
Ahora como sabemos que g alcanza su mínimo y su máximo en K=[-1,1]. Sea u* ∈ K donde g alcanza el mínimo , como h(R^2) = K , existe (x*,y*) = u*. Para ese par se cumple f(x*,y*) = g(u*), de modo que f alcanza su mínimo en R^2.

Como la función $tan\ (u)$ es creciente , la función $tan^2(u)$ va a crecer de igual manera para valores tanto negativos como positivos de u, o sea para |u| , así que en el intervalo [-1,1] la función $tan^2(u)$ alcanza su máximo en los extremos u=±1. Por lo que $g(u) = \ - tan^2(u)$ alcanza su mínimo en u=±1.

Alcanzando estos valores de u :

. $u = 1$ cuando $sen(x) = sen(y) = 1$ o sea cuando,
$x = y = (\pi/2) + 2k\pi$.

. $u =- 1$ cuando $sen(x) = sen(y) =- 1$ o sea cuando,
$x = y = (3\pi/2) + 2k\pi$.

Por tanto el problema si tiene solución , de hecho por la forma de x e y existen infinitos óptimos globales y el valor de evaluar estos en la función nos da ≈−2.4255.

<u>Analicemos ahora la convexidad:</u>

Para verificar que nuestra función sea convexa debemos ver que su hessiana sea semidefinida positiva.
Sea $u(x, y) = 0.5sen(x) + 0.5sen(y)$, entonces
$f(x, y) = \ - tan^2(u)$.

Derivadas parciales de primer orden:

Primero derivamos respecto a u:

$$df/du = -2tan(u)\cdot sec^2(u).$$

Ahora, como:

$$\partial u/\partial x = 0.5cos(x), \quad \partial u/\partial y = 0.5cos(y),$$

tenemos:

$fx = \partial f/\partial x = -2tan(u)sec^2(u)\cdot 0.5cos(x) = -tan(u)sec^2(u)cos(x),$
$fy = -tan(u)sec^2(u)cos(y).$

Derivadas parciales de segundo orden:

Segunda derivada respecto a x:

$$fx = -cos(x)\,tan(u)sec^2(u).$$

Sea

$$g(u) = tan(u)sec^2(u).$$

Entonces

$$fx = -cos(x)\,g(u).$$

Derivando respecto a x:

$$fxx = sin(x)\,g(u) - cos(x)\,g'(u)\cdot 0.5cos(x).$$

Segunda derivada respecto a y:

Por simetría:

$$fyy = sin(y)\,g(u) - cos(y)\,g'(u)\cdot 0.5cos(y).$$

Derivada mixta:

$$fxy = \partial(-cos(x)\,g(u))/\partial y = -cos(x)\,g'(u)\cdot 0.5cos(y).$$

Y por simetría $fxy = fyx$.

Derivada de g(u):

$$g(u) = tan(u)sec^2(u).$$

Derivamos:

$$g'(u) = sec^2(u)sec^2(u) + tan(u)\cdot 2sec^2(u)tan(u).$$
$$g'(u) = sec4(u) + 2tan^2(u)sec^2(u).$$

La hessiana queda:

$$Hf(x, y) = \begin{bmatrix} sin(x)\, g(u) - 0.5cos^2(x)\, g'(u) & -0.5cos(x)cos(y)\, g'(u) \\ -0.5cos(x)cos(y)\, g'(u) & sin(y)\, g(u) - 0.5cos^2(y)\, g'(u) \end{bmatrix}$$

Nuestra hessiana será semidefinida positiva , semidefinida negativa o indefinida en dependencia del punto en que se evalúe, por ejemplo :
En $(x, y) = (0, 0)$:
$u = 0u = 0,\ entonces\ g(0) = 0g(0) = 0,\ g'(0) = 1$:
La Hessiana se reduce a

$$Hf(x, y) = \begin{bmatrix} -1/2 & -1/2 \\ -1/2 & -1/2 \end{bmatrix}$$

Sus autovalores son {0,−1}, por lo que es semidefinida negativa.
Con esto concluimos que no estamos en presencia de un problema convexo.

Existen extremos locales además de los óptimos globales ?

Notemos que como estamos trabajando en un problema irrestricto los extremos locales serán aquellos cuyo $\nabla f(x\ *) = 0$.
Las condiciones de primer orden dan:

$$fx =- cos(x)\, tan(u)sec^2(u), \qquad fy =- cos(y)\, tan(u)sec^2(u)$$

Por tanto, los puntos críticos aparecen cuando:
. $u = 0u = 0 \Leftrightarrow sinx + siny = 0$.
. $o\ bien\ cosx = 0\ y\ cosy = 0\ simultáneamente \Leftrightarrow x = \pi/2 + k\pi,\ y = \pi/2 + m\pi$.

Estos puntos con gradiente 0 cumplen condiciones necesarias, las condiciones suficientes nos la da comprobar la hessiana de cada uno de ellos : Hessiana definida positiva → mínimo local , Hessiana definida negativa → máximo local y Hessiana indefinida → punto de silla.

Descripción de los algoritmos utilizados:

Método de Newton

a) Fundamentación teórica:

El método de **Newton** es un algoritmo de optimización **de segundo orden** que aprovecha información de curvatura (la **Hessiana**) para aproximar la función objetivo mediante un modelo cuadrático local.

En cada iteración, se construye una **aproximación cuadrática de segundo orden** de la función alrededor del punto actual x_k:

$$m_k(p) = f(x_k) + \nabla f(x_k)^T p + 0.5 p^T \nabla^2 f(x_k) p$$

El objetivo es encontrar el desplazamiento p_k que minimiza $m_k(p)$. Derivando e igualando a cero:

$$\nabla m_k(p) = \nabla f(x_k) + \nabla^2 f(x_k) p = 0$$

Se obtiene el **sistema lineal de Newton**:

$$\nabla^2 f(x_k) p_k = -\nabla f(x_k)$$

y la **actualización del punto**:

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

donde $\alpha_k \in (0,1]$ es el tamaño de paso (normalmente determinado por una **búsqueda en línea** tipo *backtracking* para asegurar convergencia global).

b) Propiedades:

. **Orden de convergencia:** cuadrático (muy rápido) cerca del óptimo.

. **Requiere:** el gradiente $\nabla f(x_k)$ y la Hessiana $\nabla^2 f(x_k)$.

. **Dirección de búsqueda:** p_k es dirección de **descenso** si la Hessiana es **definida positiva**.

. **Costo computacional:** elevado si la dimensión n es grande (cálculo e inversión de Hessiana).

c) Ventajas

. Convergencia muy rápida cuando se está cerca de un punto mínimo.

. Aprovecha información de curvatura (segundas derivadas).

. Precisión elevada en funciones suaves.

d) Desventajas

Si la Hessiana **no es definida positiva**, el método puede **divergir** o dirigirse hacia un máximo o punto de silla.

El costo de invertir la Hessiana en cada paso puede ser alto.

No es adecuado si la función tiene regiones planas o mal condicionadas.

e) Adaptación al problema

En este trabajo, el método de Newton se aplicó a la función:

$$f(x, y) = - tan^2(0.5sen(x) + 0.5sen(y))$$

que es **suave** y **periódica**, pero **no convexa** (tiene múltiples mínimos y máximos).

Por ello se implementó una **versión amortiguada (damped Newton)** con **búsqueda en línea** para controlar el tamaño de paso α_k, garantizando que $f(x_{k+1})<f(x_k)$ y evitando saltos hacia regiones donde tan(u) diverge.

Gracias a esto, el método logra converger rápidamente a un mínimo global (valor $f^* \approx -2.4255$) cuando el punto inicial está razonablemente cerca de un valle de la función.

Método Quasi-Newton (BFGS)

a) Fundamentación teórica

Los métodos **Quasi-Newton** son una mejora del método de Newton.
El objetivo es obtener la **dirección de descenso** de Newton sin necesidad de calcular ni invertir explícitamente la Hessiana.

La idea básica es mantener una **aproximación B_k**(o su inversa $H_k = B_k^{-1}$) de la Hessiana, actualizándola iterativamente a partir de la información de gradientes entre pasos consecutivos.

Sea:

$$s_k = x_{k+1} - x_k \,, \qquad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

El método **BFGS** (Broyden–Fletcher–Goldfarb–Shanno) actualiza la **matriz inversa de la Hessiana** mediante:

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

donde $\rho_k = 1/(y_k^T s_k)$.

La dirección de búsqueda se obtiene como:

$$p_k = -H_k \nabla f(x_k)$$

y el punto se actualiza con búsqueda en línea:

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

b) Propiedades

. Convergencia: superlineal (más rápida que el gradiente descendente, algo más lenta que Newton).

. No requiere segundas derivadas exactas, solo gradientes.

. Garantiza que H_k permanezca **simétrica y definida positiva** si $y_k^T s_k > 0$.

. Dirección siempre descendente, por lo que es muy estable.

c) Ventajas

. Evita el cálculo e inversión directa de la Hessiana.

. Estable incluso en funciones no convexas.

. Comportamiento muy robusto: funciona bien con diferentes puntos iniciales.

. Generalmente converge en pocas iteraciones.

d) Desventajas

. Necesita almacenamiento de la matriz H_k (de tamaño n×n).

Puede degradarse si los gradientes son inestables o si $y_k^T s_k$ se aproxima a cero.

Ligeramente más lento que Newton en funciones estrictamente convexas.

e) Adaptación al problema

Para la función estudiada, f(x,y), el método BFGS es **particularmente adecuado** porque:

. La función es **suave** y tiene derivadas exactas conocidas.

. Presenta **múltiples mínimos locales**, y BFGS es capaz de evitar divergencias gracias a su propiedad de mantener H_k definida positiva.

. No requiere calcular la Hessiana completa, lo que reduce errores numéricos por las regiones donde tan(u) cambia rápidamente.

En las pruebas, BFGS alcanzó el mismo valor óptimo que Newton ($f^* \approx -2.4255$) pero con **mayor estabilidad y menos sensibilidad al punto inicial**.

Comparación conceptual entre ambos métodos:

Característica	***Newton***	***BFGS***
Orden de convergencia	Cuadrático	Superlineal
Usa Hessiana exacta	Sí	No(la aproxima)
Costo por iteración	Alto (inversión de Hessiana)	Medio
Robustez numérica	Sensible al punto inicial	Alta
Dirección de descenso garantizada	Solo si Hessiana es pos. definida	Sí, si $y_k^T s_k > 0$
Comportamiento en este problema	Muy rápido si inicia cerca del mínimo	Estable y consistente en todo el dominio

Justificación de la elección:

Ambos métodos fueron seleccionados porque:

1. **La función f(x,y)** es **suave y diferenciable** (ideal para métodos de derivadas).

2. **No es convexa**, por lo que el gradiente descendente simple podría estancarse fácilmente.

3. **Newton** permite observar la dinámica de segundo orden (curvatura).

4. **BFGS** ofrece una alternativa más **robusta y práctica** que mantiene buena velocidad de convergencia sin depender de la Hessiana exacta.

Por ello, aplicar **Newton** y **BFGS** permite comparar la velocidad, estabilidad y precisión en la búsqueda de los óptimos globales de la función.

Comparación de resultados:

Para evaluar el desempeño de los métodos **Newton** y **Quasi-Newton (BFGS)** sobre la función:

$$f(x, y) = -tan^2(0.5sen(x) + 0.5sen(y))$$

se realizaron experimentos variando:

. Los **puntos iniciales** (x_0,y_0).

. Los **tamaños de paso (α)**.

. máximo de iteraciones : 100

Resultados numéricos obtenidos:

Se probaron 4 configuraciones representativas de puntos iniciales:

1. **Cerca del mínimo global:** (x_0,y_0) = (1.5,1.5)

2. **Lejos del mínimo, en otra "cuenca" de atracción:** (x_0,y_0)=(4.5,4.5).

3. **Región plana o inestable:** (x_0,y_0)=(0,π).
4. **No cumple sen(x)=sen(y):** (x_0,y_0)=(0.1,2.0).

Método	(x_0,y_0)	α inicial	Iteraciones	f(x*, y*)	Observaciones
Newton	(0.1, 2.0)	1.0 (fijo)	12	0.0	Converge a un máximo local.
Newton damped	(0.1, 2.0)	adaptativo	100	-0.30	divergencia detectada por iteraciones máximas sin lograr progresar
BFGS	(0.1, 2.0)	adaptativo	8	-2.4255	Estable; alcanza mínimo global.
Newton	(1.5, 1.5)	2.0 (fijo)	41	0.0	Converge a máximo local no global debido a que el paso grande se salta el global y cae en un valle plano de la función .
Newton	(1.5,1.5)	1 (fijo)	3	-2.4255	Converge estable al óptimo global(no hace salto hacia región crítica).
Newton damped	(1.5, 1.5)	adaptativo	3	-2.4255	Convergencia rápida (Newton cerca del mínimo global muestra su

					cuasicuadrática).
BFGS	(1.5,1.5)	adaptativo	4	-2.4255	Robusto; converge sin problemas al óptimo global desde este inicio.
Newton	(4.5,4.5)	1.0 (fijo)	6	0.0	Converge a máximo local no global debido a que el paso grande se salta el global y cae en un valle plano de la función .
Newton	(4.5,4.5)	0.5 (fijo)	18	-2.4255	Converge al óptimo global sin problemas.
Newton damped	(4.5,4.5)	adaptativo	3	-2.4255	Converge al óptimo global sin problemas.
BFGS	(4.5,4.5)	adaptativo	4	-2.4255	Converge al óptimo global sin problemas.
Newton	(0,π)	1.0 (fijo)	0	0.00	converge hacia máximo local porque el gradiente inicial lleva allí.
Newton damped	(0,π)	adaptativo	0	0.00	converge hacia el máximo local porque el gradiente inicial lleva allí.
BFGS	(0,π)	adaptativo	0	0.00	converge hacia el máximo local porque el gradiente inicial lleva allí.

Efecto del tamaño de paso (α):

En el método Newton Damped y en BFGS, la búsqueda lineal adaptativa (por Backtracking) ajusta el paso automáticamente, garantizando estabilidad sin necesidad de calibración manual

Para el método de Newton tradicional, se tiene que:

α=1.0 (paso completo) es eficiente cuando el punto inicial está cerca del óptimo pero pueden ocurrir saltos grandes que nos lleven a zonas problemáticas como en el caso de (1.5,1.5).

α<=0.3 La posibilidad de ir a una zona crítica disminuye pero aumenta el número de iteraciones significativamente.

Conclusiones:

Si no se controla el paso, Newton puede “saltar” a regiones donde tan(u) diverge (especialmente si sen(x) + sen(y) ≈ π). Por otro lado, un tamaño de paso muy pequeño aumenta las iteraciones y con ello la complejidad temporal.

.

Efecto de los puntos iniciales:

Dado que la función es **periódica y no convexa**, el punto inicial define en qué **mínimo global equivalente** termina el método.
Tanto Newton como BFGS convergen a valores óptimos idénticos, pero:

. Newton puede divergir si inicia en una zona donde la Hessiana cambia de signo.

. BFGS, al mantener H_k definida positiva, muestra un comportamiento más **robusto** y consistente.

Conclusiones de la comparación:

. Newton es el más rápido en converger, pero sensible a la elección del punto inicial y a la curvatura de la función.

. **BFGS** es más robusto y estable, logrando el mismo mínimo aunque requiera más iteraciones.

. Los diferentes puntos de partida llevan a mínimos equivalentes debido a la periodicidad de f, confirmando la existencia de **múltiples óptimos globales**.

. El tamaño de paso (α) tiene impacto crítico en Newton, mientras que BFGS ajusta automáticamente su escala.

En este problema específico, BFGS ofrece mejor combinación de estabilidad, precisión y robustez que Newton, pero como usando Newton Damped nos quitamos los problemas que conlleva Newton (debido a que se ajusta el tamaño de paso) , entonces son igual de buenos ambos para el análisis del problema.

Enlace al codigo de los algoritmos en github: https://github.com/ronaldcbmtnz/tarea_optimizacion.