(1) Calcule y clasifique los puntos críticos de la siguiente función

$$f(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2,$$

Muestra la función en Python.

Primero hay que calcular las derivadas parciales para encontrar los puntos:

$$\frac{\partial f}{\partial x_1} = 4x_1(x_1^2 + x_2^2 - 1),$$
$$\frac{\partial f}{\partial x_2} = 4x_2(x_1^2 + 2x_2^2 - 2).$$

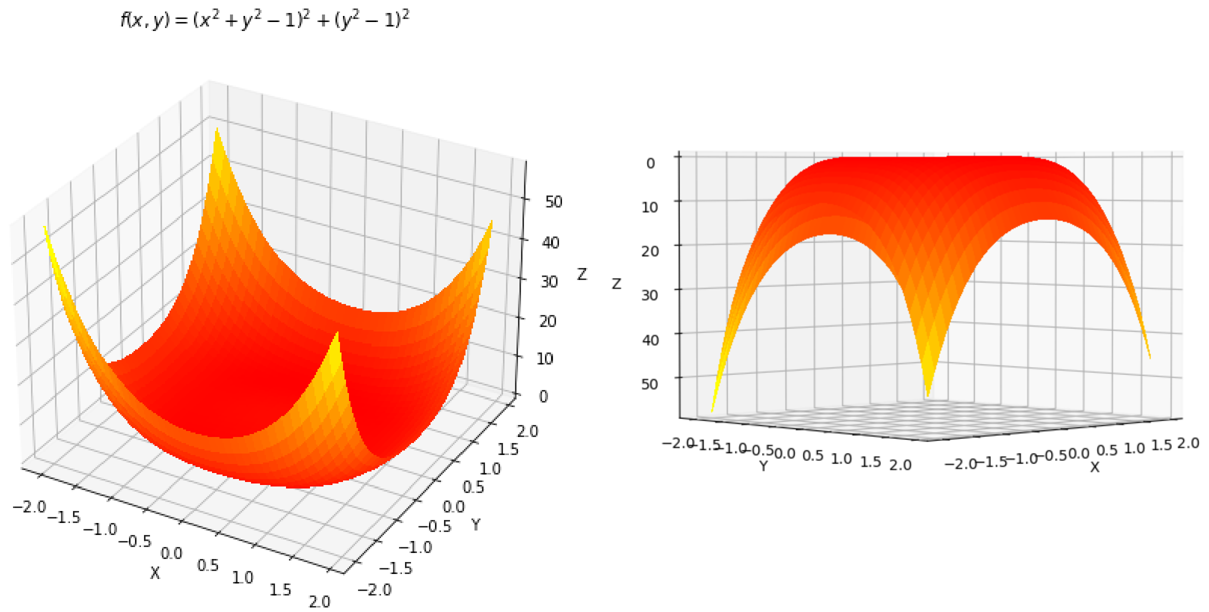
Igualando a cero, se encuentra los puntos: (0, 0), (1, 0), (-1, 0), (0, -1) y (0. 1). Para encontrar de qué tipo son, se utiliza el Hessiano de la función:

$$\begin{bmatrix} 12x_1^2 + 4x_2^2 - 4 & 8x_1x_2 \\ 8x_1x_2 & 4x_1^2 + 24x_2^2 - 8 \end{bmatrix}$$

y se evalúa en los puntos:

$$H(0,0) = \begin{pmatrix} -4 & 0 \\ 0 & -8 \end{pmatrix}, H(1,0) = \begin{pmatrix} 8 & 0 \\ 0 & -4 \end{pmatrix}, H(-1,0) = \begin{pmatrix} 8 & 0 \\ 0 & -4 \end{pmatrix},$$
$$H(0,-1) = \begin{pmatrix} 0 & 0 \\ 0 & 16 \end{pmatrix}, H(0,1) = \begin{pmatrix} 0 & 0 \\ 0 & 16 \end{pmatrix}$$

Así, se encuentra que el punto (0, 0) es un máximo, los puntos (1, 0) y (-1, 0) son puntos silla y los puntos restantes no se pueden clasificar con el criterio del Hessiano, así, graficando la función;



Evaluando la función en los puntos sin clasificar, se encuentra que ambos valen cero, y por lo tanto, son puntos mínimos.

(2) Sea A una matriz definida positiva, muestra que

$$1. A_{ij} < \frac{A_{ii} + A_{jj}}{2},$$
$$2. A_{ij}^2 < A_{ii}A_{jj}$$

Utilizando la definición de matriz definida postiva $x^T Ax > 0$, así, para un vector $x = [x_1, x_2, ..., x_n]$,

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{pmatrix} A_{ii} & A_{ij} \\ A_{ji} & A_{jj} \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} > 0$$

y escogiendo valores para el vector $x = [0, ..., 1, -1, ..., 0]$, se encuentra:

$$A_{ii} - 2A_{ij} + A_{jj} > 0,$$
$$\frac{A_{ii} + A_{jj}}{2} > A_{ij}.$$

Para el segundo inciso, también se propondrá un vector específico para x = [0, 0, ..., λ_1, λ_2, ..., 0], así, se encontrará un término similar al de la demostración anterior:

$$\lambda_1^2 A_{ii} + 2\lambda_1\lambda_2 A_{ij} + \lambda_2^2 A_{jj} > 0,$$

y como es una ecuación de segundo orden y es mayor a cero, significa que su discriminante debe ser menor a cero. Calculando el discriminante:

$$4A_{ij}^2 - 4(A_{ii}A_{jj}) < 0,$$
$$A_{ij}^2 < A_{ii}A_{jj}.$$

(3) Sea $f \to \mathbb{R}^n \to$ una función convexa. Muestra que para todo x, y se cumple:

$$f(y) \geq f(x) + \alpha(f(x) - f(z))$$

donde $\alpha > 0$ y $z = x + \frac{1}{\alpha}(x - y)$.

Utilizando la desigualdad ya demostrada en la tarea 2:

$$f(\lambda_1 x_1 + \lambda_2 x_2) \geq \lambda_1 f(x_1) + \lambda_2 f(x_2),$$

donde $\lambda_1 + \lambda_2 = 1, \lambda_1 > 0$ y $\lambda_2 \leq 0$. Haciendo un cambio de variable: $\lambda_1 = 1 + \frac{1}{\alpha}$, $\lambda_2 = -\frac{1}{\alpha}$,

$$f\left[\left(1 + \frac{1}{\alpha}\right)x + \left(-\frac{1}{\alpha}\right)y\right] \geq \left(1 + \frac{1}{\alpha}\right)f(x) + \left(-\frac{1}{\alpha}\right)f(y),$$
$$f\left[x + \frac{1}{\alpha}(x - y)\right] \geq f(x) + \frac{1}{\alpha}(f(x) - f(y)),$$
$$f(z) - f(x) \geq \frac{1}{\alpha}(f(x) - f(y)),$$
$$\alpha(f(z) - f(x)) \geq f(x) - f(y),$$
$$f(y) \geq f(x) + \alpha(f(x) - f(z))$$

(4) Sea

$$f(x) = \frac{1}{2}x^T \begin{bmatrix} 3/2 & 2 \\ 0 & 3/2 \end{bmatrix} x + x^T \begin{bmatrix} 3 \\ -1 \end{bmatrix} - 22,$$

1. Si se usa el algoritmo de gradiente descendete con tamaño de paso fijo para minimizar la función anterior, diga el rango de valores que puede tomar el tamaño de paso para que el algoritmo converja al minimizador.
2. Calcula el tamaño de paso exacto α_0 si el punto inicial es $x_0 = [0, 0]^T$.

Para poder utilizar el teorema que nos da el rango de α, $0 < \alpha < \frac{2}{\text{Eigenvalor más grande de Q}}$, es necesario que la matriz sea simétrica, y como se mencionó en clase, para convertirla simétrica es suficiente que $Q = \frac{A+A^T}{2}$, así,

$$Q = \frac{1}{2}\left\{\begin{bmatrix} \frac{3}{2} & 2 \\ 0 & \frac{3}{2} \end{bmatrix} + \begin{bmatrix} \frac{3}{2} & 9 \\ 2 & \frac{3}{2} \end{bmatrix}\right\} = \begin{bmatrix} \frac{3}{2} & 1 \\ 1 & \frac{3}{2} \end{bmatrix},$$

así, calculando los eigenvalores, se encuentra: $\lambda_1 = \frac{1}{2}$ y $\lambda_2 = \frac{5}{2}$, por lo tanto, el rango será: $0 < \alpha < \frac{4}{5}$.

Para calcular el tamaño de paso exacto dado el punto inicial, se utilizará la fórmula:

$$\alpha = \frac{\nabla^T f(x)\nabla f(x)}{\nabla^T f(x) Q \nabla f(x)}.$$

El gradiente de la función para $x = [x_1, x_2]$, es igual a:

$$\nabla^T f(x_1, x_2) = [\frac{3}{2}x_1 + x_2 + 3, x_1 + \frac{3}{2}x_2 - 1],$$
$$\nabla^T f(0, 0) = [3, -1],$$

reemplazando, se encuentra que $\alpha = \frac{10}{9}$.

(5) Muestra que $\|Bx\| \geq \|x\| \,/\, \|B^{-1}\|$ para cualquier matriz no singular B.

$$\|B^{-1}Bx\| = \|(B^{-1})(Bx)\|,$$

Utilizando la desigualdad de Cauchy: ,

$$\begin{aligned}
\|(B^{-1})(Bx)\| &\leq \|B^{-1}\| \, \|Bx\|, \\
\|x\| &\leq \|B^{-1}\| \, \|Bx\|, \\
\frac{\|x\|}{\|B^{-1}\|} &\leq \|Bx\|.
\end{aligned}$$

(6) Muestra que $logx \leq x - 1$

Primero, definimos a f(x) y su derivada,

$$\begin{aligned}
f(x) &= logx - x + 1, \\
f'(x) &= \frac{1}{x} - 1 = \frac{1-x}{x},
\end{aligned}$$

por su derivada sabemos que $f'(x) \leq 0$ para $x \geq 1$, ya que f(1) = 0, y $f(2) > 0, f(3) > 0, \ldots$, por lo tanto;

$$\begin{aligned}
logx - x + 1 &\leq 0, \\
logx &\leq x - 1.
\end{aligned}$$

(7) Muestra que $logx \geq 2(x-1)$ para $x \in [0{,}5, 1]$.

Otra forma de hacerlo, es también calcular la segunda derivada para saber qué tipo de punto crítico es:

$$\begin{aligned}
f(x) &= \frac{1}{x} - 2x + 2, \\
f'(x) &= \frac{1-2x}{x}, \\
f''(x) &= -\frac{1}{x^2},
\end{aligned}$$

por lo tanto, en el punto $f(\frac{1}{2}) > 0$ y f(1) = 0, entonces se tiene un máximo y como no hay otro punto crítico, se tiene un mínimo así: $f(x) \geq$, es decir: $logx \geq 2(x-1)$.

(8) Sea $\phi(a) := f(x_k + \alpha d_k)$ para $\alpha > 0$, $x_k, d_k \in \mathbb{R}^n$ y f es diferenciable. Sean los tamaños de paso a y b con $0 < a < b$ de modo que a cumple con la primera condición de Wolfe pero no cumple con la segunda y b no cumple con la primera condición de Wolfe, es decir, consideremos:

$$\begin{aligned}
f(x_k + ad_k) &< f(x_k) + c_1 a \nabla f(x_k)^T d_k, \\
\nabla f(x_k + ad_k)^T d_k &< c_2 \nabla f(x_k)^T d_k, \\
f(x_k + bd_k) &> f(x_k + c_1 b \nabla f(x_k)^T d_k)
\end{aligned}$$

Muestra que existe un tamaño de paso $\xi \in (a, b)$ que cumple con las dos condiciones de Wolfe.
La primera condición de Wolfe equivale a $\phi(\alpha) - \phi(0) - c_1 a \phi'(0) < 0$, por lo tanto, $phi(\alpha)$ es continua y se puede aplicar el teorema del valor medio a la función, que indica que si una función es continua y diferenciable, existe un punto c en el intervalo (s, t) tal que su derivada f'(c) es igual a:

$$f'(c) = \frac{f(t) - f(s)}{t - s}$$

Así:

$$\begin{aligned}
\phi'(\xi) &= \frac{f(x_k + td_k) - f(x_k + ad_k)}{t - a}, \\
&= \frac{f(x_k) + c_1 t \nabla f(x_k)^T d_k - f(x_k + ad_k)}{t - a}, \\
&> \frac{f(x_k) + c_1 t \nabla f(x_k)^T d_k - f(x_k) - c_1 a \nabla f(x)^T d_k}{t - a}, \\
&= \frac{c_1 (t - a) \nabla f(x)^T d_k}{t - a}, \\
&= c_1 \nabla f(x_k)^T d_k
\end{aligned}$$

como sabemos que $0 < c_1 < c_2 < 1$, se tiene que $c_1 \nabla f(x_k)^T d_k < 0$, por lo tanto, $f(x_k + \xi d_k) d_k > c_1 \nabla f(x)^T d_k > c_2 \nabla f(x)^T d_k$ que es la segunda condición de Wolfe.