Universidad de la República
Facultad de Ingeniería

Teoría y Algoritmia de Optimización

Año 2023

Entregable 1

JUAN MANUEL VARELA

Autor:

· Juan Manuel Varela

Docentes:
Ignacio Ramírez
Matías Valdes

4 de septiembre de 2023

i) He leído y estoy de acuerdo con las Instrucciones especificadas en la carátula obligatorio. ii) He resuelto por mi propia cuenta los ejercicios, sin recurrir a informes de otros compañeros, o soluciones existentes. iii) Soy el único autor de este trabajo. El informe y todo programa implementado como parte de la resolución del obligatorio son de mi autoría y no incluyen partes ni fragmentos tomados de otros informes u otras fuentes, salvo las excepciones mencionadas.

Ejercicio 1 - Convexidad

a) Para probar que $g(\mathbf{x})$ es convexa, se consideran dos puntos $\mathbf{x}, \mathbf{y}$ y un número λ tal que $0 \leq \lambda \leq 1$. Como cada $f_i(\mathbf{x})$ es convexa, entonces por definición:

$$f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f_i(\mathbf{x}) + (1-\lambda) f_i(\mathbf{y})$$

Como los w_i son no negativos:

$$w_i f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda w_i f_i(\mathbf{x}) + (1-\lambda) w_i f_i(\mathbf{y})$$

Sumando sobre todas las i:

$$\sum_i w_i f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \sum_i \lambda w_i f_i(\mathbf{x}) + (1-\lambda) w_i f_i(\mathbf{y})$$

Lo que equivale a:

$$g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda g(\mathbf{x}) + (1-\lambda) g(\mathbf{y})$$

Por lo tanto $g(\mathbf{x})$ es convexa.

b) Si se considera la función lineal $\mathbf{Ax} + \mathbf{b}$. Para cualquier $\mathbf{x}, \mathbf{y}$ y cualquier λ en el intervalo [0,1] se tiene que:

$$A(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + \mathbf{b} = \lambda(\mathbf{Ax} + \mathbf{b}) + (1-\lambda)(\mathbf{Ay} + \mathbf{b})$$

Por lo tanto:

$$l(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = f_1(\lambda(\mathbf{Ax} + \mathbf{b}) + (1-\lambda)(\mathbf{Ay} + \mathbf{b}))$$

Como f_1 es convexa, para cualquier $\mathbf{x}, \mathbf{y}$ y cualquier λ en el intervalo [0,1], por definición:

$$f_1(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f_1(\mathbf{x}) + (1-\lambda) f_1(\mathbf{y})$$

En particular:

$$f_1(\lambda(\mathbf{Ax} + \mathbf{b}) + (1-\lambda)(\mathbf{Ay} + \mathbf{b})) \leq \lambda f_1(\mathbf{Ax} + \mathbf{b}) + (1-\lambda) f_1(\mathbf{Ay} + \mathbf{b})$$

Esto es equivalente a:

$$l(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda l(\mathbf{x}) + (1-\lambda) l(\mathbf{y})$$

Por lo tanto $l(\mathbf{x})$ es convexa.

c) Para probar que Y es convexo, se consideran dos puntos $\mathbf{x}, \mathbf{y}$ en Y. Como Y es la intersecćon de los X_i, $\mathbf{x}$ e $\mathbf{y}$ están además en cada uno de los X_i.
Se tiene que cada X_i es convexo, por lo tanto, por definición $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ también está en cada X_i. Entonces, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ está en $Y \Rightarrow Y$ es convexo.

d) Basta con probar que para cualquier $\mathbf{x}, \mathbf{y}$ pertenecientes a $B(\mathbf{c}, r)$, y cualquier λ en el intervalo [0,1], el punto $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ también pertenece a $B(\mathbf{c}, r)$.

Geométricamente se puede ver que:

$$\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y} - \mathbf{c}\| \leq \lambda\|\mathbf{x} - \mathbf{c}\| + (1-\lambda)\|\mathbf{y} - \mathbf{c}\|$$

Como $\mathbf{x}, \mathbf{y} \in B(\mathbf{c}, r)$, se sabe que:

$$|\mathbf{x} - \mathbf{c}\| \leq r$$
$$\|\mathbf{y} - \mathbf{c}\| \leq r$$

Sustituyendo en la desigualdad anterior:

$$\lambda\|\mathbf{x} - \mathbf{c}\| + (1-\lambda)\|\mathbf{y} - \mathbf{c}\| \leq \lambda r + (1-\lambda)r = r$$

Por lo tanto, $\|\lambda\mathbf{x} + (1-\lambda)\mathbf{y} - \mathbf{c}\| \leq r$, lo que implica que $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in B(\mathbf{c}, r)$.

Entonces la bola es convexa.

Ejercicio 2 - Interpretación geométrica

a) Para probar que la función de costo es convexa en su dominio, primero se utiliza la propiedad $log\ a + log\ b = log(ab)$. Por esto, la función se puede reescribir como:

$$-log(x+y) - log(x-y)$$

Se sabe que la función $-log(x)$ es convexa y que tanto $f(\mathbf{x}) = x + y$ como $g(\mathbf{x}) = x - y$ son funciones lineales. Por lo probado en el ejercicio 1 parte b), la composición de una función convexa con una función lineal es convexa, y por lo probado en la parte a), la suma ponderada de funciones convexas es convexa. Por lo tanto $-log(x^2 - y^2)$ es convexa

b) Usando los resultados del ejercicio 1:

- $x^2 + y^2 \leq 1$ define una bola Euclídea, que por lo probado en la parte d), es convexa.
- $2x - y \leq 0$ define un semiplano, que es convexo.
- $y \geq \frac{1}{2}$ define otro semiplano, que es convexo.
- $x \geq 0$ define otro semiplano, que es convexo.

Por lo probado en la parte c) la intersección de conjuntos convexos es convexa. Por lo tanto, X es convexo.

c) Se muestra en la Figura 1 el conjunto X, representado en azul, y las curvas de nivel, representadas por colores que van desde el amarillo hasta el negro. Como se muestra en la escala de la derecha, el amarillo corresponde a los valores funcionales más altos y el negro a los más bajos.

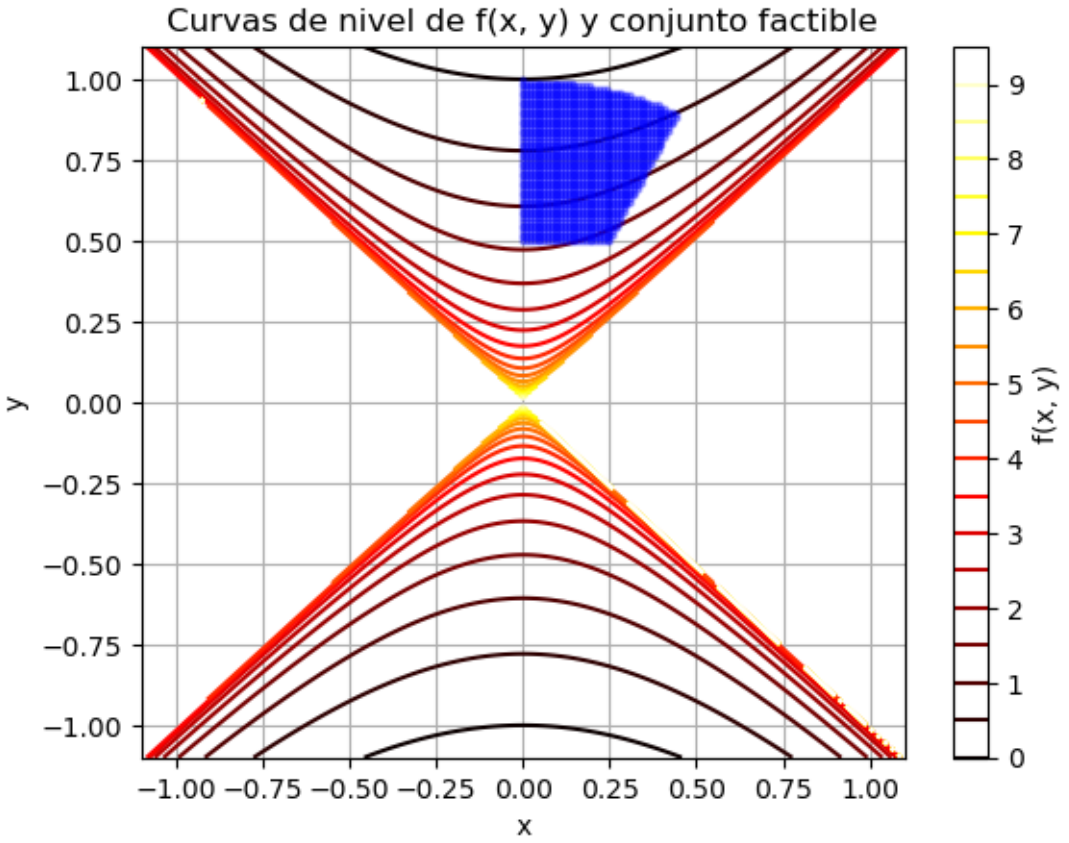

Figura 1: Curvas de nivel de función de costo y conjunto factible

A partir de esta representación se puede ver que la función crece cuando se acerca a las rectas $y = x$ e $y = -x$ y decrece cuanto más se aleja de ellas. Tambíen que la función no está definida en $\mathbb{R}$ cuando $-x \leq y \leq x$.

d) Observando el dibujo, la solución del problema sería el punto más bajo en el conjunto X. En este caso la curva de nivel donde la función de costo vale 0 toca el conjunto en un único punto, $(x^*, y^*) = (0, 1)$, todos los demás puntos del conjunto están por encima de esta curva de nivel, por lo que $(x^*, y^*) = (0, 1)$ es la solución del problema.

Ejercicio 3 - Puntos críticos y óptimos globales

a) En la Figura 2 se muestra la gráfica de la función objetivo $f(x) = 4x^4 - x^3 - 4x^2 + 1$

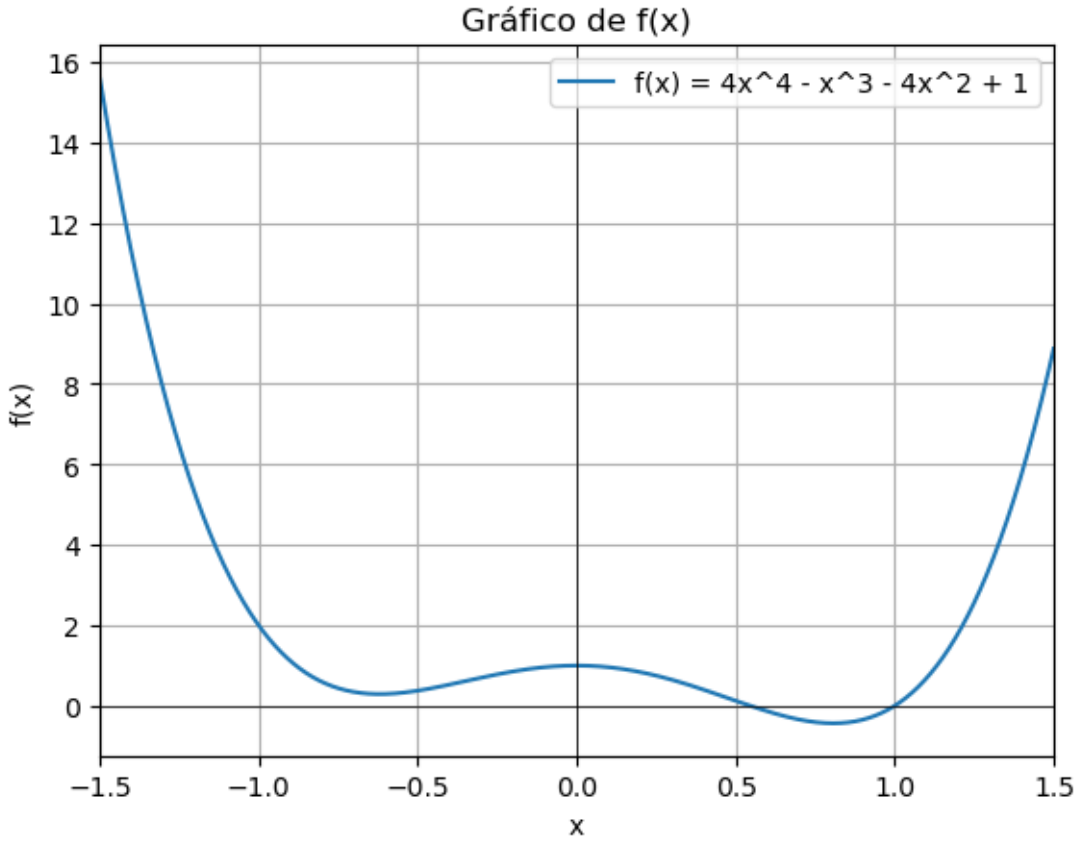

Figura 2: Gráfica de la función f(x)

En la gráfica se puede identificar un mínimo global que se da en $x \approx 0{,}8$, en este punto se anula el gradiente. Además, hay otros dos puntos donde se anula el gradiente: $x = 0$, que corresponde a un máximo local; y en $x \approx 0{,}6$, que corresponde a un mínimo local.

b) En la Figura 3 se muestra la gráfica de la función objetivo $g(x) = x^3$

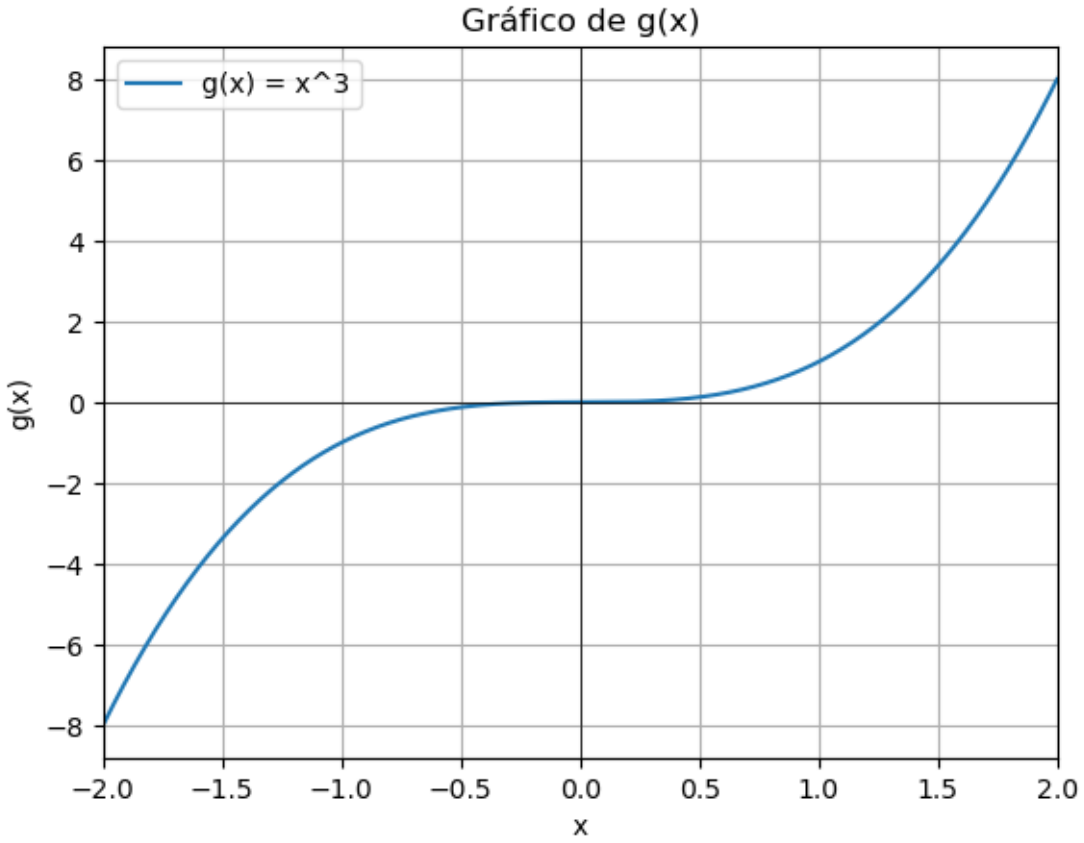

Figura 3: Gráfica de la función g(x)

En la gráfica se puede identificar un punto donde se anula el gradiente en $x = 0$, este punto corresponde a un punto silla porque la función crece hacia la derecha y decrece hacia la izquierda. El mínimo global se da en el extremo del conjunto factible $x = -1$.

c) En las Figuras 4, 5, 6 se muestran la gráficas de la función objetivo $h(x) = (x - a)^2 + 1$ con $a = 0$, $a = -1{,}5$ y $a = 0{,}5$ respectivamente.

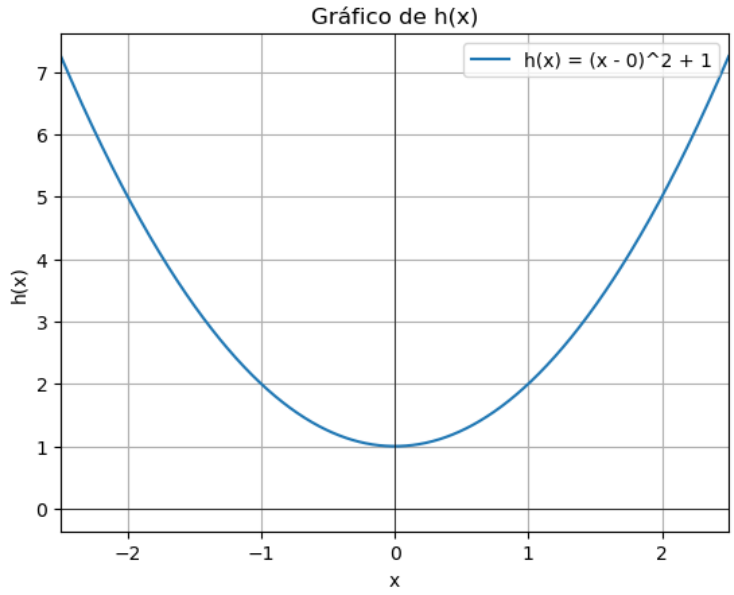

Figura 4: Gráfica de la función h(x), con a=0

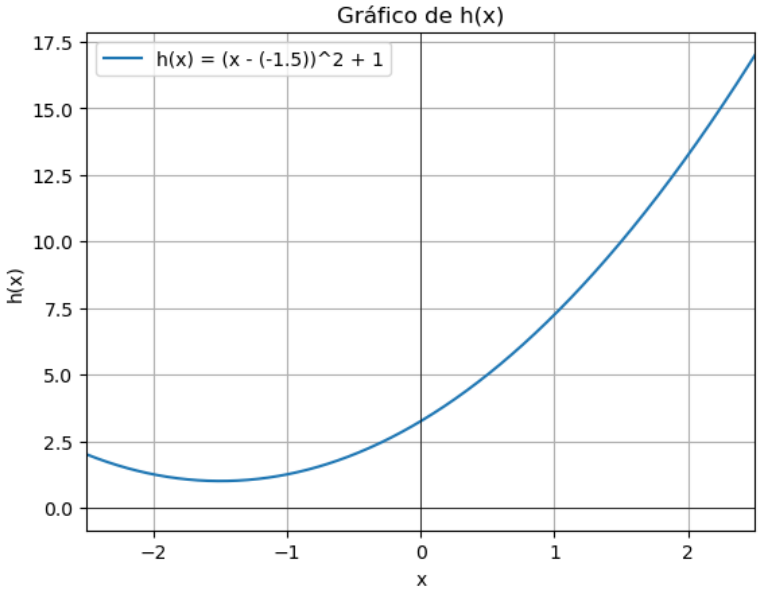

Figura 5: Gráfica de la función h(x), con a=-1.5

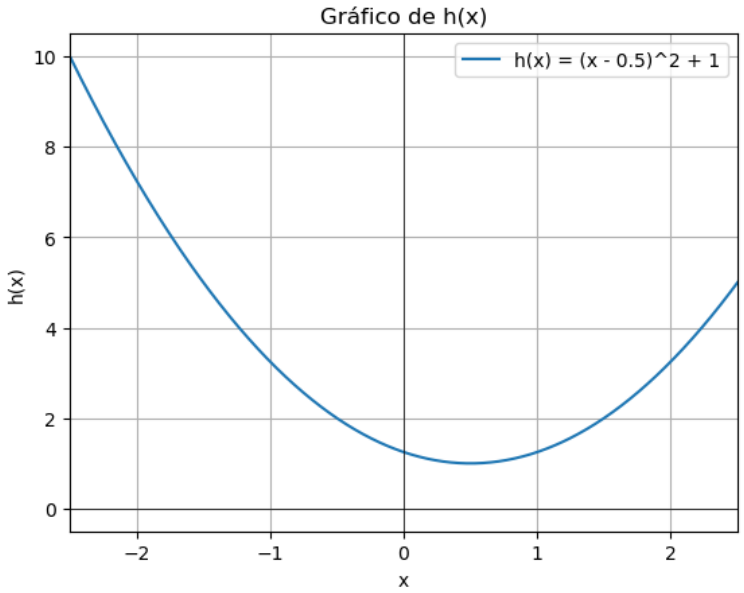

Figura 6: Gráfica de la función h(x), con a=0.5

En la gráficas se puede ver que el gradiente en todos los casos se anula en $x = a$, y que este punto corresponde al mínimo global de la función objetivo. Por lo tanto, cuando $-1 \leq a \leq 1$ el mínimo global se dará en $x = a$, esto ocurre en las Figuras 4 y 6. En cambio, cuando $a < -1$ el mínimo global se dará en el extremo del conjunto factible $x = -1$, esto ocurre en la Figura 5. De igual forma, cuando $a > 1$ el mínimo global se dará en el otro extremo $x = 1$.

d) En la Figura 7 se muestran las curvas de nivel de la función objetivo $f(x, y) = \left\|(x, y) - (\bar{x}, \bar{y})\right\|^2 + 1$ y el conjunto factible $B = [0, 1]^2$

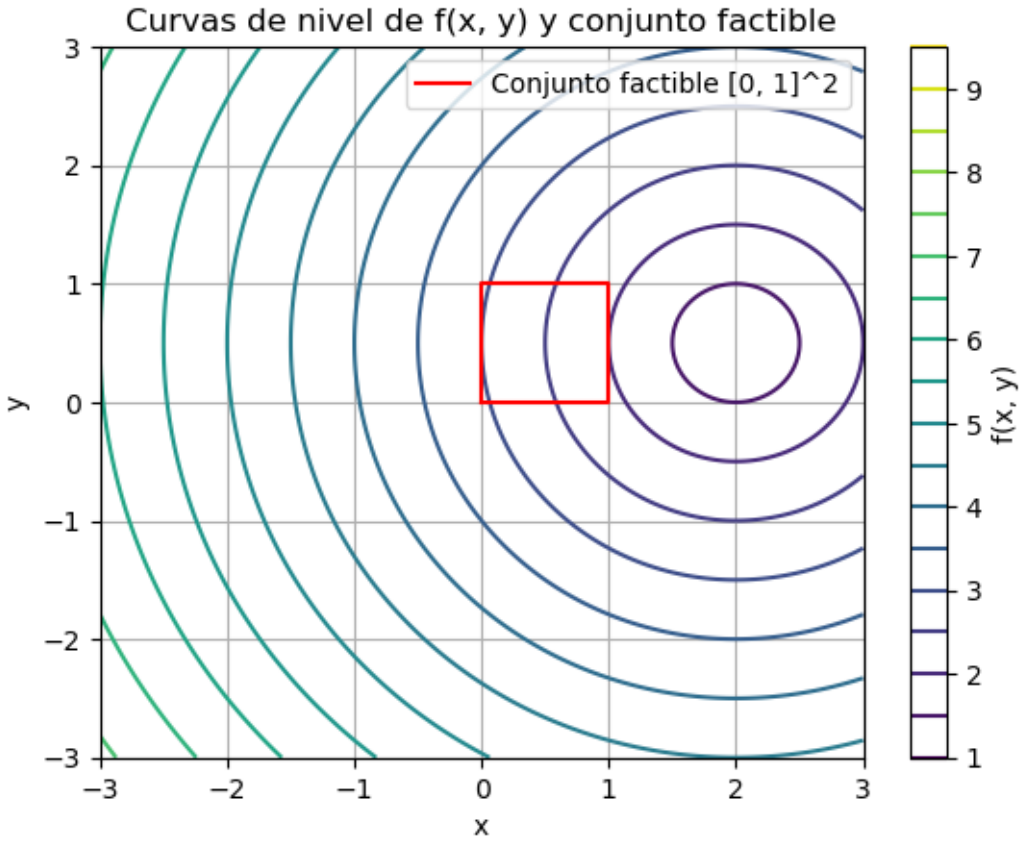

Figura 7: Gráfica de la función f(x,y)

A partir de las curvas de nivel se puede deducir que el gradiente de la función objetivo se anula en $(x, y) = (2, 1/2)$, y que este punto corresponde al mínimo global de la función objetivo. Esto tiene sentido ya que el punto más cercano a $(\bar{x}, \bar{y})$ es él mismo. Al restringirse al conjunto factible, el mínimo global se da en el borde del mismo, en el punto $(x, y) = (1, 1/2)$, ya que este punto toca la curva de nivel donde la función vale 2, todos los demás puntos del conjunto están por encima de esta curva de nivel.
Se sabe que cualquier norma en $\mathbb{R}^n$ es convexa, por lo que la función objetivo es convexa. Gráficamente se puede ver que el conjunto factible es convexo porque para dos puntos cualquiera dentro del conjunto, cualquier punto del segmento que los une estará dentro del conjunto.

Ejercicio 4 - Decenso por gradiente

a) La condición de optimalidad se obtiene derivando la función objetivo con respecto a $\mathbf{x}$ e igualandola a cero:

$$\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0$$

Por lo tanto, la solución analítica $\mathbf{x}^*$ es:

$$\mathbf{x}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

b) Se muestran las matrices calculadas en cada paso paso:

$$\mathbf{A}^T = \begin{bmatrix} -41 & -46 & -5 & -55 & -55 \\ 20 & -8 & -33 & 1 & -6 \end{bmatrix}$$

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 9872 & -12 \\ -12 & 1590 \end{bmatrix}$$

$$(\mathbf{A}^T\mathbf{A})^{-1} = \begin{bmatrix} \frac{265}{2616056} & \frac{1}{1308028} \\ \frac{1}{1308028} & \frac{617}{981021} \end{bmatrix}$$

$$\mathbf{A}^T\mathbf{b} = \begin{bmatrix} -1313 \\ 205 \end{bmatrix}$$

$$\mathbf{x}^* = \begin{bmatrix} \frac{-347535}{2616056} \\ \frac{502001}{3924084} \end{bmatrix}$$

Por lo tanto $\mathbf{x}^* \approx \begin{bmatrix} -0{,}132846926 \\ 0{,}127928199 \end{bmatrix}$

c) , d) y e) Se implementa y prueba el método de decenso por gradiente con los distintos tipos de paso. A continuación se muestra el código utilizado para resolver el ejercicio.

```
1     def descenso_por_gradiente(A, b, x_aster, paso="fijo", max_iter=1000,
      tol=1e-6):
2         x = np.zeros(A.shape[1]) # inicializacion
3         errores = []  # lista para almacenar errores relativos en cada
      iteracion
4         t_inicial = time.time()
5
6         for k in range(max_iter):
7             gradiente = A.T @ (A @ x - b)
8
9             # Se calcula el error relativo
10            error_relativo = np.linalg.norm(x_aster - x) / np.linalg.norm(
      x_aster)
11            errores.append(error_relativo)
12
13            # Si el gradiente es muy pequeno, se detiene
14            if np.linalg.norm(gradiente) < tol:
15                break
16
17            # Definir el tamano de paso
18            if paso == "fijo":
19                s = 1 / (2 * np.linalg.norm(A)**2)
20            elif paso == "decreciente":
21                s = 0.001 / (k+1)
22            elif paso == "exacto":
```

```
                # Para el paso exacto, hay que resolver un problema de optimizacion unidimensional
                s = (gradiente.T @ gradiente) / (gradiente.T @ A.T @ A @ gradiente)
            elif paso == "armijo":
                sigma = 0.1
                beta = 0.5
                s = 1  # Empieza con un tamano de paso de 1
                while np.linalg.norm(A @ (x - s * gradiente) - b)**2 > np.linalg.norm(A @ x - b)**2 - sigma * s * gradiente.T @ gradiente:
                    s *= beta  # Se reduce el tamano de paso hasta que sea suficientemente bueno

            # Actualizar x usando el gradiente y el tamano de paso
            x = x - s * gradiente

        t_final = time.time()
        t_total = t_final-t_inicial

        return x, errores, k+1, t_total

```

Se muestran a continuación los resultados obtenidos, utilizando un máximo de iteraciónes de 1000, y una norma del gradiente menor a 10^{-6} como condición de parada:

- Paso fijo: $\mathbf{x}^* = \begin{bmatrix} -0{,}132846930 \\ 0{,}127928200 \end{bmatrix}$
- Paso decreciente: $\mathbf{x}^* = \begin{bmatrix} -0{,}132846930 \\ 0{,}127928790 \end{bmatrix}$
- Paso exacto: $\mathbf{x}^* = \begin{bmatrix} -0{,}132846930 \\ 0{,}127928200 \end{bmatrix}$
- Regla de Armijo ($\sigma = 0{,}1; \beta = 0{,}5$): $\mathbf{x}^* = \begin{bmatrix} -0{,}132846930 \\ 0{,}127928200 \end{bmatrix}$

En todos los casos el resultado es prácticamente el mismo, y se acerca mucho al resultado obtenido al resolver el problema analíticamente.

En la Figura 8 se grafica el error relativo $\left\|\mathbf{x}^* - \mathbf{x}^t\right\| / \left\|\mathbf{x}^*\right\|$ vs. iteración t para los cuatro tipos de paso.

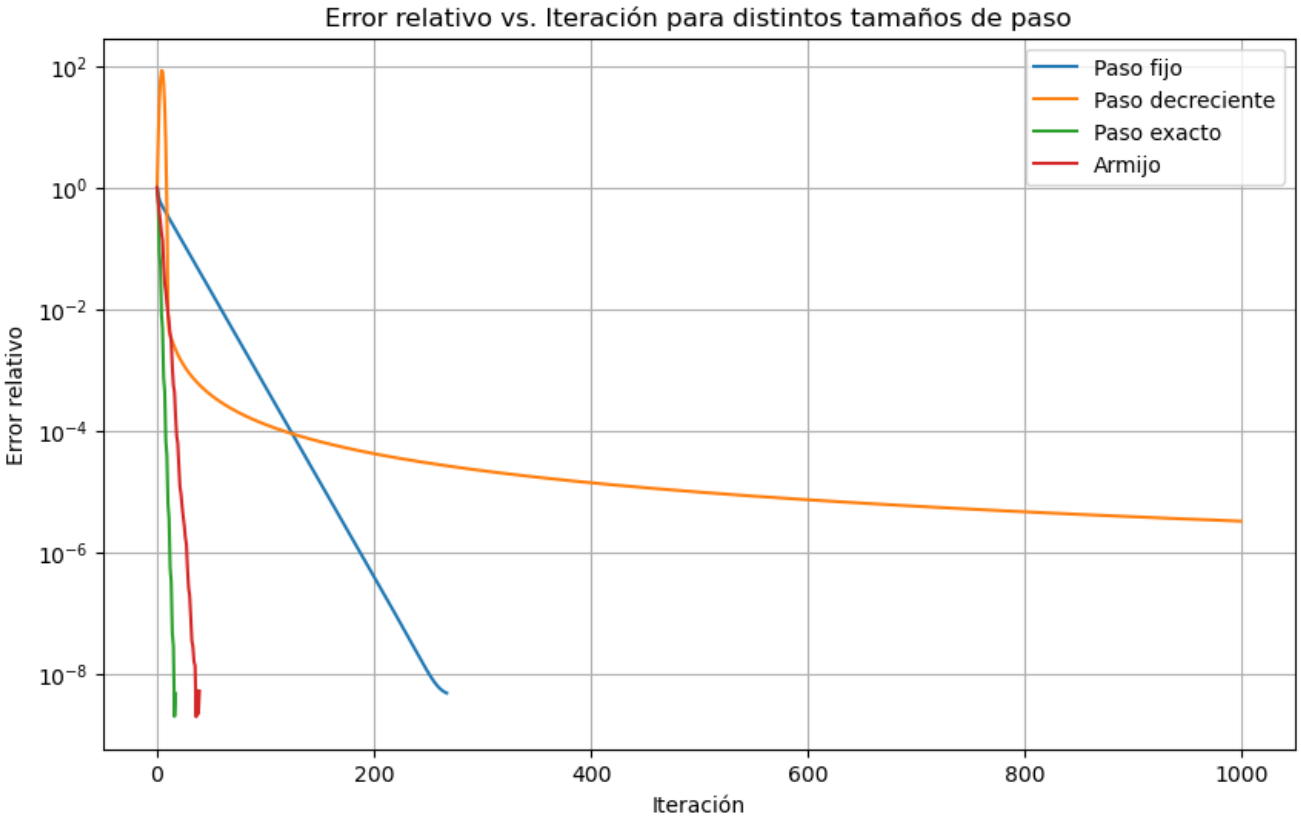

Figura 8: Error relativo respecto al óptimo para cada iteración, por tipo de paso

Se aprecia la diferencia en velocidad de convergencia para los distintos tipos de paso. Se tiene que si se utiliza paso exacto, se converge en la menor cantidad de iteraciones, seguido por Armijo, luego paso fijo y por último paso decreciente, destacando para este caso, que se llegó al máximo de iteraciones sin cumplir la condición de parada.

A continuación, en la Tabla 1 se presenta una tabla comparativa con la cantidad de iteraciones en cada caso, el tiempo total consumido y el tiempo promedio por iteracón.

paso	iteraciones	tiempo total (ms)	tiempo por iteración (ms)
fijo	268	13.224	0.049
decreciente	1000	19.661	0.020
exacto	18	1.001	0.056
Armijo	40	7.396	0.185

Tabla 1: Rendimiento para cada tipo de paso

El tiempo consumido para el tipo de paso exacto es el menor con gran diferencia. Esto probablemente se deba a que las operaciones que se realizan para resolver el problema de optimización unidimensional no son muy costosas computacionalmente, ya que se pudo resolver de manera analítica, esto queda reflejado en el tiempo por iteración, que es apenas mayor al del paso fijo. Esto podría variar mucho con otro problema, donde podría ser necesario utilizar algún

algoritmo de resolución de problemas de optimización unidimensionales como los métodos de interpolación cúbica y cuadrática mostrados en el apéndice C del libro "Nonlinear programming"de Dimitri P. Bertsekas. Se puede ver además, que al tomar siempre el mejor paso posible la cantidad de iteraciones se reduce al extremo.

El tiempo consumido utilizando la regla de Armijo es el segundo menor. Aquí se puede apreciar que el tiempo por iteración es el mayor, sin embargo reduce significativamente la cantidad de iteraciones al tomar un paso en donde se utiliza información del punto actual en cada iteración.

El tiempo consumido para el tipo de paso fijo y decreciente es mayor, esto se debe principalmente a que la cantidad de iteraciones se hace mas grande por utilizar un tamaño de paso que no requiere información nueva en cada iteración, ni realiza prácticamente ningún cálculo.

Si bien para este problema todos los tipos de paso tuvieron resultados satisfactorios, queda claro que dependiendo del tipo de problema y los recursos disponibles puede ser más conveniente utilizar un tipo u otro. En general, elegir el paso según la regla de Armijo parece ser la mejor opción dado que converge más rápidamente que el paso fijo y decreciente, se llega a una solución igual de buena que con paso exacto y se ahorra el de resolver un problema de optimización unidimensional en cada iteracion, con la complejidad y limitaciones que eso conlleva.

Ejercicio 5 - Mínimos cuadrados con restricciones

a) Para determinar que el problema es convexo basta con verificar que la función objetivo y las restricciones son convexas.

La matriz hessiana de la función objetivo es:

$$H(x,y) = \begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix}$$

Los valores propios de la matriz son 4 y 16, ambos positivos, por lo tanto la hessiana es definida positiva. Esta es condición suficiente para decir que la función objetivo es convexa.

Por otro lado, la restricción es una bola de centro $(0,0)$ y radio R. Como fue probado en el Ejercicio 1, parte d), toda bola Euclidea es convexa.

Por lo tanto, el problema es convexo.

b) Para mostrar que la restricción $x^2 + y^2 \leq R^2$ está activa para $R < \frac{1}{2}$, primero se encuentra el punto crítico de la función objetivo sin considerar la restricción. Esto significa hallar el punto donde se anula el gradiente.

El gradiente de la función es:

$$\bigtriangledown f(x,y) = (10x + 5 - 6y, 10y - 3 - 6x)$$

Resolviendo $\bigtriangledown f(x,y) = (0,0)$, se halla el punto crítico:

$$(x,y) = (-\frac{1}{2}, 0)$$

Este punto está exactamente a distancia $1/2$ del origen, por lo tanto, la restricción está activa para $R < \frac{1}{2}$.

c) En la Figura 9 se muestran las sucesiones de puntos obtenidas por el método y la región factible, con ambas elecciones de paso. Se hace un "zoom" en la zona cercana al óptimo para distinguir mejor las sucesiones, el único punto que queda fuera de la gráfica es el inicial $(0, 0)$, que fue elegido arbitrariamente.

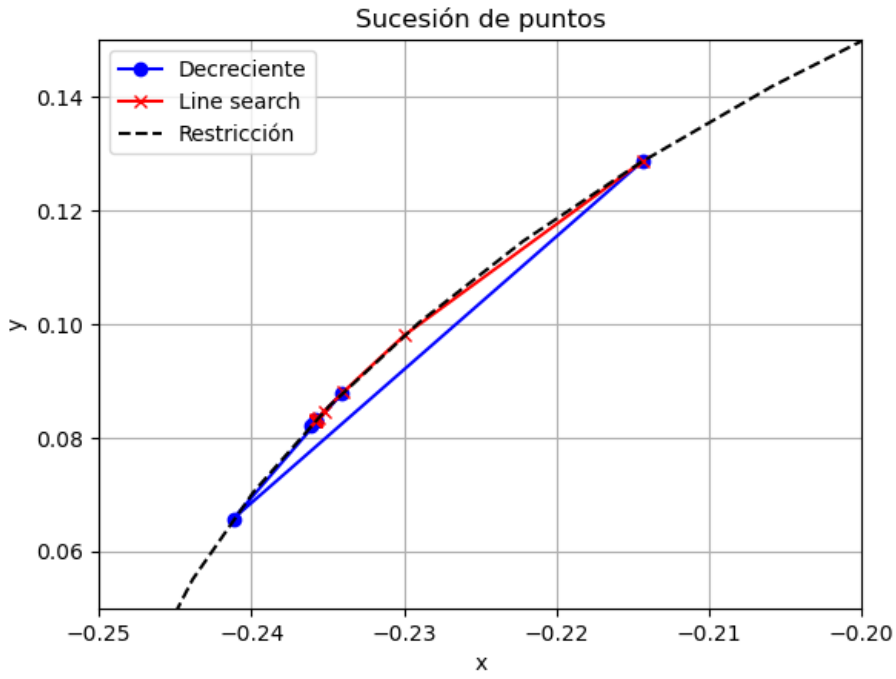

Figura 9: Sucesión de puntos (x^k, y^k) y región factible

En las Figuras 10 y 11 se grafica el valor de la función de costo y la distancia entre puntos sucesivos para cada iteración respectivamente, distinguiendo por método de elección de paso.

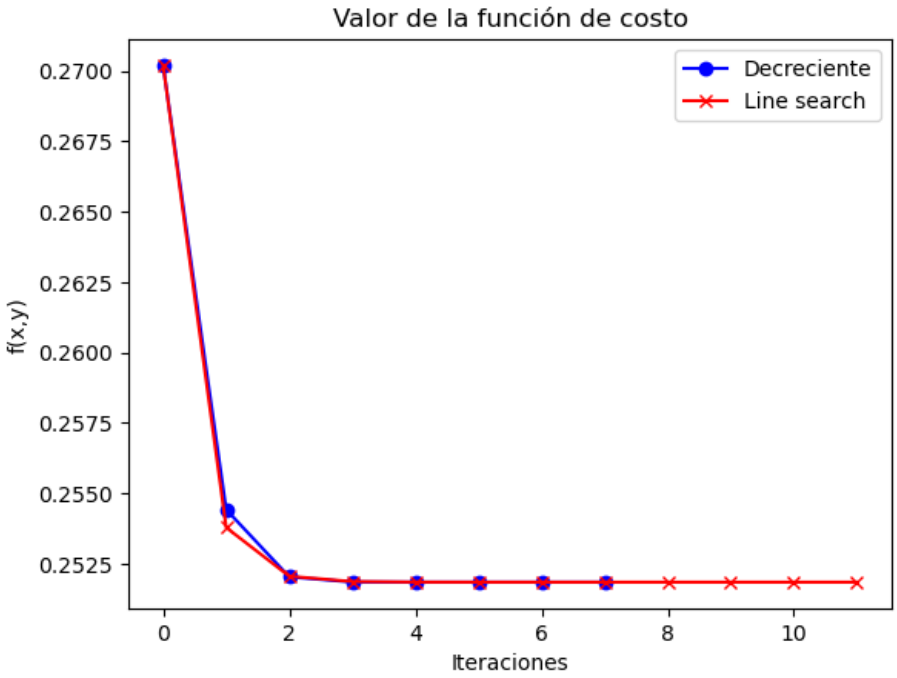

Figura 10: Valor de la función de costo para cada iteración

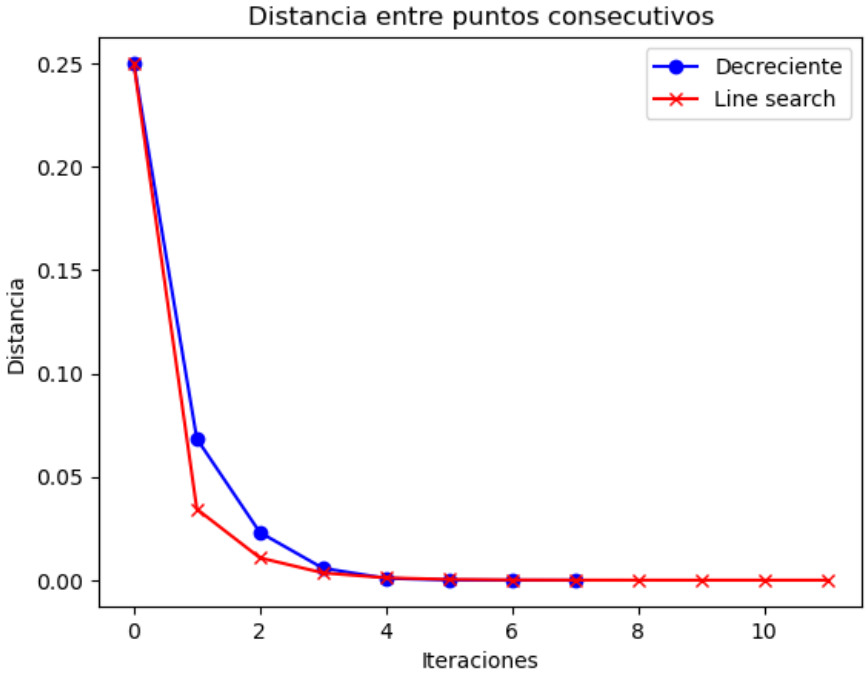

Figura 11: Distancia entre puntos sucesivos para cada iteración

Se puede observar que, como es de esperar, utilizar line search para elegir el paso resulta en un camino más directo hacia el óptimo. Esto queda evidenciado en las sucesiones de puntos obtenidas, donde se ve que los puntos azules hacen cierto zigzagueo antes de llegar al óptimo.

Además, las gráficas de la función de costo y la distancia entre puntos apoyan esta idea, ya que muestran que con line search estos valores se acercan a cero más rápido.

Los puntos obtenidos son:

- paso decreciente: $\begin{bmatrix} -0{,}23580882 \\ 0{,}08303133 \end{bmatrix}$
- line search: $\begin{bmatrix} -0{,}23580856 \\ 0{,}08303206 \end{bmatrix}$

Estos puntos son muy cercanos entre sí, y a su vez muy cercanos al óptimo $(-0{,}23580879, 08303138)$ calculado de forma analítica, utilizando el método de los multiplicadores de Lagrange.

Para determinar el paso según line search se calculó $f(s) = f(x_k - s \bigtriangledown f(x_k))$, para un x_k genérico, se calculó $\frac{df(s)}{ds}$, se igualó a 0 y se despejó s. Con la fórmula obtenida se creó una función para definir el paso en cada iteración.

La condición de parada utilizada fue que la distancia entre puntos sucesivos sea menor a 10^{-6}, en ambos casos se llegó a esta condición muy rápdidamente, en el orden de la decena de iteraciones. El hecho de que la condición de parada se haya cumplido antes para el caso de line search puede ser la explicación de que que el resultado obtenido por paso decreciente sea algo más cercano al resultado esperado.