INTRODUCCIÓN A LA OPTIMIZACIÓN CON ALGORITMOS

Índice

1. Minimización irrestricta **6**
1.1. Condiciones de Optimalidad 6
1.2. Métodos de descenso 9
1.3. Búsqueda lineal 11
1.4. Dirección de Newton 17
1.5. Métodos de Quasi-Newton 22
1.6. Regiones de confianza 25

2. Teoría básica de optimización con restricciones **31**
2.1. Minimización con restricciones de igualdad 31
2.2. Minimización con restricciones de desigualdad 39
2.3. Condiciones necesarias y suficientes de optimalidad para problemas con restricciones de igualdad y desigualdad 46
2.4. Problemas convexos 47

3. Métodos numéricos para problemas generales de optimización **49**
3.1. Método del gradiente proyectado 49
3.2. Método de Conjuntos activos 52
3.3. Métodos de penalidad y barrera 57
3.4. Método de Lagrangiano Aumentado 68

Notación y consideraciones iniciales.

1. Denotamos $x \in \mathbb{R}^n$ al vector $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

2. Dada una sucesión de puntos $\{x_k\}_k$ en $\mathbb{R}^n$. Decimos que $\{x_k\}$ converge a x, y escribimos $x_k \to x$ si, para todo $\varepsilon > 0$ existe k_0 tal que
$$\|x_k - x\| \leq \varepsilon, \forall k \geq k_0.$$
Decimos que $\tilde{x} \in \mathbb{R}^n$ es un punto de acumulación de $\{x_k\}$ (o punto límite) si existe una subsucesión infinita de índices $K = \{k_1, k_2, \ldots\}$ tal que $\lim_{k \in K} x_k = \tilde{x}$.

3. Toda sucesión $\{x_k\}$ acotada ($\|x_k\| \leq M$) tiene una subsucesión convergente.

4. Una función $f : \mathbb{R}^n \to \mathbb{R}$ es continua en x_0 si $lim_{x \to x_0} f(x) = f(x_0)$. Equivale a que, para todo $\varepsilon > 0$ existe $\delta > 0$ tal que
$$\text{si } \|x - x_0\| < \delta \text{ entonces } |f(x) - f(x_0)| \leq \varepsilon.$$

5. Dada $g : \mathbb{R}^n \to \mathbb{R}$ decimos que g es Lipschitz continua si existe $\mathcal{L} > 0$ tal que $|g(x) - g(y)| \leq \mathcal{L}\|x - y\|, \forall x, y \in \mathbb{R}^n$.

6. Si $\varphi : \mathbb{R} \to \mathbb{R}$ es una función de una variable, se define la derivada de φ como
$$\varphi'(t) = \frac{d\varphi}{dt}(t) = \lim_{h \to 0} \frac{\varphi(t+h) - \varphi(t)}{h}$$

7. Si $f : \mathbb{R}^n \to \mathbb{R}$, el gradiente de f es el vector formado por las derivadas parciales de f, $\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$ y $\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_1 x_n} & \cdots & \frac{\partial^2 f(x)}{\partial^2 x_n} \end{pmatrix}$ es la matriz Hessiana de f.

8. Si $F : \mathbb{R}^n \to \mathbb{R}^m, F(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$ entonces $F'(x) = JF(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix} \in \mathbb{R}^{m \times n}$ se denomina matriz Jacobiana de F.

9. Si las derivadas parciales de f existen y son continuas decimos que f es diferenciable. En este caso, la derivada direccional de f en x_0 en la dirección u, $\|u\| = 1$ se define como
$$D_u f(x_0) = \lim_{h \to 0} \frac{f(x_0 + hu) - f(x_0)}{h} = \nabla f(x_0)^T u.$$

10. Denotamos por $\mathcal{C}^n(\mathbb{R}^p)$ al conjunto de las funciones en $\mathbb{R}^p$ que son continuas y tienen derivadas parciales continuas hasta orden n.

11. Teorema del valor medio. Dada una función $\varphi : \mathbb{R} \to \mathbb{R}$ y dos números reales $t_1, t_2, t_2 > t_1$, existe $\xi \in (t_1, t_2)$ tal que
$$\varphi^{'}(\xi) = \frac{\varphi(t_2) - \varphi(t_1)}{t_2 - t_1}.$$
Una extensión de este resultado para funciones de varias variables establece que existe $\alpha \in (0, 1)$ tal que
$$\nabla f(x_0 + \alpha d)^T d = f(x_0 + d) - f(x_0).$$
Luego, si f tiene derivadas segundas continuas se obtiene su desarrollo de Taylor alrededor de x_0:
$$f(x) = f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0) + o(\|x - x_0\|^2)$$
donde $\frac{o(\|x-x_0\|^2)}{\|x-x_0\|^2} \to 0$.

12. Dado $x \in \Omega, \varepsilon > 0, B(x, \varepsilon) = \{y \in \mathbb{R}^n : \|x - y\| < \varepsilon\}$.

13. Teorema de la Función Implícita. Sea $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ una función diferenciable en un entorno de un punto $(z^*, 0)$ tal que

    a) $h(z^*, 0) = 0$ para algún $z^* \in \mathbb{R}^n$,
    b) $\nabla_z h(z^*, 0)$ es no singular.

    Entonces existen entornos de z^* y de $t = 0$ y una función diferenciable $z(t)$ definida en forma implícita tal que $h(z(t), t) = 0$.

14. Si $x, y \in \mathbb{R}^n, x \leq y$ representa $x_i \leq y_i$ para todo $i = 1, \ldots, n$.

15. Dada una matriz $A \in \mathbb{R}^{n\times n}$ simétrica, decimos que es semidefinida positiva (respectivamente definida positiva) si $x^T Ax \geq 0, \forall x$ (respectivamente $x^T Ax > 0, \forall x \neq 0$). Denotamos por $A \geq 0$ (respectivamente $A > 0$) una matriz semidefinida positiva(respectivamente definida positiva). Si A es simétrica definida positiva sabemos que sus autovalores $\lambda_1, \ldots, \lambda_n$ son todos positivos y se tiene que, para todo $x \in \mathbb{R}^n$:
$$\lambda_{min}(A)\|x\|^2 \leq x^T Ax \leq \lambda_{max}(A)\|x\|^2.$$

16. Norma de matrices. Dada una matriz $A \in \mathbb{R}^{n\times n}$ se define $\|A\| = \sup\limits_{x\in\mathbb{R}^n, x\neq 0} \dfrac{\|Ax\|}{\|x\|} = \sup\limits_{\|x\|=1} \|Ax\|$. Luego, las normas usuales de vectores $\|x\|_1 = \sum\limits_{i=1}^{n} |x_i|, \|x\|_2 = \sqrt{\sum\limits_{i=1}^{n} (x_i)^2}$ y $\|x\|_\infty = \text{máx}_{i=1,\ldots,n} |x_i|$ inducen normas de matrices y se puede probar que $\|A\|_1 = \text{máx}_{j=1,\ldots,n} \sum\limits_{i=1}^{n} |a_{ij}|, \|A\|_\infty = \text{máx}_{i=1,\ldots,n} \sum\limits_{j=1}^{n} |a_{ij}|$ y $\|A\|_2 = (\lambda_{max}(A^T A))^{1/2}$ donde λ_{max} es el mayor autovalor de la matriz.

17. Dado el espacio $\mathbb{R}^n$, un subconjunto $S \subset \mathbb{R}^n$ es un subespacio de $\mathbb{R}^n$ si $\alpha x + \beta y \in S, \forall \alpha, \beta \in \mathbb{R}, x, y \in S$.

18. Dado un subespacio S, los vectores $v_1, \ldots, v_m$ en S son linealmente independientes si para cualquier combinación

$$\alpha_1 v_1 + \ldots \alpha_m v_m = 0 \text{ implica } \alpha_i = 0 \forall i = 1, \ldots, m.$$

en caso constrario decimos que los vectores son linealmente dependientes.

19. Dado un subespacio S, se define el espacio ortogonal a $S \in \mathbb{R}^n$ como el subespacio

$$S^\perp = \{x \in \mathbb{R}^n : x^T y = 0, \forall y \in S\}.$$

Se tiene que $S \oplus S^\perp = \mathbb{R}^n$.

20. Si $A \in \mathbb{R}^{m \times n}$ es una matriz real, se define el núcleo como el subespacio

$$Nu(A) = \{u \in \mathbb{R}^n : Au = 0\} \subset \mathbb{R}^n,$$

y el rango o imagen de A como el subespacio

$$Im(A) = R(A) = \{w \in \mathbb{R}^m : w = Au \text{ para algún vector } u\} \subset \mathbb{R}^m.$$

Se tiene que $R(A)^\perp = Nu(A^T)$.

El teorema fundamental de algebra lineal establece que

$$Nu(A) \oplus Im(A^T) = \mathbb{R}^n.$$

21. Si $A \in \mathbb{R}^{m \times n}, m < n$ que tiene rango m entonces la matriz $A^T A$ es no singular.

1. Minimización irrestricta

En esta sección presentaremos condiciones de optimalidad y métodos prácticos para resolver problemas de minimización sin restricciones.

1.1. Condiciones de Optimalidad

Consideramos el problema general de optimización no lineal:

$$\begin{array}{c}\text{Minimizar } f(x)\\ \text{sujeto a } x \in \Omega \subset \mathbb{R}^n\end{array} \tag{1}$$

para $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciable en un conjunto abierto que contiene a Ω. La función f será llamada función objetivo.

Consideramos al conjunto $\mathbb{R}$ con el orden usual, de esta manera tiene sentido hablar de minimizador, aunque sabemos que no siempre puede garantizarse existencia de un minimizador cuando $\Omega = \mathbb{R}^n$.

Pueden considerarse diferentes representaciones del conjunto Ω:

1. $\Omega = \mathbb{R}^n$ representa el problema de minimización irrestricta.

2. $\Omega = \{x \in \mathbb{R}^n; Ax = b\}, A \in \mathbb{R}^{m\times n}$, problema de minimización con restricciones lineales de igualdad.

   Si $\Omega = \{x \in \mathbb{R}^n; Ax \leq b\}, A \in \mathbb{R}^{m\times n}$, problema de minimización con restricciones lineales de desigualdad.

   Si $\Omega = \{x \in \mathbb{R}^n; Ax = b, Cx \leq d\}, A \in \mathbb{R}^{m\times n}, C \in \mathbb{R}^{p\times n}$, problema de minimización con restricciones lineales.

   Si además, $f(x) = e^T x$ el problema se llama problema de programación lineal.

   Si, $f(x) = x^T Hx + e^T x$ con H matriz simétrica, el problema se llama problema de programación cuadrática.

3. Si $\Omega = \{x \in \mathbb{R}^n; h(x) = 0\}$ con $h : \mathbb{R}^n \to \mathbb{R}^m$ el problema es de programación no lineal con restricciones de igualdad.

4. Si $\Omega = \{x \in \mathbb{R}^n; h(x) = 0, g(x) \leq 0\}$ con $h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$ el problema es de programación no lineal con restricciones de igualdad y desigualdad o problema general de programación no lineal.

Definición 1.1. *Decimos que $x^* \in \Omega$ es un minimizador global de $f(x)$ en Ω si*

$$f(x^*) \leq f(x), \forall x \in \Omega.$$

Decimos que x^* es un minimizador local de $f(x)$ en Ω si

$$f(x^*) \leq f(x), \forall x \in \Omega \cap B(x^*, \varepsilon).$$

Estudiaremos a continuación condiciones necesarias y suficientes para minimizadores locales de una función sin restricciones. Estas condiciones pueden ser de primer o de segundo orden dependiendo del número de derivadas que pueden tenerse en cuenta. Vamos a suponer siempre que f tiene, al menos, derivadas parciales continuas de primer orden. Cuando sea necesario el uso de derivadas de orden superior se establecerá explicitamente.

Suponemos conocidos los siguientes resultados para funciones de una variable:

R1 Sea $f : \mathbb{R} \to \mathbb{R}, C^1$. Si x^* es un minimizador local de f en $\mathbb{R}$ entonces $f'(x^*) = 0$.

R2 Sea $f : \mathbb{R} \to \mathbb{R}, C^2$. Si x^* es un minimizador local de f en $\mathbb{R}$ entonces $f'(x^*) = 0$ y $f''(x^*) \geq 0$.

Teorema 1.2. Condición necesaria de primer orden. *Sea $f : \mathbb{R}^n \to \mathbb{R}, C^1$. Si x^* es un minimizador local de f en $\mathbb{R}^n$ entonces $\nabla f(x^*) = 0$.*

Dem.

Sea $d \in \mathbb{R}^n$ arbitraria. Definimos $\gamma(t) = x^* + td$ y $\varphi(t) = f(\gamma(t))$, $\varphi : \mathbb{R} \to \mathbb{R}$. Como x^* es un minimizador local de f entonces $t = 0$ es un minimizador local de φ. Observar la figura 1. Luego, por **R1** tenemos que $\varphi'(0) = 0$.

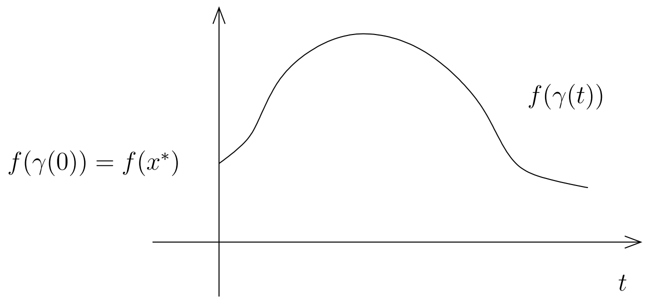


Figura 1: El valor $t = 0$ es un minimizador de $\varphi(t)$.

Luego, usando regla de la cadena, sabemos que $\varphi'(t) = \nabla f(\gamma(t))^T \gamma'(t)$. Evaluando en $t = 0$ resulta $0 = \varphi'(0) = \nabla f(x^*)^T \gamma'(0) = \nabla f(x^*)^T d$. Como d es arbitrario, considerando $d = \nabla f(x^*)$ si $\nabla f(x^*) \neq 0$ obtenemos que $\|\nabla f(x^*)\| = 0$. Luego, debe ser $\nabla f(x^*) = 0$. □

La derivada $\varphi'(0) = \nabla f(x^*)^T d$ es una derivada direccional de f en x^* en la dirección d y mide la variación de la función en esa dirección.

Definición 1.3. *Un punto x^* que cumple $\nabla f(x^*) = 0$ se denomina punto estacionario de f.*

Teorema 1.4. Condición necesaria de segundo orden. *Sea $f : \mathbb{R}^n \to \mathbb{R}, C^2$. Si x^* es un minimizador local de f entonces $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es semidefinida positiva.*

Dem.

La primera parte se obtiene del teorema anterior. Para probar la segunda parte, consideramos nuevamente la función $\varphi(t) = f(\gamma(t))$, $\varphi : \mathbb{R} \to \mathbb{R}$ para $\gamma(t) = x^* + td$. Luego, por **R2** tenemos que $\varphi^{''}(0) \geq 0$. Usando regla de la cadena,

$$\varphi^{'}(t) = \nabla f(\gamma(t))^T \gamma^{'}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^* + td)d_i.$$

$$\begin{aligned} \varphi^{''}(t) = &\left(\frac{\partial^2 f}{\partial x_1^2}(x^* + td)d_1 + \frac{\partial^2 f}{\partial x_1 \partial x_2}(x^* + td)d_2 + \ldots + \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^* + td)d_n \right) d_1 + \\ &\ldots + \left(\frac{\partial^2 f}{\partial x_1 \partial x_n}(x^* + td)d_1 + \ldots + \frac{\partial^2 f}{\partial x_n^2}(x^* + td)d_n \right) d_n = \\ &d^T \nabla^2 f(x^* + td)d. \end{aligned}$$

Asi, vemos que $\varphi^{''}(0)$ mide la curvatura de f en x^* en la dirección d. La condición $\varphi^{''}(0) \geq 0$ equivale a $d^T \nabla^2 f(x^*) d \geq 0, \forall d \in \mathbb{R}^n$. Luego, en ese caso tenemos que $\nabla^2 f(x^*)$ es semidefinida positiva como queriamos demostrar. □

La condición del teorema anterior es necesaria pero no suficiente, considerar por ejemplo, la función $f(x, y) = x^3 + y^3$ en el origen.

Teorema 1.5. Condición suficiente. *Sean f una función $\mathcal{C}^2$ y x^* un punto estacionario de f. Si $\nabla^2 f(x^*)$ es definido positivo entonces x^* es un minimizador local de f.*

Dem.

Consideremos el desarrollo de Taylor de f alrededor de x^*:

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*) + o(\|x - x^*\|^2).$$

Luego, $\forall x \neq x^*$ se tiene que

$$\frac{f(x) - f(x^*)}{\|x - x^*\|^2} = \frac{1}{2} \frac{(x - x^*)}{\|x - x^*\|}^T \nabla^2 f(x^*) \frac{(x - x^*)}{\|x - x^*\|} + \frac{o(\|x - x^*\|^2)}{\|x - x^*\|^2}.$$

Como $\nabla^2 f(x^*)$ es definido positivo tenemos que sus autovalores cumplen $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$. Luego,

$$d^T \nabla^2 f(x^*) d \geq \lambda_n \|d\|^2.$$

Esto implica que

$$\frac{f(x) - f(x^*)}{\|x - x^*\|^2} \geq \lambda_n + \frac{o(\|x - x^*\|^2)}{\|x - x^*\|^2}.$$

Como $\frac{o(\|x - x^*\|^2)}{\|x - x^*\|^2}$ tiende a 0 cuando $x \to x^*$ tenemos que, para x suficientemente próximo a x^*, $\frac{o(\|x - x^*\|^2)}{\|x - x^*\|^2} \geq -\frac{\lambda_n}{4}$ y obtenemos que

$$\frac{f(x) - f(x^*)}{\|x - x^*\|^2} \geq \frac{\lambda_n}{2} > 0$$

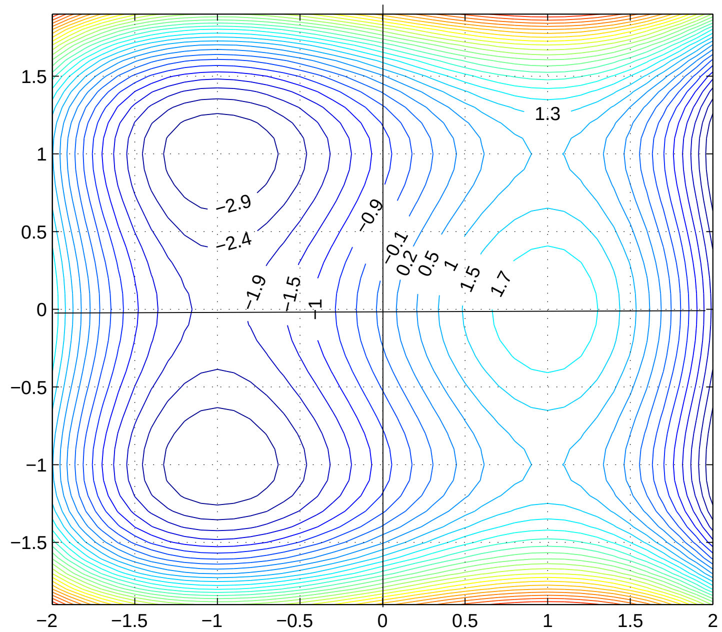
como queriamos probar. Luego, x^* es un minimizador local estricto. □

Ejercicios.

1. Encontrar, si existen, los maximizadores y minimizadores locales de la función $f(x_1, x_2) = x_1^3 + x_2^3 - 3x_1x_2$. Existen maximizadores y minimizadores globales?.

2. Sea $f(x_1, x_2) = (x_2 - x_1^2)^2 + x_1^5$. Mostrar que f tiene un único punto estacionario que no es maximizador ni minimizador de f.

3. En la siguiente figura se presentan las curvas de nivel de la función $f(x, y) = 3x - x^3 - 2y^2 + y^4$. Usar las curvas de nivel para predecir si los puntos estacionarios son minimizadores, maximizadores o puntos de silla.



1.2. Métodos de descenso

Por el teorema que nos da condiciones necesarias de primer orden sabemos que si $\nabla f(x) \neq 0$ entonces x no es minimizador local de f y existe $z \in \mathbb{R}^n$ tal que $f(z) \leq f(x)$. Luego, para buscar un minimizador local, la estrategia más natural es, dado $x_k \in \mathbb{R}^n$, buscar x_{k+1} tal que $f(x_{k+1}) < f(x_k)$. Una manera de realizar esto es tomar una dirección d_k a partir de x_k en la que sepamos que f desciende y moverse un cierto paso t_k en esa dirección para obtener x_{k+1} de la forma $x_{k+1} = x_k + t_k d_k$ de modo que $f(x_{k+1}) < f(x_k)$. Los métodos que realizan este procedimiento sucesivamente se denominan *métodos de descenso.*

Definición 1.6. *Una dirección $d \in \mathbb{R}^n$ es una dirección de descenso de f en x si existe $\varepsilon > 0$ tal que $f(x + td) < f(x), \forall t \in (0, \varepsilon]$.*

En la dirección d se obtiene un decrecimiento de f, si $t > \varepsilon$ no hay más decrecimiento.

Lema 1.1. *Sea f diferenciable en x.*

1. *Si $\nabla f(x)^T d < 0$ entonces d es una dirección de descenso de f en x.*

2. Si d es una dirección de descenso de f en x entonces $\nabla f(x)^T d \leq 0$.

Dem.

Para la parte 1., como f es diferenciable, tenemos que

$$\nabla f(x)^T d = \lim_{t \to 0} \frac{f(x+td) - f(x)}{t} < 0.$$

Entonces, para $t > 0$ suficientemente pequeño se tiene que $f(x+td) < f(x)$ como queriamos probar.

Para la parte 2., tenemos que

$$f(x+td) - f(x) = \nabla f(x)^T dt + o(t) < 0$$

luego, para $t > 0$,

$$\frac{f(x+td) - f(x)}{t} = \nabla f(x)^T d + \frac{o(t)}{t} < 0.$$

Haciendo $t \to 0^+$ obtenemos que $\nabla f(x)^T d \leq 0$. □

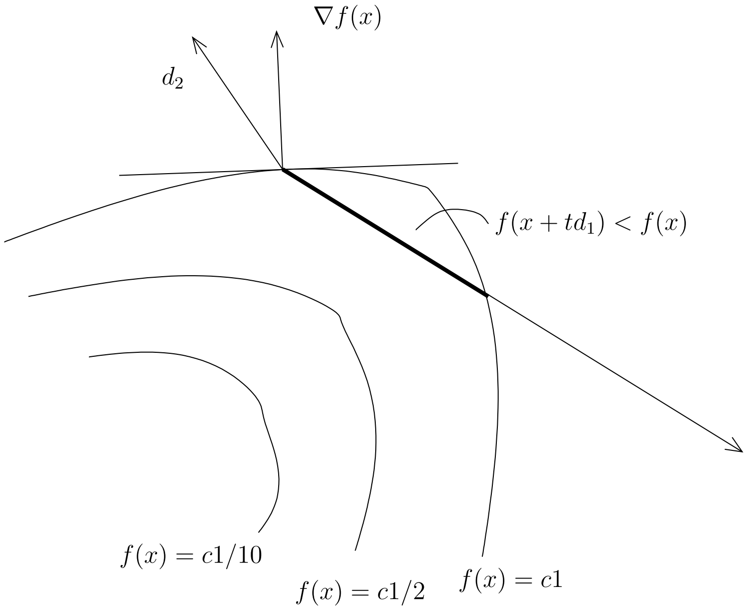


Figura 2: En el ejemplo, podemos ver las curvas de nivel de cierta función f. d_1 es dirección de descenso para f a partir de x pero d_2 no lo es.

En la figura 2 se muestran dos direcciones para una función f y sus curvas de nivel.

Observación. Por un lado, si $\nabla f(x) \neq 0$ entonces el producto $\nabla f(x)^T d < 0$ quiere decir que en esa dirección podemos encontrar puntos donde el valor funcional de f es menor que en $f(x)$.

Por otro lado, $\nabla f(x)^T d < 0$ quiere decir que el ángulo que se forma entre d y $\nabla f(x)$ es mayor a 90^o. Asi, la dirección de máximo descenso se obtiene cuando el ángulo toma el valor de 180^o. Esta dirección es $d = -\nabla f(x)$ y es llamada *dirección de Cauchy* o dirección de máximo descenso. Tiene la desventaja de que, cuando las curvas de nivel de la función objetivo son muy alargadas, la dirección es “casi” ortogonal a la dirección que hace llegar al minimizador rapidamente generando un proceso en zig-zag.

El esquema iterativo o modelo de algoritmo general basado en direcciones de descenso consiste en los siguientes pasos:

Algoritmo general. Dado $x_0 \in \mathbb{R}^n$ inicial, en cada paso k:

1. definir una dirección d_k de descenso para f en x_k;
2. encontrar $t_k > 0$ (tamaño del paso) tal que $f(x_k + t_k d_k) < f(x_k)$;
3. definir $x_{k+1} = x_k + t_k d_k$;
4. Determinar si x_{k+1} es solución.

Como las direcciones son de descenso, siempre es posible encontrar una sucesión de pasos t_k en las condiciones planteadas anteriormente. En estas condiciones, la sucesión de valores funcionales $f(x_k)$ es decreciente, esto quiere decir que, en los diferentes pasos iterativos los puntos x_k se encuentran en curvas de nivel asociadas a valores cada vez menores. Esto justifica la esperanza de que la sucesión x_k pueda converger a una solución (aunque en general, solo podremos garantizar convergencia a puntos estacionarios.)

En lo que sigue de esta sección intentaremos definir los pasos que están en abierto en el esquema general.

El punto inicial x_0 debe ser un dato del problema.

El criterio de finalización para determinar si un nuevo punto es solución puede basarse en la pregunta de si $\|\nabla f(x_{k+1})\| \leq \varepsilon$ para algun parámetro ε inicial y fijo.

Nos resta estudiar como calcular posibles direcciones d_k y como calcular las longitudes de paso t_k.

Supongamos que d_k es una dirección de descenso de f a partir de x_k, existen diferentes reglas para determinar t_k, estas reglas se denominan reglas de búsqueda lineal.

1.3. Búsqueda lineal

- Búsqueda lineal exacta.

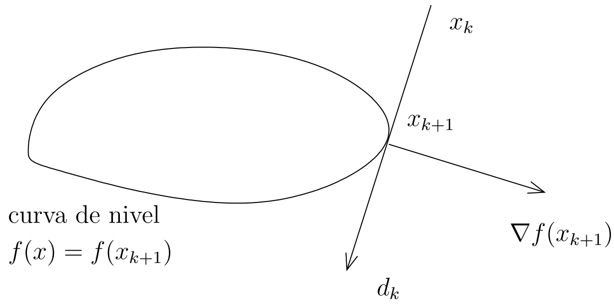
Esta estrategia consiste en buscar el minimizador de f en la semirecta $x_k + td_k$ para $t \geq 0$, es decir:

$$t_k = argmin_{t\geq 0} f(x_k + td_k).$$

Vimos que, si f es diferenciable tenemos que, en la dirección d_k se cumple que

$$\varphi^{'}(t_k) = \nabla f(x_{k+1})^T d_k = 0.$$

Si $\nabla f(x_{k+1}) \neq 0$ entonces x_{k+1} es el punto intersección de la semirecta que parte de x_k en la dirección d_k con la curva de nivel de f que pasa por x_{k+1}. Si bien es la longitud de paso optimo puede ser muy difícil de calcular en forma exacta.

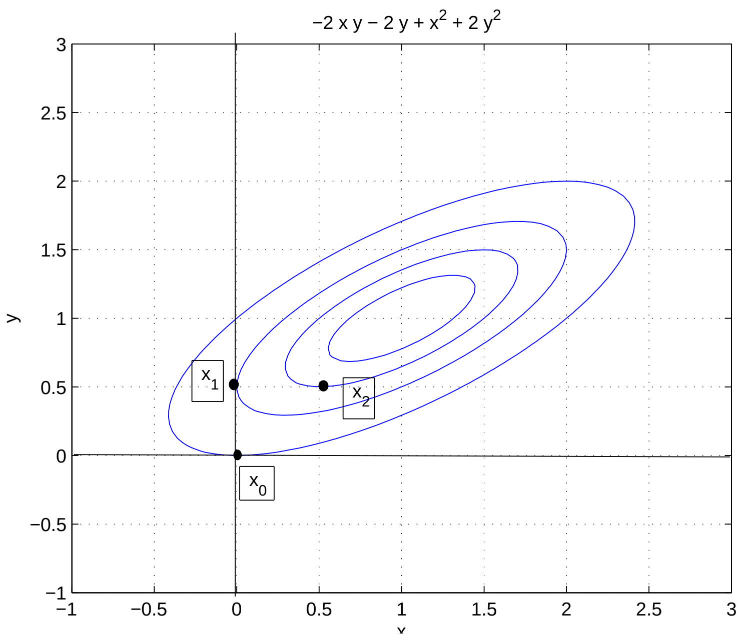


t_k es el mejor paso ya que es el que da el valor mínimo de f en la dirección

Veamos la expresión de esta longitud de paso en el caso en que f es una cuadrática. Sea $f(x) = \frac{1}{2}x^T Ax + b^T x$. Dada una dirección d, buscamos que valor t tal que $\nabla f(x + td)^T d = 0$. Como $\nabla f(x) = Ax + b$ tenemos que $(A(x + td) + b)^T d = 0$. Entonces $x^T Ad + td^T Ad + b^T d = 0$, luego $t = -\dfrac{\nabla f(x)^T d}{d^T Ad}$.

Ejemplo. Realizar dos iteraciones del Método del gradiente con búsqueda exacta a partir del punto $x_0 = (0, 0)$ para la función $f(x, y) = -2xy - 2y + x^2 + 2y^2$.

A partir de x_0 se obtiene que $d_0 = -\nabla f(x_0) = (0\ 2)^T$. Luego, $g(t) = f(x_0 + td_0) = -4t + 8t^2$ y tiene su minimizador en $t = \frac{1}{4}$. Luego, $x_1 = x_0 + td_0 = (0\ \frac{1}{2})$. Luego, $d_1 = -\nabla f(x_1) = (1\ 0)^T$, $g(t) = f(x_1 + td_1) = -t + t^2 - 0{,}5$ y tiene su minimizador en $t = \frac{1}{2}$. Luego, $x_2 = x_1 + td_1 = (\frac{1}{2}\ \frac{1}{2})$. Observar los iterados obtenidos en la figura en la que se observan las curvas de nivel de la función objetivo.



- Regla de Armijo

Esta búsqueda es inexacta y se basa en calcular una longitud de paso que de un "descenso suficiente" de f en relación al valor $f(x_k)$.

Consideramos f diferenciable, x_k, d_k dados y fijos. Definimos $\varphi(t) = f(x_k + td_k)$. Para $\sigma_1 \in (0,1)$ fijo, consideramos los t tales que

$$\varphi(t) \leq \varphi(0) + \sigma_1 t\varphi^{'}(0).$$

Luego, en términos de f se tiene la denominada condición de Armijo

$$f(x_k + td_k) \leq f(x_k) + \sigma_1 t\nabla f(x_k)^T d_k. \tag{2}$$

La reducción que exige esta condición es proporcional a la longitud del paso t_k y al valor de la derivada direccional. Ver figura 3.

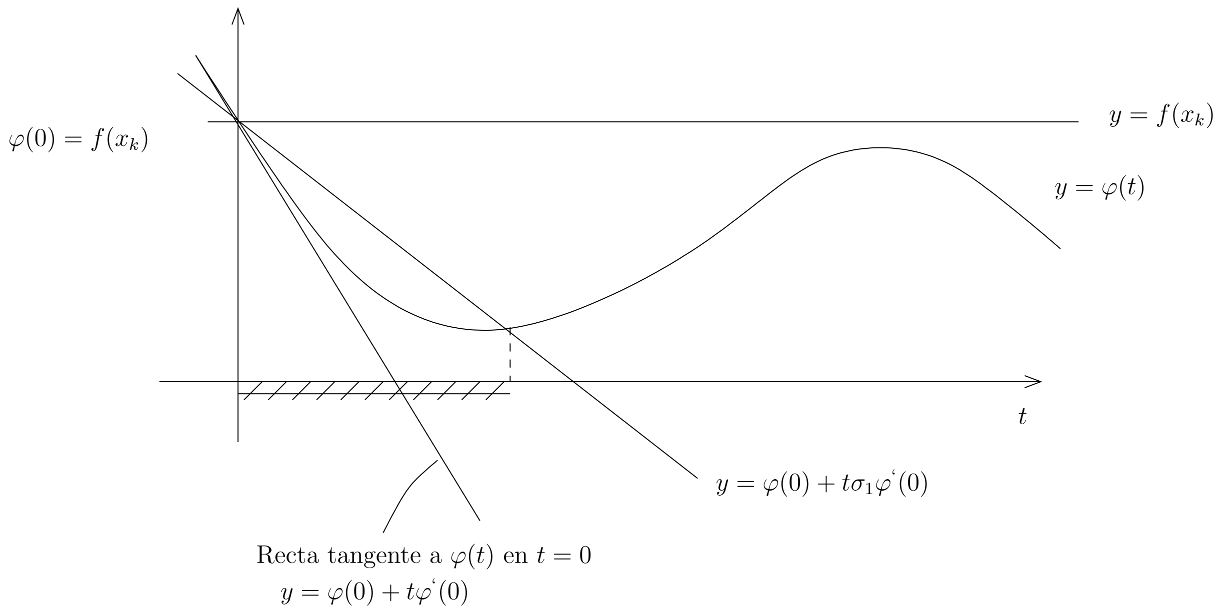


Figura 3: Regla de Armijo.

En general, en un proceso algoritmico, se procede de la siguiente manera:

1. Comenzar con $t = 1$.

2. Si $f(x_k + td_k) \leq f(x_k) + \sigma_1 t\nabla f(x_k)^T d_k$ entonces definir $t_k = t$;

3. si no, reducir t con algún criterio hasta que se cumpla la condición.

Posibles maneras de reducir t:

1. hacer $t = t/2$;

2. calcular $t \in [0.1t, 0.9t]$. (puede ser el punto medio del intervalo)

Lema 1.2. *Dados x_k, d_k tales que $\nabla f(x_k)^T d_k < 0, \sigma_1 \in (0,1)$ fijo entonces existe $\varepsilon = \varepsilon(\sigma_1)$ tal que (2) se cumple para todo $t \in (0, \varepsilon]$.*

Dem.

Como $\nabla f(x_k)^T d_k = \lim_{t\to 0} \dfrac{f(x_k+td_k)-f(x_k)}{t} < 0$ entonces $\lim_{t\to 0} \dfrac{f(x_k+td_k)-f(x_k)}{t\nabla f(x_k)^T d_k} = 1$.

Por definición de límite, tenemos que existe $\varepsilon > 0$ tal que, para todo $t \in (0, \varepsilon]$

$$\frac{f(x_k+td_k)-f(x_k)}{t\nabla f(x_k)^T d_k} \geq \sigma_1$$

como queriamos probar. □

Si bien esta búsqueda inexacta es la más popular, puede no asegurar que se haga un progreso razonable ya que se pueden tomar pasos (valores de t_k) demasiado chicos.

Ejercicios.

1. Sean $f : \mathbb{R}^n \to \mathbb{R}, x, d \in \mathbb{R}^n, \lambda > 0$ tales que $x + \lambda d$ cumple la condición de Armijo. Sea $0 < \mu < \lambda$. Cumple μ la condición de Armijo? Pruebelo o de un contraejemplo.

2. Mostrar que si f es una función cuadrática entonces se verifica Armijo con $\sigma_1 = 0.5$.

- Regla de Wolfe

Esta condición, además de pedir la condición de Armijo, exige otra condición para evitar valores de t_k muy pequenõs, es decir, garantiza mayor desplazamiento, la posibilidad de dar pasos más largos.

Se pide que el valor de t deseado cumpla además la condición $\varphi'(t) \geq \sigma_2 \varphi'(0)$. En términos de f esta condición equivale a

$$\nabla f(x_k+td_k)^T d_k \geq \sigma_2 \nabla f(x_k)^T d_k. \tag{3}$$

Es decir, la pendiente de la recta tangente a $\varphi(t)$ en el nuevo t debe ser mayor o igual a una proporción de la pendiente a $\varphi(t)$ en $t = 0$. En general se consideran valores $0 < \sigma_1 < \sigma_2 < 1$ con $\sigma_1 < 1/2$. Ver figura 4.

Luego, cuando hablamos de Regla de Wolfe nos referimos a las condiciones (2) y (3) juntas.

Proposición 1.1. *Sea f una función $\mathcal{C}^1$ y acotada inferiormente en la recta $x_k + td_k$ siendo d_k tal que $\nabla f(x_k)^T d_k < 0$ y $0 < \sigma_1 < \sigma_2 < 1, \sigma_1 < 1/2$. Entonces existe t_k que verifica la regla de Wolfe.*

Dem.

Como la función $\varphi(t) = f(x_k + td_k)$ es acotada inferiormente existe t_f el primer valor para el cual φ intersecta a la recta $y = f(x_k) + t\sigma_1 \nabla f(x_k)^T d_k$. Luego, $f(x_k + t_f d_k) = f(x_k) + t_f \sigma_1 \nabla f(x_k)^T d_k$.

Usando el Teorema del Valor Medio tenemos que existe t_c entre 0 y t_f tal que

$$f(x_k+t_f d_k) - f(x_k) = t_f \nabla f(x_k + t_c d_k)^T d_k$$

y entonces

$$\nabla f(x_k + t_c d_k)^T d_k = \sigma_1 \nabla f(x_k)^T d_k > \sigma_2 \nabla f(x_k)^T d_k.$$

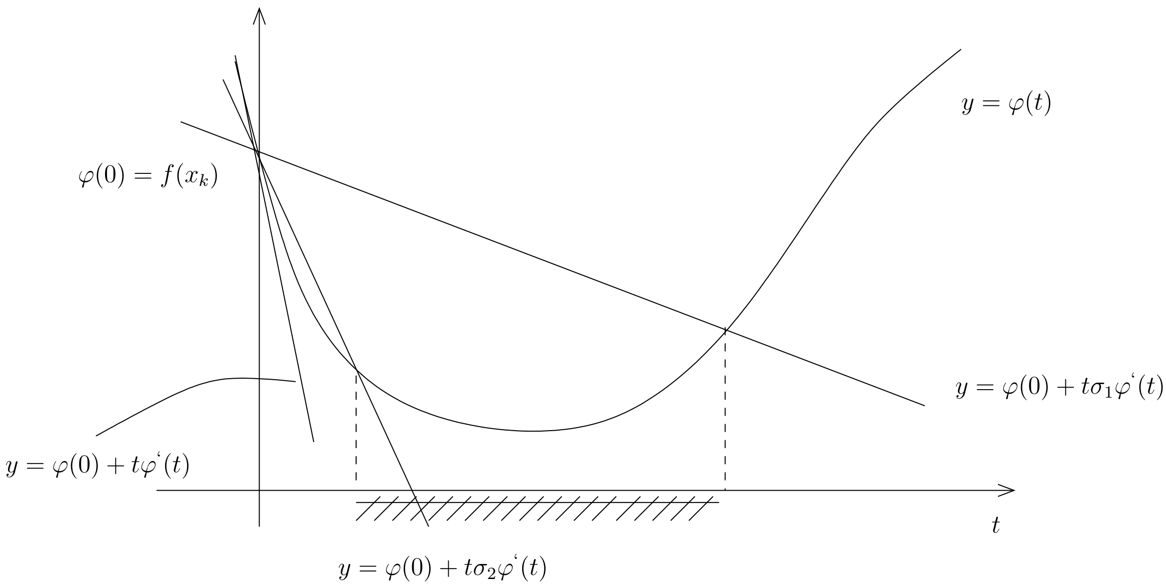


Figura 4: Regla de Wolfe.

Luego, existe $t_c < t_f$ tal que

$$\nabla f(x_k + t_c d_k)^T d_k > \sigma_2 \nabla f(x_k)^T d_k$$

y además, por definición de t_f tenemos que t_c cumple también la condición de Armijo. $\square$

Procedimiento práctico para verificar la regla de Wolfe.

Consideramos fijos $d_k, \nabla f(x_k), x_k, 0 < \sigma_1 < \sigma_2 < 1, \sigma_1 < 1/2$. Definimos un intervalo $[t_{min}, t_{max}]$. Tomamos un $t_{inicial} \in (t_{min}, t_{max})$.

En un paso iterativo iterativo hacemos lo siguiente:

1. Si t cumple (2) entonces ir a [2.].

Si no vale (2), definir $t_{max} = t$ y calcular un $t_{nuevo} \in (t_{min}, t_{max})$ e ir a [1.] con $t = t_{nuevo}$.

2. Si t cumple (3) terminar.

Si no vale (3), definir $t_{min} = t$ y calcular un $t_{nuevo} \in (t_{min}, t_{max})$ e ir a [1.] con $t = t_{nuevo}$.

El no cumplimiento de (2) significa que el actual valor de t es demasiado grande, por esta razón se modifica el t_{max}. El no cumplimiento de (3) significa que el actual valor de t es demasiado pequeño, por esta razón se modifica el t_{min}.

- Sobre la convergencia de las sucesiones

Consideremos un esquema iterativo basado en direcciones de descenso con búsqueda lineal. Dado que estamos solo haciendo uso del gradiente de la función objetivo no podriamos garantizar más que convergencia a un punto estacionario, es decir, a un punto de gradiente nulo.

Teorema 1.7. *Sea f una función $\mathcal{C}^1$ y acotada inferiormente en el conjunto $\mathcal{A} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$. Suponemos que se genera una sucesión $\{x_k\}$ cumpliendo las condiciones (2) y (3) usando d_k de descenso. Si $\nabla f(x_k)$ es Lipschitz continuo en $\mathcal{A}$ con constante L entonces*

$$\sum cos^2(\theta_k)\|\nabla f(x_k)\|^2 < \infty \tag{4}$$

siendo $cos(\theta_k)$ el ángulo que forman d_k y $-\nabla f(x_k)$.

Una demostración puede encontrarse en [1].

La condición (4) se llama condición de Zoutendijk.

Observemos que, si d_k es tal que $cos(\theta_k) > \mu > 0$ entonces se tiene, por (4), que $\|\nabla f(x_k)\| \to 0$. Luego, si $\{x_k\}$ tiene una subsucesión convergente con punto límite x_* entonces $\nabla f(x_*) = 0$, es decir, ese punto límite es un punto estacionario de f.

Por ejemplo, si elegimos $d_k = -\nabla f(x_k)$, la dirección de máximo decrecimiento o dirección de Cauchy, se tiene un método globalmente convergente:

Método del gradiente o Método de Cauchy

1. Dado x_0 inicial, en cada paso k hacer los siguientes pasos.
2. Calcular $d_k = -\nabla f(x_k)$.
3. Calcular t_k mediante Wolfe en la dirección d_k y definir $x_{k+1} = x_k + t_k d_k$. Si $\nabla f(x_{k+1}) = 0$ parar. Si no, continuar el proceso.

Si el algoritmo tiene una subsucesión de iterados $k \in K$ para los que $cos(\theta_k) > \mu > 0$ entonces $\|\nabla f(x_k)\| \to_{k \in K} 0$. Por ejemplo, este caso aparece cuando se considera la dirección de Cauchy cada una cantidad fija de iteraciones.

Si d_k es calculada mediante una fórmula del tipo $d_k = -B_k^{-1}\nabla f(x_k)$ siendo B_k simétrica definida positiva tenemos que d_k es una dirección de descenso que además cumple la condición de que el coseno esta acotado. Se puede verificar que

$$cos(\theta_k) = -\frac{\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|\|d_k\|} = -\frac{\nabla f(x_k)^T B_k^{-1}\nabla f(x_k)}{\|\nabla f(x_k)\|\|B_k^{-1}\nabla f(x_k)\|} \geq \frac{\lambda_{min}(B_k^{-1})}{\|B_k^{-1}\|} = \frac{1}{\|B_k\|\|B_k\|^{-1}} = \mathcal{K}_{B_k}.$$

$\mathcal{K}_{B_k}$ es el número de condición de B_k. Luego, si este número esta acotado inferiormente se tiene una cota para el coseno lo que garantiza convergencia global si se utiliza la condición de Wolfe.

Si en lugar de utilizar Wolfe se utilizara Armijo en la búsqueda del Método del Gradiente también se puede probar convergencia:

Método de Cauchy con búsqueda de Armijo

1. Dado x_0 inicial, en cada paso k hacer los siguientes pasos.
2. Calcular $d_k = -\nabla f(x_k)$.
3. Calcular t_k mediante Armijo en la dirección d_k y definir $x_{k+1} = x_k + t_k d_k$. Si $\nabla f(x_{k+1}) = 0$ parar. Si no, continuar el proceso.

Teorema 1.8. *Sea f una función $\mathcal{C}^1$ y acotada inferiormente. Entonces, todo punto de acumulación de la sucesión $\{x_k\}$ generada por el método del gradiente con búsqueda de Armijo es un punto estacionario de f.*

Ejercicio. Considere la función $f(x, y) = x - y + 2x^2 + 2xy + y^2$.

1. Muestre que $d = (-1, 0)$ es una dirección de descenso para f en $(0, 0)$. Analizar cual es el paso óptimo que se puede dar en esa dirección para hacer decrecer el valor de f utilizando búsqueda exacta.

2. Para la dirección de máximo decrecimiento en $(0, 0)$ determinar el intervalo de paso máximo que se puede dar en esa dirección a partir de $(0, 0)$ para hacer decrecer el valor de f utilizando la regla de Armijo con parámtero $\sigma_1 = 1/4$.

1.4. Dirección de Newton

En su forma clásica, el método de Newton es un método para resolver sistemas de ecuaciones no lineales, es decir, no fue originariamente pensado como un método de minimización.

Originalmente se introdujo para resolver el problema de encontrar $x \in \mathbb{R}^n$ tal que $F(x) = 0$ donde $F : \mathbb{R}^n \to \mathbb{R}^n$. Si x_k es una aproximación de la solución x^* del sistema, en un entorno de x_k podemos aproximar la ecuación $F(x) = 0$ por su linealización (polinomio de Taylor de orden 1 de F alrededor de x_k):

$$F(x_k) + JF(x_k)(x - x_k) = 0 \tag{5}$$

y buscar el punto que resuelve este sistema.

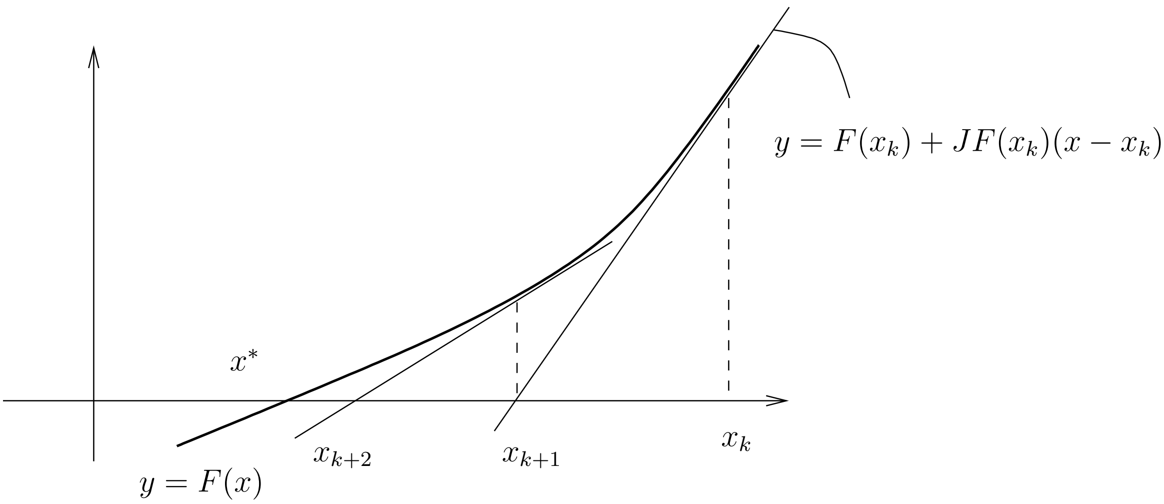


Figura 5: Método de Newton para ecuaciones.

Método de Newton para sistemas de ecuaciones:

1. Dado x_0 inicial,

2. calcular x_{k+1} solución de (5),

3. hacer $k = k + 1$ y repetir.

Observar la figura 5. El sistema (5) es un sistema lineal de ecuaciones con matriz $JF(x_k)$ que, para que tenga solución, debe ser $JF(x_k)$ no singular. Resolver este sistema lineal es una de las principales desventajas del método de Newton. Se puede realizar mediante factorizaciones de la matriz o usando métodos iterativos que resuelven sistemas lineales de ecuaciones.

Nos interesa ahora usar la filosofía del método de Newton para minimización irrestricta. Sea f una función $\mathcal{C}^2$. La aproximación cuadrática de f en un punto x_k es $q_k(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) + 1/2(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$. Podemos buscar la dirección d_k que nos lleve al minimizador de q_k. Sea $d = x - x_k$, luego $q_k(x) = q_k(x_k + d) = f(x_k) + \nabla f(x_k)^T d + 1/2 d^T \nabla^2 f(x_k) d$. Buscamos d_k tal que $\nabla q_k(x_k + d) = 0$, luego, $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$. Esta dirección se denomina dirección de Newton.

Una vez obtenida la dirección de Newton, nos preguntamos si es de descenso. Infelizmente d_k puede no ser de descenso si $\nabla^2 f(x_k)$ no es definida positiva. Por ejemplo la función $f(x, y) = (1/2)(x^2 - y^2)$ en el punto $x_0 = (0, 1)$ verifica

$$\nabla f(x_0) = (0 \ \ -1)^T, \nabla^2 f(x_0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

En este caso la dirección de Newton es $d_0 = (0 \ -1)$ y $\nabla f(x_0)^T d_0 = 1 > 0$.

Ejemplo 2. Para la función $f(x, y) = -2xy - 2y + x^2 + 2y^2$, comenzando desde $x_0 = (0, 0)$ tenemos que $d_0 = -(\nabla^2 f(x_0))^{-1} \nabla f(x_0) = (1, 1)$ y $x_1 = x_0 + d_0 = (1, 1)$ que es la solución. Es de esperar que esto suceda ya que la función f es cuadrática.

Método de Newton con búsqueda lineal de Armijo.

1. Dado x_0 inicial, en cada paso k hacer los siguientes pasos.
2. Calcular $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$.
3. Calcular t_k mediante Armijo en la dirección d_k y definir $x_{k+1} = x_k + t_k d_k$. Si $\nabla f(x_{k+1}) = 0$ parar. Si no, continuar el proceso.

La dirección del segundo paso del método se encuentra resolviendo el sistema lineal $\nabla^2 f(x_k) d_k = -\nabla f(x_k)$. Notar que este paso puede no estar bien definido si $\nabla^2 f(x_k)$ es singular.

Teorema 1.9. *Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función $\mathcal{C}^3$. Sea x_* es un minimizador local de f tal que $\nabla^2 f(x_*)$ es definida positiva. Entonces existe $\varepsilon > 0$ tal que, si $\|x_0 - x^*\| < \varepsilon$ y $t_k = 1$ para todo k en la búsqueda de Armijo entonces la sucesión $\{x_k\}$ generada por el método anterior verifica:*

1. *$\nabla^2 f(x_k)$ es definida positiva para todo k,*
2. *$x_k \to x^*$.*

Dem.

Ver Luenberguer, [2]. □

Este es un teorema de convergencia local que nos dice que si empezamos con x_0 suficientemente cerca del minimizador local x^* entonces:

1. Los sistemas que definen el cálculo de las direcciones estan bien definidos,
2. toda la sucesión converge al minimizador local.

El método de Newton se puede modificar también cuando hay no singularidad de la matriz Hessiana.

Newton con modificación de la diagonal.

1. Dado x_0 inicial, en cada paso k hacer los siguientes pasos.
2. Si $\nabla^2 f(x_k) > 0$ entonces $B_k = \nabla^2 f(x_k)$.
3. Si no, $B_k = \nabla^2 f(x_k) + \mu_k I$ con μ_k elegido de modo que B_k sea definida positiva.
4. Calcular $d_k = -(B_k)^{-1}\nabla f(x_k)$.
5. Definir $x_{k+1} = x_k + t_k d_k$ mediante búsqueda de Wolfe. Si $\nabla f(x_{k+1}) = 0$ parar. Si no, continuar el proceso.

Este es un método práctico donde se modifica la diagonal de la matriz Hessiana de modo de obtener una nueva matriz semejante que sea definida positiva.

Teorema 1.10. *Sea f una función $\mathcal{C}^2$ y supongamos que cada matriz B_k verifica $\|B_k\|\|B_k^{-1}\| \leq C$ para $C > 0$ y para todo k. Entonces todo punto de acumulación de x_k es punto estacionario de f.*

Dem.

Sale usando la condición de Zoutendijk.

Vimos que la dirección de Newton es calculada como solución de un sistema lineal de $n \times n$:

$$\nabla^2 f(x_k) d = -\nabla f(x_k).$$

Una manera de resolver este sistema puede ser mediante factorizaciones de $\nabla^2 f(x_k)$ aunque, en principio, no sepamos si esta matriz es definida positiva. Podriamos por ejemplo, intentar la factorización de Cholesky de la matriz. Si es posible, se tiene que $\nabla^2 f(x_k) = LDL^T$ y la dirección se encuentra mediante la solución de dos sistemas triangulares. Si no es posible obtener la factorización de Cholesky de la matriz Hessiana quiere decir que esta matriz no es definida positiva y en ese caso una posibilidad es aumentar la diagonal de $\nabla^2 f(x_k)$ como se propuso previamente. Esta nueva matriz es definida positiva y tiene factorización de Cholesky.

Otra manera es usar un método iterativo y terminar en alguna iteración con alguna solución inexacta pero aproximada. Muchas de estas reglas iterativas están basadas en la medición del residuo del sistema

$$r_k = \nabla^2 f(x_k) d_k + \nabla f(x_k)$$

donde d_k es la dirección buscada llamada dirección de Newton inexacta.

Como r_k cambia si f se multiplica por una constante (esto quiere decir que no es invariante) entonces se considera su tamaño relativo al tamaño de f que es determinada mediante $\|\nabla f(x_k)\|$. Asi, el proceso iterativo se finaliza cuando se encuentra una dirección d_k que cumple

$$\|r_k\| \leq \eta_k \|\nabla f(x_k)\|$$

donde $\{\eta_k\}$ es tal que $0 < \eta_k < 1$.

Si $\|r_k\|$ es chico, entonces d_k cumple $\nabla^2 f(x_k) d_k \approx -\nabla f(x_k)$.

Método de Newton Inexacto.

1. Dado x_0 inicial, en cada paso k hacer los siguientes pasos.

2. Calcular d_k tal que $\|r_k\| \leq \eta_k \|\nabla f(x_k)\|$ y $d_k^T \nabla f(x_k) < 0$.

3. Definir $x_{k+1} = x_k + t_k d_k$ mediante búsqueda de Wolfe. Si $\nabla f(x_{k+1}) = 0$ parar. Si no, calcular $\nabla^2 f(x_{k+1})$ y continuar el proceso.

Una posible elección de η_k es $\eta_k = \min\{0{,}5, \|\nabla f(x_k)\|\}$. Una manera de calular d_k es aplicar el método de gradientes conjugados al problema de minimizar la cuadrática $q(d) = 1/2 d^T \nabla^2 f(x_k)^T d + \nabla f(x_k) d$. El método de Newton que utiliza gradientes conjugados para el cálculo de la dirección se denomina *Método de Newton Truncado.*

Brevemente, el *método de Gradientes Conjugados* es un método iterativo para resolver sistemas lineales de ecuaciones del tipo

$$Ax = b, \qquad A \in \mathbb{R}^{n\times n}, \text{ simétrica definida positiva.}$$

Resolver este problema es equivalente a encontrar un minimizador local de la cuadrática

$$q(x) = \frac{1}{2} x^T A x - b^T x.$$

Es decir, el método de Gradientes Conjugados puede ser interpretado tanto como un algoritmo para resolver sistemas lineales de ecuaciones como un algoritmo para encontrar el minimizador de cuadráticas convexas.

En cada paso k del método de gradientes conjugados, dado el iterado p_k, la dirección d_k se obtiene como una combinación de la dirección previa d_{k-1} con menos el gradiente de la cuadrática en p_k $(-\nabla q(p_k))$ es decir

$$d_k = -\nabla q(p_k) + \beta_k d_{k-1}$$

donde el parámetro β_k de la combinación se define de modo que la dirección d_k sea A-conjugada con d_{k-1}, es decir, queremos que

$$d_k^T A d_{k-1} = 0.$$

Luego,

$$\beta_k = \frac{\nabla q(p_k)^T A d_{k-1}}{d_{k-1}^T A d_{k-1}}$$

Método de Gradientes Conjugados.

Dados: $p_0 = 0$, $g_0 = \nabla q(p_0), d_0 = -g_0, \varepsilon > 0$. Denotamos $g_k = \nabla q(p_k)$.

$(*)$ Si $\|g_k\| < \varepsilon$ parar.

Sino (Busca el minimizador de $q(x)$ en la dirección d_k)

Define el paso en la dirección $d_k, t_k = -\dfrac{g_k^T d_k}{d_k^T A d_k}$

Define el nuevo punto $p_{k+1} = p_k + t_k d_k$

Calcula el nuevo gradiente $g_{k+1} = \nabla q(p_{k+1}) = g_k + t_k A d_k$

Define el parámetro $\beta_{k+1} = \dfrac{\nabla q(p_{k+1})^T A d_k}{d_k^T A d_k}$ (hace que las direcciones sean conjugadas)

Define la nueva dirección de búsqueda $d_{k+1} = -\nabla q(p_{k+1}) + \beta_{k+1} d_k$. Ir a $(*)$

Se demuestran las siguientes propiedades.

Teorema 1.11. *Suponemos que p_k no es la solución, es decir $g_k \neq 0$, entonces*

i) $\overline{\{g_0, g_1, \ldots, g_k\}} = \overline{\{g_0, Ag_0, \ldots, A^k g_0\}}$

ii) $\overline{\{d_0, d_1, \ldots, d_k\}} = \overline{\{g_0, Ag_0, \ldots, A^k g_0\}}$

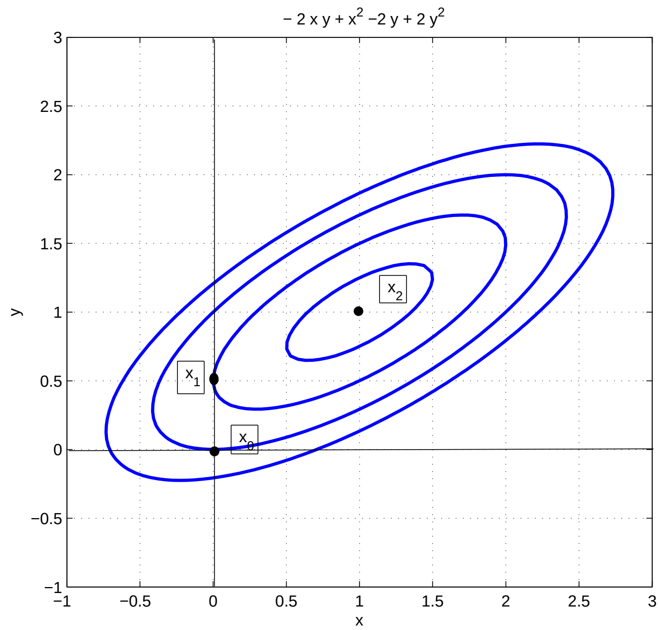
iii) Las direcciones son A conjugadas: $d_k^T A d_i = 0, \forall i = 0, \ldots, k-1.$

iv) $g_k^T d_i = 0, \forall i = 0, \ldots, k-1$

v) $g_k^T g_i = 0, \forall i = 0, \ldots, k-1$

vi) $p_k \to x^*$ *en a lo sumo n pasos.*

Veamos el proceso del algoritmo cuando se aplica a la función cuadrática $f(x, y) = -2xy - 2y + x^2 + 2y^2$. a partir del punto $p_0 = (0, 0)$. Para esta función tenemos que $\nabla f(x, y) = \begin{pmatrix} -2y + 2x \\ -2x - 2 + 4y \end{pmatrix}, A = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$. Aplicando el método se obtiene la siguiente sucesión de iterados: $d_0 = (0\ 2)^T, t_o = \frac{1}{4}, p_1 = (0\ \frac{1}{2})^T, g_1 = (-1\ 0)^T, \beta_1 = \frac{1}{4}, d_1 = (1\ \frac{1}{2})^T, t_1 = 1, p_2 = x^*$.



Para mayor información sobre este método ver [2, 1].

En una iteración del método de Newton Truncado, para calcular la dirección d_k^{NT} del paso 2., se utiliza mientras se pueda Gradientes Conjugados. Definimos la cuadrática $q_k(p) = \frac{1}{2} p^T \nabla^2 f(x_k) p + \nabla f(x_k)^T p$. Luego, mientras se pueda se utilizan las iteraciones de Gradientes

Conjugados. Decimos esto ya que este es un método que se aplica cuando la matriz es definida positiva y en este caso no sabemos si $\nabla^2 f(x_k)$ lo es. Luego, una vez calculada una dirección d_j de Gradientes Conjugados preguntamos si es una dirección de curvatura positiva para la matriz o no. Si se tiene que $d_j^T \nabla^2 f(x_k) d_j \leq 0$ el proceso para porque no se puede seguir utilizando Gradientes Conjugados, en este caso, si es la primera iteración ($j = 0$) el algoritmo da como solución la dirección $d_k^{NT} = -\nabla f(x_k)$, si $j > 0$ da como solución la dirección $d_k^{NT} = p_j$ que es el mejor punto de la cuadrática que se calculó hasta el momento. Si $d_j^T \nabla^2 f(x_k) d_j > 0$ el proceso sigue como el Gradientes Conjugados hasta que se cumple la condición $\|g_{j+1}\| \leq \eta_k \|g_0\|$ que equivale a $\|\nabla q_k(p_{j+1})\| \leq \eta_k \|\nabla f(x_k)\|$. En este caso el algoritmo da como solución la dirección $d_k^{NT} = p_{j+1}$.

Luego, el algoritmo interno para cuando encuentra una dirección de curvatura negativa para la matriz o cuando se cumple una de las condiciones del paso 2. Se puede demostrar que en cualquiera de los dos casos la dirección resultante es una dirección de descenso. Esto se debe a propiedades del método de Gradientes Conjugados. Se puede demostrar además que $cos(\theta_k) \geq C_k$ y luego, si C_k es acotado para todo k se demuestra convergencia cuando se utiliza búsqueda de Wolfe gracias a la condición de Zoutendijk.

Ejercicios.

1. Considere la función $f(x, y) = x - y + 2x^2 + 2xy + y^2$. Utilizar el método de gradientes conjugados para encontrar la solución comenzando en $(0, 0)$.

2. Considere la función $f(x, y) = (x - 2y)^2 + x^4$. Calcular la dirección de Newton en el punto $(2, 1)$. Cumple el valor $t = 1$ la regla de Armijo con parámtero $\sigma_1 = 1/5$?

3. Considere el siguiente método:

   - Dado x_k. Calcular d_k como se indica a continuación.
   - Hacer $t = 1$.
     Si $f(x_k + td_k) \leq f(x_k) + \frac{1}{2} t d_k^T \nabla f(x_k)$ (*) hacer $x_{k+1} = x_k + td_k$,
     Sino, reemplazar t por $t/2$ hasta que se verifique (*)

   Sea $f(x, y) = x^2 + y^2 - xy, x_0 = (2, 0)$

   a) Dibuje algunas curvas de nivel de f.

   b) Hacer dos iteraciones del método utilizando la dirección de Cauchy. Dibuje los iterados obtenidos en el plano en el cual estan las curvas de nivel de f. Que observa?

   c) Resuelva el problema mediante el uso de la dirección de Newton. Que observa? Porque?

1.5. Métodos de Quasi-Newton

En el método del gradiente elegimos $d_k = -I \nabla f(x_k)$. En el método de Newton la dirección d_k verifica $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$. Una desventaja al usar la verdadera matriz Hessiana es que las derivadas segundas pueden ser muy dificiles de calcular manualmente. Una ventaja es que hay convergencia local que esta bien justificada al considerar una aproximación cuadrática de la verdadera función.

Los métodos de Quasi-Newton se definen de manera que se parezcan lo mas posible al método de Newton. Asi, en lugar de considerar el modelo cuadrático usando la verdadera matriz Hessiana definen un modelo cuadrático aproximado considerando la función

$$m_k(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) + 1/2(x - x_k)^T B_k(x - x_k)$$

donde B_k es simétrica definida positiva que, de ser posible, represente una buena aproximación del verdadero Hessiano. Un proceso iterativo general debería ser como el siguiente:

Método General de Quasi-Newton.

1. Dado x_0 inicial, en cada paso k hacer los siguientes pasos.
2. Dada $B_k > 0$, calcular d_k tal que $d_k = -B_k^{-1}\nabla f(x_k)$.
3. Definir $x_{k+1} = x_k + t_k d_k$ mediante búsqueda lineal. Si $\nabla f(x_{k+1}) = 0$ parar. Si no, calcular una nueva matriz B_{k+1} y continuar el proceso.

El nuevo modelo en el punto x_{k+1} es

$$m_{k+1}(x) = f(x_{k+1}) + \nabla f(x_{k+1})^T(x - x_{k+1}) + 1/2(x - x_{k+1})^T B_{k+1}(x - x_{k+1}).$$

Luego, se requiere que la nueva matriz sea tal que $m_{k+1}(x)$ y $f(x)$ tengan gradientes conicidentes en x_k y en x_{k+1}, es decir, como,

$$\nabla m_{k+1}(x) = \nabla f(x_{k+1}) + B_{k+1}(x - x_{k+1})$$

y queremos que $\nabla m_{k+1}(x_k) = \nabla f(x_k)$ debe verificarse que

$$B_{k+1}(x_k - x_{k+1}) = \nabla f(x_k) - \nabla f(x_{k+1}). \tag{6}$$

La ecuación (6) se llama *ecuación secante.* Si definimos $s_k = x_{k+1} - x_k, y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ la ecuación secante se escribe como $B_{k+1}s_k = y_k$.

Al pre multiplicar por s_k se obtiene, usando el Teorema del Valor Medio

$$s_k^T B_{k+1}s_k = s_k^T(\nabla f(x_{k+1}) - \nabla f(x_k)) = s_k^T\nabla^2 f(\bar{x})s_k$$

para $\bar{x}$ un punto intermedio entre x_k y x_{k+1}. Por lo tanto, vemos que la matriz B_{k+1} que cumple la ecuación secante tiene la misma curvatura del Hessiano en un punto intermedio a lo largo del segmento que une x_{k+1} con x_k.

Para encontrar una matriz que cumpla la ecuación secante hay que resolver un sistema de n ecuaciones con $n \times (n + 1)/2$ incógnitas (las entradas de la matriz simétrica B). Luego, hay infinitas matrices que cumplen esta ecuación.

Observar que si B_{k+1} es definida positiva, se tiene que $s_k^T B_{k+1}s_k > 0$ entonces $s_k^T y_k > 0$.

- Fórmulas de Adaptadas Secantes

Es muy frecuente proponer que las matrices B_{k+1} se obtengan a partir de B_k mediante modificaciones de rango 1 o 2.

Las adaptadas de rango 2 tienen la forma

$$B_{k+1} = B_k + \Delta B_k^{'} + \Delta B_k^{''}$$

donde $\Delta B_k^{'}, \Delta B_k^{''}$ son matrices simétricas de rango 1.

Como se debe verificar (6) tenemos que

$$\Delta B_k^{'} s_k + \Delta B_k^{''} s_k = y_k - B_k s_k.$$

Existen muchas maneras de que se cumpla esta condición, una elección posible es

$$\Delta B_k^{'} s_k = y_k, \qquad \Delta B_k^{''} s_k = -B_k s_k.$$

Si queremos que las matrices sean de rango uno entonces podemos elegir

$$\Delta B_k^{'} = \frac{y_k^T y_k}{y_k^T s_k}, \qquad \Delta B_k^{''} = -\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}$$

luego,

$$B_{k+1} = B_k + \frac{y_k^T y_k}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \tag{7}$$

que se denomina fórmula BFGS (Broyden-Fletcher-Goldfarb-Shanno) y es la actualización secante más popular.

Teorema 1.12. *Si B_k es simétrica definida positiva y $s_k^T y_k > 0$ entonces B_{k+1} es simétrica definida positiva.*

Dem.

Es claro que B_{k+1} es simétrica. Sea $z \neq 0$. Luego

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

Como B_k es definida positiva, tenemos que existen matrices U y D tales que $U^T U = I$ y D es una matriz diagonal con $d_{ii} > 0$ tales que $B_k = U^T D U$. Los elementos d_{ii} son los autovalores de la matriz y U es la matriz de autovectores. Luego, podemos escribir $B_k = U^T D^{1/2} U U^T D^{1/2} = B_k^{1/2} B_k^{1/2}$ donde $D^{1/2}$ es la matriz diagonal que tiene como entradas en su diagonal los elementos $\sqrt{d_{ii}}$.

Sean $u = B_k^{1/2} s_k, v = B_k^{1/2} z$. Por la desigualdad de Cauchy-Schwarz sabemos que $|u^T v|^2 \leq \|u\|^2 \|v\|^2$. Luego

$$|z^T B_k^{1/2} B_k^{1/2} s_k|^2 \leq \|B_k^{1/2} s_k\|^2 \|B_k^{1/2} z\|^2.$$

Entonces

$$(z^T B_k s_k)^2 \leq s_k^T B_k s_k z^T B_k z$$

lo que demuestra, usando la hipótesis el teorema que B_{k+1} es semidefinida positiva.

Además, $z^T B_k z = \dfrac{(z^T B_k s_k)^2}{s_k^T B_k s_k}$ si y solo si u y v son múltiplos si y solo si $s_k = \alpha z$ lo que implica que $\alpha z^T y_k = s_k^T y_k \neq 0$ y en ese caso se tiene que $z^T B_{k+1} z = \dfrac{(z^T y_k)^2}{s_k^T y_k} > 0$. □

Luego, si B_k es definida positiva y podemos garantizar que en cada paso $s_k^T y_k > 0$ entonces la siguiente matriz definida mediante la fórmula (7) también será definida positiva y todas las direcciones generadas por un método de Quasi-Newton serán de descenso. Asi, el método está bien definido.

Se tienen los siguientes resultados:

Teorema 1.13. *Consideremos un método de Quasi-Newton que utiliza la fórmula de adaptada secante de rango 2 (7) para las matrices con búsqueda inexacta o de Wolfe. Luego, se tiene que* $s_k^T y_k > 0$ *para todo* k.

Dem.

Supongamos primero que se realiza búsqueda exacta. Luego, $t_k = argmin_{t>0} f(x_k + td_k)$. Si $\varphi(t) = f(x_k + td_k)$ entonces tenemos que $\varphi'(t_k) = 0$, luego, $\nabla f(x_{k+1})^T d_k = 0$.

Asi, como $s_k = x_{k+1} - x_k = t_k d_k$,

$$s_k^T y_k = s_k^T(\nabla f(x_{k+1}) - \nabla f(x_k)) = t_k d_k^T(\nabla f(x_{k+1}) - \nabla f(x_k)) = -t_k d_k^T \nabla f(x_k) = t_k \nabla f(x_k) B_k^{-1} \nabla f(x_k).$$

Por lo tanto, si $B_0 > 0$ entonces $s_0^T y_0 > 0$ entonces B_1 esta bien definida, es definida positiva, entonces $s_1^T y_1 > 0$ y B_2 es definida positiva y asi continua el proceso.

Supongamos ahora que se realiza búsqueda de Wolfe. Luego, por la segunda condición se tiene que, como $\nabla f(x_{k+1})^T d_k > \sigma_2 \nabla f(x_k)^T d_k$,

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k > (\sigma_2 - 1) \nabla f(x_k)^T d_k.$$

Luego, como $d_k = \frac{s_k}{t_k}$ obtenemos que

$$s_k^T y_k = t_k (\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k > t_k(\sigma_2 - 1) \nabla f(x_k)^T d_k.$$

De la misma forma que antes, si $B_0 > 0$ entonces $\nabla f(x_0)^T d_0 < 0$, como $\sigma_2 - 1 < 0$, tenemos que $s_0^T y_0 > 0$ entonces B_1 esta bien definida, es definida positiva, entonces $\nabla f(x_1)^T d_1 < 0$ y $s_1^T y_1 > 0$ y B_2 es definida positiva y asi continua el proceso. □

Ejemplo. Aplicar el método de Quasi-Newton con la fórmula (7) usando búsqueda exacta a partir del punto $x_0 = (0, 0)$ para la función $f(x, y) = -2xy - 2y + x^2 + 2y^2$ comenzando con matriz $B_0 = I$. Observar que se obtiene la siguiente sucesión de iterados $t_0 = \frac{1}{4}, x_1 = (0\ \frac{1}{2})^T, B_1 = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}, d_1 = (1\ \frac{1}{2})^T, t_1 = 1, x_2 = (1\ 1)^T$.

Para el método de quasi-Newton que utiliza la adaptada de rango 2 se puede demostrar el siguiente teorema de convergencia:

Teorema 1.14. *Sea* $f \in \mathcal{C}^2$. *Suponemos que* $S = \{x : f(x) \leq f(x_0)\}$ *es acotado y convexo. Suponemos que* $\nabla^2 f(x) > 0$ *para todo* $x \in \mathbb{R}^n$. *Si* $\{x_k\}$ *es una sucesión generada mediante un método de Quasi-Newton con adaptada (7) con búsqueda de Wolfe entonces* $x_k \to x^*$ *donde* x^* *es el único minimizador global de* f.

Dem.

Ver el libro [3]. □

Existen otras fórmulas de adapatadas secantes de rango 2 y de rango 1 en la literatura, [3, 1].

1.6. Regiones de confianza

Tanto los métodos de búsqueda lineal como los métodos de regiones de confianza generan sus pasos con la ayuda de un modelo cuadrático de la función objetivo, pero usan el modelo de manera diferente. Los métodos de búsqueda lineal los usan para generar una dirección de búsqueda y después enfocan su esfuerzo en encontrar un paso apropiado t en esa dirección. Los métodos de regiones de confianza definen una región alrededor del iterado actual dentro de la cual confian en el modelo como una representación adecuada de la función objetivo y eligen el paso como un minimizador aproximado del modelo en la región de confianza. Podriamos decir que estos últimos eligen la dirección y la longitud de paso simultáneamente.

El tamaño de la región de confianza esta sujeta a la efectividad del paso. Si la región es muy chiquita, el algoritmo puede perder la posibilidad de realizar un paso sustancial para moverse más cerca de un minimizador de la función objetivo. Si la región es muy grande, el minimizador del modelo puede estar lejos del minimizador de la función objetivo en la región, entonces deberiamos reducir la región e intentar otra vez.

Un excelente tratamiento de este método puede encontrarse en el libro [4].

El método de regiones de confianza obtiene el nuevo iterado respecto del anterior como

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

donde p_k es una solución del subproblema cuadrático

$$\begin{array}{c} \text{Minimizar } m_k(p) \\ \text{sujeto a } \|p\|_2 \leq \Delta_k \end{array} \tag{8}$$

donde $m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p$ siendo B_k una matriz simétrica y $\Delta_k > 0$ es el radio de la región de confianza. No se exige que la matriz B_k sea definida positiva ya que al considerar una región de confianza, el subproblema siempre tiene solución. Puede elegirse $B_k = \nabla^2 f(x_k)$ o puede adaptarse recursivamente.

Precisamos explicitar cuando este nuevo iterado es aceptado y como cambia el radio de la región de confianza para la siguiente iteración. Tanto la estrategia de adaptación del radio como la aceptación del punto estan basadas en la siguiente relación entre lo que baja el modelo con lo que baja la función:

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)} = \frac{\text{Ared}}{\text{Pred}}. \tag{9}$$

El numerador se denomina *reducción actual* y el denominador *reducción predecida por el modelo*. Como se minimiza el modelo, la reducción predecida es siempre no negativa.

Supongamos que se ha calculado una solución p_k del subproblema (8). Definimos $x_{new} = x_k + p_k$.

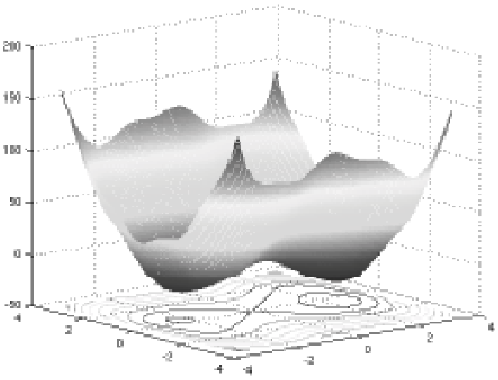
1. Aceptación del nuevo punto.
    - Si $\rho_k > 0{,}1$ entonces $x_{k+1} = x_{new}$.
    - Sino, $x_{k+1} = x_k$.
2. Adaptación del radio de la región de confianza:
    - Si $\rho_k > 0{,}75$ y $\|p_k\| \geq 0{,}8\Delta_k$ entonces $\Delta_{k+1} = 2\Delta_k$.
    - Si $\rho_k > 0{,}75$ y $\|p_k\| \ngeq 0{,}8\Delta_k$ entonces $\Delta_{k+1} = \Delta_k$.
    - Si $0{,}1 < \rho_k < 0{,}75$ entonces $\Delta_{k+1} = \Delta_k$. Si $\rho_k \leq 0{,}1$ entonces $\Delta_{k+1} = \frac{\Delta_k}{2}$.

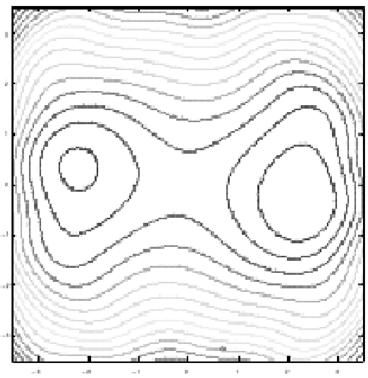
Ejemplo de aplicación del Método de Regiones de Confianza

Consideremos la función

$$f(x, y) = -10x^2 + 10y^2 + 4sen(xy) - 2x + x^4$$

Observar la gráfica de la función y sus curvas de nivel.

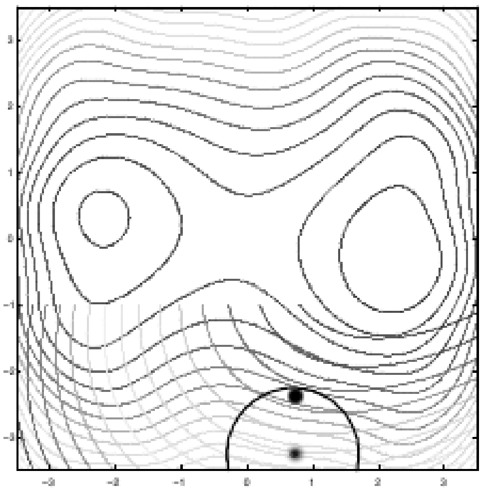




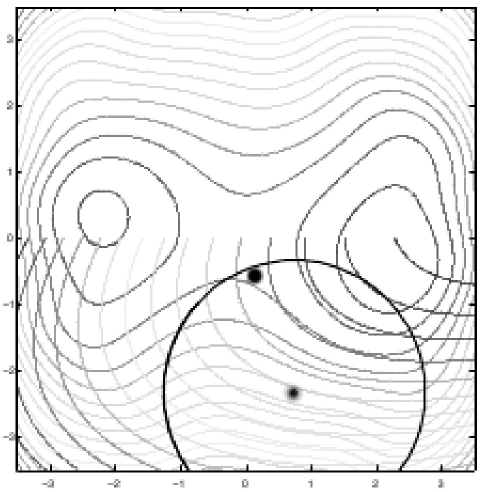
La función tiene dos minimizadores locales: $(-2{,}20, 0{,}32)$ y $(2{,}30, -0{,}34)$.

Se utilizará el Método de Regiones de Confianza para encontrar alguno de los minimizadores locales de f. Comenzamos con el punto incial $x_0 = (0{,}71, -3{,}27)$.

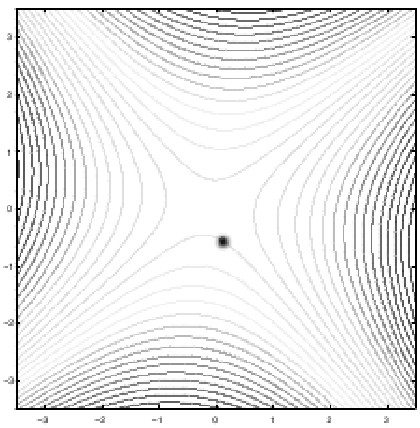
Comenzando en este punto, se genera un modelo m_0 alrededor de x_0 y se busca el minimizador del modelo con el radio de confianza inicial $\Delta_0 = 1$. En la siguiente figura se observan las curvas de nivel de f y de m_0.

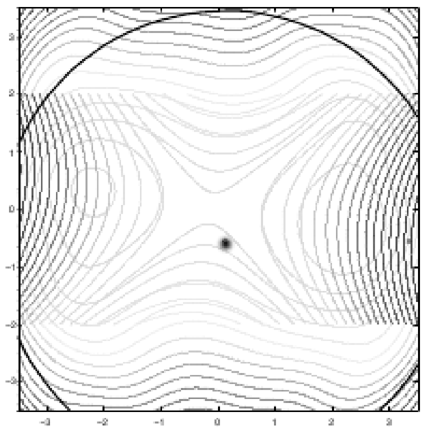


En difuso se observa el punto x_0, en negro una solución aproximada del modelo dentro de la región de confianza. El cociente $\rho_0 = \frac{A_{red}}{P_{red}} = 0{,}998 > 0{,}1$ entonces, el nuevo punto es aceptado como proximo iterado, $x_1 = x_0 + p_0$. Además, como $\rho_0 > 0{,}75$ y $\|p_0\| \geq 0{,}8\Delta_0$ entonces se aumenta el radio de confianza al valor $\Delta_1 = 2\Delta_0 = 2$. En la siguiente figura se observa el nuevo modelo m_1 alrededor del nuevo iterado x_1



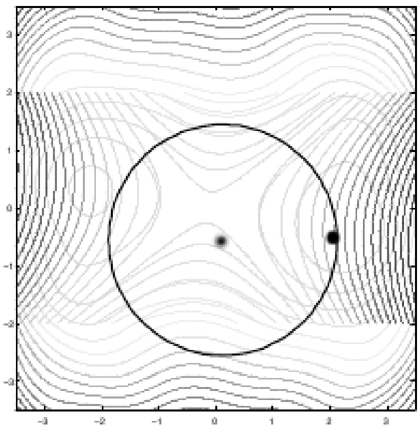
En difuso se observa el punto x_1 (centro de la región de confianza), en negro una solución aproximada del modelo dentro de la región de confianza. El nuevo cociente $\rho_1 = \frac{A_{red}}{P_{red}} = 1{,}354 > 0{,}1$ entonces el nuevo punto es aceptado como proximo iterado $x_2 = x_1 + p_1$ y como $\rho_1 > 0{,}75$ y $\|p_1\| \geq 0{,}8\Delta_1$ se aumenta el radio de confianza al valor $\Delta_2 = 2\Delta_1 = 4$. En la siguiente figura se observan las curvas de nivel del nuevo modelo m_2 alrededor del nuevo iterado x_2 y las curvas de nivel del modelo junto con las curvas de nivel de la función objetivo.



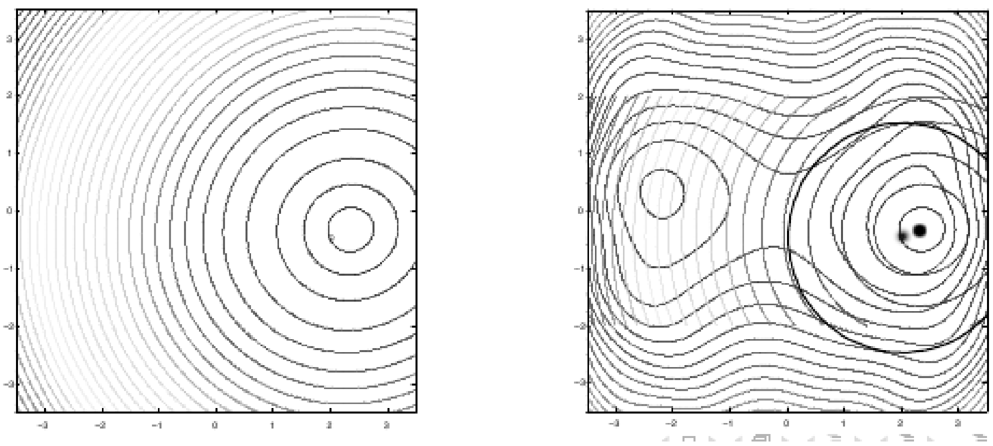


Graficamente se observa que el modelo ya no es una buena representación de la función cuando aumentamos el radio de la región de confianza. Como $\rho_2 = \frac{A_{red}}{P_{red}} = -0{,}004 \not> 0{,}1$ entonces el proximo iterado no cambia, $x_3 = x_2$ y se reduce el radio de la región de confianza a $\Delta_3 =$

$1/2\Delta_2 = 2$. En la siguiente figura se observa el nuevo modelo m_3 alrededor del nuevo iterado x_3



En difuso se observa el punto x_3, en negro una solución aproximada del modelo dentro de la región de confianza. El cociente $\rho_3 = \frac{A_{red}}{P_{red}} = 0{,}649 > 0{,}1$ entonces, el nuevo punto es aceptado como proximo iterado, $x_4 = x_3 + p_3$ y, como $\rho_3 \leq 0{,}75$ se mantiene sin cambiar el radio de confianza, $\Delta_4 = \Delta_3 = 2$. En la siguiente figura se observan las curvas de nivel del nuevo modelo m_4 alrededor del nuevo iterado x_4 y las curvas de nivel del modelo junto con las curvas de nivel de la función objetivo.



En difuso se observa el punto x_4, en negro una solución aproximada del modelo dentro de la región de confianza. El cociente $\rho_4 = \frac{A_{red}}{P_{red}} = 0{,}857 > 0{,}1$ entonces, el nuevo punto es aceptado como proximo iterado $x_5 = x_4 + p_4$ y como $\rho_4 > 0{,}75$ y $\|p_4\| \leq 0{,}8\Delta_4$ no se aumenta el radio de confianza, $\Delta_5 = \Delta_4 = 2$. Luego, en esta iteración el algoritmo encuentra aproximadamente uno de los minimizadores de la función objetivo $x^* = (2{,}303, -0{,}341)$. Observar que $\rho_5 \approx 1$.

Tabla con los valores obtenidos

k	Δ_k	p_k	$f(x_k + p_k)$	$\frac{A_{red}}{P_{red}}$	x_{k+1}
0	1	(0.05, 0.93)	43.742	0.998	$x_0 + p_0$
1	2	(-0.62, 1.78)	2.306	1.354	$x_1 + p_1$
2	4	(3.21, 0,00)	6.295	-0.004	x_2
3	2	(1.90, 0.08)	-29.293	0.649	$x_2 + p_2$
4	2	(0.32, 0.15)	-31.131	0.857	$x_3 + p_3$
5	2	(-0.03,-0.02)	-31.176	1.009	$x_4 + p_4$

El minimizador del modelo dentro de la región de confianza en la dirección de $-\nabla f(x_k)$ se denomina *punto de Cauchy*. Este punto es importante para garantizar convergencia de un método de regiones de confianza.

Denominemos $g_k = -\nabla f(x_k)$. Si tenemos que $g_k B_k g_k \leq 0$ entonces la cuadrática es indefinida en esa dirección y el minimizador se encuentra en el borde. Si $g_k^T B_k g_k > 0$ entonces hay que calcular la longitud de paso máxima que se puede dar en esa dirección sin salir de la región. Si el minimizador esta fuera de la región nuevamente el punto de Cauchy estará en el borde. Es decir, si $t_k = \dfrac{\|g_k\|^2}{g_k^T B_k g_k}$ entonces

$$p_k^C = \begin{cases} \dfrac{\Delta_k}{\|g_k\|}(-g_k) & \text{si } t_k \geq \dfrac{\Delta_k}{\|g_k\|} \\ t_k(-g_k) & \text{si } t_k < \dfrac{\Delta_k}{\|g_k\|} \text{ y } g_k^T B_k g_k > 0 \\ t_k(-g_k) & \text{si } g_k^T B_k g_k \leq 0 \end{cases}$$

Este punto tiene importancia crucial para decidir si una solución aproximada del subproblema es aceptable. Concretamente, un método de regiones de confianza será globalmente convergente si el paso p_k produce una reducción en el modelo m_k que sea menor que un múltiplo fijo de la reducción que se obtiene con en punto de Cauchy.

Se puede demostrar la siguiente proposición.

Proposición 1.2. *La reducción que se obtiene en el modelo con p_k^C es la siguiente:*

$$m_k(0) - m_k(p_k^C) \geq \frac{1}{2}\|g_k\| \operatorname{mín}\left\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right\}.$$

Luego, se puede demostrar el siguiente teorema, una demostración puede encontrarse en [4].

Teorema 1.15. *Sea $x_0 \in \mathbb{R}^n$. Suponemos que f es dos veces continuamente diferenciable y acotada inferiormente en $\{x : f(x) \leq f(x_0)\}$, que $\|B_k\| < \gamma$ y que p_k cumple*

$$m_k(0) - m_k(p_k) \geq C_1\|g_k\| \operatorname{mín}\left\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right\}$$

entonces $\lim_{k\to\infty} \|g_k\| = 0$.

2. Teoría básica de optimización con restricciones

En este capítulo vamos a analizar casos en los cuales el conjunto factible no es necesariamente todo $\mathbb{R}^n$. Este tipo de problemas se denominan *Problemas con restricciones.* Un problema general de optimización con restricciones será de la forma

$$\text{Minimizar } f(x) \text{ sujeto a } h(x) = 0, g(x) \leq 0 \tag{10}$$

donde las funciones $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$ son todas diferenciables hasta el orden que sea necesario. La función f se denomina función objetivo y el conjunto $\{x : h_i(x) = 0, g_j(x) \leq 0, i = 1, ..., m; j = 1, ..., p, \}$ se denomina conjunto factible.

En principio y para fijas ideas analizaremos las condiciones de optimalidad para problemas que presentan solo restricciones de igualdad, luego para problemas con restricciones de desigualdad y finalmente extenderemos los resultados para problemas del tipo (10). Un excelente tratamiento de este tema puede encontrarse en el libro [5].

2.1. Minimización con restricciones de igualdad

Consideramos problemas de la forma

$$\text{Minimizar } f(x) \text{ sujeto a } h(x) = 0 \tag{11}$$

donde las funciones $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m$ son todas diferenciables hasta el orden que sea necesario.

Observemos que si consideramos una restricción de igualdad en $\mathbb{R}^2$ entonces el conjunto factible $\Omega = \{(x, y) \in \mathbb{R}^2 : h(x, y) = 0\}$ es una curva.

Si consideramos solo una restricción de igualdad en $\mathbb{R}^3$ entonces el conjunto factible $\Omega = \{(x, y, z) \in \mathbb{R}^3 : h(x, y, z) = 0\}$ es una superficie. Si consideramos dos restricciones de igualdad en $\mathbb{R}^3$ entonces el conjunto factible $\Omega = \{(x, y, z) \in \mathbb{R}^3 : h_1(x, y, z) = 0, h_2(x, y, z) = 0\}$ es la intersección de dos superficies que generalmente es una curva en el espacio.

Consideremos por ejemplo el problema

$$\text{Minimizar } -x - y \text{ sujeto a } x^2 + y^2 = 1. \tag{12}$$

En la figura 6 se observa el conjunto factible (en línea continua) y las curvas de nivel de la función objetivo f (en línea punteada). Sabemos que f crece en la dirección del gradiente $\nabla f(x, y) = (-1 \ -1)^T$ y luego el minimizador de f se encuentra sobre la recta $y = x$ dentro de la región factible. Luego, la solución del problema es $x^* = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Observemos que en la solución

$$\nabla f(x^*) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \nabla h(x^*) = \begin{pmatrix} \frac{2}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{pmatrix},$$

es decir, se tiene que $-\nabla f(x^*) = \frac{2}{\sqrt{2}} \nabla h(x^*)$.

Ejercicio. Observar que un comportamiento semejante ocurre para los gradientes del problema

$$\text{Minimizar } \frac{1}{2}[(x - 2)^2 + (y - 2)^2] \text{ sujeto a } x^2 + y^2 = 1.$$

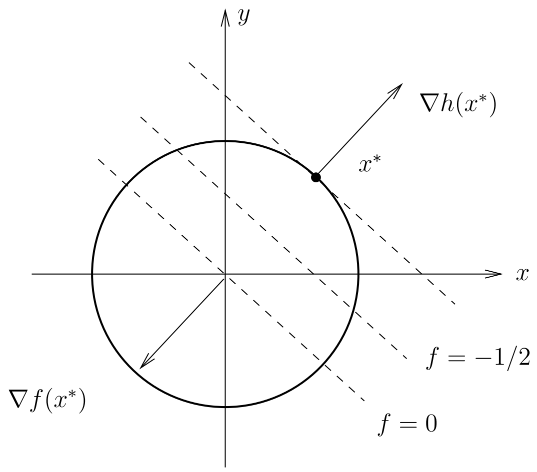


Figura 6: Curvas de nivel y conjunto factible del problema (12).

Consideremos ahora el siguiente problema en $\mathbb{R}^3$:

$$\text{Minimizar } x^2 + y^2 + z^2 \text{ sujeto a } x + y + 2z = 4, y = x. \tag{13}$$

Este problema es una formulación del problema de encontrar un punto sobre la intersección de los dos planos que se encuentra a menor distancia al origen. Definimos $h_1(x, y, z) = x + y + 2z - 4, h_2(x, y, z) = y - x$. La intersección de las dos superficies es una curva en el espacio que puede parametrizarse como el conjunto de los puntos de la forma $(t, t, 2-t)$. Observar la figura 7. De esta manera, se obtiene que $f(t, t, 2-t) = 3t^2 - 4t - 4$ y tiene su minimizador global en $t = \frac{2}{3}$ obteniendo de esta manera la solución $x^* = (\frac{2}{3}, \frac{2}{3}, \frac{4}{3})$. Verificar en este caso que se tiene $-\nabla f(x^*) = \lambda_1 \nabla h_1(x^*) + \lambda_2 \nabla h_2(x^*)$.

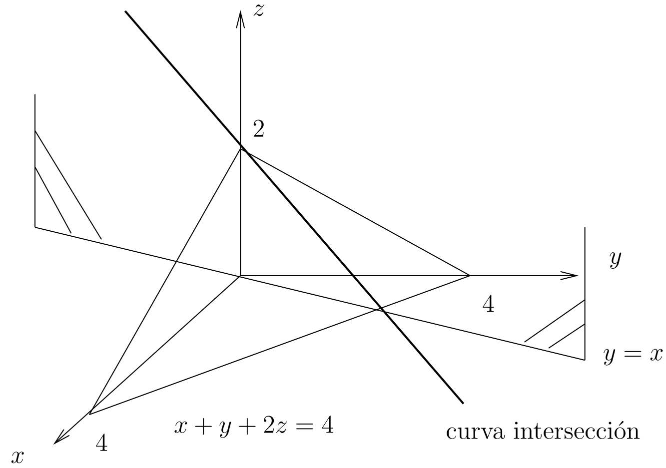


Figura 7: Conjunto factible del problema (13).

Lamentablemente esta situación no sucede siempre en un minimizador local. Considerar el problema de encontrar un punto sobre la curva $y^2 = (x - 1)^3$ que se encuentra a menor distancia al origen. Graficamente se observa que la solución de este problema es el punto $(1, 0)$, sin embargo, en este punto no se verifica que existe λ tal que $-\nabla f(x^*) = \lambda \nabla h(x^*)$. Observar la figura 8.

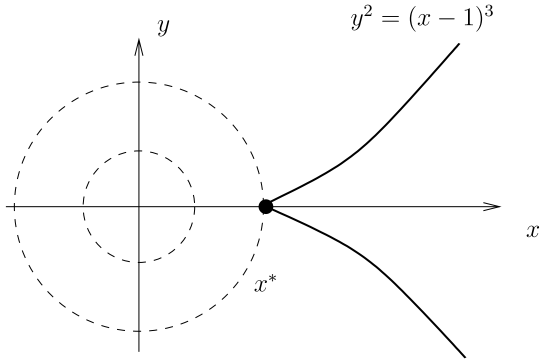


Figura 8: Curvas de nivel y conjunto factible del problema.

Para que se verifique una combinación entre el gradiente de la función y de las restricciones en un minimizador local es necesario que se cumpla alguna condición extra. Para esto, veamos que sucede con curvas contenidas en la región.

Dado un punto factible $\tilde{x}$, los caminos que pasan por el punto contenidos en región factible son curvas. Sabemos que las curvas $r(t)$ diferenciables contenidas en la región factible que pasan por $\tilde{x}$ tienen la propiedad de que $r^{'}(t)$ es un vector tangente al arco en el punto y es tangente a la superficie, es decir, se encuentra en el plano tangente a la superficie en $\tilde{x}$. Ver figura 9.

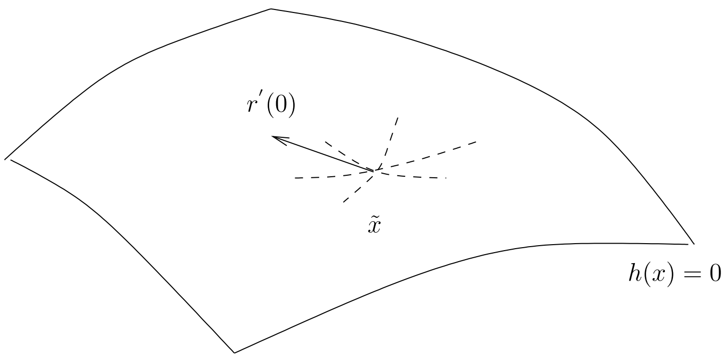


Figura 9: Vector tangente a una curva factible.

Si $\Omega = \{x \in \mathbb{R}^n : h(x) = 0\}$ y $r(t) : (a, b) \to \Omega$ es una parametrización de un arco factible entonces $\forall t \in (a, b)$

$$h(r(t)) = 0.$$

Derivando en función a t (mediante regla de la cadena) obtenemos que

$$Jh(r(t)).r^{'}(t) = 0 \qquad \forall t \in (a, b), \tag{14}$$

es decir

$$\nabla h_i(r(t))^T.r^{'}(t) = 0 \qquad \forall t \in (a, b).$$

Si $r(0) = \tilde{x}$ entonces $\nabla h_i(\tilde{x})^T.r^{'}(0) = 0$.

Es decir, dado un arco factible se tiene que el vector tangente al arco en $\tilde{x}$ es ortogonal a los gradientes de las restricciones de igualdad evaluados en $\tilde{x}$.

Ahora: que sucede con la función objetivo sobre este tipo de curvas si $\tilde{x}$ es un minimizador local?

Sabemos que

$$f(r(t)) = f(\tilde{x}) + \nabla f(\tilde{x})^T r^{'}(0)t + o(t)$$

luego, como $\tilde{x}$ es un minimizador se tiene que

$$\nabla f(\tilde{x})^T r^{'}(0)t + o(t) = f(r(t)) - f(\tilde{x}) \geq 0.$$

Dividiendo por $t > 0$ y haciendo t tender a cero obtenemos que $\nabla f(\tilde{x})^T r^{'}(0) \geq 0$. Dividiendo por $t < 0$ y haciendo t tender a cero obtenemos que $\nabla f(\tilde{x})^T r^{'}(0) \leq 0$, luego

$$\nabla f(\tilde{x})^T r^{'}(0) = 0.$$

Teorema 2.1. *Si x^* es un minimizador local del problema (11) entonces $\nabla f(\tilde{x})^T d = 0$, para todo d tal que $d = r^{'}(0)$ siendo r una curva factible tal que $r(0) = x^*$.*

Definición 2.2. *El conjunto $T_{x^*} = \{d \in \mathbb{R}^n :$ existe $r(t)$ diferenciable tal que $r(0) = x^*, r^{'}(0) = d\}$ se define como tangente al conjunto factible en x^*.*

Vimos que si $d \in T_{x^*}$ entonces $\nabla h_i(x^*)^T d = 0, \forall i = 1, \ldots, m$.

El teorema 2.1 da una condición necesaria de optimalidad de primer orden aunque poco práctica ya que no es fácil detectar todas las posibles curvas factible r. Es por esto que nos gustaria caracterizar cuando es posible asegurar que si x^* es un minimizador local entonces existe una combinación del gradiente de f en función de los gradientes de las restricciones.

Teorema 2.3. *Si x^* es un minimizador local de (11) tal que $\{\nabla h_1(x^*), \ldots, \nabla h_m(x^*)\}$ son linealmente independientes entonces existen únicos escalares $\lambda_1, \ldots, \lambda_m$ tales que*

$$-\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*). \tag{15}$$

Para demostrar este teorema necesitamos un resultado previo.

Teorema 2.4. *Si $\{\nabla h_1(x^*), \ldots, \nabla h_m(x^*)\}$ son linealmente independientes entonces el espacio $T_{x^*} = \{d \in \mathbb{R}^n : \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m\}$.*

Dem.

Vimos que $T_{x^*} \subset \{d \in \mathbb{R}^n : \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m\}$ vale siempre. Veamos la otra contención.

Sea d tal que $\nabla h_i(x^*)^T d = 0, i = 1, \ldots, m$. Consideramos el sistema de ecuaciones

$$F(t, u) = h(x^* + td + Jh(x^*)^T u) = 0$$

de m ecuaciones con $m + 1$ incógnitas. Tenemos que

1. $F \in \mathcal{C}^1$,

2. $F(0, 0) = h(x^*) = 0$,

3. $\dfrac{\partial(F_1,\ldots,F_m)}{\partial(u_1,\ldots,u_m)}(0,0) = Jh(x^*)Jh(x^*)^T$.

Por hipótesis sabemos que $\{\nabla h_1(x^*),\ldots,\nabla h_m(x^*)\}$ son linealmente independientes, luego, $\dfrac{\partial(F_1,\ldots,F_m)}{\partial(u_1,\ldots,u_m)}(0,0) = Jh(x^*)Jh(x^*)^T$ es no singular y podemos aplicar el Teorema de la Función Implícita. Luego, existen entornos de $t=0$ y de $u=0$ y una única función $u(t)$ diferenciable tal que $u(0)=0$ y $F(t,u(t)) = h(x^*+td+Jh(x^*)^Tu(t)) = 0$ en los entornos. Definimos la curva factible diferenciable

$$r(t) = x^* + td + Jh(x^*)^T u(t).$$

Tenemos que $r(0) = x^*$, $r'(t) = d + Jh(x^*)^T u'(t)$ y se tiene que $r'(0) = d + Jh(x^*)^T u'(0)$. Luego, $r'(0) = d$ (como queremos probar) si y solo si $Jh(x^*)^T u'(0) = 0$. Veamos entonces que efectivamente se tiene que $u'(0) = 0$.

Tenemos que $h(x^* + td + Jh(x^*)^T u(t)) = 0$. Derivando respecto de t y evaluando en $t=0$ obtenemos las igualdades

$$Jh(x^* + td + Jh(x^*)^T u(t))(d + Jh(x^*)^T u'(t)) = 0$$

$$Jh(x^*)(d + Jh(x^*)^T u'(0)) = 0$$

que equivale a

$$Jh(x^*)d + Jh(x^*)Jh(x^*)^T u'(0) = 0.$$

Luego, como $\nabla h_i(x^*)^T d = 0, i = 1,\ldots,m$ se tiene que

$$Jh(x^*)Jh(x^*)^T u'(0) = 0$$

pero, como $\{\nabla h_1(x^*),\ldots,\nabla h_m(x^*)\}$ son linealmente independientes la matriz $Jh(x^*)Jh(x^*)^T$ es no singular lo que implica que $u'(0) = 0$ como queríamos demostrar. □

A partir de esta propiedad podemos demostrar el teorema 2.3.

Demostración del teorema 2.3. Queremos probar que la igualdad (15), es decir

$$-\nabla f(x^*) = \begin{pmatrix} \frac{\partial h_1(x^*)}{\partial x_1} & \cdots & \frac{\partial h_m(x^*)}{\partial x_1} \\ \vdots & \vdots & \vdots \\ \frac{\partial h_1(x^*)}{\partial x_n} & \cdots & \frac{\partial h_m(x^*)}{\partial x_n} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} = (Jh(x^*))^T \lambda.$$

Como $\{\nabla h_1(x^*),\ldots,\nabla h_m(x^*)\}$ son linealmente independientes tenemos que $Jh(x^*)$ tiene rango m entonces la matriz $Jh(x^*)(Jh(x^*))^T \in \mathbb{R}^{m\times m}$ es no singular. Luego, como $\mathbb{R}^m = Nu(Jh(x^*)) \oplus R((Jh(x^*))^T)$, si suponemos que $-\nabla f(x^*) \notin R((Jh(x^*))^T)$ entonces $-\nabla f(x^*) = u + v$ con u, v únicos tales que $u \in Nu(Jh(x^*)), v \in R((Jh(x^*))^T)$ con $u \neq 0$. Luego, como $u \in Nu(Jh(x^*))$ entonces $\nabla h_i(x^*)^T u = 0, \forall i = 1,\ldots,m$ y, por el teorema anterior $u \in T_{x^*}$.

Pero $u \neq 0$ entonces

$$-\nabla f(x^*)^T u = u^T u + u^T v = \|u\|^2 > 0$$

lo que es absurdo por el teorema 2.1. □

Ejercicio. Demostrar que los multiplicadores de Lagrange son únicos.

Definición 2.5. *Sea x^* un punto factible.*

1. *Decimos que x^* es un punto regular cuando los gradientes $\{\nabla h_1(x^*), \ldots, \nabla h_m(x^*)\}$ son linealmente independientes. En este caso decimos que x^* verifica la condición de regularidad.*

2. *Definimos $T^{lin}(x^*) = \{d \in \mathbb{R}^n : \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m\}$ como cono tangente linealizado en x^*.*

3. *La condición (15) se denomina condición de Lagrange y los escalares $\lambda_1, \ldots, \lambda_m$ se denominan multiplicadores de Lagrange.*

4. *Una condición de calidad es una condición sobre los puntos factibles que, cuando es verificada por un minimizador local implica la condición de Lagrange.*

Observación. Regularidad es una condición de calidad de primer orden.

Que sucede en problemas donde la solución no es regular?

Ejemplos.
Consideremos el problema

$$\text{Minimizar } x^2 + y^2 \ \text{ sujeto a } y - 1 = 0, y - x^2 + 1 = 0.$$

Este problema tiene un único punto factible $x^* = (0, 1)$ que es la solución del problema. Si definimos $h_1(x, y) = y - 1, h_2(x, y) = y - x^2 + 1$, se tiene que

$$\nabla f(x^*) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad \nabla h_1(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \nabla h_2(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

luego, los gradientes de las restricciones son linealmente dependientes, es decir x^* no es regular, pero existen infinitos λ_1, λ_2 que verifican la condición de Lagrange.

Consideremos el problema

$$\text{Minimizar } x^2 + (y + 1)^2 \ \text{ sujeto a } x + y = 0, (x - 1)^2 + (y - 1)^2 = 2.$$

Nuevamente este problema tiene un único punto factible $x^* = (0, 0)$ que es la solución del problema. Si definimos $h_1(x, y) = x + y, h_2(x, y) = (x - 1)^2 + (y - 1)^2 - 2$, se tiene que

$$\nabla f(x^*) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad \nabla h_1(x^*) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \nabla h_2(x^*) = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$$

luego, los gradientes de las restricciones son linealmente dependientes, es decir x^* no es regular, pero no existen escalares λ_1, λ_2 que verifican la condición de Lagrange ($\lambda_1 + \lambda_2 = -2$).

Teorema 2.6. *Si x^* es un minimizador local del problema (11) y cumple $T(x^*) = T^{lin}(x^*)$ entonces existen escalares $\lambda_1, \ldots, \lambda_m$ tales que se verifica la condición (15).*
Dem.

Tenemos que $T(x^*) = T^{lin}(x^*) = Nu(Jh(x^*))$. Luego, por el teorema 2.1 sabemos que $-\nabla f(x^*) \in (T_{x^*})^\perp = (Nu(Jh(x^*)))^\perp = R(Jh(x^*)^T)$ lo que demuestra el teorema. $\square$

Definición 2.7. *La condición $T(x^*) = T^{lin}(x^*)$ se denomina condición de Kuhn-Tucker y es una condición de calidad, como muestra la propiedad anterior.*

Se define la función $l(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$ y se denomina función de Lagrange. Por lo que vimos anteriormente tenemos que, si x^* es un minimizador local de (11) y cumple alguna condición de calidad (regularidad por ejemplo) entonces existe $\lambda^* \in \mathbb{R}^m$ tal que $\nabla l(x^*, \lambda^*) = 0$. Es decir, el par (x^*, λ^*) es un punto estacionario de la función de Lagrange.

Tanto regularidad como Kuhn-Tucker son condiciones de calidad, es decir, condiciones que implican la existencia de multiplicadores de Lagrange en una solución. Existen en la literatura otras condiciones de calidad entre estas dos.

Como en el caso de minimización irrestricta, existen condiciones necesarias y suficientes de segundo orden.

Teorema 2.8. Condición necesaria de segundo orden. *Sea x^* un minimizador local regular del problema (11) entonces existe $\lambda^* \in \mathbb{R}^m$ tal que*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0$$

y

$$d^T\left(\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*)\right)d \geq 0, \quad \forall d \in T^{lin}(x^*). \tag{16}$$

Dem.

Por hipótesis y por el teorema 2.3 sabemos que existe $\lambda^* \in \mathbb{R}^m$ verificando (15). Además, como $T^{lin}(x^*) = T(x^*)$, si $d \in T^{lin}(x^*)$ entonces exite $r(t)$ una curva factible y diferenciable tal que $r(0) = 0, r^{'}(0) = d$. Luego, por el desarrollo de Taylor para $l(r(t), \lambda^*)$ alrededor de $t = 0$.

$$l(r(t), \lambda^*) = l(x^*, \lambda^*) + \nabla l(x^*, \lambda^*) r^{'}(0) t + \frac{1}{2} t^2 r^{'}(0) \nabla^2_{xx} l(x^*, \lambda^*) r^{'}(0) + o(t^2).$$

Asi, como $\nabla l(x^*, \lambda^*) = 0$ y por definición de l se tiene

$$f(r(t)) = f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \frac{1}{2} t^2 r^{'}(0) \nabla^2_{xx} l(x^*, \lambda^*) r^{'}(0) + o(t^2).$$

Como x^* es minimizador local y $r(t)$ es factible se tiene que

$$f(r(t)) \geq f(x^*) \quad \forall t \in (-\varepsilon, \varepsilon).$$

Entonces

$$\frac{1}{2} t^2 r^{'}(0) \nabla^2_{xx} l(x^*, \lambda^*) r^{'}(0) + o(t^2) \geq 0 \quad \forall t \in (-\varepsilon, \varepsilon).$$

Dividiendo por $t^2 \neq 0$ y tomando límite cuando t tiende a cero obtenemos (16). $\square$

Teorema 2.9. Condición suficiente de segundo orden *Sea x^* tal que $h(x^*) = 0$. Sea $\lambda^* \in \mathbb{R}^m$ tales que*

$$\nabla l(x^*, \lambda^*) = 0$$

y

$$v^T \nabla^2 l(x^*, \lambda^*) v > 0, \quad \forall v \in T^{lin}(x^*) - \{0\}$$

entonces x^* es un minimizador local del problema (11).

Dem.

Suponemos que x^* no es minimizador local. Entonces, existe y_k tal que $h(y_k) = 0, f(y_k) \leq f(x^*)$ y $\|y_k - x^*\| \leq \frac{1}{k}$.

Sea $\delta_k = \|y_k - x^*\|$. Definimos $s_k = \dfrac{y_k - x^*}{\delta_k}$ para todo k tal que $\delta_k \neq 0$. Luego, s_k es una sucesión acotada y tiene una subsucesión convergente. Sean $K_1 \subset \mathbb{N}, s^*$ tales que $s_k \to_{k \in K_1} s^*$.

Tenemos que

$$h_i(y_k) - h_i(x^*) = \delta_k \nabla h_i(x^*)^T s_k + o(\delta_k \|s_k\|).$$

Dividiendo por δ_k para todo k tal que $\delta_k \neq 0$ obtenemos que $\nabla h_i(x^*)^T s^* = 0$. Luego, $s^* \in T^{lin}(x^*)$.

Usando el desarrollo de Taylor de segundo orden para f y para cada h_i alrededor de x^* y evaluando en y_k obtenemos que

$$f(y_k) - f(x^*) = \delta_k \nabla f(x^*)^T s_k + \frac{1}{2} \delta_k^2 s_k^T \nabla^2 f(x^*) s_k + o(\delta_k^2)$$

$$h_i(y_k) - h_i(x^*) = \delta_k \nabla h_i(x^*)^T s_k + \frac{1}{2} \delta_k^2 s_k^T \nabla^2 h_i(x^*) s_k + o(\delta_k^2).$$

Luego, como $h_i(y_k) = 0$, haciendo $f(y_k) + \sum_{i=1}^{m} \lambda_i^* h_i(y_k)$ obtenemos que

$$f(y_k) - f(x^*) = f(y_k) - f(x^*) + \sum_{i=1}^{m} \lambda_i^* h_i(y_k) =$$

$$\delta_k \nabla_x l(x^*, \lambda^*)^T s_k + \frac{1}{2} \delta_k^2 s_k^T \nabla_{xx}^2 l(x^*, \lambda^*) s_k + o(\delta_k^2).$$

Como $\nabla_x l(x^*, \lambda^*) = 0$ y $f(y_k) \leq f(x^*)$ tenemos que

$$\frac{1}{2} \delta_k^2 s_k^T \nabla^2 l(x^*, \lambda^*) s_k + o(\delta_k^2) \leq 0.$$

Dividiendo por δ_k^2 y haciendo $k \in K_1$ tender a infinito obtenemos que

$$(s^*)^T \nabla^2 l(x^*, \lambda^*) s^* \leq 0$$

que es un absurdo. □

Ejercicios.

1. Considere el problema

$$\begin{array}{ll} \text{Min} & f(x) \\ \text{s.a} & h(x) = 0. \end{array}$$

Sea x^* un minimizador local regular del problema tal que $\nabla f(x^*) = 0$. Calcule los multiplicadores de Lagrange. Que tipo de solución es x^*?

2. Sean $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m, f, h \in \mathcal{C}^2$. Sea x^* tal que $h(x^*) = 0, \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) = 0$ y $\nabla^2 f(x^*) > 0$. Esto implica que x^* es un minimizador local de f sujeto a $h(x) = 0$? Demuestrelo o de un contraejemplo.

3. Considere el problema

$$\begin{array}{ll} \text{Min} & x_1^2 + 4x_1x_2 + x_2^2 \\ \text{s.a} & x_1^2 + x_2^2 = 1 \end{array}$$

   a) Usando las condiciones KKT, encontrar una solución del problema. Interpretar las condiciones KKT geometricamente en x^*. Como son las curvas de nivel de f?

   b) Analizar si se verifican las condiciones de segundo orden.

   c) Tiene el problema una única solución?

2.2. Minimización con restricciones de desigualdad

Consideramos problemas de la forma

$$\text{Minimizar } f(x) \text{ sujeto a } g(x) \leq 0 \tag{17}$$

donde las funciones $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \to \mathbb{R}^p$ son todas diferenciables hasta el orden que sea necesario.

Consideremos por ejemplo el problema

$$\text{Minimizar } -x-y \text{ sujeto a } x^2 + y^2 \leq 1.$$

Podemos observar en la figura que la solución del problema es, como encontramos para el problema con restricciones de igualdad, el punto $x^* = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

Si consideramos el problema

$$\text{Minimizar } -x-y \text{ sujeto a } x^2 + y^2 \leq 1, 2y \geq x$$

vemos en la figura que la solución sigue siendo la misma. Es decir, la restricción $2y \geq x$ no interfiere en el cálculo del mínimo.

Definición 2.10. *Dado un punto factible x, definimos el conjunto de restricciones de desigualdad que son activas en x:*

$$A(x) = \{i \in \{1, \ldots, p\} : g_i(x) = 0\}$$

Si $i \notin A(x)$ entonces decimos que esa restricción es inactiva en x. Es el caso de la restricción $2y \geq x$ en x^* en el ejemplo anterior.

Si $g_i(x) > 0$ decimos que esa restricción es violada en x.

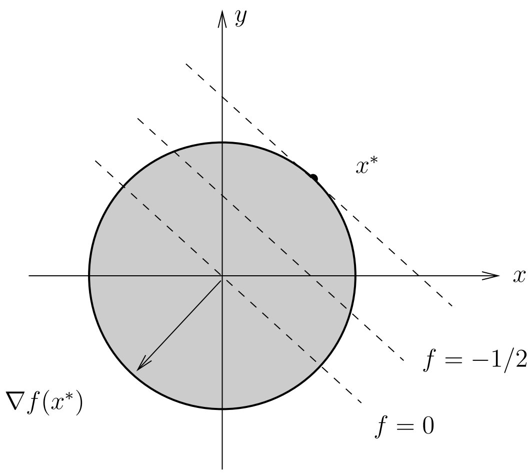


Figura 10: Solución del problema con restricciones de desigualdad.

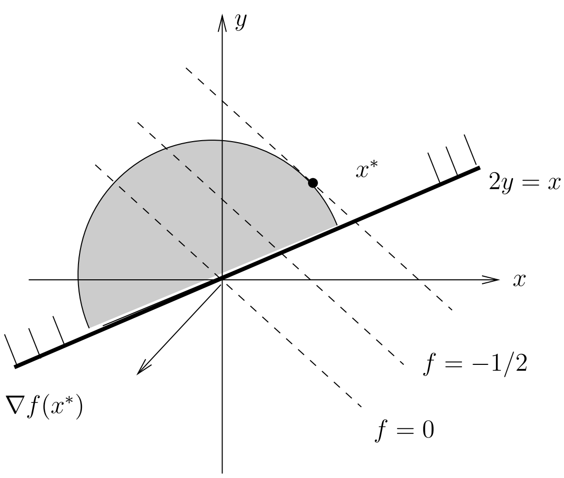


Figura 11: Solución del problema con restricciones de desigualdad.

Proposición 2.1. *Si x^* es un minimizador del problema (17) entonces x^* es un minimizador del problema con restricciones de igualdad*

$$\textit{Minimizar } f(x) \;\; \textit{sujeto a } g_i(x) = 0, i \in A(x^*). \tag{18}$$

Luego, tenemos que, si los gradientes $\{\nabla g_i(x^*)\}_{i \in A(x^*)}$ son linealmente independientes entonces existen escalares λ_i tales que

$$-\nabla f(x^*) = \sum_{i \in A(x^*)} \lambda_i \nabla g_i(x^*). \tag{19}$$

Veremos que, como las restricciones son de desigualdad, nos importa el signo de los multiplicadores.

Consideremos el problema

$$\text{Minimizar } (x-1)^2 + (y-1)^2 \;\; \text{sujeto a } x^2 + y^2 \leq 1, 2y \geq x^2.$$

Observar el punto $\tilde{x}$ en la figura 12.

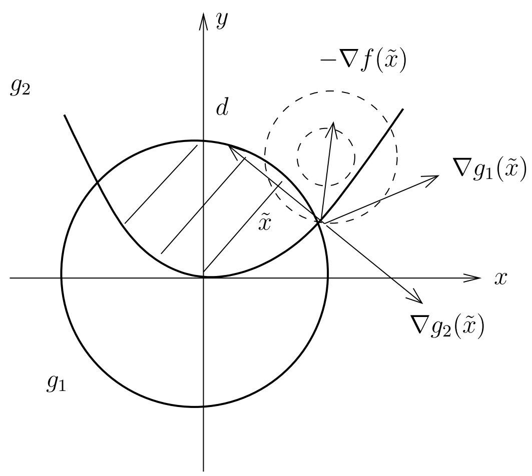


Figura 12: Importancia del signo de los multiplicadores.

Vemos que, en ese punto se tiene que

$$-\nabla f(\tilde{x}) = \mu_1 \nabla g_1(\tilde{x}) + \mu_2 \nabla g_2(\tilde{x}) \qquad \text{con } \mu_1 > 0, \mu_2 < 0.$$

Sea d una dirección tal que $\nabla g_1(\tilde{x})^T d = 0, \nabla g_2(\tilde{x})^T d < 0$. Luego,

$$-\nabla f(\tilde{x})^T d = \mu_1 \nabla g_1(\tilde{x})^T d + \mu_2 \nabla g_2(\tilde{x})^T d > 0$$

es decir, si abandonamos g_2 moviendonos en forma ortogonal al $\nabla g_1(\tilde{x})$ podemos hacer decrecer el valor de f. Esto muestra que no alcanza con ver la verificación de (19) sino que además es importante analizar el signo de los multiplicadores. Un multiplicador negativo dice que el punto no es solución.

Teorema 2.11. *Si x^* es un minimizador local de (17) tal que $\{\nabla g_i(x^*)\}_{i \in A(x^*)}$ son linealmente independientes entonces existen únicos escalares $\{\mu_i\}_{i \in A(x^*)}$ tales que $\mu_i \geq 0$*

$$-\nabla f(x^*) = \sum_{i \in A(x^*)} \mu_i \nabla g_i(x^*). \tag{20}$$

Para demostrar este teorema necesitamos un resultado previo.

Analicemos que sucede con curvas factibles. Sea $r(t)$ una curva factible diferenciable tal que $r(0) = x^*$. Si $i \in A(x^*)$ entonces

$$g_i(r(t)) = g_i(x^*) + \nabla g_i(x^*)^T r'(0)t + o(t)$$

entonces, como $g_i(r(t)) \leq 0$

$$\nabla g_i(x^*)^T r'(0)t + o(t) \leq 0$$

dividiendo por $t > 0$ y haciendo t tender a cero obtenemos que

$$\nabla g_i(x^*)^T r'(0) \leq 0.$$

Luego, en el caso de restricciones de desigualdad tenemos que

$$T_{x^*} \subset \{d \in \mathbb{R}^n : \nabla g_i(x^*)^T d \leq 0, i \in A(x^*)\}.$$

Definimos

$$T^{lin}(x^*) = \{d \in \mathbb{R}^n : \nabla g_i(x^*)^T d \leq 0, i \in A(x^*)\}$$

como tangente linealizado en x^* para problemas con restricciones de desigualdad.

Como sucede en el caso de restricciones de igualdad, se puede demostrar que si $\{\nabla g_i(x^*)\}_{i \in A(x^*)}$ son linealmente independientes entonces $T^{lin}(x^*) \subset T_{x^*}$.

Demostración del teorema 2.11.

Sabemos que x^* es un minimizador local del problema (18), luego, por hipótesis existen μ_i tales que

$$-\nabla f(x^*) = \sum_{i \in A(x^*)} \mu_i \nabla g_i(x^*).$$

Supongamos que algún $\mu_j < 0$. Por la independencia lineal existe $d \neq 0$ tal que $\nabla g_i(x^*)^T d = 0, \forall i \neq j, \nabla g_j(x^*)^T d < 0$. Entonces

$$-\nabla f(x^*)^T d = \mu_j \nabla g_j(x^*)^T d > 0.$$

Como $d \in T^{lin}(x^*) = T_{x^*}$ tenemos que existe $r(t)$ factible diferenciable tal que $r(0) = x^*, r'(0) = d$. Luego, usando Taylor,

$$\lim_{t \to 0} \frac{f(r(t)) - f(x^*)}{t} = \nabla f(x^*)^T d < 0$$

lo que implica que existe $t > 0$ tal que $f(r(t)) < f(x^*)$ absurdo. Luego, debe ser $\mu_i \geq 0, \forall i \in A(x^*)$. $\square$

Ejercicios.

1. Demostrar que si x^* es un minimizador local de f sujeto a $g_i(x) \leq 0, i = 1, \ldots, p$ y $A(x^*)$ es el conjunto de restricciones activas entonces x^* es un minimizador local del siguiente problema con restricciones de igualdad:

$$\begin{array}{ll} \text{Min} & f(x) \\ \text{s.a} & g_i(x) = 0, i \in A(x^*). \end{array}$$

2. Como en el caso de restricciones de igualdad, demostrar que si los gradientes $\nabla g_i(x^*), i \in A(x^*)$ son linealmente independientes entonces $T(x^*) = T^{lin}(x^*)$.

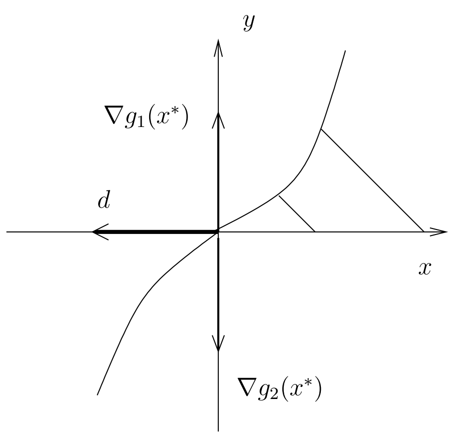
Definición 2.12. *1. La condición de que los gradientes $\{\nabla g_i(x^*)\}_{i \in A(x^*)}$ sean linealmente independientes se denomina regularidad. Es una condición de calidad para el problema con restricciones de desigualdad.*

2. Las condiciones

$$\begin{cases} -\nabla f(x^*) = \displaystyle\sum_{i=1}^{p} \mu_i \nabla g_i(x^*) \\ \mu_i \geq 0, g_i(x^*) \leq 0, \mu_i g_i(x^*) = 0 \end{cases}$$

se denominan condiciones de Karush-Kuhn-Tucker (KKT). La segunda condición se llama condición de complementariedad. Los escalares μ_i se denominan multiplicadores de Lagrange.

La inclusión $T^{lin}(x^*) \subset T_{x^*}$ no vale en general: consideremos las restricciones de desigualdad $g_1(x, y) = y - x^3, g_2(x, y) = -y$ en el punto factible $x^* = (0, 0)$.



Los gradientes son linealmente dependientes en x^* y se tiene que $d = (-1, 0) \in T^{lin}(x^*)$ pero $d \notin T_{x^*}$.

Ejercicio. Considere el problema

$$\begin{array}{ll} \text{Min} & x_1 \\ \text{s.a} & (x_1-1)^2+(x_2-1)^2 \le 1 \\ & (x_1-1)^2+(x_2+1)^2 \le 1 \end{array}$$

Se cumplen las condiciones KKT en la solución? Observe que las condiciones KKT no son condiciones necesarias de optimalidad para problemas convexos.

Analizemos las condiciones de segundo orden para problemas con restricciones de desigualdad.

Consideremos el problema

$$\text{Minimizar } x - x^2 \ \text{ sujeto a } x \ge 0.$$

La función objetivo es una parábola con ramas hacia abajo que tiene su máximo en $x = 1/2$. Tenemos que $x = 0$ es un minimizador local del problema. Además, $\mu = 1$ es un multiplicador de Lagrange. Consideremos la función de Lagrange $l(x,\mu) = x - x^2 + \mu(-x)$, luego, para todo $d \in T^{lin}(0) = \{d : -d \le 0\}$ tenemos que $d^T\nabla^2 l(x,\mu)d < 0$. Luego, $x = 0$ es un minimizador local pero $d^T\nabla^2 l(x,\mu)d \ngeq 0, \forall d \in T^{lin}(0)$.

Vemos entonces que $T^{lin}(x^*)$ no es el espacio adecuado para analizar condiciones de segundo orden para problemas con restricciones de desigualdad.

Consideremos el problema reducido (18). Si x^* es un minimizador local de (17) entonces, usando el teorema sobre condiciones necesarias de segundo orden para problemas con restricciones de igualdad de la sección anterior tenemos el siguiente teorema sobre condiciones necesarias de segundo orden para el problema (17).

Teorema 2.13. *Sea x^* un minimizador local de (17) tal que $\{\nabla g_i(x^*)\}_{i\in A(x^*)}$ son linealmente independientes. Entonces existe $\mu \in \mathbb{R}^p$ tal que se verifican las condiciones KKT y*

$$d^T\nabla^2 l(x^*,\mu)d \ge 0 \qquad \forall d \in \tilde{T}^{lin}(x^*) = \{d \in \mathbb{R}^n : \nabla g_i(x^*)^T d = 0, \forall i \in A(x^*)\}.$$

Lamentablemente $\tilde{T}^{lin}(x^*)$ no sirve para analizar las condiciones suficientes de segundo orden:

Consideremos lo problema

$$\text{Minimizar } 1/2(x^2 - y^2) \ \text{ sujeto a } y \le 0.$$

Observar en la figura las curvas de nivel de la función objetivo.

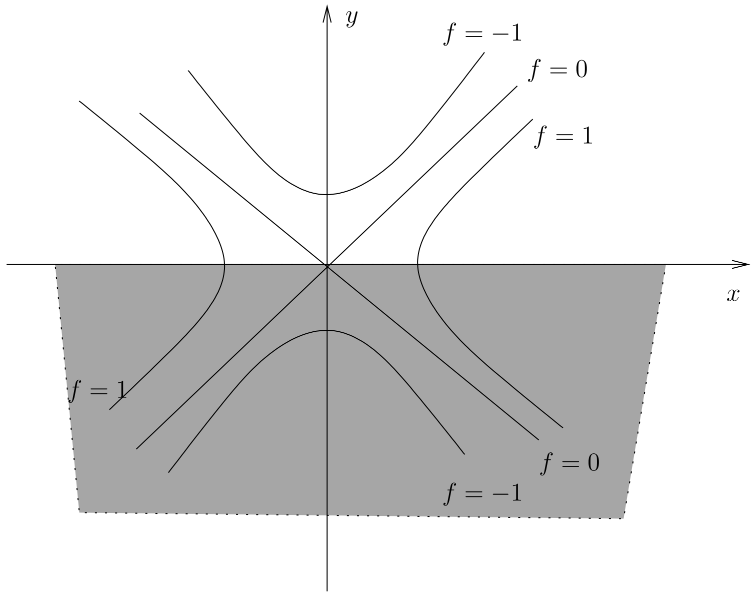
Vemos que $x = (0,0)$ no es solución, sin embargo, con el escalar $\mu = 0$ se tienen las condiciones KKT. Se tiene además que

$$\nabla^2 l(x,\mu) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \tilde{T}^{lin}(x) = \{d = (d_1, d_1) : d_2 = 0\}$$

lo que implica que $d^T\nabla^2 l(x,\mu)d \ge 0, \forall d \in \tilde{T}^{lin}(x)$ siendo que x no es minimizador local.

Dado x^* y un multiplicador de Lagrange asociado μ^*, definimos el espacio tangente

$$T_2(x^*) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} \nabla g_i(x^*)^T d = 0, & i \in A(x^*) : \mu_i > 0 \\ \nabla g_i(x^*)^T d \le 0, & i \in A(x^*) : \mu_i = 0. \end{array} \right\}$$



Teorema 2.14. Condición suficiente de segundo orden. *Sea x^* un punto factible tal que existe $\mu \in \mathbb{R}^p$ tal que se verifican las condiciones KKT y*

$$d^T \nabla^2 l(x^*, \mu) d > 0 \qquad \forall d \in T_2(x^*)$$

entonces x^* es un minimizador local del problema (17).
Dem.

Seguir la idea de la demostración que se hizo para restricciones de igualdad. □

Utilizando este tangente se puede obtener un nuevo teorema sobre condición necesaria de segundo orden.

Teorema 2.15. Condición necesaria de segundo orden. *Sea x^* un minimizador local de (17) regular. Sea $\mu \in \mathbb{R}^p$ tal que se verifican las condiciones KKT, entonces*

$$d^T \nabla^2 l(x^*, \mu) d \geq 0 \qquad \forall d \in T_2(x^*).$$

Para una demostración ver [5].

Ejercicios.

1. Considere el problema

$$\begin{array}{ll} \text{Min} & (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2 \\ \text{s.a} & x_1^2 - x_2 \leq 0 \\ & x_1 + x_2 \leq 6 \end{array}$$

   a) Escribir las condiciones KKT y verificar que estas condiciones se cumplen en el punto $x^* = (\frac{3}{2}, \frac{9}{4})$.

   b) Interpretar las condiciones KKT geometricamente en x^*.

   c) Analizar las condiciones suficientes de segundo orden.

d) Mostrar geometricamente que x^* es el único minimizador global.

2. Considere el problema

$$\begin{array}{ll} \text{Min} & x_1 x_2 \\ \text{s.a} & x_1 \leq x_2 \end{array}$$

a) Mostrar que se cumplen las condiciones KKT en el punto $x^* = (0, 0)$.

b) Analizar si se verifican las condiciones necesarias de segundo orden en el tangente $T(x^*) = \{d : \nabla g_i(x^*)^T d = 0, \forall i \in A(x^*)\} - \{0\}$.

c) Analizar si se verifican las condiciones suficientes de segundo orden.

d) Es x^* es un minimizador local?

3. Mostrar que cualquier punto factible del conjunto definido por $\{x \in \mathbb{R}^n : x \geq 0\}$ es regular.

2.3. Condiciones necesarias y suficientes de optimalidad para problemas con restricciones de igualdad y desigualdad

En esta sección consideramos el problema general de la forma

$$\text{Minimizar } f(x) \text{ sujeto a } h(x) = 0, g(x) \leq 0 \tag{21}$$

donde las funciones $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$ son todas diferenciables hasta el orden que sea necesario.

Generalizando los resultados obtenidos en las secciones anteriores obtenemos los siguientes resultados y definiciones.

Definición 2.16. *Dado un punto factible x^*,*

1. *Decimos que x^* es regular si los gradientes $\{\nabla h_1(x^*), \ldots, \nabla h_m(x^*)\} \cup \{\nabla g_i(x^*)_{i \in A(x^*)}\}$ son linealmente independientes.*

2. *Definimos el tangente linealizado en x^* al conjunto $T^{lin}(x^*) = \{d \in \mathbb{R}^n : \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m; \nabla g_i(x^*)^T d \leq 0, i \in A(x^*)\}$.*

3. *Decimos que x^* cumple la condición de Kuhn-Tucker si $T_{x^*} = T^{lin}(x^*)$.*

4. *La función $l(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{p} \mu_i g_i(x)$ se denomina función de Lagrange.*

Proposición 2.2. *Si x^* es regular entonces x^* cumple la condición de Kuhn-Tucker.*

Teorema 2.17. Condiciones KKT. *Si x^* es un minimizador local regular de (21) entonces existen únicos escalares $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_p$ tales que*

$$\left\{ \begin{array}{c} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \sum_{i=1}^{p} \mu_i \nabla g_i(x^*) = 0 \\ \mu_i \geq 0, g_i(x^*) \leq 0, \mu_i g_i(x^*) = 0 \end{array} \right. \tag{22}$$

El sistema (22) se denomina sistema KKT. Una solución del sistema es un punto *KKT* o *punto estacionario.*

Teorema 2.18. Condiciones necesarias de segundo orden *Sea x^* un minimizador local regular de (21). Entonces existen escalares $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_p$ tal que se verifican las condiciones KKT y*

$$d^T \nabla^2 l(x^*, \lambda, \mu) d \geq 0 \qquad \forall d : \begin{array}{c} \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m; \\ \nabla g_i(x^*)^T d = 0, \forall i \in A(x^*). \end{array}$$

Teorema 2.19. Condición suficiente de segundo orden. *Sea x^* un punto factible tal que existen $\lambda \in \mathbb{R}, \mu \in \mathbb{R}^p$ tal que se verifican las condiciones KKT y, $\forall d$ tal que*

$$\begin{array}{ll} \nabla h_i(x^*)^T d = 0, & i = 1, \ldots, m \\ \nabla g_i(x^*)^T d = 0, & i \in A(x^*) : \mu_i > 0 \\ \nabla g_i(x^*)^T d < 0, & i \in A(x^*) : \mu_i = 0. \end{array}$$

se cumple que

$$d^T \nabla^2 l(x^*, \lambda, \mu) d > 0$$

entonces x^* es un minimizador local del problema (17).

Ejercicios.

1. Considere el problema

$$\begin{array}{cc} \text{Min} & x_1 + x_2 \\ \text{s.a} & x_1 x_2 = 0 \\ & x_1 \geq 0, x_2 \geq 0 \end{array}$$

   Mostrar que $x^* = (0, 0)$ es KKT pero no cumple la condición de Kuhn-Tucker ($T(x^*) = T^{lin}(x^*)$). Esto implica que Kuhn-Tucker no es la condición de calidad más débil que existe en la literatura.

2.4. Problemas convexos

Definición 2.20. *1. Decimos que un conjunto Ω es convexo si $\alpha x + (1 - \alpha) y \in \Omega$ para todo $x, y \in \Omega, \alpha \in [0, 1]$.*

2. Un función f es convexa en Ω si

$$f(\alpha x + (1 - \alpha) y) \leq \alpha f(x) + (1 - \alpha) f(y)$$

para todo $x, y \in \Omega, \alpha \in [0, 1]$.

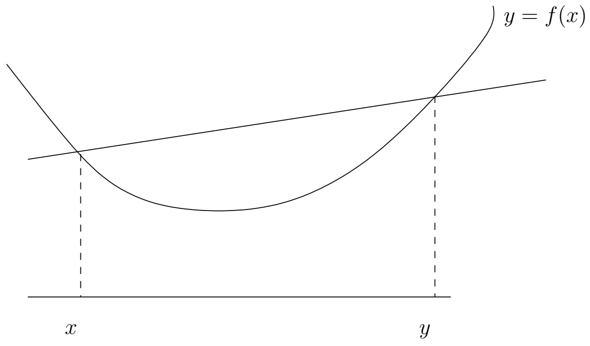
Quiere decir que el segmento de línea que conexta $(x, f(x))$ con $(y, f(y))$ está por arriba del gráfico de la función. Observar la siguiente figura.

Un problema de optimización convexa es de la forma

$$\begin{array}{c} \text{Minimizar } f(x) \\ \text{sujeto a } x \in \Omega, \end{array} \tag{23}$$

donde tanto f como Ω son convexos.

El siguiente teorema resulta ser muy importante en programación lineal que es un caso particular de optimización con conjunto convexo.



Teorema 2.21. *Sea x^* un minimizador local de (23). Entonces x^* es un minimizador global.*
Dem.

Suponemos que x^* es un minimizador local pero no global. Entonces existe $y \in \Omega$ tal que $f(y) \leq f(x^*)$. Sea $\alpha \in (0,1)$,

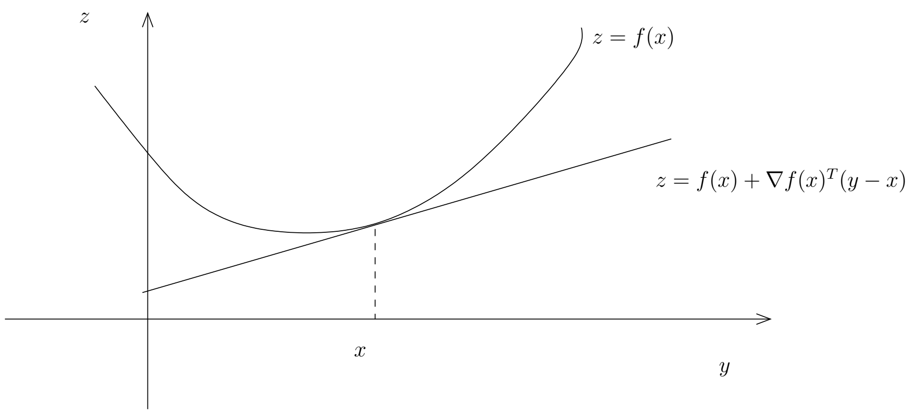
$$f(\alpha x^* + (1-\alpha)y) \leq \alpha f(x^*) + (1-\alpha)f(y) < \alpha f(x^*) + (1-\alpha)f(x^*) = f(x^*).$$

Entonces, hay punto muy cercano a x^* (α cercano a 1) en Ω en los cuales f vale menos. Esto contradice el hecho de que x^* es minimizador local. $\square$

Algunas propiedades y equivalencias importantes se obtienen cuando f convexa es derivable.

Proposición 2.3. *Si $f \in \mathcal{C}^1(\Omega)$, Ω convexo. Entonces*
f es convexa en Ω si y solo si $f(y) \geq f(x) + \nabla f(x)^T(y-x), \forall x, y \in \Omega$.

Ejercicio. Demostrar la proposición anterior. Observar la siguiente figura.



Para funciones escalares que tienen derivadas hasta orden 2, hay una caracterización equivalente a la de convexidad que suele ser muy útil:

- $f : \mathbb{R} \to \mathbb{R}, f \in \mathcal{C}^2$ es convexa si y solo si $f''(x) \geq 0, \forall x$.

Proposición 2.4. *Sea $f : \mathbb{R}^n \to \mathbb{R}, f \in \mathcal{C}^2$, f es convexa si y solo si $\nabla^2 f(x) \geq 0, \forall x$.*
Dem.

Supongamos que $\nabla^2 f(x) \geq 0, \forall x$. Sean x, y, tenemos que existe a entre x e y tal que

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(a)(y - x)$$

entonces

$$f(y) - (f(x) + \nabla f(x)^T (y - x)) = \frac{1}{2}(y - x)^T \nabla^2 f(a)(y - x) \geq 0$$

luego, f es convexa usando la proposición anterior. □

Ejercicio. Demostrar la recíproca del teorema anterior.

3. Métodos numéricos para problemas generales de optimización

En esta sección presentamos algunos de los métodos más conocidos para resolver problemas generales del optimización.

3.1. Método del gradiente proyectado

Este es un método para resolver problemas con restricciones de caja o de cotas en las variables. Este tipo de restricciones son más faciles que restricciones no lineales generales por eso se tratan separadamente.

Un problema general con este tipo de restricciones se modeliza de la forma

$$\begin{array}{c} \text{Minimizar } f(x) \\ \text{sujeto a } l \leq x \leq u, \end{array} \tag{24}$$

para $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciable, $l, u \in \mathbb{R}^n, l \leq u$ pueden tener alguna coordenada ∞.

Para fijas ideas consideraremos el problema

$$\begin{array}{c} \text{Minimizar } f(x) \\ \text{sujeto a } x \geq 0. \end{array} \tag{25}$$

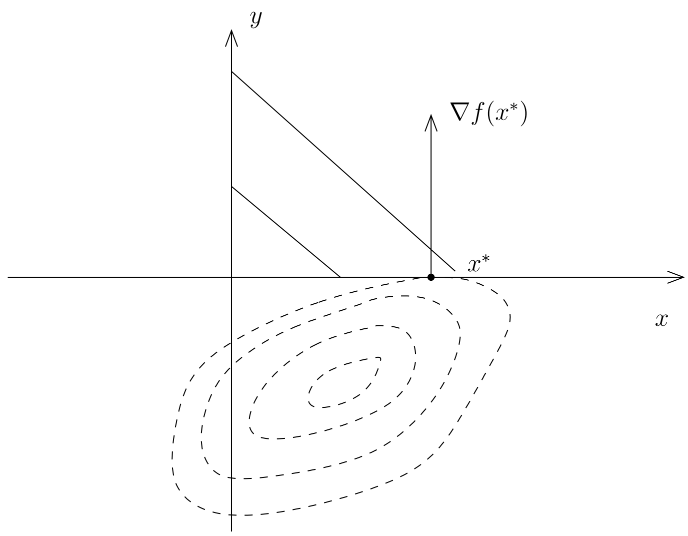
Como cualquier punto factible es regular tenemos que, si x^* es solución del problema entonces existe $\mu \geq 0$ tal que

$$\begin{array}{c} -\nabla f(x^*) = \displaystyle\sum_{i=1}^{p} \mu_i(-e_i) \\ \mu_i x_i^* = 0. \end{array}$$

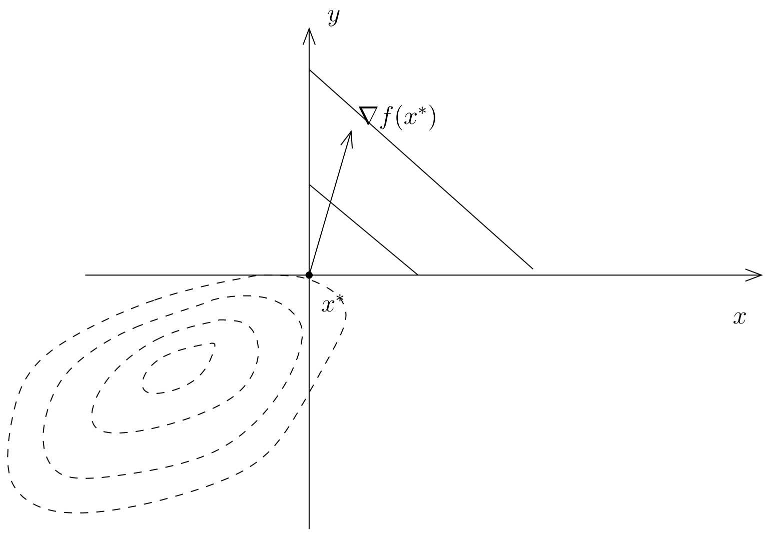
Observar que, si la i-ésima restricción es activa ($x_i^* = 0$) entonces, de la i-ésima ecuación del sistema KKT se obtiene que $-\frac{\partial f}{\partial x_i}(x^*) = -\mu_i$. Si la i-ésima restricción es inactiva ($x_i^* > 0$) entonces $\mu_i = 0$ entonces $\frac{\partial f}{\partial x_i}(x^*) = 0$. Luego, si x^* es un minimizador local del problema entonces se cumple que

$$\begin{array}{lll} \frac{\partial f}{\partial x_i}(x^*) = 0, & & \text{si } x_i^* > 0 \\ \frac{\partial f}{\partial x_i}(x^*) = \mu_i & \geq 0, & \text{si } x_i^* = 0 \end{array}$$

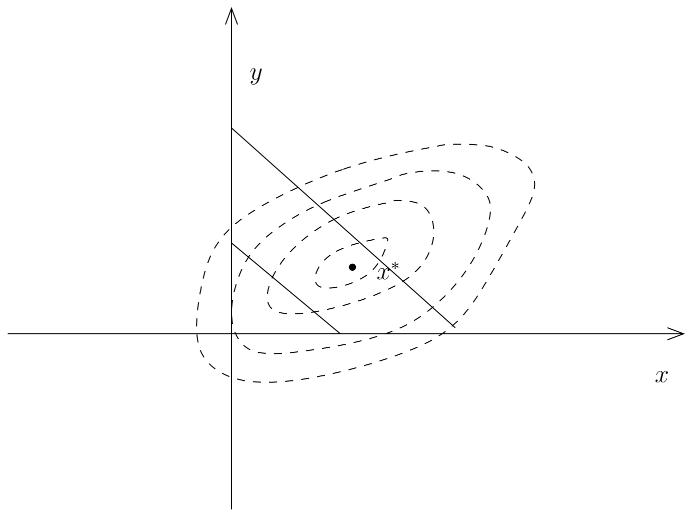
Observar los siguientes ejemplos.



En x^* se tiene:
$x_1^* > 0$ y $\frac{\partial f}{\partial x_1}(x^*) = 0$,
$x_2^* = 0$ y $\frac{\partial f}{\partial x_2}(x^*) > 0$



En x^* se tiene:
$x_1^* = 0$ y $\frac{\partial f}{\partial x_1}(x^*) > 0$,
$x_2^* = 0$ y $\frac{\partial f}{\partial x_2}(x^*) > 0$



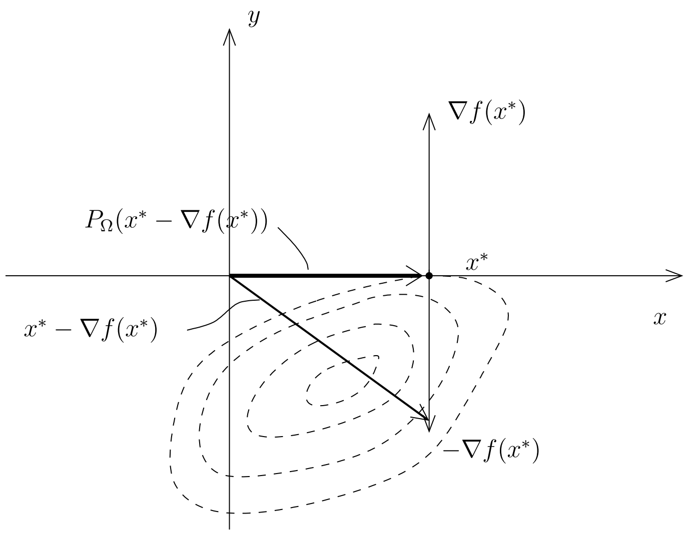
En x^* se tiene:
$x_1^* > 0$ y $\frac{\partial f}{\partial x_1}(x^*) = 0$,
$x_2^* > 0$ y $\frac{\partial f}{\partial x_2}(x^*) = 0$

Dado $\Omega = \{x \in \mathbb{R}^n : x \geq 0\}$, se define la proyección de un punto x sobre Ω como el vector $y = P_\Omega(x)$ cuyas componentes son

$$y_i = \begin{cases} 0 & \text{si } x_i < 0 \\ x_i & \text{si } x_i \geq 0. \end{cases}$$

Ejercicio. Mostrar que x^* es KKT para el problema (25) si y solo si $P_\Omega(x^* - \nabla f(x^*)) - x^* = 0$.

Geometricamente:



Esta condición de optimalidad sugiere el método del gradiente proyectado:

Método del gradiente proyectado.

1. Dados x_0 inicial. En un $x_k \in \Omega$ hacer los siguientes pasos:

2. Calcular

   $y_k = x_k - \nabla f(x_k)$
   $z_k = P_\Omega(y_k)$
   $d_k = z_k - x_k$

   Si $d_k = 0$ parar. (x_k es KKT)

   Si no, hacer búsqueda de Armijo en la dirección d_k:

3. definir $t = 1$,

$$\text{si } f(x_k + td_k) \leq f(x_k) + \sigma_1 t \nabla f(x_k)^T d_k, \text{ entonces} \tag{26}$$

   a) si $x_k + td_k \in \Omega$ definir $t_k = t$ e ir a 2.

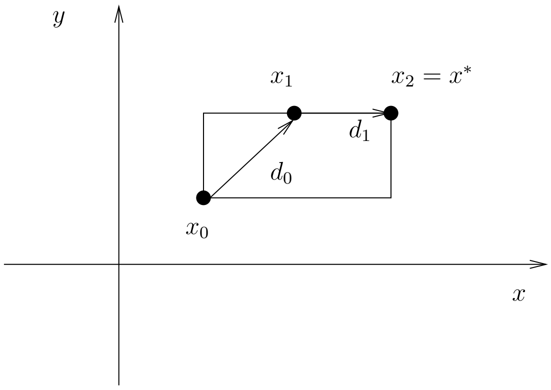
   b) si no, hacer $t = t/2$ e ir a (26).

Observaciones. Si $\Omega = \mathbb{R}^n$ es el método del gradiente con búsqueda de Armijo. El algoritmo es aplicable para las cajas $\Omega = \{x \in \mathbb{R}^n : x \geq 0\}$ o la más general $\Omega = \{x \in \mathbb{R}^n : l \leq x \leq u\}$.

Ejemplo. Hacer algunas iteraciones del método del gradiente proyectado para resolver el problema

$$\text{Minimizar } -x - y \text{ sujeto a } 1 \leq x \leq 3, l \leq y \leq 2.$$

En la figura se observan los iterados comenzando desde $x_0 = (1, 1)$.



Teorema 3.1. *Se puede demostrar lo siguiente:*

1. *Para todo k, d_k es una dirección de descenso para f en x_k.*

2. *Si $z_k = x_k$ para algún k entonces x_k es KKT.*

3. *Todo punto de acumulación es KKT.*

La idea general del método de gradiente proyectado se puede extender para resolver el problema con restricciones lineales de igualdad:

$$\begin{array}{c} \text{Minimizar } f(x) \\ \text{sujeto a } Ax = b, \end{array} \tag{27}$$

donde $A \in \mathbb{R}^{m \times n}, m < n$, rango($A$)$= m$. En este caso se tiene que x^* es KKT si y solo si $P_{Nu(A)}(-\nabla f(x^*)) = 0$.

Ejercicio.
Resolver los siguientes problemas, a partir del punto inicial mencionado, aplicando el método del gradiente proyectado con búsqueda de Armijo presentado previamente.

$$a)\ \begin{array}{ll} \text{Min} & x_1 - x_2 \\ \text{s.a} & 1 \leq x_1 \leq 3 \\ & 1 \leq x_2 \leq 2. \end{array} \qquad x_0 = (3,1)$$

$$b)\ \begin{array}{ll} \text{Min} & x_1^2 + x_2^2 \\ \text{s.a} & 0 \leq x_1 \leq 4 \\ & 1 \leq x_2 \leq 3. \end{array} \qquad x_0 = (4,3)$$

3.2. Método de Conjuntos activos

Este es un método para resolver problemas cuadráticos de la forma

$$\begin{array}{c} \text{Minimizar } \frac{1}{2}x^T Q x + c^T x \\ \text{sujeto a } a_i^T x \leq b_i, i = 1, \ldots, p \end{array} \tag{28}$$

donde Q es una matriz simétrica. El conjunto factible $\{x \in \mathbb{R}^n : a_i^T x \leq b_i, i = 1, \ldots, p\}$ también se representa en forma más compacta de la forma $Ax \leq b$ donde A es la matriz en $\mathbb{R}^{p \times n}$ cuyas filas son los vectores a_i^T.

Sabemos que un punto x^* es KKT si y solo si existen $\mu \in \mathbb{R}^p, \mu \geq 0$ tal que

$$\begin{array}{l} Qx^* + c + A^T \mu = 0 \\ a_i^T x \leq b_i, \mu_i(a_i^T x - b_i) = 0. \end{array}$$

Observar que, si A tiene rango completo (p), como

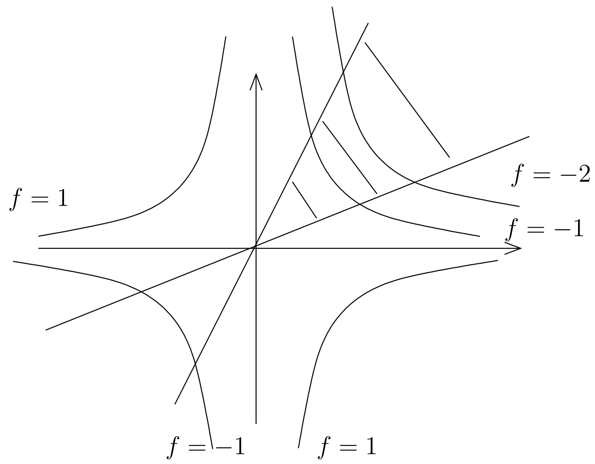
$$\nabla q(x^*) + A^T \mu = 0$$

implica que $\mu = -(AA^T)^{-1} A \nabla q(x^*)$.

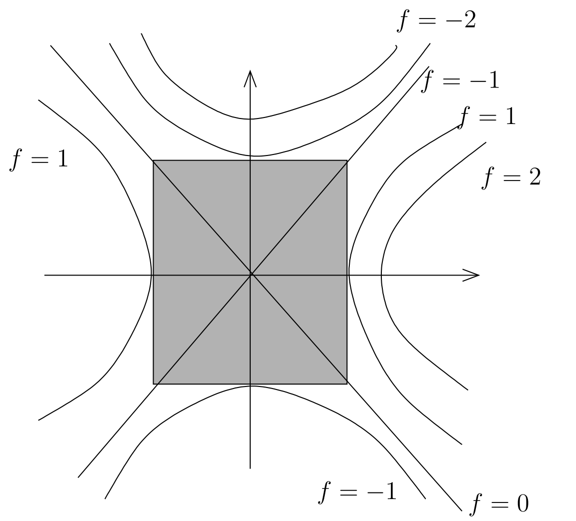
La región factible es un poliedro, definido por la intersección de los semiespacios $a_i^T x \leq b_i$. Puede ser vacío, acotado o no acotado.

Si la matriz $Q > 0$ entonces la cuadrátca es convexa y existe un único minimizador global.

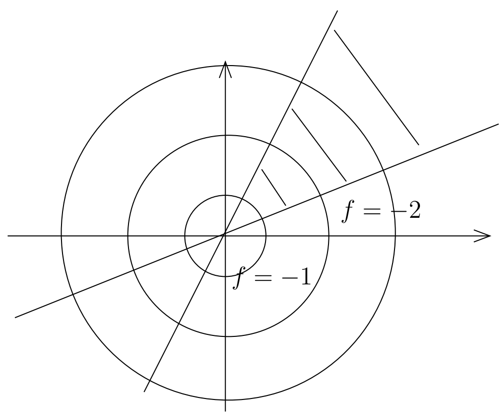
Si $Q \not> 0$ entonces, puede no haber minimizador local o puede haber más de un minimizador local que cumple las condiciones necesarias de segundo orden.



$f(x,y) = -xy$ es indefinida y el problema no tiene solución.



$f(x,y) = x^2 - y^2$ es indefinida y el problema tiene 2 minimizadores locales.



$f(x,y) = -(x^2+y^2)$ es definida negativa y el problema no tiene solución.

Veamos como funciona un m étodo de restricciones activas. Dado un punto x_0 factible, en un iterado x_k se identifican las restricciones que son activas en ese punto:

$$\mathcal{A}(x_k) = \{i : a_i^T x_k = b_i\}$$

y se plantea el subproblema

$$\begin{array}{c} \text{Minimizar } q(x) \\ \text{sujeto a } x \in W_k = \{x \in \mathbb{R}^n : a_i^T x = b_i, i \in \mathcal{A}(x_k)\}. \end{array} \tag{29}$$

Definimos la matriz $A_k = [a_i]_{i \in \mathcal{A}(x_k)}$ la matriz formada a partir de A por las filas que se corresponden con índices en $\mathcal{A}(x_k)$. Si $\tilde{x}$ es un solución del problema anterior entonces $d_k = \tilde{x} - x_k$ es solución del problema

$$\begin{array}{c} \text{Minimizar } \frac{1}{2} d^T Q d + \nabla q(x_k)^T d \\ \text{sujeto a } A_k d = 0, \end{array} \tag{30}$$

Consideramos dos posibilidades:

1. $d_k = 0$.

En este caso, $\tilde{x} = x_k$ y como x_k es solución de (29), por las condiciones KKT sabemos que existe μ tal que $\nabla q(x_k) + A_k^T \mu = 0$. Si $\{a_i\}_{i \in \mathcal{A}(x_k)}$ son linealmente independientes entonces $\mu = -(A_k A_k^T)^{-1} A_k \nabla q(x_k)$.

Si $\mu \geq 0$ entonces x_k es KKT para (28).

Si algún $\mu_i < 0, i \in \mathcal{A}(x_k)$ entonces x_k no es solución y para llegar a un minimizador tenemos que abandonar alguna restricción del conjunto $\mathcal{A}(x_k)$ para la cual $\mu_i < 0$.

2. $d_k \neq 0$.

En este caso buscamos el paso máximo que se puede dar en esa dirección sin salir de la región factible. Aqui intervienen las restricciones que no se encuentran en el conjunto de trabajo.

Si $j \notin \mathcal{A}(x_k)$ entonces $a_j^T(x_k + td_k) = a_j^T x_k + t a_j^T d_k$.

Si $a_j^T d_k \leq 0$ entonces $a_j^T(x_k + td_k) = a_j^T x_k + t a_j^T d_k \leq b_j$ y no se alcanza el borde de la región.

Si $a_j^T d_k > 0$ entonces $a_j^T(x_k + td_k) = a_j^T x_k + t a_j^T d_k = b_j$ si y solo si $t = \frac{b_j - a_j^T d_k}{a_j^T x_k}$.

Luego, el paso máximo que se puede dar en la dirección es

$$t_{max} = \min_{j \notin \mathcal{A}(x_k), a_j^T d_k > 0} \left\{ \frac{b_j - a_j^T x_k}{a_j^T d_k} \right\}.$$

Luego, definimos $t_k = \max\{1, t_{max}\}$.

Si $t_k = 1$ entonces no se ha alcanzado ninguna restricción y se define $x_{k+1} = x_k, W_{k+1} = W_k$.

Si $t_k < 1$ entonces se ha alcanzado alguna restricción, supongamos que fue la restricción j y debemos incorporar esa restricción al nuevo conjunto de trabajo y se define $x_{k+1} = x_k + t_k d_k, W_{k+1} = W_k \cup \{j\}$.

Algoritmo de conjuntos activos.

Sea x_0 inicial, factible, $\mathcal{A}(x_0), [a_i]_{i \in \mathcal{A}(x_0)}$ linealmente independientes.

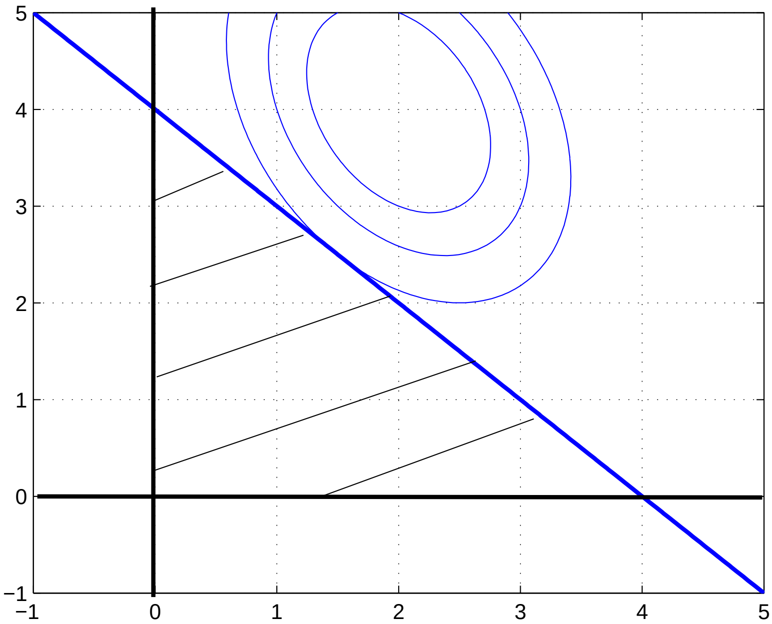


Figura 13: Curvas de nivel y región factible del problema (31).

1. Resolver el subproblema (30) correspondiente.

   Si $d = 0$ ir a [3.]. Sino

2. Calcular t_{max} y t_k.

   Si $t_k = 1$ entonces definir $x_{k+1} = x_k, W_{k+1} = W_k$ e ir a [1.].

   Si $t_k < 1$ entonces definir $x_{k+1} = x_k + t_k d_k, W_{k+1} = W_k \cup \{j\}$ siendo j el índice para el cual $t_k = \frac{b_j - a_j^T x_k}{a_j^T d_k}$ e ir a [1.].

3. Si $d = 0$, calcular $\mu = -(A_k A_k^T)^{-1} A_k \nabla q(x_k)$.

   Si $\mu \geq 0$ terminar. Sino, definir $x_{k+1} = x_k, \mathcal{A}(x_{k+1}) = \mathcal{A}(x_k) - \{j\}$, siendo j tal que la coordenada j de μ es negativa e ir a [1.].

Veamos el proceso en el siguiente ejemplo:

$$\begin{array}{c} \text{Minimizar } 2x^2 + xy + y^2 - 12x - 10y \\ \text{sujeto a } x + y \leq 4, x \geq 0, y \geq 0. \end{array} \tag{31}$$

Observar las curvas de nivel y la región factible en la figura 13.

Definimos las restricciones como (1)$x+y \leq 4$, (2) $x \geq 0$,(3) $y \geq 0$. Luego $A = \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}$.

Supongamos que comenzamos con el punto $x_0 = (0,0)$. En ese punto tenemos que $\mathcal{A}(x_0) = \{2,3\}, W_0 = \{(x,y) : x = 0, y = 0\}$, luego la solución del primer subproblema es $(0,0)$ que se corresponde con $d = (0,0)$. Calculamos los multiplicadores en esa solución. Tenemos que

$A_0 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. Luego, $-\nabla q(x_0) = (12\ 10)^T, \mu \leq 0$. ($-\nabla q(x_0)$ no se puede escribir como combinación positiva de a_2, a_3.)

Entonces, hay que abandonar una restricción, abandonamos por ejemplo la segunda restricción.

Luego, $x_1 = x_0, \mathcal{A}(x_1) = \{3\}, W_1 = \{(x,y) : y = 0\}$ y tenemos un nuevo subproblema. Buscar el minimizador del nuevo subproblema equivale a encontrar d solución de

$$\begin{array}{c} \text{Minimizar } \frac{1}{2}d^TQd + \nabla q(x_1)^T d \\ \text{sujeto a } A_1 d = 0. \end{array}$$

Si $d = (d_1, d_2), A_1 d = 0$ equivale a $d_2 = 0$ luego, la función objetivo del subproblema es, en realidad, una función de una variable d_1 y tiene su menor valor en $d_1 = 3$. Luego, $d = (3,0), x_2 = x_1 + d = (3,0)$. Buscamos ahora el minimizador en esa dirección, en este caso es el mismo punto porque la restricción es $x + y \leq 4$ (si fuera $x + y \leq 2$ deberiamos parar en el punto $(2,0)$).

Continuando con el proceso se obtienen los siguientes iterados $x_3 = (3,0), x_4 = (\frac{8}{3}, \frac{4}{3}), x_5 = (\frac{3}{2}, \frac{5}{2}) = x^*$.

Se pueden demostrar las siguientes propiedades para este método.

Proposición 3.1. *Dados $x_k, \mathcal{A}(x_k), Q > 0$.*

1. *La solución d del problema (30) es factible y de descenso en x_k.*
2. *Si $[a_i]_{i \in \mathcal{A}(x_k)}$ es linealmente independiente entonces $[a_i]_{i \in \mathcal{A}(x_{k+1})}$ es linealmente independiente.*

Teorema 3.2. *Si $Q > 0$ en un número finito de iteraciones el método encuentra un punto estacionario.*

Luego, vemos que el método se reduce a resolver una sucesión finita de problemas con restricciones de igualdad. Se puede extender a problemas con una función objetivo general, no necesariamente cuadrática.

Ejercicio.

1. Resuelva los siguientes problemas utilizando el método de restricciones activas comenzando desde los puntos indicados. Analizar el proceso graficamente cuando sea posible.

a)
$$\begin{array}{ll} \text{Min} & (x_1 - 1)^2 + (x_2 - 2,5)^2 \\ \text{s.a} & -x_1 + 2x_2 \leq 2 \\ & x_1 + 2x_2 \leq 6 \\ & x_1 - 2x_2 \leq 2 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

$x_0 = (2,0)$

b)
$$\begin{array}{ll} \text{Min} & x_1^2 - x_1x_2 + x_2^2 - 3x_1 \\ \text{s.a} & x_1 + x_2 \leq 4 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

$x_0 = (0,0)$

3.3. Métodos de penalidad y barrera

Este tipo de métodos buscan la solución del problema con restricciones no lineales generales mediante el límite de soluciones de problemas sin restricciones (o con restricciones más fáciles).

En el caso de métodos de penalidad la función original es reemplazada por una función que se forma mediante la suma de la función objetivo más un término que penaliza el hecho de que la restricción no se verifique.

En el caso de métodos de barrera, a la función objetivo se le suma un término que favorece puntos interiores de la región factible sobre los puntos que estan en la frontera.

Asociados a estos métodos hay un parámetro, de penalidad o de barrera, que determina la severidad del castigo.

Para fijar ideas, consideremos de forma general el problema de Minimizar $f(x)$ sujeto a $x \in S$, donde S es el conjunto factible que puede estar formado por restricciones de igualdad y desigualdad generales.

La idea del método de penalidad es reeemplazar el problema original por un problema sin restricciones de la forma Minimizar $f(x) + \rho P(x)$ donde $\rho > 0$ es constante (parámetro de penalidad) y P es una función de $\mathbb{R}^n$ que tiene por objetivo penalizar las restricciones que definen S. P debe cumplir:

- ser continua,
- $P(x) \geq 0$,
- $P(x) = 0$ si y solo si $x \in S$.

Por ejemplo, si $S = \{x : h_i(x) = 0, i = 1, \ldots, m\}$ pueden considerarse las funciones

1. $P(x) = \displaystyle\sum_{i=1}^{m} (h_i(x))^2$, llamada función de penalidad cuadrática,
2. $P(x) = \displaystyle\sum_{i=1}^{m} |h_i(x)|$.

Si $S = \{x : g_i(x) \leq 0, i = 1, \ldots, p\}$ pueden considerarse las funciones

1. $P(x) = \displaystyle\sum_{i=1}^{p} (\max\{0, g_i(x)\})^2$, llamada función de penalidad cuadrática,
2. $P(x) = \displaystyle\sum_{i=1}^{p} \max\{0, g_i(x)\}$.

La idea es, una vez elegida un función de penalidad, resolver una sucesión de problemas irrestrictos:

$$\text{Minimizar } f(x) + \rho_k P(x)$$

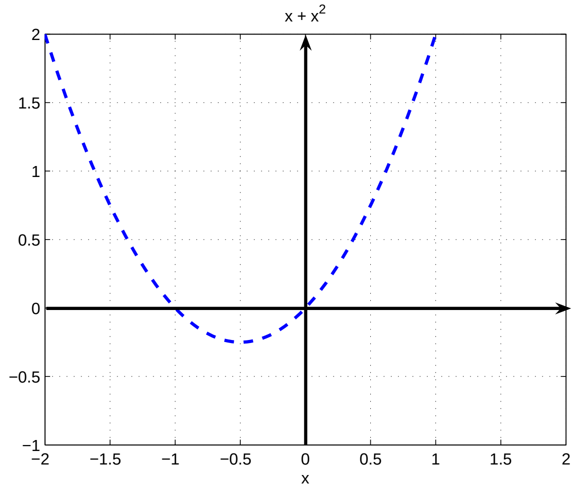
para una sucesión $\{\rho_k\}$ que tiende a infinito. Es decir, a medida que ρ_k aumenta, la no verificación de las restricciones se torna más cara. Podemos esperar que las soluciones de los problemas irrestrictos se aproximen la conjunto factible cuando $\rho_k \to \infty$ y sus puntos limites sean solución del problema original.

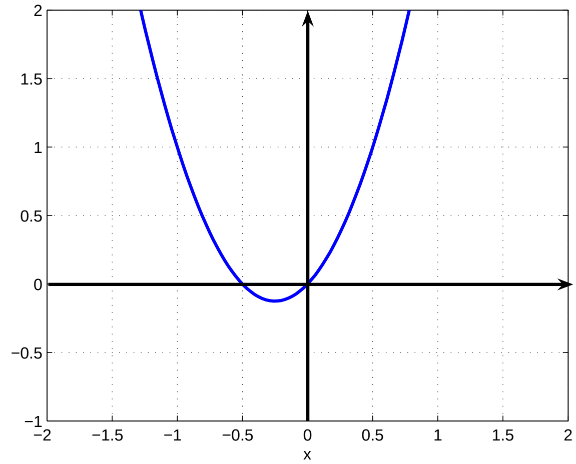
Consideremos el problema con restricciones de igualdad

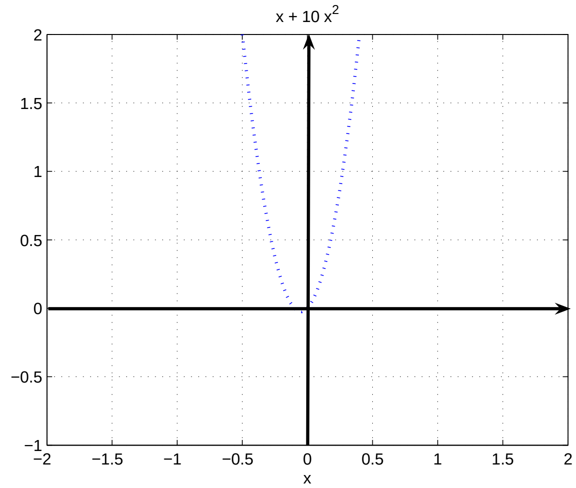
$$\begin{array}{ll} \text{Min} & x \\ \text{s.a} & x = 0 \end{array}$$

y la función de penalidad cuadrática $P(x) = x^2$.

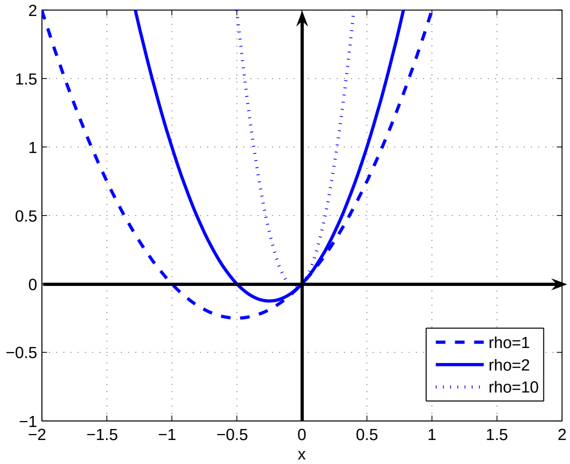
Las siguientes son las gráficas de la función $P(x,\rho) = f(x) + \rho P(x)$ para los valores de $\rho = 1, 2, 10$







Observar las tres gráficas juntas:

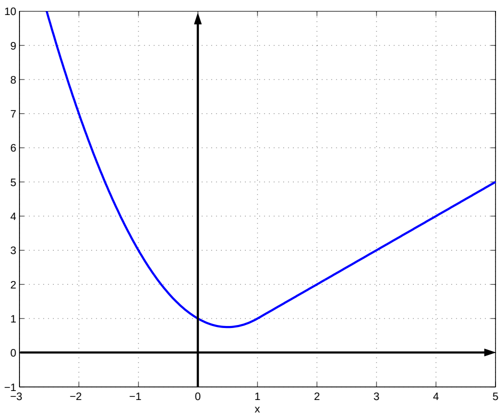


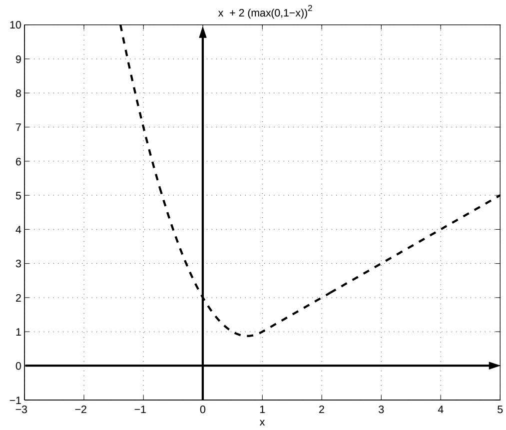
Consideremos el problema con restricciones de desigualdad

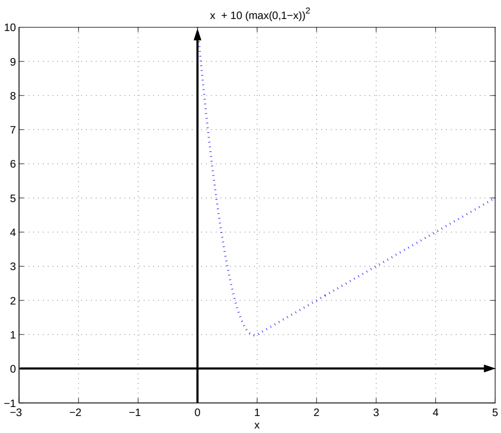
$$\begin{array}{cc} \text{Min} & x \\ \text{s.a} & x \geq 1 \end{array}$$

y la función de penalidad cuadrática $P(x) = \text{máx}\{0, 1-x\}^2$.

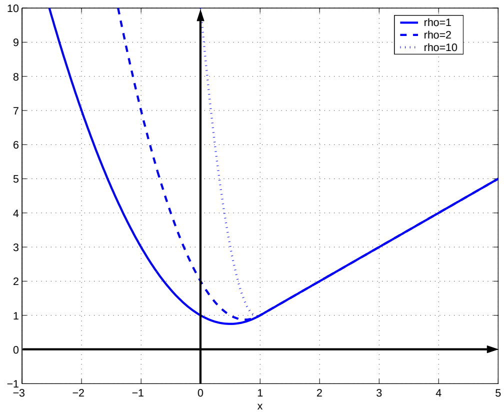
Las siguientes son las gráficas de la función $P(x,\rho) = f(x) + \rho P(x)$ para los valores de $\rho = 1, 2, 10$



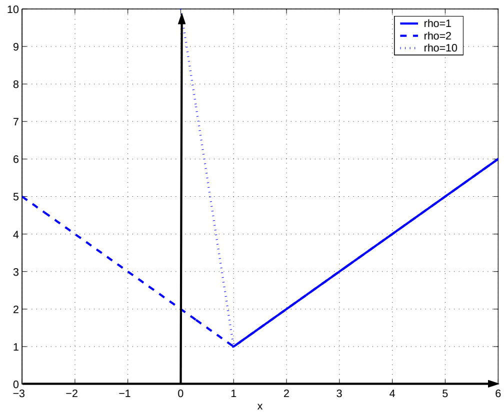




Observar las tres gráficas juntas:



Para el mismo problema, observemos las gráficas para la función de penalidad $P(x) = max\{0, 1 - x\}$.



Lema 3.1. *Definimos* $\varphi(x, \rho) = f(x) + \rho P(x)$*. Si* $\{x_k\}$ *es la sucesión generada por el método de penalidad utilizando una sucesión creciente de parámetros* ρ_k*, entonces*

1. $\varphi(x_k, \rho_k) \leq \varphi(x_{k+1}, \rho_{k+1})$

2. $P(x_k) \geq P(x_{k+1})$

3. $f(x_k) \leq f(x_{k+1})$.

Dem.

1. $\varphi(x_{k+1}, \rho_{k+1}) = f(x_{k+1}) + \rho_{k+1}P(x_{k+1}) \geq f(x_{k+1}) + \rho_k P(x_{k+1}) \geq f(x_k) + \rho_k P(x_k) = \varphi(x_k, \rho_k)$, ya que x_k es el minimizador de $f(x) + \rho_k P(x)$.

2. Por definición de x_k tenemos que

$$f(x_k) + \rho_k P(x_k) \leq f(x_{k+1}) + \rho_k P(x_{k+1}).$$

Por definición de x_{k+1} tenemos que

$$f(x_{k+1}) + \rho_{k+1}P(x_{k+1}) \leq f(x_k) + \rho_{k+1}P(x_k).$$

Sumando,

$$\rho_k P(x_k) + \rho_{k+1} P(x_{k+1}) \le \rho_k P(x_{k+1}) + \rho_{k+1} P(x_k).$$

Luego,

$$(\rho_{k+1} - \rho_k) P(x_{k+1}) \le (\rho_{k+1} - \rho_k) P(x_k)$$

y se obtiene el resultado ya que $\rho_{k+1} - \rho_k \ge 0$.

3. Usando 2.

$$f(x_k) + \rho_k P(x_k) \le f(x_{k+1}) + \rho_k P(x_{k+1}) \le f(x_{k+1}) + \rho_k P(x_k),$$

luego se obtiene 3. □

Lema 3.2. *Si x^* es solución del problema original entonces*

$$f(x^*) \ge \varphi(x_k, \rho_k) \ge f(x_k).$$

Dem.

$$f(x^*) = f(x^*) + \rho_k P(x^*) \ge f(x_k) + \rho_k P(x_k) \ge f(x_k).$$

□

Teorema 3.3. *Sea $\{x_k\}$ una sucesión generada por el método de penalidad. Entonces, todo punto limite de $\{x_k\}$ es solución.*

Dem.

Sea x^* punto limite de la sucesión. Entonces existe $K \subset \mathbb{N}$ tal que $x_k \to_{k\in K} x^*$. Por continuidad sabemos que $f(x_k) \to_{k\in K} f(x^*)$.

Sea $f^* = \text{mín}_{x\in S} f(x)$ el valor optimo del problema original.

Por los lemas previos, la sucesión $\{\varphi(x_k, \rho_k)\}$ es no decreciente y acotada por el valor f^*. Entonces

$$\lim_{k\in K} \varphi(x_k, \rho_k) \le f^*.$$

Luego,

$$\lim_{k\in K} \rho_k P(x_k) = \lim_{k\in K} \varphi(x_k, \rho_k) - f(x_k) = \alpha^* - f(x^*)$$

siendo $\alpha^* = \lim_{k\in K} \varphi(x_k, \rho_k) \le f^*$.

Como $P(x_k) \ge 0$ y $\rho_k \to \infty$ tenemos que debe ser $\lim_{k\in K} P(x_k) = 0$. Luego, por continuidad $P(x^*) = 0$ es decir, x^* es factible.

Además, tenemos que $f(x_k) \le f^*$. Entonces, $f(x^*) = \lim_{k\in K} f(x_k) \le f^*$. esto implica que $f(x^*) = f^*$. □

Consideremos la función de penalidad cuadrática para el problema con restricciones de igualdad

$$\text{Minimizar } f(x) \text{ sujeto a } h(x) = 0. \tag{32}$$

El subproblema que aparece en cada paso del método es

$$\text{Minimizar } f(x) + \frac{\rho_k}{2} \sum_{i=1}^{m} (h_i(x))^2.$$

Por un lado, un punto x_k estacionario de este subproblema verifica:

$$\nabla f(x_k) + \rho_k \sum_{i=1}^{m} h_i(x_k)\nabla h_i(x_k) = 0. \tag{33}$$

Por otro lado, para el problema general (32), buscamos un $\lambda \in \mathbb{R}^m$ tal que

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0. \tag{34}$$

El siguiente teorema nos da condiciones necesarias bajo las cuales $\rho_k h_i(x_k) \to \lambda_i$.

Teorema 3.4. *Sea $\{x_k\}$ la sucesión generada por el método de penalidad con la función cuadrática. Si $x_k \to x^*$ solución regular entonces $\lambda_k = \rho_k h(x_k)$ converge al multiplicador de Lagrange asociado.*

Dem.

De la igualdad (33) tenemos que

$$\nabla f(x_k) + (Jh(x_k))^T \lambda_k = 0. \tag{35}$$

Como x^* es regular sabemos que existe un único λ que verifica (34), luego

$$\lambda = -(Jh(x^*)Jh(x^*)^T)^{-1} Jh(x^*)\nabla f(x^*).$$

Como $Jh(x^*)$ tiene filas linealmente independientes, por continuidad, $Jh(x_k)$ también tiene filas linealmente independientes para k suficientemente grande. Luego, por (35) tenemos que

$$\lambda_k = -(Jh(x_k)Jh(x_k)^T)^{-1} Jh(x_k)\nabla f(x_k)$$

y por continuidad se demuestra el teorema. $\square$

Este teorema se puede generalizar para desigualdades usando la función de penalidad cuadrática.

Teorema 3.5. *Sea $\{x_k\}$ la sucesión tal que $\nabla f(x_k) + \sum_{i=1}^{p} \rho_k \max\{0, g_i(x_k)\}\nabla g_i(x^*) = 0$. Si $x_k \to x^*$ solución regular del problema de Minimizar $f(x)$ sujeto a $g_i(x) \leq 0, i = 1, \ldots, p$ entonces $\mu_k = \rho_k \max\{0, g(x_k)\}$ converge al multiplicador de Lagrange asociado.*

Los métodos de barrera se aplican exclusivamente a problemas con restricciones de desigualdad.

Consideramos el problema

$$\text{Minimizar } f(x) \ \text{ sujeto a } g_i(x) \leq 0, i = 1, \ldots, p \tag{36}$$

y suponemos que el interior relativo del conjunto factible es no vacio: $int(S) = \{x \in \mathbb{R}^n : g_i(x) < 0\} \neq \emptyset$. Definimos por Ω al conjunto factible.

En este caso la función objetivo se reemplaza por una función de la forma $f(x) + \mu B(x)$ donde la función $B(x)$ definida para $x \in int(S)$ debe cumplir:

1. B debe ser continua en $int(S)$,
2. $B(x) \geq 0, \forall x \in int(S)$,
3. Si $\{x_k\} \subset \Omega, g_i(x_k) < 0$ y, para algún $i \in \{1, \ldots, p\}$ $\lim_{k\to\infty} g_i(x_k) = 0$ entonces $\lim_{k\to\infty} B(x_k) = \infty$.

La última condición establece que la barrera se acerca a infinito cuando un punto se acerca a la frontera de la región.

Algunas funciones de barrera conocidas en la literatura:

1. $B(x) = -\sum_{i=1}^{p} \frac{1}{g_i(x)}$ es llamada *barrera inversa.*
2. $B(x) = -\sum_{i=1}^{p} ln(-g_i(x))$ es llamada *barrera logaritmica.*

La idea es, una vez elegida una función barrera, resolver una sucesión de problemas de la forma

$$\text{Minimizar } f(x) + \mu_k B(x), \text{sujeto a } x \in int(S)$$

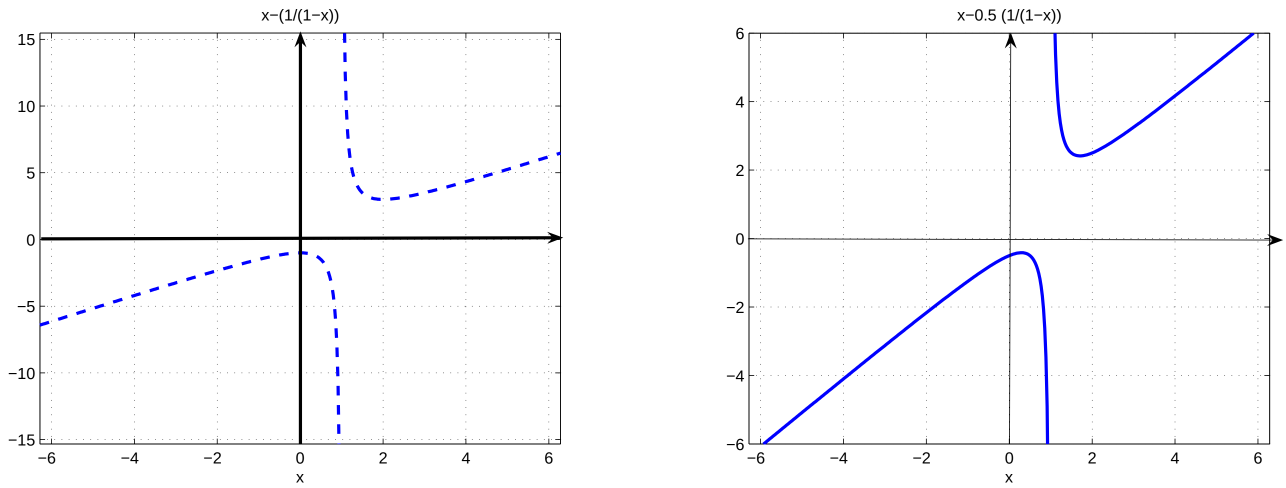
donde μ_k es una sucesión decreciente de escalares no negativos que tiende a cero.

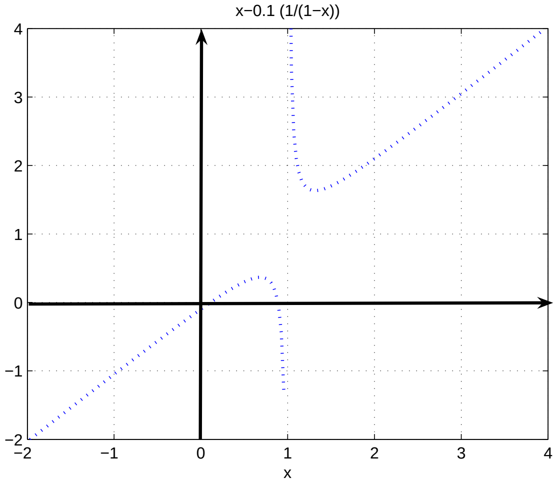
Consideremos el problema

$$\begin{array}{ll} \text{Min} & x \\ \text{s.a} & x \geq 1. \end{array}$$

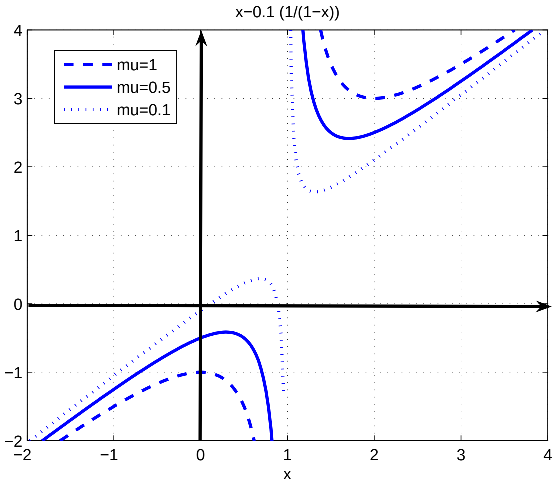
Consideramos la función barrera inversa $B(x) = -\frac{1}{1-x}$.

Las siguiente son la gráficas de la función $Q(x,\mu) = f(x) + \mu B(x)$ para los valores de $\mu = 1; 0{,}5; 0{,}1$



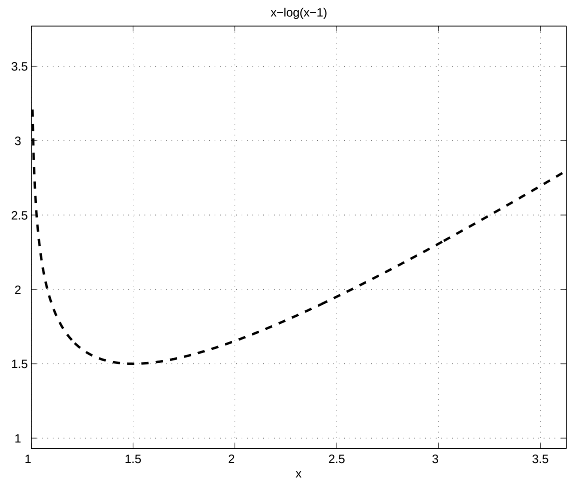


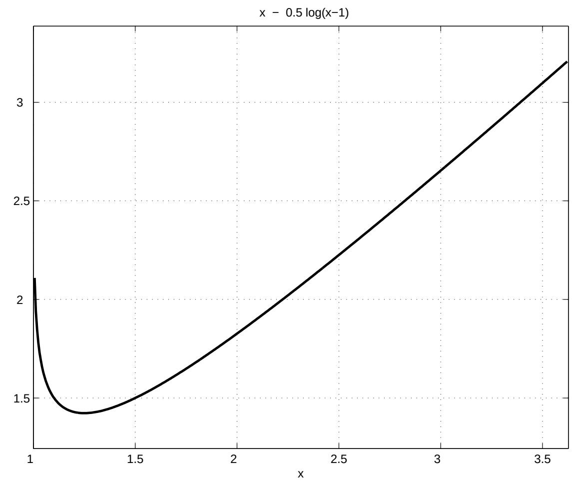
Observar las tres gráficas juntas:

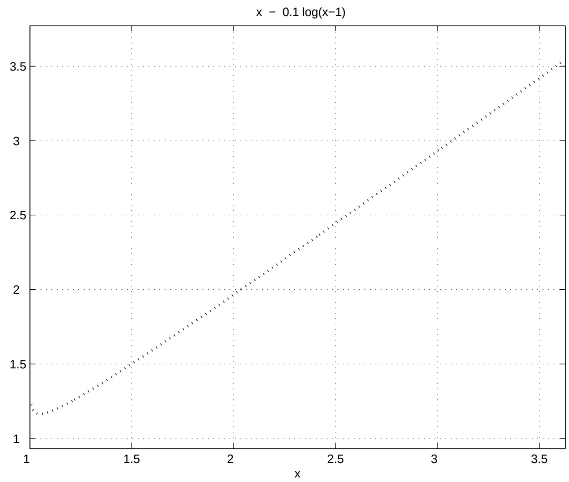


Veamos las gráficas para el mismo problema utilizando la función barrera logaritmica. $B(x) = -log(x-1)$.

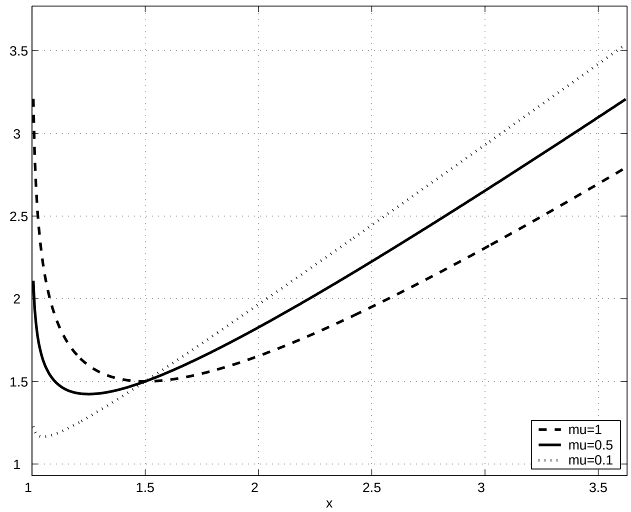
Las siguiente son la gráficas de la función $Q(x, \mu) = f(x) + \mu B(x)$ para los valores de $\mu = 1, 0{,}5, 0{,}1$







Observar las tres gráficas juntas:

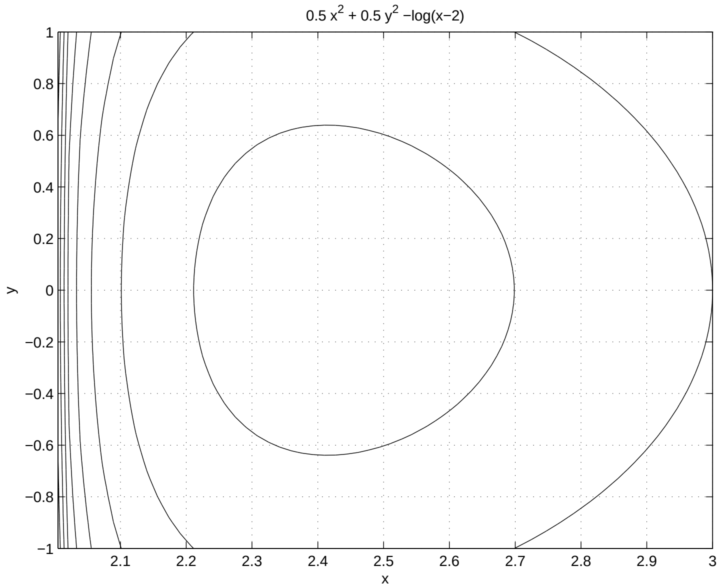


Consideremos el problema

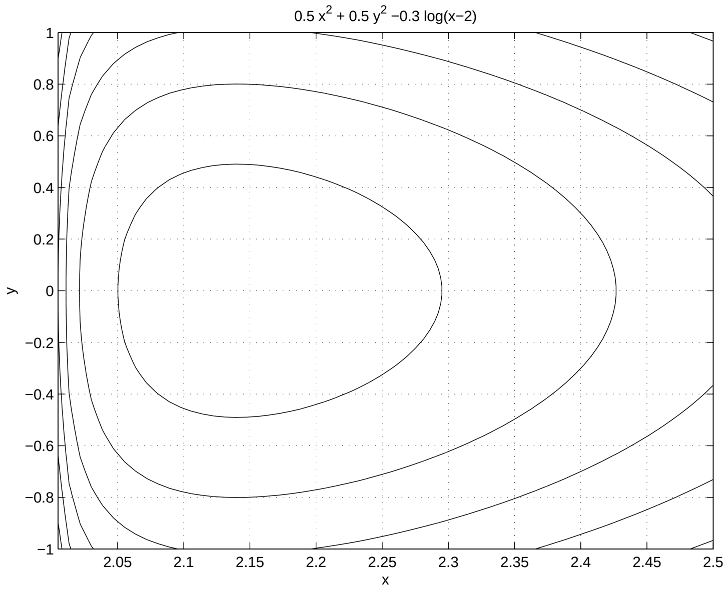
$$\begin{array}{ll} \text{Min} & 0{,}5(x^2+y^2) \\ \text{s.a} & x \geq 2. \end{array}$$

y la función barrera logaritmica $B(x) = -ln(x-2)$.

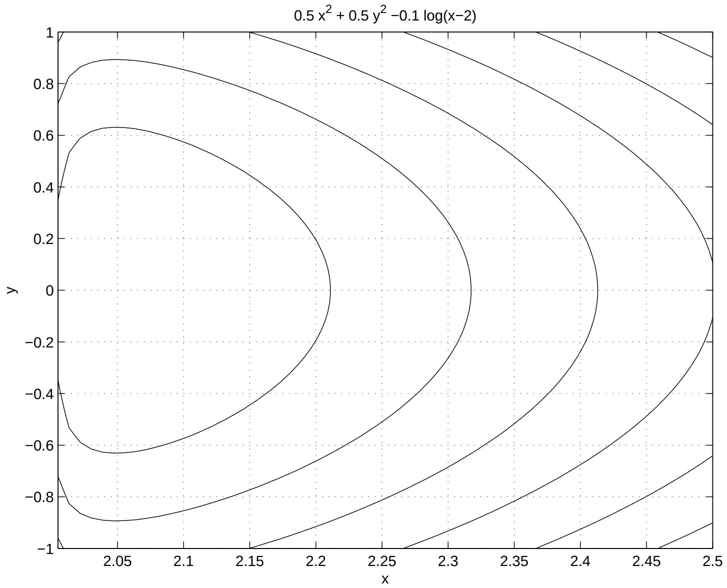
Las siguientes son las curvas de nivel de la función $Q(x,\mu) = f(x) + \mu B(x)$ para el valor de $\mu = 1$



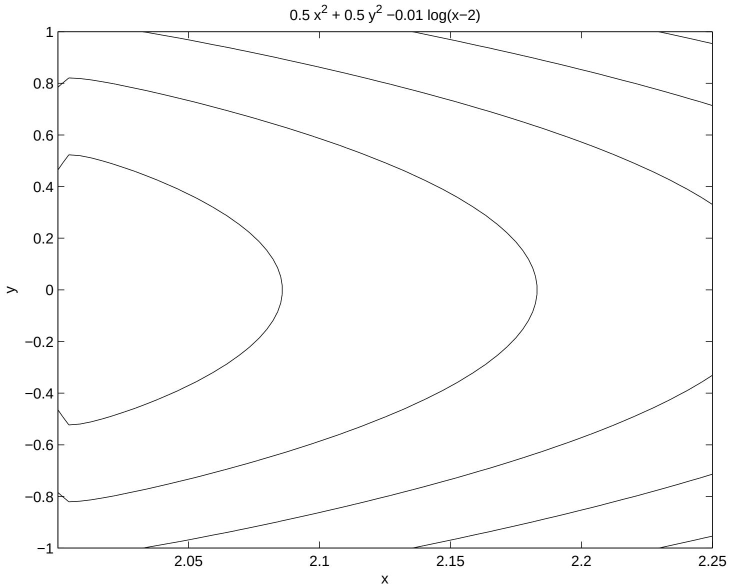
Las siguientes son las curvas de nivel de la función $Q(x, \mu) = f(x) + \mu B(x)$ para el valor de $\mu = 0{,}3$



Las siguientes son las curvas de nivel de la función $Q(x, \mu) = f(x) + \mu B(x)$ para el valor de $\mu = 0{,}1$



Las siguientes son las curvas de nivel de la función $Q(x,\mu) = f(x) + \mu B(x)$ para el valor de $\mu = 0{,}01$



De manera similar al método de penalidad se obtienen los siguientes resultados.

Lema 3.3. *Definimos $Q(x,\mu) = f(x) + \mu B(x)$. Si $\{x_k\}$ es la sucesión generada por el método de barrera utilizando una sucesión decreciente de parámetros μ_k, entonces*

1. $Q(x_{k+1}, \mu_{k+1}) \leq Q(x_k, \mu_k)$
2. $B(x_k) \leq B(x_{k+1})$
3. $f(x_{k+1}) \leq f(x_k)$.

Teorema 3.6. *Sea $\{x_k\}$ una sucesión generada por el método de barrera. Entonces, todo punto limite de $\{x_k\}$ es solución.*

Ejercicios.

1. Considere el problema
$$\begin{array}{ll} \text{Min} & x_1^2 - x_2 \\ \text{s.a} & x_1 + x_2 = 6, x_1 \geq 0. \end{array}$$
Mostrar que si se utiliza el método de penalidad con la función de penalidad cuadrática, entonces $(x_1^k, x_2^k) \to x^*$ cuando $\rho_k \to \infty$.

2. Considere el problema
$$\begin{array}{ll} \text{Min} & 2x_1^2 + 9x_2 \\ \text{s.a} & x_1 + x_2 \geq 4. \end{array}$$
Mostrar que si se utiliza el método de barrera inversa, entonces $(x_1^k, x_2^k) \to x^*$ cuando $\mu_k \to 0$.

3. Considere el problema
$$\begin{array}{ll} \text{Min} & -30x_1 + 3x_1^2 - 8x_2 + 2x_2^2 \\ \text{s.a} & 3x_1 + 2x_2 \leq 6. \end{array}$$
Calcular la solución mediante la aplicación del método de barrera logaritmica. Calcular el multiplicador de Lagrange asociado.

3.4. Método de Lagrangiano Aumentado

Un defecto importante de los métodos de penalidad es la necesidad de hacer $\rho_k \to \infty$ para probar convergencia a la solución. Esto genera problemas numéricos en la resolución de los subproblemas.

Consideremos el problema con restricciones de igualdad (11). Las condiciones KKT para este problema vienen dadas por:

$$\begin{aligned} \nabla f(x) + Jh(x)^T\lambda &= 0 \\ h(x) &= 0. \end{aligned}$$

Sabemos que, si x^* es un punto KKT y λ es el multiplicador entonces (x^*, λ) es un punto estacionario de la función de Lagrange $l(x, \lambda)$. Lamentablemente, aunque conocieramos de antemano λ, x^* puede no ser un minimizador de $l(x, \lambda)$.

Consideremos por ejemplo el problema de minimizar x^3 sujeto a $x + 1 = 0$. Tenemos que $x^* = -1, \lambda = -3$ y $l''(x^*, \lambda) < 0$. Sin embargo se puede demostrar el siguiente resultado.

Teorema 3.7. *Si x^* cumple las condiciones suficientes de segundo orden y λ es el multiplicador de Lagrange asociado existe $\bar{\rho} \geq 0$ tal que la función*

$$\bar{l}(x) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \frac{\rho}{2} \sum_{i=1}^{m} (h_i(x))^2$$

tiene un minimizador local en x^* para todo $\rho \geq \bar{\rho}$.

Dem.

Ver el libro [6]. □

Luego, si conocieramos de antemano los multiplicadores, bastaría un valor finito de ρ para transformar el problema original en un problema sin restricciones.

El problema (11) es equivalente, para λ fijo, al problema

$$\text{Minimizar } f(x) + \sum_{i=1}^{m} h_i(x) \text{ sujeto a } h(x) = 0. \tag{37}$$

Si aplicamos el método de penalidad cuadrática a este problema tenemos en cada iteración el subproblema

$$\text{Minimizar} f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \frac{\rho}{2} \sum_{i=1}^{m} (h_i(x))^2$$

que, para diferentes λ son subproblemas diferentes. Si buscamos un punto estacionario tenemos que

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \rho \sum_{i=1}^{m} h_i(x) \nabla h_i(x) = 0$$

que equivale a

$$\nabla f(x) + \sum_{i=1}^{m} (\lambda_i + \rho h_i(x)) \nabla h_i(x) = 0.$$

Comparando con el sistema KKT del problema original, se puede deducir que, para λ fijo, el escalar $\lambda + \rho h(x)$ podría ser una buena estimativa para llegar a calcular el multiplicador asociado. Esto suigere el siguiente método.

Método de Lagrangiano Aumentado.

Dados x_0 inicial, $\rho_1 > 0, \lambda_1 \in \mathbb{R}^m$.

1. Calcular x_k minimizador local de $f(x) + \sum_{i=1}^{m} [\lambda_i]_k h_i(x) + \frac{\rho_k}{2} \sum_{i=1}^{m} (h_i(x))^2$.
2. Si $\|h(x_k)\| > 0{,}1 \|h(x_{k-1})\|$ entonces $\rho_{k+1} = 10\rho_k$.
3. Definir $\lambda_{k+1} = \lambda_k + \rho_k h(x_k)$ y repetir.

La función $L(x, \lambda, \rho)$ se denomina *función Lagrangiano Aumentado.*

La solución de cada subproblema del paso 1. es un punto x_k que cumple

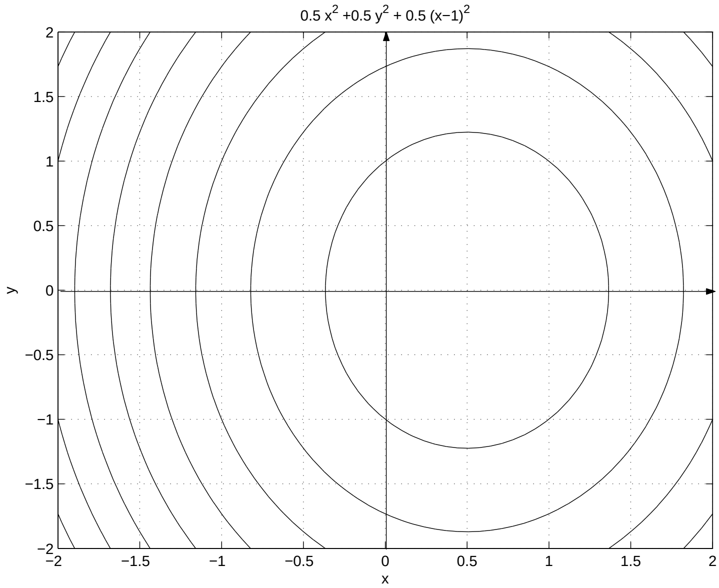
$$\nabla f(x_k) + \sum_{i=1}^{m} ([\lambda_k]_i + \rho_k h_i(x_k)) \nabla h_i(x_k) = 0$$

pero puede no ser factible, o puede, no ser “tan factible” como lo era x_{k+1}. Por eso, en el paso 2. comparamos si $\|h(x_k)\|$ decrece comparada con el valor previo $\|h(x_{k-1})\|$.

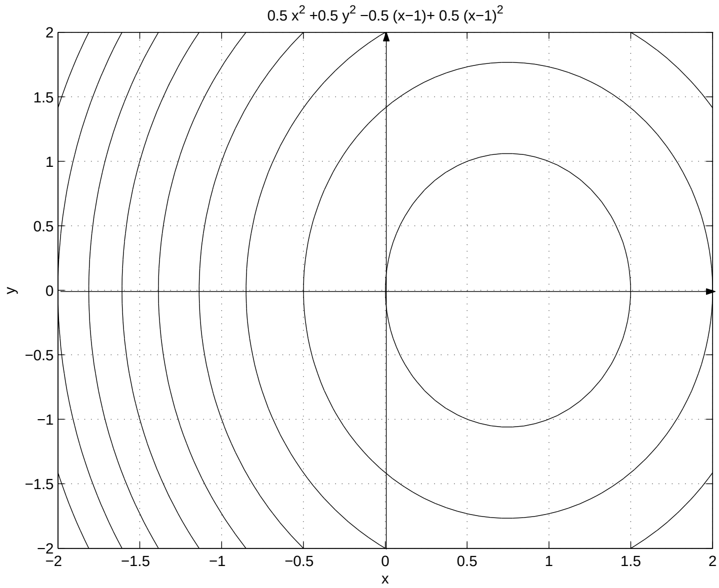
Consideremos el problema

$$\begin{array}{ll} \text{Min} & \frac{1}{2}(x^2 + y^2) \\ \text{s.a} & x = 1. \end{array}$$

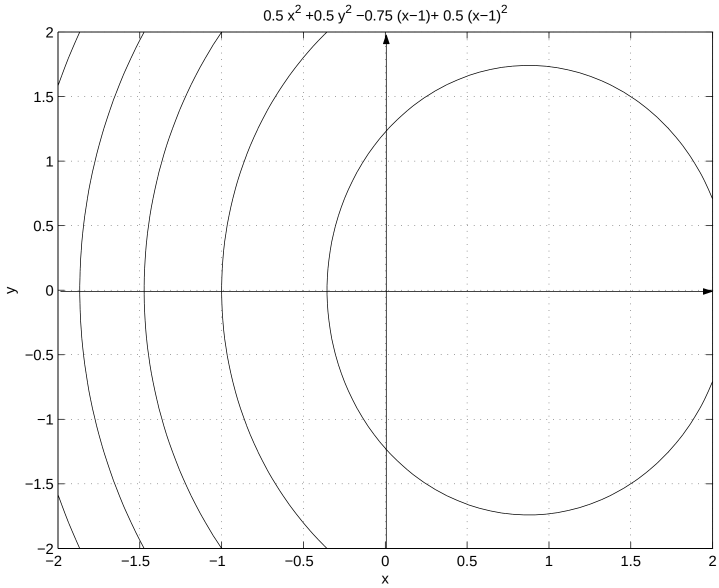
Comenzando con $x_0 = (0, 0), \lambda_1 = 0, \rho_1 = 1$, las siguientes son la curvas de nivel de la función Lagrangiano aumentado $L(x, \lambda_1, \rho_1)$.



Para esta función, la solución es el punto $x_1 = (0{,}5, 0)$. Se tiene $\rho_2 = 1, \lambda_2 = -0{,}5$.
Las siguientes son la curvas de nivel de la función Lagrangiano aumentado $L(x, \lambda_2, \rho_2)$.



Para esta función, la solución es el punto $x_2 = (0{,}75, 0)$. Se tiene $\rho_3 = 1, \lambda_3 = -0{,}75$.
Las siguientes son la curvas de nivel de la función Lagrangiano aumentado $L(x, \lambda_3, \rho_3)$.



Teorema 3.8. *Sea x_k minimizador global de la función $L(x, \lambda_k, \rho_k)$. Suponemos que λ_k es acotada y que $\{\rho_k\}$ es una sucesión creciente que tiende a ∞. Entonces, todo punto limite de $\{x_k\}$ es minimizador global de f sujeto a $h(x) = 0$.*
Dem.

Se puede encontrar un prueba en [6].

Proposición 3.2. *Sea x_k tal que $\|\nabla_x L(x, \lambda_k, \rho_k)\| \leq \varepsilon_k$, λ_k acotada, $\{\rho_k\}$ es una sucesión creciente que tiende a ∞ y $\varepsilon_k \to 0$. Si $\{x_k\}_{k\in K}$ es una subsucesión tal que $x_k \to_{k\in K} x^*$ para x^* regular entonces x^* es KKT y $\lambda_k + \rho_k h(x_k) \to_{k\in K} \lambda$ multiplicador de Lagrange asosiado.*

Este método se puede extender para problemas que además tienen restricciones de desigualdad, [6].

Ejercicio. Considere el problema

$$\begin{array}{ll} \text{Min} & 2x_1^2 + 2x_1x_2 + x_2^2 - 2x_2 \\ \text{s.a} & \qquad\qquad x_1 = 0. \end{array}$$

1. Calcule la solución (x^*, λ^*).
2. Existe $\bar{\rho} > 0$ para el cual x^* es minimizador de la función Lagrangiano aumentada $L(x, \lambda^*, \rho)$ para todo $\rho \geq \bar{\rho}$? Si existe, calcularlo.
3. Hacer 3 iteraciones el método de Lagrangiano aumentado comenzando con $x_0 = (0, 0), \rho_0 = 1, \lambda_0 = 0$. Hay convergencia?

Referencias

[1] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer, 1999.

[2] D. G. Luenberger. *Linear and Nonlinear Programming, 2nd Edition.* Addison- Wesley Inc., Reading, Massachusetts, 1984.

[3] J. Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations.* SIAM, 1996.

[4] A. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust region methods.* MPS-SIAM Series on Optimization, 2000.

[5] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming: Theory and Algorithms.* Wiley, third edition, 2006.

[6] D.P. Bertsekas. *Nonlinear Programming: 2nd Edition.* Athena Scientific, 1999.