Programación No Lineal

INFO-1159

Dr. Julio Rojas Mora

Departamento de Ingeniería Informática

UNIVERSIDAD CATÓLICA DE TEMUCO

Problema de optimización general

Definición

Un problema generalizado de optimización busca n variables de decisión $x_1, x_2, \ldots, x_n$ de una región factible, de tal manera que se optimiza (maximiza o minimiza) un función objetivo dada:

$$f(x_1, x_2, \ldots, x_n)$$

de las variables de decisión.

Problema de programación no lineal (NLP)

Definición

Un problema de programación no lineal (NLP) es un problema de optimización general en el que la función objetivo es no lineal y/o la región factible está definida por restricciones no lineales. Por lo tanto, un programa no lineal es definido de manera general como:

$$\text{Maximizar } f(x_1, x_2, \ldots, x_n)$$

sujeto a:

$$\begin{aligned} g_1(x_1, x_2, \ldots, x_n) &\le b_1 \\ &\vdots \\ g_m(x_1, x_2, \ldots, x_n) &\le b_m \end{aligned}$$

Problema de programación no lineal (NLP)

Observación

Agregando variables de holgura y utilizando notación vectorial, tenemos que:

$$\text{Maximizar } f(\mathbf{x})$$

$$\text{sujeto a:}$$

$$g_j(\mathbf{x}) + h_j(\mathbf{x}) = b_j,\ \forall j = 1, \dots, m$$

$$x_i \geq 0,\ \forall i = 1, \dots, n$$

Problema de programación no lineal (NLP)

Ejemplo

Un inversor tiene US$5000 para invertir en dos potenciales instrumentos financieros, siendo x_1 y x_2 la cantidad invertida en cada uno de ellos. De datos históricos se espera que la inversión en el primero de ellos tenga una tasa de retorno de 20 %, mientras que la del segundo sea de 16 %. El riesgo, medido como la varianza del retorno total de la inversión, es dado por la función $2x_1^2 + x_2^2 + (x_1 + x_2)^2$. Por tanto, el riesgo es una función tanto de la inversión en cada instrumento separado como de la inversión total. El inversor quisiera maximizar su retorno esperado a la vez que minimiza su riesgo.

UNIVERSIDAD CATÓLICA DE TEMUCO

Solución gráfica de un NLP

Solución

De manera general, no se podrían satisfacer ambos objetivos de manera simultánea. No obstante el retorno de la inversión y el riesgo pueden ser combinados en una función objetivo, obteniendo el siguiente modelo:

$$\text{Maximizar } f(\mathbf{x}) = 1{,}20x_1 + 1{,}16x_2 - \theta\left(2x_1^2 + x_2^2 + (x_1 + x_2)^2\right)$$

sujeto a:

$$\begin{aligned} g_1(x) = & x_1 + x_2 \leq 5000 \\ & x_1, x_2 \geq 0 \qquad \text{(esto es equivalente a} \\ & \qquad g_2(x) = -x_1 \leq 0, \\ & \qquad g_3(x) = -x_2 \leq 0) \end{aligned}$$

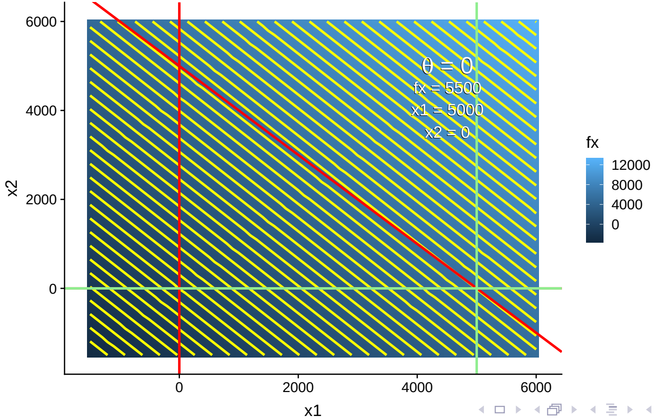
donde θ refleja la compensación entre el riesgo y el retorno de la inversión, es decir, la aversión al riesgo del inversor.

CATÓLICA DE TEMUCO

Solución gráfica de un NLP

Ejemplo

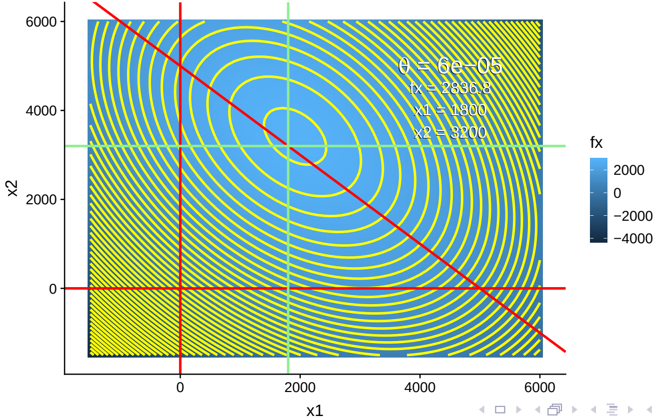
Cuando $\theta = 0$, el problema es se convierte en uno de programación líneal, por lo que la solución está en uno de los vértices.



Solución gráfica de un NLP

Ejemplo

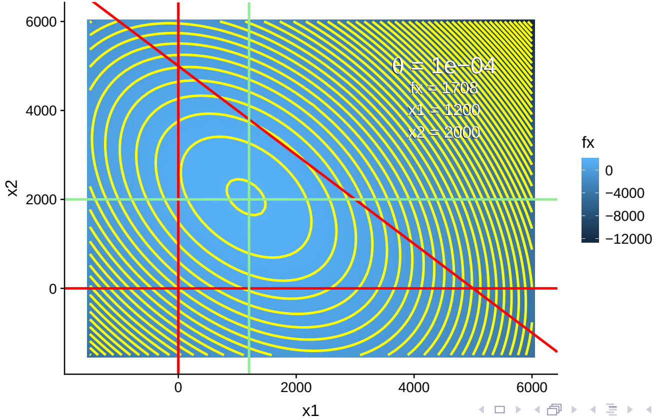
Cuando $\theta = 0{,}00006$, el problema tiene una solución sobre la restricción, pero no en uno de los vértices.



Solución gráfica de un NLP

Ejemplo

Cuando $\theta = 0{,}0001$, el problema tiene una solución en el interior de la región factible, pero no en uno de sus vértices ni sobre las restricciones.



Solución gráfica de un NLP

Definición

Podemos observar que, al resolver gráficamente un NLP la solución óptima puede ubicarse:

1. En un punto interior de la región factible.
2. En un punto de la frontera de la región factible, pero este punto no es un vértice.
3. En un vértice de la región factible.

Observación

En consecuencia, no pueden aplicarse métodos como el simplex para buscar las soluciones en uno de los vértices de la región factible.

Solución de un NLP por fuerza bruta

Ejercicio

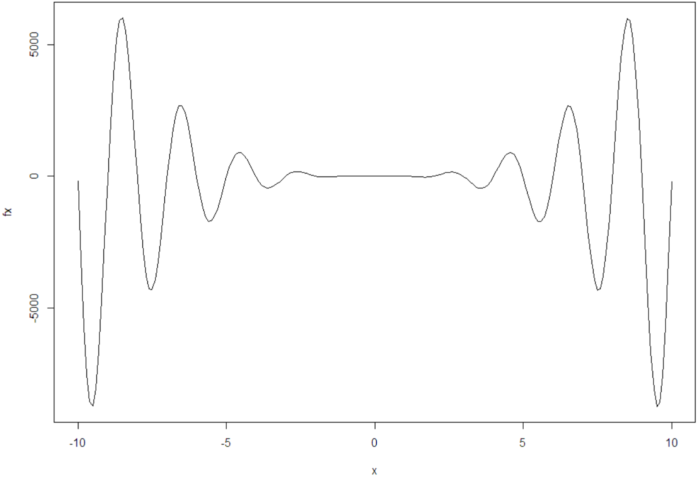
Dados θ y $\Delta x_1 = \Delta x_2 = \Delta x$, realice un algoritmo de fuerza bruta que busque la solución del problema de inversión en el intervalo $x_1, x_2 \in [0, 5]$.

1. Leer θ y $\Delta x_1 = \Delta x_2 = \Delta x$
2. Repita para $x_1 \in [0, 5]$ cada Δx
   2.1 Repita para $x_2 \in [0, 5]$ cada Δx
      2.1.1 Si $x_1 + x_2 \leq 5$ guardar la tupla $\langle \mathbf{x}, f(\mathbf{x}) \rangle$ en una lista
3. $\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmax}} f(\mathbf{x})$
4. Reportar $\langle \mathbf{x}^\star, f(\mathbf{x}^\star) \rangle$

Problema

¿Por qué no guardar el máximo dentro de los ciclos? ¿De qué depende la complejidad de la solución del problema?

Problema de programación no lineal (NLP)



Tipos de soluciones de un NLP

Definición

El conjunto de soluciones:

$$X_f = \{\mathsf{x}_f : \mathsf{x}_f \in \mathbb{R}^n\}$$

se define como el conjunto de soluciones factibles de $f(\mathsf{x})$ si toda solución x_f es factible.

Definición

El conjunto de soluciones factibles:

$$V_{\mathsf{x}^\star} = \{\mathsf{x}_f : d(\mathsf{x}_f, \mathsf{x}^\star) \leq \epsilon\},$$

se define como una vecindad o entorno $V_{\mathsf{x}^\star}$ de $\mathsf{x}^\star$.

Tipos de soluciones de un NLP

Definición

Una solución $\mathbf{x}^\star$ de un NLP se define como:

1. Un máximo global, si $f(\mathbf{x}^\star) \geq f(\mathbf{x}_f)$, $\forall \mathbf{x} \in \mathbb{R}^n$.
2. Un máximo local, si en $\forall \mathbf{x} \in V_{\mathbf{x}^\star}$, $f(\mathbf{x}^\star) \geq f(\mathbf{x})$.
3. Un máximo local estricto, si en $\forall \mathbf{x} \in V_{\mathbf{x}^\star} \backslash \mathbf{x}^\star$, $f(\mathbf{x}^\star) > f(\mathbf{x})$.
4. Un máximo local aislado, si en $\forall \mathbf{x} \in V_{\mathbf{x}^\star}$, $\mathbf{x}^\star$ es el único máximo local.

UNIVERSIDAD CATÓLICA DE TEMUCO

Convexidad y concavidad

Definición

Se dice que una función $f(\mathbf{x})$ es convexa si, para cada $\mathbf{x}_a$ y $\mathbf{x}_b$, y para cada $\lambda \in [0, 1]$;

$$f(\lambda \mathbf{x}_a + (1-\lambda)\mathbf{x}_b) \leq \lambda f(\mathbf{x}_a) + (1-\lambda) f(\mathbf{x}_b).$$

Definición

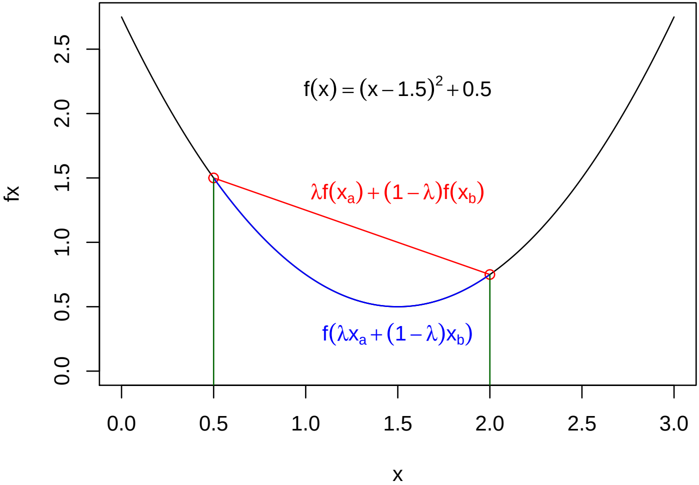
Se dice que una función $f(\mathbf{x})$ es estrictamente convexa si, para cada $\mathbf{x}_a$ y $\mathbf{x}_b$, y para cada $\lambda \in [0, 1]$;

$$f(\lambda \mathbf{x}_a + (1-\lambda)\mathbf{x}_b) < \lambda f(\mathbf{x}_a) + (1-\lambda) f(\mathbf{x}_b).$$

Observación

El lado izquierdo de las definiciones es la evaluación de la función, mientras que el lado derecho es la interpolación lineal entre $\mathbf{x}_a$ y $\mathbf{x}_b$.

Convexidad y concavidad



Convexidad y concavidad

Definición

Se dice que una función $f(\mathbf{x})$ es cóncava si, para cada $\mathbf{x}_a$ y $\mathbf{x}_b$, y para cada $\lambda \in [0, 1]$:

$$f(\lambda \mathbf{x}_a + (1-\lambda)\mathbf{x}_b) \geq \lambda f(\mathbf{x}_a) + (1-\lambda) f(\mathbf{x}_b).$$

Definición

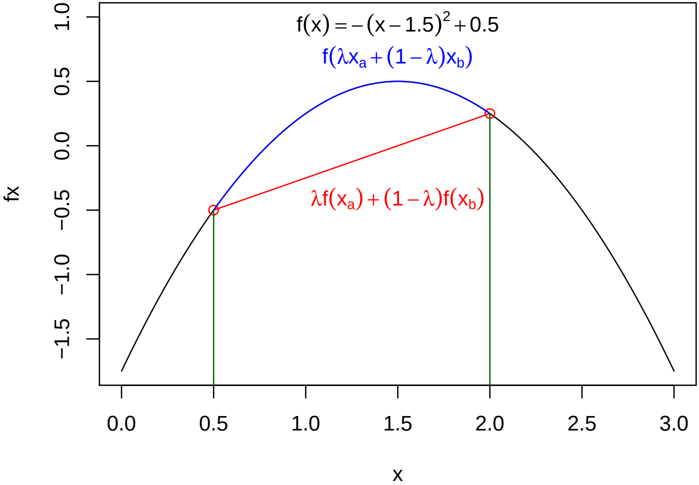
Se dice que una función $f(\mathbf{x})$ es estrictamente concava si, para cada $\mathbf{x}_a$ y $\mathbf{x}_b$, y para cada $\lambda \in [0, 1]$:

$$f(\lambda \mathbf{x}_a + (1-\lambda)\mathbf{x}_b) > \lambda f(\mathbf{x}_a) + (1-\lambda) f(\mathbf{x}_b).$$

Observación

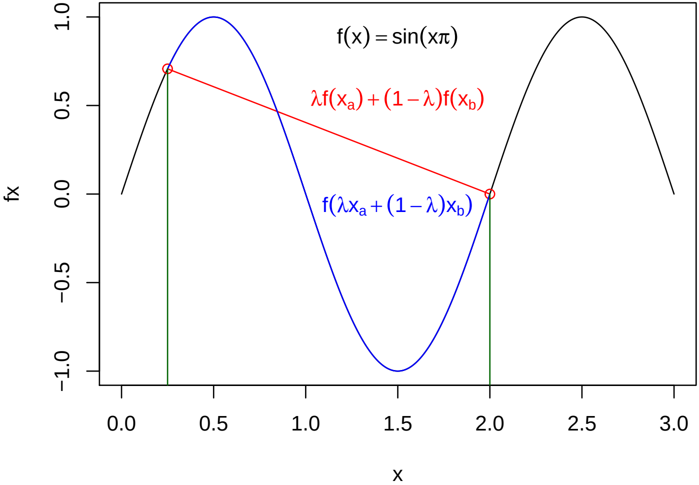
El lado izquierdo de las definiciones es la evaluación de la función, mientras que el lado derecho es la interpolación lineal entre $\mathbf{x}_a$ y $\mathbf{x}_b$.

Convexidad y concavidad



UNIVERSIDAD CATÓLICA DE TEMUCO

Convexidad y concavidad



Convexidad y concavidad

Ejercicio

Dada una función, los puntos x_a y x_b, y el valor de $\Delta\lambda$, realice un programa en el lenguaje de su preferencia que determine si la función es cóncava, convexa o ninguna de las dos en el intervalo $[x_a, x_b]$, utilizando el método descrito anteriormente. Como requerimientos optativos, utilice un *parser* para leer la función y grafíquela, señalando en color azul el valor de la misma y en color rojo el valor de la interpolación, ambos en el intervalo señalado.

Derivada de una función

Definición

Dada $f(\mathbf{x})$, una función continuamente diferenciable en $\mathbf{x}$, la derivada $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$ se define a partir del valor de la secante como:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

Derivada de una función

Ejercicio

Dada una función, el punto x y el valor de Δx, realice un programa en el lenguaje de su preferencia que calcule la derivada $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$.

Ejercicio

Dada una función, la derivada de la función $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$, y un valor de error absoluto ϵ, realice un programa en el lenguaje de su preferencia que calcule el valor de Δx tal que:

$$\left| \frac{\mathrm{d}f(x)}{\mathrm{d}x} - \lim_{\Delta \mathbf{x} \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \right| < \epsilon.$$

UNIVERSIDAD CATÓLICA DE TEMUCO

Derivada de una función

Observación

Si Δx es muy pequeño, el valor de la aproximación tendrá un error muy grande debido al redondeo interno del computador. Por contra, si es muy grande, el valor de la pendiente de la secante será mejor estimado, pero tendrá una diferencia grande con el valor de la tangente. Como regla general, se estima Δx a partir del ϵ de la máquina:

$$\Delta x = \sqrt{\epsilon} \cdot x,$$

donde para un valor de punto flotante de 64 bits, $\epsilon \approx 2{,}22 \times 10^{-16}$.

Gradiente de una función

Definición

Dada $f(\mathbf{x})$, una función continuamente diferenciable en $\mathbf{x}$, el gradiente $\nabla f(\mathbf{x})$ de $f(\mathbf{x})$:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

es el vector de derivadas parciales de $f(\mathbf{x})$ con respecto a cada una de las variables de $\mathbf{x}$, manteniendo constantes las demás.

Gradiente de una función

Ejercicio

Dada una función, el vector $\mathbf{x}$ y el vector $\Delta\mathbf{x}$, realice un programa en el lenguaje de su preferencia que calcule el gradiente $\nabla f(\mathbf{x})$ de $f(\mathbf{x})$.

Ejercicio

Utilice wxMaxima o Maple para calcular el gradiente de la función:

$$f(x, y) = x^2 + y^2 + pxy$$

Matriz hessiana de una función

Definición

Dada $f(\mathbf{x})$, una función continuamente diferenciable dos veces en $\mathbf{x}$, la matriz hessiana $\nabla^2 f(\mathbf{x})$ de $f(\mathbf{x})$:

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

es la matriz de derivadas parciales de $f(\mathbf{x})$ con respecto a dos variables de $\mathbf{x}$ a la vez o una de las variables dos veces, manteniendo constantes las demás.

Teorema de Taylor

Teorema

Suponga que la función $f : \mathbb{R}^n \to \mathbb{R}$ *es continuamente diferenciable y que* $\Delta \mathsf{x} \in \mathbb{R}^n$. *Entonces tenemos que:*

$$f(\mathsf{x} + \Delta \mathsf{x}) = f(\mathsf{x}) + \nabla f(\mathsf{x} + t\Delta \mathsf{x})^T \Delta \mathsf{x},$$

para cualquier $t \in (0, 1)$. *Aún más, si* f *es dos veces continuamente diferenciable, tenemos que:*

$$\nabla f(\mathsf{x} + \Delta \mathsf{x}) = \nabla f(\mathsf{x}) + \int_0^1 \nabla^2 f(\mathsf{x} + t\Delta \mathsf{x})\Delta \mathsf{x}\, dt$$

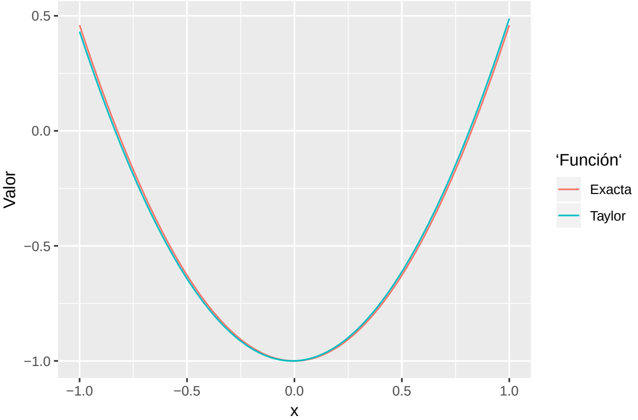
y por tanto:

$$f(\mathsf{x} + \Delta \mathsf{x}) = f(\mathsf{x}) + \nabla f(\mathsf{x})^T \Delta \mathsf{x} + \frac{1}{2}\Delta \mathsf{x}^T \nabla^2 f(\mathsf{x} + t\Delta \mathsf{x})\Delta \mathsf{x}.$$

Teorema de Taylor

Observación

El Teorema de Taylor es una aproximación de la función $f(\mathbf{x})$ en el entorno $V_{\mathbf{x}}$.



Problema

Demuestre el teorema de Taylor para la función $f(x) = x^2 - \cos(x)$ en $x = 0$ con $\Delta x = 0{,}01$ y $t = 1$.

Teorema de Taylor

Solución

$\nabla f(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x} = 2x + \sin(x)$ y $\nabla^2 f(x) = \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} = 2 + \cos(x)$

$$
\begin{aligned}
f(\mathbf{x} + \Delta\mathbf{x}) =& f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x} + t\Delta\mathbf{x})\Delta\mathbf{x}. \\
(x+\Delta x)^2 - \cos(x+\Delta x) =& x^2 - \cos(x) + (2x + \sin(x))\Delta x + \\
& \frac{1}{2}\Delta x(2 + \cos(x+\Delta x))\Delta x \\
-0{,}99985 =& -0{,}99985
\end{aligned}
$$

Teorema de Taylor

Problema

Demuestre el teorema de Taylor para la función $f(\mathbf{x}) = 1{,}20x_1 + 1{,}16x_2 - \frac{1}{2}\left(2x_1^2 + x_2^2 + (x_1 + x_2)^2\right)$ en $\mathbf{x}^T = [0{,}23, 0{,}41]$ con $\Delta\mathbf{x}^T = [0{,}01, 0{,}01]$ y $t = 0{,}01$.

Problema

Realice un programa que grafique una función univariante dada (en rojo) y su aproximación mediante el teorema de Taylor (en azul), dadas la función, sus primeras dos derivadas, un intervalo $[x_a, x_b]$ y los valores de Δx y t.

Condiciones necesarias de primer orden

Teorema

Si $\mathbf{x}^\star$ *es un mínimo local y* $f(\mathbf{x})$ *es continuamente diferenciable en* $V_{\mathbf{x}^\star}$*, entonces* $\nabla f(\mathbf{x}^\star) = 0$.

Demostración.

Suponga, por contradicción, que $\nabla f(\mathbf{x}^\star) \neq 0$. Defina el vector $\Delta\mathbf{x} = -\nabla f(\mathbf{x}^\star)$ y note que

$$\Delta\mathbf{x}^T \nabla f(\mathbf{x}^\star) = -\nabla f(\mathbf{x}^\star)^T \nabla f(\mathbf{x}^\star) = -\left\|\nabla f(\mathbf{x}^\star)\right\|^2 < 0.$$

Debido a que $\Delta f(\mathbf{x})$ es continua en $\mathbf{x}^\star$, existe un escalar $\tau > 0$ tal que:

$$\Delta\mathbf{x}^T \nabla f(\mathbf{x}^\star + t\Delta\mathbf{x}) < 0, \ \forall t \in (0, \tau].$$

Para cualquier $\bar{t} \in (0, \tau]$, tenemos por el teorema de Taylor que:

$$f(\mathbf{x}^\star + \bar{t}\Delta\mathbf{x}) = f(\mathbf{x}^\star) + \bar{t}\Delta\mathbf{x}^T \nabla f(\mathbf{x}^\star + t\Delta\mathbf{x}), \ \forall t \in (0, \tau].$$

Por lo tanto, $f(\mathbf{x}^\star + \bar{t}\Delta\mathbf{x}) < f(\mathbf{x}^\star)$, $\forall \bar{t} \in (0, \tau)$. Hemos encontrado una dirección que se aleja de $\mathbf{x}^\star$ sobre la que $f(\mathbf{x})$ disminuye, por lo que $\mathbf{x}^\star$ no es el óptimo local, lo que es contradictorio. $\square$

Valores propios y definición de una matriz

Definición

Sea $\mathbf{M} = \left[m_{i,j}\right]_{n\times n}$, $\mathbf{v} = [v_i]_n$ y λ es un escalar que satisface $\mathbf{M}\mathbf{v} = \lambda\mathbf{v}$, entonces λ es llamado el valor propio asociado con el vector propio $\mathbf{v}$ de $\mathbf{M}$. Los valores propios de $\mathbf{M}$ son las raíces de la ecuación característica $\det\left(\mathbf{M} - \lambda\mathbf{I}\right) = 0$, donde $\mathbf{I}$ es la matriz unitaria de tamaño $n \times n$.

Definición

Una matriz $\mathbf{M} = \left[m_{i,j}\right]_{n\times n}$ es:

1. Semidefinida positiva si y solo si todos sus valores propios $\lambda \geq 0$.
2. Semidefinida negativa si y solo si todos sus valores propios $\lambda \leq 0$.
3. Indefinida si y solo si algunos valores propios son positivos, mientras el resto son negativos.

Valores propios y definición de una matriz

Definición

Una matriz $\mathbf{M} = \left[m_{i,j}\right]_{n\times n}$ es:

1. Semidefinida positiva si y solo si $\mathbf{x}^T\mathbf{M}\mathbf{x} \geq 0,\ \forall\mathbf{x} \in \mathbb{R}\backslash 0$.
2. Semidefinida negativa si y solo si $\mathbf{x}^T\mathbf{M}\mathbf{x} \leq 0,\ \forall\mathbf{x} \in \mathbb{R}\backslash 0$.
3. Indefinida si y solo si $\mathbf{x}^T\mathbf{M}\mathbf{x} = 0,\ \forall\mathbf{x} \in \mathbb{R}\backslash 0$.

Condiciones necesarias de segundo orden

Teorema

Si $\mathbf{x}^\star$ es un mínimo local de $f(\mathbf{x})$ y $\nabla^2 f(\mathbf{x})$ existe y es continua en $V_{\mathbf{x}^\star}$, entonces $\nabla f(\mathbf{x}^\star) = 0$ y $\nabla^2 f(\mathbf{x}^\star)$ es semidefinida positiva.

Demostración.

Sabemos por el Teorema anterior que $\nabla f(\mathbf{x}^\star) = 0$. Por contradicción, asuma que $\nabla^2 f(\mathbf{x}^\star)$ no es semidefinida positiva. Entonces, podemos escoger un vector $\Delta\mathbf{x}$ tal que $\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x}^\star)\Delta\mathbf{x} < 0$, y, dado que $\nabla^2 f(\mathbf{x})$ es continua en $V(\mathbf{x}^\star)$, habrá un escalar $\tau > 0$ tal que $\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x}^\star + t\Delta\mathbf{x})\Delta\mathbf{x} < 0$, $\forall t \in [0, T]$. Al realizar una expansión por serie de Taylor en $V(\mathbf{x}^\star)$, tenemos que para todo $\bar{t} \in (0, T]$ y algún $t \in (0, \bar{t})$:

$$f(\mathbf{x}^\star + \bar{t}\Delta\mathbf{x}) = f(\mathbf{x}^\star) + \bar{t}\Delta\mathbf{x}^T \nabla f(\mathbf{x}^\star) + \frac{1}{2}\bar{t}^2 \Delta\mathbf{x}^T \nabla^2 f(\mathbf{x}^\star + t\Delta\mathbf{x})\Delta\mathbf{x} < f(\mathbf{x}^\star).$$

Al igual que en el Teorema anterior, hemos encontrado una dirección desde $\mathbf{x}^\star$ sobre la que $f(\mathbf{x})$ es decreciente, y, nuevamente, $\mathbf{x}^\star$ no es un optimo local.

$\square$

Condiciones suficientes de segundo orden

Teorema

Suponga que $\nabla^2 f(\mathbf{x})$ *es continua en* $V_{\mathbf{x}^\star}$*, que* $\nabla f(\mathbf{x}^\star) = 0$ *y que* $\nabla^2 f(\mathbf{x}^\star)$ *es definida positiva, entonces* $\mathbf{x}^\star$ *es un mínimo local estricto de* $f(\mathbf{x})$*.*

Demostración.

Debido a $\nabla^2 f(\mathbf{x})$ es continua y definida positiva en $V_{\mathbf{x}^\star}$, podemos escoger un radio $r > 0$, tal que $\nabla^2 f(\mathbf{x})$ siga siendo definida positiva para todos los $\mathbf{x}$ en la bola abierta $\mathcal{D} = \{\mathbf{z} : \|\mathbf{z} - \mathbf{x}\| < r\}$. Tomando cualquier vector $\Delta\mathbf{x}$ con $\|\Delta\mathbf{x}\| < r$, tenemos que $\mathbf{x}^\star + \Delta\mathbf{x} \in \mathcal{D}$ y:

$$\begin{aligned} f(\mathbf{x}^\star + \Delta\mathbf{x}) &= f(\mathbf{x}^\star) + \Delta\mathbf{x}^T \nabla f(\mathbf{x}^\star) + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{z})\Delta\mathbf{x} \\ &= f(\mathbf{x}^\star) + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{z})\Delta\mathbf{x}, \end{aligned}$$

donde $\mathbf{z} = \mathbf{x}^\star + t\Delta\mathbf{x}$ para algún $t \in [0, 1]$. Dado que $\mathbf{z} \in \mathcal{D}$, tenemos que $\Delta\mathbf{x}^T \nabla^2 f(\mathbf{z})\Delta\mathbf{x} > 0$, y, por lo tanto, $f(\mathbf{x}^\star + \Delta\mathbf{x}) > f(\mathbf{x}^\star)$.

□

Condiciones suficientes de segundo orden

Observación

Las condiciones suficientes de segundo orden garantizan que el mínimo sea un mínimo local estricto.

Observación

Note, además, que las condiciones suficientes de segundo orden no son necesarias. Un punto $\mathbf{x}^\star$ puede ser un mínimo local estricto y, sin embargo, puede no satisfacer las condiciones suficientes.

Ejemplo

La función $f(x) = x^4$ tiene un mínimo local estricto en $\mathbf{x}^\star = 0$. No obstante, en ese punto, la matriz Hessiana se desvanece (se iguala a cero) y, por lo tanto, no es definida positiva.

Matriz hessiana de una función: óptimo local

Definición

Dado $\nabla^2 f(\mathbf{x})$, el criterio $H(\mathbf{x})$ para una función bivariante $f(\mathbf{x})$ se define como:

$$H(\mathbf{x}) = \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} \cdot \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} - \left(\frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2}\right)^2 .$$

Por tanto:

1. Si $H(\mathbf{x}) < 0$, entonces el punto $\mathbf{x}$ es un punto de silla.
2. Si $H(\mathbf{x}) > 0$ y $\frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} < 0$, entonces el punto $\mathbf{x}$ es un máximo local.
3. Si $H(\mathbf{x}) > 0$ y $\frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} > 0$, entonces el punto $\mathbf{x}$ es un mínimo local.
4. Si $H(\mathbf{x}) = 0$, no existe información suficiente sobre el punto $\mathbf{x}$.

Matriz hessiana de una función: óptimo local

Ejercicio

Dada la función bivariante $f(\mathbf{x}) = x_1^2 + 4x_2^2 - 2x_1^2 x_2 + 4$,encuentre los puntos críticos en el intervalo $x_1, x_2 \in [0, 1]$.

Matriz hessiana de una función: óptimo local

Solución

Se debe resolver el gradiente igualado a cero:

$$\nabla \mathbf{x} = \begin{bmatrix} 2x_1 - 4x_1x_2 \\ 8x_2 - 2x_1^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Utilizando un solver encontramos que las soluciones son:

$$\mathbf{x}_1^\star = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \mathbf{x}_2^\star = \begin{bmatrix} -\sqrt{2} \\ \frac{1}{2} \end{bmatrix}; \mathbf{x}_3^\star = \begin{bmatrix} \sqrt{2} \\ \frac{1}{2} \end{bmatrix}$$

Por lo que en el intervalo deseado la única solución posible es $\mathbf{x}_1^\star$

Matriz hessiana de una función: óptimo local

Ejercicio

Dada la función bivariante $f(\mathbf{x})$, los valores de $\Delta\mathbf{x}$, y el punto $\mathbf{x}$, determine, programe en el lenguaje de su preferencia una función que mediante el criterio $H(\mathbf{x})$ determine si el punto $\mathbf{x}$ es un mínimo o máximo local, un punto de silla, o si no hay información suficiente.

Óptimo global

Teorema

Cuando $f(\mathbf{x})$ es convexa, cualquier mínimo local $\mathbf{x}^\star$ es un mínimo global de $f(\mathbf{x})$.

Demostración.

Suponga que $\mathbf{x}^\star$ es un mínimo local, pero no uno global. Por lo tanto, podemos encontrar un punto $\mathbf{z} \in \mathbb{R}^n$ para el que $f(\mathbf{z}) < f(\mathbf{x}^\star)$. Considere el segmento que une $\mathbf{x}^\star$ a $\mathbf{z}$, es decir:

$$\mathbf{x} = \lambda\mathbf{z} + (1-\lambda)\mathbf{x}^\star, \ \forall\lambda \in (0,1].$$

Por la propiedad de convexidad de $f(\mathbf{x})$, tenemos que:

$$f(\mathbf{x}) \leq \lambda f(\mathbf{z}) + (1-\lambda) f(\mathbf{x}^\star) < f(\mathbf{x}^\star).$$

Cualquier vecindad $V_{\mathbf{x}^\star}$ contiene un trozo del segmento $\mathbf{x}$, así que cualquier $\mathbf{x} \in V_{\mathbf{x}^\star}$ satisface la desigualdad y, por lo tanto, $\mathbf{x}^\star$ no es un mínimo local, por lo que se contradice el supuesto y $\mathbf{x}^\star$ es un óptimo global.

□

Óptimo global

Teorema

Si $f(\mathbf{x})$ es es convexa y diferenciable, cualquier punto estacionario $\mathbf{x}^\star$ ($\nabla f(\mathbf{x}^\star) = 0$) es un mínimo global de $f(\mathbf{x})$.

Demostración.

Suponga que $\mathbf{x}^\star$ no es un mínimo global. Dado un punto $\mathbf{z} \in \mathbb{R}^n$ para el que $f(\mathbf{z}) < f(\mathbf{x}^\star)$, considere el segmento que une $\mathbf{x}^\star$ a $\mathbf{z}$, es decir:

$$\mathbf{x} = \lambda \mathbf{z} + (1 - \lambda)\mathbf{x}^\star, \ \forall \lambda \in (0, 1].$$

Por la propiedad de convexidad de $f(\mathbf{x})$, tenemos que:

$$\begin{aligned}
\nabla f(\mathbf{x}^\star)^T(\mathbf{z} - \mathbf{x}^\star) &= \left.\frac{\mathrm{d}}{\mathrm{d}\lambda} f(\mathbf{x}^\star + \lambda(\mathbf{z} - \mathbf{x}^\star))\right|_{\lambda=0} \\
&= \lim_{\lambda \to 0} \frac{f(\mathbf{x}^\star + \lambda(\mathbf{z} - \mathbf{x}^\star)) - f(\mathbf{x}^\star)}{\lambda} \\
&\leq \lim_{\lambda \to 0} \frac{\lambda f(\mathbf{z}) + (1 - \lambda) f(\mathbf{x}^\star) - f(\mathbf{x}^\star)}{\lambda} \\
&= f(\mathbf{z}) - f(\mathbf{x}^\star) < 0
\end{aligned}$$

Por lo tanto, $\nabla f(\mathbf{x}^\star) \neq 0$ y $\mathbf{x}^\star$ no es un punto estacionario.

$\square$

Algoritmo del gradiente descendente: t fijo

Definición

Es un algoritmo que utiliza la dirección del gradiente para encontrar puntos óptimos de la función objetivo.

1. Dado un punto $\mathbf{x}$ en el dominio de $f(\mathbf{x})$ y un escalar t
2. Repita mientras $\|\nabla f(\mathbf{x}_n)\|_2 \geq \eta$
   2.1 $\Delta\mathbf{x}_n := -\nabla f(\mathbf{x}_n)$
   2.2 $\mathbf{x}_{n+1} := \mathbf{x}_n + t\Delta\mathbf{x}_n$

Ejercicio

Implemente el algoritmo para funciones bivariantes utilizando el lenguaje de su preferencia y librerías matemáticas que incorporen derivadas numéricas o el cálculo del gradiente.

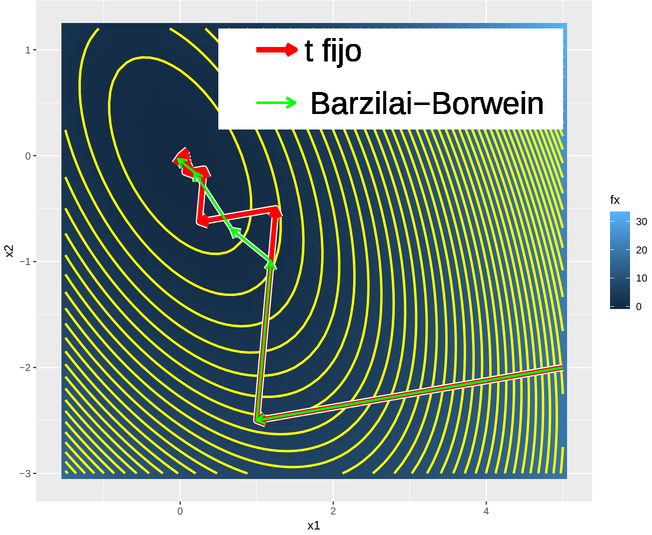
Algoritmo del gradiente descendente: actualización de t por el método de Barzilai-Borwein

1. Dado un punto $\mathbf{x}$ en el dominio de $f(\mathbf{x})$ y un escalar t_1
2. Repita mientras $\|\nabla f(\mathbf{x}_n)\|_2 \geq \eta$
   2.1 $\Delta \mathbf{x}_n := -\nabla f(\mathbf{x}_n)$
   2.2 $t_n := \frac{\left|(\mathbf{x}_n - \mathbf{x}_{n-1})^T [\nabla f(\mathbf{x}_n) - \nabla f(\mathbf{x}_{n-1})]\right|}{\|\nabla f(\mathbf{x}_n) - \nabla f(\mathbf{x}_{n-1})\|^2}$
   2.3 $\mathbf{x}_{n+1} := \mathbf{x}_n + t_n \Delta \mathbf{x}_n$

Ejercicio

Implemente el algoritmo para funciones bivariantes utilizando el lenguaje de su preferencia y librerías matemáticas que implementen derivadas numéricas o el cálculo del gradiente.

Algoritmo del gradiente descendente



Método de Newton

1. Dado un punto $\mathbf{x}$ en el dominio de $f(\mathbf{x})$, un escalar $t > 0$ y una cota de error $\epsilon > 0$
2. Repita
   2.1 $\mathbf{x}_{n+1} := \mathbf{x}_n - t\left(\nabla^2 \mathbf{x}_n\right)^{-1} \nabla \mathbf{x}$
   2.2 Si $\|\nabla \mathbf{x}_{n+1}\| < \|\nabla \mathbf{x}_{n+1}\|$
      2.2.1 $g(0) := \|\nabla \mathbf{x}_{n+1}\|$
      2.2.2 $g'(0) := -\frac{\|\nabla \mathbf{x}_n\|}{\|\nabla \mathbf{x}_{n+1} - \nabla \mathbf{x}_n\|}$
      2.2.3 $g(1) := \|\nabla \mathbf{x}_{n+1}\|$
      2.2.4 $t' := \text{máx}\left(\frac{-g'(0)}{2 \cdot [g(1) - g(0) - g'(0)]}, 0{,}1\right)$
      2.2.5 $\mathbf{x}_{n+1} = [1 - t']\,\mathbf{x}_n + t'\mathbf{x}_{n+1}$
3. Hasta que $\|\mathbf{x}_{n+1} - \mathbf{x}_n\| \leq \epsilon$

Ejercicio

Implemente el algoritmo para funciones univariantes utilizando el lenguaje de su preferencia y librerías matemáticas para el cálculo de derivadas numéricas o explícitas. ¿Cuál es el problema más importante en el algoritmo para el método de Newton con múltiples variables?

Multiplicadores de Lagrange

Definición

Es un método que utiliza el gradiente para determinar puntos críticos de funciones objetivos no lineales sujetas a restricciones no lineales. Los pasos que utiliza son los siguientes:

1. Dada una funcion $f(\mathsf{x})$ y una restricción $g(\mathsf{x}) = c$, defina la función de Lagrange como:

$$\mathcal{L}(\mathsf{x}, \lambda) = f(\mathsf{x}) - \lambda(g(\mathsf{x}) - c),$$

donde λ es denominado como el "multiplicador de Lagrange".

2. Resuelva:

$$\nabla\mathcal{L}(\mathsf{x}, \lambda) = 0$$

Multiplicadores de Lagrange

Problema

Dada una lámina de cartón de área c, determine el volumen máximo que puede tener una caja construida con la misma.

Solución

El modelo sería:
Maximizar $f(x, y, z) = x \cdot y \cdot z$
Sujeto a:
$2(x \cdot y + x \cdot z + y \cdot z) = c$,
por tanto:

$$\mathcal{L}(x, y, z, \lambda) = x \cdot y \cdot z + \lambda \cdot \left(x \cdot y + x \cdot z + y \cdot z - \frac{c}{2}\right)$$

UNIVERSIDAD CATÓLICA DE TEMUCO

Multiplicadores de Lagrange

Problema

Dada una lámina de cartón de área c, determine el volumen máximo que puede tener una caja construida con la misma.

Solución

$$\mathcal{L}(x, y, z, \lambda) = xyz - \lambda \cdot (2xy + 2xz + 2yz - c)$$

$$\nabla \mathcal{L}(x, y, z, \lambda) = \begin{bmatrix} yz - \lambda(2y + 2z) \\ xz - \lambda(2x + 2z) \\ xy - \lambda(2x + 2y) \\ 2xy + 2xz + 2yz - c \end{bmatrix}$$

Multiplicadores de Lagrange

Problema

Dada una lámina de cartón de área c, determine el volumen máximo que puede tener una caja construida con la misma.

Solución

$$
\begin{aligned}
yz - \lambda(2y + 2z) &= 0 \\
xz - \lambda(2x + 2z) &= 0 \\
xy - \lambda(2x + 2y) &= 0 \\
2xy + 2xz + 2yz - c &= 0
\end{aligned}
$$

La solución es:

$$\lambda = \frac{\sqrt{6c}}{24} \; ; \; x = y = z = \frac{\sqrt{6c}}{6} \; ; \; f(x, y, z) = \frac{c^{\frac{3}{2}}\sqrt{6}}{36}$$

Multiplicadores de Lagrange

(%i38) to_poly_solve([y·z−l·(2·y+2·z)=0, x·z−l·(2·x+2·z)=0, x·y−l·(2·x+2·y)=0, 2·x·y+2·x·z+2·y·z=c], [x,y,z,l]);

(%o38) $\%union\left(\left[l=-\frac{\sqrt{c}}{4\sqrt{6}},x=-\frac{\sqrt{c}}{\sqrt{6}},y=-\frac{\sqrt{c}}{\sqrt{6}},z=-\frac{\sqrt{c}}{\sqrt{6}}\right],\left[l=\frac{\sqrt{c}}{4\sqrt{6}},x=\frac{\sqrt{c}}{\sqrt{6}},y=\frac{\sqrt{c}}{\sqrt{6}},z=\frac{\sqrt{c}}{\sqrt{6}}\right]\right)$

Multiplicadores de Lagrange

WolframAlpha computational intelligence.

y*z-2*l*y-2*l*z=0,x*z-2*l*x-2*l*z=0,x*y-2*l*x-2*l*y=0,2*x*y+2*x*z+2*y*z=c,c>0,x>0,y>0,z>0

Browse Examples　Surprise Me

Input:

$$\{y\,z-2\,l\,y-2\,l\,z=0,\ x\,z-2\,l\,x-2\,l\,z=0,\ x\,y-2\,l\,x-2\,l\,y=0,\ 2\,x\,y+2\,x\,z+2\,y\,z=c,\ c>0,\ x>0,\ y>0,\ z>0\}$$

Real solution:

Approximate form

$$c>0,\quad l=\frac{\sqrt{c}}{4\sqrt{6}},\quad x=\frac{\sqrt{c}}{\sqrt{6}},\quad y=\frac{\sqrt{c}}{\sqrt{6}},\quad z=\frac{\sqrt{c}}{\sqrt{6}}$$

UNIVERSIDAD CATÓLICA DE TEMUCO

Multiplicadores de Lagrange

$solve(\{y \cdot z - \text{lambda} \cdot (2 \cdot y + 2 \cdot z) = 0, x \cdot z - \text{lambda} \cdot (2 \cdot x + 2 \cdot z) = 0, x \cdot y - \text{lambda} \cdot (2 \cdot x + 2 \cdot y) = 0, 2 \cdot x \cdot y + 2 \cdot x \cdot z + 2 \cdot y \cdot z = c\}, \{x, y, z, \text{lambda}\})$

$$\left\{\lambda = \frac{c}{6\left(\frac{c}{3\,RootOf(6\_Z^2 - c)} + 2\,RootOf(6\_Z^2 - c)\right)}, x = RootOf(6\_Z^2 - c), y = \frac{c}{6\,RootOf(6\_Z^2 - c)}, z = \frac{2c}{3\left(\frac{c}{3\,RootOf(6\_Z^2 - c)} + 2\,RootOf(6\_Z^2 - c)\right)}\right\}$$

$allvalues(RootOf(6\_Z^2 - c))$

$$\frac{\sqrt{6}\sqrt{c}}{6}, -\frac{\sqrt{6}\sqrt{c}}{6}$$

$$\frac{c}{6\,\frac{\sqrt{6}\sqrt{c}}{6}}$$

$$\frac{\sqrt{6}\sqrt{c}}{6}$$

$$\frac{2c}{3\left(\frac{c}{3\,\frac{\sqrt{6}\sqrt{c}}{6}} + 2\,\frac{\sqrt{6}\sqrt{c}}{6}\right)}$$

$$\frac{\sqrt{6}\sqrt{c}}{6}$$

$$\frac{c}{6\left(\frac{c}{3\,\frac{\sqrt{6}\sqrt{c}}{6}} + 2\,\frac{\sqrt{6}\sqrt{c}}{6}\right)}$$

$$\frac{\sqrt{6}\sqrt{c}}{24}$$

$$\left(\frac{\sqrt{6}\sqrt{c}}{6}\right)^3$$

$$\frac{\sqrt{6}\,c^{3/2}}{36}$$

$$\xrightarrow{\text{differentiate w.r.t. c}}$$

$$\frac{\sqrt{6}\sqrt{c}}{24}$$

UNIVERSIDAD CATÓLICA DE TEMUCO

Interpretación de los multiplicadores de Lagrange

Definición

El multiplicador de Lagrange λ es el valor marginal asociado a la relajación de una restricción, es decir, es el precio sombra de la restricción:

$$\lambda = \frac{\partial f(\mathbf{x}^{\star})}{\partial c}$$

Ejemplo

En el problema de la caja, $\lambda = \frac{\sqrt{6c}}{24}$, es decir, que por cada unidad que se aumente la superficie de la lámina de cartón, el volumen de la caja aumentará en $\frac{\sqrt{6c}}{24}$.

Multiplicadores de Lagrange: múltiples restricciones

Definición

Para múltiples restricciones, el método de los multiplicadores de Lagrange resuelve el siguiente sistema de ecuaciones

$$\nabla f(\mathbf{x}) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\mathbf{x}) = 0$$

$$g_j(\mathbf{x}) - c_j = 0, \forall j = 1, \ldots, m.$$

Multiplicadores de Lagrange: múltiples restricciones

Problema

Dada una lámina de cartón de área c, determine el volumen máximo que puede tener una caja construida con la misma con una base igual a $xy = b$.

Solución

El modelo sería:
Maximizar $f(x, y, z) = xyz$
Sujeto a:
$2(xy + xz + yz) = c$
$xy = b$

Multiplicadores de Lagrange: múltiples restricciones

Problema

Dada una lámina de cartón de área c, determine el volumen máximo que puede tener una caja construida con la misma con una base igual a $xy = b$.

Solución

$$
\begin{aligned}
yz - \lambda_1 (2y + 2z) - \lambda_2 y &= 0 \\
xz - \lambda_1 (2x + 2z) - \lambda_2 x &= 0 \\
xy - \lambda_1 (2x + 2y) &= 0 \\
2xy + 2xz + 2yz - c &= 0 \\
xy - b &= 0
\end{aligned}
$$

Multiplicadores de Lagrange: múltiples restricciones

(%i39) to_poly_solve([y·z−l1·(2·y+2·z)−l2·y=0, x·z−l1·(2·x+2·z)−l2·x=0, x·y−l1·(2·x+2·y)=0, 2·x·y+2·x·z+2·y·z=c,x·y=b], [x,y,z,l1,l2]);

(%o39) $\%union\left(\left[l1=-\frac{\sqrt{b}}{4},l2=-\frac{c-6\,b}{8\sqrt{b}},x=-\sqrt{b},y=-\sqrt{b},z=-\frac{c-2\,b}{4\sqrt{b}}\right],\left[l1=\frac{\sqrt{b}}{4},l2=\frac{c-6\,b}{8\sqrt{b}},x=\sqrt{b},y=\sqrt{b},z=\frac{c-2\,b}{4\sqrt{b}}\right]\right)$

Multiplicadores de Lagrange: múltiples restricciones

WolframAlpha computational intelligence.

y*z-2*l1*y-2*l1*z-l2*y=0,x*z-2*l1*x-2*l1*z-l2*x=0,x*y-2*l1*x-2*l1*y=0,2*x*y+2*x*z+2*y*z=

Browse Examples   Surprise Me

Input:

$$\{y\,z - 2\,l1\,y - 2\,l1\,z - l2\,y = 0,\ x\,z - 2\,l1\,x - 2\,l1\,z - l2\,x = 0,\ x\,y - 2\,l1\,x - 2\,l1\,y = 0,\ 2\,x\,y + 2\,x\,z + 2\,y\,z = c,\ x\,y = b,\ b > 0,\ c > 0,\ x > 0,\ y > 0,\ z > 0\}$$

Open code

Alternate forms:

$$\{y\,(z - l2) = 2\,l1\,(y + z),\ x\,(z - l2) = 2\,l1\,(x + z),\ x\,y = 2\,l1\,(x + y),\ c = 2\,(y\,z + x\,(y + z)),\ b = x\,y,\ b > 0,\ c > 0,\ x > 0,\ y > 0,\ z > 0\}$$

$$\{z\,(y - 2\,l1) = (2\,l1 + l2)\,y,\ z\,(x - 2\,l1) = (2\,l1 + l2)\,x,\ y\,(x - 2\,l1) = 2\,l1\,x,\ c = 2\,y\,z + x\,(2\,y + 2\,z),\ b = x\,y,\ b > 0,\ c > 0,\ x > 0,\ y > 0,\ z > 0\}$$

Real solution:

$$b > 0, \quad c > 2\,b, \quad l1 = \frac{\sqrt{b}}{4}, \quad l2 = \frac{c - 6\,b}{8\,\sqrt{b}}, \quad x = \sqrt{b}, \quad y = \sqrt{b}, \quad z = \frac{c - 2\,b}{4\,\sqrt{b}}$$

Download Page

POWERED BY THE WOLFRAM LANGUAGE

UNIVERSIDAD CATÓLICA DE TEMUCO

Librería NLOpt

Ejercicio

Mediante el lenguaje de su preferencia, utilizar la librería NLOpt para resolver el problema de la caja de cartón con una y dos restricciones, asumiendo un valor de $c > 2b$ para el segundo caso.