Optimización II. Capítulo 12. "Teoría de la Optimización con Restricciones"

Oscar S. Dalmau Cedeño
dalmau@cimat.mx

19 de septiembre de 2018

Cono Tangente

Cono Tangente

Sea $C \subseteq \mathbb{R}^n$ un conjunto no vacío y $\boldsymbol{x} \in C$. Un **cono tangente** al conjunto C en el punto x es el conjunto

$$T_C(\boldsymbol{x}) = \{\boldsymbol{d} \in \mathbb{R}^n |\ d = \lim_{k\to\infty} \frac{\boldsymbol{z}_k - \boldsymbol{x}}{t_k};\ \boldsymbol{z}_k \in C\}$$

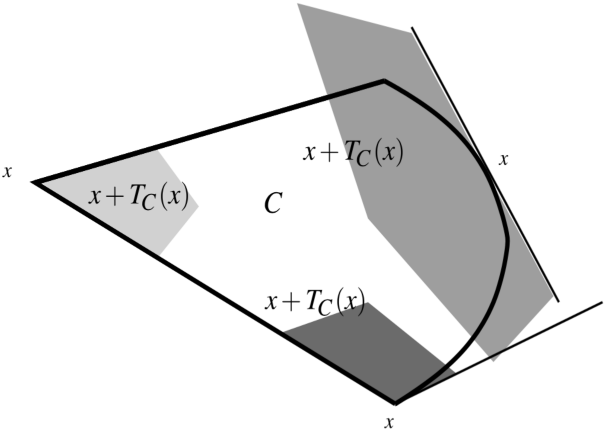
donde $\lim_{k\to\infty} \boldsymbol{z}_k = \boldsymbol{x}$, $\lim_{k\to\infty} t_k = 0$ y $t_k \geq 0$.

A los vectores de este conjunto se llaman **vectores tangentes** al conjunto C en $\boldsymbol{x}$.

Cono Tangente

- Si $\boldsymbol{x} \in Interior(C)$, entonces $N_C(\boldsymbol{x}^*) = \{0\}$ y $T_C(\boldsymbol{x}^*) = \mathbb{R}^n$.
- Se puede verificar que $T_\Omega(\boldsymbol{x})$ es un cono
- A partir de ahora usaremos como C al conjunto factible Ω, ie $T_\Omega(\boldsymbol{x})$.

Cono Tangente



Direcciones factibles linearizadas

Direcciones factibles linearizadas

Dado un punto factible $\boldsymbol{x} \in \Omega$ y el conjunto de restricciones activas $\mathcal{A}(\boldsymbol{x})$. El **conjunto de direcciones factibles linearizadas** $\mathcal{F}(\boldsymbol{x})$ es el conjunto

$$\mathcal{F}(\boldsymbol{x}^*) = \left\{ \boldsymbol{d} \in \mathbb{R}^n \middle| \begin{array}{ll} \boldsymbol{x}^T \nabla c_i(\boldsymbol{x}^*) = 0 & \text{si } i \in \mathcal{E} \text{ y} \\ \boldsymbol{d}^T \nabla c_i(\boldsymbol{x}^*) \geq 0 & \text{si } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} \end{array} \right\}.$$

- Es decir, es el conjunto de las direcciones factibles (condición de factibilidad) en las restricciones activas $\mathcal{A}(\boldsymbol{x})$.
- Se puede verificar que $\mathcal{F}(\boldsymbol{x})$ es un cono
- Si $\mathcal{A}(\boldsymbol{x}^*) = \emptyset$ entonces $\mathcal{F}(\boldsymbol{x}^*) = \mathbb{R}^n$

Comentario

- La definición del cono tangente $T_\Omega(\boldsymbol{x})$ depende de la geometria de Ω.
- La definición de las **direcciones factibles linearizadas** $\mathcal{F}(\boldsymbol{x})$ depende de las expresiones algebraicas de las restricciones.

Relación entre el cono tangente y las direcciones factibles linearizadas

Lema

Sea $\boldsymbol{x}^*$ un punto factible, entonces

1. $T_\Omega(\boldsymbol{x}^*) \subset \mathcal{F}(\boldsymbol{x}^*)$
2. Si se cumple la condición LICQ en $\boldsymbol{x}^*$ entonces $T_\Omega(\boldsymbol{x}^*) = \mathcal{F}(\boldsymbol{x}^*)$

Prueba de la segunda parte

Sea $\boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*)$, i.e., $\boldsymbol{d}^T \nabla c_i(\boldsymbol{x}^*) = 0$ para $i \in \mathcal{E}$ y $\boldsymbol{d}^T \nabla c_i(\boldsymbol{x}^*) \geq 0$ para $i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}$.
Sea la matriz $A(\boldsymbol{x}^*)^T = [\nabla c_i(\boldsymbol{x}^*)]$ para $i \in \mathcal{A}(\boldsymbol{x}^*)$. Dado que se cumple la condición LICQ por hipótesis, la matriz $A(\boldsymbol{x}^*) \in \mathbb{R}^{m \times n}$, $m < n$ es de rango completo, ie, rango m.
Sea $Z \in \mathbb{R}^{n \times (n-m)}$ una base del espacio nulo de la matriz $A(\boldsymbol{x}^*)$ es decir, $A(\boldsymbol{x}^*)Z = 0 \in \mathbb{R}^{m \times (n-m)}$.

$$B = \begin{bmatrix} A(\boldsymbol{x}^*) \\ Z^T \end{bmatrix}$$

Prueba de la segunda parte

Construyamos la función $F(t, \boldsymbol{z}) : \mathbb{R}^1 \times \mathbb{R}^n \to \mathbb{R}^n$

$$F(t, \boldsymbol{z}) = \begin{bmatrix} c(\boldsymbol{z}) - tA(\boldsymbol{x}^*)\boldsymbol{d} \\ Z^T(\boldsymbol{z} - \boldsymbol{x}^* - t\boldsymbol{d}) \end{bmatrix}$$

y la curva de nivel o parametrización $F(t, \boldsymbol{z}) = 0$. Debido a que

$$D_z F(t, \boldsymbol{z}) = \begin{bmatrix} A(\boldsymbol{z}) \\ Z^T \end{bmatrix}$$

cumple $D_z F(0, \boldsymbol{x}^*) = B$ es invertible. Usando el teorema de la función implicita, existe una funcion $z(t)$ para t suficientemente pequeño tal que $F(t, \boldsymbol{z}(t)) = 0$.

Prueba de la segunda parte

Definamos: $\boldsymbol{z}_k \stackrel{def}{=} \boldsymbol{z}(t_k)$, con $t_k > 0$ entonces $F(t_k, \boldsymbol{z}_k) = 0$.

$$F(t_k, \boldsymbol{z}_k) = \begin{bmatrix} c(\boldsymbol{z}_k) - t_k A(\boldsymbol{x}^*)\boldsymbol{d} \\ Z^T(\boldsymbol{z} - \boldsymbol{x}^* - t_k \boldsymbol{d}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{aligned} c_i(\boldsymbol{z}_k) - t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} &= 0 \text{ para } i \in \mathcal{E} \\ c_i(\boldsymbol{z}_k) - t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} &= 0 \text{ para } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} \end{aligned}$$

Luego

$$\begin{aligned} c_i(\boldsymbol{z}_k) &= t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} = 0 \text{ para } i \in \mathcal{E} \\ c_i(\boldsymbol{z}_k) &= t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0 \text{ para } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} \end{aligned}$$

Esto ultimo pues $\boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*)$ y $t_k > 0$, ie, $\boldsymbol{z}_k$ es factible en $\mathcal{A}(\boldsymbol{x}^*)$

Prueba de la segunda parte

Por otro lado, usando Taylor y como $c(\boldsymbol{x}^*) = 0$

$$
\begin{aligned}
0 = F(t_k, \boldsymbol{z}_k) &= \begin{bmatrix} c(\boldsymbol{z}_k) - t_k A(\boldsymbol{x}^*)\boldsymbol{d} \\ Z^T(\boldsymbol{z}_k - \boldsymbol{x}^* - t_k \boldsymbol{d}) \end{bmatrix} \\
&= \begin{bmatrix} A(\boldsymbol{x}^*)(\boldsymbol{z}_k - \boldsymbol{x}^*) + o(\|\boldsymbol{z}_k - \boldsymbol{x}^*\|) - t_k A(\boldsymbol{x}^*)\boldsymbol{d} \\ Z^T(\boldsymbol{z}_k - \boldsymbol{x}^* - t_k \boldsymbol{d}) \end{bmatrix} \\
&= \begin{bmatrix} A(\boldsymbol{x}^*) \\ Z^T \end{bmatrix} (\boldsymbol{z}_k - \boldsymbol{x}^* - t_k \boldsymbol{d}) + o(\|z_k - \boldsymbol{x}^*\|)
\end{aligned}
$$

y como B es invertible $0 = \boldsymbol{z}_k - \boldsymbol{x}^* - t_k\boldsymbol{d} + o(\|\boldsymbol{z}_k - \boldsymbol{x}^*\|)$ es decir $\boldsymbol{d} \in T_\Omega(\boldsymbol{x}^*)$ y por tanto $\mathcal{F}(\boldsymbol{x}^*) \subset T_\Omega(\boldsymbol{x}^*)$, ie, $\mathcal{F}(\boldsymbol{x}^*) = T_\Omega(\boldsymbol{x}^*)$

Nota

- Se habia probabo que la secuencia z_k satisface

$$\begin{aligned} c_i(\boldsymbol{z}_k) &= t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} = 0 \text{ para } i \in \mathcal{E} \\ c_i(\boldsymbol{z}_k) &= t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0 \text{ para } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} \end{aligned}$$

  Esto ultimo pues $\boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*)$ y $t_k > 0$, es decir $\boldsymbol{z}_k$ es factible en $\mathcal{A}(\boldsymbol{x}^*)$

- La secuencia $\boldsymbol{z}_k$ es factible para las restricciones con indices en $\mathcal{I} \setminus \mathcal{A}(\boldsymbol{x}^*)$??
  Sea $i \notin \mathcal{A}(\boldsymbol{x}^*)$ entonces $c_i(\boldsymbol{x}^*) > 0$ pues i no se activó.

Nota: Idea 1

Sea $\boldsymbol{y}$ el punto mas cercano a $\boldsymbol{x}^*$ tal que $c_i(\boldsymbol{y}) = 0$ y definamos ν la distancia de $\boldsymbol{x}^*$ a $\boldsymbol{y}$, ie $\nu = \|\boldsymbol{x}^* - \boldsymbol{y}\|$.
Como $\boldsymbol{z}_k \to \boldsymbol{x}^*$ entonces: Para toda $\epsilon > 0$ existe $K > 0$ tal que para $k > K$ se cumple $\|\boldsymbol{z}_k - \boldsymbol{x}^*\| < \epsilon$. Tomando por ejemplo $\epsilon = 0{,}5\nu$, se sigue que $c_i(\boldsymbol{z}_k) > 0$ para $k > K$. Luego podemos obtener una subsuseción factible.

Nota: Idea 2

Usemos Taylor, y usemos que $\boldsymbol{z}_k = \boldsymbol{x}^* + t_k \boldsymbol{d} + o(t_k)$, $t_k > 0$

$$c_i(\boldsymbol{z}_k) \quad = \quad c_i(\boldsymbol{x}^*) + t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} + o(t_k)$$

Dado que $c_i(\boldsymbol{x}^*) > 0$ podemos podemos escoger t_k suficientemente pequeña, ie, $k > K$; tal que $c_i(\boldsymbol{x}^*) + t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} > 0$ y por tanto $c_i(\boldsymbol{z}_k) > 0$ para $k > K$ pues ambas expresiones tienen el mismo signo,.. ver Notas mas adelante!

Teorema

Una solución local del problema original es un punto x en el que todas las secuencias factibles satisfacen la propiedad $f(\boldsymbol{z}_k) \geq f(\boldsymbol{x})$ para k suficientemente grande.
El siguiente resultado muestra que si dichas secuencias existen entonces el producto interior de las direcciones límites y el gradiente de la función objetivo es no negativo

Teorema

If $\boldsymbol{x}^*$ es una solución local del problema original, entonces $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0$, para toda $\boldsymbol{d} \in T_\Omega(\boldsymbol{x}^*)$.

Prueba

Sea $\boldsymbol{d} \in T_{\Omega}(\boldsymbol{x}^*)$ y sean las correspondiente secuencia factible $\boldsymbol{z}_k \to \boldsymbol{x}^*$ y $t_k \to 0$, $t_k > 0$. Luego

$$\boldsymbol{z}_k \quad = \quad \boldsymbol{x}^* + t_k \boldsymbol{d} + o(t_k)$$

Usando Taylor

$$f(\boldsymbol{z}_k) \quad = \quad f(\boldsymbol{x}^*) + t_k \nabla f(\boldsymbol{x}^*) \boldsymbol{d} + o(t_k)$$

Supongamos, por el absurdo, que $\nabla f(\boldsymbol{x}^*) \boldsymbol{d} < 0$ luego $f(\boldsymbol{z}_k) < f(\boldsymbol{x}^*)$. Para k suficientemente grande $\boldsymbol{z}_k$ esta en una vecindad de $\boldsymbol{x}$, lo que contradice que $\boldsymbol{x}^*$ es un mínimo! lqqd

Lema de Farkas

Consideremos el cono $\mathcal{K} = \{B\boldsymbol{y} | \ \boldsymbol{y} \succeq \mathbf{0}\}$, $B \in \mathbb{R}^{n \times m}$, $\boldsymbol{y} \in \mathbb{R}^m$.

Lema de Farkas

Sea el **cono** K definido como arriba. Dado un vector $g \in \mathbb{R}^n$ entonces se cumple exactamente una de las siguientes condiciones pero no ambas

1. $\boldsymbol{g} \in \mathcal{K}$ (ie, $\boldsymbol{g} = B\boldsymbol{y}$ con $\boldsymbol{y} \succeq \mathbf{0}$)
2. Existe $\boldsymbol{d} \in \mathbb{R}^n$ tal que $\boldsymbol{g}^T\boldsymbol{d} < 0$, $B^T\boldsymbol{d} \succeq \mathbf{0}$

Lema de Farkas: Primera parte

Veamos que ambas alternativas no se puedes cumplir

- Si ambas alternativas se cumplen (usando el absurdo) entonces

$$\begin{aligned} \boldsymbol{g} &= B\boldsymbol{y}, \ \boldsymbol{y} \succeq \mathbf{0} \\ B^T\boldsymbol{d} &\succeq \mathbf{0} \\ \boldsymbol{g}^T\boldsymbol{d} &< 0 \end{aligned}$$

  A partir de la primera relación y usando el hecho que $B^T\boldsymbol{d} \succeq \mathbf{0}$ y $\boldsymbol{y} \succeq \mathbf{0}$.

$$\boldsymbol{g}^T\boldsymbol{d} = (B\boldsymbol{y})^T\boldsymbol{d} = \boldsymbol{y}^T(B^T\boldsymbol{d}) \geq 0$$

  que entra en contradición con la relación $\boldsymbol{g}^T\boldsymbol{d} < 0$.
- Por tanto, ambas alternativas no se pueden cumplir

Lema de Farkas: Comentario

- Si $\boldsymbol{d} \in \mathbb{R}^n$ satisface $B^T\boldsymbol{d} \succeq \boldsymbol{0}$ entonces $\boldsymbol{d}$ esta en el dual de $\mathcal{K}$, ie

$$\langle B\boldsymbol{y}, \boldsymbol{d}\rangle = (B\boldsymbol{y})^T\boldsymbol{d} \quad = \quad \boldsymbol{y}^T(B^T\boldsymbol{d}) \geq 0$$

  pues $\boldsymbol{y} \succeq \boldsymbol{0}$.
- Por lo anterior, es imposible que la proyección de $\boldsymbol{d}$ sobre un vector de $\mathcal{K}$ sea negativa!!, es decir, no se puede cumplir que

$$(B\boldsymbol{y})^T\boldsymbol{d} \quad < \quad 0$$

Por lo que nuevamente se confirma el Lema de Farkas, que no se pueden cumplir al mismo tiempo ambas alternativas.

Lema de Farkas: Segunda parte

Para la segunda parte se debe demostrar la existencia, es decir si $\boldsymbol{g} \notin \mathcal{K}$ entonces existe $\boldsymbol{d} \in \mathbb{R}^n$ tal que $\boldsymbol{g}^T \boldsymbol{d} < 0$. (En el Nocedal se muestra una forma de calcular $\boldsymbol{d}$)

1. Primero se prueba que $\mathcal{K}$ es cerrado, por lo que esta bien definido el punto mas proximo o distancia al cono
2. Se resuelve el problema

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s} \in \mathcal{K}} \|\boldsymbol{s} - \boldsymbol{g}\|$$

3. El vector $\hat{s}$ cumple las siguientes propiedades

$$\begin{aligned} \hat{\boldsymbol{s}}^T(\hat{\boldsymbol{s}} - \boldsymbol{g}) &= 0 \\ \boldsymbol{s}^T(\hat{\boldsymbol{s}} - \boldsymbol{g}) &\geq 0 \end{aligned}$$

Lema de Farkas: Segunda parte

1. Se define $\boldsymbol{d} \stackrel{def}{=} \hat{\boldsymbol{s}} - \boldsymbol{g}$. Reescribiendo las relaciones anteriores

$$\begin{aligned} \hat{\boldsymbol{s}}^T \boldsymbol{d} &= 0 \\ \boldsymbol{s}^T \boldsymbol{d} &\geq 0 \end{aligned}$$

   Además ,..., $\boldsymbol{g} = \hat{\boldsymbol{s}} - \boldsymbol{d}$

2. Se puede verificar que el vector definido arriba satisface la condiciones de la segunda parte del Lema de Farkas
3. Usando las condiciones anteriores para $\hat{\boldsymbol{s}}$

$$\boldsymbol{d}^T \boldsymbol{g} \quad = \quad \boldsymbol{d}^T (\hat{\boldsymbol{s}} - \boldsymbol{d}) = \hat{\boldsymbol{s}}^T \boldsymbol{d} - \boldsymbol{d}^T \boldsymbol{d} = -\|\boldsymbol{d}\|^2 < 0$$

   pues $\boldsymbol{d} \neq 0$ debido a que $\boldsymbol{g} \notin \mathcal{K}$, ie, $\boldsymbol{d} = 0$ ssi $\boldsymbol{g} \in \mathcal{K}$, o si $\hat{\boldsymbol{s}} = \boldsymbol{g}$

Lema de Farkas: Segunda parte

1. Como para toda $\boldsymbol{s} \in \mathcal{K}$, se tiene $\boldsymbol{s}^T(\hat{\boldsymbol{s}} - \boldsymbol{g}) \geq 0$ entonces $\boldsymbol{s}^T\boldsymbol{d} \geq 0$. Luego,

$$\boldsymbol{s}^T\boldsymbol{d} \;=\; (B\boldsymbol{y})^T\boldsymbol{d} = \boldsymbol{y}^T(B^T\boldsymbol{d}) \geq 0$$

   para toda como $\boldsymbol{y} \succeq 0$. Finalmente, $\boldsymbol{y}^T(B^T\boldsymbol{d}) \geq 0$ ssi $B^T\boldsymbol{d} \geq 0$

2. De lo contrario, si existe una entrada de $B^T\boldsymbol{d}$ menor que cero, es posible escoger $\boldsymbol{y} \succeq 0$ tal que $\boldsymbol{y}^T(B^T\boldsymbol{d}) < 0$
3. Por ejemplo, $y_i = 0$ para las componentes positivas de $B^T\boldsymbol{d}$ y $y_i > 0$ para las negativas, y se obtendria $\boldsymbol{y}^T(B^T\boldsymbol{d}) < 0$ que es una contradiccion

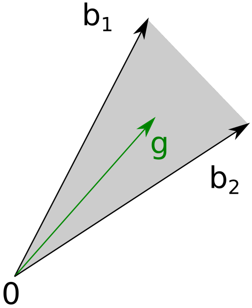
Lema de Farkas

Reescribiendo las alternativas del Lema de Farkas:

1. $\boldsymbol{g} = \sum_{i=1}^{m} y_i \boldsymbol{b}_i, \; y_i \geq 0$
2. Existe $\boldsymbol{d} \in \mathbb{R}^n$ tal que $\boldsymbol{g}^T \boldsymbol{d} < 0$, $\boldsymbol{b}_i^T \boldsymbol{d} \geq \mathbf{0}$

Lema de Farkas: Alternativa 1

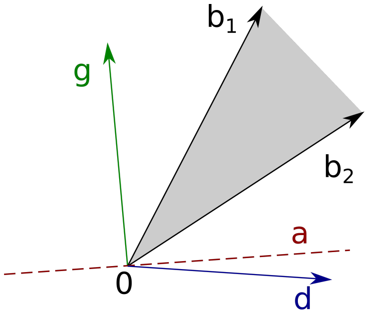
1. $\boldsymbol{g} = \sum_{i=1}^{m} y_i \boldsymbol{b}_i,\ y_i \geq 0$ (Alternativa 1)



$\boldsymbol{g} = y_1 \boldsymbol{b}_1 + y_2 \boldsymbol{b}_2$; $y_1, y_2 \geq 0$.

Lema de Farkas: Alternativa 2

1. Existe $\boldsymbol{d} \in \mathbb{R}^n$ tal que $\boldsymbol{g}^T \boldsymbol{d} < 0$, $\boldsymbol{b}_i^T \boldsymbol{d} \geq \boldsymbol{0}$ (Alternativa 2)



$$\boldsymbol{b}_1^T \boldsymbol{d} \geq 0,\ \boldsymbol{b}_2^T \boldsymbol{d} \geq 0,\ \boldsymbol{g}^T \boldsymbol{d} < 0.$$

Es decir, existe una recta en la dirección $\boldsymbol{a}$ ortogonal a $\boldsymbol{g}$; ie, $\boldsymbol{g}^T \boldsymbol{a} = 0$; separa al cono del vector $\boldsymbol{d}$.

Lema de Farkas

Vamos a definir $\boldsymbol{g} = \nabla f(\boldsymbol{x}^*)$ y $\boldsymbol{b}_i = \nabla c_i(\boldsymbol{x}^*)$, $y_i = \lambda_i^*$ entonces se cumple unas de las siguientes alternativas

1. $\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^{m} \lambda_i^* \nabla c_i(\boldsymbol{x}^*),\ \lambda_i^* \geq 0$
2. Existe $\boldsymbol{d} \in \mathbb{R}^n$ tal que $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} < 0$, $\nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} \geq \mathbf{0}$

Lema de Farkas: Condición de Primer Orden

1. $\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^{m} \lambda_i^* \nabla c_i(\boldsymbol{x}^*), \ \lambda_i^* \geq 0$
2. Existe $\boldsymbol{d} \in \mathbb{R}^n$ tal que $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} < 0$, $\nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d} \geq \mathbf{0}$

Comentario: Como $\boldsymbol{d} \in T_\Omega(\boldsymbol{x}^*)$ entonces por el Teorema $\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0$, luego se cumple la condición 1 que coincide con la KKT1. Ver detalles de la demostración en el Nocedal pagina 329.

Direcciones factibles linearizadas

Recuerda que ...

Dado un punto factible $x \in \Omega$ y el conjunto de restricciones activas $\mathcal{A}(\boldsymbol{x})$. El conjunto de direcciones factibles linearizadas $\mathcal{F}(\boldsymbol{x})$ es el conjunto

$$\mathcal{F}(\boldsymbol{x}^*) = \left\{ d \in \mathbb{R}^n \middle| \begin{array}{ll} \boldsymbol{d}^T \nabla c_i(\boldsymbol{x}^*) = 0 & \text{si } i \in \mathcal{E} \text{ y} \\ \boldsymbol{d}^T \nabla c_i(\boldsymbol{x}^*) \geq 0 & \text{si } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} \end{array} \right\}.$$

Nota: Si $\mathcal{A}(\boldsymbol{x}^*) = \emptyset$ entonces $\mathcal{F}(\boldsymbol{x}^*) = \mathbb{R}^n$

Direcciones factibles linearizadas

- Si $\boldsymbol{x}^*$ satisface LICQ entonces $\mathcal{F}(\boldsymbol{x}^*) = T_\Omega(\boldsymbol{x}^*)$
- Si $\boldsymbol{w} \in T_\Omega(\boldsymbol{x}^*)$ entonces $\boldsymbol{w}^T \nabla f(\boldsymbol{x}^*) \geq 0$
- Luego si $\boldsymbol{w} \in \mathcal{F}(\boldsymbol{x}^*)$ entonces $\boldsymbol{w}^T \nabla f(\boldsymbol{x}^*) \geq 0$

Direcciones factibles linearizadas

- Si $\boldsymbol{w}^T \nabla f(\boldsymbol{x}^*) > 0$ entonces es claro que la función crece, y por tanto $f(\boldsymbol{x}^*) < f(\boldsymbol{x}^* + t_k \boldsymbol{w})$ con $t_k > 0$
- Sin embargo, si $\boldsymbol{w}^T \nabla f(\boldsymbol{x}^*) = 0$ hasta ahora no podemos determinar si la función crece o decrece en esta direccion.
- Por lo anterior, nos interesa analizar las direcciones donde no podemos decir si la función crece o decrece ,i.e., $\boldsymbol{w}^T \nabla f(\boldsymbol{x}^*) = 0$. Para ello se define el cono crítico.
- Si $\nabla f(\boldsymbol{x}^*) = 0$ entonces $\boldsymbol{w} \in \mathcal{F}(\boldsymbol{x}^*) = T_\Omega(\boldsymbol{x}^*)$

Cono Crítico

Cono Crítico

Se define cono crítico al conjunto definido como sigue

$$\mathcal{C}(\boldsymbol{x}^*, \lambda^*) = \{\boldsymbol{w} \in \mathcal{F}(\boldsymbol{x}^*) | \text{ y en el caso } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}, \text{ con } \lambda_i > 0 \text{ se cumple } \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) = 0\}$$

- Si $\nabla f(\boldsymbol{x}^*) = 0$ entonces $\boldsymbol{w} \in \mathcal{F}(\boldsymbol{x}^*) = T_\Omega(\boldsymbol{x}^*)$, ie,

$$\mathcal{C}(\boldsymbol{x}^*, \lambda^*) = \mathcal{F}(\boldsymbol{x}^*) = T_\Omega(\boldsymbol{x}^*)$$

- Si $\boldsymbol{x}^* \in Interior(\Omega)$ entonces $\mathcal{A}(\boldsymbol{x}^*) = \emptyset$ y $\mathcal{C}(\boldsymbol{x}^*, \lambda^*) = \mathcal{F}(\boldsymbol{x}^*) = T_\Omega(\boldsymbol{x}^*) = \mathbb{R}^n$

Cono Crítico

Es decir

Cono Crítico

$$\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*) \quad \Leftrightarrow \quad \begin{cases} \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) = 0 & \text{si } i \in \mathcal{E} \\ \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) = 0 & \text{si } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}, \lambda_i^* > 0 \\ \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) \geq 0 & \text{si } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}, \lambda_i^* = 0 \end{cases}$$

Cono Crítico

Note que si $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$ entonces $\lambda_i^* \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) = 0$, para $i \in \mathcal{E} \cup \mathcal{I}$, i.e, Si $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$, es decir, $i \in \mathcal{A}(\boldsymbol{x}^*)$

$$\begin{aligned} \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) &= 0 \quad \text{si } i \in \mathcal{E} \\ \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) &= 0 \quad \text{si } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}, \lambda_i^* > 0 \\ \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) &\geq 0 \quad \text{si } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}, \lambda_i^* = 0 \end{aligned}$$

Por otro lado, si $i \notin \mathcal{A}(\boldsymbol{x}^*)$ entonces $\lambda_i^* = 0$, por lo que

$$\boldsymbol{w}^T \nabla f(\boldsymbol{x}^*) = \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^*) = 0$$

Es decir, son las direcciones donde no podemos decidir si la función crece o decrece.

Condición necesaria de segundo orden

Sea $\boldsymbol{x}^*$ una solución del problema original y Además satisface la condición LICQ. Si λ^* satisface las KKTs, entonces

$$\boldsymbol{w}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \lambda^*) \boldsymbol{w} \quad \geq \quad 0, \text{ para todo } \boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$$

Condición necesaria de segundo orden: Comentarios

- Por la condición LICQ se sabe que $T_\Omega(\boldsymbol{x}^*) = \mathcal{F}(\boldsymbol{x}^*)$
- Sea $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*) \subset \mathcal{F}(\boldsymbol{x}^*)$
- Luego $\boldsymbol{w} \in T_\Omega(\boldsymbol{x}^*) = \mathcal{F}(\boldsymbol{x}^*)$ y por tanto, es una dirección límite, ie, existe una sucesion $\boldsymbol{z}_k$ factible tal que

$$\boldsymbol{w} \quad = \quad \lim_{k\to\infty} \frac{\boldsymbol{z}_k - \boldsymbol{x}^*}{t_k}$$

donde $t_k \to 0$ y $t_k > 0$. (Ver lamina 10, la demostración de la segunda parte del Lema de la lamina 7 de esta presentacion)

Condición necesaria de segundo orden: Comentarios

- Podemos escribir entonces

$$\boldsymbol{z}_k - \boldsymbol{x}^* \;=\; t_k \boldsymbol{w} + o(t_k)$$

Condición necesaria de segundo orden: Comentarios

- (Ver lamina 10) La secuencia $\boldsymbol{z}_k$ es factible, ie, satisface

$$
\begin{aligned}
c_i(\boldsymbol{z}_k) &= t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{w} = 0 \text{ para } i \in \mathcal{E} \\
c_i(\boldsymbol{z}_k) &= t_k \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{w} \geq 0 \text{ para } i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I}
\end{aligned}
$$

- Evaluando el Lagrangiano en $\boldsymbol{z}_k, \lambda^*$

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{z}_k, \lambda^*) &= f(\boldsymbol{z}_k) - \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* c_i(\boldsymbol{z}_k) \\
&= f(\boldsymbol{z}_k) - t_k \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{w} \\
&= f(\boldsymbol{z}_k) - t_k \overbrace{\nabla f(\boldsymbol{x}^*)^T \boldsymbol{w}}^{=0} \\
&= f(\boldsymbol{z}_k)
\end{aligned}
$$

Condición necesaria de segundo orden: Comentarios

- Evaluando el Lagrangiano en $\boldsymbol{x}^*, \lambda^*$ y como $c_i(\boldsymbol{x}^*) = 0$ para $i \in \mathcal{A}(\boldsymbol{x}^*)$ o por condiciones de complementariedad para toda $i \in \mathcal{E} \cup \mathcal{I}$

$$\begin{aligned}\mathcal{L}(\boldsymbol{x}^*, \lambda^*) &= f(\boldsymbol{x}^*) - \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \overbrace{c_i(\boldsymbol{x}^*)}^{=0} \\ &= f(\boldsymbol{x}^*)\end{aligned}$$

Condición necesaria de segundo orden: Resumen

Resumen

$$
\begin{array}{rcl}
\boldsymbol{z}_k - \boldsymbol{x}^* & = & t_k \boldsymbol{w} + o(t_k) \\
\mathcal{L}(\boldsymbol{z}_k, \lambda^*) & = & f(\boldsymbol{z}_k) \\
\mathcal{L}(\boldsymbol{x}^*, \lambda^*) & = & f(\boldsymbol{x}^*)
\end{array}
$$

Condición necesaria de segundo orden: Taylor

Apliquemos Taylor al Lagrangiano

$$\overbrace{\mathcal{L}(\boldsymbol{z}_k, \lambda^*)}^{f(\boldsymbol{z}_k)} = \overbrace{\mathcal{L}(\boldsymbol{x}^*, \lambda^*)}^{f(\boldsymbol{x}^*)} + \overbrace{\nabla_x \mathcal{L}(\boldsymbol{x}^*, \lambda^*)^T}^{0} \overbrace{(\boldsymbol{z}_k - \boldsymbol{x}^*)}^{t_k \boldsymbol{w} + o(t_k)} + \\ + \frac{1}{2} \overbrace{(\boldsymbol{z}_k - \boldsymbol{x}^*)^T}^{t_k \boldsymbol{w} + o(t_k)} \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \lambda^*) \overbrace{(\boldsymbol{z}_k - \boldsymbol{x}^*)}^{t_k \boldsymbol{w} + o(t_k)} + o(\overbrace{\|\boldsymbol{z}_k - \boldsymbol{x}^*\|^2}^{t_k^2})$$

$$\overbrace{f(\boldsymbol{z}_k) - f(\boldsymbol{x}^*)}^{\geq 0} = \frac{1}{2} t_k^2 \boldsymbol{w}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \lambda^*) \boldsymbol{w} + o(t_k^2)$$

Luego

$$\boldsymbol{w}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \lambda^*) \boldsymbol{w} \quad \geq \quad 0$$

Condición suficiente de segundo orden

Suponga que $\boldsymbol{x}^*$ es un punto factible y que existe λ^* satisface las KKTs. Si

$$\boldsymbol{w}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*, \lambda^*) \boldsymbol{w} \quad > \quad 0, \text{ para todo } \boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*), \boldsymbol{w} \neq \mathbf{0}$$

entonces $\boldsymbol{x}^*$ es un mínimo local del problema original.

Ver detalles en el Nocedal

Ejemplo 1

$$\text{min } -0{,}1(x_1-4)^2+x_2^2, \text{ s.t. } x_1^2+x_2^2-1\geq 0$$

Lagrangiano

$$\mathcal{L}(x_1,x_2,\lambda) = -0{,}1(x_1-4)^2+x_2^2-\lambda(x_1^2+x_2^2-1)$$

KKTs

$$\begin{aligned}
\nabla_x\mathcal{L}(x_1,x_2,\lambda) &= 0\\
x_1^2+x_2^2-1 &\geq 0\\
\lambda &\geq 0\\
\lambda(x_1^2+x_2^2-1) &= 0
\end{aligned}$$

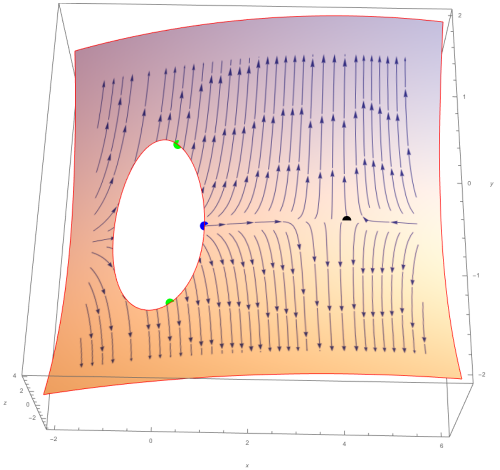
Ejemplo 1:KKTs

$$\begin{aligned}
-0{,}2(x_1 - 4) - 2\lambda x_1 &= 0 \\
2x_2 - 2\lambda x_2 &= 0 \\
x_1^2 + x_2^2 - 1 &\geq 0 \\
\lambda &\geq 0 \\
\lambda(x_1^2 + x_2^2 - 1) &= 0
\end{aligned}$$

1. Si $\lambda = 0$ entonces $x_1 = 4$ y $x_2 = 0$
2. Si $\lambda = 1$ entonces $x_1^2 + x_2^2 - 1 = 0$ y por tanto $x_1 = \frac{4}{11}$ y $x_2 = \pm\frac{\sqrt{105}}{11}$
3. Si $\lambda > 0$ y $\lambda \neq 1$, entonces $x_1^2 + x_2^2 - 1 = 0$, y por tanto $x_2 = 0$, $x_1 = 1$, $\lambda = 0{,}3$ o $x_2 = 0$, $x_1 = -1$, $\lambda = -0{,}5$ (imposible pues $\lambda > 0$)

Ejemplo 1: Puntos crítcos

Ver en **Mathematica**: ejemplo1Tema1Clase3.nb



Ejemplo 1

- Caso: $\boldsymbol{x}^* = [1, 0]^T$, $\lambda^* = 0{,}3$

El **conjunto activo** es $\mathcal{A}(\boldsymbol{x}^*) = \{1\}$
Cono crítico: como $\lambda^* = 0{,}3 > 0$ y $\mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} = \{1\}$ entonces, si $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$ entonces se tiene que cumplir $\boldsymbol{w}^T \nabla c(\boldsymbol{x}^*) = 0$..

$$\nabla c(\boldsymbol{x}^*) \quad = \quad 2[\boldsymbol{x}_1^*, \boldsymbol{x}_2^*]^T = [2, 0]^T$$

Luego de $\boldsymbol{w}^T \nabla c(\boldsymbol{x}^*) = 0$ se tiene que $w_1 = 0$, por lo que

$$\mathcal{C}(\boldsymbol{x}^*, \lambda^*) \quad = \quad \{[0, w_2]^T |\, w_2 \in \mathbb{R}\}$$

Ejemplo 1

El Lagrangiano es:

$$\begin{aligned}
\nabla^2_{xx}\mathcal{L}(\boldsymbol{x},\lambda) &= \begin{bmatrix} -0{,}2-2\lambda & 0 \\ 0 & 2-2\lambda \end{bmatrix} \\
\nabla^2_{xx}\mathcal{L}(\boldsymbol{x}^*,\lambda^*) &= \begin{bmatrix} -0{,}8 & 0 \\ 0 & 1{,}4 \end{bmatrix}
\end{aligned}$$

Finalmente, para $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*,\lambda^*)$ se tiene que $\boldsymbol{w} = [0, w_2]^T \neq 0$ y por tanto

$$\boldsymbol{w}^T\nabla^2_{xx}\mathcal{L}(\boldsymbol{x}^*,\lambda^*)\boldsymbol{w} = 1{,}4w_2^2 > 0,$$

por lo que $\boldsymbol{x}^* = [1,0]^T$ es un mínimo local.

Ejemplo 1

- Caso: $\boldsymbol{x}^* = [\frac{4}{11}, \pm\frac{\sqrt{105}}{11}]^T$, $\lambda^* = 1$

El **conjunto activo** es $\mathcal{A}(\boldsymbol{x}^*) = \{1\}$
Cono crítico: como $\lambda^* = 1 > 0$ y $\mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} = \{1\}$ entonces, si $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*, \lambda^*)$ entonces se tiene que cumplir $\boldsymbol{w}^T \nabla c(\boldsymbol{x}^*) = 0$..

$$\nabla c(\boldsymbol{x}^*) \quad = \quad 2[\boldsymbol{x}_1^*, \boldsymbol{x}_2^*]^T = 2[\frac{4}{11}, \pm\frac{\sqrt{105}}{11}]^T$$

Luego de $\boldsymbol{w}^T \nabla c(\boldsymbol{x}^*) = 0$, por lo que

$$\mathcal{C}(\boldsymbol{x}^*, \lambda^*) \quad = \quad \{\boldsymbol{w} = w_1[\sqrt{105}, \mp 4]^T |\, w_1 \in \mathbb{R}\}$$

Ejemplo 1

El Lagrangiano es:

$$\begin{aligned} \nabla^2_{xx}\mathcal{L}(\boldsymbol{x},\lambda) &= \begin{bmatrix} -0{,}2-2\lambda & 0 \\ 0 & 2-2\lambda \end{bmatrix} \\ \nabla^2_{xx}\mathcal{L}(\boldsymbol{x}^*,\lambda^*) &= \begin{bmatrix} -2{,}2 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Finalmente, para $\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^*,\lambda^*)$ se tiene que $\boldsymbol{w} = w_1[\sqrt{105}, \mp 4]^T \neq 0$ y por tanto

$$\boldsymbol{w}^T \nabla^2_{xx}\mathcal{L}(\boldsymbol{x}^*,\lambda^*)\boldsymbol{w} = -231 w_1^2 < 0,$$

por lo que en $\boldsymbol{x}^* = [\frac{4}{11}, \pm\frac{\sqrt{105}}{11}]^T$ hay máximos locales en el cono critico $\mathcal{C}(\boldsymbol{x}^*,\lambda^*)$,... son maximos locales del problema?

Ejemplo 1

- Caso: $\boldsymbol{x}^* = [4, 0]^T$, $\lambda^* = 0$

El **conjunto activo** es $\mathcal{A}(\boldsymbol{x}^*) = \emptyset$, pues $c(\boldsymbol{x}^*) = 15 > 0$
Cono crítico: Por lo que $\boldsymbol{x}^* \in Interior(\Omega)$ y por tanto

$$\mathcal{C}(\boldsymbol{x}^*, \lambda^*) \;=\; T_\Omega(\boldsymbol{x}^*) = \mathcal{F}(\boldsymbol{x}^*) = \mathbb{R}^n$$

Además, se cumple que $\nabla f(\boldsymbol{x}^*) = 0$ pues $\lambda^* = 0$. Es decir, este caso, se analiza como un problema sin restricciones

Ejemplo 1

El Lagrangiano es:

$$\begin{aligned}
\nabla^2_{xx}\mathcal{L}(\boldsymbol{x},\lambda) &= \begin{bmatrix} -0{,}2-2\lambda & 0 \\ 0 & 2-2\lambda \end{bmatrix} \\
\nabla^2_{xx}\mathcal{L}(\boldsymbol{x}^*,\lambda^*) &= \begin{bmatrix} -0{,}2 & 0 \\ 0 & 2 \end{bmatrix}
\end{aligned}$$

por lo que en $\boldsymbol{x}^* = [4,0]^T$ hay un punto silla.

Nota

Si $t_k > 0$, $t_k \to 0$, $z_k = x + ct_k + o(t_k)$, $b \neq 0$ y

$$f(z_k) \quad = \quad a + t_k b + o(t_k)$$

Entonces $\text{sgn}(f(z_k) - a) = \text{sgn}(b)$.
Ideas generales:

$$\lim_{k\to\infty} \frac{f(z_k) - a}{t_k} \quad = \quad \lim_{k\to\infty} \left(b + \frac{o(t_k)}{t_k} \right) = b$$

Luego por definición de límite de sucesiones, ie,

$$\forall \epsilon > 0 (\exists N \in \mathbb{N} (\forall n \in \mathbb{N} (n \geq N \implies |x_n - x| < \epsilon)))$$

Nota

Existe K tal que para tada $k > K$

$$\left| \frac{f(z_k) - a}{t_k} - b \right| \quad \leq \quad \epsilon$$

Si $b > 0$ seleccionemos por ejemplo: $\epsilon = b/2$

$$\begin{aligned} -\epsilon \quad &\leq \quad \frac{f(z_k) - a}{t_k} - b \\ f(z_k) - a \quad &\geq \quad t_k(b - \epsilon) = t_k b/2 > 0 \end{aligned}$$

luego $f(z_k) - a > 0$ y en este caso sgn$(f(z_k) - a) =$ sgn(b).

Nota

Existe K tal que para tada $k > K$

$$\left| \frac{f(z_k) - a}{t_k} - b \right| \quad \leq \quad \epsilon$$

Si $b < 0$ seleccionemos por ejemplo: $\epsilon = -b/2$

$$\begin{aligned} \frac{f(z_k) - a}{t_k} - b \quad &< \quad \epsilon \\ f(z_k) - a \quad &< \quad t_k(b + \epsilon) = t_k b/2 < 0 \end{aligned}$$

luego $f(z_k) - a < 0$ y en este caso sgn$(f(z_k) - a) =$ sgn(b).

Nota

Sea A de rango completo y Z una base del espacio nulo de A, ie, $AZ = 0$, entonces la matriz

$$\left[\begin{array}{c} A \\ Z^T \end{array} \right]$$

es invertible.
Basta probar que la solución del sistema

$$\left[\begin{array}{cc} A^T, & Z \end{array} \right] x \;\; = \;\; 0$$

es unica, ie, que $x = 0$

Nota

Basta probar que la solución del sistema

$$\left[\ A^T,\ Z\ \right] x \quad = \quad 0$$

es unica, ie, que $x = 0$. (Otra via: Notar que AA^T y $Z^T Z$ tienen inversa!), como $AZ = 0$

$$\begin{array}{rcl} A^T x_a + Z x_z & = & 0 \\ AA^T x_a + AZ x_z & = & 0 \\ x_a^T AA^T x_a & = & 0 \\ \|A^T x_a\| & = & 0 \\ A^T x_a & = & 0 \end{array}$$

y $x_a = 0$ pues las columnas de A^T son una base, de forma similar $x_b = 0$. luego $x = 0$ lqqd!