Optimización II. Capítulo 12. “Teoría de la Optimización con Restricciones”

Oscar S. Dalmau Cedeño
dalmau@cimat.mx

10 de septiembre de 2018

Problema de optimización

Formulación general de un problema de minimización con restricciones.

$$\min_{x\in\mathbb{R}^n} f(x) \quad s.a: \begin{cases} c_i(x) = 0, & i \in \mathcal{E}; \\ c_i(x) \geq 0, & i \in \mathcal{I}. \end{cases} \tag{1}$$

Se define el ***conjunto factible*** Ω como el conjunto de puntos x que satisfacen las restricciones, es decir:

$$\Omega = \left\{ x \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I} \right\} \tag{2}$$

El ***conjunto activo*** $\mathcal{A}(x)$ en cualquier punto factible x, es la unión del conjunto $\mathcal{E}$ con el conjunto de índices de las restricciones de desigualdad que se activan:

$$\mathcal{A}(x) = \mathcal{E} \cup \{ i \in \mathcal{I} \mid c_i(x) = 0 \} \tag{3}$$

Ejemplo 1: Una restricción de igualdad

Veamos un caso mas general

$$\begin{aligned} & \text{min } f(x) \\ s.t.: \quad & c(x) = 0 \end{aligned}$$

En resumen: Si existe una d dirección de descenso que mantiene factibilidad entonces tiene que cumplir que:

$$\nabla f(x)^T d < 0 \text{ (condición de descenso)} \quad (4)$$
$$\nabla c(x)^T d = 0 \text{ (condición de factibilidad)} \quad (5)$$

Ejemplo 1: Una restricción de igualdad

Veamos un caso mas general

$$\begin{aligned} &\text{min } f(x) \\ s.t.: \quad & c(x) = 0 \end{aligned}$$

En resumen: Si NO existe una d dirección de descenso que mantiene factibilidad entonces, se llego al óptimo y

$$\begin{aligned} \nabla f(x^*)^T d &= 0 \text{ (no cumple condición de descenso)} \\ \nabla c(x^*)^T d &= 0 \text{ (condición de factibilidad)} \end{aligned}$$

Luego $\nabla f(x^*) || \nabla c(x^*)$ y por tanto $\nabla f(x^*) = \lambda^* \nabla c(x^*)$ (en 2 dimensiones) Nota: Si $\nabla f(x^*)^T d > 0$, entonces $-d$ satisface las condiciones (4)-(5)

Ejemplo 2: Una restricción de desigualdad

Asumamos que $\nabla f(x)^T d < 0$, (por tanto, $\nabla f(x) \neq 0$,

Caso I, $c(x) > 0$: Es decir, x esta en el interior de la región factible. Entonces podemos escoger d de longitud suficientemente pequeña tal que

$$\begin{aligned} c(x) + \nabla c(x)^T d &\geq 0 \text{ (condición de factibilidad)} \\ \nabla f(x)^T d &< 0 \text{ (condición de descenso)} \end{aligned}$$

Caso II, $c(x) = 0$: Es decir, x esta en la frontera de la región factible. Luego, de $c(x) + \nabla c(x)^T d \geq 0$ tenemos

$$\begin{aligned} \nabla c(x)^T d &\geq 0 \text{ (condición de factibilidad, caso II)} \\ \nabla f(x)^T d &< 0 \text{ (condición de descenso)} \end{aligned}$$

Ejemplo 2: Una restricción de desigualdad

- No hay dirección de descenso cuando

$$\begin{aligned} \nabla f(x) &= \lambda \nabla c(x), \text{ con} \\ \lambda &\geq 0 \end{aligned}$$

y en este caso se tiene un óptimo. λ debe ser mayor o igual a cero pues $\nabla f(x)$ y $\nabla c(x)$ tienen la misma dirección.

Condición LICQ

Definición 12.1 (LICQ).
Dado un punto $\boldsymbol{x}^*$ y el conjunto activo $\mathcal{A}(\boldsymbol{x}^*)$, se dice que se satisface la *calificación de independencia lineal de los gradientes de las restricciones* (LICQ) si el conjunto de los gradientes de las restricciones activas $\{\nabla c_i(\boldsymbol{x}^*), i \in \mathcal{A}(\boldsymbol{x}^*)\}$ son linealmente independientes.

Lagrangiano

El Lagrangiano del problema original se define:

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x) \tag{6}$$

al vector $\lambda = [\lambda_1, \lambda_2, \cdots, \lambda_{|\mathcal{E} \cup \mathcal{I}|}]^T$ se le llama Multiplicador de Lagrange.

Teorema 12.1 (Condiciones Necesarias de Primer Orden)
Suponga que $\boldsymbol{x}^*$ es una solución local de (1) y que la *LICQ* se satisface en $\boldsymbol{x}^*$. Entonces existe un *vector multiplicador de la Lagrange* λ^*, con las componentes λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, tal que se satisfacen las siguientes condiciones en $(\boldsymbol{x}^*, \lambda^*)$

$$\begin{aligned}
\nabla_x \mathcal{L}(\boldsymbol{x}^*, \lambda^*) &= 0 && (7)\\
c_i(\boldsymbol{x}^*) &= 0, \; i \in \mathcal{E} && (8)\\
c_i(\boldsymbol{x}^*) &\geq 0, \; i \in \mathcal{I} && (9)\\
\lambda_i^* &\geq 0, \; i \in \mathcal{I} && (10)\\
\lambda_i^* c_i(\boldsymbol{x}^*) &= 0, \; i \in \mathcal{E} \cup \mathcal{I} && (11)
\end{aligned}$$

Condiciones de *Karush–Kuhn–Tucker* (condiciones KKT).

- Teniendo en cuenta las condiciones anteriores, se pueden omitir algunos términos en la primera condición.

$$0 = \nabla_x \mathcal{L}(\boldsymbol{x}^*, \lambda^*) = \nabla f(\boldsymbol{x}^*) - \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \nabla c_i(\boldsymbol{x}^*). \tag{12}$$

  Por (KKT5) $\lambda_i^* = 0$ si $c_i(\boldsymbol{x}^*) > 0$, ie $i \notin \mathcal{A}(\boldsymbol{x}^*)$
- En el Teorema anterior una de las hipótesis es que $\boldsymbol{x}^*$ es una solución local del problema original.
- La unicidad de los multiplicadores de Lagrange es para x^* conocido o dado. Sin embargo, en caso de tener otro mínimo, el multiplicador de lagrange puede ser diferente al anterior, es decir, es único para cada x^*.
- A la condición (KKT5) se lo conoce como ***Condición de Complementariedad***.

KKTs

$$\nabla f(\boldsymbol{x}^*) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \nabla c_i(\boldsymbol{x}^*) = 0 \qquad \text{(KKT1)}$$

$$c_i(\boldsymbol{x}^*) = 0, \;\; i \in \mathcal{E} \qquad \text{(KKT2)}$$

$$c_i(\boldsymbol{x}^*) \geq 0, \;\; i \in \mathcal{I} \qquad \text{(KKT3)}$$

$$\lambda_i^* \geq 0, \;\; i \in \mathcal{I} \qquad \text{(KKT4)}$$

$$\lambda_i^* c_i(\boldsymbol{x}^*) = 0, \;\; i \in \mathcal{E} \cup \mathcal{I} \qquad \text{(KKT5)}$$

Nota: Si

KKTs

Si conocemos el conjunto $\mathcal{A}(\boldsymbol{x}^*)$ entonces $\lambda_i^* = 0$ para $i \notin \mathcal{A}(\boldsymbol{x}^*)$, y podermos quitar los indices $i \notin \mathcal{A}(\boldsymbol{x}^*)$ de la funcion, luego tenemos la siguiente simplificación

$$\begin{aligned}
\nabla f(\boldsymbol{x}^*) - \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \nabla c_i(\boldsymbol{x}^*) &= 0 && (13)\\
c_i(\boldsymbol{x}^*) &= 0,\ i \in \mathcal{A}(\boldsymbol{x}^*) && (14)\\
\lambda_i^* &\geq 0,\ i \in \mathcal{A}(\boldsymbol{x}^*) \cap \mathcal{I} && (15)\\
c_i(\boldsymbol{x}^*) &> 0,\ i \notin \mathcal{A}(\boldsymbol{x}^*) && (16)\\
\lambda_i^* &= 0,\ i \notin \mathcal{A}(\boldsymbol{x}^*) && (17)
\end{aligned}$$

Nota

En muchos textos se dice: que si un multiplicador de Lagrange i se anula, ie, $\lambda_i^* = 0$, la restricción correspondiente se puede eliminar del problema de optimización, por ejemplo: En Practical optimization Methods With mathematica Applications. M. Asghar Bhatti, pag 148, se dice:

- ‘Inactive inequality constraints: constraint for which $c_i(x^*) > 0$. These constraints do not determine the optimum and hence, are not needed in developing optimality conditions”
- La afirmación anterior es válida cuando se analiza la condicion de extremo de un punto en particular x^*.
- Si $c_i(x^*) > 0$ siempre podemos encontrar una vecindad alrededor de x^*, tal que $c_i(x) > 0$ para toda x en dicha vecindad. (propiedad de conservación del signo para funciones continuas, ver notas Optimización I).

Nota

- Por lo que la restriccion no contaria en el analisis de la condicion de extremo, debido a que (el extremo local) es una condicion local.
- La afirmación anterior no significa, que las restricciones que no se activan se pueden eliminar en el problema general. Pues estas restricciones, son condiciones de optimalidad que deben ser satisfechas, ie, también determinan el conjunto factible.
- Es decir, si se eliminan las restricciones inactivas, el problema que resulta no es necesariamente equivalente al problema original. Ver el problema (4) mas abajo.

Problemas

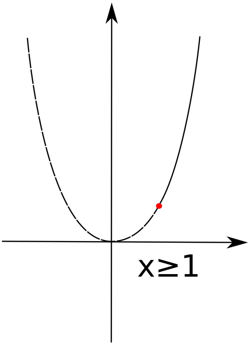
$$\min\ f(x) = \frac{1}{2}x^2;\ s.a:\ x \geq 1 \qquad \text{(Problema 1)}$$

$$\min\ f(x) = \frac{1}{2}x^2;\ s.a:\ x \geq 0 \qquad \text{(Problema 2)}$$

$$\min\ f(x) = \frac{1}{2}x^2;\ s.a:\ x \geq -1 \qquad \text{(Problema 3)}$$

$$\min\ f(x) = \frac{1}{4}(x^4 + 4x^3 - 8x^2);\ s.a:\ x \geq \frac{1}{2} \qquad \text{(Problema 4)}$$

Problema 1



$$\min \frac{1}{2}x^2;\ x \geq 1$$

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^2 - \lambda(x-1)$$

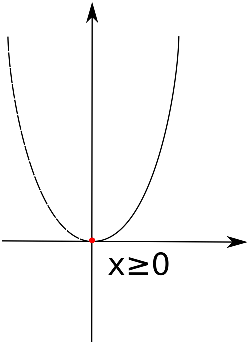
$$\nabla_x \mathcal{L}(x,\lambda) = x - \lambda = 0$$

$$x \geq 1;\ \lambda \geq 0$$

$$\lambda(x-1) = 0$$

Resolviendo las KKTs, $x^* = \lambda^* = 1$ y no se puede eliminar la restricción pues está activa, ie, $x^* - 1 = 0$ y $\lambda^* = 1 > 0$

Problema 2



$$\min \frac{1}{2}x^2;\ x \geq 0$$

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^2 - \lambda x$$

$$\nabla_x \mathcal{L}(x,\lambda) = x - \lambda = 0$$

$$x \geq 0;\ \lambda \geq 0$$

$$\lambda x = 0$$

Resolviendo las KKTs, $x^* = \lambda^* = 0$ y como $\lambda^* = 0$ se puede eliminar la restricción $x \geq 0$, suponiendo que $x^* = 0$ es dado. En el caso anterior, las condiciones KKTs son equivalentes a resolver el problema sin restricciones.

Problema 2

Suponiendo que $x^* = 0$ es conocido, las KKTs son:

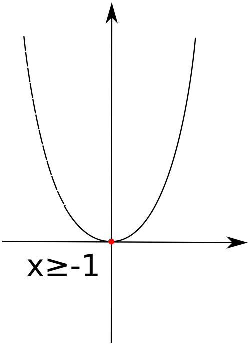
$$\begin{aligned} x^* - \lambda^* &= 0 \\ x^* &\geq 0 \\ \lambda^* &\geq 0 \\ \lambda^* x^* &= 0 \end{aligned}$$

.

Es lo mismo que las condiciones de optimalidad sin restricciones

$$\begin{aligned} f(x) &= \frac{1}{2}x^2 \\ f'(x^*) = x^* &= 0 \end{aligned}$$

Problema 3



$$\min \frac{1}{2}x^2;\ x \geq -1$$

$$\mathcal{L}(x,\lambda) = \frac{1}{2}x^2 - \lambda(x+1)$$

$$\nabla_x \mathcal{L}(x,\lambda) = x - \lambda = 0$$

$$x \geq -1;\ \lambda \geq 0$$

$$\lambda(x+1) = 0$$

Resolviendo las KKTs, $x^* = \lambda^* = 0$ y se puede eliminar la restricción $x \geq -1$ pues no está activa, ie, $x^* + 1 > 0$ y $\lambda^* = 0$

Problema 3

Suponiendo que $x^* = 0$ es conocido, las KKTs son:

$$
\begin{aligned}
x^* - \lambda^* &= 0 \\
x^* &\geq -1 \\
\lambda^* &\geq 0 \\
\lambda^*(x^* + 1) &= 0
\end{aligned}
$$

.

Es lo mismo que las condiciones de optimalidad sin restricciones

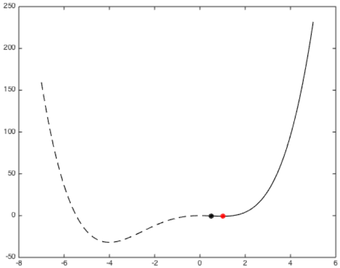
$$
\begin{aligned}
f(x) &= \frac{1}{2}x^2 \\
f'(x^*) = x^* &= 0
\end{aligned}
$$

Problema 4

$$\min\ f(x) = \frac{1}{4}(x^4 + 4x^3 - 8x^2); \quad (18)$$

$$s.a : x \geq \frac{1}{2} \quad (19)$$

El óptimo es $x^* = 1$



Problema 4

$$\min f(x) = \frac{1}{4}(x^4 + 4x^3 - 8x^2); \tag{20}$$

$$s.a : x \geq \frac{1}{2} \tag{21}$$

El Lagrangiano es $\mathcal{L}(x, \lambda) = f(x) - \lambda(x - \frac{1}{2})$ y las condiciones KKTs

$$x^3 + 3x^2 - 4x - \lambda = 0 \tag{KKT1}$$

$$x \geq \frac{1}{2} \tag{KKT2}$$

$$\lambda \geq 0 \tag{KKT3}$$

$$\lambda(x - \frac{1}{2}) = 0 \tag{KKT4}$$

Problema 4

Se puede ver que para $x^* = 1$ y $\lambda^* = 0$ se cumplen las KKTs

$$(x^*)^3 + 3(x^*)^2 - 4x^* - \lambda^* = 0 \qquad \text{(KKT1)}$$

$$x^* \geq \frac{1}{2} \qquad \text{(KKT3)}$$

$$\lambda^* \geq 0 \qquad \text{(KKT4)}$$

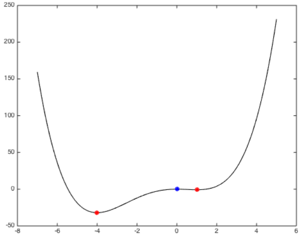
$$\lambda^*(x^* - \frac{1}{2}) = 0 \qquad \text{(KKT5)}$$

Ademas: Hay una sola solución y la restricción KKT2 es inactiva, ie, $x^* > \frac{1}{2}$.

Problema 4

En claro que la restricción $x^* > \frac{1}{2}$ no influye en la determinación del óptimo local $x^* = 1$ y se puede eliminar del problema. Al eliminar esta restricción se obtiene un problema sin restricciones

$$\min f(x) = \frac{1}{4}(x^4 + 4x^3 - 8x^2) \tag{22}$$



Y ahora se tienen dos mínimos locales $x^* = 1$ y $x^* = 4$. Por lo que al eliminar la restricción el nuevo problema no es equivalente, y para obtener el óptimo se debe seleccionar el que satisface $x > \frac{1}{2}$ concluyendo que $x^* = 1$

Resumen

- Si $c_i(x^*) > 0$, es decir, si la restricción i-esima no se activa en el optimo x^*, la restricción correspondiente se puede eliminar de las KKTs, ie, se obtiene KKTs *modificadas*.
- Al eliminar la restricción i-ésima no activa, el problema resultante no es equivalente, ie, la solución de las KKTs *modificadas* contiene optimo x^*, pero pueden aparecer nuevas soluciones que no son óptimos del problema original.
- Para obtener las soluciones del problema original, seleccionamos las soluciones de las KKTs *modificadas* que satisfacen todas las restricciones, ie, incluyendo las restricciones no activas!

Resumen

- Por otro lado, si $\lambda_i^* > 0; i \in \mathcal{I}$, entonces, se activa la restricción correspondiente, ie, $c_i(x^*) = 0$.
- Lo anterior es un indicativo de que la restricción es muy relevante, y si es eliminada de las KKTs, el optimo se pierde y no es posible recuperar dicha solución a partir de las KKTs modificadas.
- Ver el Problema 1 discutido anteriormente. En este ejemplo, si se elimina la restricción $x \geq 1$ se obtiene un problema sin restricciones cuyo optimo es $x^* = 0$, es decir, no podremos recuperar la solución $x^* = 1$ del problema con restricciones.

Ejemplo 1: KKT

Escriba las condiciones de optimalidad del siguiente problema (KKTs)

$$\begin{aligned} &\text{min } f(x) = x_1 + x_2 + x_3 \\ s.t.: \quad & x_1^2 + x_2^2 + x_3^2 = 1 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

Ejemplo 1: KKT

$$\mathcal{L}(x,\lambda) \quad = \quad x_1 + x_2 + x_3 - \sum_{i=1}^{3} \lambda_i x_i - \lambda_4(x_1^2 + x_2^2 + x_3^2 - 1)$$

KKTs: Sea $\lambda = [\lambda_1, \lambda_2, \lambda_3]^T$

$$\nabla_x \mathcal{L} = \mathbf{1} - \lambda - 2\lambda_4 x = 0 \qquad \text{(KKT1)}$$

$$x_1^2 + x_2^2 + x_3^2 = 1 \qquad \text{(KKT2)}$$

$$x_i \geq 0 \qquad \text{(KKT3)}$$

$$\lambda_i \geq 0 \qquad \text{(KKT4)}$$

$$\lambda_i x_i = 0, \text{ con } i = 1, 2, 3 \qquad \text{(KKT5)}$$

Ejemplo 2: KKT

Escriba las condiciones de optimalidad del siguiente problema (KKTs)

$$\begin{aligned} &\text{min } \frac{1}{2}\|\omega\|^2 \\ s.t.: \quad & y_i(\omega^T x_i + b) \geq 1 \end{aligned}$$

dados (x_i, y_i), $x_i \in R^n$, $y_i \in \{-1, 1\}$, $i = 1, 2, \cdots, N$.

Ejemplo 2: KKTs

Lagrangiano

$$\mathcal{L}(\omega, b, \lambda) \quad = \quad f(\omega) - \sum_{i=1}^{N} \lambda_i \left(y_i(\omega^T x_i + b) - 1 \right)$$

Ejemplo 2: KKTs

$$\nabla_{\omega,b}\mathcal{L}(\omega,b,\lambda)=0 \tag{KKT1}$$

$$y_i(\omega^T x_i+b)\geq 1,\ i=1,2\cdots,N; \tag{KKT3}$$

$$\lambda_i\geq 0,\ \ i=1,2\cdots,N; \tag{KKT4}$$

$$\lambda_i[y_i(\omega^T x_i+b)-1]=0,\ i=1,2\cdots,N; \tag{KKT5}$$

Ejemplo 2: KKTs

$$\omega - \sum_{i=1}^{N} \lambda_i y_i x_i = 0 \tag{KKT1-a}$$

$$-\sum_{i=1}^{N} \lambda_i y_i = 0 \tag{KKT1-b}$$

$$y_i(\omega^T x_i + b) \geq 1,\ i = 1, 2 \cdots, N; \tag{KKT3}$$

$$\lambda_i \geq 0,\ \ i = 1, 2 \cdots, N; \tag{KKT4}$$

$$\lambda_i[y_i(\omega^T x_i + b) - 1] = 0,\ i = 1, 2 \cdots, N; \tag{KKT5}$$

Cono

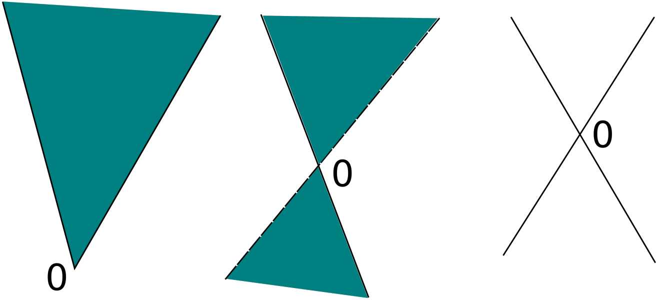
Cono

Sea $C \subseteq R^n$, decimos que C es un cono si $\lambda x \in C$ para todo $\lambda > 0$ y cada $x \in C$.
Un cono C es convexo si $\alpha x + \beta y \in C$, para escalares positivos $\alpha, \beta > 0$ y cualquier $x, y \in C$

- Los conos pueden ser convexos o no
- Si el cono es cerrado entonces $0 \in C$
- Ejemplos:
  1. $\{(x_1, x_2) |\ x_1 > 0$ y $x_2 \geq 0\}$
  2. $\{(x_1, x_2) |\ x_1 = 0$ o $x_2 = 0\}$
  3. $\{(x_1, x_2) |\ x_1 x_2 \geq 0\}$
  4. $Ax \geq 0$ donde A es una matriz, es un cono poliédrico convexo

Cono

a) Covexo b) No convexo c) No convexo

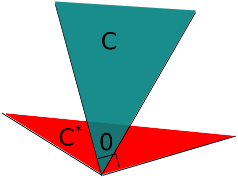


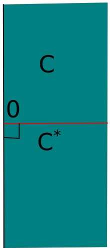
Cono Dual

Cono Dual

Sea $C \subseteq R^n$ un cono no vacío. Se llama cono dual o cono conjugado al conjunto

$$C^* = \{s \in R^n |\ \langle s, x\rangle \geq 0;\ \forall x \in C\}$$





Cono Polar

Cono Polar

Sea $C \subseteq R^n$ un cono no vacío. El polar de C es el conjunto

$$C^o = \{s \in R^n | \langle s, y \rangle \leq 0; \ \forall y \in C\}$$

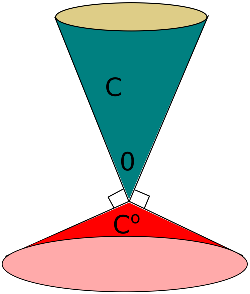
- La definición de Cono Polar depende del producto interior, de modo que si cambia el producto escalar cambia el cono polar
- Ejemplos

$$\begin{aligned} \langle s, x \rangle &= s^T x \\ \langle s, x \rangle &= s^T \Lambda x \end{aligned}$$

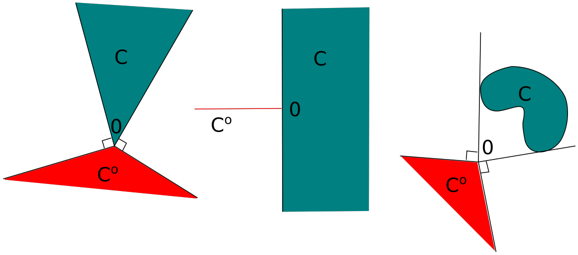
donde $\Lambda = \text{diag}\{\lambda_1, \cdots, \lambda_n\}$ donde $\lambda_i > 0$.

Cono Polar

- El polar de un cono convexo cerrado no vacío C el un cono convexo cerrado no vacío C^o, y $C^{oo} = C$.
- El cono polar is igual al negativo del cono dual, ie, $C^o = -C^*$



Cono Polar



Nota: La definición de cono polar se puede extender al caso de conjuntos, como en el tercer ejemplo.

Cono Normal

Cono Normal

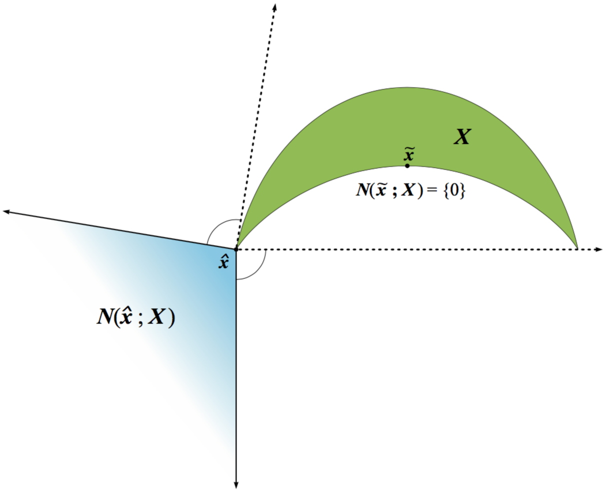
Sea $C \subseteq R^n$ un conjunto no vacío y $x \in C$. Un cono normal al conjunto C en el punto x es el conjunto

$$N_C(x) = \{s \in R^n | \; \langle s, y - x\rangle \leq 0; \; \forall y \in C\}$$

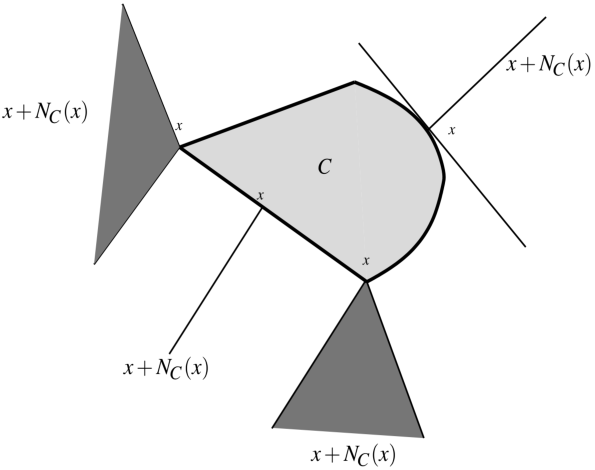
A los vectores de este conjunto se llaman vectores normales al conjunto C en x.

Si $x \in Interior(C)$, entonces $N_C(x) = \{0\}$.
El cono $N_C(x)$ es el lo mismo que el Polar C^o de C, considerando el sistema de coordenadas centrado en x.

Cono Normal



Cono Normal



Cono Tangente

Cono Tangente

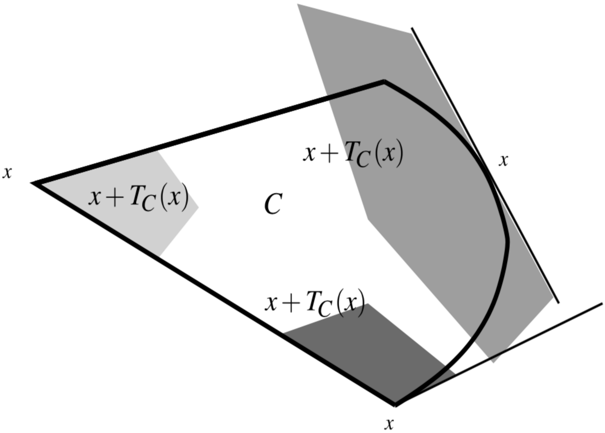
Sea $C \subseteq R^n$ un conjunto no vacío y $x \in C$. Un cono tangente al conjunto C en el punto x es el conjunto

$$T_C(x) = \{s \in R^n | \; s = \lim_{k\to\infty} \frac{x_k - x}{\lambda_k}; \; x_k \in C\}$$

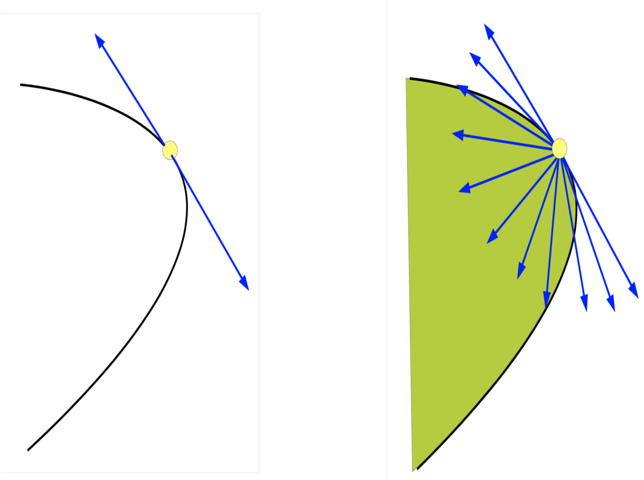
donde $\lim_{k\to\infty} x_k = x$, $\lim_{k\to\infty} \lambda_k = 0$ y $\lambda_k \geq 0$.
A los vectores de este conjunto se llaman vectores tangentes al conjunto C en x.

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

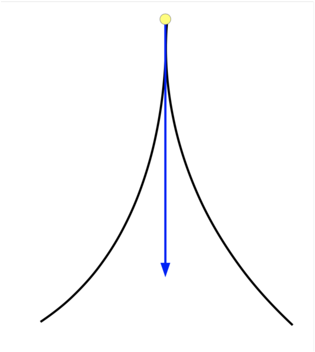
Cono Tangente

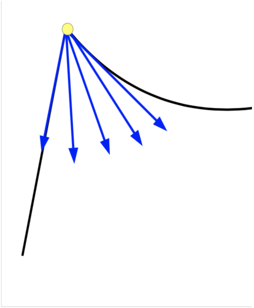


Cono Tangente



Cono Tangente





Direcciones factibles linearizadas

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

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

Es decir, es el conjunto de las direcciones factibles (condición de factibilidad) en las restricciones activas $\mathcal{A}(x)$.
Se puede verificar que $\mathcal{F}(x)$ es un cono

Comentario

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

Ejemplos:

$$\text{min } x_1 + x_2 \;\; s.t. \; x_1^2 + x_2^2 - 2 = 0 \qquad \text{(Ejemplo 1)}$$

$$\text{min } x_1 + x_2 \;\; s.t. \; (x_1^2 + x_2^2 - 2)^2 = 0 \qquad \text{(Ejemplo 2)}$$

Los problemas anteriores son equivalentes

Ejemplo 1

$$\text{min } x_1 + x_2 \;\; s.t. \; x_1^2 + x_2^2 - 2 = 0 \qquad \text{(Ejemplo 1)}$$

Cono Tangente: Consideremos las secuencias factibles,

$$\begin{aligned} z_k &= [-\sqrt{2 - t_k^2}, t_k]^T \\ y_k &= [-\sqrt{2 - t_k^2}, -t_k]^T \end{aligned}$$

con $t_k = \frac{1}{k} > 0$ y $t_k \to 0$, sea el punto $\hat{x} = [-\sqrt{2}, 0]^T$, entonces

$$\text{lim } \frac{z_k - x}{t_k} = [0, 1]^T; \;\; \text{lim } \frac{y_k - x}{t_k} = [0, -1]^T$$

Ejemplo 1

$$\text{min } x_1 + x_2 \;\; s.t. \; x_1^2 + x_2^2 - 2 = 0 \qquad \text{(Ejemplo 1)}$$

Cono Tangente: Luego el cono tangente en el punto $\hat{x} = [-\sqrt{2}, 0]^T$ es el conjunto

$$T_\Omega(\hat{x}) \;\; = \;\; \{d = [0, d_2] |\; d_2 \in R\}$$

Dado que las restricciones $x_1^2 + x_2^2 - 2 = 0$ y $(x_1^2 + x_2^2 - 2)^2 = 0$ representan el mismo conjunto, ie, el circulo de radio $\sqrt{2}$ con centro en el origen, entonces el cono tangente es el mismo en ambos casos.

Ejemplo 1

$$\min x_1 + x_2 \;\; s.t. \; x_1^2 + x_2^2 - 2 = 0 \qquad \text{(Ejemplo 1)}$$

Direcciones factibles linearizadas $\mathcal{F}(\hat{x})$ en el punto $\hat{x} = [-\sqrt{2}, 0]^T$

$$\nabla c(x) \quad = \quad 2x$$

Como $\mathcal{F}(\hat{x}) = \{d | \; d^T \nabla c(\hat{x}) = 0\}$ entonces

$$0 \quad = \quad d^T \nabla c(\hat{x}) = -2\sqrt{2} d_1 = 0$$

es decir, $d_1 = 0$, luego $d = [0, d_2]$ con $d_2 \in R$ y por tanto $\mathcal{F}(\hat{x}) = T_\Omega(\hat{x})$ para el ejemplo 1.

Ejemplo 2

$$\text{min } x_1 + x_2 \;\; s.t. \; (x_1^2 + x_2^2 - 2)^2 = 0 \qquad \text{(Ejemplo 2)}$$

Direcciones factibles linearizadas $\mathcal{F}(\hat{x})$ en el punto $\hat{x} = [-\sqrt{2}, 0]^T$

$$\nabla c(x) \;\; = \;\; 2(\|x\|^2 - 2)x$$

Como $\mathcal{F}(\hat{x}) = \{d |\; d^T \nabla c(\hat{x}) = 0\}$ entonces

$$0 \;\; = \;\; d^T \nabla c(\hat{x}) = 2(\|\hat{x}\|^2 - 2) d^T \hat{x} == 0$$

pues $\|\hat{x}\|^2 - 2 = 0$, luego $d \in R^2$,ie, $\mathcal{F}(\hat{x}) = R^2$, y por tanto $\mathcal{F}(\hat{x}) \neq T_\Omega(\hat{x})$ para el ejemplo 2.

Relacion entre el cono tangente y las direcciones factibles linearizadas

Lema

Sea x^* un punto factible, entonces

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

Prueba de la primera parte

Si $d \in T_\Omega(x^*)$ entonces existen secuencias factibles $z_k \to x^*$ y $t_k \to 0$, $t_k > 0$ tal que

$$\lim \frac{z_k - x^*}{t_k} \;=\; d$$

Luego $z_k = x^* + t_k d + o(t_k)$, por definición de $o(\cdot)$.
Para la prueba consideremos los dos casos: $i \in \mathcal{E}$ y $i \in \mathcal{A}(x^*) \cap \mathcal{I}$.

Prueba de la primera parte

$i \in \mathcal{E}$: $c_i(z_k) = 0$ pues la secuencia z_k es factible, y usando Taylor

$$\begin{aligned} c_i(z_k) &= c_i(x^* + t_k d + o(t_k)) \\ &= c_i(x^*) + \nabla c_i(x^*)^T (t_k d + o(t_k))) + o(t_k) \\ &= t_k \nabla c_i(x^*)^T d + o(t_k) \end{aligned}$$

Luego $0 = t_k \nabla c_i(x^*)^T d + o(t_k)$ dividiendo por t_k

$$0 = \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}$$

y pasando al limite $\nabla c_i(x^*)^T d = 0$, es decir, $d \in \mathcal{F}(x^*)$

Prueba de la primera parte

$i \in \mathcal{A}(x^*) \cap \mathcal{I}$: $c_i(z_k) \geq 0$ pues la secuencia z_k es factible, y $c_i(x^*) = 0$. Usando Taylor

$$\begin{aligned} c_i(z_k) &= c_i(x^* + t_k d + o(t_k)) \\ &= c_i(x^*) + \nabla c_i(x^*)^T (t_k d + o(t_k))) + o(t_k) \\ &= t_k \nabla c_i(x^*)^T d + o(t_k) \end{aligned}$$

Luego $0 \leq t_k \nabla c_i(x^*)^T d + o(t_k)$ dividiendo por t_k

$$0 \leq \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}$$

y pasando al limite $\nabla c_i(x^*)^T d \geq 0$, es decir, $d \in \mathcal{F}(x^*)$.

Prueba de la primera parte

Basado en la prueba de los casos anteriores, se concluye la primera parte, ie, $T_\Omega(x^*) \subset \mathcal{F}(x^*)$

Prueba de la segunda parte

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

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

Prueba de la segunda parte

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

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

y la curva de nivel o parametrizacion $F(t,z) = 0$. Debido a que

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

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

Prueba de la segunda parte

Como $F(t_k, z_k) = 0$, $z_k = z(t_k)$, entonces

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

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

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

Esto ultimo pues $d \in \mathcal{F}(x^*)$ y $t_k > 0$

Prueba de la segunda parte

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

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

y como B es invertible $0 = z_k - x^* - t_k d + o(\|z_k - x^*\|)$ es decir $z_k \in T_\Omega(x^*)$

Teorema

Una solucion local del problema original es un punto x en el que todas las secuencias factibles satisfacen la propiedad $f(z_k) \geq f(x)$ para k suficientemente grande.
El siguiente resultado muestra que si dichas secuencias existen entonces el producto interior de las direcciones limites y el gradiente de la funcion objetivo es no negativo

Teorema

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

Prueba

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

$$z_k \quad = \quad x^* + t_k d + o(t_k)$$

Usando Taylor

$$f(z_k) \quad = \quad f(x^*) + t_k \nabla f(x^*) d + o(t_k)$$

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

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:

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

Luego por definicion de limite 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 $k > K$

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

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

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

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

Nota

Existe K tal que para tada $k > K$

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

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

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

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