Teoría y Algoritmia de Optimización

Métodos de Lagrange

Agenda

- Teorema de Lagrange: Condiciones necesarias
- El Lagrangeano
- Condiciones suficientes de optimalidad
- Análisis de sensibilidad
- Métodos basados en el Lagrangeano

"Liberté, egalité, optimalité" - Joseph-Louis Lagrange
También conocido como Giuseppe Lodovico Lagrangia

El Lagrangeano y las restricciones de igualdad

Geometría de problemas con restricciones de igualdad

$$\min_{x} \ f(x)$$

$$\text{sujeto a} \ \ h(x) = 0$$

espacio reservado para dibujo: curvas de nivel

Direcciones de variación factibles

$$\min_{x} \quad f(x)$$

$$\text{sujeto a} \quad h(x) = 0$$

- h(x) diferenciable

$$\{d : \nabla h(x^*)^T d = 0\}$$

espacio reservado para dibujo: direcciones factibles

Optimalidad y direcciones factibles

$$\min_{x} \ f(x)$$

$$\text{sujeto a} \ \ h(x) = 0$$

x^* óptimo, f(·) diferenciable, *d* dirección factible:

$$f(x^* + d) - f(x^*) = \nabla f(x^*)^T d + o(d) \to 0$$

En el límite: $\nabla f(x^*)^T d = 0$

Intuición, en resumen

1. $h(.)$ dif., d dir. factible: $\nabla h(x^*)^T d = 0$
2. x^* óptimo, $f(\cdot)$ dif., d dir. factible: $\nabla f(x^*)^T d = 0$
3. Combinando 1 y 2: $\nabla h(x^*) || \nabla f(x^*)$

$$\exists \lambda^* : \nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$$

Intuición, en resumen

espacio reservado para dibujo: intuición, en resumen

$$\exists \lambda^* : \nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$$

Regularidad ~ Regularité

espacio reservado para dibujo: regularidad

$$\{\nabla h_i(x^*) : i = 1, \ldots, m\} \ \ell.i. \qquad \exists\lambda : \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) = 0$$

Condiciones de segundo orden

espacio reservado para dibujo: segundo orden

$$d^T \left[\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x) \right] d \geq 0 \qquad \forall\, d : \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m$$

Condiciones Necesarias de Optimalidad

Condiciones necesarias de optimalidad

Primer orden

Si se cumple:

- x^* es mínimo local
- *h(x)* y *f(x)* diferenciables
- $\{\nabla h_i(x^*) : i = 1, \ldots, m\}$ $\ell.i.$

Entonces:

- $\exists! \lambda : \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) = 0$

Segundo orden

Si además:

- *h(x)* y *f(x)* doblemente dif.

Entonces:

$$d^T \left[\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x) \right] d \geq 0$$

$$\forall\, d : \nabla h_i(x^*)^T d = 0, i = 1, \ldots, m$$

Condiciones necesarias: demostración

Técnica

1. Aproximamos problema original **con restricciones**
2. Por sucesión de problemas auxiliares **sin restricciones**
3. Estudiamos **sucesión de soluciones** $\{x^k\}$ **del problema auxiliar**
4. Comparamos las condiciones opt. de x^* y $\{x^k\}$ en el límite

Condiciones necesarias: paso 1: problema auxiliar

Por **hipótesis**:

$$x^* = \arg\min_x f(x) \text{ s.a. } h(x) = 0$$

Función de costo **auxiliar**:

$$F^k(x) = f(x) + \frac{k}{2}\|h(x)\|_2^2 + \frac{\alpha}{2}\|x - x^*\|_2^2$$

Definimos la bola:

$$S = \{x : \|x - x^*\|_2 \leq \epsilon\}$$

Definimos:

$$x^k = \arg\min_x F^k(x) \quad \text{s.a.} \quad x \in S$$

espacio reservado para dibujo

Condiciones necesarias: paso 2: límites

Como $x^k = \arg\min_x F^k(x)$ s.a. $x \in S$, se cumple:

$$F^k(x^k) = f(x^k) + \frac{k}{2}\|h(x^k)\|_2^2 + \frac{\alpha}{2}\|x^k - x^*\|_2^2 \leq F^k(x^*) = f(x^*)$$

Como $f(x^*)$ es **acotada** en S:

$$\lim_k \|h(x^k)\| = 0$$

Por tanto, todo punto límite $\bar{x}$ de $\{x^k\}$ cumple $h(\bar{x}) = 0$

Condiciones necesarias: paso 2

Como $x^k = \arg\min F^k(x)$ s.a $x \in S$, se cumple:

$$F^k(x^k) = f(x^k) + \frac{k}{2}\|h(x^k)\|_2^2 + \frac{\alpha}{2}\|x^k - x^*\|_2^2 \leq F^k(x^*) = f(x^*)$$

En particular

$$f(x^k) + \frac{\alpha}{2}\|x^k - x^*\|_2^2 \leq F^k(x^*) = f(x^*)$$

En el límite $k \to \infty$

$$f(\bar{x}) + \frac{\alpha}{2}\|\bar{x} - x^*\|_2^2 \leq f(x^*)$$

Condiciones necesarias: paso 2

De lo anterior, tenemos **i)** $f(\bar{x}) + \frac{\alpha}{2}\|\bar{x} - x^*\|_2^2 \leq f(x^*)$

Como $h(\bar{x}) = 0$, $\bar{x}$ es **factible,** y $\bar{x} \in S$ tenemos que **ii)** $f(x^*) \leq f(\bar{x})$

De **(i)** y **(ii)**: $f(x^*) = f(\bar{x}) \Rightarrow \|x^* - \bar{x}\| = 0 \Rightarrow x^* = \bar{x}$

Entonces, efectivamente vemos que $\{x^k\}$ **converge a** x^*

Entonces x^k pertenece al **interior de *S*** para *k* suficientemente grande

O sea que x^k (a partir de cierto *k*) es solución del problema **sin restricciones:**

$$\min_x F^k(x)$$

Condiciones necesarias: paso 2: resumen

- Obtuvimos sucesión $\{x^k\}$ que converge a x^*
- A partir de cierto k, x^k es solución del problema **sin restricciones**: $\min_x F^k(x)$
- Ahora podemos analizar las condiciones de optimalidad **sin restricciones** de x^k
- Como x^k converge a x^*, eso nos dará las condiciones del problema original **con restricciones**

espacio reservado para dibujo

Condiciones necesarias: paso 3

De la optimalidad de x^k tenemos **(i)**:

$$0 = \nabla F^k(x^k) = \nabla f(x^k) + \nabla h(x^k) kh(x^k) + \alpha(x^k - x^*)$$

Nuestro objetivo es llegar a **(ii)**:

$$0 = \nabla f(x^*) + \nabla h(x^*) \lambda^*$$

Cómo vamos de **(i)** a **(ii)**?

- En el límite de *k*, el último término ya vimos que va a 0
- Lo que resta es mostrar que: $\lambda^* = kh(x^*)$

Condiciones necesarias: paso 3

Vamos a despejar $\lambda^* = kh(x^*)$

$$\nabla f(x^k) + \nabla h(x^k) kh(x^k) + \alpha(x^k - x^*) = 0$$

$$\nabla h(x^k)^T \left[\nabla f(x^k) + \nabla h(x^k) kh(x^k) + \alpha(x^k - x^*)\right] = 0$$

$$\nabla h(x^k)^T \nabla h(x^k) kh(x^k) = -\nabla h(x^k)^T \left[\nabla f(x^k) + \alpha(x^k - x^*)\right]$$

Condiciones necesarias: paso 3

Teníamos

$$\nabla h(x^k)^T \nabla h(x^k) kh(x^k) = -\nabla h(x^k)^T \left[\nabla f(x^k) + \alpha(x^k - x^*)\right]$$

Como x^* es **regular,** $\nabla h(x^*)$ tiene rango m

Por continuidad, $\nabla h(x^k)$ también tiene rango m para k suf. grande

Entonces $\nabla h(x^k)^T \nabla h(x^k)$ es **invertible,** por lo que podemos despejar:

$$kh(x^k) = -\left[\nabla h(x^k)^T \nabla h(x^k)\right]^{-1} \nabla h(x^k)^T \left[\nabla f(x^k) + \alpha(x^k - x^*)\right]$$

Condiciones necesarias: primer orden

Tomando el límite en *k* y definiendo $\lambda^* = kh(x^*)$ llegamos a que:

$$\lambda^* = -\left[\nabla h(x^*)^T \nabla h(x^*)\right]^{-1} \nabla h(x^*)^T \nabla f(x^*)$$

Lo que prueba la condición necesaria de primer orden:

$$\exists! \, \lambda^* : \nabla f(x^*) + {\lambda^*}^T \nabla h(x^*)$$

Nos queda ver las condiciones de segundo orden...

Condiciones necesarias: segundo orden

- Ya vimos que, para *k* suficientemente grande, x^k es un mínimo local del problema **sin restricciones:** $\min_x F^k(x)$
- Entonces, en esas condiciones, cumple las condiciones de segundo orden:

$$d^T \nabla^2 F^k(x^k) d \geq 0, \ \forall\, d$$

- Desarrollando la Hessiana, tenemos que, para cualquier *d*:

$$d^T \left[\nabla^2 f^k(x^k) + \sum_{i=1}^{m} h_i(x^k) \nabla^2 h_i(x^k) + \alpha I \right] d \geq 0, \ \forall\, d$$

Condiciones necesarias: segundo orden

- Tenemos que: $d^T \left[\nabla^2 f^k(x^k) + \sum_{i=1}^{m} h_i(x^k) \nabla^2 h_i(x^k) + \alpha I \right] d \geq 0, \, \forall \, d$
- En particular, se cumple para:

$$d^T \left[\nabla^2 f^k(x^k) + \sum_{i=1}^{m} h_i(x^k) \nabla^2 h_i(x^k) + \alpha I \right] d \geq 0, \; \forall \; \{d : \nabla h(x^k) d = 0\}$$

- Llevando *k* al límite, tenemos: i) $x^k \to x^*$, y ii) $kh(x^k) \to \lambda^*$, o sea:

$$d^T \left[\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) + \alpha I \right] d \geq 0, \; \forall \; \{d : \nabla h(x^*) d = 0\}$$

Condiciones necesarias: segundo orden

- Tomando $\alpha \to 0$ (algo que podemos hacer sin problemas) llegamos a la condición de segundo orden que queríamos demostrar:

$$d^T \left[\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) \right] d \geq 0, \; \forall \; \{d : \nabla h(x^*) d = 0\}$$

El Lagrangeano

El Lagrangeano

$$L(x, \lambda) = f(x) + \lambda h(x)$$

Aplicaciones:

1. Análisis de condiciones de optimalidad
2. Análisis de sensibilidad
3. Algoritmos de optimización
4. Dualidad

El Lagrangeano: caso general

$$x^* = \arg\min_x f(x)$$

$$\text{s. a.}$$

$$h_i(x) = 0, i = 1, \ldots, m$$

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

Lagrangeano y Condiciones necesarias

$$x^* = \arg\min_x f(x)$$

$$\text{s. a.}$$

$$h_i(x) = 0, i = 1, \ldots, m$$

$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

El Lagrangeano y condiciones necesarias

$$x^* = \arg\min_x f(x)$$
s. a.
$$h_i(x) = 0, i = 1, \ldots, m$$

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

$i)\ \nabla_x L(x^*, \lambda^*) = 0$

$ii)\ \nabla_\lambda L(x^*, \lambda^*) = 0$

$iii)\ d^T \nabla^2_{xx} L(x^*, \lambda^*) d \geq 0$

$\forall\, d : \nabla h_i(x^*)^T d = 0, \forall\, i$

El Lagrangeano y condiciones necesarias

Observación:

(x^*, λ^*) cumple con las condiciones necesarias del problema:

$$\min_{(x,\lambda)} L(x, \lambda) \;\; \text{s.a.} \;\; h(x) = 0$$

Condiciones Suficientes de Optimalidad

Condiciones suficientes de optimalidad

$$x^* = \quad \arg\min_x f(x)$$
$$\text{s. a.}$$
$$h_i(x) = 0, i = 1, \ldots, m$$

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

$i)\ \nabla_x L(x^*, \lambda^*) = 0$

$ii)\ \nabla_\lambda L(x^*, \lambda^*) = 0$

$iii)\ d^T \nabla^2_{xx} L(x^*, \lambda^*) d > 0$

$\forall\, d \neq 0 : \nabla h(x^*)^T d = 0$

Nota: No es necesario exigir **regularidad**!

Condiciones suficientes: demostración

La **técnica** es muy similar a la que usamos para el caso de condiciones necesarias:

- Definimos una función auxiliar con **penalizaciones**
  - Se penaliza fuertemente alejarse de la restricción: $h(x) = 0$
- Analizamos las condiciones suficientes de optimalidad de la función auxiliar en el caso **sin restricciones**
- Con la elección apropiada de parámetros, veremos que esas condiciones implican también la optimalidad para el problema objetivo **con restricciones**

Condiciones suficientes: lema auxiliar

- Al igual que en el teorema anterior, necesitamos soluciones **únicas**
- Por eso, necesitamos construir un problema **estrictamente convexo**
- El siguiente resultado nos muestra cómo hacerlo (no lo vamos a demostrar)

Lema de la "positividad definida complementaria"

$$\left.\begin{array}{l} P = P^T,\ Q = Q^T,\ Q \succeq 0 \\ x^T P x > 0,\ \forall x : x^T Q x = 0 \end{array}\right\} \Rightarrow \exists c : P + cQ \succ 0$$

“Positividad definida complementaria”

espacio reservado para dibujo: positividad complementaria

El Lagrangeano aumentado

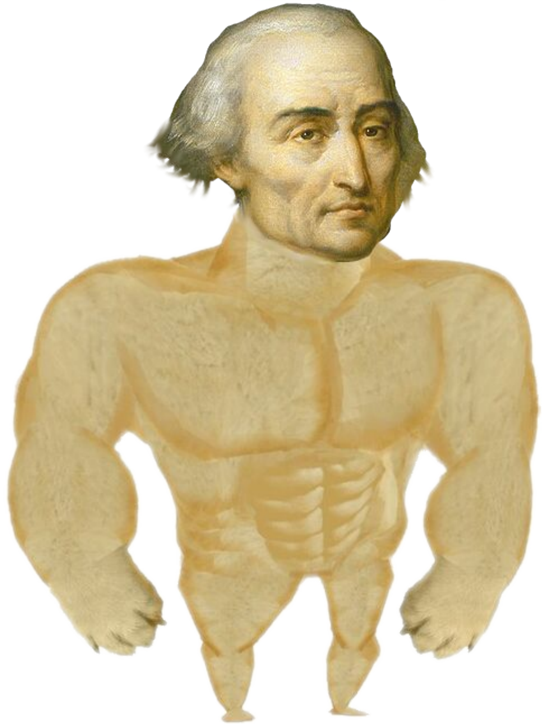
El Lagrangeano Aumentado es la función:

$$L_c(x,\lambda) = f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|_2^2$$

Es el Lagrangeano del **problema**

$$\min f(x) + \frac{c}{2}\|h(x))\|_2^2 \text{ s.a. } h(x) = 0$$

(el valor de **c** es el que nos va a dar la positividad complementaria más adelante)



Demostración (1)

Primera observación: el **problema auxiliar**

$$\min f(x) + \frac{c}{2}\|h(x))\|_2^2 \ \text{ s.a. } \ h(x) = 0$$

es (trivialmente) equivalente al **problema original**:

$$\min f(x) \ \text{ s.a. } \ h(x) = 0$$

Demostración (2)

- Gradiente de $L_c(x, \lambda)$:

$$\nabla_x L_c(x, \lambda) = \nabla f(x) + \nabla h(x)[\lambda + ch(x)]$$

- Por hipótesis tenemos que i) $h(x^*) = 0$ y ii) $\nabla_x L(x^*, \lambda^*) = 0$
- Sustituyendo en el gradiente:

$$\nabla_x L_c(x^*, \lambda^*) = \nabla f(x^*) + \nabla h(x^*)^T \lambda = \nabla_x L(x^*, \lambda^*) = 0$$

Demostración (3)

- Hessiana de $L_c(x, \lambda)$:

$$\nabla^2_{xx} L_c(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^{m} (\lambda_i + c h_i(x)) \nabla^2 h(x) + c \nabla h(x) \nabla h(x)^T$$

- Por hipótesis tenemos que i) $h(x^*) = 0$ y ii) $\nabla_x L(x^*, \lambda^*) = 0$. Sustituyendo:

$$\nabla^2_{xx} L_c(x^*, \lambda^*) = \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h(x^*) + c \nabla h(x^*) \nabla h(x^*)^T$$

$$\nabla^2_{xx} L_c(x^*, \lambda^*) = \nabla^2_{xx} L(x^*, \lambda^*) + c \nabla h(x^*) \nabla h(x^*)^T$$

Demostración (4)

Sigamos con la Hessiana. Llegamos a que:

$$\nabla^2_{xx} L_c(x^*, \lambda^*) = \nabla^2_{xx} L(x^*, \lambda^*) + c\nabla h(x^*)\nabla h(x^*)^T$$

También por hipótesis, tenemos que:

$$d^T \nabla^2_{xx} L(x^*, \lambda^*) d \; > \; 0 \; \forall d \neq 0 : h(x^*)^T d = 0$$

O lo que es lo mismo:

$$d^T \nabla^2_{xx} L(x^*, \lambda^*) d \; > \; 0 \; \forall d \neq 0 : d^T \nabla h(x^*) \nabla h(x^*)^T d = 0$$

Demostración (5)

Siguiendo con la Hessiana, ahora tenemos:

$$\nabla^2_{xx} L_c(x^*, \lambda^*) = \nabla^2_{xx} L(x^*, \lambda^*) + c \nabla h(x^*) \nabla h(x^*)^T$$

$$d^T \nabla^2_{xx} L(x^*, \lambda^*) d \; > \; 0 \; \forall d \neq 0 : d^T \nabla h(x^*) \nabla h(x^*)^T d = 0$$

Identificando:

$$P = \nabla^2_{xx} L(x^*, \lambda^*) \,, \quad Q = \nabla h(x^*) \nabla h(x^*)^T$$

Estamos en las condiciones del lema de “positividad definida complementaria”:

$$\exists c_0 : \nabla^2_{xx} L_c(x^*, \lambda^*) \succ 0, \; \forall \, c > c_0$$

Demostración (6)

Llegamos a que:

$$L_c(x) \geq L_c(x^*) + \frac{\gamma}{2}\|x - x^*\|_2^2, \;\; \forall\, x : \|x - x^*\| \leq \epsilon$$

Nuevamente, por hipótesis: $h(x^*) = 0 \Rightarrow L_c(x^*, \lambda^*) = f(x^*)$

De esas dos cosas, es inmediato que:

$$f(x) \geq f(x^*) + \frac{\gamma}{2}\|x - x^*\|_2^2, \;\; \forall\, x : h(x) = 0 \text{ y } \|x - x^*\| \leq \epsilon$$

O sea, x^* es un mínimo local estricto del problema original **con restricciones**.

Análisis de Sensibilidad

Condiciones de desigualdad

Condiciones de desigualdad

Vamos a estudiar la optimalidad de problemas del tipo

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & h_1(x) = 0, h_2(x) = 0, \ldots, h_m(x) = 0 \\ & g_1(x) \leq 0, g_2(x) \leq 0, \ldots, g_r(x) \leq 0 \end{array}$$

Nuevamente, la estrategia será reducir este escenario a un caso que ya conocemos, el de restricciones de igualdad.

El conjunto de restricciones activas (active set)

Para un punto factible cualquiera x, el *active set* se define como:

$$A(x) = \{j : g_j(x) = 0\}$$

espacio reservado para dibujo: active set

Restricciones activas y optimalidad

Definición

Si $j \notin A(x)$ decimos que la restricción $g_j(x)$ está *inactiva*

Observación

Si x^* es un mínimo local, entonces:

- Las restricciones inactivas *no importan (***es como si no existieran***)*
- Las restricciones activas son como **restricciones de igualdad**

Restricciones activas y optimalidad

Si x^* es un mínimo local, entonces:

- Las restricciones inactivas *no importan* (**es como si no existieran**)
- Las restricciones activas son como **restricciones de igualdad**

Observación: x^* también es mínimo local del problema:

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & h_1(x) = 0, h_2(x) = 0, \ldots, h_m(x) = 0 \\ & g_j(x) = 0, j \in A(x^*) \end{array}$$

Restricciones activas y optimalidad

Si x^* es un mínimo local y es además es regular de:

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & h_1(x) = 0, h_2(x) = 0, \ldots, h_m(x) = 0 \\ & g_j(x) = 0, j \in A(x^*) \end{array}$$

Entonces cumple las condiciones de optimalidad de Lagrange, es decir,

existen $\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*, \mu_j^*, j \in A(x^*)$ tales que:

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

Hacia las condiciones de optimalidad

Si x^* es un mínimo local y es además es regular, tenemos que

existen $\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*, \mu_j^*, j \in A(x^*)$ tales que:

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

Ahora definimos $\mu_j^* = 0, j \notin A(x^*)$, lo que nos permite expresar lo anterior como:

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

Hacia las condiciones de optimalidad

Importante: $\mu_j^* \geq 0$

espacio reservado para dibujo: no negatividad de los mu

Condiciones de Karush–Kuhn–Tucker (KKT)

Sea x^* min. local regular del problema:

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & h_1(x)=0, h_2(x)=0,\ldots,h_m(x)=0 \\ & g_1(x)\le 0, g_2(x)\le 0,\ldots,g_r(x)\le 0 \end{array}$$

Entonces existen $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*),\ \mu^* = (\mu_1^*, \mu_2^*, \ldots, \mu_r^*)$ tales que

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0 \qquad \begin{array}{ll} \mu_j^* \ge 0, & j \in A(x^*) \\ \mu_j^* = 0, & j \notin A(x^*) \end{array}$$

Si, además, f, g, h son C^2, entonces se cumple

$$d^T \nabla^2 L(x^*, \lambda^*, \mu^*) d \ge 0\,,\ \forall\, d : \left\{ \begin{array}{ll} \nabla h_i(x^*)^T d = 0, & \forall i \\ \nabla g_j(x^*)^T d = 0, & \forall j \in A(x^*) \end{array} \right.$$

Conversión de desigualdad en igualdad

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & h_1(x) = 0, h_2(x) = 0, \ldots, h_m(x) = 0 \\ & g_1(x) \leq 0, g_2(x) \leq 0, \ldots, g_r(x) \leq 0 \end{array}$$

Introducimos variables auxiliares $z_1, z_2, \ldots, z_r$:

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & h_1(x) = 0, h_2(x) = 0, \ldots, h_m(x) = 0 \\ & g_1(x) + z_1^2 = 0, g_2(x) + z_2^2 = 0, \ldots, g_r(x) + z_r^2 = 0 \end{array}$$

Conversión de igualdad en desigualdad

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & h_1(x) = 0, h_2(x) = 0, \ldots, h_m(x) = 0 \end{array}$$

Lo podemos reescribir como:

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & h_1(x) \leq 0, h_2(x) \leq 0, \ldots, h_m(x) \leq 0 \\ & -h_1(x) \leq 0, -h_2(x) \leq 0, \ldots, -h_m(x) \leq 0 \end{array}$$

Comentarios finales sobre condiciones de optimalidad

- La mayor limitante del método de Lagrange es la exigencia de **regularidad**
- Existen numerosos resultados (desde 1948 en adelante) que definen condiciones similares a las de Lagrange para casos particulares de problemas no regulares.
- No vamos a entrar en esos casos; el lector interesado puede recurrir al libro del curso (Bertsekas, Sección 3.3.3 en adelante).

Métodos de Optimización basados en el Lagrangeano

Métodos de Multiplicadores de Lagrange

- Vimos que podemos *caracterizar* el óptimo de un problema **con restricciones**...
- ... como el límite de una secuencia de problemas **sin restricciones**
- ¿Podemos usar esa misma idea para *hallar* ese óptimo?
- Concretamente: ¿podemos obtener el óptimo del problema con restricciones...
- ... como el límite de una secuencia de óptimos de problemas sin restricciones?

Spoiler: *si estamos en este slide, obviamente es porque la respuesta es sí*

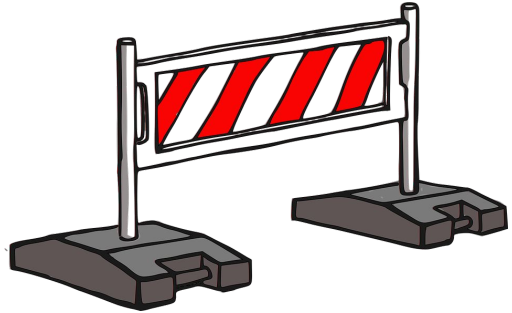
Métodos basados en los multiplicadores de Lagrange

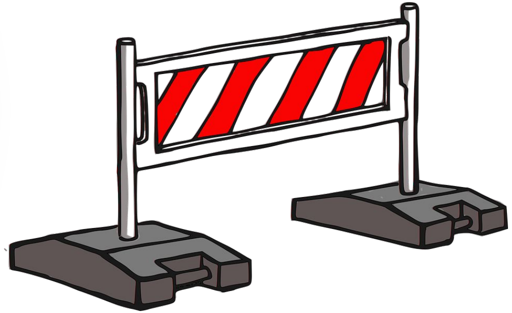
Tres tipos

- Métodos de barrera / punto interior
- Métodos de penalización (cuadrática)
- Método de los multiplicadores

Dos variantes

- Métodos exactos
- Métodos inexactos





Métodos de barrera

Métodos de punto interior

Idea

- Se aplican a problemas con restricciones de **desigualdad**
- Obtener una secuencia de soluciones que convergen a la solución
- Los puntos de la secuencia son **interiores** al conjunto factible

Métodos de barrera

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & g_1(x) \leq 0, g_2(x) \leq 0, \ldots, g_r(x) \leq 0 \end{array}$$

Resolvemos secuencia de problemas:

$$\text{minimizar} \;\; f(x) + \epsilon^k B_i(x)$$

Donde $0 < \epsilon^k < \epsilon^{k+1}$ y la *función de barrera* $B_i(x)$ es tal que

$B_i(x) \to \infty$ cuando $g_i(x) \to 0^-$

Ejemplos más comunes: $B_i(x) = -\log(-g_i(x))$ $\quad$ $B_i(x) = -1/g_i(x)$

Métodos de punto interior

Intuición

espacio reservado para dibujo: métodos de barrera

Método de barrera: aspectos prácticos

- Bastante fácil de implementar
- **Warm-restarts**:
  - Usar solución del paso *k* como punto inicial para el paso *k+1*!
  - Clave para la eficiencia!
  - Esto se aplica a todos los métodos que siguen (y a muchos otros)
- Al igual que otros métodos de penalización, puede sufrir de problemas de estabilidad numérica en los bordes

Método de penalización cuadrática

Métodos de penalización

Idea

- Se aplican a problemas con restricciones de **igualdad**
- Obtener una secuencia de soluciones que convergen a la solución
- Se sustituye la restricción por un término que penaliza su violación
- Cada nuevo problema hace pesar más la violación de la restricción

Método de penalización cuadrática exacto

Intuición

- Para c muy grande, el valor de $h(x)$ en $L_c(x,\lambda) = f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|_2^2$ va a tener un peso relativo muy grande
- Entonces, el mínimo global $x_c = \arg\min_x L_c(x,\lambda)$ va a estar muy cerca de

  $x^* = \arg\min_x f(x,\lambda)$ s.a. $h(x) = 0$ para *cualquier* valor de λ
- De hecho, se puede usar $\lambda = 0$
  - (en sus orígenes (1960s), sólo se usaba $\lambda = 0$)

Método de penalización cuadrática exacto

Pseudocódigo

1. Resolver $x_k = \arg\min_x L_{c^k}(x, \lambda^k)$ para $L_c(x, \lambda) = f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|_2^2$
2. Verificar condiciones de optimalidad (KKT)
   a. Se cumplen: terminar
   b. No se cumplen: incrementar k; aumentar c^k; volver a paso 1

Métodos de penalización

Aspectos prácticos

- Muy fácil de implementar
- Mal condicionado cuando c se hace muy grande

Método de penalización cuadrática inexacto

- El método de penalización cuadrática exacto converge a un óptimo local del problema original
- La demostración es muy simple
- Sin embargo, en la práctica, lo más común es se llegar a un x_k que cumple

$$\|\nabla_x L_{c^k}(x^k, \lambda^k)\| \leq \epsilon^k$$

- Veremos que, también en este caso, se llega al óptimo del problema original
  - (en el límite, por supuesto).

Convergencia del método de penalización *inexacto*

Si se cumple

- f(.) y g(.) C^2
- $\|\nabla_x L_{c^k}(x^k, \lambda^k)\| \leq \epsilon^k$ donde:
  - $\{\lambda^k\}$ es acotada
  - $0 < c^k < c^{k+1} \quad c^k \to \infty$
  - $\epsilon^k \to 0$
- $\{x^k\} \to x^*$ con $\mathrm{ran}(\nabla h(x^*)) = m$

Entonces

- $\{\lambda^k + c^k h(x^k)\}_K \to \lambda^*$
- $\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0, \quad h(x^*) = 0$

Demostración

1. Desarrollamos: $\nabla L_c^k(x^k, \lambda^k) = \nabla f(x^k) + (\lambda^k + ch(x^k))\nabla h(x^k)$

2. Definimos: $\tilde{\lambda}^k = \lambda^k + ch(x^k)$

3. Reescribimos $\nabla L_c^k(x^k, \lambda^k) = \nabla f(x^k) + \tilde{\lambda}^k \nabla h(x^k)$

4. Despejamos $\tilde{\lambda}^k = \left[\nabla h(x^k)^t \nabla h(x^k)\right]^{-1} \nabla h(x^k)^T \left[\nabla L_c^k(x^k, \lambda^k) - \nabla f(x^k)\right]$

Demostración

5. Teníamos: $\tilde{\lambda}^k = \left[\nabla h(x^k)^t \nabla h(x^k)\right]^{-1} \nabla h(x^k)^T \left[\nabla L_c^k(x^k, \lambda^k) - \nabla f(x^k)\right]$

6. Por hipótesis: $\nabla L_c^k(x^k, \lambda^k) \to 0$ entonces $\tilde{\lambda}^k \to \lambda^*$ en donde

$$\tilde{\lambda}^* = -\left[\nabla h(x^*)^t \nabla h(x^*)\right]^{-1} \nabla h(x^*)^T \nabla f(x^*)$$

7. En el límite: $\nabla L_{c^k}(x^*, \lambda^*) = \nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$

Demostración

8. Como $\{\lambda^k\}$ es acotada, y $\{\lambda^k + c^k h(x^k)\} \to \lambda^*$; entonces $\{c^k h(x^k)\}$ también es acotada.

9. Como $c^k \to \infty$, se debe cumplir $h(x^k) \to 0$

10. Como $\{x^k\} \to x^*$ entonces, en el límite, $h(x*) = 0$

Método de los multiplicadores

Métodos de los multiplicadores

Idea

- Se aplican a problemas con restricciones de **igualdad**
- Obtener una secuencia de soluciones que convergen a la solución
- Se sustituye la restricción por un término que penaliza su violación
- Cada nuevo problema tiene una mejor estimación del vector de multiplicadores
- En el límite, con los multiplicadores óptimos, la solución es tb. la óptima

Métodos de los multiplicadores

Intuición (ma non troppo)

- De la demostración anterior, vimos que $\tilde{\lambda}^k = \lambda^k + ch(x^k) \to \lambda^*$
- Entonces, podemos definir $\lambda^{k+1} = \lambda^k + ch(x^k) \approx \lambda^*$
- Eso es todo!
- **Mucho** más estable que el anterior
- Convergencia *lineal*
- Combinado con *c* creciente, convergencia *superlineal*

Método de penalización cuadrática exacto

Pseudocódigo

1. Elegir punto, constante y multiplicador iniciales y poner k=0
2. Resolver $x_k = \arg\min_x L_{c^k}(x, \lambda^k)$ para $L_c(x,\lambda) = f(x) + \lambda^T h(x) + \frac{c}{2}\|h(x)\|_2^2$
3. Verificar condiciones de optimalidad (KKT)
    a. Se cumplen: TERMINAR
4. Incrementar *k*
5. Incrementar c^k, actualizar $\lambda^{k+1} = \lambda^k + ch(x^k)$
6. Volver a 1

Comentarios finales

Métodos de los multiplicadores

- Vimos los tres métodos más básicos (y usados)
- Hay **muchos** otros métodos. Por ejemplo:
  - Métodos de penalización *exponencial* (resuelven problema de pen. cuadrática)
  - Métodos de penalización / multiplicadores de segundo orden
  - Métodos de penalización *exactos*
  - Resolución directa del sistema de ecuaciones de optimalidad
- En particular, en el **Tema 3** del curso veremos los **métodos duales**
- Al final del **Tema 4**, combinaremos el método de multiplicadores con los **métodos proximales** para desarrollar el método **ADMM**, de mucha vigencia