

# Optimizacion - EX

Nombre:  
Sebastian Lorca

Date:  
June 30, 2024

Profesor:  
Raimundo Cuadrado

Email:  
[sdlorca@uc.cl](mailto:sdlorca@uc.cl)

Sigla:  
ICS1113

Semestre:  
1/2024

Contents

<b>1</b>	<b>Programación Entera</b>	<b>2</b>
1.1	Relajación Lineal . . . . .	2
1.2	Branch and Bound . . . . .	2
1.2.1	Branch . . . . .	2
1.2.2	Bound . . . . .	3
1.3	Branch and Cut . . . . .	3
1.3.1	Cortes de Gomory . . . . .	3
1.3.2	Cortes de Cover . . . . .	3
<b>2</b>	<b>Programación No Lineal</b>	<b>4</b>
2.1	Optimización sin restricciones . . . . .	4
2.1.1	Condición Necesaria de Primer Orden . . . . .	4
2.1.2	Condición Necesaria de Segundo Orden . . . . .	4
2.1.3	Condición Suficiente de Segundo Orden . . . . .	4
2.1.4	Relación con Convexidad . . . . .	4
2.1.5	Método del Gradiente (o de Cauchy) (NO ENTRA) . . . . .	4
2.2	Optimización con restricciones . . . . .	5
2.2.1	Caracterización de Puntos Óptimos . . . . .	5
2.2.2	Condiciones de Lagrange . . . . .	5
2.2.3	Condiciones de Karush-Kuhn-Tucker . . . . .	6
<b>3</b>	<b>Problemas</b>	<b>6</b>
3.1	Lagrange . . . . .	6
3.2	KKT . . . . .	6
<b>4</b>	<b>Glosario</b>	<b>8</b>

# 1 Programación Entera

Un problema general de programación entera es:

$$\begin{array}{ll} \min & c^T x \\ \text{s.a.} & Ax = b \\ & x \geq 0 \\ & x \in \mathbb{Z}^n \end{array}$$

## 1.1 Relajación Lineal

La relajación lineal de un problema de programación entera es:

$$\begin{array}{lll} \min & z^* = c^T x & \sim & z^o = c^T x \\ \text{s.a.} & Ax = b & & Ax = b \\ & x \geq 0 & & x \geq 0 \\ & x \in \mathbb{Z}^n & & x \in \mathbb{R}^n \end{array}$$

Como el problema se relaja, se tiene que  $z^o \leq z^*$ , ya que el problema relajado tiene un dominio factible mayor.

## 1.2 Branch and Bound

Asumiendo un problema de programación entera  $S$  acotado, se puede ocupar la técnica de relajación lineal y obtener:

$$\begin{array}{ll} P_0) & z^o = c^T x \\ & Ax = b \\ & x \geq 0 \\ & x \in \mathbb{R}^n \end{array}$$

Y ocupando Simplex resolver el problema relajado, es decir,  $z_o = c^T x_o$ . Si  $x_o$  tiene alguna entrada fraccionaria, se puede ocupar la técnica de **Branch and Bound**.

### 1.2.1 Branch

Si  $x_o$  tiene alguna entrada fraccionaria, se puede seleccionar una variable  $x_j$  con valor fraccionario y se puede crear dos subproblemas:

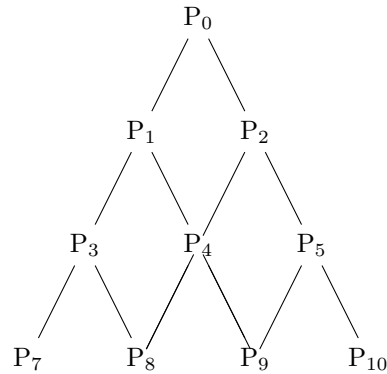
$$\begin{array}{ll} P_1) & z^o = c^T x \\ & Ax = b \\ & x \geq 0 \\ & x \in \mathbb{R}^n \\ & x_j \leq \lfloor x_j \rfloor \end{array}$$

$$\begin{array}{ll} P_2) & z^o = c^T x \\ & Ax = b \\ & x \geq 0 \\ & x \in \mathbb{R}^n \\ & x_j \geq \lceil x_j \rceil \end{array}$$

### 1.2.2 Bound

Si  $x_o$  no tiene ninguna entrada fraccionaria, se puede comparar  $z^o$  con el mejor valor entero encontrado hasta el momento, el *incumbente*. Si  $z^o$  es mejor, se actualiza el mejor valor entero encontrado.

Repitiendo el proceso de Branch and Bound, se puede encontrar la solución óptima del problema de programación entera generando un árbol de soluciones:



## 1.3 Branch and Cut

La técnica de **Branch and Cut** es una extensión de **Branch and Bound** que incluye cortes para acelerar la convergencia del algoritmo.

### 1.3.1 Cortes de Gomory

1. Se expresan los coeficientes de cada restricción como la suma de un entero y una fracción no negativa.
2. Se eliminan los términos enteros y se cambia  $=$  por  $\geq$ .

**Ejemplo:**

$$0x_1 - \frac{3}{4}x_2 + \frac{1}{3}x_3 - 3\frac{1}{4}x_4 = 10\frac{3}{7}$$

Paso 1:

$$(-1 + \frac{1}{4})x_2 + (0 + \frac{1}{3})x_3 + (-4 + \frac{3}{4})x_4 = 10 + \frac{3}{7}$$

Paso 2:

$$\frac{1}{4}x_2 + \frac{1}{3}x_3 + \frac{3}{4}x_4 \geq \frac{3}{7}$$

Más información [acá](#)

### 1.3.2 Cortes de Cover

Para restricciones que son la suma de variables binarias con coeficientes no negativos que sea menor o igual a un valor no negativo, entonces es posible agregar un corte de Cover.

Se expresa de la forma:

$$\sum_{j \in J} a_j x_j \leq b$$

Donde  $a_j \geq 0$  y  $b \geq 0$   $x \in \{0, 1\}^n$

Y el corte de Cover se expresa como:

$$\sum_{j \in J} x_j \leq |J| - 1$$

Información de MIP [acá](#).

## 2 Programación No Lineal

### 2.1 Optimización sin restricciones

Para clasificar funciones, se denominaran de clase  $C^k$  si todas sus derivadas parciales de orden  $k$  son continuas<sup>1</sup>.

#### 2.1.1 Condición Necesaria de Primer Orden

**Teorema de Fermat:** (Sea  $f(x)$  continuamente derivable) Si  $f(x)$  tiene un máximo o mínimo local en  $x_0$ , y  $f(x)$  es derivable en  $x_0$ , entonces  $f'(x_0) = 0$ .

**Teorema de Fermat generalizado:** (Sea  $f(x)$  continuamente derivable) Si  $f(x)$  tiene un máximo o mínimo local en  $x_0$ , y  $f(x)$  es derivable en  $x_0$ , entonces  $\nabla f(x_0) = 0$ .

#### 2.1.2 Condición Necesaria de Segundo Orden

Si  $x^*$  es mínimo local de  $f$ , entonces  $\nabla^2 f(x^*)$  es semidefinida positiva. Esto es llamado el **Criterio de la Segunda Derivada**.

#### 2.1.3 Condición Suficiente de Segundo Orden

Si  $f : D \rightarrow \mathbb{R}$  de clase  $C^2$ . Sea  $x^* \in D$  con  $\nabla f(x^*) = 0$  (punto crítico):

1. Si  $\nabla^2 f(x^*) = H(f(x^*))$  es definida positiva, entonces  $x^*$  es un mínimo local *estricto* de  $f$ .
2. Si  $\nabla^2 f(x^*) = H(f(x^*))$  es definida negativa, entonces  $x^*$  es un máximo local *estricto* de  $f$ . Para que sea definida negativa, las menores principales impares deben ser negativas y las pares positivas.
3. Si  $\nabla^2 f(x^*) = H(f(x^*))$  es indefinida, entonces  $x^*$  es un punto silla.

Si  $f$  es convexa, cualquier mínimo local  $x^*$  es mínimo global de  $f$ . Si esta es diferenciable, entonces cualquier punto crítico  $x^*$  es mínimo global de  $f$ .

#### 2.1.4 Relación con Convexidad

**Teorema:** Sea  $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  dos veces continuamente diferenciable, con  $D$  convexo. Entonces:

1.  $f$  es convexa si y solo si  $\nabla^2 f(x)$  es semidefinida positiva para todo  $x \in D$ .
2. Si  $\nabla^2 f(x)$  es definida positiva para todo  $x \in D$ , entonces  $f$  es estrictamente convexa.

Heloox

#### 2.1.5 Método del Gradiente (o de Cauchy) **(NO ENTRA)**

El método del gradiente es un método iterativo para encontrar un mínimo local de una función. Se dice que "la función decrece en la dirección del negativo del gradiente". Así:

$$f(x + \lambda d) < f(x) \quad \forall \lambda \in (0, r]$$

Donde  $d$  es la dirección de descenso. Si  $d^T \nabla f(x) < 0$ , entonces  $d$  es una dirección de descenso.

**Algoritmo:**

1. Sea  $x^0 \in \mathbb{R}^n$  un punto inicial,  $k = 0$ .
2. Sea  $d^k = -\nabla f(x^k)$  la dirección de descenso.
3. Si  $\|\nabla f(x^k)\| < \varepsilon$ , entonces  $x^k$  es un mínimo local.
4. Sino, resolver aproximadamente  $\min_{\lambda > 0} f(x^k + \lambda d^k)$ .
5. Sea  $x^{k+1} = x^k + \lambda d^k$ .  $k = k + 1$  (volver al paso 2).

Bajo condiciones razonables el modelo converge a un punto estacionario (punto crítico) de  $f$  (o a un punto cercano con el criterio de término). En caso que la función objetivo sea convexa, el método converge a un mínimo global. **Beneficio:** Es un método simple.

Si  $x^*$  es un mínimo local, entonces existe  $r > 0$  tal que si el punto inicial  $x^o$  cumple  $\|x^o - x^*\| < r$ , entonces el método converge a  $x^*$ . A esto se le llama **Convergencia Local**.

<sup>1</sup>Una función derivable es continua, pero una función continua no necesariamente es derivable.

## 2.2 Optimización con restricciones

### 2.2.1 Caracterización de Puntos Óptimos

**Definición:** Sea  $x \in S$ , un vector  $d \in \mathbb{R}^n$  es una dirección factible en  $S$  con respecto a  $x$  si existe  $\varepsilon > 0$  tal que  $x + td \in S, \forall t \in (0, \varepsilon]$ .

**Condiciones de Regularidad:** Sea  $x \in S$  e  $I(x) \neq \emptyset : \{i : g_i(x) = 0\}$  (conjunto de restricciones activas). Se dice que las funciones  $g_i, i \in I(x)$  cumplen la condición de regularidad en  $x$  si el conjunto de direcciones factibles es adherente al **cono tangente**<sup>2</sup> en  $x$ .

Así, se dice que  $x$  es **regular** para las restricciones de  $P$  si los vectores gradientes de las restricciones activas en  $x$  son linealmente independientes. También se pueden ocupar los tres criterios siguientes:

1. Se dice que  $x$  es regular si los gradientes de las restricciones activas en  $x$  son linealmente independientes.
2. Se dice que  $x$  es regular si el rango de la matriz jacobiana de  $g(x)$  en  $x$  tiene rango máximo<sup>3</sup>.
3. (Slater-Usawa) Se dice que  $x$  es regular si  $f$  y  $g_i$  son funciones convexas y además existe  $x^*$  tal que  $g_i(x^*) < 0, i \in I(x)$  para restricciones no lineales y  $g_i(x^*) \leq 0, i \in I(x)$  para restricciones lineales. Si se cumple, **todo el problema es regular**.

**1era Condición de Optimalidad:** Sea  $x^*$  un mínimo local de  $f$  en  $S$ . Si  $f$  es diferenciable en  $x^*$ , entonces  $\nabla f(x^*)^T d \geq 0$  para toda dirección factible  $d$  en  $S$  con respecto a  $x^*$ .

**1nda Condición de Optimalidad:** Si  $x^*$  es un mínimo local de  $P$  y se cumple la condición de regularidad en  $x^*$ , entonces:

$$\nabla^T f(x^*)d \leq 0, \quad \forall d \neq 0 : \nabla g_i(x^*)^T d \leq 0, i \in I(x^*)$$

Esto es,  $d$  es una dirección factible en  $S$  con respecto a  $x^*$ .

### 2.2.2 Condiciones de Lagrange

El **Teorema de Lagrange** para un problema:

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeto a} && g_i(x) = a_i \quad \forall i = 1, \dots, m \end{aligned}$$

Se puede escribir como:

$$L(x, \lambda) = f(x) + \sum_{i=1}^r \lambda_i (g_i(x) - a_i)$$

Donde  $\lambda_i$  son los **Multiplicadores de Lagrange**, que se entienden como aquellos que miden la sensibilidad del valor óptimo de la función objetivo al cambio de los recursos  $b_i$  de las restricciones en el problema. Estos multiplicadores, para un problema de  $m$  restricciones, se expresan como  $L(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_n) - \sum_{i=1}^m \lambda_i g_i(x_1, x_2, \dots, x_n)$ .

La solución óptima para los multiplicadores de lagrange se determina con las derivadas parciales de cada  $x_n$  y  $\lambda_i$  e igualando a cero. Así, se obtiene un sistema de ecuaciones que se resuelve para encontrar los valores óptimos de  $x_n$  y  $\lambda_i$ .

Para encontrar los puntos críticos de  $L(x, \lambda)$ , se deben resolver las siguientes ecuaciones:

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^r \lambda_i \nabla g_i(x) &= 0 \\ g_i(x) &= a_i \quad \forall i = 1, \dots, m \end{aligned}$$

Si  $x^*$  es un mínimo local...

Primero se deben evaluar las condiciones de regularidad. Para esto, se debe evaluar la matriz jacobiana de  $g(x)$  en  $x^*$ , y si esta matriz es de rango completo, entonces se cumple la condición de regularidad.

<sup>2</sup>El Cono Tangente es el cono generado por los gradientes de las restricciones que son activas en  $x$ .

<sup>3</sup>Al reducir la matriz Jacobiana, toda fila y columna tiene pivote

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

$$g_i(x^*) = a_i \quad \forall i = 1, \dots, m$$

Luego, se debe evaluar la matriz hessiana de  $L(x, \lambda)$  en  $(x^*, \lambda^*)$ , y si esta matriz es definida positiva, entonces  $x^*$  es un mínimo local.

Revisar 3.1

### 2.2.3 Condiciones de Karush-Kuhn-Tucker

Las condiciones de Lagrange son útiles para restricciones de igualdad, pero no para restricciones de desigualdad. Para esto, se ocupan las condiciones de Karush-Kuhn-Tucker (KKT).

Sea  $x \in D$ . Denotemos por  $I(x) = i : g_i(x) = b_i$  el conjunto de índices de las restricciones activas en el punto  $x$ .

#### Teorema de Karush-Kuhn-Tucker:

Un conjunto  $K \subset \mathbb{R}^n$  es convexo si y solo si para todo  $x_1, x_2 \in K$  y  $\lambda \in [0, 1]$ , se tiene que  $\lambda x_1 + (1-\lambda)x_2 \in K$ .

(Sea  $f(x)$  y  $g(x)$  continuamente derivables) Si  $x_0$  es un mínimo local de  $f(x)$  sujeto a  $g(x) = 0$ , entonces existen  $\lambda_0, \lambda_1, \dots, \lambda_m$  tales que:

$$\nabla f(x_0) + \sum_{i=1}^m \lambda_i \nabla g_i(x_0) = 0$$

$$\lambda_i g_i(x_0) = 0 \quad \forall i = 1, \dots, m$$

$$\lambda_i \geq 0 \quad \forall i = 1, \dots, m$$

## 3 Problemas

### 3.1 Lagrange

Extrahído de [libro de problemas de optimización](#).

$$z = f(x_1, x_2) = 2x_1x_2$$

$$\text{s.a. } x_1^2 + x_2^2 = 1$$

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda(g(x_1, x_2) - 1)$$

$$= 2x_1x_2 + \lambda(x_1^2 + x_2^2 - 1)$$

$$\frac{\partial L}{\partial x_1} = 2x_2 + 2\lambda x_1 = 0$$

$$\frac{\partial L}{\partial x_2} = 2x_1 + 2\lambda x_2 = 0$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_2^2 - 1 = 0$$

Resolviendo el sistema de ecuaciones, se obtiene que  $x_1 = \pm \frac{1}{\sqrt{2}}$  y  $x_2 = \pm \frac{1}{\sqrt{2}}$ .

### 3.2 KKT

Demuestre que para  $f$  diferenciable y  $x^*$  mínimo local, entonces  $\nabla f(x^*) = 0$ .

Asumiendo que  $\nabla f(x^*) \neq 0$ , entonces:

$$\begin{aligned} h &= -\nabla f(x^*) \\ g(t) &= f(x^* + th) \quad \frac{g}{dt} \\ g'(0) &= \nabla f(x^*)^T h \\ g'(0) &= \nabla f(x^*)^T (-\nabla f(x^*)) = -\|\nabla f(x^*)\|^2 < 0 \\ \therefore g'(0) &< 0 \end{aligned}$$

Como  $x^*$  es mínimo local, no puede haber un valor más pequeño, luego  $\Rightarrow \Leftarrow$ . Por lo tanto:  $\nabla f(x^*) = 0$ .

Al resolver un problema de Programación lineal hasta el óptimo se obtiene una solución en donde una variable no básica tiene costo reducido nulo. ¿Significa necesariamente esto que existe otro punto extremo del problema que también es solución óptima?

No, no necesariamente. Si bien es cierto que si una variable no básica tiene costo reducido nulo, entonces el problema tiene soluciones múltiples, no necesariamente todas las soluciones múltiples son óptimas.

$$\begin{aligned} \min \quad & x^2 + y^2 \\ \text{s.a.} \quad & xy \geq 4 \\ & x + y = 5 \end{aligned}$$

Condición de regularidad:

$$J[g(x^*)] = \begin{bmatrix} y & x \\ 1 & 1 \end{bmatrix}$$

Para todo  $x = y$  tengo rango completo. Así, el único punto no regular es (2.5, 2.5). Así, encontramos todos los puntos KKT.

Ahora estandarizamos el problema (despejamos a 0):

$$\begin{aligned} \min \quad & x^2 + y^2 \\ \text{s.a.} \quad & xy - 4 \geq 0 \\ & x + y - 5 = 0 \end{aligned}$$

Lagrangiano:

$$L(x, y, \mu, \lambda) = x^2 + y^2 + \mu(xy - 4) + \lambda(x + y - 5)$$

$$\frac{\partial L}{\partial x} = 2x + \mu y + \lambda = 0$$

$$\frac{\partial L}{\partial y} = 2y + \mu x + \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = x + y - 5 = 0$$

$$\frac{\partial L}{\partial \mu} \leq 0 \rightarrow 4 - xy \leq 0$$

$$\mu \frac{\partial L}{\partial \mu} = 0 \rightarrow \mu(xy - 4) = 0$$

$$\mu \geq 0$$

Reemplazamos  $\mu$  por 0:



$$\begin{aligned}
2x + \lambda &= 0 \\
2y + \lambda &= 0 \\
x + y - 5 &= 0 \\
x = y &= \frac{5}{2} \\
\lambda &= -5
\end{aligned}$$

Caso 2:  $\mu \neq 0$

$$\begin{aligned}
2x + \mu y + \lambda &= 0 \\
2y + \mu x + \lambda &= 0 \\
x + y - 5 &= 0 \\
xy - 4 &= 0 \\
\mu(xy - 4) &= 0 \\
xy &= 4
\end{aligned}$$

Se obtiene  $x = 4$ ,  $y = 1$  y  $\mu = 2$ . También se tiene  $x = 1$ ,  $y = 4$  y  $\mu = -1$ .

## 4 Glosario

### Derivadas parciales

Representan la tasa de cambio de una función en una dirección específica.  $\frac{\partial f}{\partial x_i}$

### Vector gradiente

Es un vector que contiene todas las derivadas parciales de una función.  $\nabla f(x) = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$

### Hessiano

Es la matriz de todas las segundas derivadas parciales de una función.  $H(x) = \nabla^2 f(x) = \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)$

### Matriz Semi-Definida Positiva

Una matriz  $A$  es semi-definida positiva si para todo vector  $x$ , se cumple que  $x^T A x \geq 0$ . Esto se puede demostrar si todos los valores propios de  $A$  son mayores o iguales a cero.

### Demostrar que una matriz es semi-definida positiva

Para demostrar que una matriz es semi-definida positiva, se pueden ocupar dos métodos:

**Método 1:** Calcular los valores propios de la matriz y verificar que todos son mayores o iguales a cero.

**Método 2:** Calcular el determinante de todas las submatrices principales de la matriz. Si todos los determinantes son mayores o iguales a cero, entonces la matriz es semi-definida positiva.

Las submatrices principales (o *menores principales*) son:

$$\begin{aligned}
A_1 &= [a_{11}] \\
A_2 &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\
A_n &= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}
\end{aligned}$$

## Expansión de Taylor

Es una aproximación de una función en un punto  $x_0$  mediante una serie de potencias.  $f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0) + \dots$

## Dirección de Descenso

Un vector  $d \neq 0$  se llama dirección de descenso en  $x$  si existe  $r > 0$  tal que  $f(x + td) < f(x)$  para todo  $t \in (0, r)$ .

Así, si se cumple que  $d^T \nabla f(x) < 0$ , entonces  $d$  es una dirección de descenso.