Escuela Técnica Superior de Ingeniería
Departamento de Organización Industrial

Optimización no lineal

José María Ferrer Caja
Universidad Pontificia Comillas

Planteamiento general

$$\boxed{\begin{array}{l} \min_{x} f(x) \\ g_i(x) \le 0 \qquad \forall i = 1,\ldots,m \\ f, g_i : \mathbb{R}^n \to \mathbb{R} \end{array}}$$

- La teoría se desarrolla para problemas de minimización, los problemas de maximización se tratan de forma análoga
- En un mismo problema puede haber restricciones de igualdad y de desigualdad
- Las funciones f, g_i pueden ser diferenciables o no serlo
- La resolución es más difícil y costosa que en optimización lineal
- Existen algoritmos específicos para ciertos problemas

Clasificación (1)

- Optimización sin restricciones

$$\boxed{\min_x f(x)}$$

  - ✓ Es importante distinguir si f es o no continua y/o diferenciable

- Optimización con restricciones lineales

$$\boxed{\begin{array}{l}\min_x f(x) \\ Ax = b\end{array}}$$

  - ✓ Existe una extensión del algoritmo simplex para resolverlo

- Programación cuadrática

$$\boxed{\begin{array}{l}\min_x c^T x + \frac{1}{2} x^T Q x \quad (Q \in \mathbb{R}^{n \times n}) \\ Ax = b\end{array}}$$

  - ✓ Aparece en muchas situaciones. Existen algoritmos eficientes

Clasificación (2)

- Programación convexa

$$\boxed{f, g_i \ \text{convexas}}$$

  - ✓ Todo mínimo local es global
  - ✓ Si el problema es de maximización, las funciones deben ser cóncavas

- Programación separable

$$\boxed{\begin{aligned} f(x) &= \sum_{j=1}^{n} f_j(x_j) \\ g_i(x) &= \sum_{j=1}^{n} g_{ij}(x_j) \end{aligned}}$$

  - ✓ Cumple la hipótesis de aditividad de la PL

Clasificación (3)

- Programación geométrica

$$\boxed{\begin{aligned} f(x), g_i(x) &= \sum_{j=1}^{n} c_j P_j(x) \\ P_j(x) &= x_1^{a_{j1}} x_2^{a_{j2}} \dots x_n^{a_{jn}}, \quad j = 1, \dots, n \end{aligned}}$$

  - ✓ Problemas de diseño en ingeniería

- Programación fraccional

$$\boxed{f(x) = \frac{f_1(x)}{f_2(x)}}$$

Expansión en serie de Taylor

- Aproximación del valor de una función por un polinomio
  - ✓ Se conoce $f(x_0)$. Se desea obtener $f(x)$

$$\boxed{f(x) = f(x_0 + p) \simeq f(x_0) + \nabla f(x_0)^T p + \frac{1}{2} p^T \nabla^2 f(x_0) p}$$

$$x, x_0 \in \mathbb{R}^n$$

$$p = x - x_0$$

∇f ⟸ Gradiente de la función

$\nabla^2 f$ ⟸ Matriz hessiana de la función

  - ✓ Una función polinómica coincide con su polinomio de Taylor

Expansión en serie de Taylor. Ejemplo

- Obtener el polinomio de Taylor de grado 2 de $f(x,y) = e^{2x-y}$
- Calcular aproximadamente $f(0.1,-0.2)$
  - ✓ Usamos el punto $(x_0,y_0) = (0,0)$ ⟹ $p^T = (x,y) - (0,0) = (x,y)$

$$f(x,y) = e^{2x-y} \Rightarrow f(0,0) = 1$$

$$\nabla f(x,y) = \begin{pmatrix} f_x \\ f_y \end{pmatrix} = \begin{pmatrix} 2e^{2x-y} \\ -e^{2x-y} \end{pmatrix} \Rightarrow \nabla f(0,0) = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

$$\nabla^2 f(x,y) = \begin{pmatrix} f_{xx} & f_{yx} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} 4e^{2x-y} & -2e^{2x-y} \\ -2e^{2x-y} & e^{2x-y} \end{pmatrix} \Rightarrow \nabla^2 f(0,0) = \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}$$

$$f(x,y) \simeq f(0,0) + \nabla f(0,0)^T p + \frac{1}{2} p^T \nabla^2 f(0,0) p =$$

$$= 1 + \begin{pmatrix} 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \boxed{1 + 2x - y + 2x^2 - 2xy + \frac{1}{2} y^2}$$

$$f(0.1,-0.2) \simeq 1 + 0.2 + 0.2 + 0.02 + 0.04 + 0.02 = \boxed{1.48}$$

Matrices simétricas. Menores

- Se llaman menores de una matriz a los determinantes de las submatrices cuadradas de la matriz
- Menores principales de una matriz (cuadrada): se obtienen cruzando k filas cualesquiera con sus respectivas columnas
  - ✓ Existen $2^n - 1$ menores principales en una matriz de orden n
- Menores de esquina de una matriz (cuadrada): se obtienen cruzando las k primeras filas con sus respectivas columnas
  - ✓ Existen n menores de esquina en una matriz de orden n
- La matriz cuadrada A (de orden n) es simétrica si $A^t = A$

Matrices definidas y semidefinidas positivas

- La matriz simétrica A (de orden n) es semidefinida positiva si $x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n$

  o bien

  todos sus autovalores son no negativos

  o bien

  todos sus menores principales son no negativos

- La matriz simétrica A (de orden n) es definida positiva si

$$x^T A x > 0 \quad \forall x \in \mathbb{R}^n - \{0\}$$

  o bien

  todos sus autovalores son positivos

  o bien

  todos sus menores de esquina son positivos

Matrices definidas y semidefinidas negativas

- La matriz simétrica A (de orden n) es semidefinida negativa si $x^T A x \leq 0 \quad \forall x \in \mathbb{R}^n$

  o bien

  todos sus autovalores son no positivos

  o bien

  -A es semidefinida positiva

- La matriz simétrica A (de orden n) es definida negativa si

  $$x^T A x < 0 \quad \forall x \in \mathbb{R}^n - \{0\}$$

  o bien

  todos sus autovalores son negativos

  o bien

  sus menores de esquina son alternativamente <0, >0, etc.

  o bien

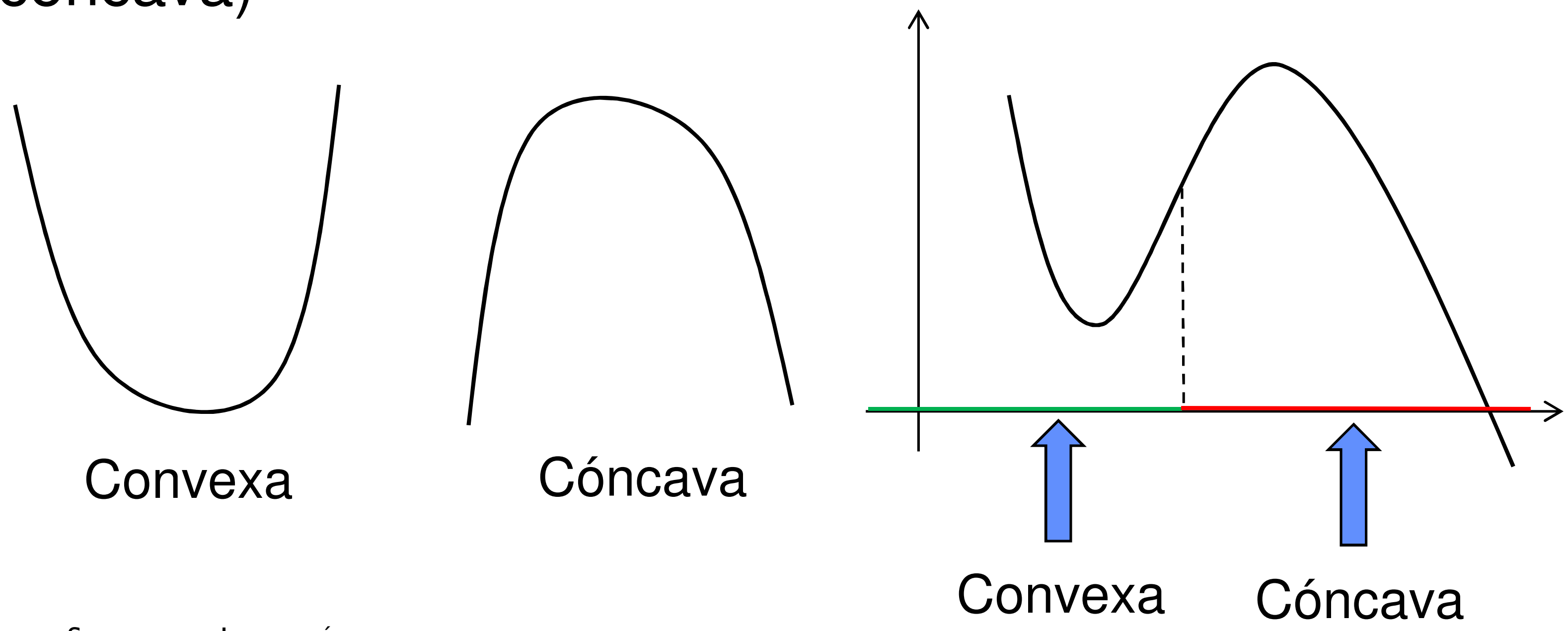
  -A es definida positiva

Funciones convexas y cóncavas

- La función f es convexa en un conjunto X si

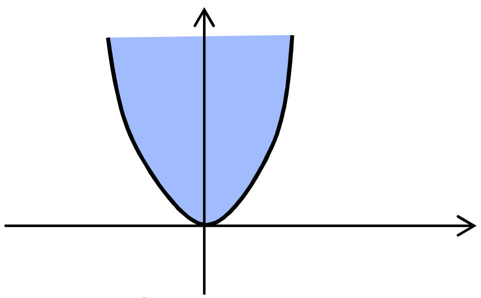
$$\forall x, y \in X, \forall \lambda \in [0,1] \quad f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

- La función f es cóncava si -f es convexa
- La función f es convexa (cóncava) en un punto x_0 si existe un entorno X de en el que la función es convexa (cóncava)



Funciones convexas y cóncavas. Hessiano

- La función f es convexa en el punto x_0 si la matriz hessiana es semidefinida positiva en x_0
- La función f es cóncava en el punto x_0 si la matriz hessiana es semidefinida negativa en x_0
- La región $f(x) \leq 0$ es convexa si la matriz hessiana es semidefinida positiva en todos sus puntos
- La región $f(x) \leq 0$ es cóncava si la matriz hessiana es semidefinida negativa en todos sus puntos



$$x^2 - y \leq 0$$

$$f(x,y) = x^2 - y$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

Semidefinida positiva

$\Downarrow$

Región convexa

Optimización sin restricciones y diferenciabilidad

- Condición necesaria de optimalidad (para funciones diferenciables)

$$\boxed{x^* \text{ es mínimo local de } f(x) \;\Rightarrow\; \nabla f(x^*) = 0}$$

- Condición necesaria de optimalidad (para funciones dos veces diferenciables)

$$\boxed{x^* \text{ es mínimo local de } f(x) \;\Rightarrow\; \begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ semidefinida positiva} \end{cases}}$$

- Condición suficiente de optimalidad (para funciones dos veces diferenciables)

$$\boxed{\left.\begin{array}{c} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ definida positiva} \end{array}\right\} \Rightarrow x^* \text{ es mínimo local estricto de } f(x)}$$

Optimización sin restricciones y convexidad

- Condición necesaria y suficiente de optimalidad (para funciones convexas y dos veces diferenciables)

$$\boxed{x^* \text{ es mínimo global de } f(x) \;\Leftrightarrow\; \nabla f(x^*) = 0}$$

✓ Ejemplo: $\min f(x,y) = x^2 + 3y^2 + 2xy - 4y$

$$\nabla f(x,y) = \begin{pmatrix} 2x + 2y \\ 6y + 2x - 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow (x,y) = (-1,1)$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 2 & 2 \\ 2 & 6 \end{pmatrix}$$ definida positiva $\forall (x,y)$

$\Downarrow$

f es convexa

$\Downarrow$

En (-1,1) se alcanza el mínimo global de f: f(-1,1) = - 2

Optimización con restricciones. Lagrangiano (1)

- Se tiene el problema

$$\boxed{\begin{array}{l} \min\limits_{x} f(x) \\ g_i(x) \le 0 \quad i = 1,\dots,m \\ h_j(x) = 0 \quad j = 1,\dots,l \\ x \in X \subseteq \mathbb{R}^n \end{array}}$$

- Su función lagrangiana o lagrangiano es:

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{l} \mu_j h_j(x)$$

- $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^l$ ⟵ Multiplicadores de Lagrange
  - ✓ $L(x,\lambda,\mu)$ proporciona una cota inferior de $f(x)$ para $\lambda \ge 0, \mu$
  - ✓ Interesa la cota inferior más ajustada $\max\limits_{\lambda \ge 0, \mu} L(x,\lambda,\mu)$

- Relajación lagrangiana $\boxed{\min\limits_{x \in X} \left\{ \max\limits_{\lambda \ge 0, \mu} L(x,\lambda,\mu) \right\}}$

Optimización con restricciones. Lagrangiano (2)

- Si todas las restricciones son de igualdad:
  - ✓ El mínimo del lagrangiano coincide con el mínimo de la función
  - ✓ En el mínimo, el gradiente de la función objetivo es combinación lineal de los gradientes de las restricciones
  - ✓ Los coeficientes de la c.l. son los multiplicadores
- Para el problema lineal $\boxed{\begin{array}{l}\min\limits_x c^T x \\ Ax = b\end{array}}$
  - ✓ Lagrangiano: $L(x,\mu) = c^T x + \mu^T (Ax - b)$
  - ✓ Aplicando la condición necesaria $\nabla L(x^*,\mu^*) = 0$

$$\begin{aligned}\nabla_x L(x,\mu) &= c + A^T \mu \\ \nabla_\mu L(x,\mu) &= Ax - b\end{aligned} \quad \Rightarrow \quad \begin{aligned} c + A^T\mu = 0 &\Rightarrow c = -A^T\mu \\ Ax - b = 0 &\Rightarrow Ax = b\end{aligned}$$

  - ✓ Los multiplicadores son las variables duales cambiadas de signo

Condiciones de optimalidad de KKT (1)

- Condiciones necesarias de Karush-Kuhn-Tucker para problemas con restricciones de desigualdad
  - ✓ Sea el problema P
    $$\begin{array}{|l|}\hline \min_x f(x) \\ g_i(x) \le 0 \quad i = 1, \ldots, m \\ x \in X \\ \hline\end{array}$$
  - ✓ $X \subseteq \mathbb{R}^n$ es abierto y no vacío
  - ✓ Sea x^* un punto factible (verifica las restricciones)
  - ✓ Sea $I = \{ i \ / \ g_i(x^*) = 0 \}$
  - ✓ Sean $f, \{ g_i, i \in I \}$ diferenciables en x^*
  - ✓ Sean $\{ g_i, i \notin I \}$ continuas en x^*
  - ✓ Sean $\{ \nabla g_i(x^*) \}_{i \in I}$ linealmente independientes

Condiciones de optimalidad de KKT (2)

- Condiciones necesarias de Karush-Kuhn-Tucker para problemas con restricciones de desigualdad
  - ✓ x^* es mínimo local de P $\Rightarrow$ existen escalares $\{\lambda_i, i \in I\}$ tales que

$$\nabla f(x^*) + \sum_{i \in I} \lambda_i \nabla g_i(x^*) = 0$$
$$\lambda_i \geq 0 \quad \forall i \in I$$

  - ✓ Si además $\{g_i, i \notin I\}$ son diferenciables en x^*

$$\boxed{\begin{aligned} &\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0 \\ &\lambda_i g_i(x^*) = 0 \quad i = 1, \ldots, m \\ &\lambda_i \geq 0 \quad i = 1, \ldots, m \end{aligned}}$$

  - ✓ Las condiciones necesarias para máximo local son idénticas salvo el signo de los multiplicadores: $\lambda_i \leq 0$

Condiciones de KKT. Ejemplo (1)

$$\min_{x,y} f(x,y) = 9y - (x-5)^2$$
$$-x^2 + y \leq 0$$
$$-x - y \leq 0$$
$$x - 1 \leq 0$$

- Condiciones de KKT $\Rightarrow$

$$\begin{cases} -2(x-5) - 2x\lambda_1 - \lambda_2 + \lambda_3 = 0 \\ 9 + \lambda_1 - \lambda_2 = 0 \\ \lambda_1(-x^2 + y) = 0 \\ \lambda_2(-x - y) = 0 \\ \lambda_3(x - 1) = 0 \\ \lambda_1, \lambda_2, \lambda_3 \geq 0 \end{cases}$$

- Además se deben cumplir las restricciones originales de factibilidad

$$\begin{cases} -x^2 + y \leq 0 \\ -x - y \leq 0 \\ x - 1 \leq 0 \end{cases}$$

Condiciones de KKT. Ejemplo (2)

- Se resuelve distinguiendo para cada multiplicador que sea nulo o no

✓ $\lambda_2 = 0 \Rightarrow \lambda_1 = -9 < 0$ Imposible. Luego se tiene $\lambda_2 \neq 0$

✓ $\lambda_2 \neq 0 \Rightarrow y = -x$

- $\lambda_1 = 0 \Rightarrow \lambda_2 = 9$. Queda el sistema $\begin{cases} -2x + 10 - 9 + \lambda_3 = 0 \\ \lambda_3(x-1) = 0 \end{cases}$
  - $\lambda_3 = 0 \Rightarrow A = (1/2, -1/2), \lambda_A = (0, 9, 0)$
  - $\lambda_3 \neq 0 \Rightarrow B = (1, -1), \lambda_B = (0, 9, 1)$
- $\lambda_1 \neq 0 \Rightarrow y = x^2 \Rightarrow x^2 = -x \Rightarrow x(x+1) = 0$
  - $x = 0 \Rightarrow y = 0 \Rightarrow C = (0, 0), \lambda_C = (1, 10, 0)$
  - $x = -1 \Rightarrow y = 1 \Rightarrow \lambda_3 = 0 \Rightarrow \lambda_1 = -3, \lambda_2 = 6$ Imposible

- Se han obtenido 3 puntos KKT candidatos a mínimos:

$$A = (1/2, -1/2), \quad B = (1, -1), \quad C = (0, 0)$$

Condiciones de KKT. Ejemplo (3)

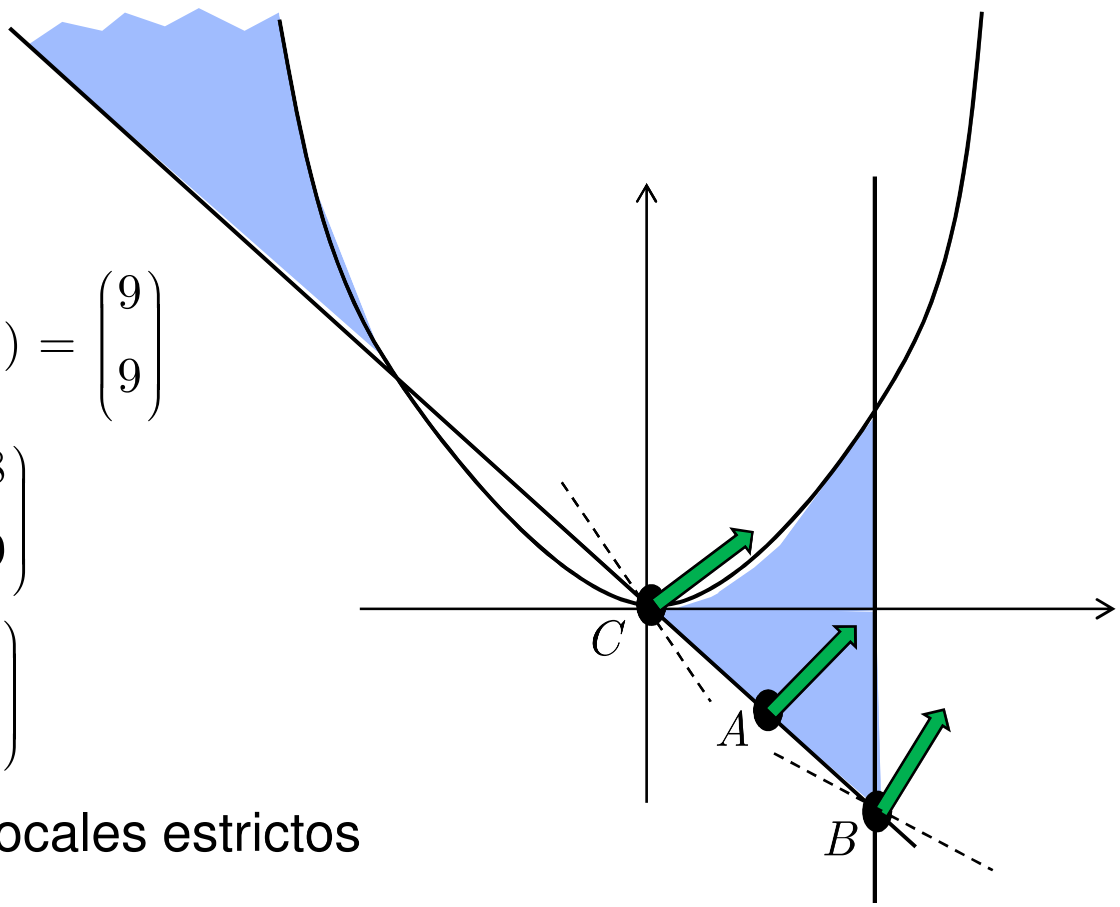
- Análisis gráfico:

$$\nabla f(x,y) = \begin{pmatrix} 10 - 2x \\ 9 \end{pmatrix}$$

$$\nabla f(A) = \nabla f(1/2, -1/2) = \begin{pmatrix} 9 \\ 9 \end{pmatrix}$$

$$\nabla f(B) = \nabla f(1, -1) = \begin{pmatrix} 8 \\ 9 \end{pmatrix}$$

$$\nabla f(C) = \nabla f(0, 0) = \begin{pmatrix} 10 \\ 9 \end{pmatrix}$$



- B y C son mínimos locales estrictos
- A no es nada
- No existe mínimo global, la función no es acotada

Condiciones de optimalidad de KKT (3)

- Condiciones suficientes de Karush-Kuhn-Tucker para problemas con restricciones de desigualdad
  - ✓ Sea x^* un punto factible (verifica las restricciones) de P
  - ✓ Sea $I = \left\{ i \ / \ g_i(x^*) = 0 \right\}$
  - ✓ Sean $f, \left\{ g_i, i \in I \right\}$ convexas y diferenciables en x^*

Si existen $\left\{ \lambda_i, i \in I \right\}$ tales que $\begin{cases} \nabla f(x^*) + \sum_{i \in I} \lambda_i \nabla g_i(x^*) = 0 \\ \lambda_i \geq 0 \quad \forall i \in I \end{cases}$

entonces x^* es mínimo local de P

Si además $f, \left\{ g_i, i \in I \right\}$ son convexas y diferenciables en todo el conjunto factible, entonces x^* es el mínimo global de P

- Para el caso de maximización, la función objetivo debe ser cóncava y los escalares $\lambda_i \leq 0 \ \forall i \in I$

Condiciones de optimalidad de KKT (4)

- Condiciones suficientes de Karush-Kuhn-Tucker para problemas con restricciones de desigualdad

  - ✓ Como alternativa a la convexidad se puede usar el lagrangiano restringido

$$L(x) = f(x) + \sum_{i \in I} \lambda_i^* g_i(x)$$

  - ✓ Si la matriz hessiana en el candidato x^*

$$\nabla^2 L(x^*) = \nabla^2 f(x^*) + \sum_{i \in I} \lambda_i^* \nabla^2 g_i(x^*)$$

    es semidefinida positiva , entonces x^* es un mínimo local de P

  - ✓ Si la matriz hessiana anterior es definida positiva, entonces x^* es un mínimo local estricto de P

Condiciones de optimalidad de KKT (5)

- Condiciones necesarias de Karush-Kuhn-Tucker para problemas con restricciones de desigualdad e igualdad
  - ✓ Sea el problema P'

$$
\boxed{\begin{array}{ll}
\min_x f(x) & \\
g_i(x) \le 0 & i = 1,\ldots,m \\
h_j(x) = 0 & j = 1,\ldots,l \\
x \in X &
\end{array}}
$$

  - ✓ $X \subseteq \mathbb{R}^n$ es abierto y no vacío
  - ✓ Sea x^* un punto factible (verifica las restricciones)
  - ✓ Sea $I = \{i \,/\, g_i(x^*) = 0\}$
  - ✓ Sean $f, \{g_i, i \in I\}$ diferenciables en x^*
  - ✓ Sean $\{g_i, i \notin I\}$ continuas en x^*
  - ✓ Sean $\{h_j, j = 1,\ldots,l\}$ continuamente diferenciables en x^*
  - ✓ Sean $\{\nabla g_i(x^*), i \in I; \nabla h_j(x^*), j = 1,\ldots,l\}$ l.i.

Condiciones de optimalidad de KKT (6)

- Condiciones necesarias de Karush-Kuhn-Tucker para problemas con restricciones de desigualdad e igualdad
  - ✓ x^* es mínimo local de P' $\Rightarrow$ existen escalares

$$\left\{\lambda_i, i \in I; \mu_j, j = 1, \ldots, l\right\}$$

tales que

$$\nabla f(x^*) + \sum_{i \in I} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{l} \mu_j \nabla h_j(x^*) = 0$$

$$\lambda_i \geq 0 \quad \forall i \in I$$

  - ✓ Si además $\left\{g_i, i \notin I\right\}$ son diferenciables en x^*

$$\boxed{\begin{aligned} &\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{l} \mu_j \nabla h_j(x^*) = 0 \\ &\lambda_i g_i(x^*) = 0 \quad i = 1, \ldots, m \\ &\lambda_i \geq 0 \quad i = 1, \ldots, m \end{aligned}}$$

Condiciones de optimalidad de KKT (7)

- Condiciones suficientes de Karush-Kuhn-Tucker para problemas con restricciones de desigualdad e igualdad
  - ✓ Sea x^* un punto factible (verifica las restricciones) de P'
  - ✓ Sea $I = \left\{ i \;/\; g_i(x^*) = 0 \right\}$
  - ✓ Sean $f, \left\{ g_i, i \in I \right\}$ convexas y diferenciables en x^*

Si existen escalares $\left\{ \lambda_i, i \in I; \mu_j, j = 1, \ldots, l \right\}$ tales que

$$\nabla f(x^*) + \sum_{i \in I} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{l} \mu_j \nabla h_j(x^*) = 0$$

$$\lambda_i \geq 0 \quad \forall i \in I$$

de modo que h_j sea convexa en x^* si $\mu_j > 0$
ó h_j sea cóncava en x^* si $\mu_j < 0$

entonces x^* es mínimo local de P'

Condiciones de optimalidad de KKT (8)

- Condiciones suficientes de Karush-Kuhn-Tucker para problemas con restricciones de desigualdad e igualdad

Si además, las condiciones sobre las funciones $f, \{g_i, i \in I\}, h_j$ se cumplen en todo el conjunto factible, entonces x^* es el mínimo global de P'