Optimización con restricciones

Prof. Cesar de Prada
ISA-UVA
prada@autom.uva.es

Indice

- Restricciones
- Problemas con restricciones de igualdad
  - Multiplicadores de Lagrange
- Problemas generales NLP
  - Condiciones de Karush-Kuhn-Tucher (KKT)
- Programación cuadrática QP
- Método de Wolfe, SLP
- Funciones de penalización

Restricciones

En la mayor parte de los problemas de toma de decisiones están presentes ligaduras entre las variables o limitaciones en las mismas

- ✓Unas debidas a las ecuaciones del modelo
- ✓Otras al rango permisible de unas variables
- ✓Otras debidas a reglas de operación, existencias, etc.

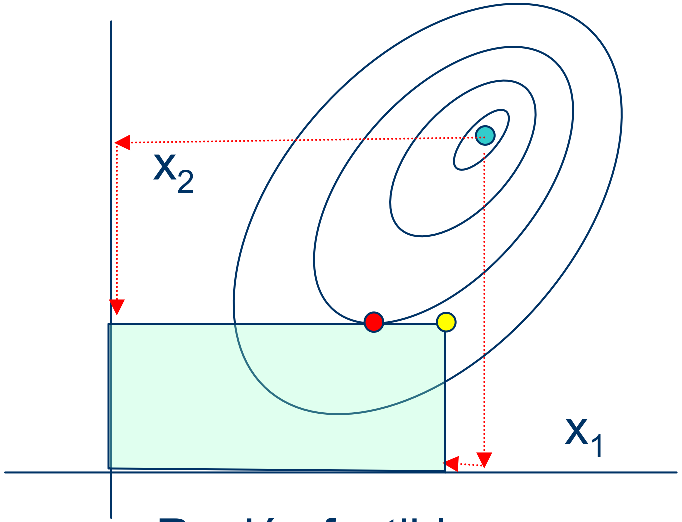
Por ello un problema de optimización a menudo se formula como:

$$\min_{x} J(\mathbf{x})$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{x}) \le \mathbf{0}$$

Problema de optimización no-lineal con restricciones NLP

Restricciones

La presencia de restricciones limita el espacio de búsqueda pero, al mismo tiempo, dificulta el encontrar la solución óptima porque se pierden algunos de los criterios de optimalidad como que el gradiente es nulo en el óptimo



Región factible

Una política conducente a calcular el óptimo sin restricciones y luego limitarlo al contorno de la región factible lleva a soluciones incorrectas

Restricciones de igualdad

Hay una categoría de problemas en los cuales las variables de decisión están sujetas solo a un conjunto de ecuaciones de igualdad:

$$\min_{\mathbf{x}} J(\mathbf{x})$$

$$\mathbf{h}(\mathbf{x}) = 0$$

$$\min_{\mathbf{x}} J(x_1, x_2, ..., x_n)$$

$$\left.\begin{array}{l} h_1(x_1, x_2, ..., x_n) = 0 \\ h_2(x_1, x_2, ..., x_n) = 0 \\ .... \\ h_m(x_1, x_2, ..., x_n) = 0 \end{array}\right\} \quad n > m$$

Si es posible despejar m variables en función de las n-m restantes y sustituirlas en J, es posible transformar el problema en uno de optimización sin restricciones con n-m variables

Restricciones de igualdad

$$\left.\begin{array}{l}\min\limits_{x_1,x_2}\,(x_1x_2+x_2^2)\\ x_1\log x_2=x_2\end{array}\right\}\Rightarrow x_1=\frac{x_2}{\log x_2}\quad\Rightarrow\min_{x_2}\frac{x_2^2}{\log x_2}+x_2^2$$

El procedimiento es similar si partiendo de los valores de $x_1,x_2,\ldots,x_{n-m}$ se puede evaluar $J(x_1,x_2,\ldots,x_n)$ por medio de una rutina de cálculo. En ese caso se puede aplicar un algoritmo de optimización sin restricciones sobre las variables $x_1,x_2,\ldots,x_{n-m}$

Sin embargo, en muchos casos, debido a la no-linealidad de las ecuaciones h_i, no es posible aislar y despejar m variables y el procedimiento no puede emplearse, o su uso implica el uso del método de Newton para resolver las ecuaciones en una rutina de cálculo.

Multiplicadores de Lagrange

Para problemas de optimización con restricciones de igualdad, el método de los multiplicadores de Lagrange proporciona condiciones necesarias que deben cumplirse en el óptimo. La idea es convertir el problema en otro sin restricciones ampliado en m variables λ_j (los multiplicadores de Lagrange) tal que su solución coincida en las variables x con el primitivo y cumpla las restricciones h(x) = 0

$$\left.\begin{array}{l} \min\limits_{\mathbf{x}} J(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) = 0 \end{array}\right\} \Rightarrow \boxed{\min_{\mathbf{x},\boldsymbol{\lambda}} L(\mathbf{x},\boldsymbol{\lambda}) = \min_{\mathbf{x},\boldsymbol{\lambda}} J(\mathbf{x}) + \boldsymbol{\lambda}'\mathbf{h}(\mathbf{x})}$$

Para todos los x que cumplan las restricciones h(x)=0 se verifica:

$$\min_{\mathbf{x},\boldsymbol{\lambda}} L(\mathbf{x},\boldsymbol{\lambda}) = \min_{\mathbf{x}} J(\mathbf{x})$$

Si x^* es optimo para el problema original, minimiza $J(x^*)$ y cumple $h(x^*) = 0$, luego también tiene que ser una solución del problema de la Lagrangiana L

Multiplicadores de Lagrange

$$\min_{\mathbf{x},\boldsymbol{\lambda}} L(\mathbf{x},\boldsymbol{\lambda}) = \min_{\mathbf{x},\boldsymbol{\lambda}} J(\mathbf{x}) + \boldsymbol{\lambda}'\mathbf{h}(\mathbf{x})$$

$L(\mathbf{x},\lambda)$ Lagrangiana

La solución del problema ampliado sin restricciones es:

$$\boxed{\left.\frac{\partial L(\mathbf{x},\boldsymbol{\lambda})}{\partial \mathbf{x}}\right|_{\mathbf{x}^*,\boldsymbol{\lambda}^*} = \mathbf{0}, \qquad \left.\frac{\partial L(\mathbf{x},\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\mathbf{x}^*,\boldsymbol{\lambda}^*} = \mathbf{0} \quad \Rightarrow \quad \mathbf{h}(\mathbf{x}^*) = \mathbf{0}}$$

Que puede resolverse mediante el método de Newton. Si tiene solución, después hay que comprobar mediante el Hessiano de L que la solución x^*,λ^* corresponde verdaderamente a un mínimo de L respecto a x

Cualificación de las restricciones

$$\frac{\partial J(\mathbf{x}^*)}{\partial \mathbf{x}} + \boldsymbol{\lambda}^{*'} \frac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{x}} = \mathbf{0} \qquad \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

$$\begin{bmatrix} \frac{\partial J(\mathbf{x})}{\partial x_1} & \frac{\partial J(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial J(\mathbf{x})}{\partial x_n} \end{bmatrix} + \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_m \end{bmatrix} \begin{bmatrix} \frac{\partial h_1(\mathbf{x})}{\partial x_1} & \frac{\partial h_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial h_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial h_2(\mathbf{x})}{\partial x_1} & \frac{\partial h_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial h_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial h_m(\mathbf{x})}{\partial x_1} & \frac{\partial h_m(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial h_m(\mathbf{x})}{\partial x_n} \end{bmatrix} = \mathbf{0}$$

Para que haya solución óptima, debe cumplirse que los gradientes $\nabla_x h_j$ sean linealmente independientes, lo que se conoce como cualificación de las restricciones

Cualificación de las restricciones

$$\frac{\partial J(\mathbf{x})}{\partial x_1} + \lambda_1 \frac{\partial h_1(\mathbf{x})}{\partial x_1} + \lambda_2 \frac{\partial h_2(\mathbf{x})}{\partial x_1} + \ldots + \lambda_m \frac{\partial h_m(\mathbf{x})}{\partial x_1} = 0$$

$$\frac{\partial J(\mathbf{x})}{\partial x_2} + \lambda_1 \frac{\partial h_1(\mathbf{x})}{\partial x_2} + \lambda_2 \frac{\partial h_2(\mathbf{x})}{\partial x_2} + \ldots + \lambda_m \frac{\partial h_m(\mathbf{x})}{\partial x_2} = 0$$

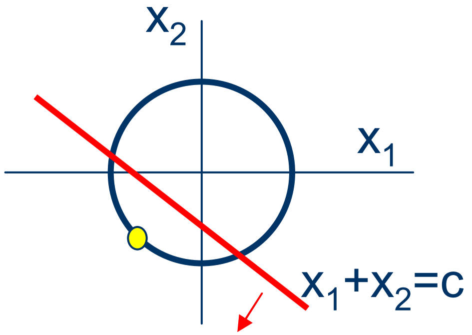
$$\ldots\ldots$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$

Para que haya solución óptima, debe cumplirse que los gradientes $\nabla_x h_j$ sean linealmente independientes, lo que se conoce como cualificación de las restricciones

Ejemplo 1

$$\left.\begin{array}{l}\min\limits_{x_1,x_2} x_1 + x_2 \\ x_1^2 + x_2^2 = 4\end{array}\right\}$$



Cumple la cualificación de las restricciones pues solo hay una

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

$$\frac{\partial L(\mathbf{x},\boldsymbol{\lambda})}{\partial x_1} = 1 + 2\lambda x_1 = 0 \quad \frac{\partial L(\mathbf{x},\boldsymbol{\lambda})}{\partial x_2} = 1 + 2\lambda x_2 = 0 \quad \frac{\partial L(\mathbf{x},\boldsymbol{\lambda})}{\partial \lambda} = x_1^2 + x_2^2 - 4 = 0$$

$$\left.\begin{array}{l} 1 + 2\lambda x_1 = 0 \\ 1 + 2\lambda x_2 = 0 \\ x_1^2 + x_2^2 - 4 = 0 \end{array}\right\} \left.\begin{array}{l} x_1 = -1/2\lambda \\ x_2 = -1/2\lambda \\ 1/4\lambda^2 + 1/4\lambda^2 = 4 \end{array}\right\} \begin{array}{l} \lambda^* = \pm 1/(2\sqrt{2}) \\ x_1 = \mp\sqrt{2}/2 \\ x_2 = \mp\sqrt{2}/2 \end{array}$$

Ejemplo 1

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

$$\frac{\partial L(\mathbf{x},\boldsymbol{\lambda})}{\partial x_1} = 1 + 2\lambda x_1 \qquad \frac{\partial L(\mathbf{x},\boldsymbol{\lambda})}{\partial x_2} = 1 + 2\lambda x_2 \qquad \frac{\partial L(\mathbf{x},\boldsymbol{\lambda})}{\partial \lambda} = x_1^2 + x_2^2 - 4$$

$$\mathbf{H}_{\mathbf{x}}(\mathbf{x},\boldsymbol{\lambda}) = \begin{bmatrix} \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 x_2} \\ \frac{\partial^2 L}{\partial x_2 x_1} & \frac{\partial^2 L}{\partial x_2^2} \end{bmatrix}_{\mathbf{x}^*,\boldsymbol{\lambda}^*} = \begin{bmatrix} 2\lambda & 0 \\ 0 & 2\lambda \end{bmatrix}_{\mathbf{x}^*,\boldsymbol{\lambda}^*} \Rightarrow \begin{matrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \text{PD} \\ \begin{bmatrix} -\sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{bmatrix} \text{ND} \end{matrix}$$

Basta que el mínimo sea respecto a las variables x

Este corresponde a λ^*=sqrt(2)/2: x_1^*= - sqrt(2)/2, x_2^* = - sqrt(2)/2

Interpretación económica de los multiplicadores de Lagrange

$$\left.\begin{array}{l}\min\limits_{\mathbf{x}} J(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) = \mathbf{b}\end{array}\right\} \Rightarrow \min_{\mathbf{x},\boldsymbol{\lambda}} L(\mathbf{x},\boldsymbol{\lambda}) = \min_{\mathbf{x},\boldsymbol{\lambda}} J(\mathbf{x}) + \boldsymbol{\lambda}'[\mathbf{h}(\mathbf{x}) - \mathbf{b}] \qquad L(\mathbf{x}^*,\boldsymbol{\lambda}^*) = J(\mathbf{x}^*)$$

¿Cómo varía J(x*) si b cambia en una unidad?

$$\left.\begin{array}{l}\dfrac{\partial \mathrm{J}(\mathbf{x}^*)}{\partial \mathbf{b}} = \dfrac{\partial \mathrm{J}(\mathbf{x}^*)}{\partial \mathbf{x}^*}\dfrac{\partial \mathbf{x}^*}{\partial \mathbf{b}} \\ \mathbf{0} = \dfrac{\partial(\mathbf{h}(\mathbf{x}^*) - \mathbf{b})}{\partial \mathbf{b}} = \dfrac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{x}^*}\dfrac{\partial \mathbf{x}^*}{\partial \mathbf{b}} - \mathbf{I}\end{array}\right\}$$

← Multiplicando por $\lambda^{*\prime}$ y sumando:

$$\frac{\partial \mathrm{J}(\mathbf{x}^*)}{\partial \mathbf{b}} = \left[\frac{\partial \mathrm{J}(\mathbf{x}^*)}{\partial \mathbf{x}^*} + \boldsymbol{\lambda}^{*\prime}\frac{\partial \mathbf{h}(\mathbf{x}^*)}{\partial \mathbf{x}^*}\right]\frac{\partial \mathbf{x}^*}{\partial \mathbf{b}} - \boldsymbol{\lambda}^{*\prime} = \left[\frac{\partial \mathrm{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*)}{\partial \mathbf{x}^*}\right]\frac{\partial \mathbf{x}^*}{\partial \mathbf{b}} - \boldsymbol{\lambda}^{*\prime} = -\boldsymbol{\lambda}^{*\prime}$$

Los multiplicadores de Lagrange óptimos (precios sombra) dan las sensibilidades opuestas de J* respecto a las restricciones b

Multiplicadores de Lagrange / Precios sombra de LP

Programación Lineal, LP:

$$\max_{\mathbf{x}} J = \mathbf{c}'\mathbf{x}$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\mathbf{x} \geq 0$$

Programación no Lineal con restricciones de igualdad

$$\left.\begin{array}{l}\min_{\mathbf{x}} J(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) = \mathbf{b}\end{array}\right\}$$

¿Cómo varia J(x^*) si b cambia en una unidad?

La respuesta (max) usando la teoría de la dualidad de LP es la solución del dual z, denominada precio sombra

$$\boxed{\frac{\partial J^*}{\partial \mathbf{b}} = \mathbf{z}^*}$$

La respuesta (min) usando los multiplicadores de Lagrange es -λ,

$$\boxed{\frac{\partial J^*}{\partial \mathbf{b}} = -\boldsymbol{\lambda}^{*\prime}}$$

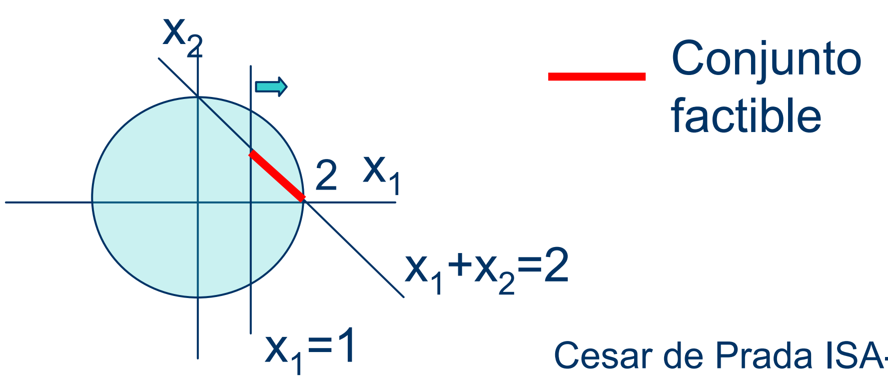
Programación no-lineal NLP

$$\min_{x} J(\mathbf{x})$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$

El caso general incluye restricciones de igualdad y desigualdad en las variables de decisión x.

Suponiendo que las funciones J, g y h son continuamente diferenciables, es posible encontrar condiciones necesarias y (en ciertos casos) suficientes de óptimo del problema NLP denominadas condiciones de Karush-Kunt-Tucker (KKT) (o solo KTC, Kunt-Tacker Conditions)

$$\left.\begin{aligned} &\min_{x_1, x_2} x_1^2 - x_2 \\ &x_1 + x_2 = 2 \\ &x_1^2 + x_1^2 \leq 4 \\ &x_1 - 1 \geq 0 \end{aligned}\right\}$$



Condiciones de optimalidad KKT

$$\min_x J(\mathbf{x})$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$

La idea fundamental en el desarrollo de las condiciones KKT parte de la Lagrangiana, considerando que, si una restricción de desigualdad está activa en el óptimo, entonces puede tratarse como una de igualdad asignándosele un multiplicador de Lagrange μ, y si no está activa entonces puede ignorarse con lo que su multiplicador μ debe hacerse cero. De esta forma μ o g deben ser cero.

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = J(\mathbf{x}) + \sum_j \lambda_j h_j(\mathbf{x}) + \sum_i \mu_i g_i(\mathbf{x}) \qquad \mu_i g_i(x) = 0$$

Por otra parte si aumentamos el lado derecho de g(x) $\leq$ 0 la región factible aumenta y, por tanto, J(x) puede tener un valor menor, de forma que la sensibilidad, medida por $-\mu$, debe ser negativa, o sea $\boldsymbol{\mu} \geq 0$.

Condiciones de optimalidad KKT

$$\min_x J(\mathbf{x})$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{x}) \le \mathbf{0}$$

De esta forma, el óptimo del problema NLP debe cumplir las condiciones de óptimo de la Lagrangiana L(**x**,**λ**,**μ**), mas las condiciones adicionales.

Para funciones continuamente diferenciables J, h y g, las condiciones **necesarias** de óptimo respecto a las variables x son:

$$\left.\begin{array}{l} \dfrac{\partial J}{\partial \mathbf{x}} + \boldsymbol{\lambda}' \dfrac{\partial \mathbf{h}}{\partial \mathbf{x}} + \boldsymbol{\mu}' \dfrac{\partial \mathbf{g}}{\partial \mathbf{x}} = \mathbf{0} \\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ \mathbf{g}(\mathbf{x}) \le \mathbf{0} \\ \mu_i g_i(\mathbf{x}) = 0 \\ \boldsymbol{\mu} \ge \mathbf{0} \end{array}\right\}$$

$$\nabla_x J(\mathbf{x}) + \sum_j \lambda_j \nabla_x h_j(\mathbf{x}) + \sum_i \mu_i \nabla_x g_i(\mathbf{x}) = 0$$
$$h_j(\mathbf{x}) = 0$$
$$g_i(\mathbf{x}) \le 0$$
$$\mu_i g_i(\mathbf{x}) = 0$$
$$\mu_i \ge 0$$

Además, $\nabla_x g_i$ y $\nabla_x h_j$ para las restricciones activas en el óptimo, deben ser también linealmente independientes,

Cualificación de las restricciones

La exigencia de que $\nabla_x g_i$ y $\nabla_x h_j$ (para las restricciones activas en el óptimo) sean linealmente independientes, (a fin de que el sistema de ecuaciones tenga solución) se conoce como cualificación de las restricciones y puede no ser fácil de verificar.

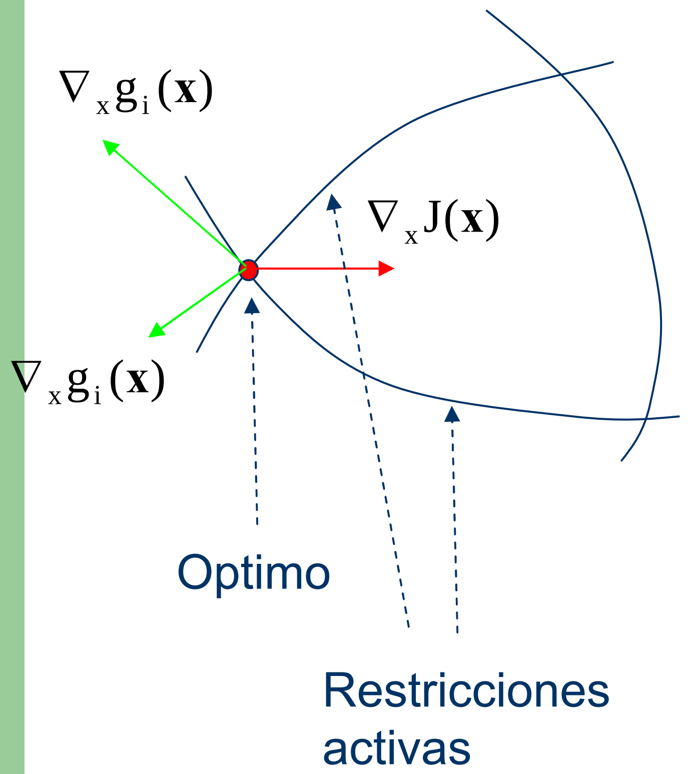
En ciertos casos se cumple siempre:

$$\begin{bmatrix} \nabla_x h(\mathbf{x}) & \nabla_x g(\mathbf{x}) \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = -\nabla_x J(\mathbf{x})$$

✓Cuando las restricciones son lineales

✓Cuando las desigualdades son convexas y las igualdades lineales y hay un punto factible en el interior de la región marcada por las desigualdades.

En el óptimo



Problema con restricciones de desigualdad

$$\nabla_x J(\mathbf{x}) = -\sum_i \mu_i \nabla_x g_i(\mathbf{x})$$

Dado que, en el óptimo, para las restricciones activas μ >0 el gradiente de J, que es una combinación lineal de los gradientes de $g_{j\ (activas)}$, debe estar en un semiplano distinto a estos últimos

Condiciones KKT de suficiencia de primer orden

$$\min_x J(\mathbf{x})$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$

Si la función J es convexa, las restricciones g_i convexas y las restricciones de igualdad h_j son lineales, entonces la región factible es convexa y si existe una solución de las condiciones KKT, entonces dicha solución es el óptimo global del problema NLP

$$\nabla_x J(\mathbf{x}) + \sum_j \lambda_j \nabla_x h_j(\mathbf{x}) + \sum_i \mu_i \nabla_x g_i(\mathbf{x}) = 0$$

$$h_j(\mathbf{x}) = 0$$

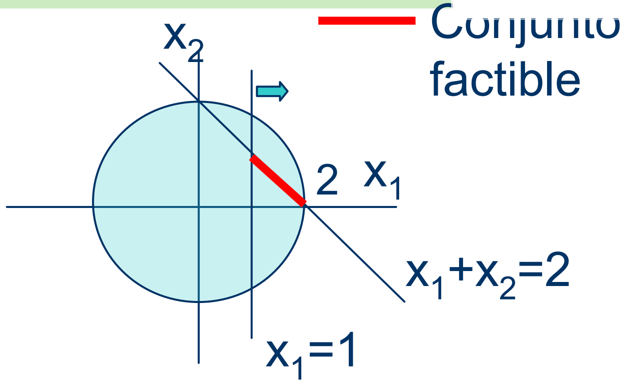
$$g_i(\mathbf{x}) \leq 0$$

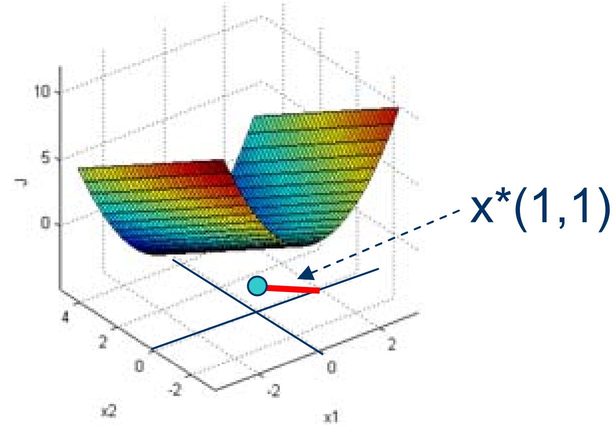
$$\mu_i g_i(\mathbf{x}) = 0$$

$$\mu_i \geq 0$$

Ejemplo 2

$$\left.\begin{aligned} &\min_{x_1,x_2} x_1^2 - x_2 \\ &x_1 + x_2 = 2 \\ &x_1^2 + x_2^2 \le 4 \\ &-x_1 + 1 \le 0 \end{aligned}\right\}$$





¿Cumple las condiciones de suficiencia de primer orden? Si

$$L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = x_1^2 - x_2 + \lambda(x_1 + x_2 - 2) + \mu_1(x_1^2 + x_2^2 - 4) + \mu_2(-x_1 + 1)$$

$$\nabla_x J(\mathbf{x}) + \sum_j \lambda_j \nabla_x h_j(\mathbf{x}) + \sum_i \mu_i \nabla_x g_i(\mathbf{x}) = 0$$

$$h_j(\mathbf{x}) = 0$$

$$g_i(\mathbf{x}) \le 0$$

$$\mu_i g_i(\mathbf{x}) = 0$$

$$\mu_i \ge 0$$

$$\Rightarrow$$

$$\begin{aligned} &2x_1 + \lambda + \mu_1 2x_1 - \mu_2 = 0 \\ &-1 + \lambda + \mu_1 2x_2 = 0 \\ &x_1 + x_2 = 2 \quad \mu_1 \ge 0 \quad \mu_2 \ge 0 \\ &x_1^2 + x_2^2 \le 4 \quad \mu_1(x_1^2 + x_2^2 - 4) = 0 \\ &-x_1 + 1 \le 0 \quad \mu_2(-x_1 + 1) = 0 \end{aligned}$$

Ejemplo 2

$$\left.\begin{array}{l}
2x_1+\lambda+\mu_1 2x_1-\mu_2=0\\
-1+\lambda+\mu_1 2x_2=0\\
x_1+x_2=2 \quad \mu_1\ge 0 \quad \mu_2\ge 0\\
x_1^2+x_2^2\le 4 \quad \mu_1(x_1^2+x_2^2-4)=0\\
-x_1+1\le 0 \quad \mu_2(-x_1+1)=0
\end{array}\right\}$$

$$x_1=\frac{\mu_2-\lambda}{2(1+\mu_1)}$$

$$x_2=\frac{1-\lambda}{2\mu_1}$$

$$\frac{\mu_2-\lambda}{2(1+\mu_1)}+\frac{1-\lambda}{2\mu_1}=2$$

$$\left[\left(\frac{\mu_2-\lambda}{2(1+\mu_1)}\right)^2+\left(\frac{1-\lambda}{2\mu_1}\right)^2-4\right]\mu_1=0$$

$$\left[\frac{-\mu_2+\lambda}{2(1+\mu_1)}+1\right]\mu_2=0$$

Deben examinarse todas las alternativas para encontrar posibles soluciones

$$
\left.
\begin{aligned}
x_1 &= \frac{\mu_2-\lambda}{2(1+\mu_1)} \\
x_2 &= \frac{1-\lambda}{2\mu_1} \\
&\frac{\mu_2-\lambda}{2(1+\mu_1)} + \frac{1-\lambda}{2\mu_1} = 2 \\
&\left[\left(\frac{\mu_2-\lambda}{2(1+\mu_1)}\right)^2 + \left(\frac{1-\lambda}{2\mu_1}\right)^2 - 4\right]\mu_1 = 0 \\
&\left[\frac{-\mu_2+\lambda}{2(1+\mu_1)} + 1\right]\mu_2 = 0
\end{aligned}
\right\}
$$

$$
\text{si } \mu_1 = 0, \mu_2 = 0 \quad \begin{cases} \dfrac{\mu_2-\lambda}{2(1+\mu_1)} + \dfrac{1-\lambda}{2\mu_1} = 2 \end{cases}
$$

$$
\text{si } \mu_1 = 0, \mu_2 \neq 0 \quad \begin{cases} \dfrac{\mu_2-\lambda}{2(1+\mu_1)} + \dfrac{1-\lambda}{2\mu_1} = 2 \\ \dfrac{-\mu_2+\lambda}{2(1+\mu_1)} + 1 = 0 \end{cases}
$$

$$
\text{si } \mu_2 = 0, \mu_1 \neq 0 \quad \Rightarrow \quad \begin{cases} \left(\dfrac{-\lambda}{2(1+\mu_1)}\right)^2 + \left(\dfrac{1-\lambda}{2\mu_1}\right)^2 - 4 = 0 \\ \dfrac{-\lambda}{2(1+\mu_1)} + \dfrac{1-\lambda}{2\mu_1} = 2 \end{cases}
$$

$$
\text{si } \mu_1 \neq 0, \mu_2 \neq 0 \quad \begin{cases} \dfrac{\mu_2-\lambda}{2(1+\mu_1)} + \dfrac{1-\lambda}{2\mu_1} = 2 \\ \left(\dfrac{\mu_2-\lambda}{2(1+\mu_1)}\right)^2 + \left(\dfrac{1-\lambda}{2\mu_1}\right)^2 - 4 = 0 \\ \dfrac{-\mu_2+\lambda}{2(1+\mu_1)} + 1 = 0 \end{cases}
$$

Detalle de la resolución μ_2=0, μ_1≠0

$$\left.\begin{array}{c}\left(\dfrac{-\lambda}{2(1+\mu_1)}\right)^2+\left(\dfrac{1-\lambda}{2\mu_1}\right)^2-4=0\\ \dfrac{-\lambda}{2(1+\mu_1)}+\dfrac{1-\lambda}{2\mu_1}=2\end{array}\right\}\quad \begin{array}{c}\dfrac{-\lambda}{2(1+\mu_1)}=2-\dfrac{1-\lambda}{2\mu_1}\\ \left(2-\dfrac{1-\lambda}{2\mu_1}\right)^2+\left(\dfrac{1-\lambda}{2\mu_1}\right)^2=4\end{array}$$

$$4+\left(\frac{1-\lambda}{2\mu_1}\right)^2-4\frac{1-\lambda}{2\mu_1}+\left(\frac{1-\lambda}{2\mu_1}\right)^2=4 \qquad ((1-\lambda)-4\mu_1)(1-\lambda)=0\begin{cases}\lambda=1\\ 1-\lambda=4\mu_1\end{cases}$$

$$\lambda=1\rightarrow\frac{-1}{2(1+\mu_1)}=2 \quad 1+\mu_1=-1/4 \quad \mu_1=-5/4\ NF$$

$$1-\lambda=4\mu_1\rightarrow\frac{4\mu_1-1}{2(1+\mu_1)}+\frac{4\mu_1}{2\mu_1}=2 \quad \frac{4\mu_1-1}{2(1+\mu_1)}+2=2\rightarrow\mu_1=1/4,\lambda=0$$

$$x_1=0,x_2=2 \quad NF$$

Ejemplo 2

$$\left.\begin{array}{c}
x_1 = \dfrac{\mu_2 - \lambda}{2(1+\mu_1)} \\
x_2 = \dfrac{1-\lambda}{2\mu_1} \\
\dfrac{\mu_2 - \lambda}{2(1+\mu_1)} + \dfrac{1-\lambda}{2\mu_1} = 2 \\
\left[\left(\dfrac{\mu_2 - \lambda}{2(1+\mu_1)}\right)^2 + \left(\dfrac{1-\lambda}{2\mu_1}\right)^2 - 4\right]\mu_1 = 0 \\
\left[\dfrac{-\mu_2 + \lambda}{2(1+\mu_1)} + 1\right]\mu_2 = 0
\end{array}\right\}
\begin{array}{l}
\text{si } \mu_1 = 0, \mu_2 = 0 \quad \{\lambda = 1, x_1 = -1/2, x_2 = 5/2 \ NF \\
\text{si } \mu_1 = 0, \mu_2 \neq 0 \quad \{\lambda = 1, \mu_2 = 3, \boxed{x_1 = 1, x_2 = 1} \\
\text{si } \mu_1 \neq 0, \mu_2 = 0 \quad \begin{cases} \mu_1 = 1/4, \lambda = 0, x_1 = 0, x_2 = 2 \ NF \\ \mu_1 = -5/4, \lambda = 1 \ NF \end{cases} \\
\text{si } \mu_1 \neq 0, \mu_2 \neq 0 \quad \{\lambda = 1; \mu_1 = 0, \mu_2 = 3, x_1 = 1, x_2 = 1
\end{array}$$

Solución óptima

Cualificación de las restricciones

$$\left.\begin{array}{l}\min\limits_{x_1,x_2} x_1^2 - x_2 \\ x_1 + x_2 = 2 \\ x_1^2 + x_2^2 \le 4 \\ -x_1 + 1 \le 0\end{array}\right\}$$

Optimo: $x_1=1, x_2=1, \lambda=1, \mu_1=0, \mu_2=3$

Restricciones activas

$\nabla_x h(x^*)$, $\nabla_x g(x^*)$ (activas)

$$\nabla_x h = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \nabla_x g = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad rango \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} = 2$$

Son linealmente independientes, luego se cumple la cualificación de las restricciones

Sensibilidad

$$\min_x J(\mathbf{x})$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{b}$$
$$\mathbf{g}(\mathbf{x}) \leq \mathbf{c}$$

A partir de la interpretación de las condiciones KKT como multiplicadores de Lagrange, se puede establecer la siguiente relación para las sensibilidades:

$$\boxed{\frac{\partial J(\mathbf{x}^*)}{\partial \mathbf{b}} = -\boldsymbol{\lambda}^{*\prime} \qquad \frac{\partial J(\mathbf{x}^*)}{\partial \mathbf{c}} = -\boldsymbol{\mu}^{*\prime}}$$

La cual nos permite deducir como cambia la función de costo cuando las restricciones se relajen en una unidad, lo cual es importante en problemas de toma de decisiones.

Ejemplo 2

$$\left.\begin{aligned}
&\min_{x_1,x_2} J(x_1,x_2) = \min_{x_1,x_2} x_1^2 - x_2 \\
&x_1 + x_2 = 2 \\
&x_1^2 + x_2^2 \le 4 \\
&-x_1 + 1 \le 0
\end{aligned}\right\}$$

Restricciones activas

Optimo: x_1=1,x_2=1,λ=1,μ_1=0,μ_2=3, J=0

¿Como varia el valor óptimo de la función de coste J si se aumenta el término derecho de las restricciones en una unidad?

Sensibilidades: -1,0,-3

Si aumentasen una unidad cada una manteniendo las otras, la primera mejoraría en una unidad el costo J, no influiría en la segunda y la tercera mejoraría J en tres unidades

Condiciones de segundo orden

Las condiciones de KKT son expresiones de primer orden que pueden dar lugar a máximos, mínimos o puntos de silla de la Lagrangiana L(x,λ, μ) en (x*,λ*,μ*).

Las condiciones de segundo orden permiten determinar si se trata de un mínimo respecto a las variables x y proporcionan condiciones **suficientes** para la optimalidad de la solución. Vienen dadas por la condición sobre el Hessiano:

$$\mathbf{z}'\nabla_x^2 L(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)\mathbf{z} > 0 \qquad \forall z \mid \begin{bmatrix} \nabla_x \mathbf{h}(\mathbf{x}^*) \\ \nabla_x \mathbf{g}_{act}(\mathbf{x}^*) \end{bmatrix} \mathbf{z} = \mathbf{0}$$

De modo que el Hessiano de L respecto a x es PD para todos los vectores z ortogonales a los gradientes de las restricciones activas en x^*. Estos vectores dan origen al plano tangente a dichas restricciones. Para la necesidad basta con que el Hessiano sea PSD. (g_{act} son restricciones activas)

Lagrangiana $L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = x_1^2 - x_2 + \lambda(x_1 + x_2 - 2) + \mu_1(x_1^2 + x_2^2 - 4) + \mu_2(-x_1 + 1)$

$$\nabla_x J(\mathbf{x}) + \sum_j \lambda_j \nabla_x h_j(\mathbf{x}) + \sum_i \mu_i \nabla_x g_i(\mathbf{x}) = 0 \quad \begin{cases} 2x_1 + \lambda + \mu_1 2x_1 - \mu_2 = 0 \\ -1 + \lambda + \mu_1 2x_2 = 0 \end{cases}$$

Ejemplo 2

Optimo: $x_1=1, x_2=1, \lambda=1, \mu_1=0, \mu_2=3$

$$\left.\begin{aligned} &\min_{x_1,x_2} x_1^2 - x_2 \\ &x_1 + x_2 = 2 \\ &x_1^2 + x_2^2 \le 4 \\ &-x_1 + 1 \le 0 \end{aligned}\right\}$$

Restricciones activas

$$\mathbf{z}'\nabla_x^2 L(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)\mathbf{z} > 0 \qquad \forall z \mid \begin{bmatrix} \nabla_x \mathbf{h}(\mathbf{x}^*) \\ \nabla_x \mathbf{g}_{act}(\mathbf{x}^*) \end{bmatrix}\mathbf{z} = \mathbf{0}$$

$$\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0 \quad z_1 + z_2 = 0$$

$$\begin{bmatrix} -1 & 0 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0 \quad -z_1 = 0$$

Vectores de planos tangentes

Hessiano respecto a x

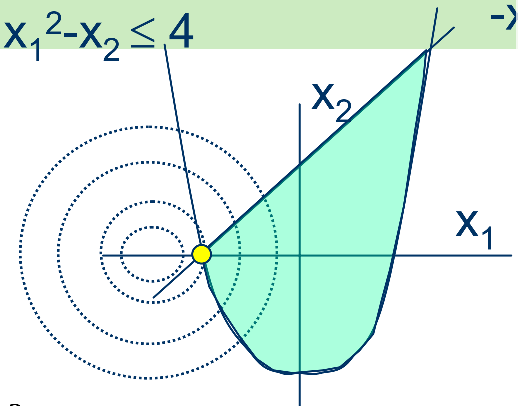
$$H_x = \begin{bmatrix} \dfrac{\partial L^2}{\partial x_1^2} & \dfrac{\partial L^2}{\partial x_1 \partial x_2} \\ \dfrac{\partial L^2}{\partial x_2 \partial x_1} & \dfrac{\partial L^2}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 + 2\mu_1 & 0 \\ 0 & 2\mu_1 \end{bmatrix} \quad \mathbf{z}' H_x^* \mathbf{z} = \begin{bmatrix} z & -z \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} z \\ -z \end{bmatrix} = 2z^2 > 0$$

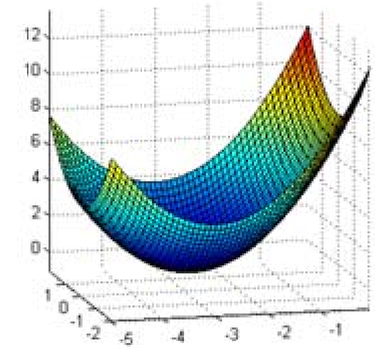
Cumple necesidad pero no la suficiencia

$$\begin{bmatrix} 0 & z \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ z \end{bmatrix} = 0$$

Ejemplo 3

$$\left.\begin{aligned} &\min_{x_1,x_2} (x_1+3)^2 + x_2^2 \\ &-x_1 + x_2 - 2 \le 0 \\ &x_1^2 - x_2 - 4 \le 0 \end{aligned}\right\}$$





$$L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = (x_1+3)^2 + x_2^2 + \mu_1(-x_1+x_2-2) + \mu_2(x_1^2 - x_2 - 4)$$

$$\nabla_x J(\mathbf{x}) + \sum_j \lambda_j \nabla_x h_j(\mathbf{x}) + \sum_i \mu_i \nabla_x g_i(\mathbf{x}) = 0$$

$$h_j(\mathbf{x}) = 0$$

$$g_i(\mathbf{x}) \le 0$$

$$\mu_i g_i(\mathbf{x}) = 0$$

$$\mu_i \ge 0$$

$$2(x_1+3) + \mu_1(-1) + \mu_2 2x_1 = 0$$

$$2x_2 + \mu_1(1) + \mu_2(-1) = 0$$

$$(-x_1 + x_2 - 2)\mu_1 = 0 \quad \mu_1 \ge 0$$

$$(x_1^2 - x_2 - 4)\mu_2 = 0 \quad \mu_2 \ge 0$$

$$-x_1 + x_2 - 2 \le 0$$

$$x_1^2 - x_2 - 4 \le 0$$

Ejemplo 3

$$\left.\begin{array}{l}
2(x_1+3)+\mu_1(-1)+\mu_2 2x_1 = 0 \\
2x_2+\mu_1(1)+\mu_2(-1) = 0 \\
(-x_1+x_2-2)\mu_1 = 0 \quad \mu_1 \geq 0 \\
(x_1^2-x_2-4)\mu_2 = 0 \quad \mu_2 \geq 0 \\
-x_1+x_2-2 \leq 0 \\
x_1^2-x_2-4 \leq 0
\end{array}\right\}
\quad
\left.\begin{array}{c}
x_1 = \dfrac{\mu_1-6}{2(1+\mu_2)} \\
x_2 = \dfrac{\mu_2-\mu_1}{2} \\
\left(\dfrac{6-\mu_1}{2(1+\mu_2)}+\dfrac{\mu_2-\mu_1}{2}-2\right)\mu_1 = 0 \\
\left(\left(\dfrac{\mu_1-6}{2(1+\mu_2)}\right)^2-\dfrac{\mu_2-\mu_1}{2}-4\right)\mu_2 = 0
\end{array}\right\}$$

$$\text{si } \mu_1 = 0, \mu_2 = 0 \quad \rightarrow x_1 = -3, x_2 = 0 \quad NF$$

$$\text{si } \mu_1 \neq 0, \mu_2 = 0 \quad \left\{ \frac{6-\mu_1}{2}+\frac{-\mu_1}{2}-2 = 0 \right. \quad \rightarrow \mu_1 = 1, x_1 = -5/2, x_2 = -1/2 \; NF$$

Ejemplo 3

$$\left.\begin{array}{r} x_1 = \dfrac{\mu_1 - 6}{2(1+\mu_2)} \\ x_2 = \dfrac{\mu_2 - \mu_1}{2} \\ \left(\dfrac{6-\mu_1}{2(1+\mu_2)} + \dfrac{\mu_2-\mu_1}{2} - 2\right)\mu_1 = 0 \\ \left(\left(\dfrac{\mu_1-6}{2(1+\mu_2)}\right)^2 - \dfrac{\mu_2-\mu_1}{2} - 4\right)\mu_2 = 0 \end{array}\right\}$$

$$\text{si } \mu_1 = 0, \mu_2 \neq 0 \quad \begin{cases} \dfrac{36}{4(1+\mu_2)^2} - \dfrac{\mu_2}{2} = 4 \\ \mu_2^3 + 10\mu_2^2 + 17\mu_2 = 10 \\ \mu_2 = -7.58, -2.87, 0.46 \\ x_1 = -2.06, x_2 = 0.23\ NF \end{cases}$$

$$\text{si } \mu_1 \neq 0, \mu_2 \neq 0 \begin{cases} \dfrac{6-\mu_1}{2(1+\mu_2)} + \dfrac{\mu_2-\mu_1}{2} - 2 = 0 \\ \left(\dfrac{\mu_1-6}{2(1+\mu_2)}\right)^2 - \dfrac{\mu_2-\mu_1}{2} - 4 = 0 \end{cases}$$

$$\left(\frac{\mu_2-\mu_1}{2} - 2\right)^2 - \frac{\mu_2-\mu_1}{2} = 4 \rightarrow (\mu_2-\mu_1)^2 = 10(\mu_2-\mu_1)$$

$$\begin{cases} \mu_2 = \mu_1 \rightarrow \mu_1 = 2/5, \mu_2 = 2/5, \boxed{x_1 = -2, x_2 = 0} \\ \mu_2 - \mu_1 = 10 \rightarrow \dfrac{6-\mu_1}{11+\mu_1} + 6 = 0 \rightarrow \mu_1 = -14.4\ NF \end{cases}$$

Optimo, por ser la región convexa y J convexa

Cualificación de las restricciones

$$\left.\begin{array}{l}\min\limits_{x_1,x_2} (x_1+3)^2 + x_2^2 \\ -x_1 + x_2 - 2 \le 0 \\ x_1^2 - x_2 - 4 \le 0\end{array}\right\}$$

Optimo: x_1=-2,x_2=0,μ_1=2/5,μ_2=2/5, J=1

Restricciones activas

$\nabla_x h(x^*)$, $\nabla_x g(x^*)$ (activas)

$$\nabla_x g_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \nabla_x g_2 = \begin{bmatrix} 2(-2) \\ -1 \end{bmatrix} \quad rango\begin{bmatrix} -1 & -4 \\ 1 & -1 \end{bmatrix} = 2$$

Son linealmente independientes, luego se cumple la cualificación de las restricciones

Sensibilidad

$$\left.\begin{array}{l}\min\limits_{x_1,x_2} (x_1+3)^2 + x_2^2 \\ -x_1 + x_2 - 2 \le 0 \\ x_1^2 - x_2 - 4 \le 0\end{array}\right\}$$

Optimo: $x_1=-2, x_2=0, \mu_1=2/5, \mu_2=2/5$, J=1

Restricciones activas

¿Como varia el valor óptimo de la función de coste J si se aumenta el término derecho de las restricciones en una unidad?

Sensibilidades: -2/5,-2/5

Si aumentasen una unidad cada una manteniendo la otra, J mejoraría en 2/5 en cada caso

Lagrangiana $$L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = (x_1+3)^2 + x_2^2 + \mu_1(-x_1+x_2-2) + \mu_2(x_1^2 - x_2 - 4)$$

$$\nabla_x J(\mathbf{x}) + \sum_j \lambda_j \nabla_x h_j(\mathbf{x}) + \sum_i \mu_i \nabla_x g_i(\mathbf{x}) = 0 \quad \begin{cases} 2(x_1+3) + \mu_1(-1) + \mu_2 2x_1 = 0 \\ 2x_2 + \mu_1(1) + \mu_2(-1) = 0 \end{cases}$$

Ejemplo 3

Optimo: $x_1=-2, x_2=0, \mu_1=2/5, \mu_2=2/5$, J=1

Condiciones de 2º orden

$$\left.\begin{array}{l} \min_{x_1,x_2} (x_1+3)^2 + x_2^2 \\ -x_1 + x_2 - 2 \le 0 \\ x_1^2 - x_2 - 4 \le 0 \end{array}\right\}$$

Restricciones activas

$$\mathbf{z}'\nabla_x^2 L(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)\mathbf{z} > 0 \qquad \forall z \mid \begin{bmatrix} \nabla_x \mathbf{h}(\mathbf{x}^*) \\ \nabla_x \mathbf{g}_{act}(\mathbf{x}^*) \end{bmatrix}\mathbf{z} = \mathbf{0}$$

$$\begin{bmatrix} -1 & 1 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0 \quad -z_1 + z_2 = 0$$

Vectores de planos tangentes

$$\begin{bmatrix} 2(-2) & -1 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0 \quad -4z_1 - z_2 = 0$$

Hessiano respecto a x

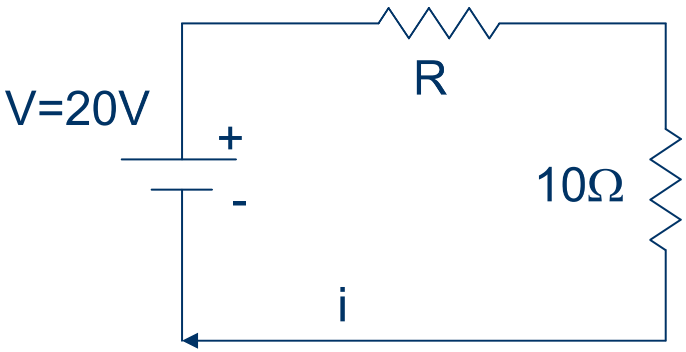
$$H_x = \begin{bmatrix} \dfrac{\partial L^2}{\partial x_1^2} & \dfrac{\partial L^2}{\partial x_1 \partial x_2} \\ \dfrac{\partial L^2}{\partial x_2 \partial x_1} & \dfrac{\partial L^2}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2+2\mu_2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\mathbf{z}' H_x^* \mathbf{z} = \begin{bmatrix} z & z \end{bmatrix}\begin{bmatrix} 14/5 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} z \\ z \end{bmatrix} = 14z^2/5 > 0$$

$$\begin{bmatrix} z & -4z \end{bmatrix}\begin{bmatrix} 14/5 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} z \\ -4z \end{bmatrix} = 14z^2/5 > 0$$

Cumple la suficiencia

Ejemplo 4



Encontrar el valor de R de modo que se maximice la potencia absorbida en dicha resistencia

$$\left.\begin{aligned} P &= I^2 R \\ 20 &= IR + 10I \end{aligned}\right\} \left\{\begin{aligned} &\min_R \frac{400R}{(R+10)^2} \\ &R \geq 0 \end{aligned}\right\}$$

Solución de problemas NLP

- La solución analítica de un problema NLP utilizando las condiciones KKT solo es posible en casos sencillos
- Existen varias alternativas prácticas para abordar la solución numérica de un problema NLP.
  - Explotar la estructura particular del mismo (QP, Wolfe,..)
  - Transformar el problema en otro equivalente sin restricciones (Funciones de penalización)
  - Sucesión de problemas aproximados mas sencillos (SLP, Cutting Plane,..)
  - Resolución aproximada secuencial de las condiciones KKT (SQP)
  - Métodos tipo gradiente (GRG)
  - Etc.

Programación Cuadrática (QP)

Al igual que existen muchos problemas prácticos de interés tipo LP, otro tipo particular de problemas de optimización con restricciones que tienen mucha aplicación son los denominados de Programación Cuadrática, o QP, en que la función de costo es cuadrática en las variables de decisión y las restricciones son lineales.

$$\min_{x} J(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x}$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

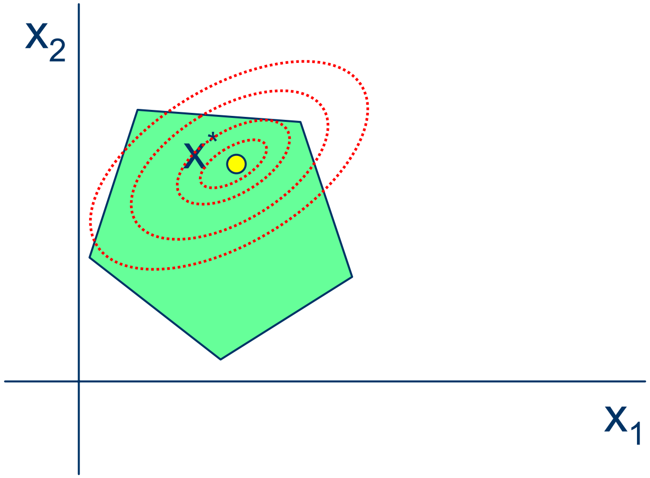
$$\mathbf{x} \geq 0$$

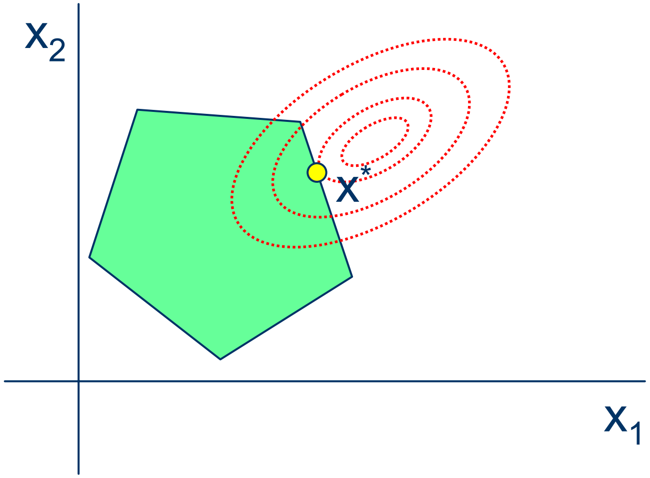
Q (n x n) simétrica

A (m x n) rango m < n

Se pueden encontrar formulaciones equivalentes que expresan el problema en términos de restricciones de desigualdad, como maximización, etc. Para ello se utilizan las mismas técnicas de variables de holgura, cambios de signo, etc. Que en la LP

Programación cuadrática





A diferencia de la programación lineal, en el caso de la programación cuadrática el óptimo puede encontrarse en el interior o en el contorno de la región factible

Programación cuadrática (QP)

$$\min_x J(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x}$$

$$\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{0}$$

$$-\mathbf{x} \leq 0$$

Técnicas de solución:

- ✓Mediante las condiciones KKT
- ✓Conjuntos activos
- ✓Métodos mas generales (Wolfe)
- ✓Otros

Al ser A de rango m, los gradientes $\nabla_x h = A$, $\nabla_x g = -I$ son independientes y se cumple la cualificación de las restricciones

Si la matriz Q es positiva semi-definida, entonces J es convexa, y como la región factible, si existe, también lo es, una solución local del problema es un óptimo global y puede encontrarse resolviendo las condiciones KKT

Si Q no es PSD, entonces pueden existir óptimos locales o no existir solución y las condiciones KKT solo son necesarias.

Resolución mediante KKT (Dantzing-Wolfe)

$$\min_x J(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x}$$

$$\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{0}$$

$$-\mathbf{x} \leq 0$$

Lagrangiana:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x} + \boldsymbol{\lambda}'(\mathbf{A}\mathbf{x} - \mathbf{b}) - \boldsymbol{\mu}'\mathbf{x}$$

Condiciones KKT:

$$\mathbf{c} + \mathbf{Q}\mathbf{x} + \mathbf{A}'\boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0}$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

$$\mathbf{x} \geq \mathbf{0}, \quad \boldsymbol{\mu} \geq \mathbf{0}$$

$$\boldsymbol{\mu}'\mathbf{x} = 0$$

Se puede encontrar una solución factible del conjunto de ecuaciones lineales en (x,λ,μ) como un problema LP en dichas variables con el algoritmo simplex tipo I, y mantener la condición μ'x=0 imponiendo como regla de cambio de pivote que no esten simultáneamente en la base columnas con componentes del vector (x,λ,μ) distintas de cero en x y en μ

Ecuaciones lineales excepto por **μ'x = 0,** ó x ó μ son cero.

Resolución mediante KKT

$$\left.\begin{aligned} &\mathbf{c} + \mathbf{Qx} + \mathbf{A}'\boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0} \\ &\mathbf{Ax} = \mathbf{b} \\ &\mathbf{x} \geq \mathbf{0}, \quad \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}\right\} \Rightarrow \begin{aligned} &\boldsymbol{\lambda} = \boldsymbol{\sigma} - \boldsymbol{\delta} \\ &\boldsymbol{\sigma} \geq \mathbf{0}, \boldsymbol{\delta} \geq \mathbf{0} \\ &\mathbf{z}' = (\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\delta}, \boldsymbol{\mu}) \end{aligned}$$

$$\boldsymbol{\mu}'\mathbf{x} = 0$$

La condición μ'x=0 se impone a traves de reglas sobre las columnas en las operaciones de pivote

$$\left.\begin{aligned} &\mathbf{c} + \mathbf{Qx} + \mathbf{A}'(\boldsymbol{\sigma} - \boldsymbol{\delta}) - \boldsymbol{\mu} = \mathbf{0} \\ &\mathbf{Ax} = \mathbf{b} \\ &\mathbf{z} \geq \mathbf{0} \end{aligned}\right\} \begin{bmatrix} \mathbf{Q} & \mathbf{A}' & -\mathbf{A}' & -\mathbf{I} \\ \mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\sigma} \\ \boldsymbol{\delta} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} -\mathbf{c} \\ \mathbf{b} \end{bmatrix} \quad \mathbf{z} \geq \mathbf{0}$$

$$\left.\begin{aligned} &\max_{\mathbf{z},\boldsymbol{\nu}} -(1,1,\ldots 1)\boldsymbol{\nu} \\ &\begin{bmatrix} \mathbf{I} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \boldsymbol{\nu} \\ \mathbf{z} \end{bmatrix} = \boldsymbol{\beta}, \quad \mathbf{z} \geq \mathbf{0}, \boldsymbol{\nu} \geq 0 \end{aligned}\right\}$$

Problema tipo Fase I del simplex cuya solución óptima, si ν^*=0, da la solución de las condiciones KKT

Conjuntos activos (Active set)

Está indicado para problemas formulados en términos de desigualdades lineales. El método hace uso del hecho que, en un determinado paso del algoritmo, las restricciones activas (cuyos índices forman el conjunto activo Λ) hay que considerarlas como igualdades, mientras que las no activas pueden eliminarse de la formulación momentáneamente

$$\min_{\mathbf{x}} \mathrm{J}(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x}$$
$$\mathbf{A}\mathbf{x} \leq \mathbf{b} = \mathbf{0}$$

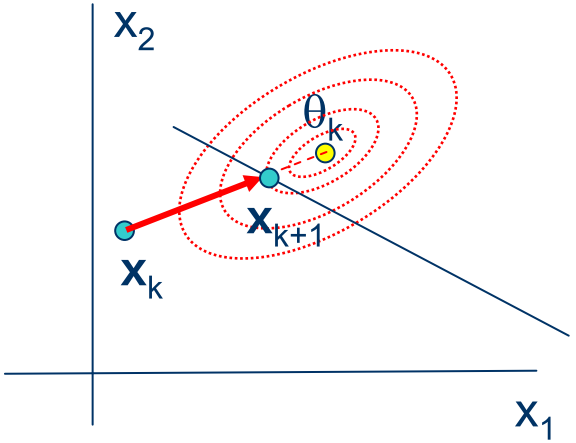
En cada paso del algoritmo el problema a resolver es del tipo:

$$\left.\begin{array}{l} \min\limits_{\mathbf{x}} J(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \dfrac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x} \\ \mathbf{a_i}'\mathbf{x} - \beta_i = 0 \quad i \in \Lambda \end{array}\right\} P_\Lambda$$

En el que no existen restricciones tipo $\leq$, y puede resolverse mas facilmente, p.e. por sustitución o multiplicadores de Lagrange

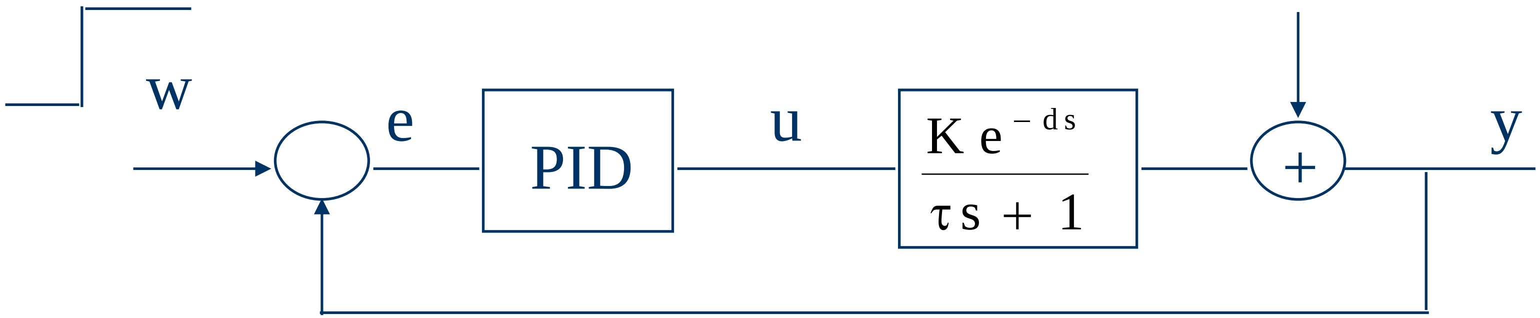
Método de los conjuntos activos

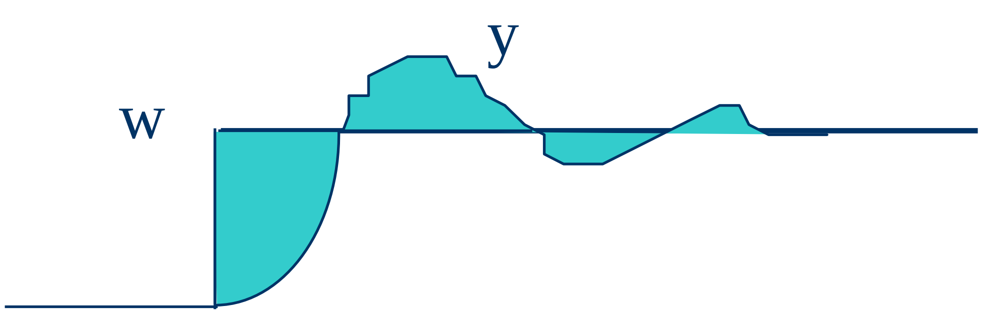
1. Escoger x_k y calcular Λ
2. Resolver el problema , P_Λ, asociado con igualdades en Λ. Sea θ_k la solución y λ_k sus multiplicadores de Lagrange
3. Calcular $x_{k+1}=x_k+\alpha_k(\theta_k-x_k)$ Si θ_k no verifica una restricción que no está en Λ (α_k <1) añadir el índice a Λ
4. Si los λ_k que corresponden a desigualdades son todos ≥ 0, entonces θ_k es óptimo. En caso contrario eliminar el índice p correspondiente al menor de ellos de Λ
5. Hacer k=k+1 y volver a 2



$$\alpha_k = \min\left(1, \min_{\substack{i\notin\Lambda \\ \mathbf{a_i}'(\boldsymbol{\theta}_k - \mathbf{x}_k)<0}} \frac{\beta_i - \mathbf{a_i}'\mathbf{x_k}}{\mathbf{a_i}'(\boldsymbol{\theta}_k - \mathbf{x}_k)}\right)$$

Ejemplo QP (Dinámico)





$$\min_{K_p, T_i, T_d} \int e(t)^2 dt \quad \text{MISE}$$

$$K_p \geq 0,\ T_i \geq 0,\ T_d \geq 0$$

error = w - y (función lineal de K_p , T_i, T_d)

Método de Wolfe

$$\min_x J(\mathbf{x})$$
$$\mathbf{Ax} \le \mathbf{b}$$
$$\mathbf{x} \ge 0$$

El algoritmo de Wolfe está orientado a resolver problemas algo mas generales que los de tipo QP: Aquellos en los que la función de costo no tiene por que ser cuadrática, aunque se mantienen las restricciones lineales

Se resuelve mediante una sucesión de problemas LP aproximados

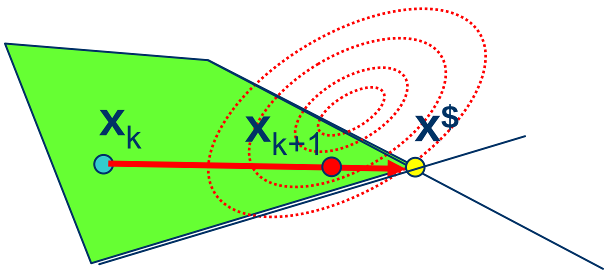
En concreto sustituye J(x) por una aproximación de primer orden, y usa el resultando de este problema para definir un dirección de búsqueda factible en la que se mejora J(x) respetando las restricciones

Método de Wolfe

$$\min_x J(\mathbf{x})$$
$$\mathbf{Ax} \le \mathbf{b}$$
$$\mathbf{x} \ge 0$$

Partiendo de un $\mathbf{x}_k$ factible, se resuelve el problema linealizado en $\mathbf{x}_k$ de tipo LP:

$$\min_{\mathbf{x}} J(\mathbf{x}_k) + \nabla_{\mathbf{x}} J(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)$$
$$\mathbf{Ax} \le \mathbf{b}$$
$$\mathbf{x} \ge 0$$



Por ser un problema LP tiene una solución óptima en un vértice $\mathbf{x}^\$$ y por tanto :

$$J(\mathbf{x}_k) = J(\mathbf{x}_k) + \nabla_x J(\mathbf{x}_k)(\mathbf{x}_k - \mathbf{x}_k) \ge J(\mathbf{x}_k) + \nabla_x J(\mathbf{x}_k)(\mathbf{x}_k^\$ - \mathbf{x}_k)$$
$$J(\mathbf{x}_k) \ge J(\mathbf{x}_k) + \nabla_x J(\mathbf{x}_k)(\mathbf{x}_k^\$ - \mathbf{x}_k) \Rightarrow \boxed{0 \ge \nabla_x J(\mathbf{x}_k)(\mathbf{x}_k^\$ - \mathbf{x}_k)}$$

Con lo cual la dirección $\mathbf{x}_k^\$ - x_k$ es una en la que J(x) disminuye

Método de Wolfe

El segundo paso del algoritmo consiste en minimizar J(**x**) en la dirección $\mathbf{x}_k^{\$} - \mathbf{x}_k$ en el segmento entre $\mathbf{x}_k$ y $\mathbf{x}_k^{\$}$ el cual es factible. Esto es una optimización escalar que se resuelve fácilmente.

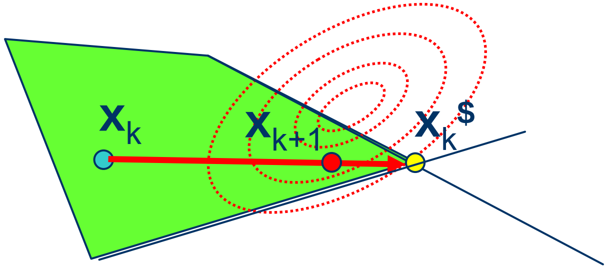
$$\min_{\sigma_k} J(\mathbf{x}_k + \sigma_k(\mathbf{x}_k^{\$} - \mathbf{x}_k))$$

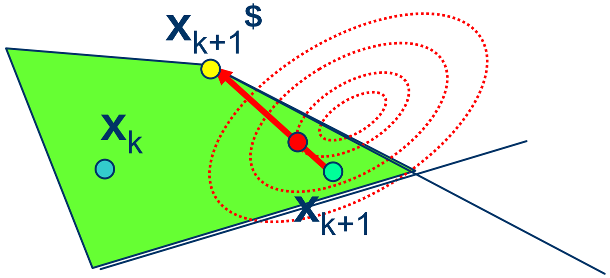
$$\mathbf{x}_{k+1} = \mathbf{x}_k + \sigma_k^*(\mathbf{x}_k^{\$} - \mathbf{x}_k)$$

De esta forma se determina un nuevo x_{k+1} iterandose hasta que se cumpla una condición del tipo:

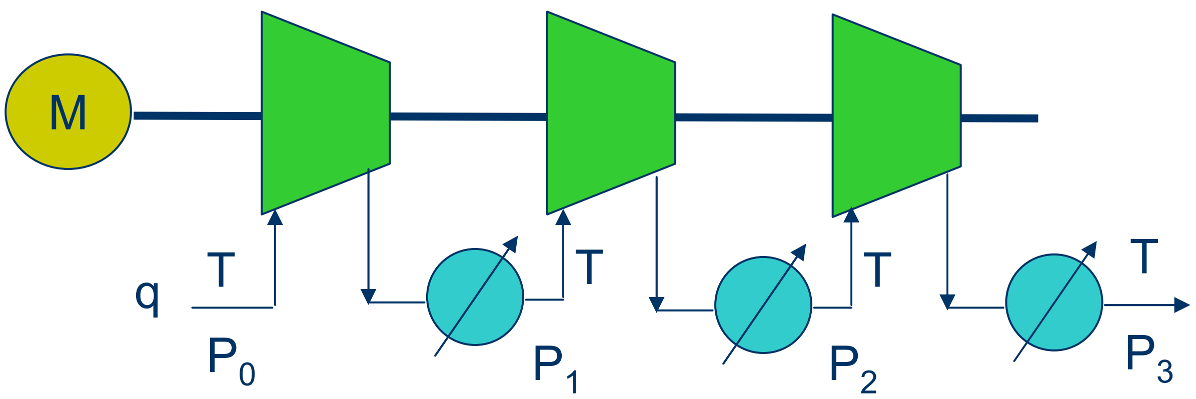
$$\frac{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|}{\varepsilon_0 + \|\mathbf{x}_k\|} \le \varepsilon_2 \qquad \frac{|J(\mathbf{x}_{k+1}) - J(\mathbf{x}_k)|}{\varepsilon_0 + |J(\mathbf{x}_k)|} \le \varepsilon_3$$

Región factible





Compresor con tres etapas



La potencia consumida por un compresor adiabático reversible en el que la entrada está a la temperatura T es:

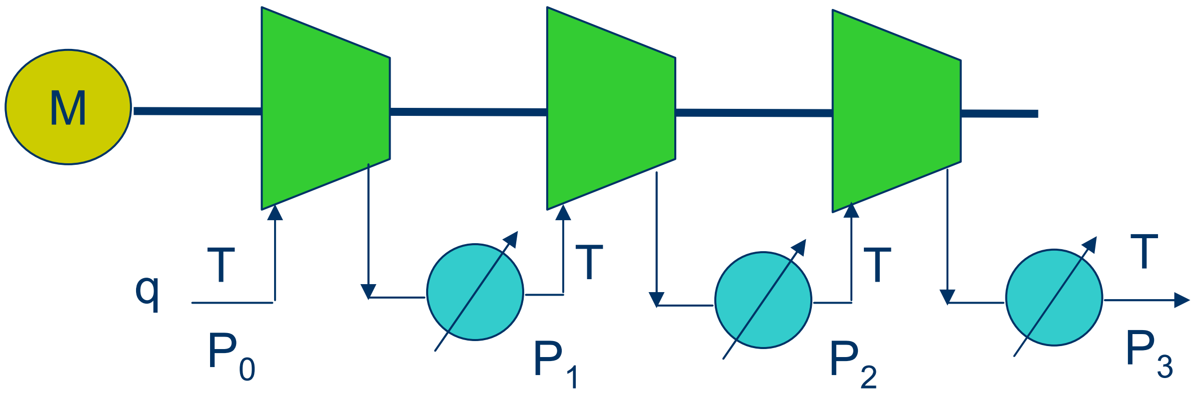
$$W = qRT \frac{\gamma}{\gamma - 1}\left[\left(\frac{P_{sal}}{P_{ent}}\right)^{\frac{\gamma-1}{\gamma}} - 1\right]$$

q moles/h T ºK γ = 4/3

Si un gas entra a 1bar y debe salir a 64 bares manteniendo q y T constantes, ¿cuales deben ser las presiones de trabajo de cada etapa intermedia para gastar la mínima energía?

$$\gamma = \frac{c_p}{c_v} \qquad R = \text{constante gases}$$

Compresor



La potencia total consumida será la suma de la que consume cada compresor

$$W_{Total} = qRT4\left[\left(\frac{P_1}{P_0}\right)^{\frac{1}{4}} + \left(\frac{P_2}{P_1}\right)^{\frac{1}{4}} + \left(\frac{P_3}{P_2}\right)^{\frac{1}{4}} - 3\right]$$

$$P_1 \geq P_0, \quad P_1 \leq P_2 \leq P_3$$

$$\min_{P_1,P_2} P_1^{\frac{1}{4}} + \left(\frac{P_2}{P_1}\right)^{\frac{1}{4}} + \left(\frac{64}{P_2}\right)^{\frac{1}{4}}$$

$$P_1 \geq 1, \quad P_1 \leq P_2 \leq 64$$

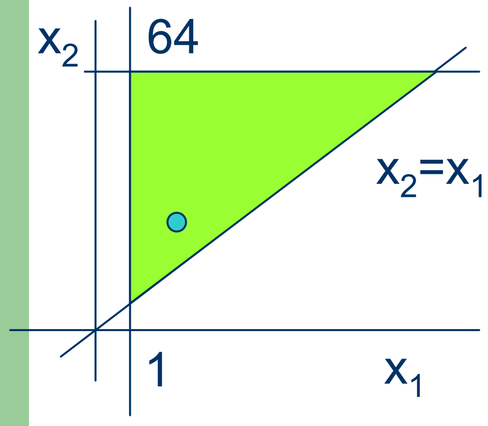
Compresor / Método de Wolfe

$$\min_{x_1,x_2} x_1^{\frac{1}{4}} + \left(\frac{x_2}{x_1}\right)^{\frac{1}{4}} + \left(\frac{64}{x_2}\right)^{\frac{1}{4}}$$

$$-x_1 \le -1$$

$$x_1 - x_2 \le 0$$

$$x_2 \le 64$$



Punto factible inicial $x_1=2$, $x_2=10$

Linealización: $J(\mathbf{x}_k) + \nabla_{\mathbf{x}} J(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)$

$$\left.\frac{\partial J}{\partial x_1}\right|_{2,10} = 0.25x_1^{-0.75}(1 - x_2^{0.25}x_1^{-0.5})\Big|_{2,10} = -0.0383$$

$$\left.\frac{\partial J}{\partial x_2}\right|_{2,10} = 0.25x_2^{-0.75}(x_1^{-0.25} - 64^{0.25}x_2^{-0.5})\Big|_{2,10} = -0.00238$$

$$\min_{x_1,x_2} (J(2,10) - 0.0383(x_1 - 2) - 0.00238(x_2 - 10))$$

$$-x_1 \le -1$$

$$x_1 - x_2 \le 0$$

$$x_2 \le 64$$

$J(2,10)=4.275$

Cesar de Prada ISA-UVA

Compresor / Método de Wolfe

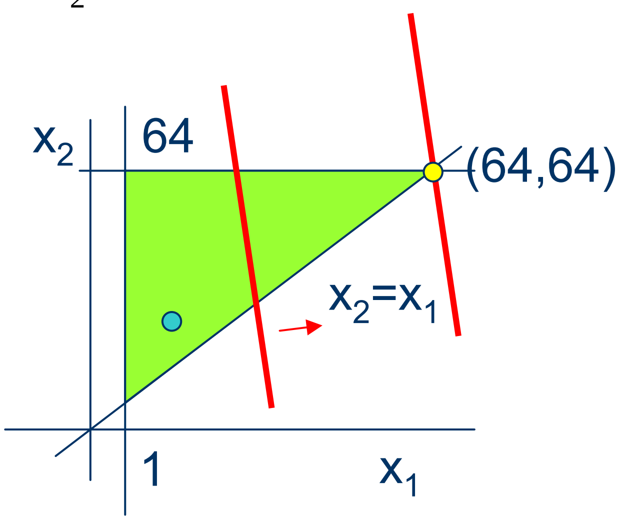
$$\min_{x_1,x_2} \; (-0.0383x_1 - 0.00238x_2)$$
$$-x_1 \le -1$$
$$x_1 - x_2 \le 0$$
$$x_2 \le 64$$

Problema LP equivalente, puede resolverse gráficamente.

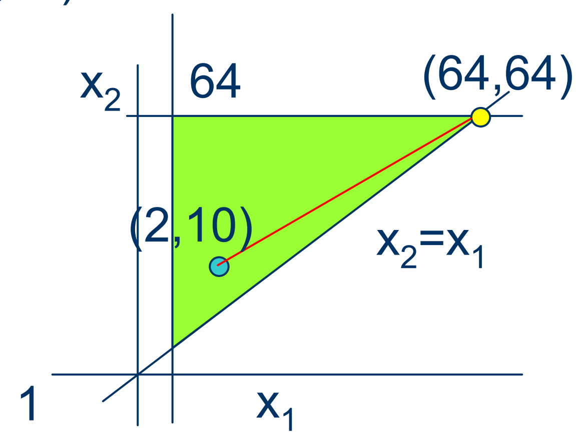
Costo: $-0.0383x_1 - 0.00238x_2 = c < 0$

$x_2 = -16.09x_1 - 420.16\,c$

Optimo LP: (64,64)

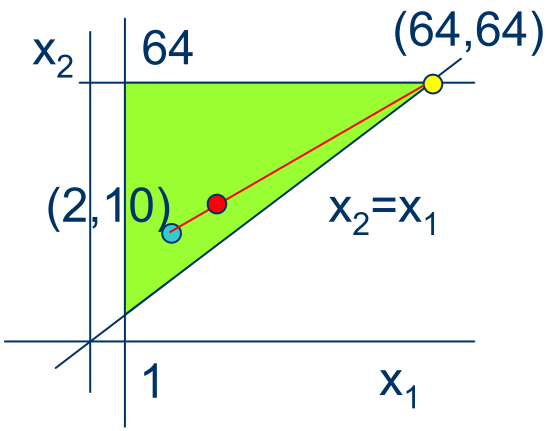


Proporciona una dirección de minimización de J



Compresor / Método de Wolfe

Paso 2: minimización en el segmento (2,10) a (64,64)



$$\min_{\alpha} J(2+\alpha(64-2), 10+\alpha(64-10))$$

$$0 \leq \alpha \leq 1$$

Puede resolverse por interpolación, sección dorada, etc. Solución: $\alpha^* = 0.027$

Nuevo punto:

x_1= 2+0.027 62=3.69

x_2= 10+0.027 54=11.47

J=4.25

Se continua iterando hasta que no haya mejora apreciable

Programación lineal sucesiva SLP

Una extensión natural del método de Wolfe aplicable al problema general NLP es linealizar J(x) y las restricciones en torno al punto x_k con lo que resulta un problema LP aproximado cuya solución x_k^* se supone estará mas próxima al óptimo NLP. A continuación se toma $x_{k+1}=x_k^*$ y se vuelve a linealizar J(x) y las restricciones iterándose hasta llegar a un punto en que no hay cambio apreciable en J y en x

$$\left.\begin{array}{l} \min\limits_{x} J(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \end{array}\right\} \approx \left.\begin{array}{l} \min\limits_{x} J(\mathbf{x}_k) + \nabla_x J(\mathbf{x}_k)(\mathbf{x}-\mathbf{x}_k) \\ \mathbf{h}(\mathbf{x}_k) + \nabla_x \mathbf{h}(\mathbf{x}_k)(\mathbf{x}-\mathbf{x}_k) = \mathbf{0} \\ \mathbf{g}(\mathbf{x}_k) + \nabla_x \mathbf{g}(\mathbf{x}_k)(\mathbf{x}-\mathbf{x}_k) \leq \mathbf{0} \end{array}\right\} \mathbf{m} \leq \mathbf{x}-\mathbf{x}_k \leq \mathbf{M}$$

Se añaden restricciones sobre las máximas desviaciones del punto de linealización que aseguran que esta es aceptable

Métodos de Penalización

La idea básica es transformar el problema NLP

$$\min_x J(\mathbf{x})$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{x}) \le \mathbf{0}$$

En otro sin restricciones tal que su solución se aproxime a la del NLP

$$\min_x V(\mathbf{x}, \eta) = \min_x J(\mathbf{x}) + \sum_i \eta_i P_i(h_i(\mathbf{x}), g_i(\mathbf{x}))$$

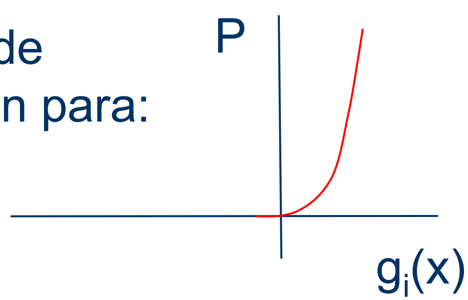
V es la función a minimizar que añade a J las funciones de penalización P_i. η_i son parámetros que se va ajustando a lo largo del algoritmo para acercar la minimización de V a la solución del problema NLP

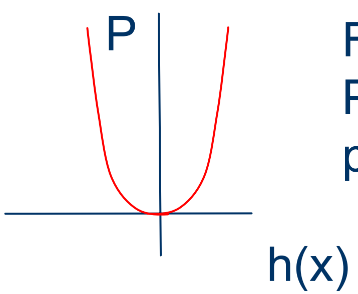
Funciones de Penalización

$$\min_x V(\mathbf{x},\eta) = \min_x J(\mathbf{x}) + \sum_i \eta_i P_i(h_i(\mathbf{x}), g_i(\mathbf{x}))$$

La principal característica de las funciones de penalización P, es que tienden a cero cuando se satisface las restricciones, pero toma un valor muy grande cuando éstas son violadas. De manera tal que el algoritmo no realiza la búsqueda del mínimo de V en esta región. Las penalizaciones P modifican la función de coste original J, incrementando su valor en aquellas regiones donde no se satisfacen las restricciones.

Funciones de Penalización para: $g_i(x) \leq 0$





Funciones de Penalización para: h(x) = 0

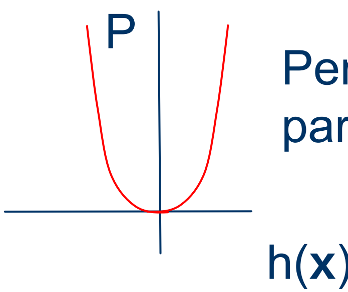
Funciones de Barrera/Penalización

$g_i(\mathbf{x}) \le 0$

Penalización externa

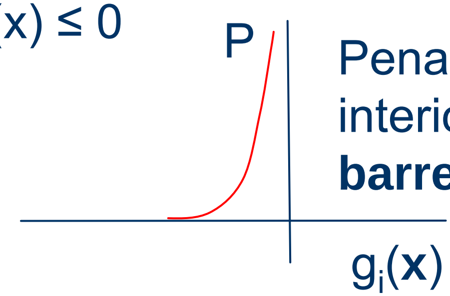
Penalización: Si se violan las restricciones, entonces P toma valores muy grandes. El óptimo puede violar ligeramente las restricciones

$h(\mathbf{x}) = 0$



Penalización parabólica

$g_i(x) \le 0$



Penalización interior o de **barrera**

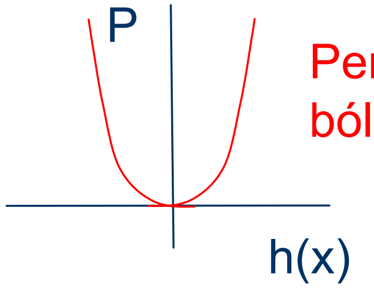
Barrera: si el valor de $g_i(\mathbf{x})$ se **aproxima** a la restricción, entonces P toma un valor muy grande. Fuerza a que x se encuentre dentro de la región factible

Funciones de Penalización

Restricciones de Igualdad $h(\mathbf{x}) = 0$

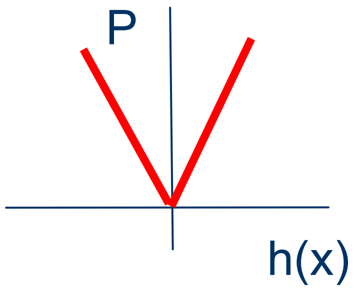
Si h(x) se desvía del valor de cero , entonces la función P crece cuadráticamente, penalizando los valores de V(x) en los puntos no factibles. El parámetro η proporciona la magnitud de la penalización.

Se debe adicionar un término $\eta(h_i(\mathbf{x}))^2$ a la función objetivo J, por cada una de las restricciones de igualdad existentes.



Penalización para bólica $\eta h(x)^2$

P, función continua con derivadas continuas

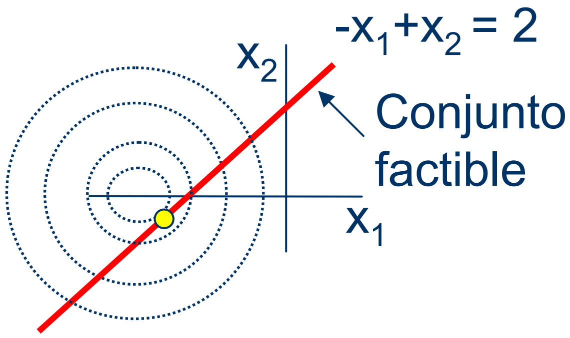


Valor Absoluto $\eta|h(x)|$, discontinuas en h(x) = 0

Ejemplo de restricción de igualdad

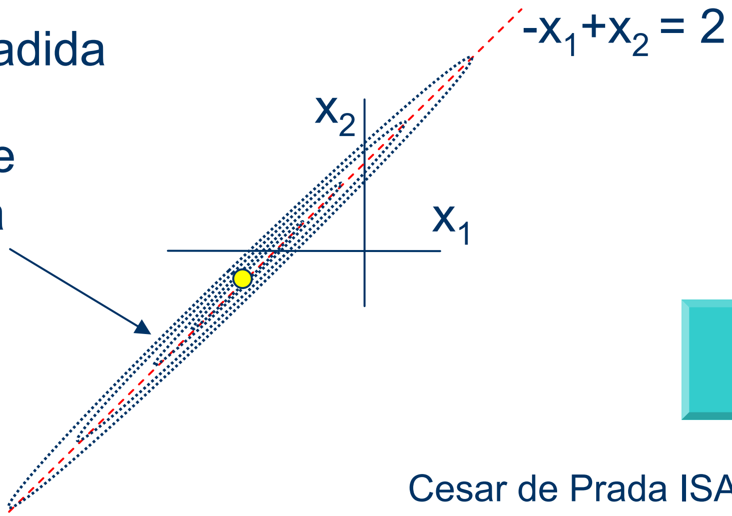
$$\left.\begin{aligned} &\min_{x_1,x_2} (x_1+3)^2 + x_2^2 \\ &-x_1 + x_2 - 2 = 0 \end{aligned}\right\}$$

Problema original



$$\min_{x_1,x_2} (x_1+3)^2 + x_2^2 + \eta(-x_1 + x_2 - 2)^2$$

Problema con la penalización añadida

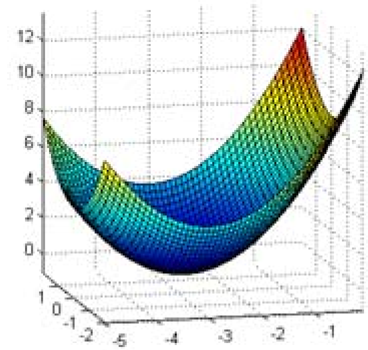


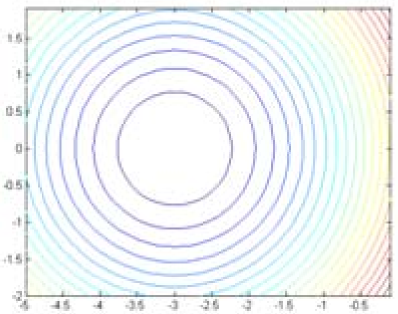
Curvas de nivel deformadas que fuerzan soluciones en torno a la restricción $-x_1+x_2=2$

(Aunque el problema es díficil numéricamente al aumentar η)

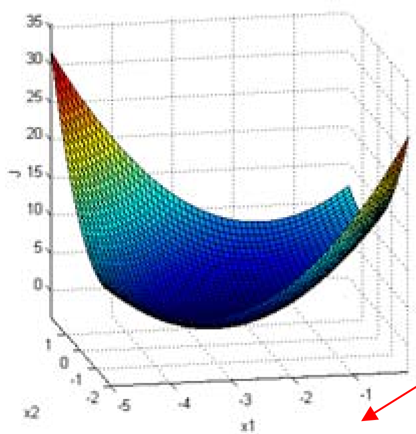
Matlab

Ejemplo 1

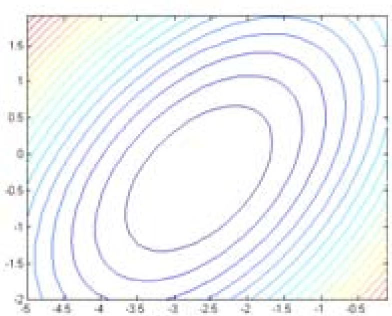


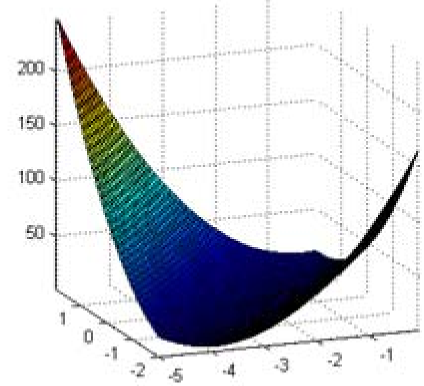


η =0 función original J(x)

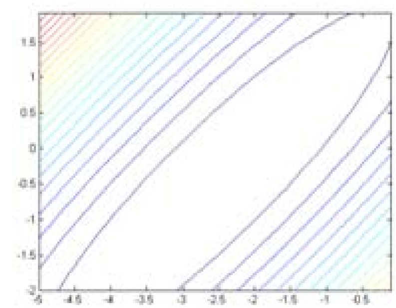


η =1 Función de Penalización V(x)=J(x)+ ηP





η =10 Función de Penalización V(x)=J(x)+ ηP



El mínimo de J(x) se encuentra alrededor de h(x)=0

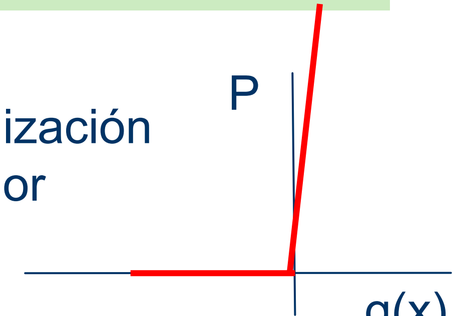
Funciones de Penalización

Restricciones de desigualdad $g(\mathbf{x}) \le 0$

Penalización Infinita

$$P(g(\mathbf{x})) = \begin{cases} 0 & \text{if } g(\mathbf{x}) \le 0 \\ 10^{20}|g(\mathbf{x})| & \text{if } g(\mathbf{x}) > 0 \end{cases}$$

Penalización Exterior

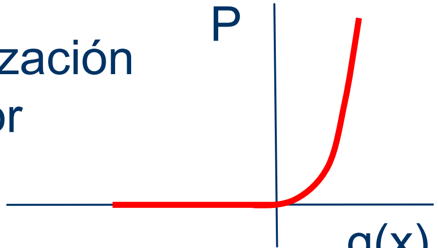


Principal inconveniente: discontinua en g(**x**) = 0

Penalización cuadrática asimétrica (brakets)

$$P(g(\mathbf{x})) = [\max(0, g(\mathbf{x}))]^2$$

Penalización exterior



Continua y con derivadas continuas

Al igual que con otras funciones de penalización: se pueden violar ligeramente las restricciones

Funciones de Penalización

Penalización Exacta

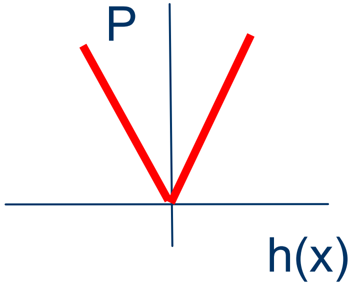
En problemas con restricciones de igualdad y desigualdad el mínimo de la función de coste:

$$\min_{\mathbf{x}} J(\mathbf{x}) + \sum_i \omega_i |h_i(\mathbf{x})| + \sum_j \sigma_i \max(0, g_j(\mathbf{x}))$$

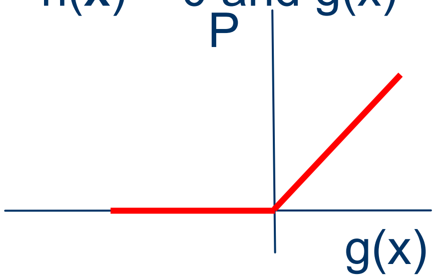
Coincide exactamente con el óptimo x^* de un problema NLP si los pesos ω_i, σ_j satisfacen:

$$0 \le \omega_i \ge |\lambda_i^*| \qquad 0 \le \sigma_i \ge |\mu_i^*|$$

donde x^*,λ^*, μ^* son la solución de las condiciones de KKT. Luenberger (1984)



Principal inconveniente: discontinuidades en h(**x**) = 0 and g(x) = 0



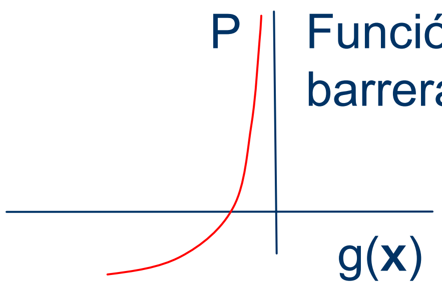
Funciones de Barrera

Barrera Logarítmica

$$P(g(\mathbf{x})) = -\ln(-g(\mathbf{x}))$$

Los puntos que se encuentran dentro de la región factible se favorecen, mientras que los que se encuentran cercanos a g(**x**) se penalizan. Se construye una barrera infinita antes de las restricciones

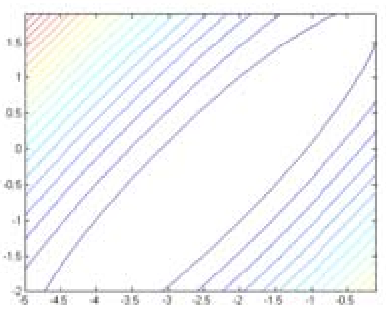
P es continua, pero si por alguna razón, las restricciones se violan, es muy difícil volver a la situación anterior

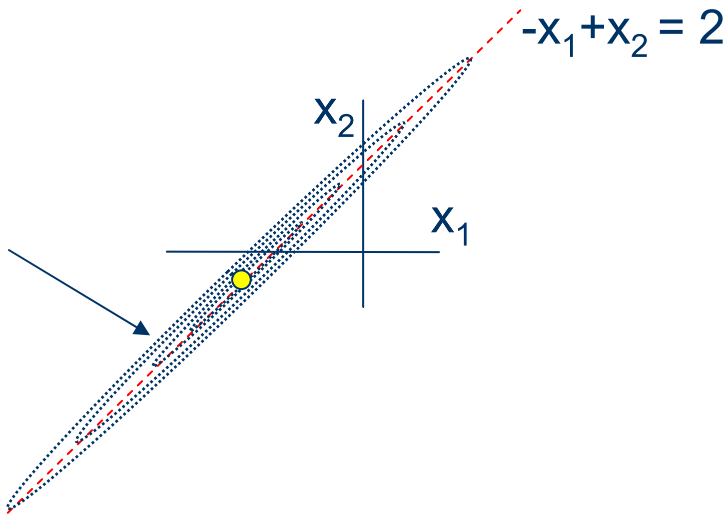


Función de barrera o interior

Nótese que con funciones de barrera, el parámetro η tiene que decrecer progresivamente para que permita un punto cercano a las restricciones

Número de Condición del Hessiano





Cuando el parámetro η se modifica para forzar a que los ptos. x_k del algoritmo estén próximos a modificar la forma de los contornos de las restricciones, el problema de mal condicionamiento se incrementa. Un número de condición de 10^5 es moderadamente grande, 10^9 es grande y 10^{14} muy grande por tanto métodos basados en la inversa del hessiano no funcionará.

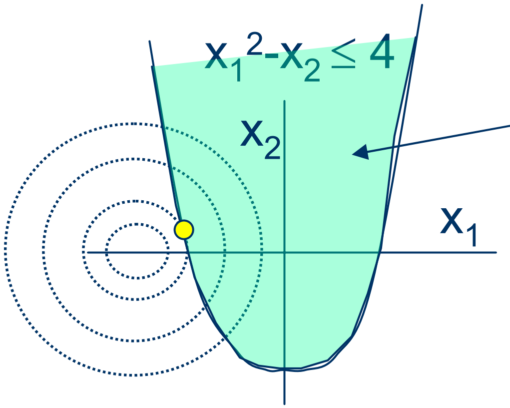
Algoritmos con funciones de penalización

1. Formular el problema NLP como uno sin restricciones con funciones de penalización añadidas $V(\mathbf{x},\eta) = J(\mathbf{x}) + \Sigma\eta_i\, P_i(g_i(\mathbf{x}),h_i(\mathbf{x}))$
2. Resolver el problema de minimizar V respecto a x con un valor de η_i fijo
3. Modificar cada η_i de acuerdo a una cierta regla, aumentándolos si P es una penalización exterior y disminuyéndolos si es una barrera interior
4. Comprobar las condiciones de finalización, si no se cumplen volver al paso 2

Ejemplo de restricciones de desigualdad

$$\left.\begin{array}{l}\min\limits_{x_1,x_2} (x_1+3)^2 + x_2^2 \\ x_1^2 - x_2 - 4 \le 0\end{array}\right\}$$

Problema original

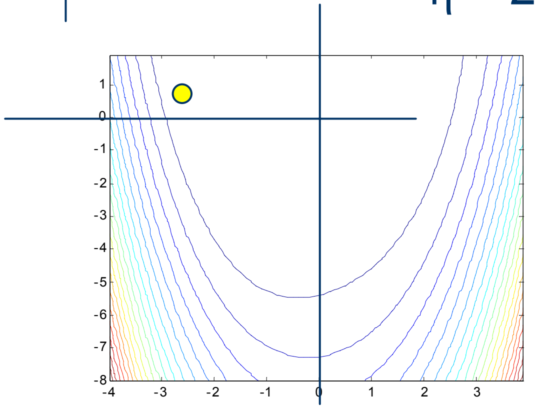


Problema con penalización añadida tipo cuadrática asimétrica

$$\min_{x_1,x_2} (x_1+3)^2 + x_2^2 + \eta P(g(x))$$

$$P(g(x)) = \begin{cases} 0 \text{ si } x_1^2 - x_2 - 4 \le 0 \\ (x_1^2 - x_2 - 4)^2 \text{ si } x_1^2 - x_2 - 4 > 0 \end{cases}$$

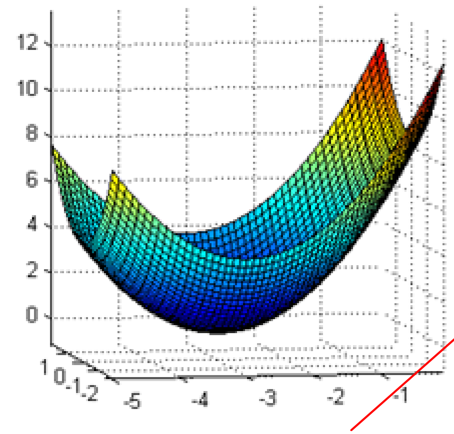
$\eta = 2$

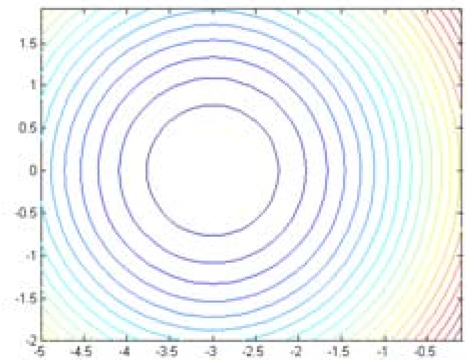


Matlab

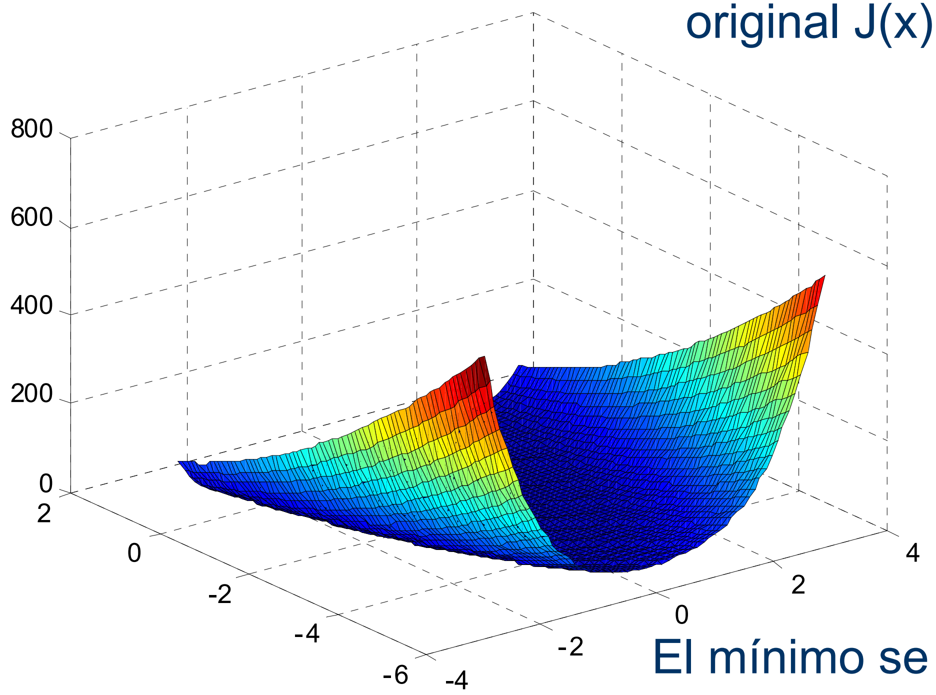
Cesar de Prada ISA-UVA

Ejemplo 2

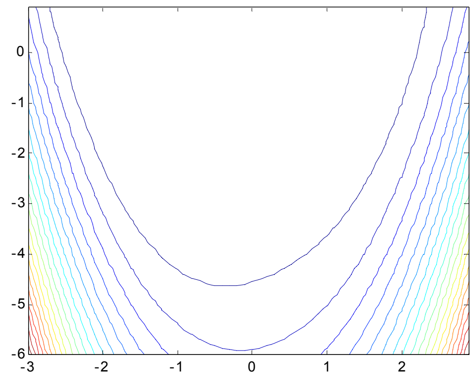




η =0 función original J(x)



η =5 función de penalización V(x)=J(x)+ ηP



El mínimo se encuentra dentro de la región $g(x)\leq 0$

Cesar de Prada ISA-UVA

Restricciones duras y blandas

- Las restricciones se suelen clasificar en duras (hard) y blandas (soft). Las primeras son aquellas que, como las leyes físicas, ciertos límites, etc., deben cumplirse estrictamente. Las segundas son aquellas en las que se permite una cierta desviación sobre el límite, como en especificaciones, capacidades,..
- El concepto de penalización es útil cuando se considera que existen restricciones blandas en las cuales se puede permitir una cierta violación de las mismas a costa de pagar un precio por ello.
- Los problemas con funciones de penalización se denominan a veces “elásticos” en contraposición a las formulaciones clasicas “inelásticas”

Factibilidad / variables de holgura

- Un procedimiento alternativo a las funciones de penalización para usar restricciones "blandas" y garantizar que existe solución factible con métodos tipo LP, QP, SLP, SQP, etc. es la introducción de variables de holgura en el lado derecho de las restricciones, que deben minimizarse, acompañadas de su correspondiente penalización en el índice J.

$$\min_{x,\varepsilon,\delta} J(\mathbf{x}) + \alpha\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} + \beta\boldsymbol{\delta}'\boldsymbol{\delta}$$

$$\mathbf{h}(\mathbf{x}) = \boldsymbol{\varepsilon}$$

$$\mathbf{g}(\mathbf{x}) \leq \boldsymbol{\delta}$$

$$\boldsymbol{\delta} \geq \mathbf{0}$$

Si existe solución factible del problema original, obviamente el optimo dara $\varepsilon = 0$, $\delta = 0$, y se tiene la misma solución. Pero si no existe, ε y δ amplian la región factible, justo lo mínimo para que exista dicha solución factible.