Escuela Técnica Superior de Ingeniería

Departamento de Organización Industrial

Optimización no lineal

Andrés Ramos

Universidad Pontificia Comillas

http://www.iit.comillas.edu/aramos/

Andres.Ramos@comillas.edu

CONTENIDO

- ➢ PROBLEMAS DE PROGRAMACIÓN NO LINEAL
- ❑ TIPOS DE PROBLEMAS NLP
- ❑ CLASIFICACIÓN DE MÉTODOS DE RESOLUCIÓN SIN RESTRICCIONES
- ❑ CONDICIONES DE OPTIMALIDAD PARA OPTIMIZACIÓN SIN RESTRICCIONES
- ❑ CONDICIONES DE OPTIMALIDAD PARA NLP
- ❑ METHODS FOR UNCONSTRAINED OPTIMIZATION (master)
- ❑ NONLINEAR PROGRAMMING METHODS (master)

Problemas de programación no lineal (i)

- Problema de transporte con descuentos por cantidad
  - ✓ El precio unitario de transporte entre un origen y un destino es decreciente en función de la cantidad a transportar.
- Problema de flujo de cargas en un sistema eléctrico
  - ✓ Las pérdidas son no lineales
- Problema de producción con elasticidad en el precio y/o en el coste
  - ✓ Curva de demanda o curva demanda-precio $p(x)$ representa el precio unitario que se necesita para poder vender x unidades. Es una función decreciente, nunca inferior al coste unitario de producción c. Los ingresos brutos (producto de cantidad producida por precio) es una expresión no lineal. Margen de contribución

$$f(x) = xp(x) - cx$$

  - ✓ Los costes no lineales pueden aparecer por una mayor eficiencia unitaria en función de la cantidad.

Problemas de programación no lineal (ii)

- Problema de selección de una cartera de inversiones

  n tipos de acciones

  x_j, j=1,…,n representan el número de acciones j que se van a incluir en la cartera

  μ_j y σ_{jj} la media y la varianza históricas del rendimiento sobre cada acción de tipo j, en donde σ_{jj} es una medida del riesgo de estas acciones. Sea σ_{ij} la covarianza del rendimiento sobre una acción de cada tipo i y j.

  $R(x)$ rendimiento esperado y su varianza $V(x)$

$$R(x) = \sum_{j=1}^{n} \mu_j x_j$$

$$V(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j$$

  la función objetivo es $f(x) = R(x) - \beta V(x)$

  siendo β el factor de aversión al riesgo.

CONTENIDO

- PROBLEMAS DE PROGRAMACIÓN NO LINEAL
- TIPOS DE PROBLEMAS NLP
- CLASIFICACIÓN DE MÉTODOS DE RESOLUCIÓN SIN RESTRICCIONES
- CONDICIONES DE OPTIMALIDAD PARA OPTIMIZACIÓN SIN RESTRICCIONES
- CONDICIONES DE OPTIMALIDAD PARA NLP
- METHODS FOR UNCONSTRAINED OPTIMIZATION (master)
- NONLINEAR PROGRAMMING METHODS (master)

Optimización no lineal (i)

- Optimización SIN restricciones

$$
\begin{aligned}
&\min_x f(x)\\
&f(x):\mathbb{R}^n \to \mathbb{R}\\
&x \in \mathbb{R}^n
\end{aligned}
$$

- Optimización CON restricciones (Programación No Lineal NLP)

$$
\begin{aligned}
&\min_x f(x)\\
&g_i(x) = 0 \quad i \in \varepsilon\\
&g_i(x) \le 0 \quad i \in \phi\\
&f(x):\mathbb{R}^n \to \mathbb{R}\\
&g_i(x):\mathbb{R}^n \to \mathbb{R}\\
&x \in \mathbb{R}^n
\end{aligned}
$$

Optimización no lineal (ii)

- Programación cuadrática

$$\min_x f(x) = \frac{1}{2} x^T Q x - b^T x$$
$$Ax = b$$

- Programación convexa

$f(x)$ es convexa (cóncava si es maximización) y $g_i(x)$ es convexa, $\forall i = 1, \ldots, m$

- Programación separable

la función se puede separar en una suma de funciones de las variables individuales $f(x) = \sum_{j=1}^{n} f_j(x_j)$

- Programación geométrica

función objetivo y restricciones toman la forma $g(x) = \sum_{j=1}^{n} c_j P_j(x)$

$$P_j(x) = x_1^{a_{j1}} x_2^{a_{j2}} \ldots x_n^{a_{jn}}, \quad j = 1, \ldots, n$$

CONTENIDO

- PROBLEMAS DE PROGRAMACIÓN NO LINEAL
- TIPOS DE PROBLEMAS NLP
- CLASIFICACIÓN DE MÉTODOS DE RESOLUCIÓN SIN RESTRICCIONES
- CONDICIONES DE OPTIMALIDAD PARA OPTIMIZACIÓN SIN RESTRICCIONES
- CONDICIONES DE OPTIMALIDAD PARA NLP
- METHODS FOR UNCONSTRAINED OPTIMIZATION (master)
- NONLINEAR PROGRAMMING METHODS (master)

Clasificación métodos optimización SIN restricciones según el uso de derivadas

- Sin derivadas
  - ✓ Necesarios cuando no se pueden calcular éstas
- Primeras derivadas (gradiente)
- Segundas derivadas (hessiano)
  - ✓ Mayor coste computacional
  - ✓ Mejores propiedades de convergencia

Expansión en serie de Taylor

- Aproxima una función f cerca de un punto dado x_0
- Se necesitan conocer las derivadas de la función

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

$p \in \mathbb{R}^n$ un vector diferente de 0

$f(x)$ valor de la función

$\nabla f(x)$ gradiente de la función

$\nabla^2 f(x)$ hessiano de la función (si f tiene segundas derivadas continuas es una matriz simétrica)

- O alternativamente

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

siendo un punto ξ entre x y x_0

Función cuadrática (i)

$$f(x) = \frac{1}{2} x^T Q x - b^T x$$

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

$$b \in \mathbb{R}^n$$

$$Q \in \mathbb{R}^n \times \mathbb{R}^n$$

- Gradiente

$$\nabla f(x) = Qx - b$$

- Hessiano

$$\nabla^2 f(x) = Q$$

Función cuadrática (ii)

$$f(x,y)=\frac{1}{2}x^2+2xy+\frac{1}{2}y^2-y+9$$

$$\nabla f(x,y)=\begin{pmatrix} x+2y \\ 2x+y-1 \end{pmatrix} \qquad \nabla^2 f(x,y)=\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

- Estamos en $x_0=(1,-1)$ $f(1,-1)=9$ y queremos saber el valor de la función en $x_1=(1.1,-0.9)$
- Evaluación directa

$$f(1.1,-0.9)=0.605-1.98+0.405+0.9+9=8.93$$

- Aproximación mediante expansión en serie de Taylor

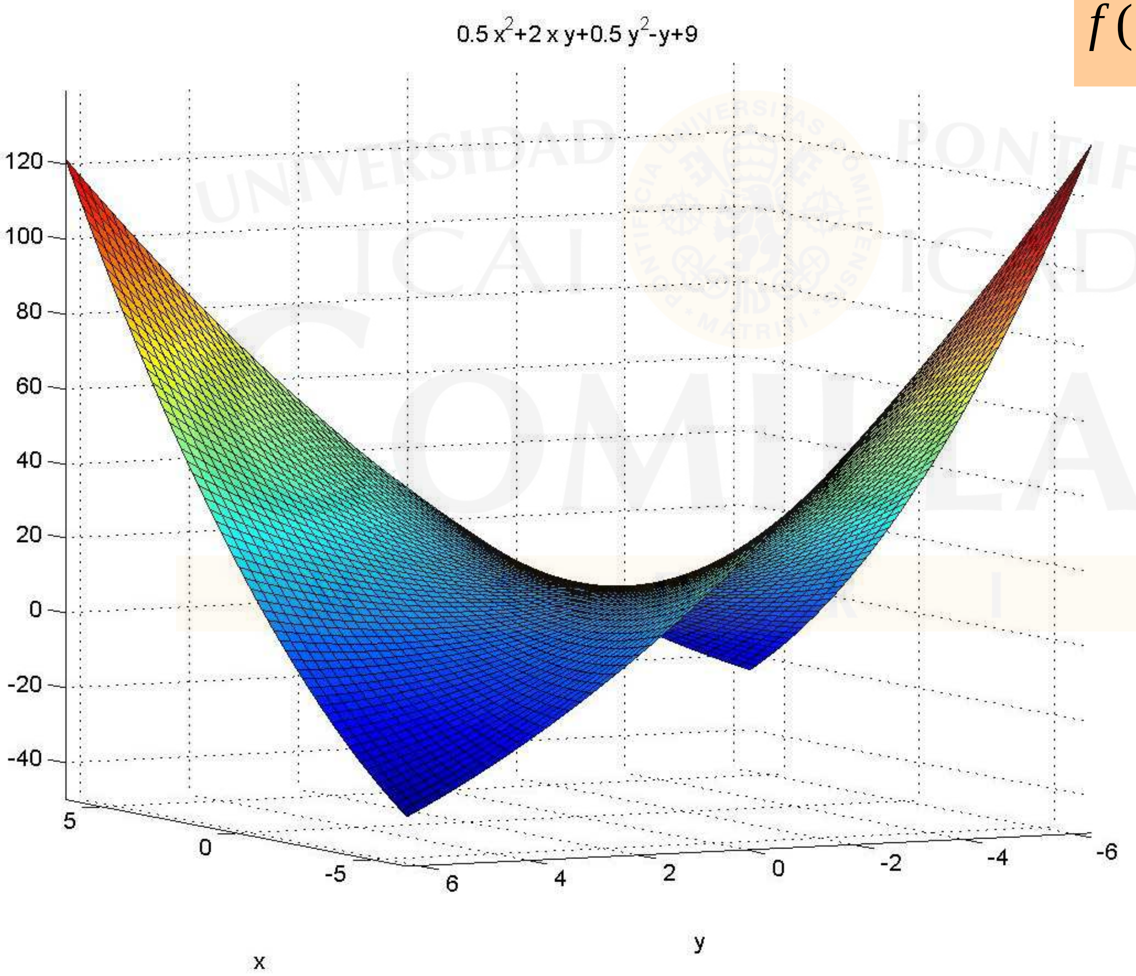
$$f(1.1,-0.9)=9+\begin{pmatrix}-1 & 0\end{pmatrix}\begin{pmatrix}0.1\\0.1\end{pmatrix}+\frac{1}{2}\begin{pmatrix}0.1 & 0.1\end{pmatrix}\begin{pmatrix}1 & 2\\2 & 1\end{pmatrix}\begin{pmatrix}0.1\\0.1\end{pmatrix}=9-0.1+0.03=8.93$$

- La aproximación de Taylor de 2º orden es exacta en una función cuadrática

Función cuadrática (iii)

```
ezsurf('0.5*x^2+2*x*y+0.5*y^2-y+9')
```

$$f(x,y) = \frac{1}{2}x^2 + 2xy + \frac{1}{2}y^2 - y + 9$$



Mínimo local, global (i)

- Sea la función *f* diferenciable con primera y segunda derivadas continuas

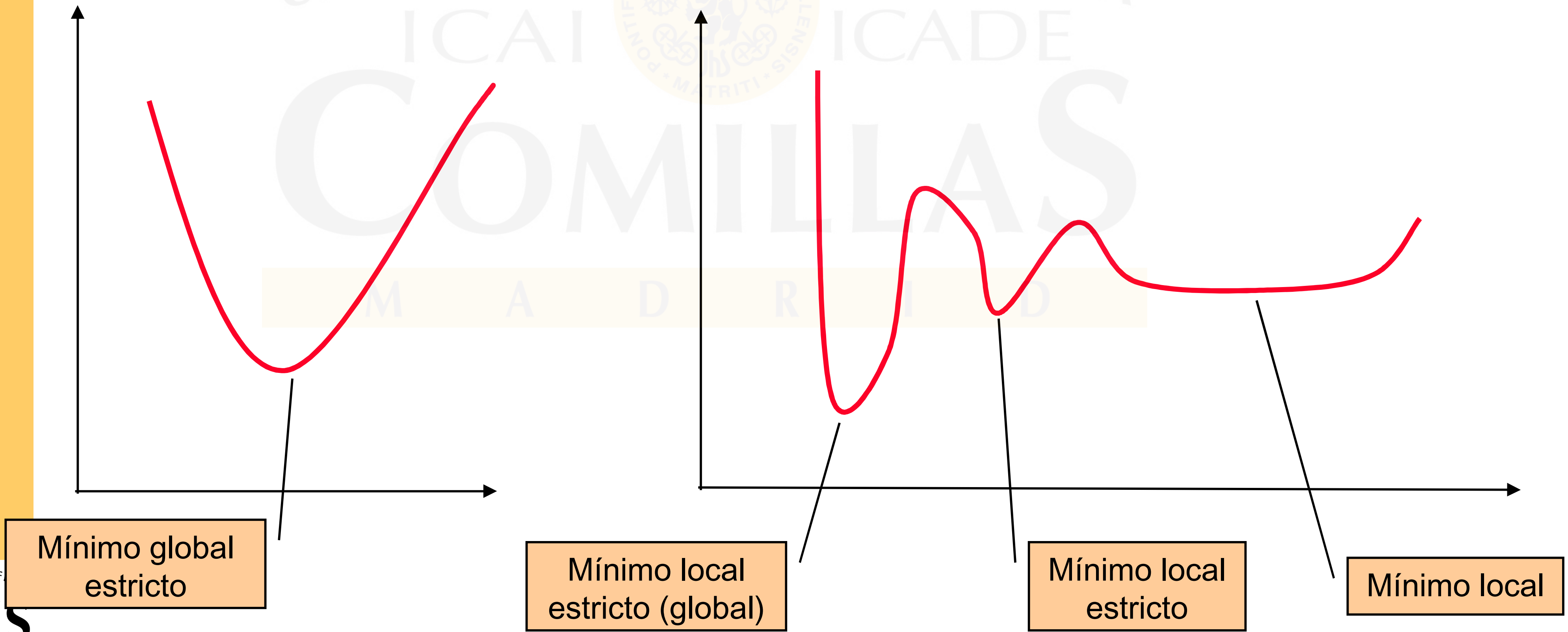
$$\min_x f(x)$$
$$f(x): \mathbb{R}^n \rightarrow \mathbb{R}$$
$$x \in \mathbb{R}^n$$

$x^* \in \mathbb{R}^n$ es el óptimo de la función

- ✓ Es mínimo global si $f(x^*) \leq f(x)$ para $x \in \mathbb{R}^n$
- ✓ Es mínimo global estricto si $f(x^*) < f(x)$ para $x \in \mathbb{R}^n$
- ✓ Es mínimo local si $f(x^*) \leq f(x)$ en su vecindad $\|x - x^*\| < \varepsilon$
  siendo ε un número positivo (típicamente pequeño) cuyo valor puede depender de x^*
- ✓ Es mínimo local estricto si $f(x^*) < f(x)$ en su vecindad $\|x - x^*\| < \varepsilon$

Mínimo local, global (ii)

- El mínimo global es difícil de encontrar. Muchos métodos son locales.
- Sólo bajo supuestos adicionales (convexidad) se puede garantizar que es global.



CONTENIDO

- PROBLEMAS DE PROGRAMACIÓN NO LINEAL
- TIPOS DE PROBLEMAS NLP
- CLASIFICACIÓN DE MÉTODOS DE RESOLUCIÓN SIN RESTRICCIONES
- ➢ CONDICIONES DE OPTIMALIDAD PARA OPTIMIZACIÓN SIN RESTRICCIONES
- CONDICIONES DE OPTIMALIDAD PARA NLP
- METHODS FOR UNCONSTRAINED OPTIMIZATION (master)
- NONLINEAR PROGRAMMING METHODS (master)

Optimización SIN restricciones
Condiciones de optimalidad (i)

$$\min_x f(x)$$
$$x \in \mathbb{R}^n$$

- Condición necesaria de 1º orden
  - ✓ Si x^* es un mínimo local de f entonces necesariamente $\nabla f(x^*) = 0$
  - ✓ Condición satisfecha también por cualquier punto estacionario
- Condición necesaria de 2º orden
  - ✓ Si x^* es un mínimo local de f entonces necesariamente $\nabla^2 f(x^*)$ es una matriz semidefinida positiva (equivalente a convexa, curvatura positiva)
- Condición suficiente de 2º orden
  - ✓ Si $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es definida positiva entonces x^* es un mínimo local estricto de f
- Condición necesaria y suficiente
  - ✓ Sea f convexa y diferenciable en x^* (si es dos veces diferenciable el hessiano será semidefinido positivo); x^* es un mínimo local si y sólo si $\nabla f(x^*) = 0$. x^* es un mínimo global si y sólo si f convexa en toda la región

Matriz definida positiva

- Una matriz *A* es definida positiva si $x^T A x > 0$ para cualquier vector no nulo *x*
- O bien si todos sus autovalores son positivos
- O bien si todos los menores **de esquina** (determinantes de orden 1,2,…,*n* (siendo *n* la dimensión de la matriz) obtenidos añadiendo filas y columnas consecutivas desde el primer elemento) son positivos.

Es definida positiva $\nabla^2 f(x,y) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

No lo es $\nabla^2 f(x_1,x_2) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

Matriz semidefinida positiva

- Una matriz A es semidefinida positiva si $x^T A x \geq 0$ para cualquier vector no nulo x
- O bien si todos sus autovalores son no negativos
- O bien si todos los menores **principales** de orden 1,2,…,n (siendo n la dimensión de la matriz) son no negativos.
- Menores principales son los determinantes de k filas cualesquiera cruzadas con sus correspondientes columnas

$\delta_1 = |1| = 1$ $\quad \delta_2 = |12| = 12$ $\quad \delta_3 = |4| = 4$

$$\delta_{1,2} = \begin{vmatrix} 1 & -3 \\ -3 & 12 \end{vmatrix} = 3 \quad \delta_{1,3} = \begin{vmatrix} 1 & -1 \\ -1 & 4 \end{vmatrix} = 3 \quad \delta_{2,3} = \begin{vmatrix} 12 & 6 \\ 6 & 4 \end{vmatrix} = 12$$

$$\delta_{1,2,3} = \begin{vmatrix} 1 & -3 & -1 \\ -3 & 12 & 6 \\ -1 & 6 & 4 \end{vmatrix} = 0$$

- Es semidefinida positiva

$$\nabla^2 f(x,y,z) = \begin{pmatrix} 1 & -3 & -1 \\ -3 & 12 & 6 \\ -1 & 6 & 4 \end{pmatrix}$$

Matriz definida negativa

- Una matriz *A* es definida negativa si $x^T A x \leq 0$ para cualquier vector no nulo *x*
- O bien todos sus autovalores son negativos
- O bien si los menores de esquina son, alternativamente, negativo, positivo, etc.
- O bien si la matriz -*A* (resultado de cambiar el signo de todos los elementos) es definida positiva.

Equivalencias

Función convexa	Función cóncava
Hessiano matriz definida positiva	Hessiano matriz definida negativa
Curvatura (2ª derivada) positiva	Curvatura (2ª derivada) negativa

Mínimo local + convexidad en toda la región	Máximo local + concavidad en toda la región
Mínimo global	Máximo global

Semidefinida positiva (negativa)	Definida positiva (negativa)
Mínimo (máximo) local	Mínimo (máximo) local estricto

Ejemplo 1: Optimización SIN restricciones

$$f(x,y)=\frac{1}{2}x^2+2xy+\frac{1}{2}y^2-y+9$$

- Gradiente = 0 es un sistema de ecuaciones

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

con solución $(x,y)=(2/3,-1/3)$

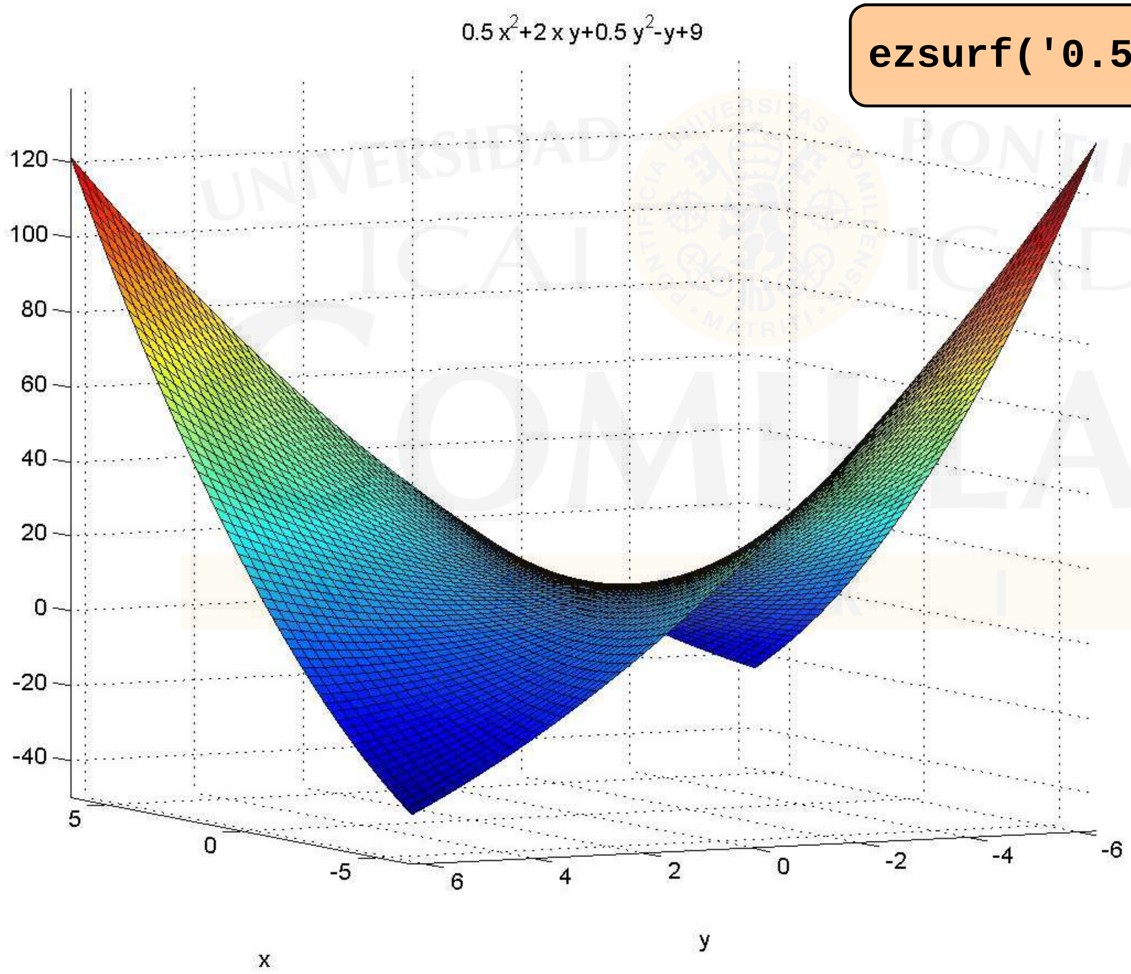
- Hessiano $\nabla^2 f(x,y)=\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ es matriz indefinida (no es ni semidefinida positiva ni semidefinida negativa)

Luego no es ni un mínimo ni máximo local

Ejemplo 1: Optimización SIN restricciones

$$f(x,y)=\frac{1}{2}x^2+2xy+\frac{1}{2}y^2-y+9$$

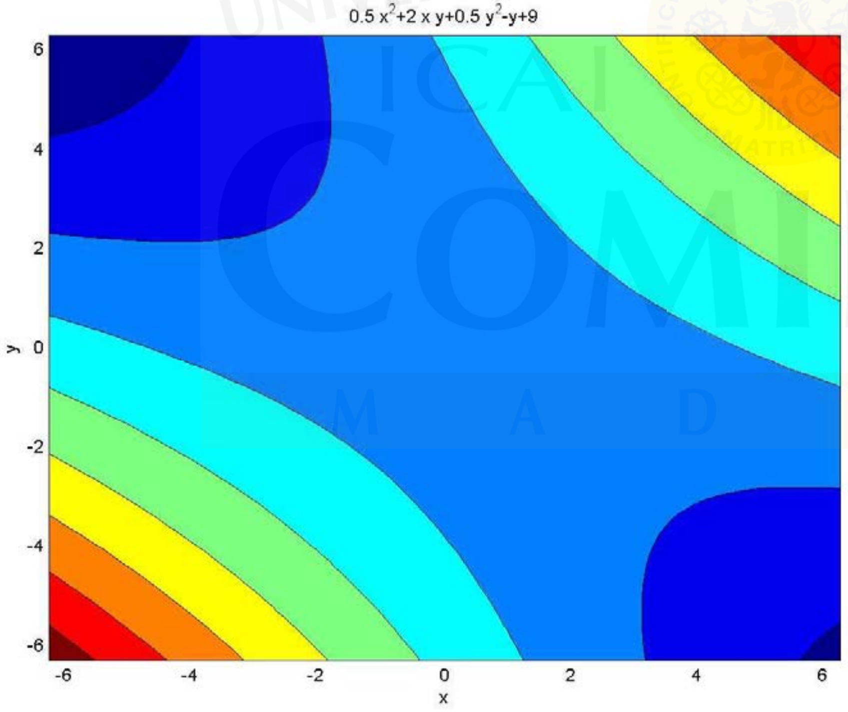
```
ezsurf('0.5*x^2+2*x*y+0.5*y^2-y+9')
```



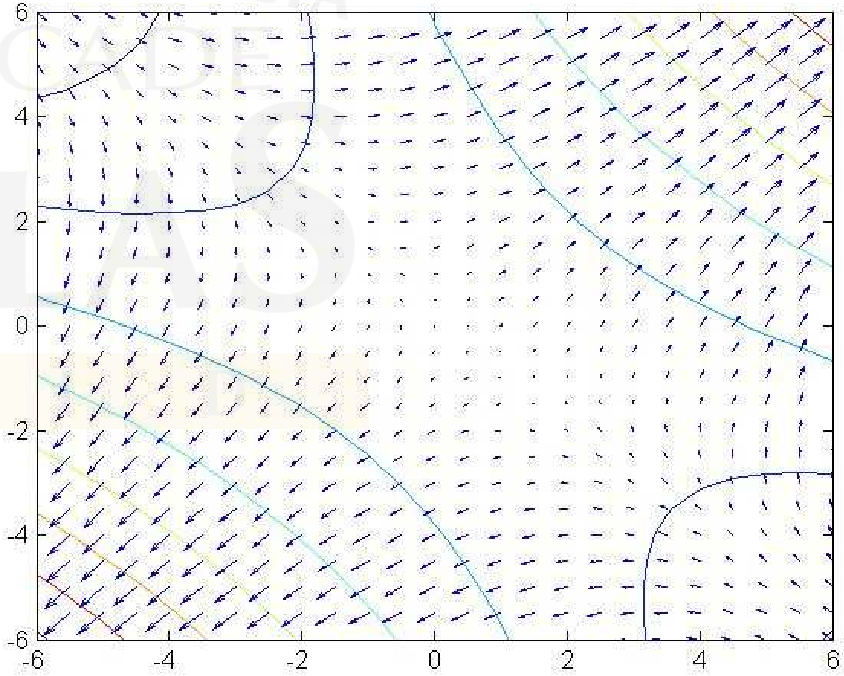
Ejemplo 1: Optimización SIN restricciones

$$f(x,y)=\frac{1}{2}x^2+2xy+\frac{1}{2}y^2-y+9$$

```
ezcontourf('0.5*x^2+2*x*y+0.5*y^2-y+9')
```

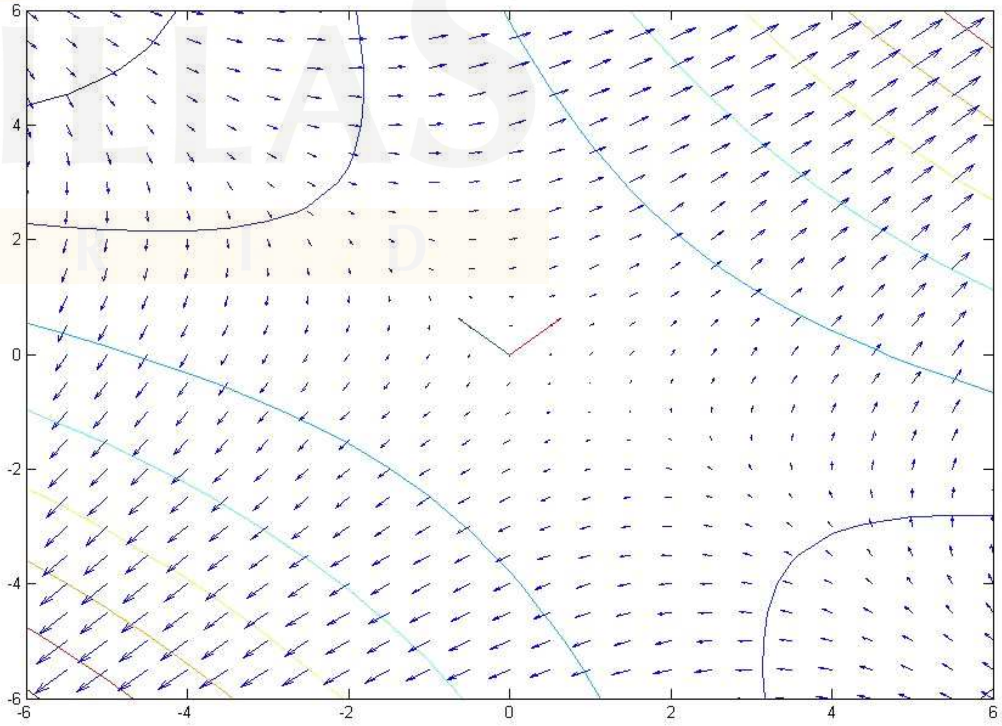



```
[x,y] = meshgrid(-6:0.5:6,-6:0.5:6);
z=0.5*x.^2+2*x.*y+0.5*y.^2-y+9;
[px,py] = gradient(z,0.5,0.5);
contour(x,y,z); hold on
quiver(x,y,px,py)
```



Autovalores y autovectores

- Hessiano $\nabla^2 f(x,y) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ es matriz indefinida
- Sus autovalores son $\lambda = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$
- Y los autovectores $v_1 = \begin{pmatrix} -0.71 \\ 0.71 \end{pmatrix}$ $v_2 = \begin{pmatrix} 0.71 \\ 0.71 \end{pmatrix}$



Ejemplo 2: Optimización SIN restricciones

$$f(x,y) = 2x^2 + y^2 - 2xy + 2x^3 + x^4$$

- Gradiente = 0 es un sistema de ecuaciones

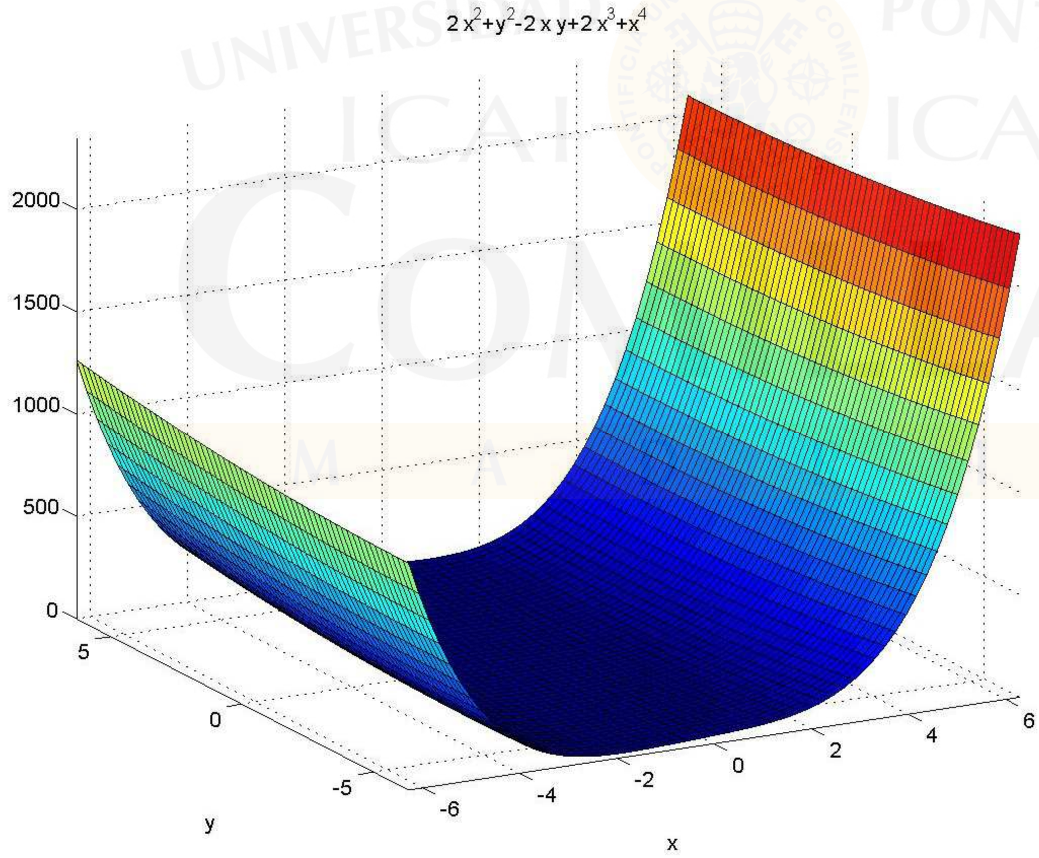
$$\nabla f(x,y) = \begin{pmatrix} 4x - 2y + 6x^2 + 4x^3 \\ 2y - 2x \end{pmatrix} = 0$$

- Hessiano $\nabla^2 f(x,y) = \begin{pmatrix} 4 + 12x + 12x^2 & -2 \\ -2 & 2 \end{pmatrix}$

Ejemplo 2: Optimización SIN restricciones

$$f(x,y) = 2x^2 + y^2 - 2xy + 2x^3 + x^4$$

```
ezsurf('2*x^2+y^2-2*x*y+2*x^3+x^4')
```

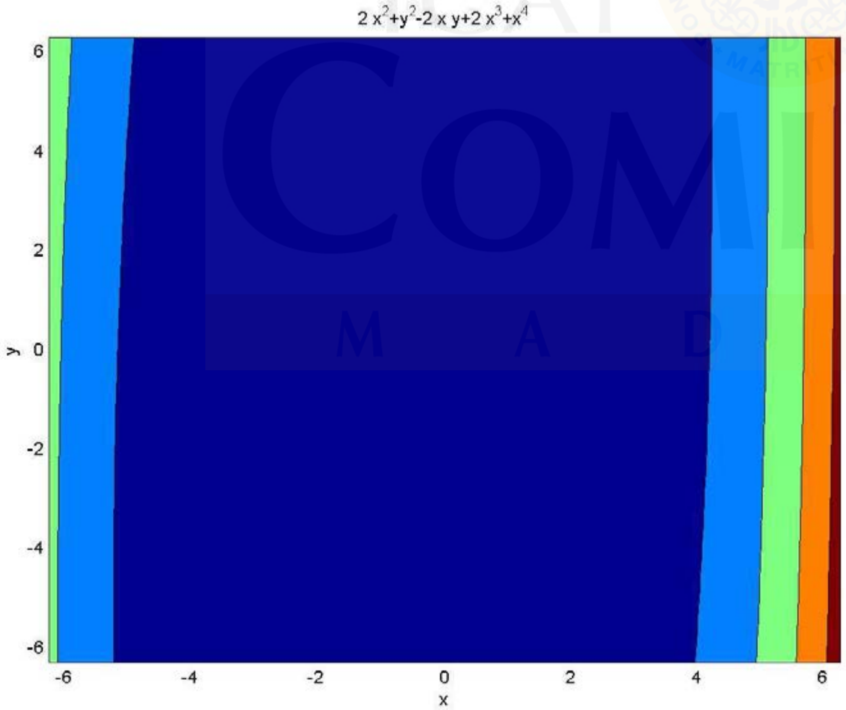


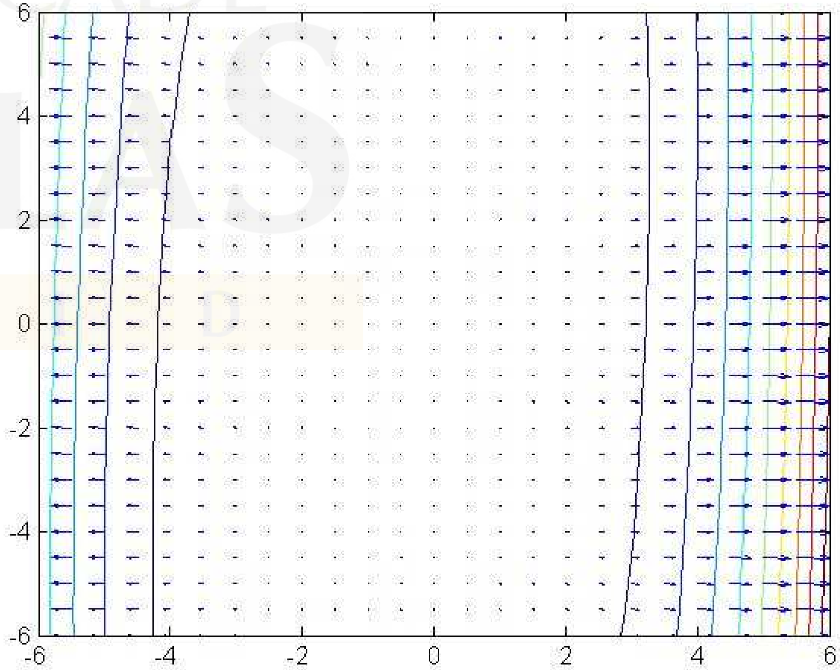
Ejemplo 2: Optimización SIN restricciones

$$f(x,y) = 2x^2 + y^2 - 2xy + 2x^3 + x^4$$

```
ezcontourf('2*x^2+y^2-2*x*y+2*x^3+x^4')
```

```
[x,y] = meshgrid(-6:0.5:6, -6:0.5:6);
z=2*x.^2+y.^2-2*x.*y+2*x.^3+x.^4;
[px,py] = gradient(z,0.5,0.5);
contour(x,y,z);
hold on, quiver(x,y,px,py)
```





Ejemplo 3: Optimización SIN restricciones

$$f(x,y) = (x-2)^2 + (y-1)^2$$

- Gradiente = 0 es un sistema de ecuaciones

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

con solución $(x,y) = (2,1)$

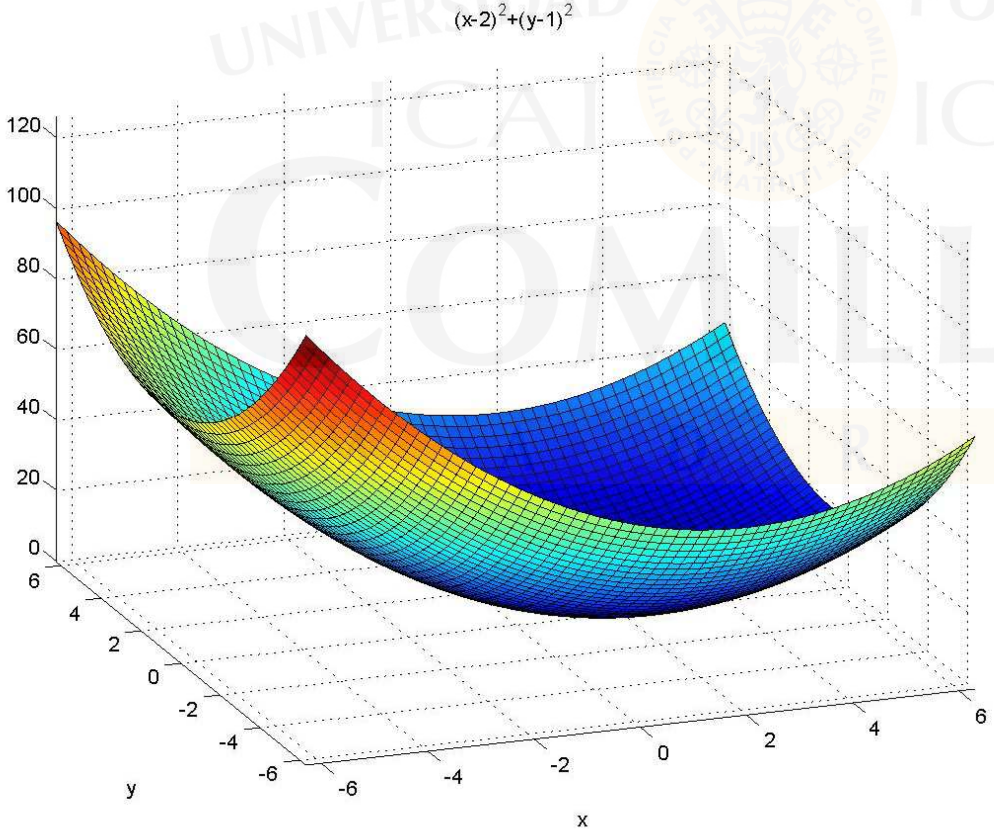
- Hessiano $\nabla^2 f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ es matriz semidefinida positiva

e independiente del punto, luego es un mínimo global

Ejemplo 3: Optimización SIN restricciones

$$f(x,y)=(x-2)^2+(y-1)^2$$

```
ezsurf('(x-2)^2+(y-1)^2')
```

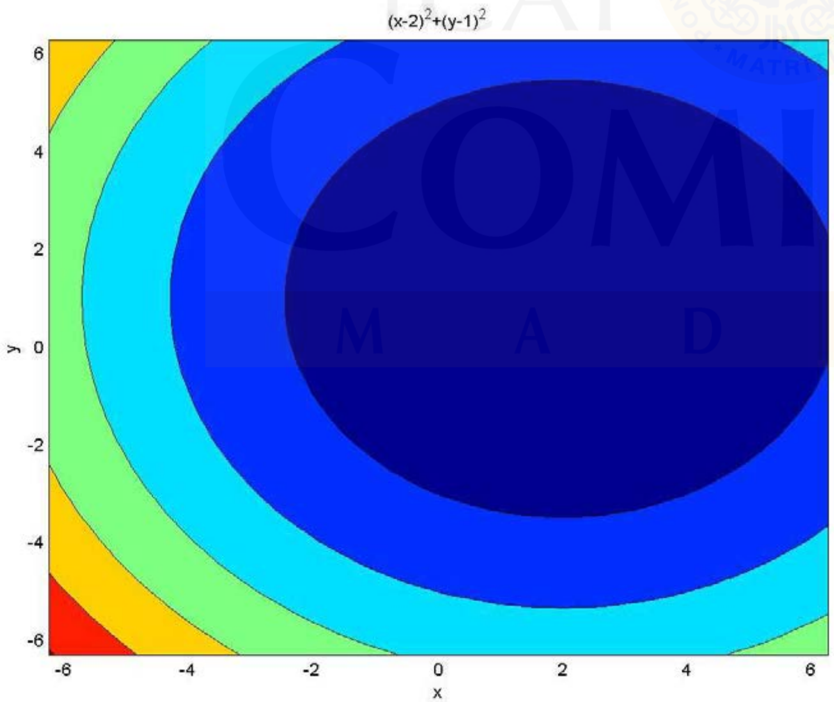


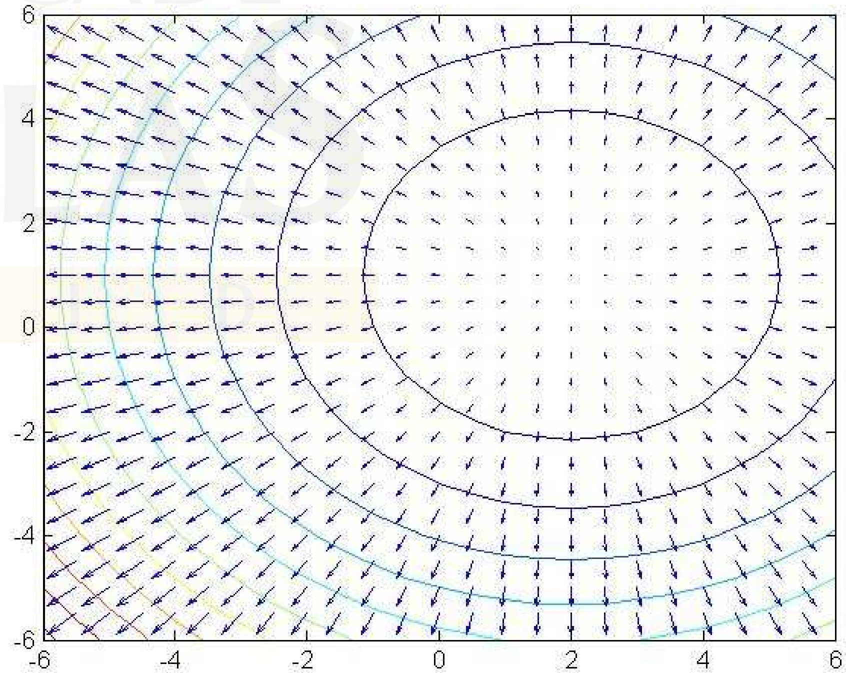
Ejemplo 3: Optimización SIN restricciones

$$f(x,y) = (x-2)^2 + (y-1)^2$$

```
[x,y] = meshgrid(-6:.5:6, -6:.5:6);
z=(x-2).^2+(y-1).^2;
[px,py] = gradient(z,0.5,0.5);
contour(x,y,z);
hold on, quiver(x,y,px,py)
```

```
ezcontourf('(x-2)^2+(y-1)^2')
```





Ejemplo 4: Optimización SIN restricciones

$$f(x,y) = 8x^2 + 3xy + 7y^2 - 25x + 31y - 29$$

- Gradiente = 0 es un sistema de ecuaciones

$$\nabla f(x,y) = \begin{pmatrix} 16x + 3y - 25 \\ 3x + 14y + 31 \end{pmatrix} = 0$$

con solución $(x,y) = (2.060, -2.656)$

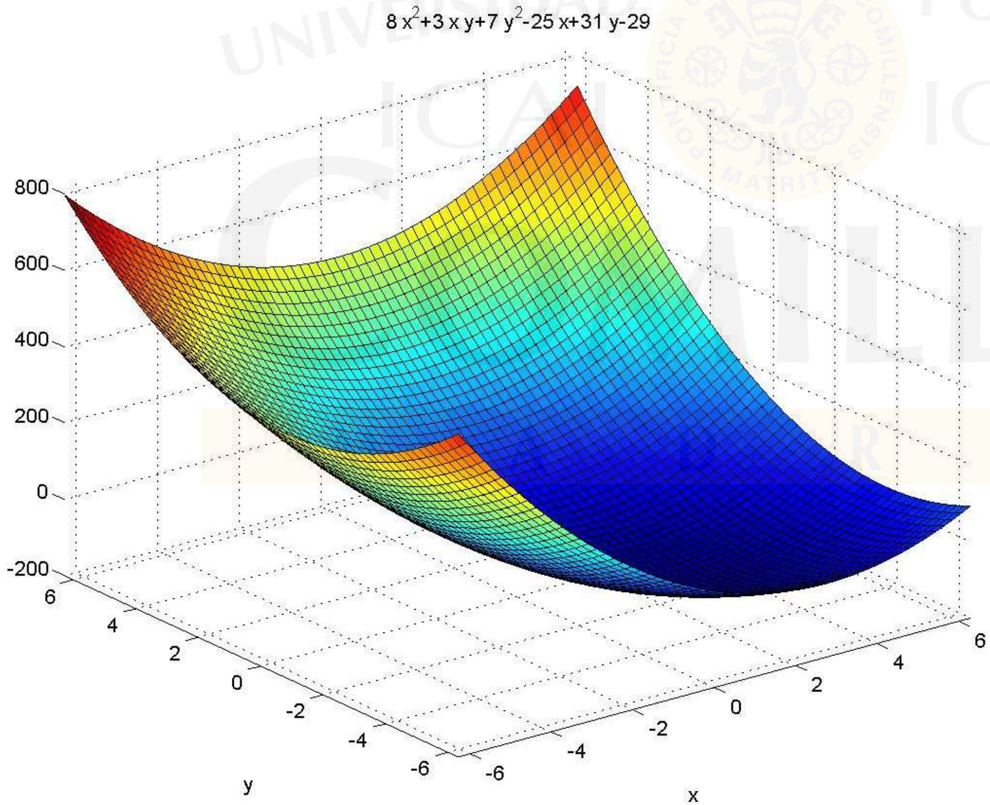
- Hessiano $\nabla^2 f(x,y) = \begin{pmatrix} 16 & 3 \\ 3 & 14 \end{pmatrix}$ es matriz semidefinida positiva

e independiente del punto, luego es un mínimo global

Ejemplo 4: Optimización SIN restricciones

$$f(x,y) = 8x^2 + 3xy + 7y^2 - 25x + 31y - 29$$

```
ezsurf('8*x^2+3*x*y+7*y^2-25*x+31*y-29')
```

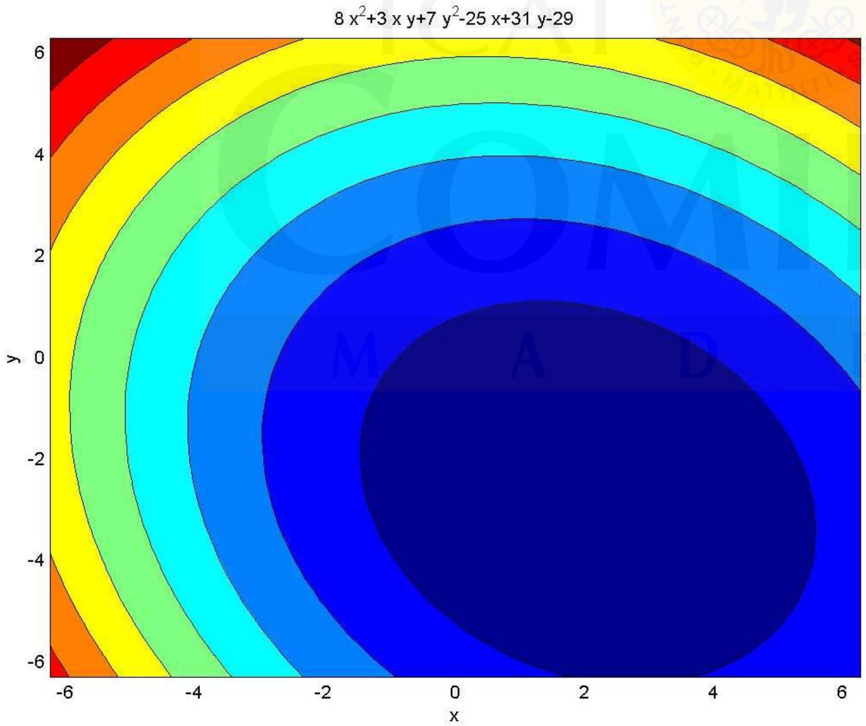


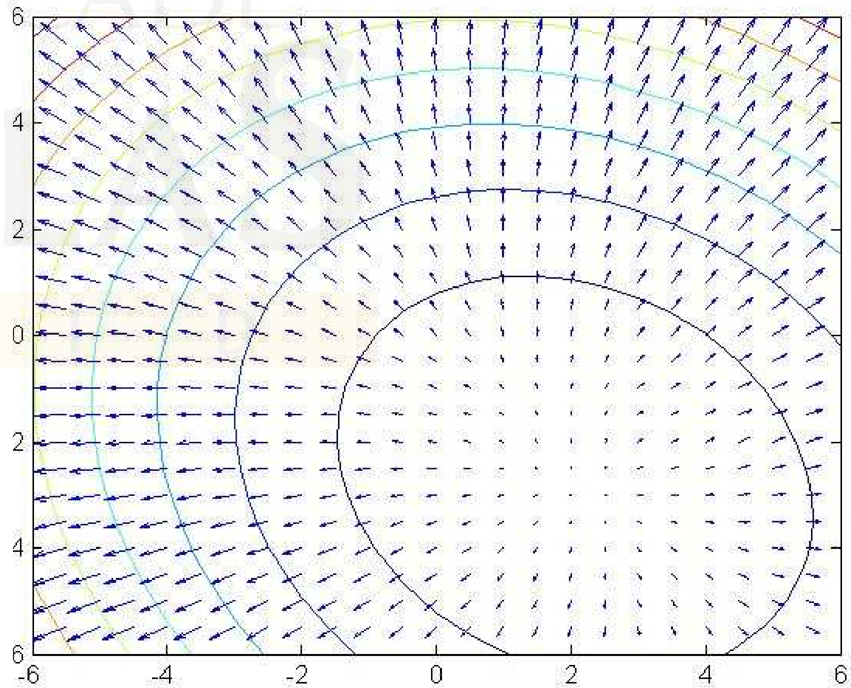
Ejemplo 4: Optimización SIN restricciones

$$f(x,y) = 8x^2 + 3xy + 7y^2 - 25x + 31y - 29$$

```
[x,y] = meshgrid(-6:.5:6, -6:.5:6);
z=8*x.^2+3*x.*y+7*y.^2-25*x+31*y-29;
[px,py] = gradient(z,0.5,0.5);
contour(x,y,z);
hold on, quiver(x,y,px,py)
```

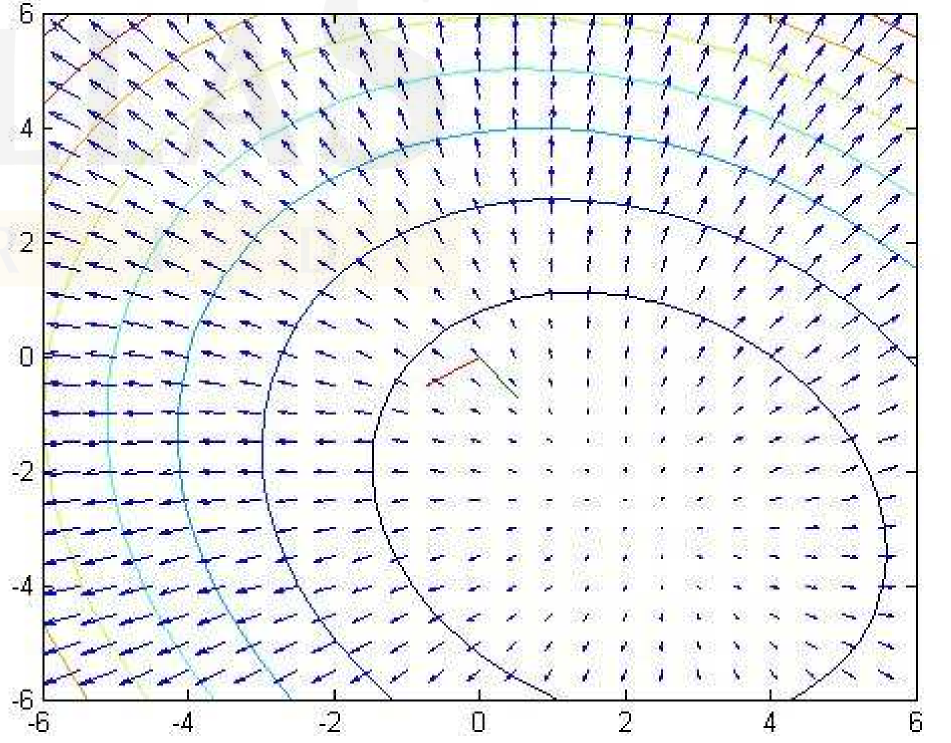
```
ezcontourf('8*x^2+3*x*y+7*y^2-25*x+31*y-29')
```





Autovalores y autovectores

- Hessiano $\nabla^2 f(x,y) = \begin{pmatrix} 16 & 3 \\ 3 & 14 \end{pmatrix}$ es matriz semidefinida positiva
- Sus autovalores son $\lambda = \begin{pmatrix} 11.84 \\ 18.16 \end{pmatrix}$
- Y los autovectores $v_1 = \begin{pmatrix} 0.58 \\ -0.81 \end{pmatrix}$ $v_2 = \begin{pmatrix} -0.81 \\ -0.58 \end{pmatrix}$



CONTENIDO

- PROBLEMAS DE PROGRAMACIÓN NO LINEAL
- TIPOS DE PROBLEMAS NLP
- CLASIFICACIÓN DE MÉTODOS DE RESOLUCIÓN
- CONDICIONES DE OPTIMALIDAD PARA OPTIMIZACIÓN SIN RESTRICCIONES
- CONDICIONES DE OPTIMALIDAD PARA NLP
- METHODS FOR UNCONSTRAINED OPTIMIZATION (master)
- NONLINEAR PROGRAMMING METHODS (master)

Lagrangiano (i)

- Sea el problema de optimización $\begin{aligned}&\min_x f(x)\\ &Ax=b\end{aligned}$

  siendo $x \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$
- Se define el lagrangiano como

$$L(x,\lambda) = f(x) + \lambda^T (Ax - b)$$

  siendo $\lambda \in \mathbb{R}^m$ los multiplicadores de Lagrange.
- El lagrangiano es un problema sin restricciones.
- El problema con restricciones se transforma en otro sin ellas con m variables adicionales.
- El mínimo de ambos problemas coincide puesto que $Ax = b$.

Lagrangiano (ii)

- Condiciones de optimalidad de primer orden

$$\nabla L(x^*,\lambda^*)=0 \quad \Rightarrow \quad \begin{cases} \nabla_x L(x^*,\lambda^*)=\nabla f(x^*)+A^T\lambda^*=0 \\ \nabla_\lambda L(x^*,\lambda^*)=Ax^*-b=0 \end{cases}$$

  y por tanto si x^* es un mínimo local debe cumplir $\nabla f(x^*)=-A^T\lambda^*$

- En un mínimo local el gradiente de la función objetivo es una combinación lineal de los gradientes de las restricciones y los multiplicadores de Lagrange son los pesos.
- Los multiplicadores representan el cambio en la función objetivo para un cambio unitario (marginal) en la cota de cada restricción. En el caso particular de LP éstos recibían el nombre de *variables duales* o *precios sombra*. Con esta formulación del lagrangiano los multiplicadores resultan con signo contrario a las variables duales.

Lagrangiano (iii)

- Sea el problema de optimización
$$\min_x f(x)$$
$$g_i(x) \le 0 \quad i = 1,\dots,m$$
$$h_j(x) = 0 \quad j = 1,\dots,l$$

  donde $f:\mathbb{R}^n \to \mathbb{R}$, $g_i:\mathbb{R}^n \to \mathbb{R}$

- Se define el lagrangiano como

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

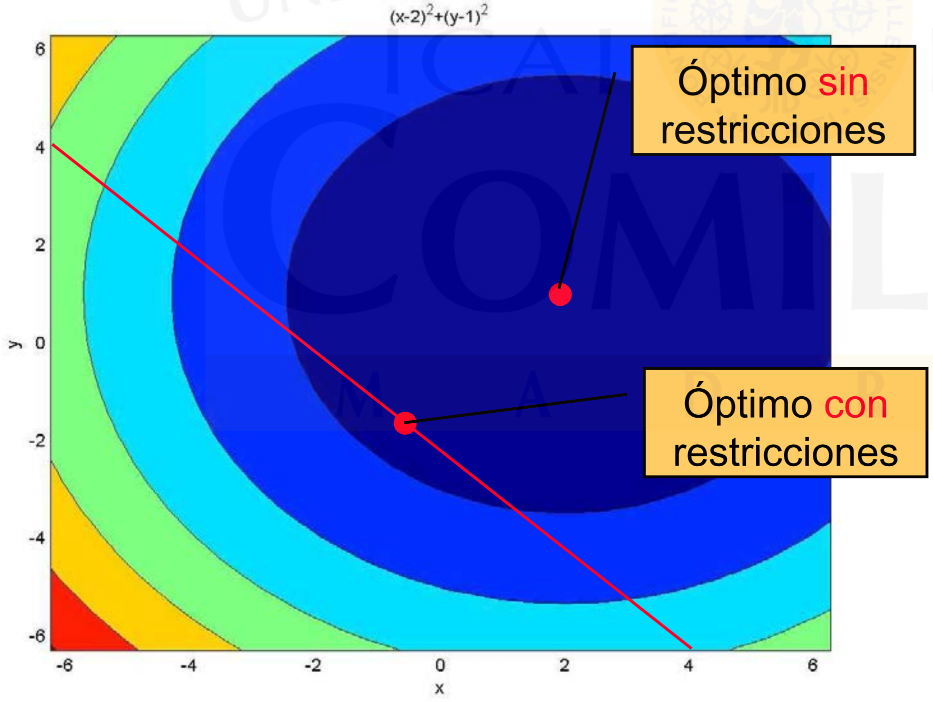
  donde $\lambda \in \mathbb{R}^m$ y $\mu \in \mathbb{R}^l$ son los multiplicadores de Lagrange.

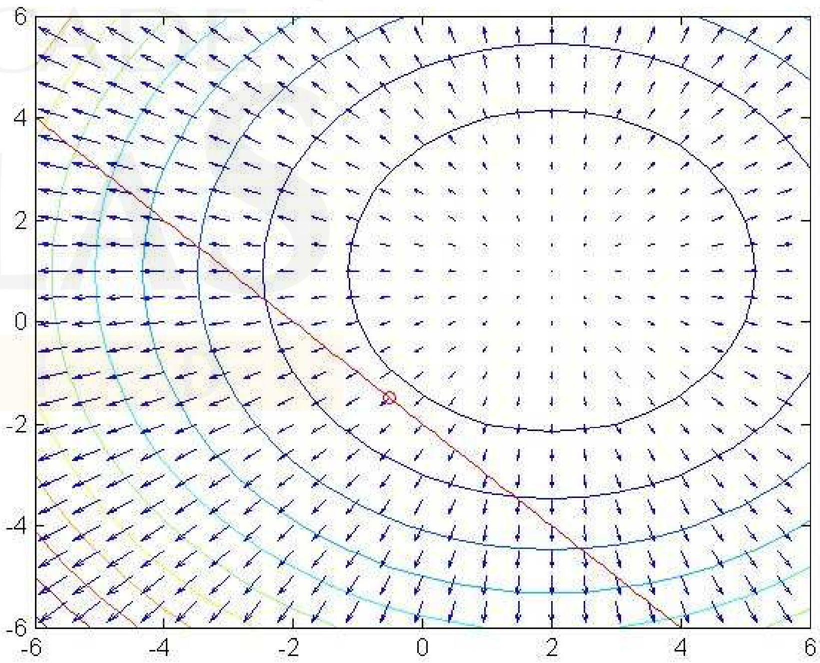
- Lagrangiano es siempre una cota inferior de $f(x)$ para valores factibles de x y valores conocidos de $\lambda \ge 0$ (no negativos) y μ (libre).

Ejemplo 1 (i)

$$\min (x-2)^2 + (y-1)^2$$
$$x + y = -2$$

- Punto óptimo (-0.5,-1.5)
- Gradiente $\nabla f(x^*, y^*) = \begin{pmatrix} 2(x-2) \\ 2(y-1) \end{pmatrix}_{(-0.5,-1.5)} = \begin{pmatrix} -5 \\ -5 \end{pmatrix} = -(1 \quad 1)^T \lambda^*$ $\quad \lambda^* = 5$

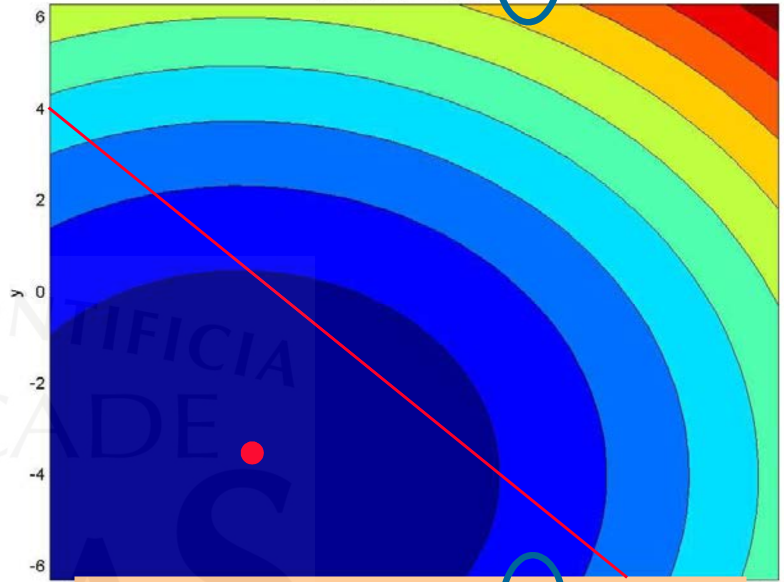




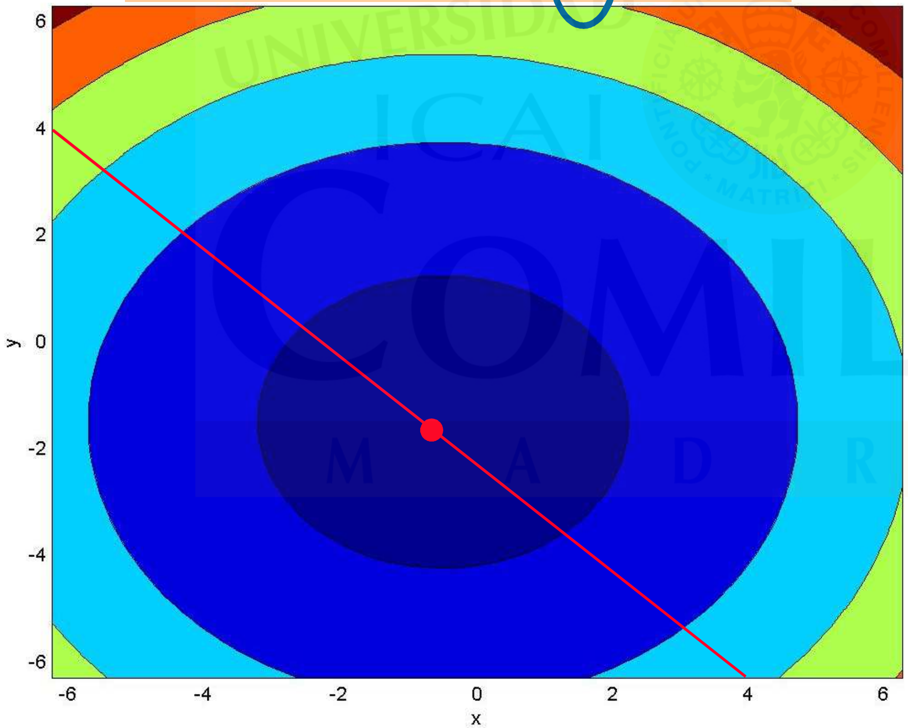
Ejemplo 1 (ii)

$$\min (x-2)^2 + (y-1)^2 + \lambda (x+y+2)$$

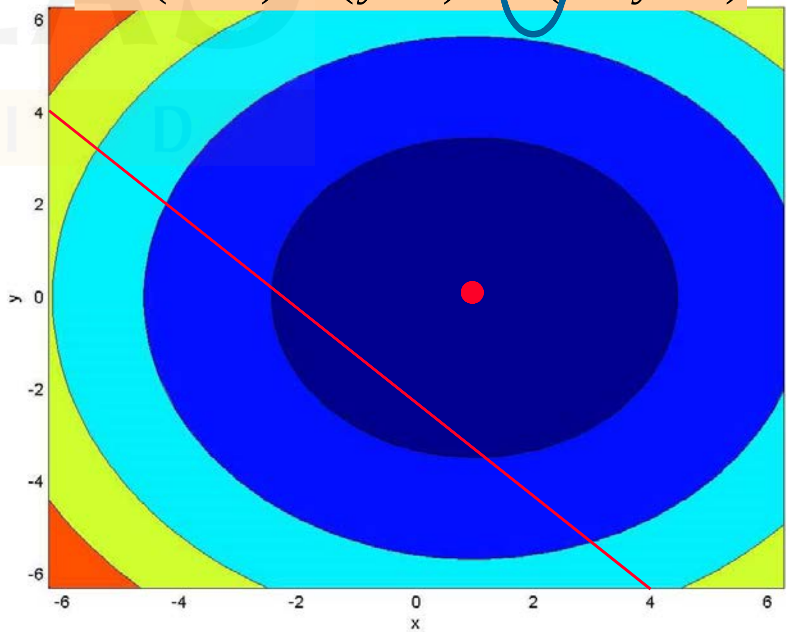
$$\min (x-2)^2 + (y-1)^2 + 10(x+y+2)$$



$$\min (x-2)^2 + (y-1)^2 + 5(x+y+2)$$

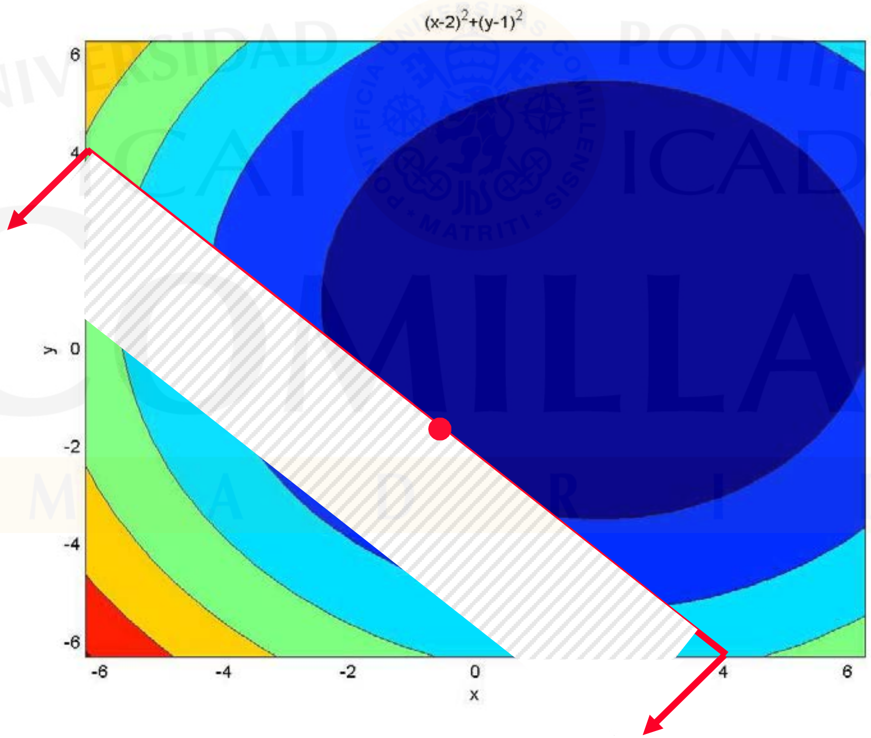


$$\min (x-2)^2 + (y-1)^2 + 2(x+y+2)$$



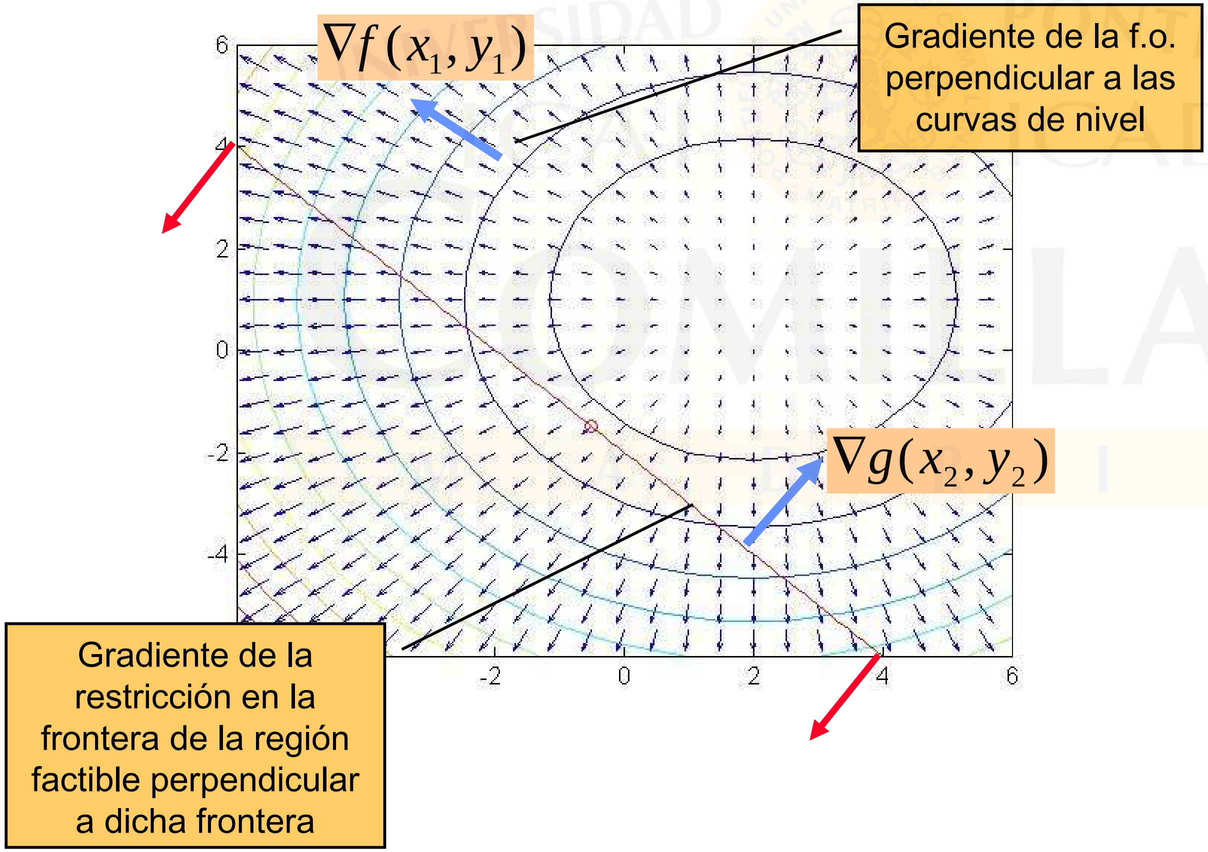
Ejemplo 2 (i)

$$\min (x-2)^2 + (y-1)^2$$
$$x + y \leq -2$$



Ejemplo 2 (ii)

$$\min (x-2)^2 + (y-1)^2$$
$$x + y \le -2$$



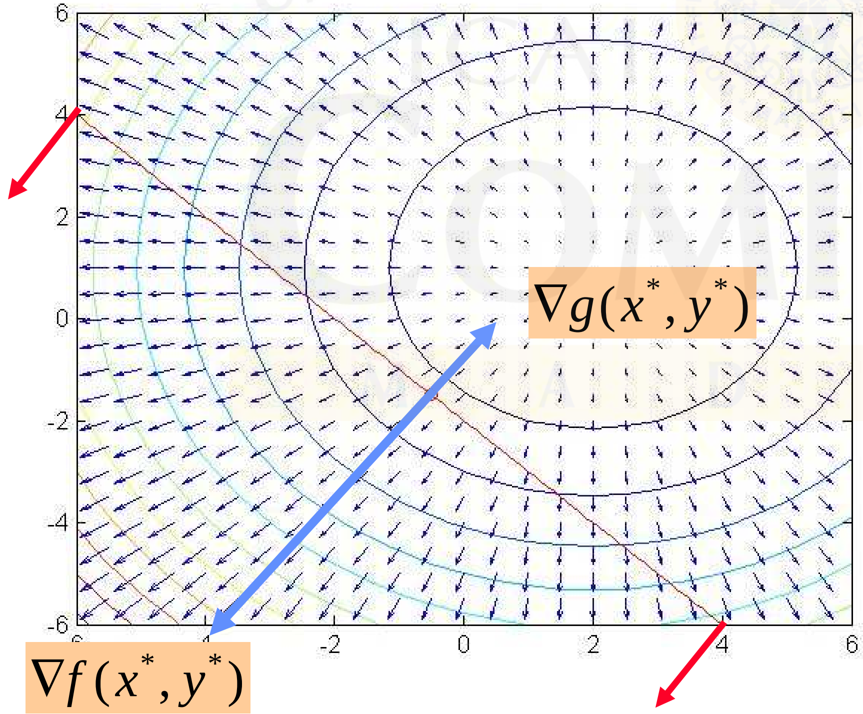
$$\nabla f(x, y) = \begin{pmatrix} 2(x-2) \\ 2(y-1) \end{pmatrix}$$

$$\nabla g(x, y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Ejemplo 2 (iii)

$$\min (x-2)^2 + (y-1)^2$$
$$x + y \le -2$$

- En el óptimo (-0.5,-1.5) ambos gradientes tienen sentidos opuestos



$$\nabla f(x^*, y^*) = \begin{pmatrix} 2(x-2) \\ 2(y-1) \end{pmatrix}_{(-0.5,-1.5)} = \begin{pmatrix} -5 \\ -5 \end{pmatrix}$$

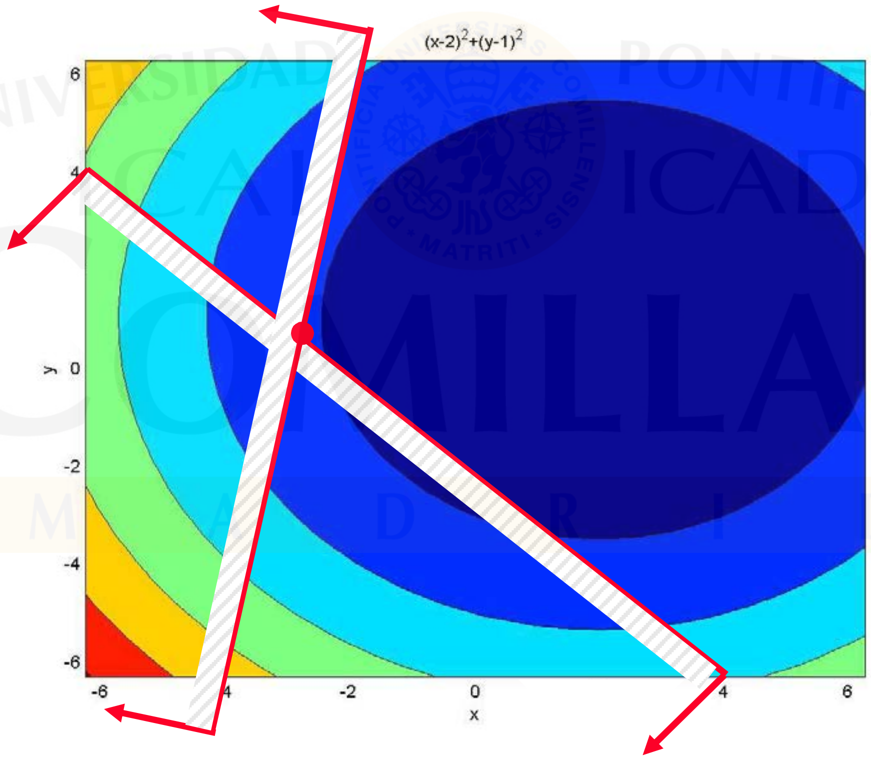
$$-\nabla g(x^*, y^*) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\nabla f(x^*, y^*) = -\lambda \nabla g(x^*, y^*)$$
$$\lambda \ge 0$$

$$\begin{pmatrix} -5 \\ -5 \end{pmatrix} = \lambda \begin{pmatrix} -1 \\ -1 \end{pmatrix} \qquad \lambda = 5$$
$$\lambda \ge 0$$

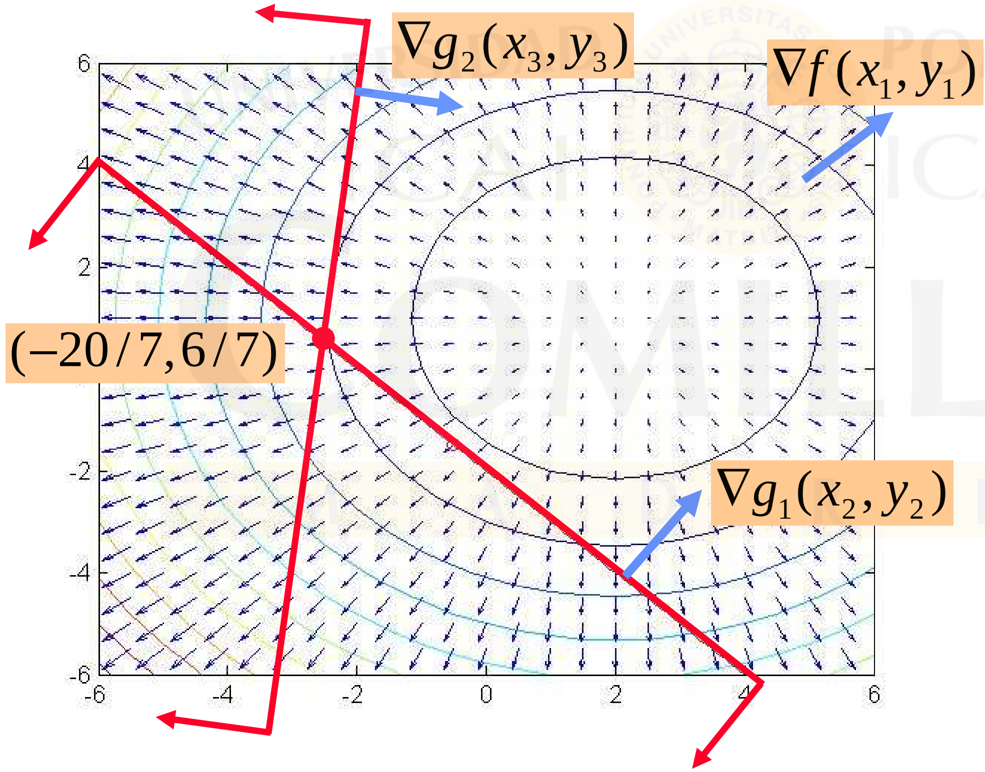
Ejemplo 3 (i)

$$\min (x-2)^2+(y-1)^2$$
$$x+y\leq -2$$
$$6x-y\leq -18$$



Ejemplo 3 (ii)

$$\min (x-2)^2+(y-1)^2$$
$$x+y\le -2$$
$$6x-y\le -18$$



$$\nabla f(x,y)=\begin{pmatrix}2(x-2)\\2(y-1)\end{pmatrix}$$

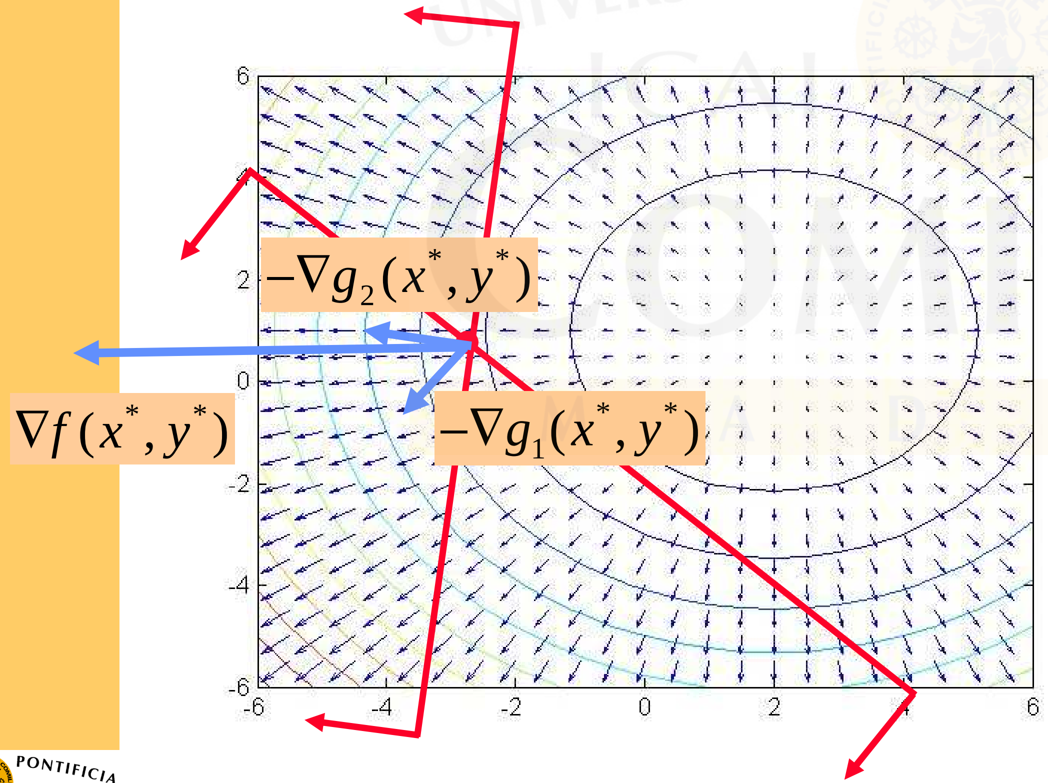
$$\nabla g_1(x,y)=\begin{pmatrix}1\\1\end{pmatrix}$$

$$\nabla g_2(x,y)=\begin{pmatrix}6\\-1\end{pmatrix}$$

Ejemplo 3 (iii)

$$\min (x-2)^2+(y-1)^2$$
$$x+y\le -2$$
$$6x-y\le -18$$

- En el óptimo (-20/7,6/7) el gradiente de la f.o. se puede expresar como combinación lineal de los gradientes de las restricciones cambiados de signo



$$\nabla f(x^*,y^*)=\begin{pmatrix}2(x-2)\\2(y-1)\end{pmatrix}_{(-20/7,6/7)}=\begin{pmatrix}-68/7\\-2/7\end{pmatrix}$$

$$-\nabla g_1(x^*,y^*)=\begin{pmatrix}-1\\-1\end{pmatrix}\qquad -\nabla g_2(x^*,y^*)=\begin{pmatrix}-6\\1\end{pmatrix}$$

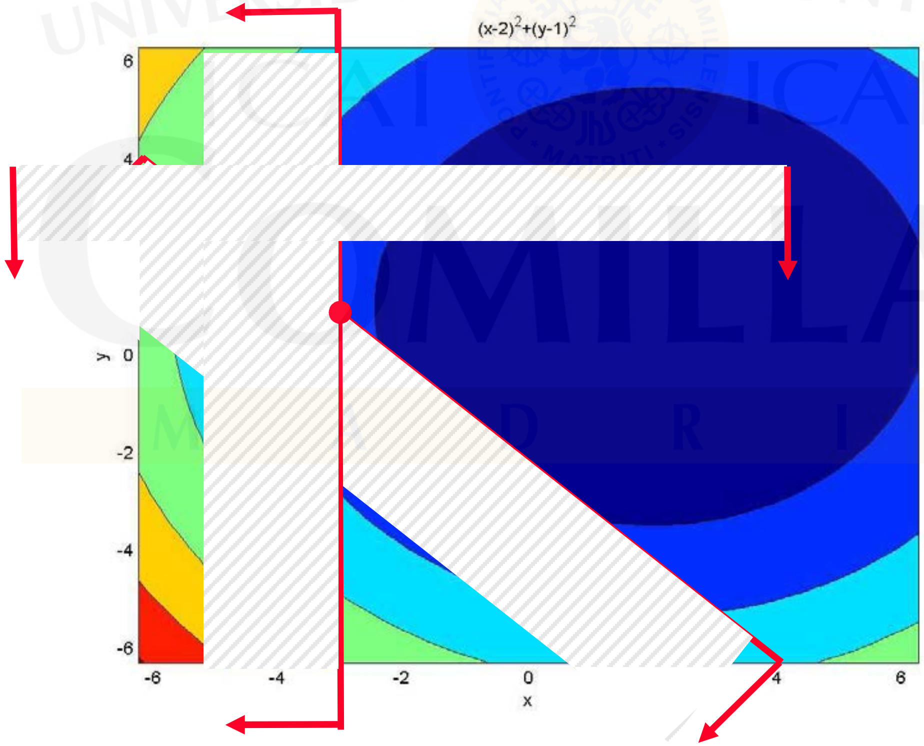
$$\nabla f(x^*,y^*)=-\lambda_1\nabla g_1(x^*,y^*)-\lambda_2\nabla g_2(x^*,y^*)$$
$$\lambda_1,\lambda_2\ge 0$$

$$\begin{pmatrix}-68/7\\-2/7\end{pmatrix}=\lambda_1\begin{pmatrix}-1\\-1\end{pmatrix}+\lambda_2\begin{pmatrix}-6\\1\end{pmatrix}$$
$$\lambda_1,\lambda_2\ge 0$$

$$\lambda_1=2.75$$
$$\lambda_2=1.125$$

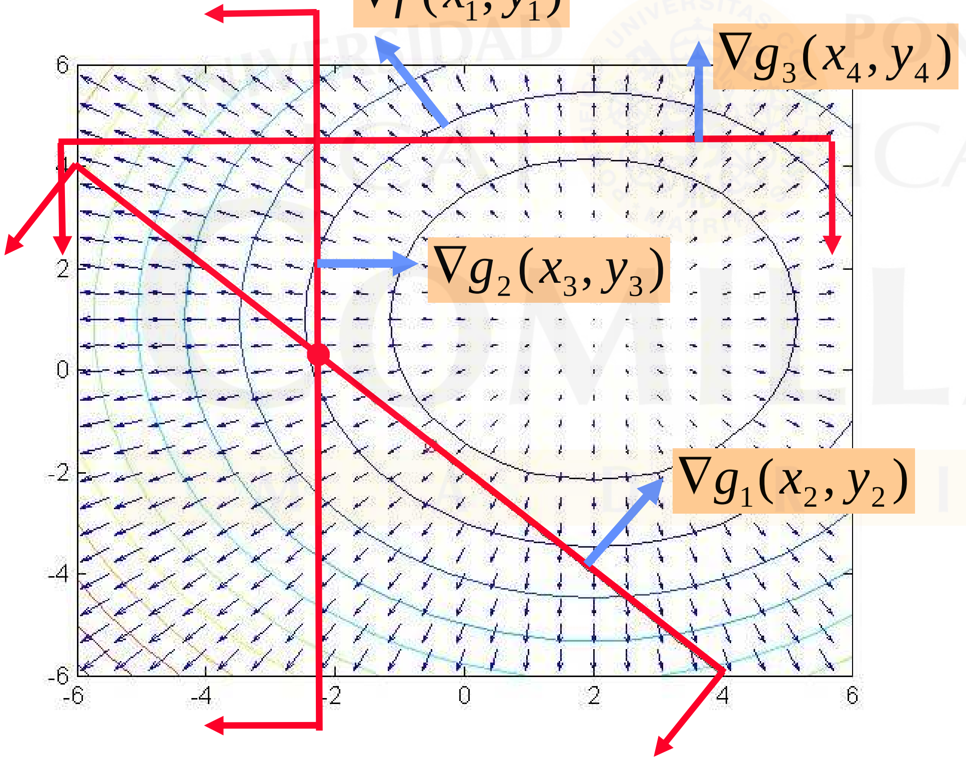
Ejemplo 4 (i)

$$\min (x-2)^2 + (y-1)^2$$
$$x + y \leq -2$$
$$x \leq -3$$
$$y \leq 4$$



Ejemplo 4 (ii)

$$\min (x-2)^2 + (y-1)^2$$
$$x + y \le -2$$
$$x \le -3$$
$$y \le 4$$



$$\nabla f(x,y) = \begin{pmatrix} 2(x-2) \\ 2(y-1) \end{pmatrix}$$

$$\nabla g_1(x,y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

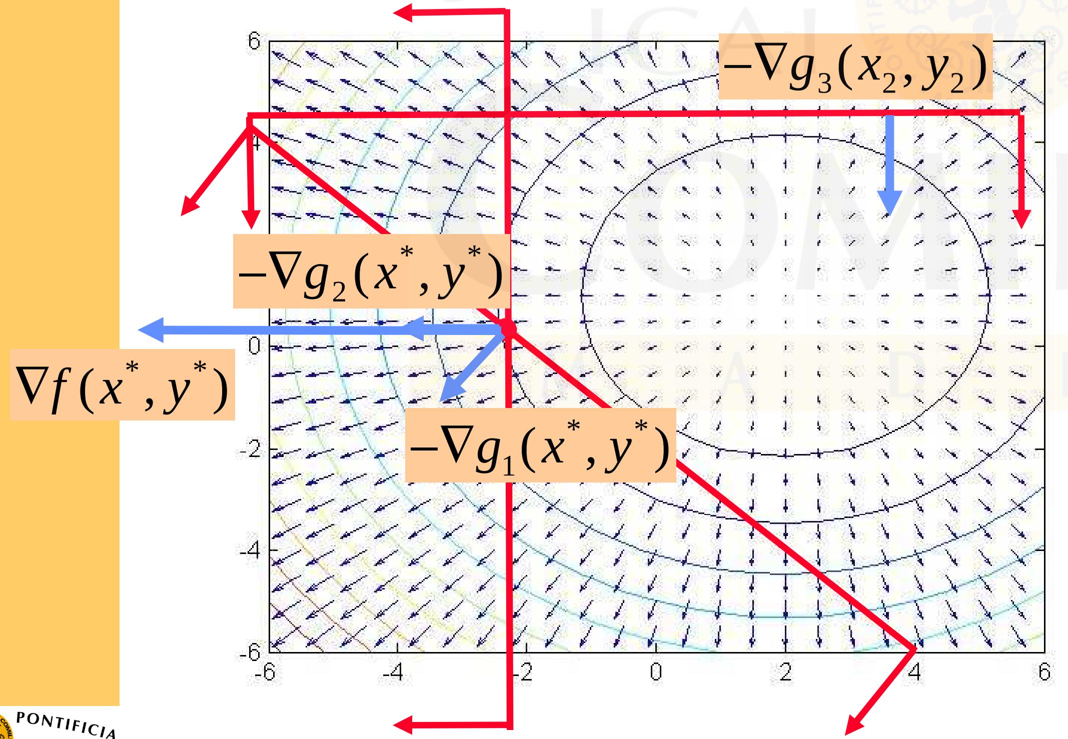
$$\nabla g_2(x,y) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\nabla g_3(x,y) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Ejemplo 4 (iii)

$$\min (x-2)^2+(y-1)^2$$
$$x+y \le -2$$
$$x \le -3$$
$$y \le 4$$

- En el óptimo (-3,1) el gradiente de la f.o. se puede expresar como combinación lineal de los gradientes de las ***restricciones activas*** cambiados de signo



$$\nabla f(x^*,y^*)=\begin{pmatrix}2(x-2)\\2(y-1)\end{pmatrix}_{(-3,1)}=\begin{pmatrix}-10\\0\end{pmatrix}$$

$$-\nabla g_1(x^*,y^*)=\begin{pmatrix}-1\\-1\end{pmatrix} \qquad -\nabla g_2(x^*,y^*)=\begin{pmatrix}-1\\0\end{pmatrix}$$

$$-\nabla g_3(x^*,y^*)=\begin{pmatrix}0\\-1\end{pmatrix}$$

$$\nabla f(x^*,y^*)=-\lambda_1\nabla g_1(x^*,y^*)-\lambda_2\nabla g_2(x^*,y^*)$$
$$\lambda_1,\lambda_2 \ge 0$$

$$\begin{pmatrix}-10\\0\end{pmatrix}=\lambda_1\begin{pmatrix}-1\\-1\end{pmatrix}+\lambda_2\begin{pmatrix}-1\\0\end{pmatrix}$$
$$\lambda_1,\lambda_2 \ge 0$$

$$\lambda_1=0$$
$$\lambda_2=10$$
$$\lambda_3=0$$

- La primera restricción es superflua. Solución degenerada.

Condiciones necesarias con restricciones de desigualdad (i)

- Sea el problema

$$\min_x f(x)$$
$$g_i(x) \le 0 \quad i = 1,\ldots,m$$

donde $f:\mathbb{R}^n \to \mathbb{R}$, $g_i:\mathbb{R}^n \to \mathbb{R}$

- Sea x^* un punto factible

$I = \{i / g_i(x^*) = 0\}$ el conjunto de restricciones activas

f y $\{g_i, i \in I\}$ diferenciables en x^*

$\{g_i, i \notin I\}$ continuas en x^*

$\{\nabla g_i(x^*)\}_{i\in I}$ linealmente independientes

Condiciones necesarias con restricciones de desigualdad (ii)

$$\min_x f(x)$$
$$g_i(x) \le 0 \quad i = 1,\dots,m$$

- Si x^* es mínimo local entonces existen unos escalares $\{\lambda_i, i \in I\}$ tales que

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

- Además si las funciones $\{g_i, i \notin I\}$ son diferenciables en x^*, si x^* es óptimo local entonces

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$$
$$\boxed{\lambda_i^* g_i(x^*) = 0 \quad i = 1,\dots,m}$$
$$\lambda_i^* \ge 0 \quad i = 1,\dots,m$$

Condición de complementariedad de holguras

- Restricción no activa → multiplicador 0.

  Restricción activa → multiplicador puede ser o no 0.

Condiciones necesarias con restricciones de desigualdad (iii)

- Sea el problema

$$\min_x f(x)$$
$$g_i(x) \le 0 \quad i = 1, \dots, m$$

- Condiciones **necesarias** de Karush-Kuhn-Tucker (KKT) de primer orden para tener un óptimo local

Gradiente de f.o.: combinación lineal de gradientes de restricciones cambiados de signo

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

$$\lambda_i^* g_i(x^*) = 0 \quad i = 1, \dots, m$$

$$g_i(x^*) \le 0 \quad i = 1, \dots, m$$

$$\lambda_i^* \ge 0 \quad i = 1, \dots, m$$

Punto factible

Condición de complementariedad de holguras
Restricción no activa λ=0
Restricción activa λ≠0

Condiciones necesarias con restricciones de desigualdad (v)

- Sea el problema
$$\min_x f(x)$$
$$g_i(x) \le 0 \quad i = 1,\dots,m$$
- El lagrangiano será
$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$
$$\lambda_i \ge 0$$
- La condición de optimalidad para el lagrangiano será
$$\nabla L(x^*,\lambda^*) = 0 \quad \Rightarrow \quad \begin{cases} \nabla_x L(x^*,\lambda^*) = \nabla f(x^*) + \sum_{i \in I} \lambda_i^* \nabla g_i(x^*) = 0 \\ \nabla_\lambda L(x^*,\lambda^*) = g_i(x^*) = 0 \quad \forall i \in I \end{cases}$$

  la 2ª corresponde a la definición de restricciones activas $\forall i \in I$
- Para considerar todas las restricciones lo expreso como
$$\lambda_i^* g_i(x^*) = 0 \quad i = 1,\dots,m$$

Condiciones suficientes con restricciones de desigualdad (i)

- Si la f.o. es no convexa o la región factible es no convexa puede haber puntos que verifiquen las condiciones necesarias.

- Sea x^* un punto factible

  $I = \{ i / g_i(x^*) = 0 \}$ el conjunto de restricciones activas

  f y $\{g_i, i \in I\}$ convexas y diferenciables en **toda la región factible**

- Si existen unos escalares $\{\lambda_i, i \in I\}$ tales que

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

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

  entonces x^* es mínimo global

Condiciones suficientes con restricciones de desigualdad (ii)

- Condición de mínimo local estricto

  Alternativamente, en lugar de poner la condición de que f y $\{g_i, i \in I\}$ sean convexas y diferenciables en x^* se puede expresar también como que el lagrangiano $L(x) = f(x) + \sum_{i \in I} \lambda_i^* g_i(x)$, siendo λ_i^* los multiplicadores de Lagrange de las restricciones, tenga un hessiano $\nabla^2 L(x^*) = \nabla^2 f(x^*) + \sum_{i \in I} \lambda_i^* \nabla^2 g_i(x^*)$ que sea una matriz definida positiva en x^*.

- Las condiciones suficientes para el caso de maximización se traducen en que la función f sea cóncava en el punto, las restricciones no cambian y los multiplicadores sean menores o iguales que 0.

Condiciones necesarias con restricciones de igualdad y desigualdad (i)

- Sea el problema

$$\min_x f(x)$$
$$g_i(x) \le 0 \quad i = 1,\dots,m$$
$$h_j(x) = 0 \quad j = 1,\dots,l$$

donde $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$

- Sea x^* un punto factible

$I = \{ i / g_i(x^*) = 0 \}$ el conjunto de restricciones activas

f y $\{g_i, i \in I\}$ diferenciables en x^*

$\{g_i, i \notin I\}$ continuas en x^*

$\{h_j, j = 1,\dots,l\}$ continuamente diferenciables en x^*

$\{\nabla g_i(x^*), i \in I; \nabla h_j(x^*), j = 1,\dots,l\}$ linealmente independientes

Condiciones necesarias con restricciones de igualdad y desigualdad (ii)

- Si x^* es mínimo local entonces existen unos escalares $\{\lambda_i, i \in I; \mu_j, j = 1,\ldots,l\}$ tales que

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

- Además si las funciones $\{g_i, i \notin I\}$ son diferenciables en x^*, si x^* es óptimo local entonces

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

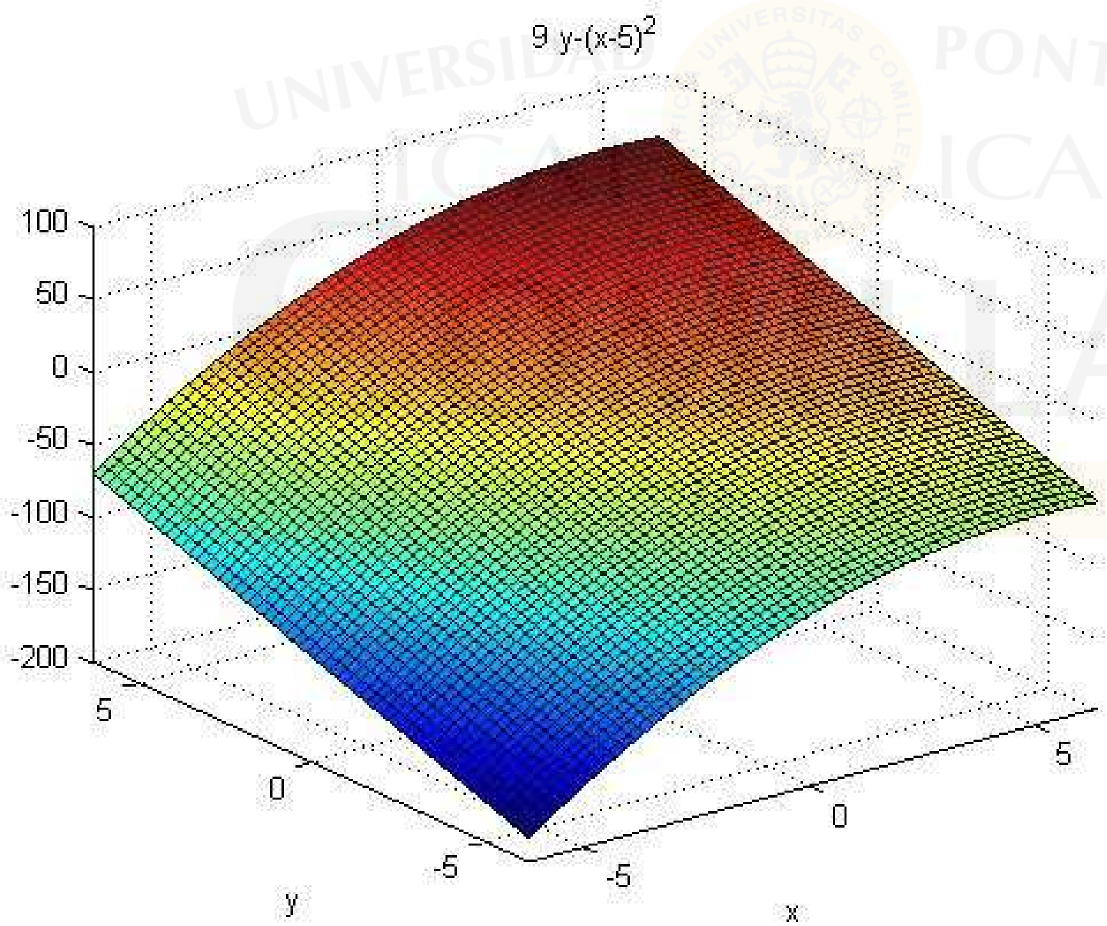
Ejemplo 5 (i)

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

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

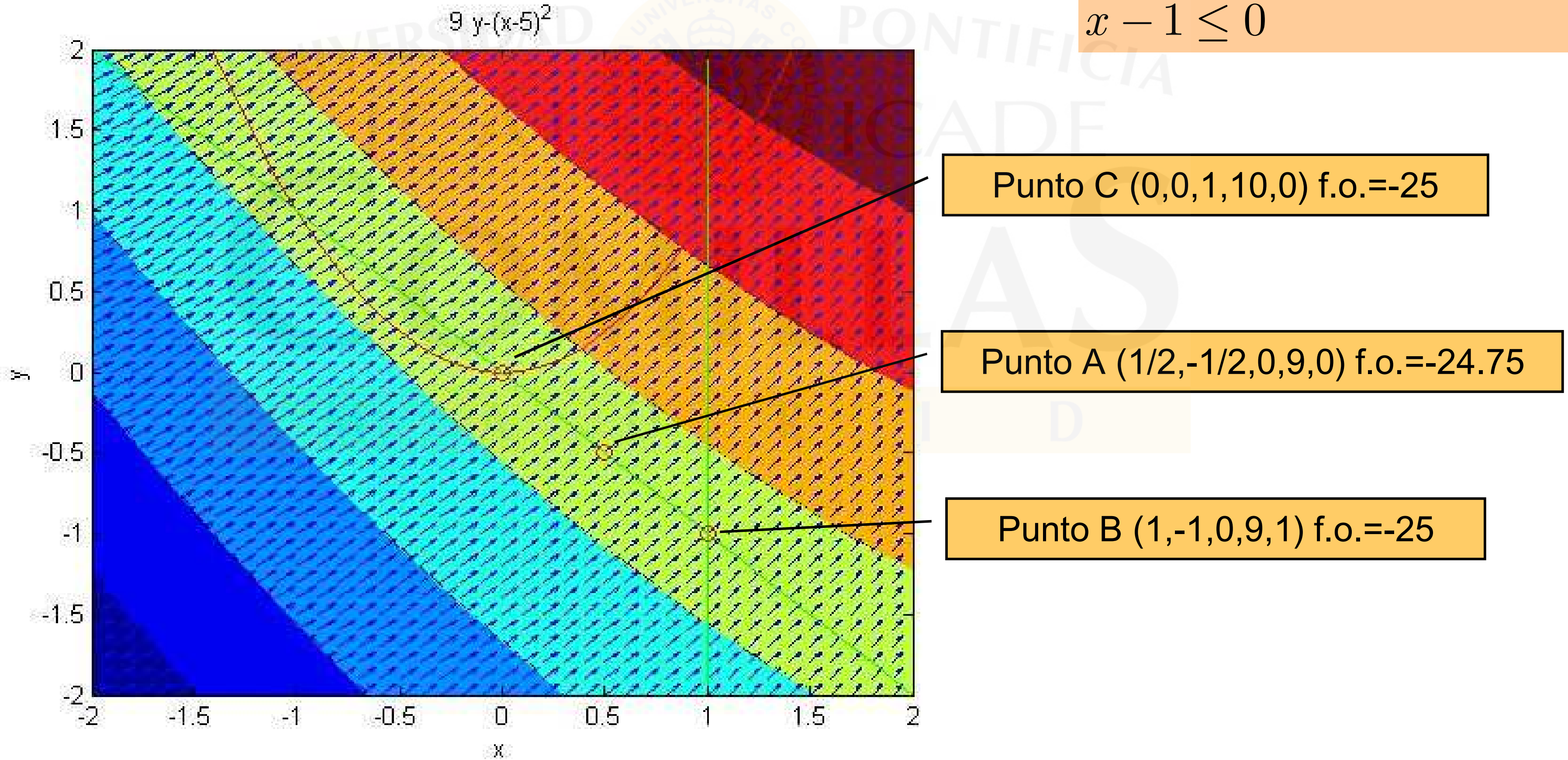
$$-x - y \leq 0$$

$$x - 1 \leq 0$$



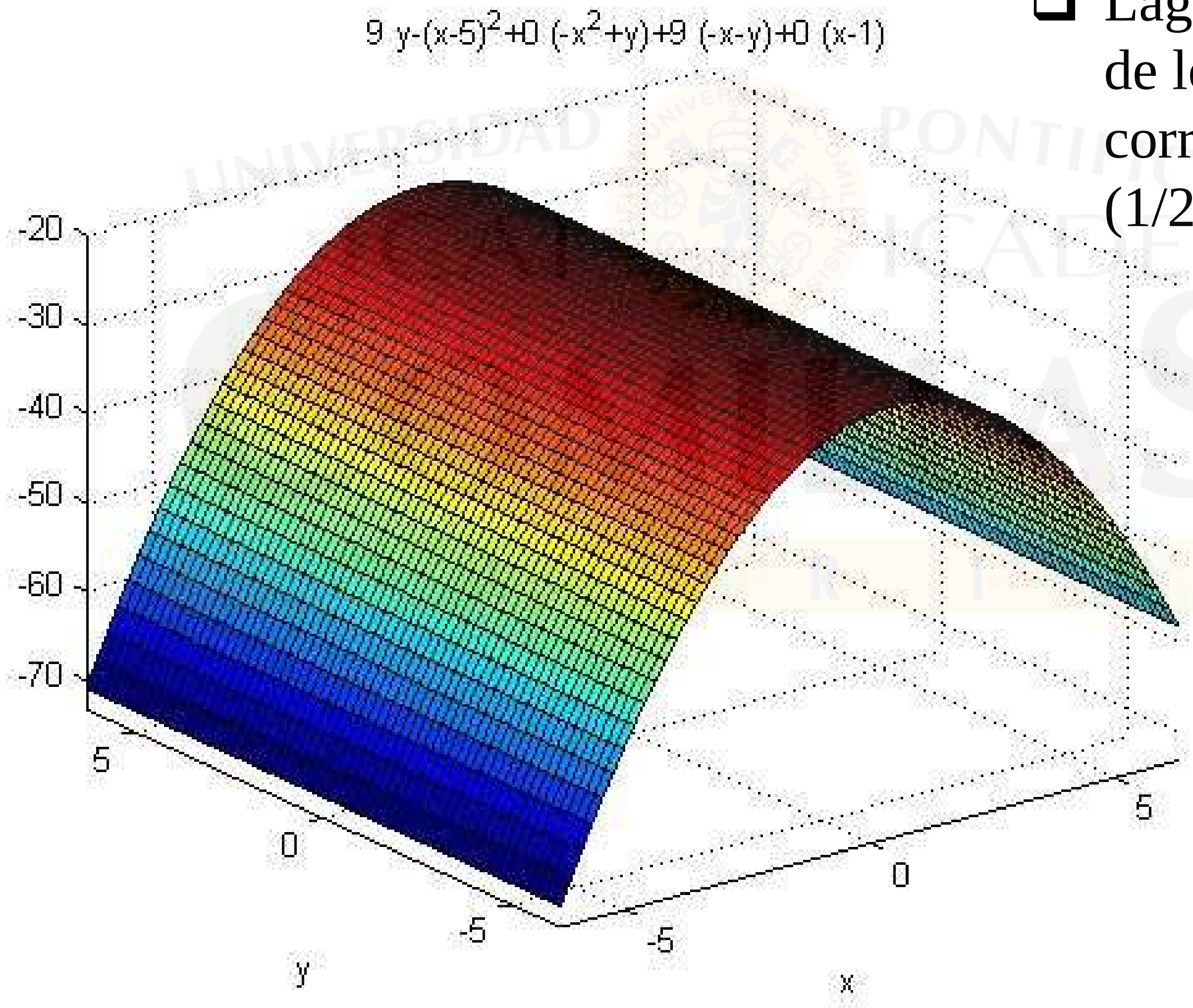
Ejemplo 5 (ii)

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



Ejemplo 5 (iii)

$$\min_{x,y,\lambda_1,\lambda_2,\lambda_3} L(x,y,\lambda_1,\lambda_2,\lambda_3) = 9y - (x-5)^2 + \lambda_1(-x^2+y) + \lambda_2(-x-y) + \lambda_3(x-1)$$



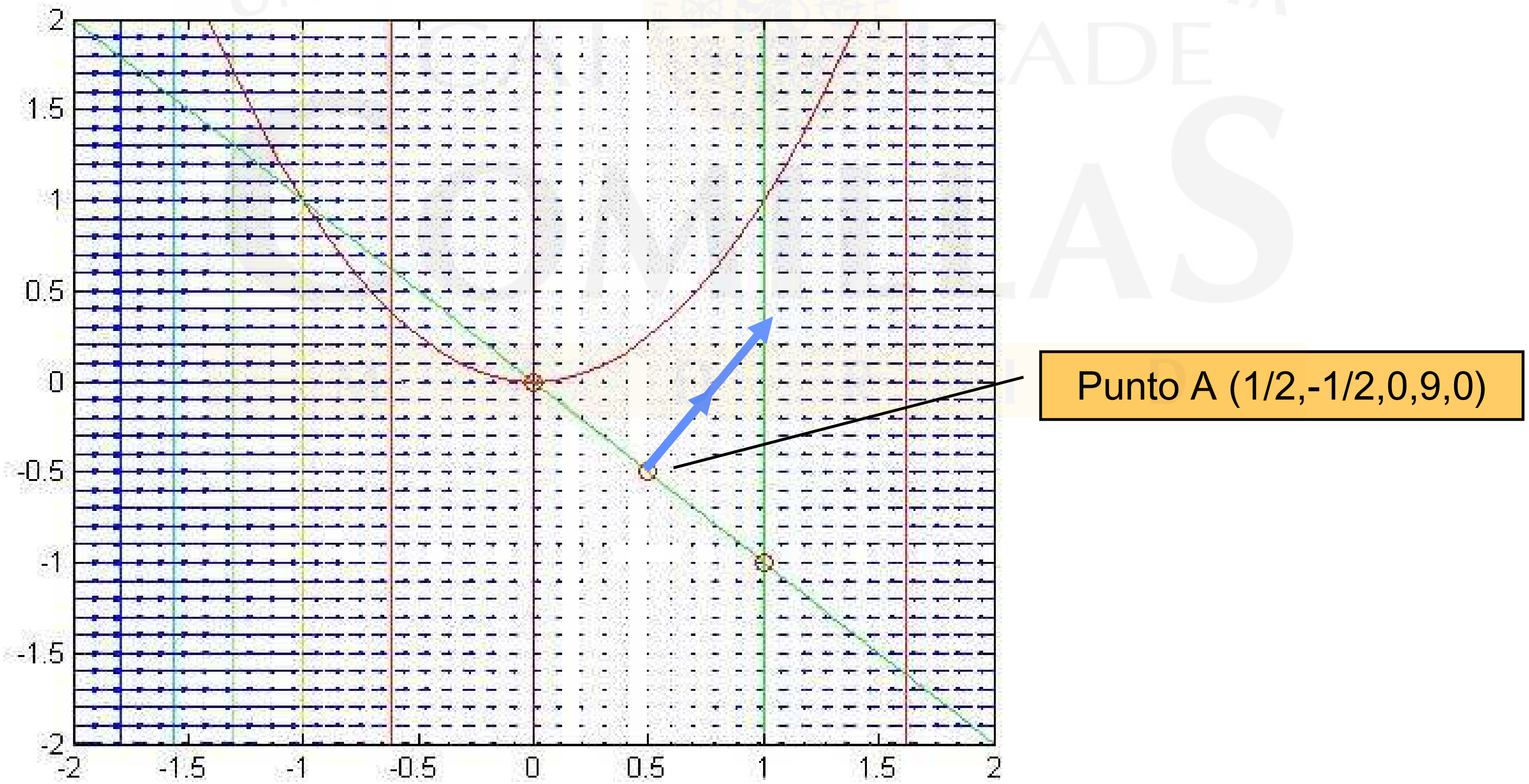
- Lagrangiano para los valores de los multiplicadores correspondientes al punto A (1/2,-1/2,0,9,0)

Función no acotada

Ejemplo 5 (iv)

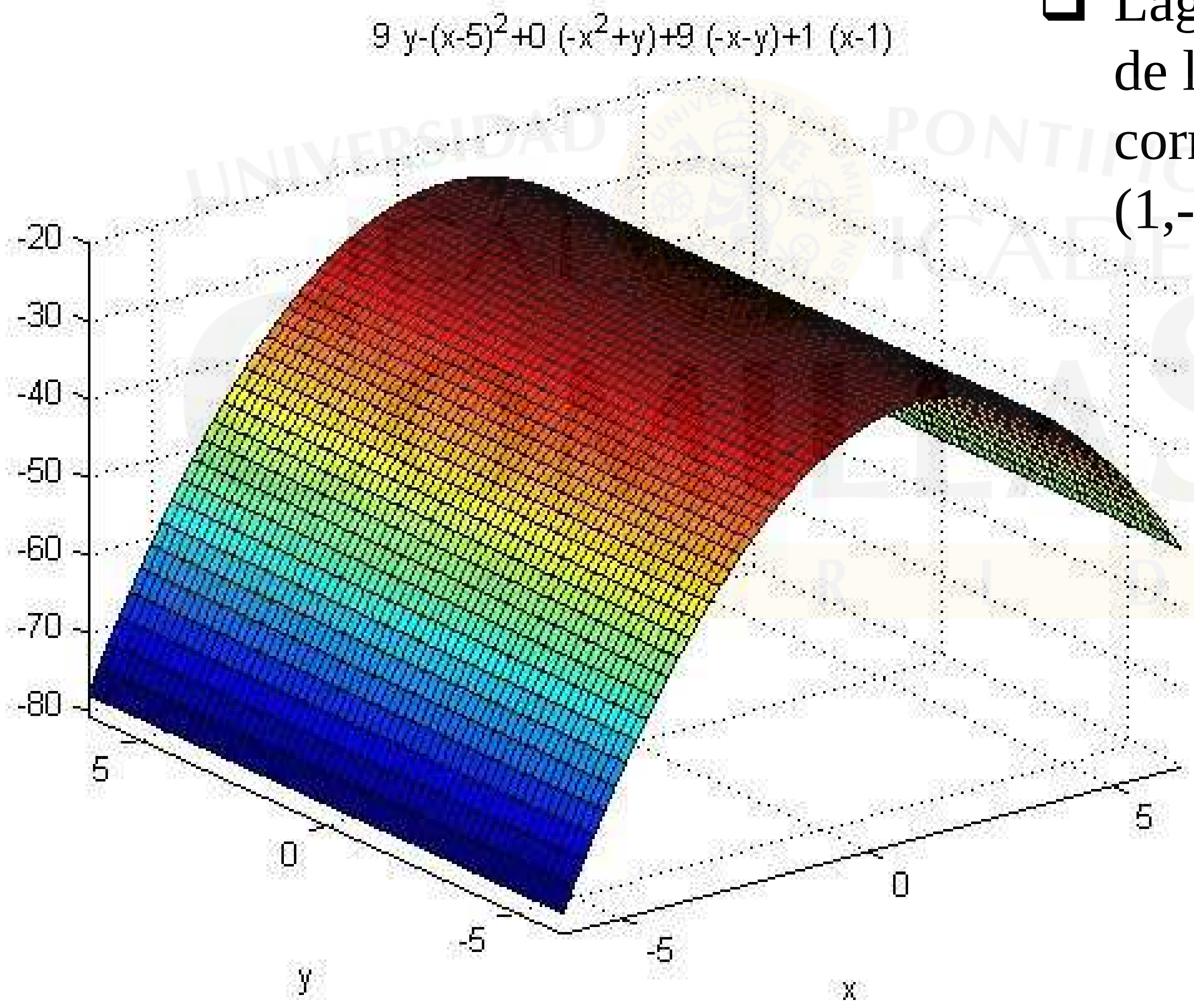
$$\min_{x,y,\lambda_1,\lambda_2,\lambda_3} L(x,y,\lambda_1,\lambda_2,\lambda_3) = 9y - (x-5)^2 + \lambda_1(-x^2+y) + \lambda_2(-x-y) + \lambda_3(x-1)$$

- Lagrangiano para los valores de los multiplicadores correspondientes al punto A (1/2,-1/2,0,9,0)



Ejemplo 5 (v)

$$\min_{x,y,\lambda_1,\lambda_2,\lambda_3} L(x,y,\lambda_1,\lambda_2,\lambda_3) = 9y-(x-5)^2+\lambda_1(-x^2+y)+\lambda_2(-x-y)+\lambda_3(x-1)$$



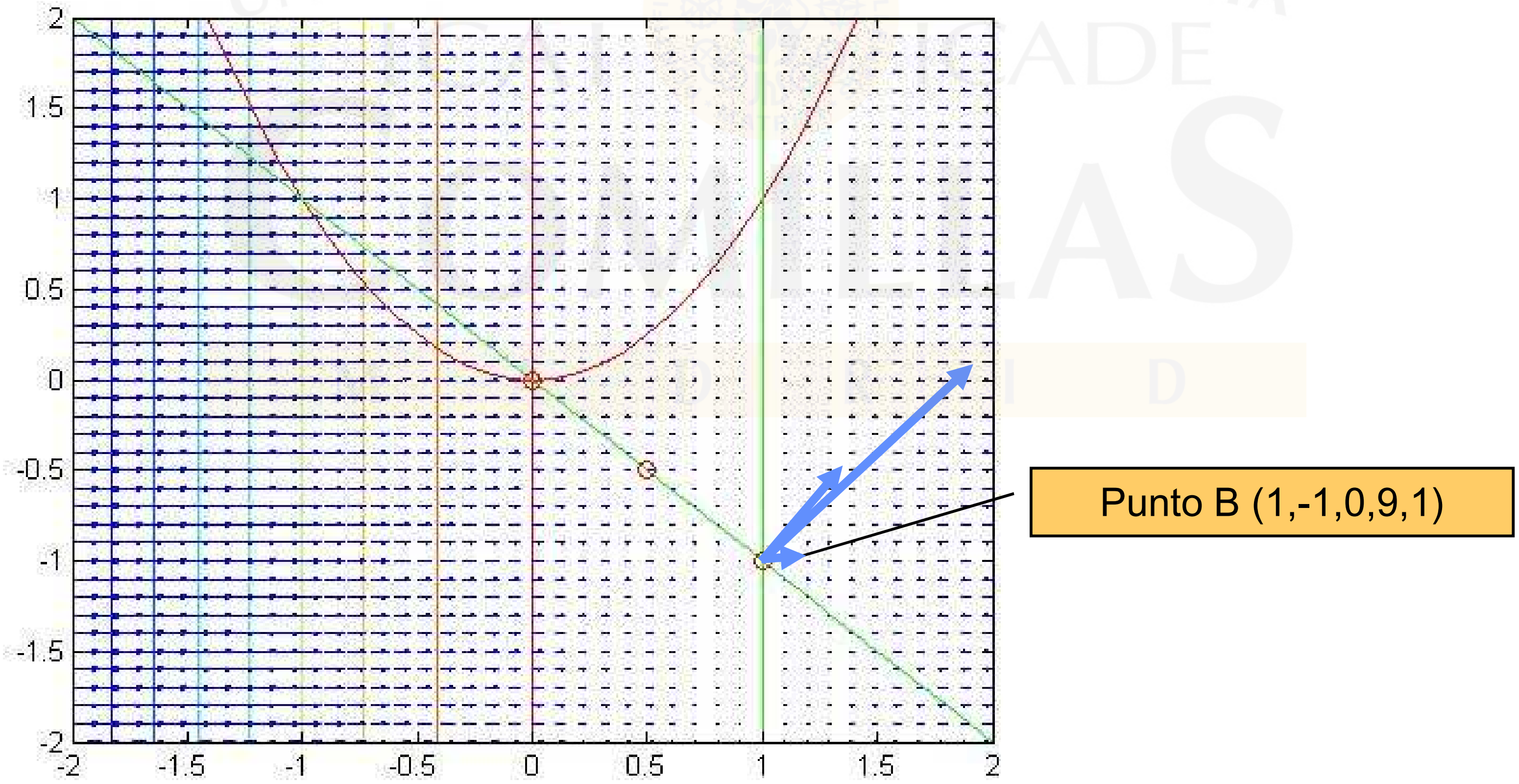
- Lagrangiano para los valores de los multiplicadores correspondientes al punto B (1,-1,0,9,1)

Función no acotada

Ejemplo 5 (vi)

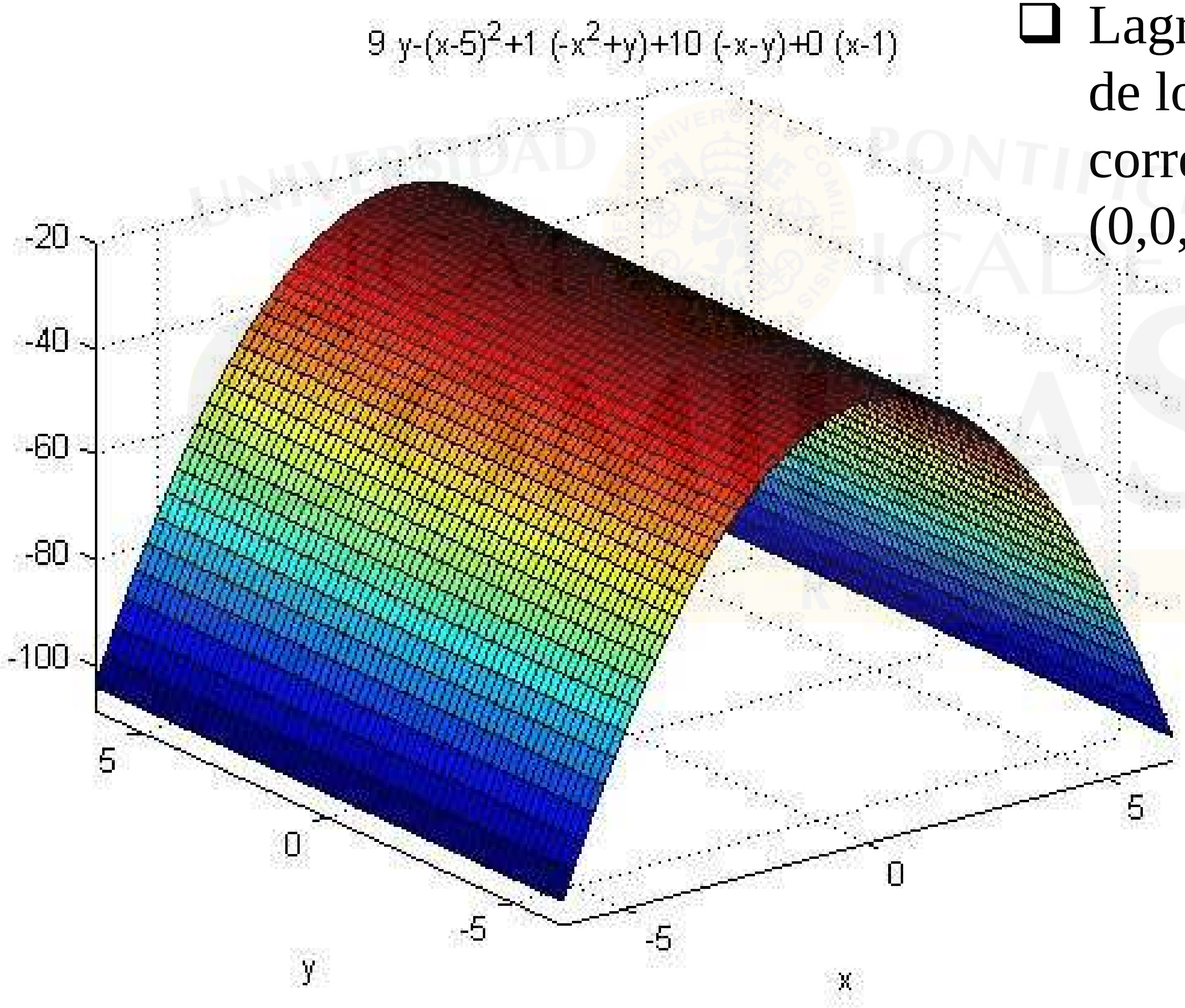
$$\min_{x,y,\lambda_1,\lambda_2,\lambda_3} L(x,y,\lambda_1,\lambda_2,\lambda_3) = 9y - (x-5)^2 + \lambda_1(-x^2+y) + \lambda_2(-x-y) + \lambda_3(x-1)$$

- Lagrangiano para los valores de los multiplicadores correspondientes al punto B (1,-1,0,9,1)



Ejemplo 5 (vii)

$$\min_{x,y,\lambda_1,\lambda_2,\lambda_3} L(x,y,\lambda_1,\lambda_2,\lambda_3) = 9y - (x-5)^2 + \lambda_1(-x^2+y) + \lambda_2(-x-y) + \lambda_3(x-1)$$



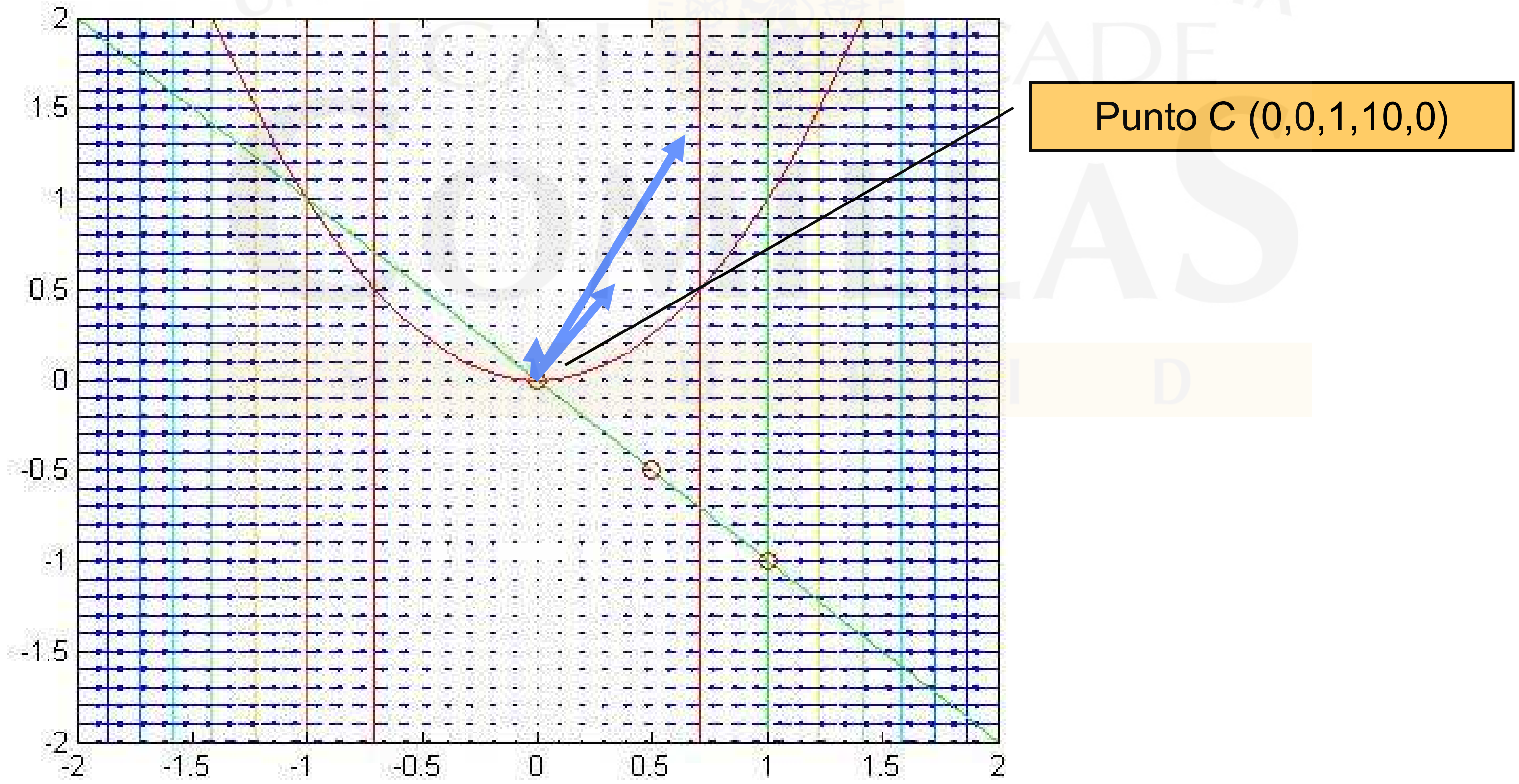
- Lagrangiano para los valores de los multiplicadores correspondientes al punto C (0,0,1,10,0)

Función no acotada

Ejemplo 5 (viii)

$$\min_{x,y,\lambda_1,\lambda_2,\lambda_3} L(x,y,\lambda_1,\lambda_2,\lambda_3) = 9y - (x-5)^2 + \lambda_1(-x^2+y) + \lambda_2(-x-y) + \lambda_3(x-1)$$

- Lagrangiano para los valores de los multiplicadores correspondientes al punto C (0,0,1,10,0)



Condiciones necesarias con restricciones de igualdad y desigualdad (iii)

- Sea el problema

$$\min_x f(x)$$
$$g_i(x) \le 0 \quad i = 1,\dots,m$$
$$h_j(x) = 0 \quad j = 1,\dots,l$$

- Condiciones necesarias de Karush-Kuhn-Tucker (KKT) de primer orden para tener un óptimo local

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

Gradiente de f.o.: combinación lineal de gradientes de restricciones cambiados de signo

$$\lambda_i^* g_i(x^*) = 0 \quad i = 1,\dots,m$$

Condición de complementariedad de holguras
Restricción no activa λ=0
Restricción activa λ≠0

$$g_i(x^*) \le 0 \quad i = 1,\dots,m$$
$$h_j(x^*) = 0 \quad j = 1,\dots,l$$

Punto factible

$$\lambda_i^* \ge 0 \quad i = 1,\dots,m$$

Condiciones suficientes con restricciones de igualdad y desigualdad (i)

- Sea x^* un punto factible

  $I=\{i/g_i(x^*)=0\}$ el conjunto de restricciones activas

  f y $\{g_i, i\in I\}$ convexas y diferenciables en **toda la región factible**

- Si existen unos escalares $\{\lambda_i, i\in I; \mu_j, j=1,\ldots,l\}$ 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 toda la región factible si $\mu_j>0$ y h_j sea cóncava en toda la región factible si $\mu_j<0$, entonces x^* es mínimo global

CONTENIDO

- ❑ PROBLEMAS DE PROGRAMACIÓN NO LINEAL
- ❑ TIPOS DE PROBLEMAS NLP
- ❑ CLASIFICACIÓN DE MÉTODOS DE RESOLUCIÓN SIN RESTRICCIONES
- ❑ CONDICIONES DE OPTIMALIDAD PARA OPTIMIZACIÓN SIN RESTRICCIONES
- ❑ CONDICIONES DE OPTIMALIDAD PARA NLP
- ➢ METHODS FOR UNCONSTRAINED OPTIMIZATION (master)
- ❑ NONLINEAR PROGRAMMING METHODS (master)

Clasificación métodos optimización SIN restricciones según el uso de derivadas

- Sin derivadas
  - ✓ Método de búsqueda aleatoria
  - ✓ Método de Hooke y Jeeves
  - ✓ Método de Rosenbrock (de las coordenadas rotativas o cíclicas)
  - ✓ Método de Nelder y Mead (símplice)
- Primeras derivadas (gradiente)
  - ✓ Método de máximo descenso (Steepest Descent)
  - ✓ Método del gradiente conjugado (Flectcher y Reeves)
- Segundas derivadas (hessiano)
  - ✓ Método de Newton
  - ✓ Métodos cuasi Newton (Broyden-Fletcher-Goldfarb-Shanno BFGS, Davidon-Fletcher-Powell DFP)

Método de Newton para función unidimensional (i)

- Los algoritmos de interpolación realizan en cada iteración una aproximación de la función *f*, en el punto x_k considerado en dicha iteración, por un polinomio de segundo o de tercer grado. Para hacerlo necesita evaluar las primeras derivadas de la función en x_k.
- Este método ajusta, en la iteración *k*, una parábola $q(x)$ a $f(x)$ y toma x_{k+1} como el vértice de dicha parábola

$$q(x_{k+1}) = f(x_k) + f'(x_k)(x_{k+1} - x_k) + \frac{1}{2} f''(x_k)(x_{k+1} - x_k)^2$$

$$q'(x_{k+1}) = 0 \Rightarrow x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

- El algoritmo converge cuadráticamente bajo ciertas condiciones, pero es muy inestable y suele ser necesario tomar precauciones e incluir protecciones.

Método de Newton para función unidimensional (ii)

$f(x) = (x-1)^3 + 2(x-1)^2 + 3$ $\quad f'(x) = 3(x-1)^2 + 4(x-1)$ $\quad f''(x) = 6(x-1) + 4$

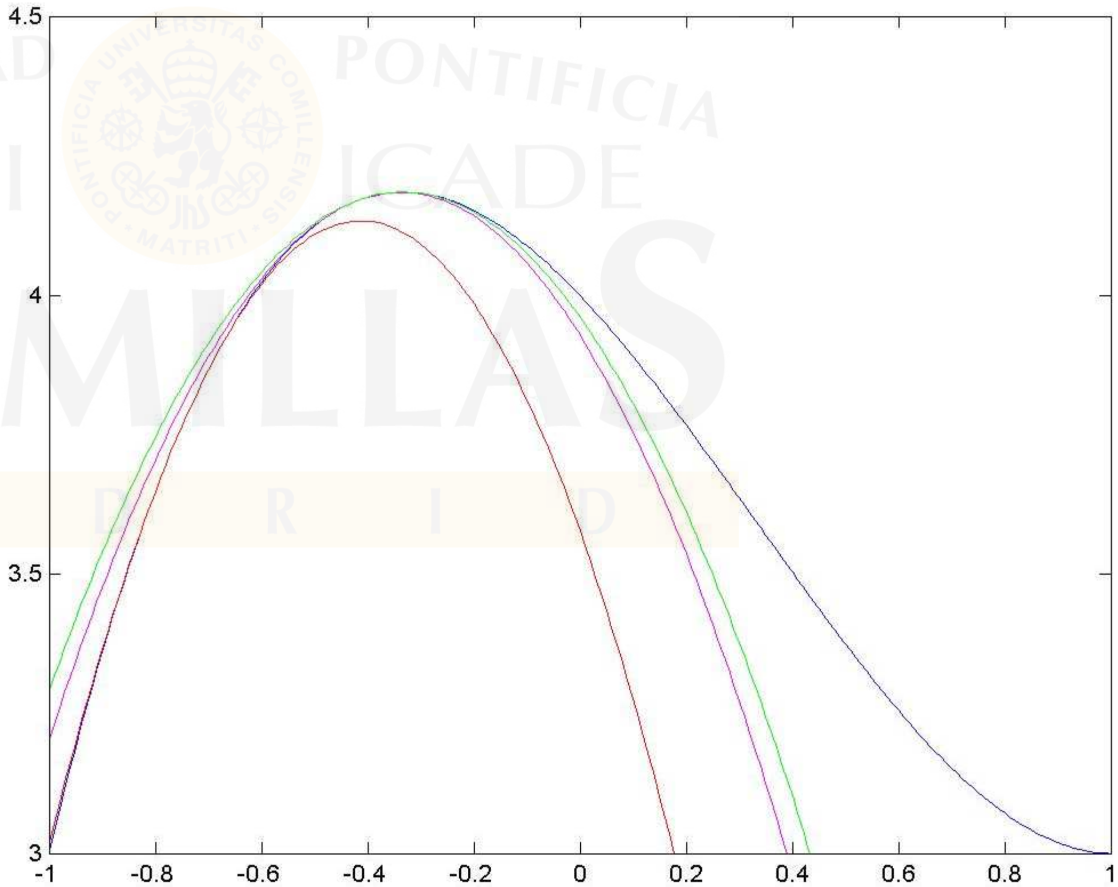
Secuencia de puntos

$x_0 = -0.75$

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = -0.4135$$

$x_2 = -0.3376$

$x_3 = -0.3333$



Procedimiento general de optimización

- Genera una secuencia de puntos hasta convergencia
  - ✓ Partir de un punto inicial x_k
  - ✓ Buscar una dirección de movimiento p_k
  - ✓ Calcular la longitud de paso α_k
  - ✓ Actualizar al nuevo punto x_{k+1}

  $$x_{k+1} = x_k + \alpha_k p_k$$

Condiciones de cada iteración

$$x_{k+1} = x_k + \alpha_k p_k$$

- Elegir el nuevo punto de manera que el valor de la función disminuya $f(x_{k+1}) < f(x_k)$
- Condiciones sobre la dirección de movimiento p_k
  - ✓ La dirección es descendente $p^T \nabla f(x_k) < 0$
  - ✓ El descenso es suficiente (vectores no ortogonales) $-\dfrac{p^T \nabla f(x_k)}{\|p\| \cdot \|\nabla f(x_k)\|} \geq \varepsilon > 0$
  - ✓ La dirección de movimiento está relacionada con gradiente $\|p\| \geq m \|\nabla f(x_k)\|$
- Condiciones sobre el escalar α_k
  - ✓ El descenso es suficiente (condición de Armijo)

    $$f(x_{k+1}) \leq f(x_k) + \mu \alpha_k p_k^T \nabla f(x_k) \qquad 0 < \mu < 1$$

  - ✓ Descenso no demasiado pequeño

    Por ejemplo, se define α_k como una secuencia 1, ½, ¼, 1/8, etc. Se van utilizando los valores de la secuencia empezando por 1. Si para α=1 se satisface la condición anterior se para, si no se utiliza el valor siguiente

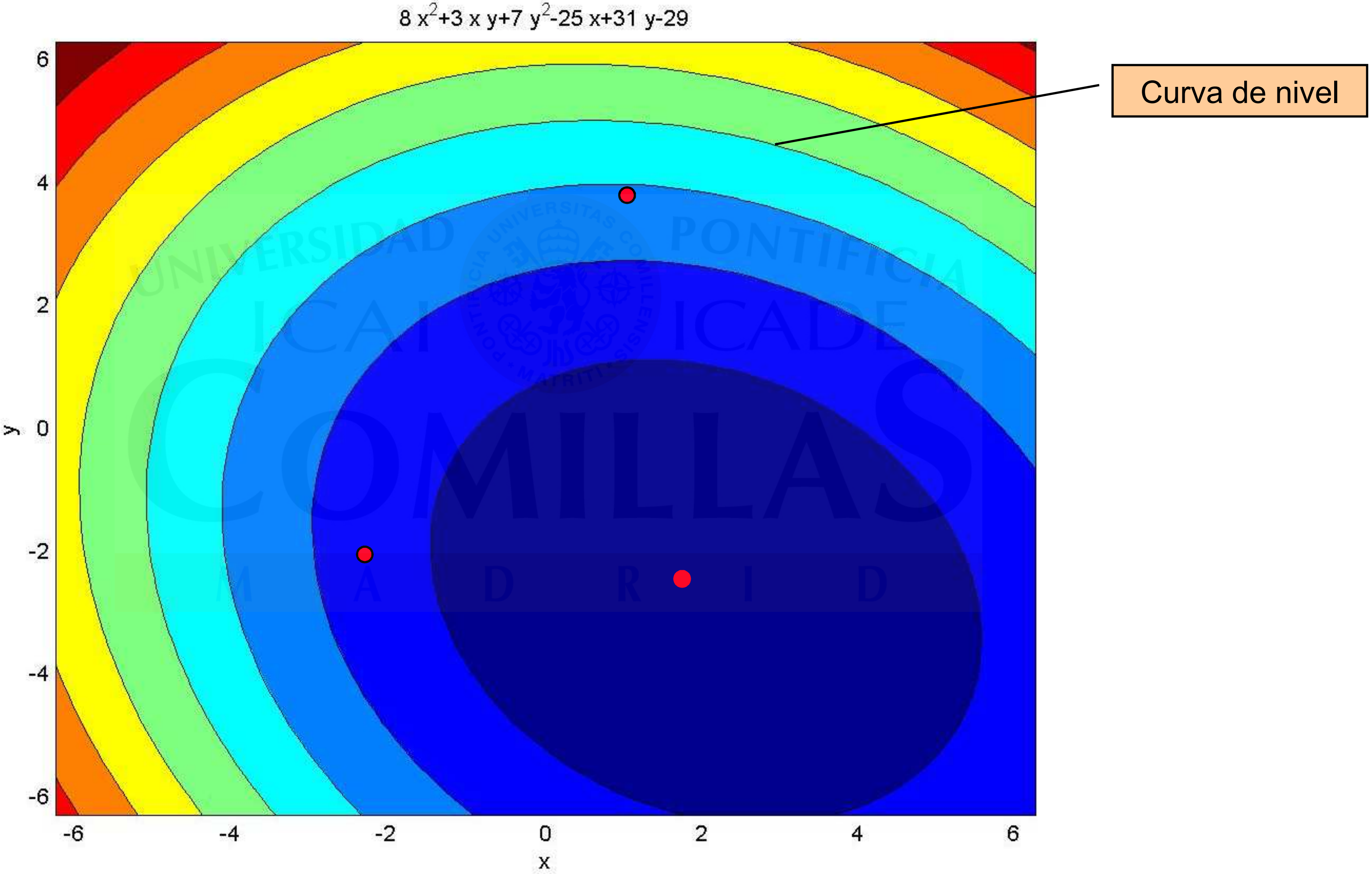
  - ✓ Minimiza el valor de la función: método de búsqueda unidimensional (*line-search methods*)

    $$\min_{\alpha > 0} \; F(\alpha) \equiv f(x_k + \alpha p_k)$$

Método de Rosenbrock o de coordenadas cíclicas

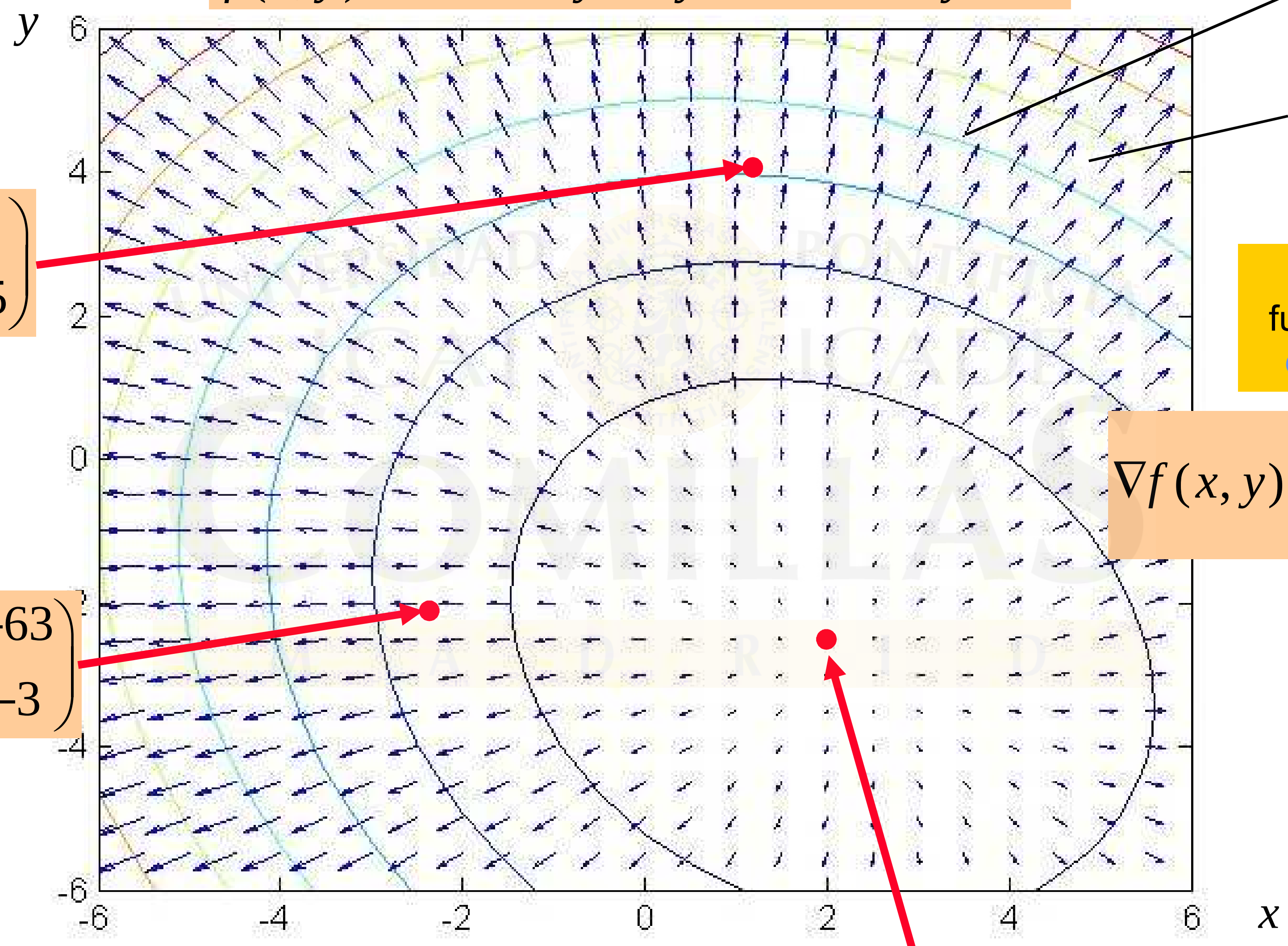
- Consiste en partir de un punto x_1 y minimizar la función f en la dirección $d_1=(1,0,0,\ldots,0)$ (minimización de una función unidimensional); alcanzado el punto x_2 que minimiza la función en esa dirección, se minimiza desde ese punto en la dirección $d_2=(0,1,0,\ldots,0)$ para determinar el punto x_3 y así sucesivamente hasta llegar al punto x_{n+1} en que se vuelve a minimizar en la dirección $d_1=(1,0,0,\ldots,0)$.
- El proceso se repite hasta alcanzar la precisión deseada. No es un método muy eficiente, los otros métodos aprovechan mejor las direcciones detectadas de mejora, pero es bueno para hacerse una primera idea de lo que son métodos de búsqueda. Cada búsqueda unidimensional puede ser con o sin derivadas.

Gradiente (i)



Gradiente (ii)

$$f(x,y) = 8x^2 + 3xy + 7y^2 - 25x + 31y - 29$$



Curva de nivel

Vector gradiente

$$\nabla f(2,3) = \begin{pmatrix} 5 \\ 77.5 \end{pmatrix}$$

El gradiente de una función da la dirección de máximo aumento

$$\nabla f(x,y) = \begin{pmatrix} 16x + 3y - 25 \\ 3x + 14y + 31 \end{pmatrix}$$

$$\nabla f(-2,-2) = \begin{pmatrix} -63 \\ -3 \end{pmatrix}$$

$$(x,y) = (2.060, -2.656)$$

Método del máximo descenso

- No requiere el uso de segundas derivadas. Por tanto, es poco costoso computacionalmente.
- Tiene una convergencia más lenta, lineal.
- En general, no se debe utilizar.
- Se aproxima la función por la serie de Taylor de primer orden
- La dirección de búsqueda resultante para minimizar al máximo la función es la opuesta al gradiente de la función

$$p_k = -\nabla f(x_k)$$

Ejemplo 1

$$f(x,y) = (x-2)^2 + (y-1)^2$$

$$\min_{\alpha>0} F(\alpha) = f(x_k + \alpha p_k)$$

$$p_k = -\nabla f(x_k, y_k) = -\begin{pmatrix} 2(x_k - 2) \\ 2(y_k - 1) \end{pmatrix}$$

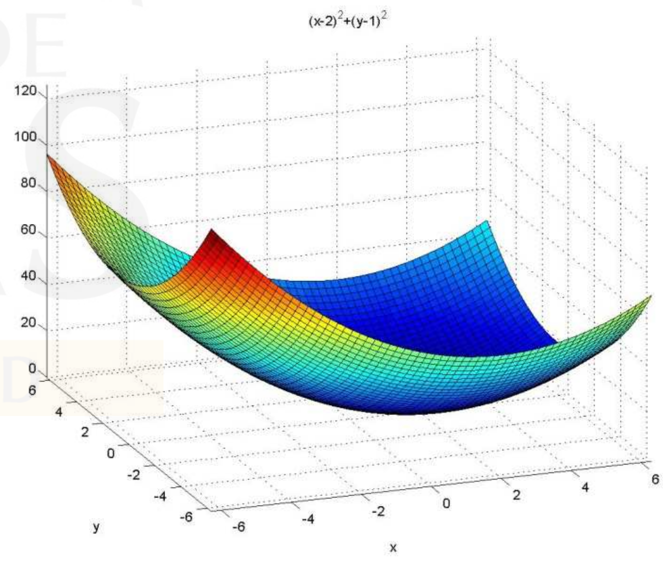
- Punto inicial x_k=(0,0)
- Dirección de movimiento p_k=(4,2)

$$\min_{\alpha} F(\alpha) = (4\alpha-2)^2 + (2\alpha-1)^2 = 20\alpha^2 - 20\alpha + 5$$

$$F'(\alpha) = 0$$

$$40\alpha - 20 = 0$$

$$\alpha = 0.5$$



- Siguiente punto $x_{k+1}= x_k + \alpha_k p_k$=(2,1)
- Dirección de movimiento p_k=(0,0)
- Se ha llegado al óptimo ya que el gradiente es 0

Ejemplo 2 (i)

$$f(x,y) = 8x^2 + 3xy + 7y^2 - 25x + 31y - 29$$

$$p_k = -\nabla f(x_k, y_k) = -\begin{pmatrix} 16x_k + 3y_k - 25 \\ 3x_k + 14y_k + 31 \end{pmatrix}$$

$$\min_\alpha F(\alpha) = f(x_k + \alpha p_k)$$

- Punto inicial x_k=(-4,4)
- Dirección de movimiento p_k=(77,-75)

$$\min_\alpha F(\alpha) = 8(-4+77\alpha)^2 + 3(-4+77\alpha)(4-75\alpha) +$$
$$7(4-75\alpha)^2 - 25(-4+77\alpha) + 31(4-75\alpha) - 29 =$$
$$69482\alpha^2 - 11554\alpha + 387$$

$$F'(\alpha) = 0$$

$$138964\alpha - 11554 = 0$$

$$\alpha = 0.0831$$

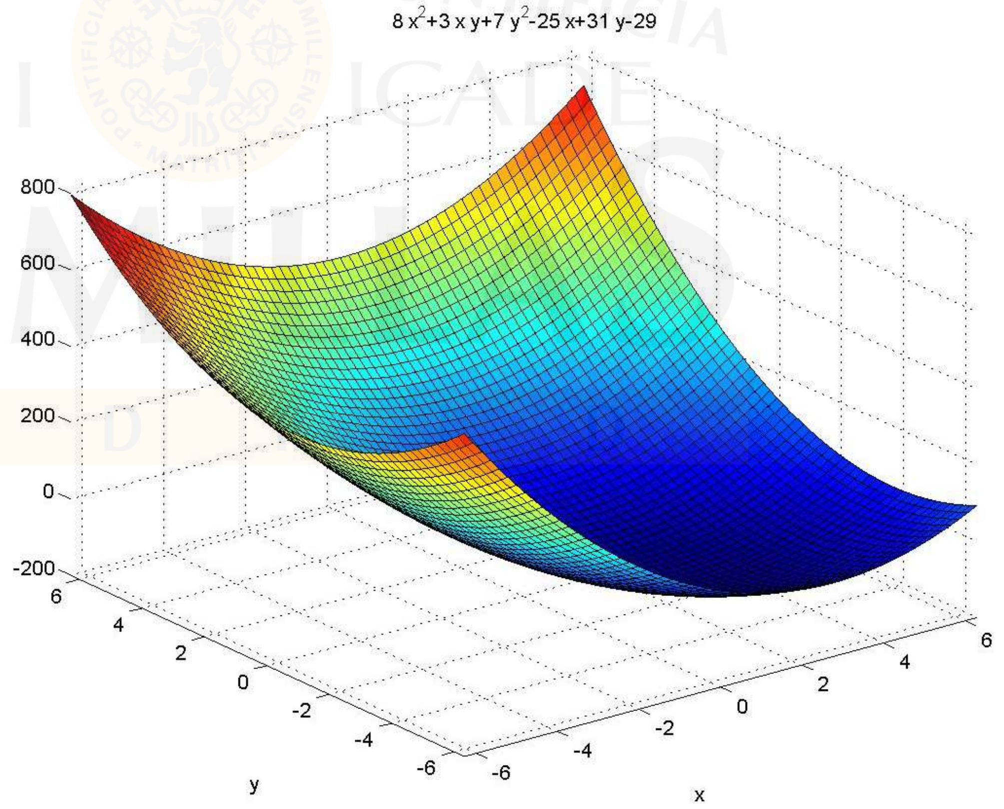
- Siguiente punto (-4+0.0831x77,4-0.0831x75)=(2.402,-2.236)

Ejemplo 2 (ii)

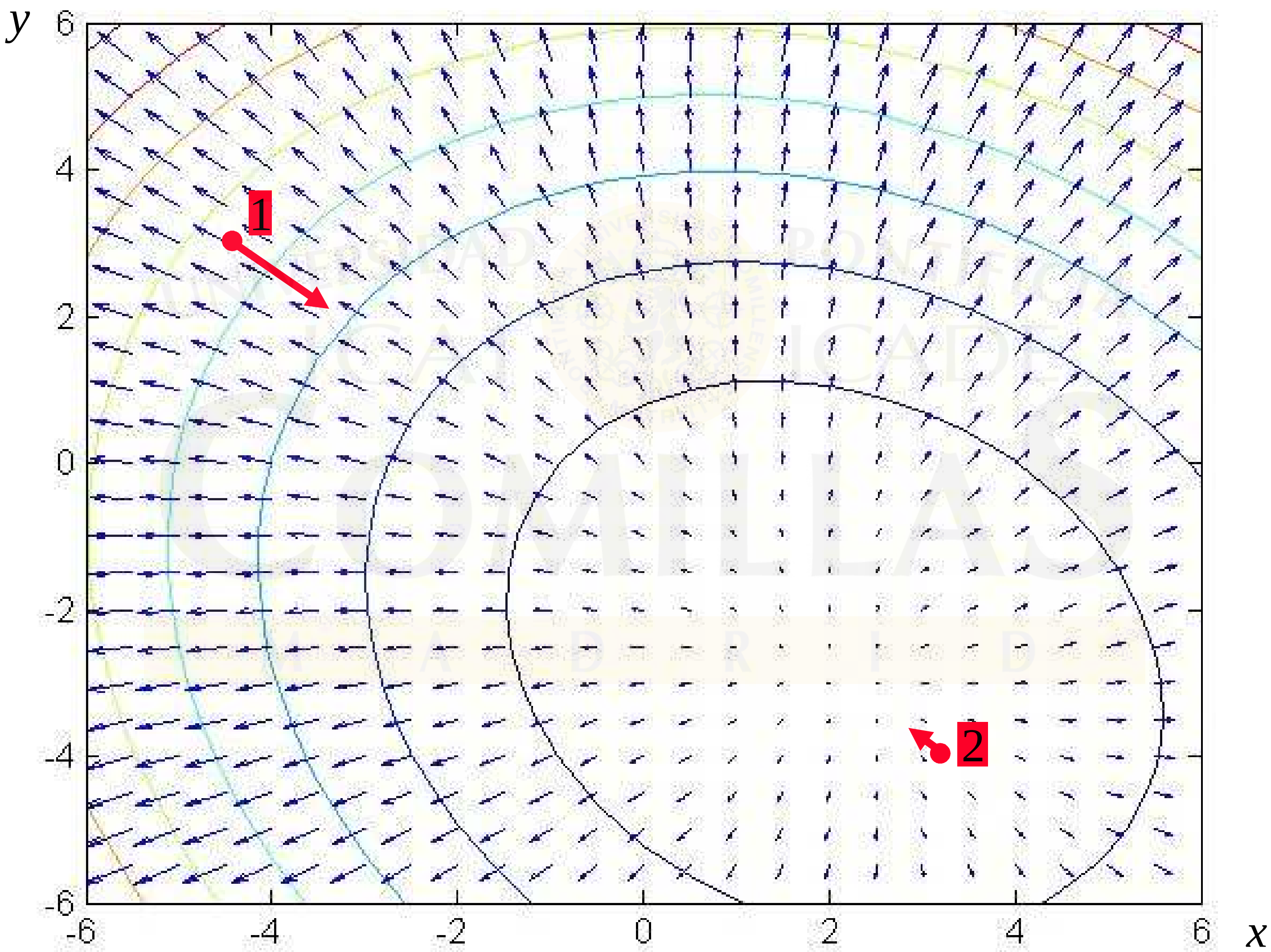
- Dirección de movimiento p_k=(-6.726,-6.905)

$$\begin{aligned}\min_{\alpha} F(\alpha) = {} & 8(2.4-6.7\alpha)^2 + 3(2.4-6.7\alpha)(-2.2-6.9\alpha) + \\ & 7(-2.2-6.9\alpha)^2 - 25(2.4-6.7\alpha) + 31(-2.2-6.9\alpha) - 29\end{aligned}$$

- Etc.



Ejemplo 2 (iii)



Ejemplo 3 (i)

$$\min_x \; f(x) = \frac{1}{2} x^T Q x - b^T x \qquad Q = \begin{pmatrix} 1 & \cdot & \cdot \\ \cdot & 5 & \cdot \\ \cdot & \cdot & 25 \end{pmatrix} \qquad b = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$$

- Para funciones cuadráticas, el óptimo se puede determinar como $\nabla f(x) = Qx - b = 0$ y, por tanto, $x^* = Q^{-1} b = \begin{pmatrix} -1 \\ -1/5 \\ -1/25 \end{pmatrix}$
- La dirección de máximo descenso es $p_k = -\nabla f(x_k) = -(Qx_k - b)$
- Si se utiliza una búsqueda unidimensional exacta el valor resultante de es

$$\alpha_k = -\frac{\nabla f(x_k)^T p_k}{p_k^T Q p_k}$$

Ejemplo 3 (ii)

- Punto inicial $x_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $f(x_0) = 0$ $\nabla f(x_0) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $\alpha_0 = 3/31 = 0.097$

- Se utiliza la norma 2 del gradiente como medida de la convergencia $\|\nabla f(x_0)\| = \sqrt{1^2 + 1^2 + 1^2} = 1.73$
- Nuevo punto

$$x_1 = x_0 + \alpha_0 p_0 = \begin{pmatrix} -0.097 \\ -0.097 \\ -0.097 \end{pmatrix} \quad f(x_1) = -0.145 \quad \nabla f(x_1) = \begin{pmatrix} 0.903 \\ 0.516 \\ -1.419 \end{pmatrix} \quad \alpha_1 = 0.059$$

- Norma 2 del gradiente $\|\nabla f(x_1)\| = 1.760$

Ejemplo 3 (iii)

- Nuevo punto $$x_2 = \begin{pmatrix} -0.150 \\ -0.127 \\ -0.013 \end{pmatrix} \qquad f(x_2) = -0.237 \qquad \nabla f(x_2) = \begin{pmatrix} 0.850 \\ 0.364 \\ 0.673 \end{pmatrix}$$
- Norma 2 del gradiente $\left\| \nabla f(x_2) \right\| = 1.144$
- Etc.

- El proceso continúa hasta que la norma 2 del gradiente se haga suficientemente pequeña (inferior a una cierta tolerancia, 10^{-8} por ejemplo). En este ejemplo se necesitan 216 iteraciones hasta alcanzar esta tolerancia.

Convergencia del método del máximo descenso para funciones cuadráticas

- La convergencia del método de máximo descenso para una función cuadrática con búsqueda unidimensional exacta es lineal. La relación de mejora entre dos iteraciones consecutivas se puede acotar superiormente de esta manera:

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \leq \left(\frac{\mathrm{cond}(Q) - 1}{\mathrm{cond}(Q) + 1} \right)^2$$

  donde el número de condición de la matriz A que se define como $\mathrm{cond}(A) \equiv \|A\| \cdot \|A^{-1}\|$ y $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$, siendo $\lambda_{\max}(A^T A)$ el autovalor máximo de la matriz $A^T A$.

- Si A es una matriz definida positiva y simétrica $\mathrm{cond}(A) = \lambda_1 / \lambda_n$ siendo λ_1 y λ_n el mayor y menor autovalor respectivamente. En el ejemplo anterior $\mathrm{cond}(Q) = 25$
- Un valor elevado del número de condición indica que la convergencia es muy lenta. Para $\mathrm{cond}(Q) = 100$ este método mejora la solución como mucho del 4 % en cada iteración.

Número de condición

- Mide la sensibilidad de la solución de un sistema de ecuaciones lineales a los errores en los datos.

Convergencia del método del máximo descenso para funciones no lineales generales

- Para funciones no lineales generales la convergencia es también lineal con esta cota superior

$$\left(\frac{\operatorname{cond}(Q)-1}{\operatorname{cond}(Q)+1}\right)^2$$

donde ahora $Q=\nabla^2 f(x^*)$ es el hessiano de la función en la solución.

Método de Newton para resolución de un sistema de ecuaciones no lineales (i)

- Resuelve iterativamente un sistema de ecuaciones no lineales

$$\begin{aligned} f_1(x_1,\dots,x_n) &= 0 \\ f_2(x_1,\dots,x_n) &= 0 \\ &\vdots \\ f_n(x_1,\dots,x_n) &= 0 \end{aligned}$$

- Aproxima la función no lineal por una función lineal en cada punto (iteración), utilizando la expansión en serie de Taylor de primer orden

$$\begin{aligned} & f(x_k + p_k) \approx f(x_k) + \nabla f(x_k)^T p_k \\ & f(x^*) \approx f(x_k) + \nabla f(x_k)^T p_k = 0 \\ & p_k = -\nabla f(x_k)^{-T} f(x_k) \\ & x_{k+1} = x_k + p_k = x_k - \nabla f(x_k)^{-T} f(x_k) \end{aligned}$$

- $\nabla f(x)^T = \left(\nabla f_1(x) \quad \nabla f_2(x) \quad \dots \quad \nabla f_n(x)\right)^T$ jacobiano de la función

Método de Newton para resolución de un sistema de ecuaciones no lineales (ii)

- Converge cuadráticamente cuando el punto está próximo a la solución
- El jacobiano de la función en cada punto debe ser no singular

Método de Newton para resolución de un sistema de ecuaciones no lineales (ii)

$$f_1(x_1, x_2) = 3x_1x_2 + 7x_1 + 2x_2 - 3 = 0$$
$$f_2(x_1, x_2) = 5x_1x_2 - 9x_1 - 4x_2 + 6 = 0$$

$$\nabla f(x_1, x_2) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_2}{\partial x_1} \\ \dfrac{\partial f_1}{\partial x_2} & \dfrac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 3x_2 + 7 & 5x_2 - 9 \\ 3x_1 + 2 & 5x_1 - 4 \end{pmatrix}$$

$$x_{k+1} = x_k - \begin{pmatrix} 3x_2 + 7 & 5x_2 - 9 \\ 3x_1 + 2 & 5x_1 - 4 \end{pmatrix}^{-T} \begin{pmatrix} 3x_1x_2 + 7x_1 + 2x_2 - 3 \\ 5x_1x_2 - 9x_1 - 4x_2 + 6 \end{pmatrix}$$

- Suponemos como punto inicial $x_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$$x_1 = x_0 - \begin{pmatrix} 3x_2 + 7 & 5x_2 - 9 \\ 3x_1 + 2 & 5x_1 - 4 \end{pmatrix}^{-T} \begin{pmatrix} 3x_1x_2 + 7x_1 + 2x_2 - 3 \\ 5x_1x_2 - 9x_1 - 4x_2 + 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 13 & 1 \\ 5 & 1 \end{pmatrix}^{-T} \begin{pmatrix} 14 \\ -1 \end{pmatrix} = \begin{pmatrix} -1.375 \\ 5.375 \end{pmatrix}$$

- Después de 8 iteraciones el punto toma el valor aproximado de $x_8 = \begin{pmatrix} 0 \\ 1.5 \end{pmatrix}$ y las funciones toman valor 0.

Método de Newton para optimización

- Se aplica el método de Newton de resolución de un sistema de ecuaciones a la condición necesaria de optimalidad de primer orden. $\nabla f(x) = 0$
- El jacobiano de esta función $\nabla f(x)$ es el hessiano $\nabla^2 f(x)$
- Iteración $x_{k+1} = x_k + p_k = x_k - \nabla^2 f(x_k)^{-T} \nabla f(x_k)$
- Donde p_k (la dirección de Newton) se obtiene mediante la resolución de un sistema de ecuaciones lineales (sistema de Newton) en lugar de calcular la inversa del hessiano.

  $\nabla^2 f(x_k) p_k = -\nabla f(x_k)$

Ejemplo 1 (i)

$$f(x,y)=(x-2)^2+(y-1)^2$$

$$\nabla f(x_k,y_k)=\begin{pmatrix}2(x_k-2)\\2(y_k-1)\end{pmatrix}$$

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

$$\nabla^2 f(x_k)p_k=-\nabla f(x_k)$$

- Punto inicial x_k=(0,0)
- Dirección de movimiento p_k=(2,1)

$$p_k=-\begin{pmatrix}2 & \cdot\\ \cdot & 2\end{pmatrix}^{-T}\begin{pmatrix}-4\\-2\end{pmatrix}=-\begin{pmatrix}0.5 & \cdot\\ \cdot & 0.5\end{pmatrix}\begin{pmatrix}-4\\-2\end{pmatrix}=\begin{pmatrix}2\\1\end{pmatrix}$$

- Cálculo de α_k

$$\min_{\alpha} F(\alpha)=(2\alpha-2)^2+(\alpha-1)^2=5\alpha^2-10\alpha+5$$

$$F'(\alpha)=0$$

$$10\alpha-10=0$$

$$\alpha=1$$

Ejemplo 1 (ii)

- Siguiente punto $x_{k+1} = x_k + \alpha_k p_k = (2,1)$
- Dirección de movimiento $p_k = (0,0)$
- Se ha llegado al óptimo ya que el gradiente es 0

Método cuasi Newton

- Disminuir el coste computacional asociado a calcular y almacenar el hessiano y a resolver el sistema de ecuaciones lineales.
- Basados en aproximar el hessiano de la función $\nabla^2 f(x_k)$ en cada punto por otra matriz B_k definida positiva más fácil de calcular. Los diferentes métodos cuasi Newton difieren en la elección de B_k y en su actualización.
- Ventajas:
  - ✓ Prescinde del cálculo de las segundas derivadas (hessiano), utiliza sólo primeras derivadas en la aproximación de B_k
  - ✓ La dirección de búsqueda se puede calcular con menor coste computacional
- Desventajas:
  - ✓ La convergencia ya no es cuadrática, iteraciones menos costosas
  - ✓ Requieren el almacenamiento de una matriz, luego no son aptos para problemas grandes

Cálculo de la matriz B_k

- Condición de la secante $\nabla^2 f(x_k)(x_k - x_{k-1}) \approx \nabla f(x_k) - \nabla f(x_{k-1})$
- Aproximación $B_k(x_k - x_{k-1}) \approx \nabla f(x_k) - \nabla f(x_{k-1})$
- Para una función cuadrática el hessiano Q satisface esta condición, luego la aproximación es exacta.
- Definiendo $s_k = x_{k+1} - x_k$ y $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ entonces $B_{k+1} s_k = y_k$
- En los métodos cuasi Newton la matriz B_k se actualiza en cada iteración

  $B_{k+1} = B_k$ + actualización
- La inicialización de B_k suele ser la matriz identidad $B_0 = I$.

Procedimiento método cuasi Newton

1. Especificación de una solución inicial x_0 y de una aproximación inicial al hessiano B_0
2. Iterar k=0,1,… hasta encontrar la solución óptima
   - ✓ Calcular la dirección de movimiento $B_k p_k = -\nabla f(x_k)$
   - ✓ Hacer una búsqueda unidimensional para determinar $x_{k+1} = x_k + \alpha_k p_k$
   - ✓ Calcular $s_k = x_{k+1} - x_k$ $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$
   - ✓ Actualizar la aproximación del hessiano

     $B_{k+1} = B_k$ + actualización

Actualizaciones de la matriz B_k

- Actualización BFGS (Broyden-Fletcher-Goldfarb-Shanno)

$$B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$

- Actualización DFP (Davidon-Fletcher-Powell)

$$B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} + (s_k^T B_k s_k) u_k u_k^T \qquad u_k = \frac{y_k}{y_k^T s_k} - \frac{B_k s_k}{s_k^T B_k s_k}$$

- En ambas actualizaciones para garantizar que B_{k+1} sigue siendo definida positiva se debe cumplir $y_k^T s_k > 0$ condición que se debe garantizar controlando la búsqueda unidimensional.
- La matriz B_k se aproximada más al hessiano en cada iteración

CONTENIDO

- PROBLEMAS DE PROGRAMACIÓN NO LINEAL
- TIPOS DE PROBLEMAS NLP
- CLASIFICACIÓN DE MÉTODOS DE RESOLUCIÓN
- CONDICIONES DE OPTIMALIDAD PARA OPTIMIZACIÓN SIN RESTRICCIONES
- CONDICIONES DE OPTIMALIDAD PARA NLP
- METHODS FOR UNCONSTRAINED OPTIMIZATION (master)
- ➢ NONLINEAR PROGRAMMING METHODS (master)

Programación no lineal

- Sea el problema

$$\min_x f(x)$$
$$g_i(x) \le 0 \quad i = 1,\dots,m$$
$$h_j(x) = 0 \quad j = 1,\dots,l$$

$$f, g_i, h_j : \mathbb{R}^n \to \mathbb{R} \qquad x \in \mathbb{R}^n$$

donde *f*, *g* y *h* son funciones diferenciables con primera y segunda derivadas continuas.

- Si la región factible es convexa y la función objetivo es convexa en toda la región factible cualquier óptimo local es también global

Métodos de programación no lineal (NLP)

1. Métodos de penalización: minimizan una función relacionada con el lagrangiano que tiene el mismo mínimo
    - Métodos de penalización y barrera
    - Método del lagrangiano aumentado
    - Método de programación cuadrática secuencial
2. Métodos factibles: mantienen factibilidad al partir de un punto factible y moverse en direcciones factibles
    - ✓ Generalización del método simplex de LP. Resuelven una secuencia de subproblemas con un conjunto de restricciones activas que cambia en cada iteración.
    - ✓ Inconvenientes: selección del conjunto de restricciones activas y dificultad de satisfacer las restricciones
        - Método del gradiente reducido

Métodos de penalización

- Resuelven un problema NLP resolviendo una secuencia de problemas de optimización sin restricciones. En el límite la solución de ambos problemas es la misma.
- En la función objetivo se incluyen unas penalizaciones que miden las violaciones de las restricciones y además unos parámetros que determinan la importancia de cada restricción.
  - ✓ Métodos de penalización (método de penalización exterior)
    - Penalizan la violación de una restricción. Se mueve por puntos infactibles.
    - Mejor para restricciones de igualdad
  - ✓ Métodos barrera (método de penalización interior)
    - Evitan que se alcance el contorno de una restricción. Puntos estrictamente factibles.
    - No son válidos para restricciones de igualdad

Método de penalización (exterior)

- Sea el problema
$$\min_x f(x)$$
$$g_i(x) = 0 \quad i = 1,\dots,m$$

- Se define la función de penalización

$$\psi(x) = \frac{1}{2}\sum_{i=1}^{m} g_i(x)^2 = \frac{1}{2} g(x)^T g(x)$$

- El problema original se transforma en

$$\min_x \pi(x,\rho_k) = f(x) + \rho_k \psi(x)$$

siendo ρ_k, el parámetro de penalización, un escalar positivo que crece monótonamente hacia ∞

- Al avanzar las iteraciones los óptimos se mueven hacia la región factible

Método de penalización (exterior)

- Condición de optimalidad de la función de penalización

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

- Definimos $\lambda_i = \lambda_i(\rho) = \rho g_i(x(\rho))$. Estimación de los multiplicadores de Lagrange.
- Entonces, las condiciones de optimalidad de la función original son

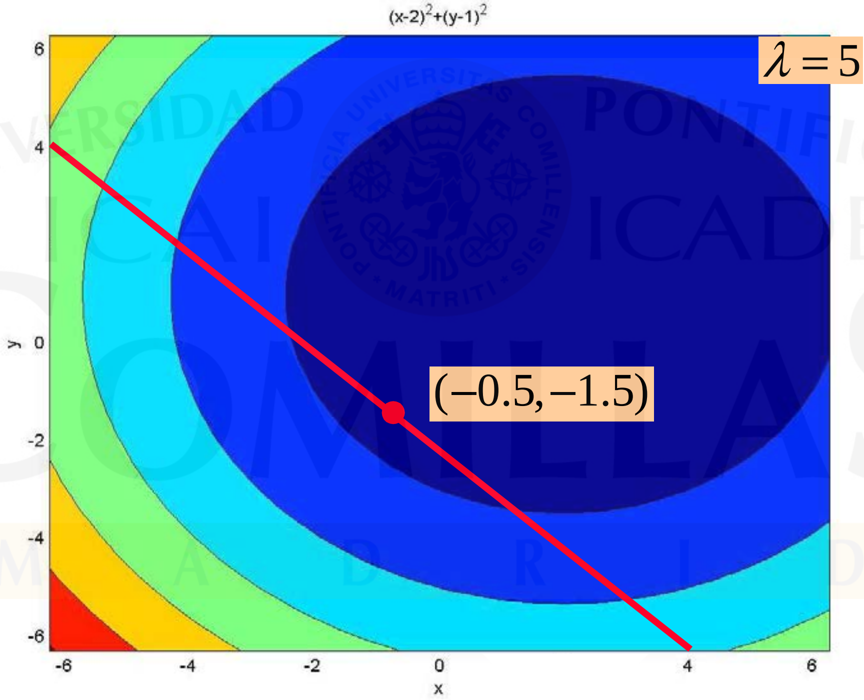
$$\begin{aligned}
&\nabla f(x(\rho)) + \sum_{i=1}^{m} \lambda_i(\rho) \nabla g_i(x(\rho)) = 0 \\
&\lambda_i(\rho) = \rho g_i(x(\rho)) \quad i = 1,\ldots,m \\
&\lambda_i(\rho) \geq 0 \quad i = 1,\ldots,m
\end{aligned}$$

Método de penalización (exterior)

- El hessiano del problema penalizado puede dar un número de condición elevado $\nabla_x^2 \pi(x(\rho), \rho)$

Ejemplo 1 (i)

$$\min (x-2)^2 + (y-1)^2$$
$$x + y = -2$$



Ejemplo 1 (ii)

$$\min (x-2)^2+(y-1)^2+\frac{1}{2}\rho(x+y+2)^2$$

- Condición de optimalidad de primer orden $\nabla_x \pi(x)=0$

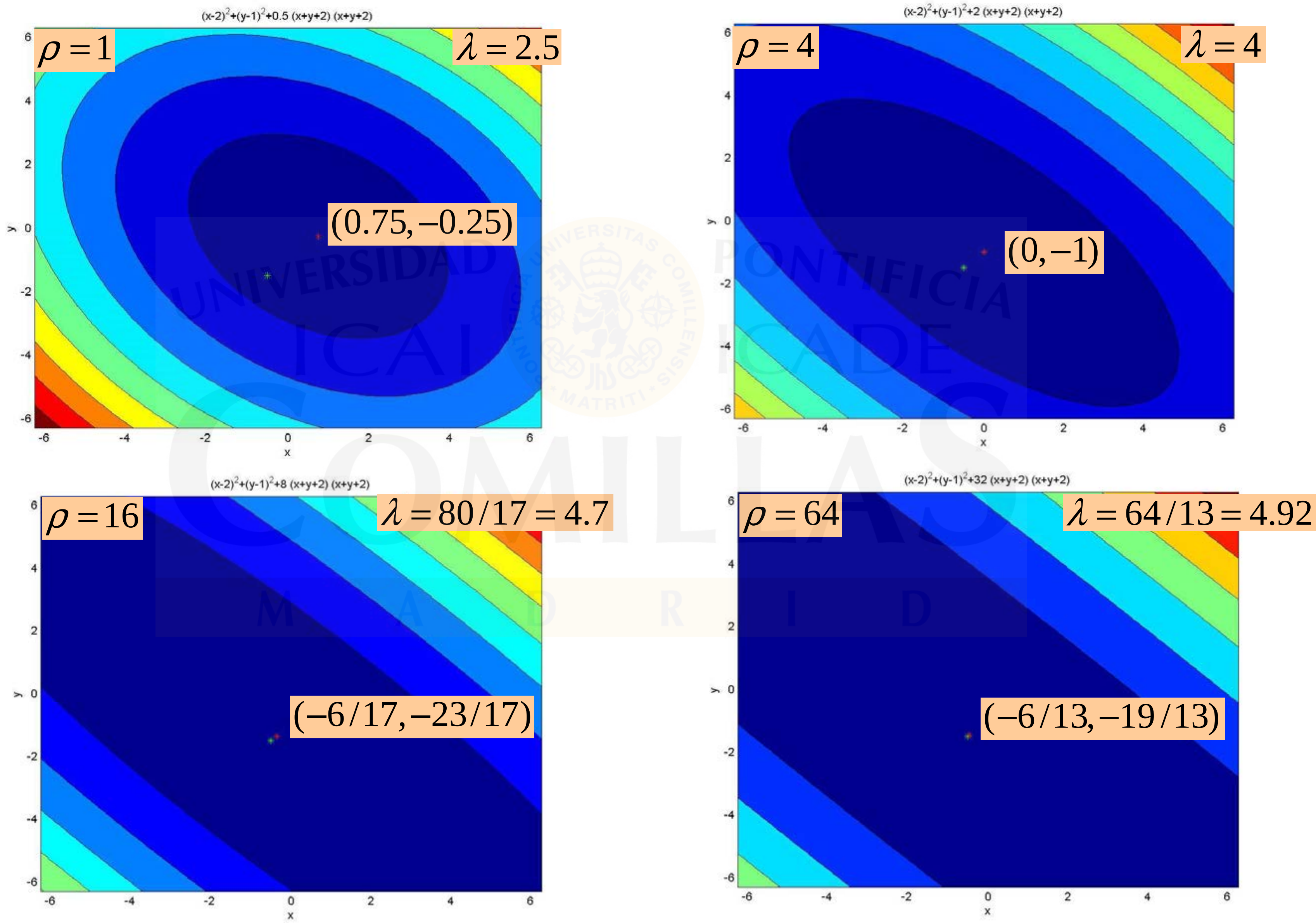
$$2(x-2)+\rho(x+y+2)=0$$
$$2(y-1)+\rho(x+y+2)=0$$

$$x=\frac{4-\rho}{2+2\rho} \qquad y=\frac{2-3\rho}{2+2\rho} \qquad \lambda=\frac{10\rho}{2+2\rho}$$

- Para ρ=1 $(x^*,y^*)=(0.75,-0.25)$ $\lambda=2.5$
- Para ρ=2 $(x^*,y^*)=(1/3,-2/3)$
- Para ρ=4 $(x^*,y^*)=(0,-1)$ $\lambda=4$
- Para ρ=8 $(x^*,y^*)=(-2/9,-11/9)$
- Para ρ=16 $(x^*,y^*)=(-6/17,-23/17)$ $\lambda=80/17=4.7$
- Para ρ=32 $(x^*,y^*)=(-14/33,-47/33)$
- Para ρ=64 $(x^*,y^*)=(-6/13,-19/13)$ $\lambda=64/13=4.92$
- Para ρ=∞ $(x^*,y^*)=(-0.5,-1.5)$ $\lambda=5$

Cada uno es una optimización sin restricciones

Ejemplo 1 (iii)



Ejemplo 1 (iv)

- El hessiano del problema penalizado es

$$\nabla_x^2 \pi(x(\rho),\rho) = \begin{pmatrix} 2+\rho & \rho \\ \rho & 2+\rho \end{pmatrix}$$

- Para ρ=1 $\quad \nabla_x^2 \pi(x(\rho),\rho) = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \quad cond(\nabla^2 \pi) = \lambda_1/\lambda_2 = 4/2 = 2$

- Para ρ=4 $\quad \nabla_x^2 \pi(x(\rho),\rho) = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix} \quad cond(\nabla^2 \pi) = \lambda_1/\lambda_2 = 10/2 = 5$

- Para ρ=16 $\quad \nabla_x^2 \pi(x(\rho),\rho) = \begin{pmatrix} 18 & 16 \\ 16 & 18 \end{pmatrix} \quad cond(\nabla^2 \pi) = \lambda_1/\lambda_2 = 34/2 = 17$

- Para ρ=64 $\quad \nabla_x^2 \pi(x(\rho),\rho) = \begin{pmatrix} 66 & 64 \\ 64 & 66 \end{pmatrix} \quad cond(\nabla^2 \pi) = \lambda_1/\lambda_2 = 130/2 = 65$

Método barrera (penalización interior)

- Sea el problema
$$\min_x f(x)$$
$$g_i(x) \geq 0 \quad i = 1, \ldots, m$$

- Se define la función barrera

$$\phi(x) = -\sum_{i=1}^{m} \log(g_i(x))$$ o bien $$\phi(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)}$$

  - ✓ Es una función continua en el interior de la región factible que se hace ∞ al acercarse al contorno

- El problema original se transforma en
$$\min_x \pi(x, \mu_k) = f(x) + \mu_k \phi(x)$$
siendo μ_k, el parámetro barrera, un escalar positivo que decrece monótonamente hacia 0

- Al disminuir el parámetro los puntos se acercan al contorno de la región factible

Método de penalización (exterior)

- Condición de optimalidad de la función barrera

$$\nabla f(x) - \mu \sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)} = 0$$

- Definimos $\lambda_i = \lambda_i(\mu) = \dfrac{-\mu}{g_i(x)}$ $\quad \lambda_i(\mu) g_i(x) = -\mu$
- Estimación de los multiplicadores de Lagrange.
- Entonces, las condiciones de optimalidad de la función original son

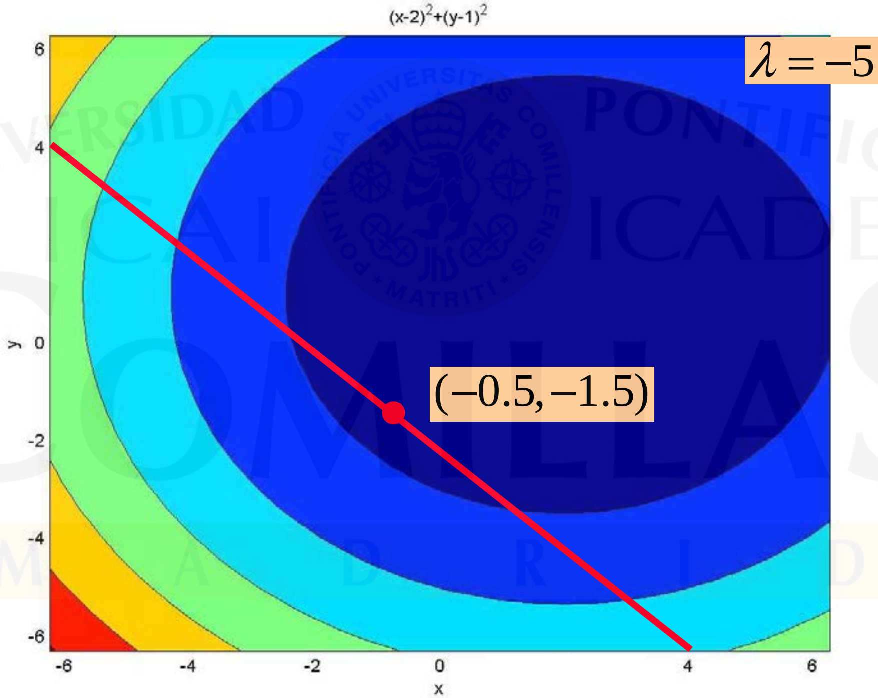
$$\begin{aligned} &\nabla f(x(\mu)) + \sum_{i=1}^{m} \lambda_i(\mu) \nabla g_i(x(\mu)) = 0 \\ &\lambda_i(\mu) g_i(x(\mu)) = -\mu \quad i = 1, \ldots, m \\ &\lambda_i(\mu) \le 0 \quad i = 1, \ldots, m \end{aligned}$$

Método de penalización (exterior)

- El hessiano del problema penalizado puede dar un número de condición elevado $\nabla_x^2 \pi(x(\mu), \mu)$

Ejemplo 1 (i)

$$\min (x-2)^2 + (y-1)^2$$
$$-x - y \geq 2$$



Ejemplo 1 (ii)

$$\min (x-2)^2 + (y-1)^2 - \mu \log(-x-y-2)$$

- Condición de optimalidad de primer orden $\nabla_x \pi(x) = 0$

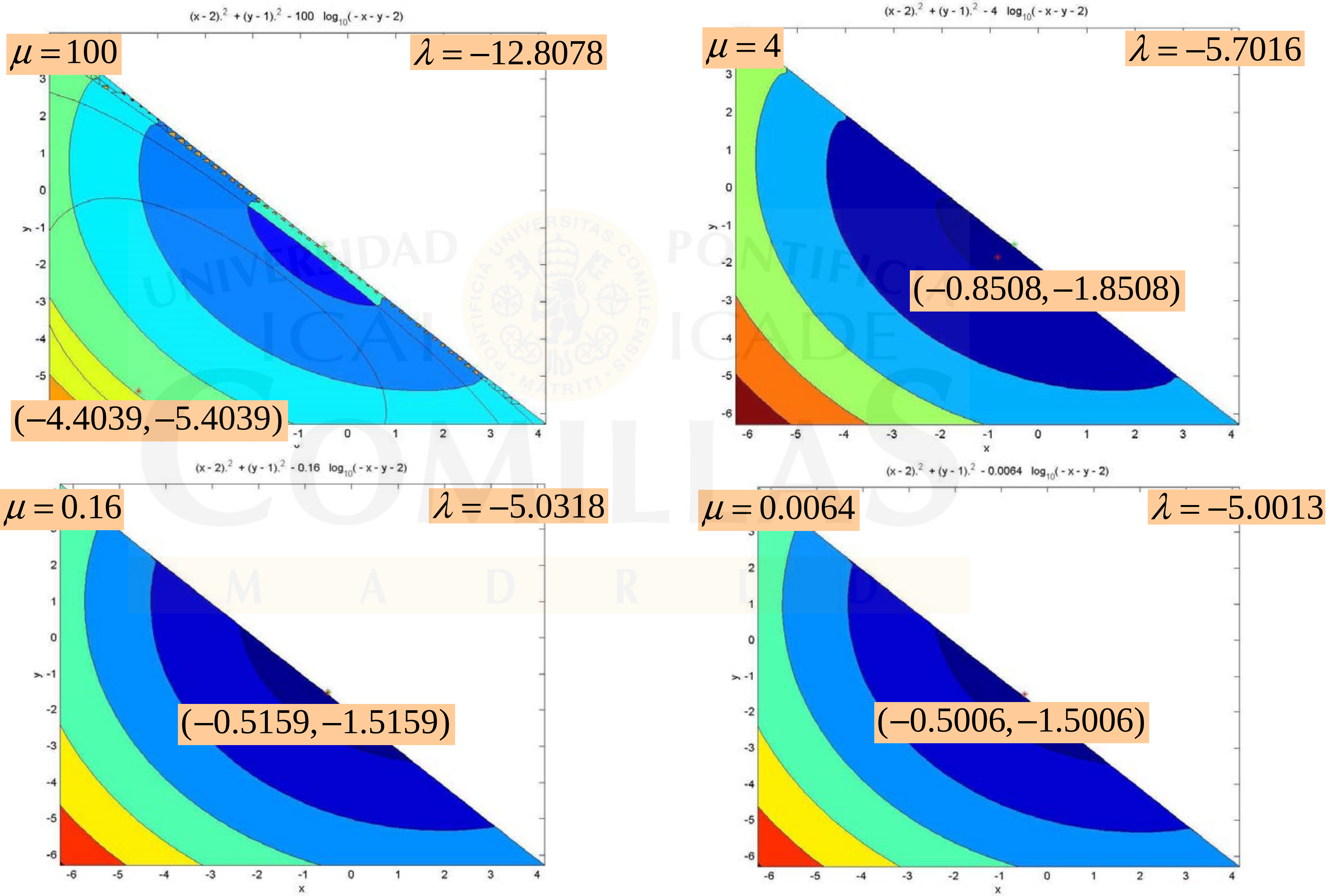
$$2(x-2) + \mu/(-x-y-2) = 0$$
$$2(y-1) + \mu/(-x-y-2) = 0$$

$$x = \frac{3}{4} \pm \frac{1}{8}\sqrt{100+16\mu} \qquad y = -\frac{1}{4} \pm \frac{1}{8}\sqrt{100+16\mu} \qquad \lambda = \frac{\mu}{x+y+2}$$

- Para μ=100 $\quad (x^*, y^*) = (-4.4039, -5.4039) \quad \lambda = -12.8078$
- Para μ=20 $\quad (x^*, y^*) = (-1.8117, -2.8117)$
- Para μ=4 $\quad (x^*, y^*) = (-0.8508, -1.8508) \quad \lambda = -5.7016$
- Para μ=0.8 $\quad (x^*, y^*) = (-0.5776, -1.5776)$
- Para μ=0.16 $\quad (x^*, y^*) = (-0.5159, -1.5159) \quad \lambda = -5.0318$
- Para μ=0.032 $\quad (x^*, y^*) = (-0.5032, -1.5032)$
- Para μ=0.0064 $\quad (x^*, y^*) = (-0.5006, -1.5006) \quad \lambda = -5.0013$
- Para μ=0 $\quad (x^*, y^*) = (-0.5, -1.5) \quad \lambda = -5$

Cada uno es una optimización sin restricciones

Ejemplo 1 (iii)



Ejemplo 1 (iv)

- El hessiano del problema barrera es

$$\nabla_x^2 \pi(x(\mu),\mu) = \begin{pmatrix} 2+\mu/(x+y+2)^2 & \mu/(x+y+2)^2 \\ \mu/(x+y+2)^2 & 2+\mu/(x+y+2)^2 \end{pmatrix}$$

- Para μ=100 $\nabla_x^2 \pi(x(\mu),\mu) = \begin{pmatrix} 3.64 & 1.64 \\ 1.64 & 3.64 \end{pmatrix}$ $cond(\nabla^2 \pi) = \lambda_1/\lambda_2 = 5.28/2 = 2.64$

- Para μ=4 $\nabla_x^2 \pi(x(\mu),\mu) = \begin{pmatrix} 10.127 & 8.127 \\ 8.127 & 10.127 \end{pmatrix}$ $cond(\nabla^2 \pi) = \lambda_1/\lambda_2 = 18.25/2 = 9.125$

- Para μ=0.16 $\nabla_x^2 \pi(x(\mu),\mu) = \begin{pmatrix} 160.24 & 158.24 \\ 158.24 & 160.24 \end{pmatrix}$ $cond(\nabla^2 \pi) = \lambda_1/\lambda_2 = 318.48/2 = 159.24$

Método del lagrangiano aumentado (i)

- El mal condicionamiento de los métodos anteriores puede mejorarse incluyendo los multiplicadores en la función de penalización.
- Los algoritmos ahora deben actualizar tanto las variables como los multiplicadores.
- La convergencia es más rápida que en los métodos anteriores.

Método del lagrangiano aumentado (ii)

- Sea el problema

$$\min_x f(x)$$
$$g_i(x) = 0 \quad i = 1,\ldots,m$$

- El óptimo del lagrangiano coincide con el del problema anterior si el punto es factible

$$\min_x L(x) = f(x) + \sum_{i=1}^{m} \lambda g_i(x)$$
$$g_i(x) = 0 \quad i = 1,\ldots,m$$

- Resolvemos este problema por el método de penalización exterior

$$\min_x A(x) = f(x) + \sum_{i=1}^{m} \lambda_i^k g_i(x) + \frac{1}{2}\rho_k \sum_{i=1}^{m} [g_i(x)]^2$$

- De forma general

$$\min A(x,\lambda,\rho) = f(x) + \lambda^T g(x) + \frac{1}{2}\rho g(x)^T g(x)$$

Método del lagrangiano aumentado (iii)

- Elegir valores de x_0, λ_0 y ρ_0
- Prueba de optimalidad. Si se verifica se detiene el algoritmo.

$$\nabla L(x_k, \lambda_k) = 0$$

- Resolver el problema no lineal sin restricciones y calcular x_{k+1}

$$\min_x \; A(x, \lambda_k, \rho_k) = f(x) + \lambda_k^T g(x) + \frac{1}{2}\rho_k g(x)^T g(x)$$

- Actualizar λ_{k+1} y ρ_{k+1}

$$\lambda_{k+1} = \lambda_k + \rho_k g(x_{k+1})$$

ρ_{k+1} ha de ser mayor que ρ_k

Programación cuadrática secuencial (i)

- Sea el problema
$$\min_x f(x)$$
$$g_i(x) = 0 \quad i = 1,\dots,m$$

- Formulamos el lagrangiano
$$\min_x L(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

- Condiciones de optimalidad de primer orden
$$\nabla L(x^*,\lambda^*) = 0 \quad \Rightarrow \quad \begin{cases} \nabla_x L(x,\lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \\ \nabla_\lambda L(x,\lambda) = g_i(x) = 0 \end{cases}$$

- Se formula el método de Newton para este sistema de ecs.
$$\begin{pmatrix} x_{k+1} \\ \lambda_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \lambda_k \end{pmatrix} + \begin{pmatrix} p_k \\ v_k \end{pmatrix}$$

donde p_k y v_k se determinan resolviendo el siguiente sistema de ecuaciones lineales

Programación cuadrática secuencial (ii)

- Sistema de ecuaciones lineales

$$\nabla^2 L(x_k,\lambda_k)\begin{pmatrix} p_k \\ v_k \end{pmatrix} = -\nabla L(x_k,\lambda_k)$$

$$\begin{pmatrix} \nabla^2_{xx} L(x_k,\lambda_k) & \nabla g(x_k) \\ \nabla g(x_k)^T & 0 \end{pmatrix}\begin{pmatrix} p_k \\ v_k \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x_k,\lambda_k) \\ g(x_k) \end{pmatrix}$$

- Representa las condiciones de optimalidad de primer orden de este problema de optimización cuadrática con restricciones lineales

$$\min_p \frac{1}{2} p^T \left[\nabla^2_{xx} L(x_k,\lambda_k)\right] p + p^T \left[\nabla_x L(x_k,\lambda_k)\right]$$

$$[\nabla g(x_k)]^T p + g(x_k) = 0 \qquad : v_k$$

- En lugar de resolver un sistema de ecs lineales se resuelve este problema de optimización cuadrática con restricciones lineales

Programación cuadrática secuencial (iii)

1. Se resuelve el problema cuadrático

$$\min_p \frac{1}{2} p^T \left[\nabla^2_{xx} L(x_k, \lambda_k)\right] p + p^T \left[\nabla_x L(x_k, \lambda_k)\right]$$

$$[\nabla g(x_k)]^T p + g(x_k) = 0 \qquad : v_k$$

2. Se obtienen (p_k, v_k)
3. Se actualizan los valores de las variables y multiplicadores

$$\begin{pmatrix} x_{k+1} \\ \lambda_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \lambda_k \end{pmatrix} + \begin{pmatrix} p_k \\ v_k \end{pmatrix}$$

Función reducida (i)

- Sea el problema $\begin{array}{l} \min\limits_{x} \quad f(x) \\ Ax = b, \quad x \in \mathbb{R}^n \end{array}$
- Supongamos un punto factible $\bar{x}$ y una dirección factible *p*, *Ap*=0, entonces cualquier punto factible se puede expresar como $x = \bar{x} + p$.

$$\begin{array}{l} \min\limits_{p} \quad f(\bar{x} + p) \\ Ap = 0, \quad p \in \mathbb{R}^n \end{array}$$

Función reducida (ii)

- Si Z es una matriz $n \times r$ ($r \geq n - m$) del espacio nulo de A (espacio definido por el conjunto de vectores ortogonales a las filas de A) la región factible también se puede expresar como

  $x = \bar{x} + Zv$, donde $v \in \mathbb{R}^r$

- Ahora el problema se transforma en una minimización de la función $\phi(v)$ que es la función reducida de f en la región factible, problema sin restricciones y de menores dimensiones

$$\min_v \quad \phi(v) = f(\bar{x} + Zv)$$
$$v \in \mathbb{R}^r$$

Cálculo de Z (método de reducción de variables)

- Sean las matrices

$$A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{m\times m}, N \in \mathbb{R}^{m\times(n-m)}, Z \in \mathbb{R}^{n\times(n-m)}$$

- Las direcciones factibles han de cumplir

$$Ap = 0 = \begin{pmatrix} B & N \end{pmatrix} \begin{pmatrix} p_B \\ p_N \end{pmatrix} = Bp_B + Np_N = 0$$

$$p = \begin{pmatrix} p_B \\ p_N \end{pmatrix} = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} p_N$$

$$Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}$$

- El nuevo punto se obtiene como

$$x = \bar{x} + p = \bar{x} + Zp_N = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix} + \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} p_N$$

Condiciones de optimalidad NLP con restricciones lineales de igualdad

- Gradiente reducido (proyectado) de f en x $\quad \nabla\phi(v) = Z^T \nabla f(x)$
- Hessiano reducido (proyectado) de f en x $\quad \nabla^2\phi(v) = Z^T \nabla^2 f(x) Z$

- Condiciones necesarias de primer y segundo orden

$$Z^T \nabla f(x^*) = 0$$
$$p^T \nabla^2 f(x^*) p \geq 0$$

siendo $p=Zv$ un vector del espacio nulo de A

- Condiciones suficientes de primer y segundo orden para mínimo local estricto

$$Ax^* = b$$
$$Z^T \nabla f(x^*) = 0$$
$$Z^T \nabla^2 f(x^*) Z \quad \text{matriz definida positiva}$$

Ejemplo

- Sea el problema

$$\min f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3$$
$$x_1 - x_2 + 2x_3 = 2$$

$$\nabla f(x) = \begin{pmatrix} 2x_1 - 2 \\ 2x_2 \\ -2x_3 + 4 \end{pmatrix}, \nabla^2 f(x) = \begin{pmatrix} 2 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & -2 \end{pmatrix}$$

$$x^* = \begin{pmatrix} 2.5 \\ -1.5 \\ -1 \end{pmatrix} \quad B = (1) \quad N = (-1 \quad 2) \quad Z = \begin{pmatrix} 1 & -2 \\ 1 & \cdot \\ \cdot & 1 \end{pmatrix}$$

$$Z^T \nabla f(x^*) = \begin{pmatrix} 1 & 1 & \cdot \\ -2 & \cdot & 1 \end{pmatrix} \begin{pmatrix} 3 \\ -3 \\ 6 \end{pmatrix} = \begin{pmatrix} \cdot \\ \cdot \end{pmatrix}, Z^T \nabla^2 f(x^*) Z = \begin{pmatrix} 1 & 1 & \cdot \\ -2 & \cdot & 1 \end{pmatrix} \begin{pmatrix} 2 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & -2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & \cdot \\ \cdot & 1 \end{pmatrix} = \begin{pmatrix} 4 & -4 \\ -4 & 6 \end{pmatrix}$$

- El hessiano de la f.o. no es una matriz definida positiva y, sin embargo, el hessiano reducido sí lo es. Luego el punto es un mínimo local estricto.

Método de Newton reducido

- Sea el problema $\begin{array}{l} \min_x \; f(x) \\ Ax = b, \quad x \in \mathbb{R}^n \end{array}$

  $x = \bar{x} + p$ siendo $\bar{x}$ punto factible inicial y p dirección factible

  $x = \bar{x} + Zv$ donde Z es una matriz n x n-m del espacio nulo de A

- La función reducida es $\begin{array}{l} \min_v \; \phi(v) = f(\bar{x} + Zv) \\ v \in \mathbb{R}^r \end{array}$

- Se aplica el método de Newton para calcular la dirección v_k

  $$\left[Z^T \nabla^2 f(x_k) Z\right]^T v = -Z^T \nabla f(x_k)$$

  resolviendo este sistema de ecuaciones lineales

- p_k se obtiene a partir de v_k , $p = Zv$. No depende matemáticamente de Z pero sí numéricamente.

  $$p_k = Zv_k = -Z\left[Z^T \nabla^2 f(x_k) Z\right]^{-T} Z^T \nabla f(x_k)$$

Método del gradiente conjugado lineal (i)

- Sea el sistema de ecuaciones

  $Ax = b$

  donde A es una matriz simétrica y definida positiva

- Resolver el sistema es equivalente a resolver

  $\min f(x) = \frac{1}{2} x^T A x - b^T x$

  ya que por condiciones de optimalidad

  $\nabla f(x) = Ax - b = 0$

  y el hessiano es definido positivo

  $\nabla^2 f(x) = A$

- Vectores conjugados

  $p_i^T A p_j = 0 \quad \forall i \neq j$

  Si $A=I$ entonces los vectores conjugados son ortogonales.

Método del gradiente conjugado lineal (ii)

- Supongamos que tenemos un punto y combinación lineal de m+1 vectores conjugados

$$y=\sum_{i=0}^{m}\alpha_i p_i$$

- Evaluando la función

$$f(y)=f\left(\sum_{i=0}^{m}\alpha_i p_i\right)=\frac{1}{2}\left(\sum_{i=0}^{m}\alpha_i p_i\right)^T A\left(\sum_{j=0}^{m}\alpha_j p_j\right)-b^T\left(\sum_{i=0}^{m}\alpha_i p_i\right)=$$

$$=\frac{1}{2}\sum_{i=0}^{m}\sum_{j=0}^{m}\alpha_i\alpha_j p_i^T A p_j-\sum_{i=0}^{m}\alpha_i b^T p_i$$

$$=\frac{1}{2}\sum_{i=0}^{m}\alpha_i^2 p_i^T A p_i-\sum_{i=0}^{m}\alpha_i b^T p_i$$

Desaparecen los productos p_iAp_j

$$=\sum_{i=0}^{m}\left(\frac{1}{2}\alpha_i^2 p_i^T A p_i-\alpha_i b^T p_i\right)$$

Método del gradiente conjugado lineal (iii)

- Minimizando la función

$$\min_y f(y) = \min_{\alpha_i} f\left(\sum_{i=0}^{m} \alpha_i p_i\right) = \sum_{i=0}^{m} \min_{\alpha_i} \left(\frac{1}{2}\alpha_i^2 p_i^T A p_i - \alpha_i b^T p_i\right)$$

- El problema de minimización multidimensional se ha transformado en $m+1$ problemas de minimización unidimensionales
- Para cada dimensión

$$\alpha_i \left(p_i^T A p_i\right) - b^T p_i = 0 \qquad \alpha_i = \frac{b^T p_i}{p_i^T A p_i}$$

- Es decir, si representamos un punto como combinación lineal de un conjunto de vectores conjugados la solución es inmediata

Método del gradiente conjugado lineal (iv)

- Determina iterativamente el conjunto de vectores conjugados y los pesos
- Método

$$x_0 = 0, \quad r_0 = b = b - Ax_0, \quad p_{-1} = 0, \quad \beta_0 = 0$$

Para $i = 0,1,\ldots$

Si $\|r_i\| < \varepsilon$ fin.

$$\text{Si } i > 0 \rightarrow \beta_i = \frac{r_i^T r_i}{r_{i-1}^T r_{i-1}}$$

$$p_i = r_i + \beta_i p_{i-1}$$ Dirección de movimiento

$$\alpha_i = \frac{r_i^T r_i}{p_i^T A p_i}$$ Longitud de paso

$$x_{i+1} = x_i + \alpha_i p_i$$ Actualización del punto

$$r_{i+1} = r_i - \alpha_i A p_i$$ Actualización del residuo

x es la solución, p son vectores conjugados, r son vectores ortogonales

Método del gradiente conjugado lineal (v)

- Ejemplo

$$A=\begin{pmatrix}1 & \cdot & \cdot\\ \cdot & 2 & \cdot\\ \cdot & \cdot & 3\end{pmatrix}\quad b=\begin{pmatrix}1\\1\\1\end{pmatrix}\qquad x_0=\begin{pmatrix}\cdot\\ \cdot\\ \cdot\end{pmatrix},\quad r_0=\begin{pmatrix}1\\1\\1\end{pmatrix},\quad p_{-1}=\begin{pmatrix}\cdot\\ \cdot\\ \cdot\end{pmatrix},\quad \beta_0=0,\quad \varepsilon=10^{-12}$$

$$\|r_0\|=1.73>\varepsilon,\qquad p_0=\begin{pmatrix}1\\1\\1\end{pmatrix},\quad \alpha_0=0.5,\quad x_1=\begin{pmatrix}0.5\\0.5\\0.5\end{pmatrix},\quad r_1=\begin{pmatrix}0.5\\ \cdot\\ -0.5\end{pmatrix}$$

$$\|r_1\|=0.71>\varepsilon,\qquad p_1=\begin{pmatrix}0.667\\0.167\\-0.333\end{pmatrix},\quad \alpha_1=0.6,\quad x_2=\begin{pmatrix}0.9\\0.6\\0.3\end{pmatrix},\quad r_2=\begin{pmatrix}0.1\\-0.2\\0.1\end{pmatrix}$$

$$\|r_2\|=0.245>\varepsilon,\qquad p_2=\begin{pmatrix}0.18\\-0.18\\0.06\end{pmatrix},\quad \alpha_2=0.556,\quad x_3=\begin{pmatrix}1\\0.5\\0.33\end{pmatrix},\quad r_3=\begin{pmatrix}\cdot\\ \cdot\\ \cdot\end{pmatrix}$$

Método del gradiente conjugado lineal (vi)

- Adecuado para la resolución de problemas de gran tamaño
  - ✓ Requieren poco almacenamiento
  - ✓ Poco cálculo por iteración
- Convergencia lineal con este ratio

$$\frac{\left\| x_{i+1} - x^* \right\|_A}{\left\| x_i - x^* \right\|_A} \le \frac{\sqrt{cond(A)} - 1}{\sqrt{cond(A)} + 1}$$

$$\left\| y \right\|_A^2 \equiv \frac{1}{2} y^T A y$$

Si $cond(A) = 1 \rightarrow$ ratio $= 0 \rightarrow$ converge en una iteración

Si $cond(A) = 100 \rightarrow$ ratio $= 0.82$

Si $cond(A) = 1000000 \rightarrow$ ratio $= 0.998$

Método del gradiente conjugado no lineal (i)

- Sólo requiere el uso de primeras derivadas

$$\min f(x)$$

$$x_0 = 0, \quad p_{-1} = 0, \quad \beta_0 = 0$$

Para $i = 0,1,\ldots$

Si $\left\|\nabla f(x_i)\right\| < \varepsilon$ fin.

Si $i > 0 \rightarrow \beta_i = \dfrac{\nabla f(x_i)^T \nabla f(x_i)}{\nabla f(x_{i-1})^T \nabla f(x_{i-1})}$

$$p_i = -\nabla f(x_i) + \beta_i p_{i-1}$$

Determinar α_i para minimizar $f(x_i + \alpha_i p_i)$

$$x_{i+1} = x_i + \alpha_i p_i$$

Dirección de movimiento

Longitud de paso

Actualización del punto

Método del gradiente conjugado no lineal (i)

- Ejemplo $\min f(x) = \frac{1}{10}(x-e)^T D(x-e) + \left(x^T x - \frac{1}{4}\right)^2$

$$e = \begin{pmatrix}1 & 1 & \cdots & 1\end{pmatrix}^T$$

$$D = \begin{pmatrix}1 & & & \\ & 2 & & \\ & & \ddots & \\ & & & n\end{pmatrix}$$

$$x_0 = \begin{pmatrix}1 & -1 & 1 & -1\end{pmatrix}^T$$

$$f(x_0) = 16.46$$

$$\nabla f(x_0) = \begin{pmatrix}15 & -15.8 & 15 & -16.6\end{pmatrix}^T$$

$$p = -\nabla f(x_0)$$

$$\alpha = 0.0625$$

$$x_1 = \begin{pmatrix}0.0625 & -0.0125 & 0.0625 & 0.0375\end{pmatrix}^T$$

$$f(x_0) = 0.985$$

$$\nabla f(x_0) = \begin{pmatrix}-0.248 & -0.393 & -0.623 & -0.806\end{pmatrix}^T$$

$$p = -\nabla f(x_0)$$

$$\alpha = 0.5$$

$$x_1 = \begin{pmatrix}0.177 & 0.194 & 0.364 & 0.451\end{pmatrix}^T$$

Precondicionamiento

- Idea básica
  - ✓ Sacar partido de información auxiliar sobre el problema para acelerar la convergencia
- Supongamos que conocemos la matriz M definida positiva
- En lugar de resolver $Ax = b$ resolvemos $M^{-1}Ax = M^{-1}b$

$$\begin{pmatrix} 2000 & 1000 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad cond(A) = 1700$$

$$M^{-1} = \begin{pmatrix} 1000 & \cdot \\ \cdot & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.001 \\ 1 \end{pmatrix} \qquad cond(M^{-1}A) = 3$$

- El objetivo es encontrar M de manera que $M \approx A$

Andrés Ramos

http://www.iit.comillas.edu/aramos/

Andres.Ramos@comillas.edu