COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI ICADE CIHS

comillas.edu

comillas.edu

Nonlinear Optimization

Andrés Ramos (Andres.Ramos@comillas.edu)
Pedro Sánchez (Pedro.Sanchez@comillas.edu)
Sonja Wogrin (Sonja.Wogrin@comillas.edu)

Departamento de Organización Industrial

Contents

1

TYPE OF NLP PROBLEMS AND SOLUTIONS

Nonlinear optimization problems (i)

- Optimization WITHOUT constraints

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

- Optimization WITH constraints (Nonlinear programming NLP)

$$\min_x f(x)$$
$$g_i(x) = 0 \quad i \in \varepsilon$$
$$g_i(x) \leq 0 \quad i \in \varphi$$
$$f(x): \mathbb{R}^n \to \mathbb{R}$$
$$g_i(x): \mathbb{R}^n \to \mathbb{R}$$
$$x \in \mathbb{R}^n$$

Nonlinear optimization problems (ii)

- Quadratic Programming

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

- Convex Programming

  $f(x)$ is convex (concave if we are maximizing) and $g_i(x)$ is convex, $\forall i = 1, \dots, m$

- Separable Programming

  The function can be separated into a sum of functions of the individual variables

$$f(x) = \sum_{j=1}^{n} f_j(x_j)$$

- Geometric Programming

  Objective function and constraints take the form $g(x) = \sum_{j=1}^{n} c_j P_j(x)$

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

Type of NLP solutions (i)
Local, global minimum

- Let the function f be continuously differentiable for the first and second order

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

$x^* \in \mathbb{R}^n$ is the optimum of the function

  - It is a global minimum if $f(x^*) \leq f(x)$ for $x \in \mathbb{R}^n$
  - It is a strict global minimum if $f(x^*) < f(x)$ for $x \in \mathbb{R}^n$
  - It is a local minimum if $f(x^*) \leq f(x)$ in its vicinity $\|x - x^*\| < \varepsilon$
    where ε is a positive number (typically small) whose exact value can depend on x^*
  - It is a strict local minimum if $f(x^*) < f(x)$ in its vicinity $\|x - x^*\| < \varepsilon$

Type of NLP solutions (i)
Local, global minimum

- Let the function f be continuously differentiable for the first and second order

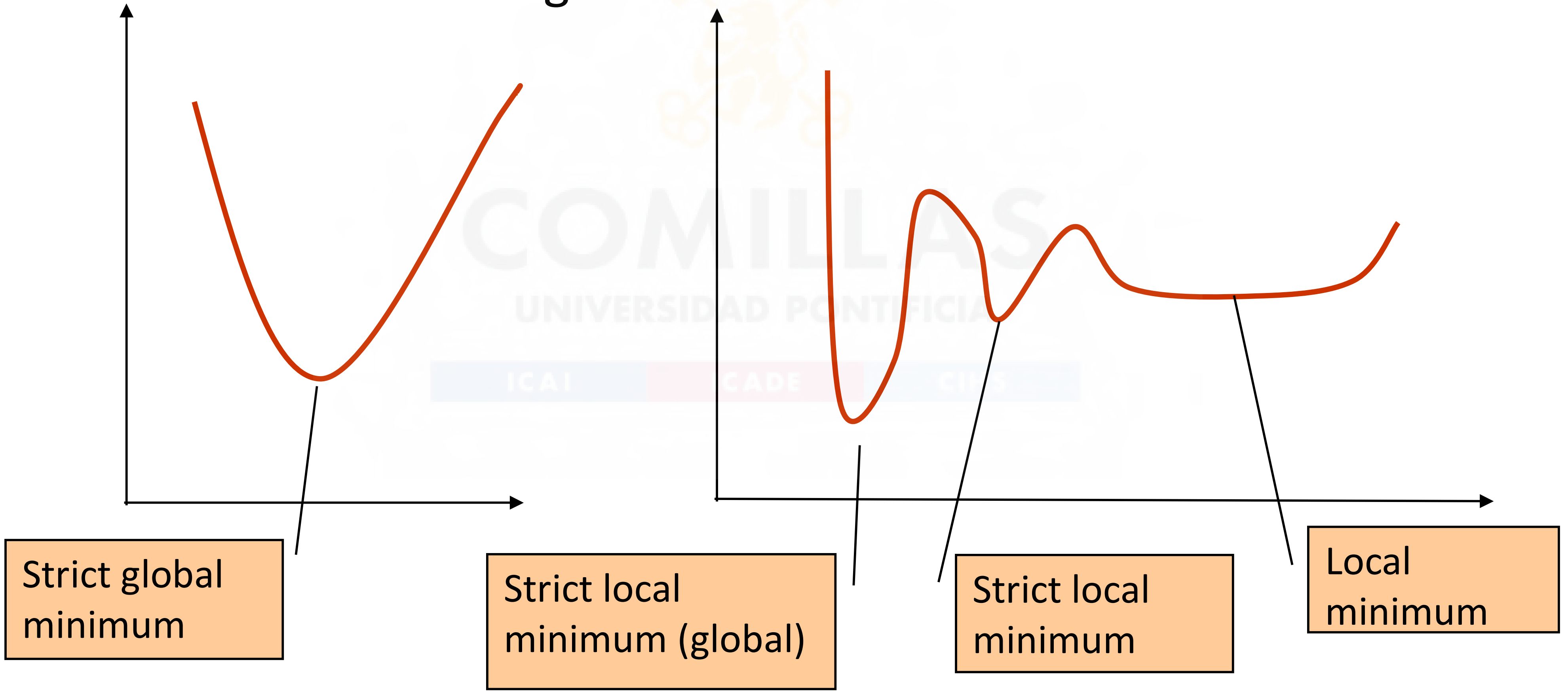
$$\min_x f(x)$$
$$f(x): \mathbb{R}^n \to \mathbb{R}$$
$$x \in \mathbb{R}^n$$

$x^* \in \mathbb{R}^n$ is the optimum of the function

- It is a global minimum if $f(x^*) \leq f(x)$ for $x \in \mathbb{R}^n$
- It is a strict global minimum if $f(x^*) < f(x)$ for $x \in \mathbb{R}^n$
- It is a local minimum if $f(x^*) \leq f(x)$ in its vicinity $\|x - x^*\| < \varepsilon$
  where ε is a positive number (typically small) whose exact value can depend on x^*
- It is a strict local minimum if $f(x^*) < f(x)$ in its vicinity $\|x - x^*\| < \varepsilon$

Type of NLP solutions (ii)
Local and global minima

- A global minimum is hard to find. Many methods are local.
- Only under additional assumptions (convexity) global minimum can be guaranteed.



Type of NLP solutions (iii)
Positive definite matrix

- A matrix A is positive definite if $x^T A x > 0$ for all non-zero vectors x
- Or rather, if all its eigenvalues are positive.
- Or rather, if all its leading principal minors (determinants of order 1,2,…,n (being n the matrix dimension) obtained by adding consecutive rows and columns starting from the first element) are positive.

This matrix is positive definite. $\nabla^2 f(x,y) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

This one is not. $\nabla^2 f(x_1,x_2) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

Type of NLP solutions (iv)
Positive semidefinite (psd) matrix

- A matrix A is positive semidefinite if $x^T A x \geq 0$ for all non-zero vectors x
- Or rather, if all its eigenvalues are nonnegative.
- Or rather, if all its first minors of order 1,2,…,n (being n the matrix dimension) are nonnegative.
- First minors are the determinants of k arbitrary rows and their corresponding columns.

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

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

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

- ❑ This matrix is positive semidefinite

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

Type of NLP solutions (v)
Negative definite matrix

- A matrix A is negative definite if $x^T A x < 0$ for all non-zero vectors x
- Or rather, if all its eigenvalues are negative.
- Or rather, if all its leading minors are, alternatively, negative, positive, etc.
- Or rather, if the matrix $-A$ (which is obtained by changing the sign of all the matrix elements) is positive definite.

Type of NLP solutions (vi)
Equivalent relations

Convex Function	Concave Function
Positive definite Hessian matrix	Negative definite Hessian matrix
Positive curvature (2^{nd} derivative)	Negative curvature (2^{nd} derivative)

Local minimum + convexity of the entire region	Local maximum + concavity of the entire region
Global Minimum	Global Maximum

Positive semidefinite (negative)	Positive definite (negative)
Local minimum (maximum)	Strict local minimum (maximum)

2

UNCONSTRAINED NONLINEAR OPTIMIZATION

UNCONSTRAINED optimization
Optimality conditions

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

- First-order necessary condition
  - If x^* is a local minimum of f, then necessarily $\nabla f(x^*) = 0$
  - That condition is satisfied for any stationary point
- Second-order necessary condition
  - If x^* is a local minimum of f, then necessarily $\nabla^2 f(x^*)$ is a positive semidefinite matrix (equivalent to convex, positive curvature)
- Second-order sufficient condition
  - If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite, then x^* is a strict local minimum of f
- Necessary and sufficient condition
  - Let be f convex and differentiable in x^* (if it is twice differentiable Hessian is positive semidefinite); x^* is a local minimum if and only if $\nabla f(x^*) = 0$. x^* is a global minimum if and only if it is convex in all the region

Example 1: Optimization WITHOUT constraints

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

- Gradient = 0 yields a system of equations

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

with solution $(x,y) = (2/3, -1/3)$

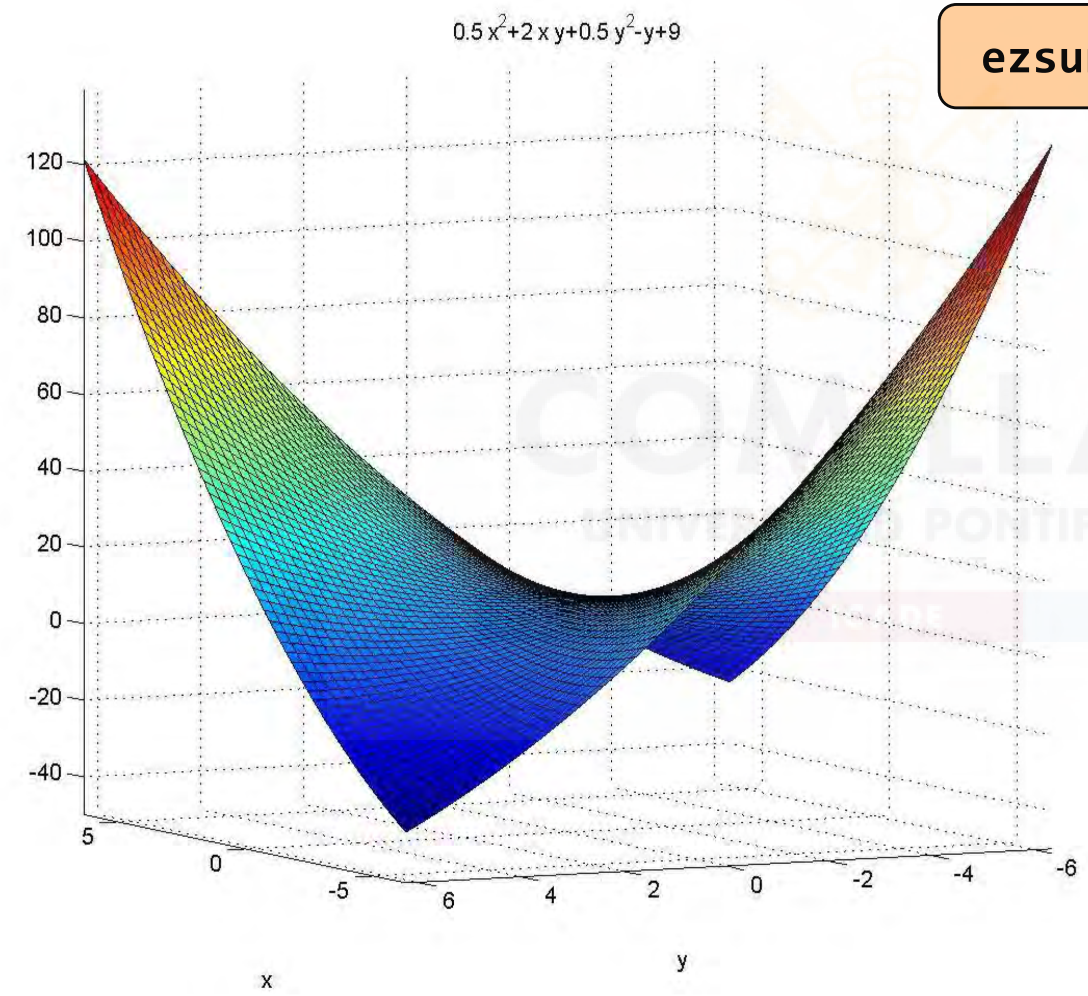
- Hessian $\nabla^2 f(x,y) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ is an indefinite matrix (it is not positive semidefinite nor negative semidefinite)

This means that it is neither a local maximum nor a minimum

Example 1: Optimization WITHOUT constraints

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

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

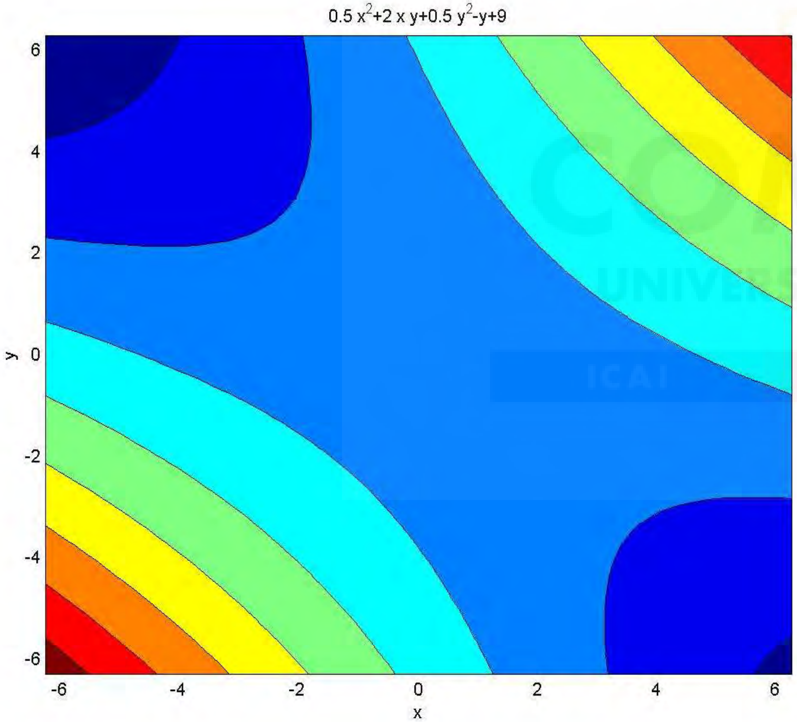


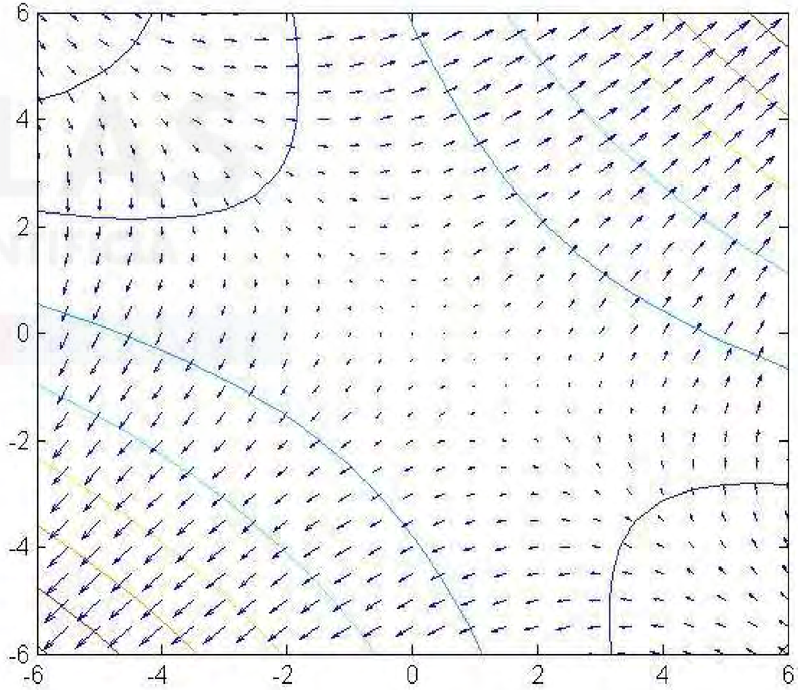
Example 1: Optimization WITHOUT constraints

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

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

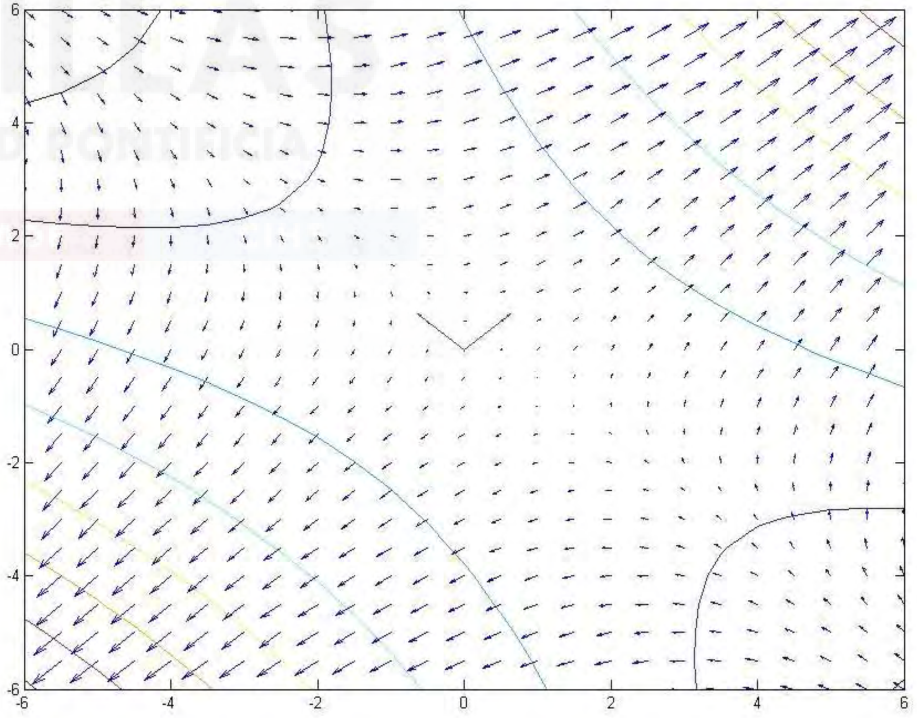
```
[x,y] = meshgrid(-6:0.5:6,-6:0.5:6);
z=0.5*x.^2+2*x.*y+0.5*y.^2-y+9;
[px,py] = gradient(z,0.5,0.5);
contour(x,y,z); hold on
quiver(x,y,px,py)
```





Example 1: Eigenvalues and eigenvectors

- Hessian $\nabla^2 f(x,y) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ is an indefinite matrix
- Its eigenvalues are $\lambda = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$
- Its eigenvectors are $v_1 = \begin{pmatrix} -0.71 \\ 0.71 \end{pmatrix}$ $v_2 = \begin{pmatrix} 0.71 \\ 0.71 \end{pmatrix}$



Example 2: Optimization WITHOUT constraints

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

- Gradient = 0 yields a system of equations

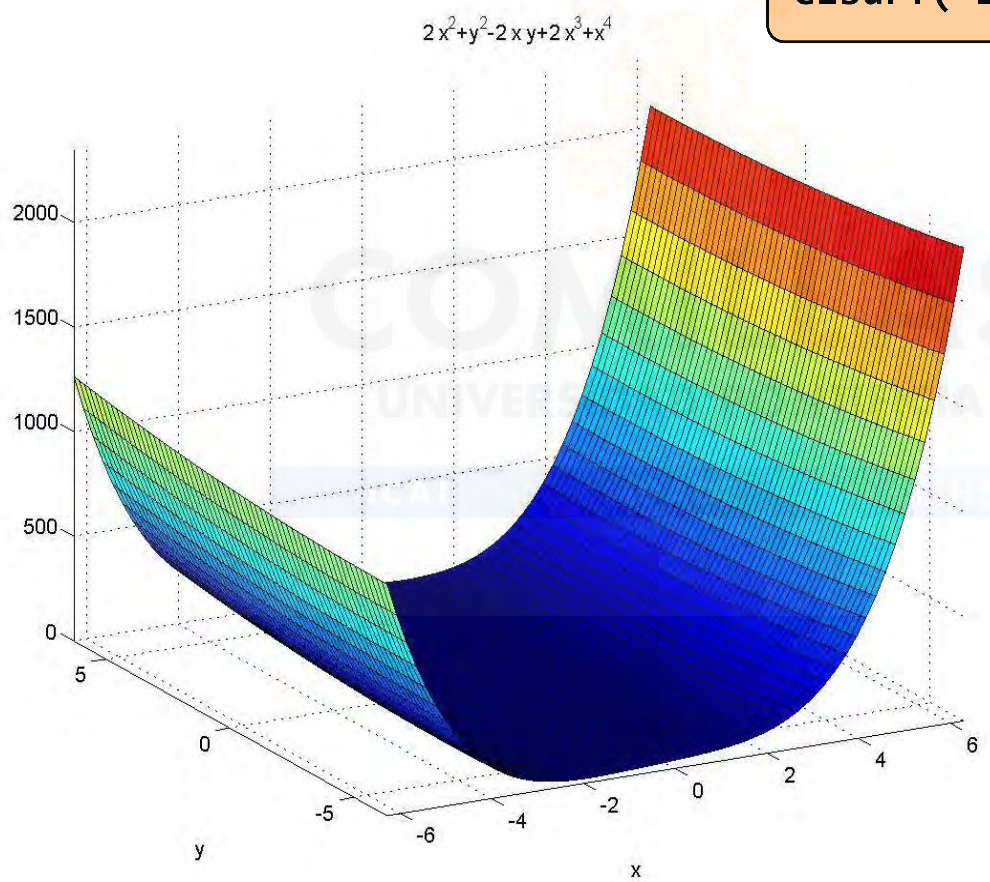
$$\nabla f(x,y) = \begin{pmatrix} 4x - 2y + 6x^2 + 4x^3 \\ 2y - 2x \end{pmatrix} = 0$$

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

Example 2: Optimization WITHOUT constraints

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

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

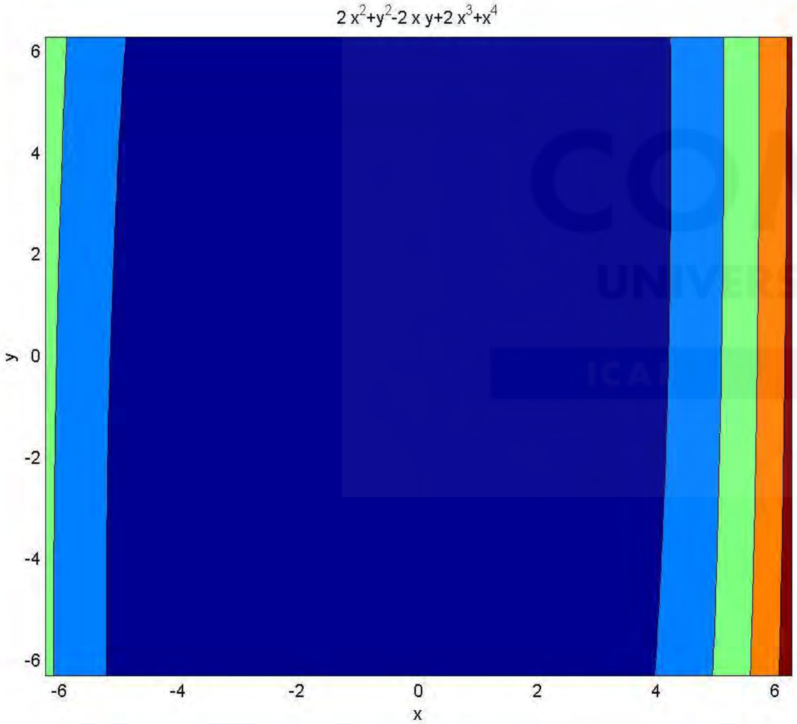


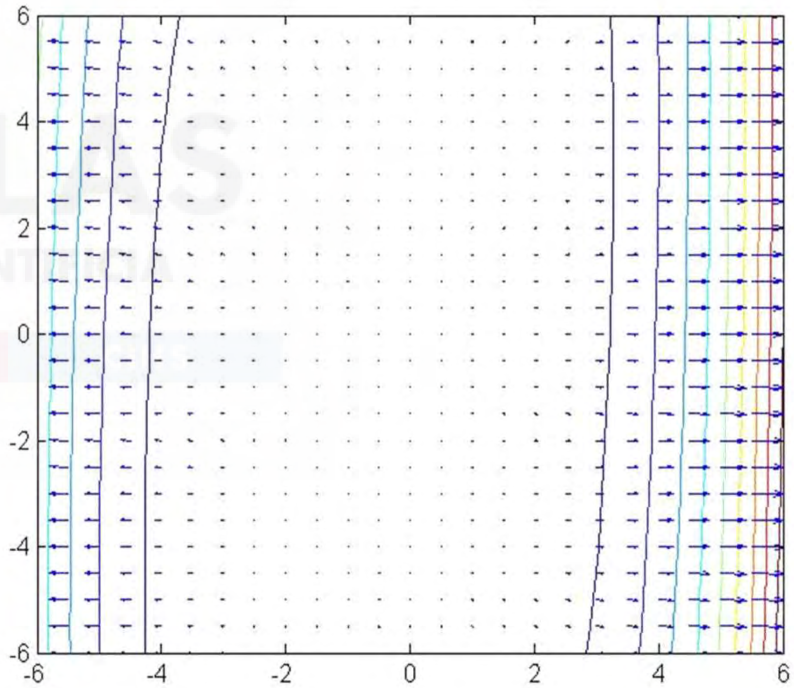
Example 2: Optimization WITHOUT constraints

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

```
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)
```





Example 3: Optimization WITHOUT constraints

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

- Gradient = 0 yields a system of equations

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

with solution $(x,y)=(2,1)$

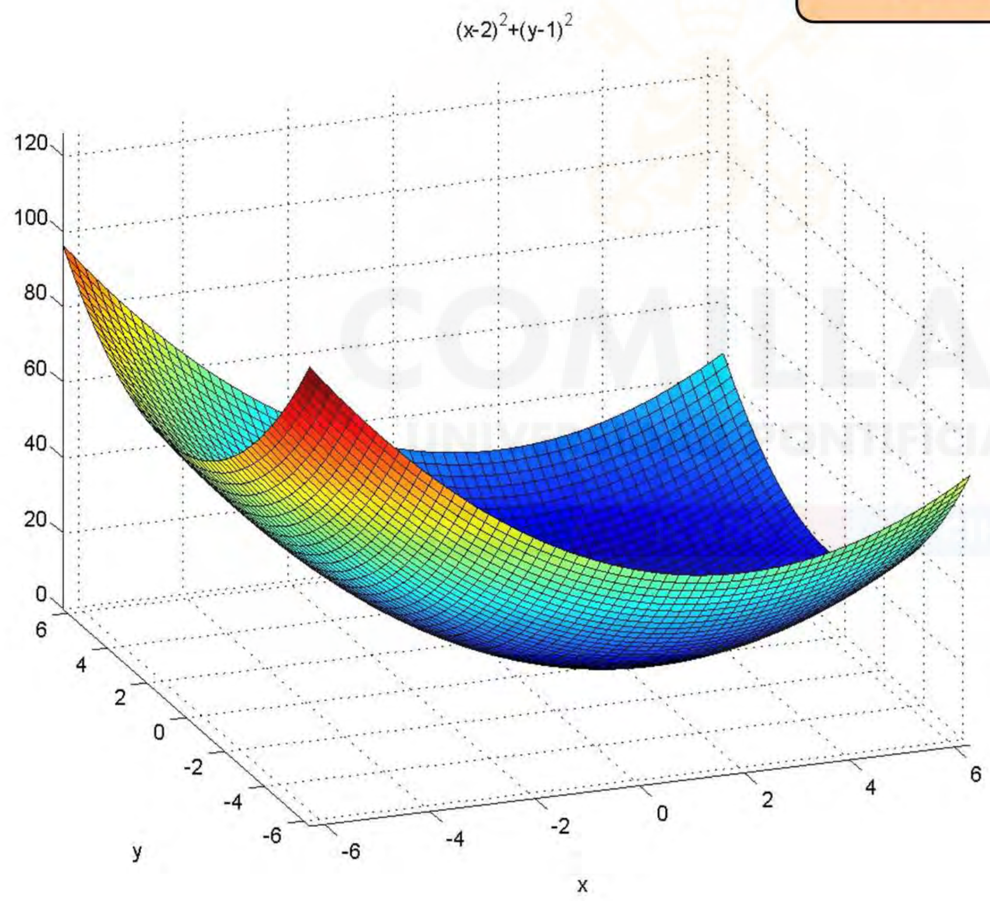
- Hessian $\nabla^2 f(x,y)=\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ is a positive definite matrix

and (independently of the point) this yields a global minimum

Example 3: Optimization WITHOUT constraints

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

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

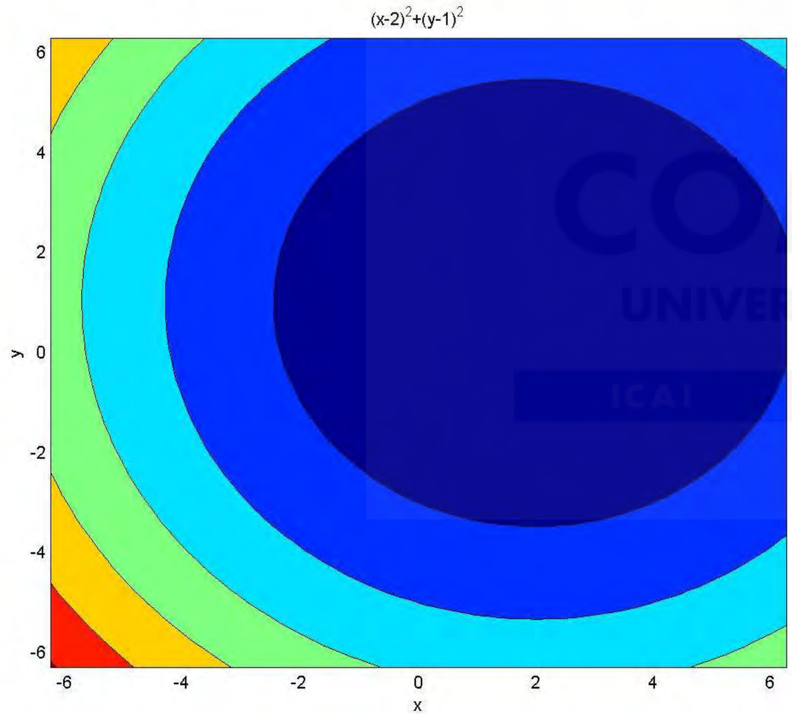


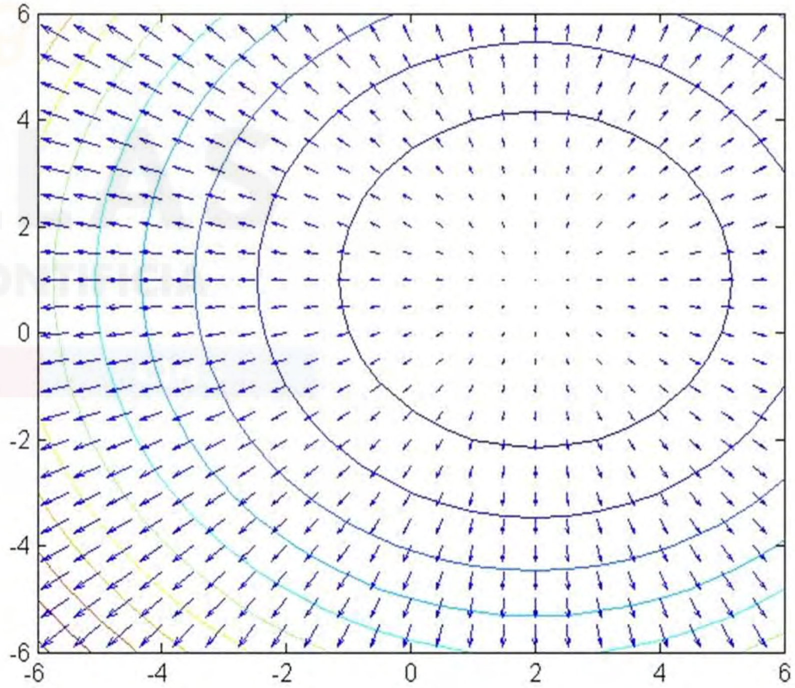
Example 3: Optimization WITHOUT constraints

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

```
ezcontourf('(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)
```





3

OPTIMALITY CONDITIONTS FOR NLP

Lagrangian (i)

- Let's see this optimization problem $\min_{x} f(x)$, $Ax = b$

  being $x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$
- We define the Lagrangian as

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

  being $\lambda \in \mathbb{R}^m$ the Lagrange multipliers.
- The Lagrangian problem is unconstrained.
- A constrained problem is transformed into an unconstrained one with m additional variables.
- The minimum of both problems coincides, given that $Ax = b$

Lagrangian (ii)

- First-order optimality conditions

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

and therefore, if x^* is a local minimum, it must satisfy $\nabla f(x^*)=-A^T\lambda^*$

- In a local minimum, the objective function gradient is a linear combination of the constraint gradients, and the Lagrange multipliers are the weights.
- Multipliers represent the change in the objective function for a marginal change in the constraint bound. In the case of LP, those are named *dual variables* or *shadow prices*. With this formulation, multipliers result in the opposite sign of the dual variables.

Lagrangian (iii)

- Let's see this optimization problem

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

  being $f: \mathbb{R}^n \to \mathbb{R}$, $g_i: \mathbb{R}^n \to \mathbb{R}$
- Let's define the Lagrangian as

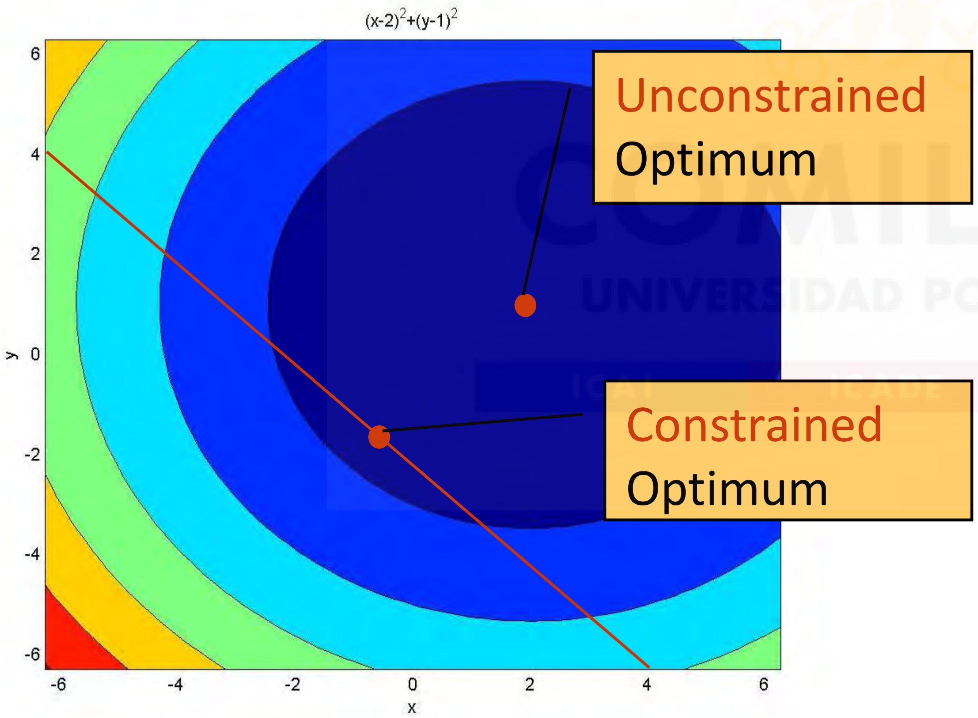
$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{l} \mu_j h_j(x)$$

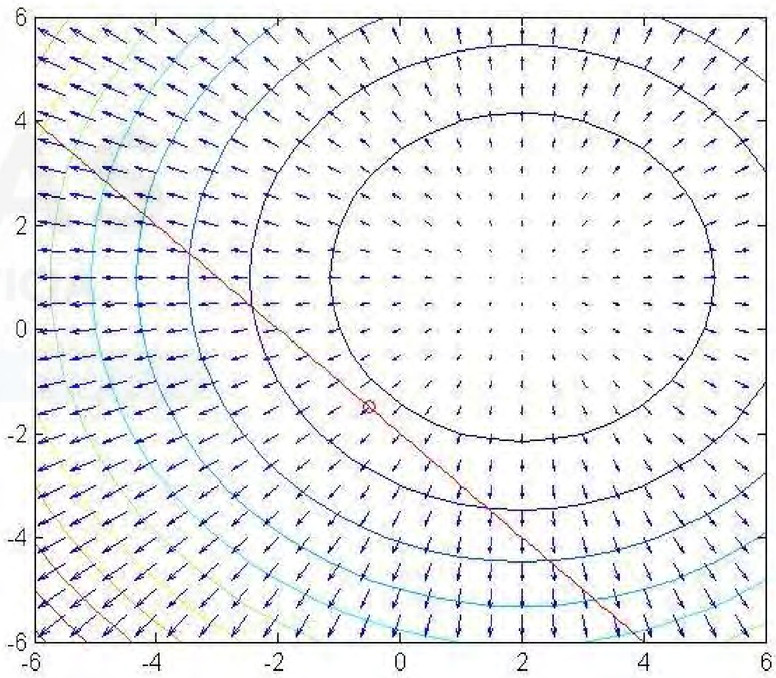
  where $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^l$ are the Lagrange multipliers.
- Lagrangian is always a lower bound of $f(x)$ for feasible values of x and known values of $\lambda \geq 0$ (nonnegative) and μ (free).

Example 1 (i)

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

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

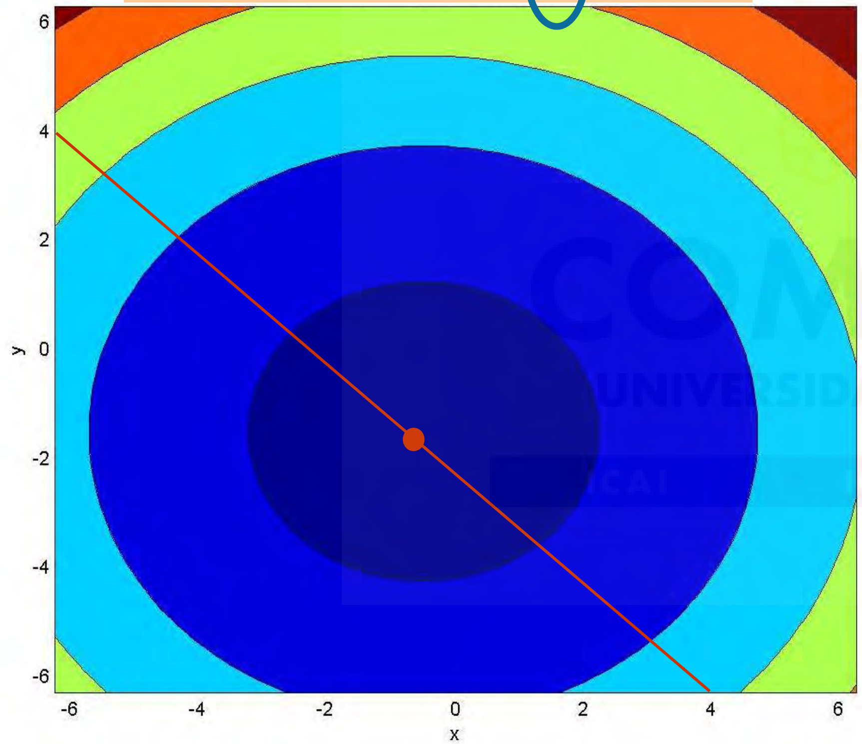




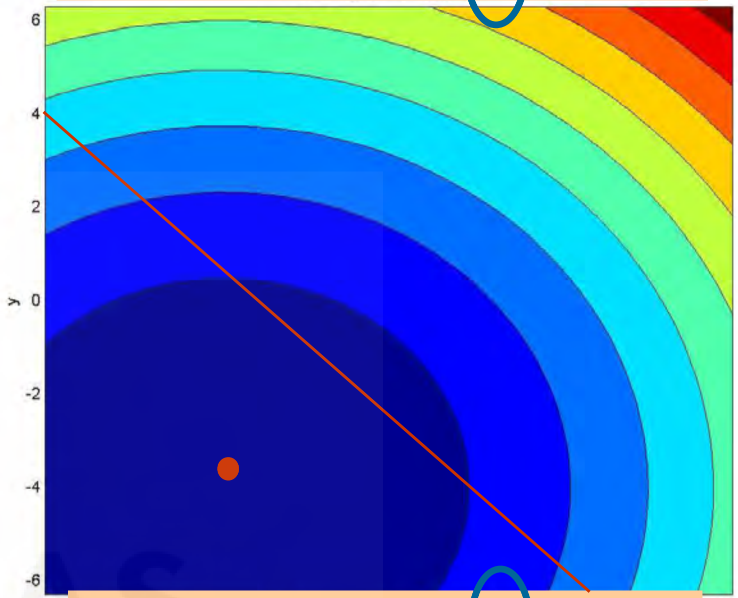
Example 1 (ii)

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

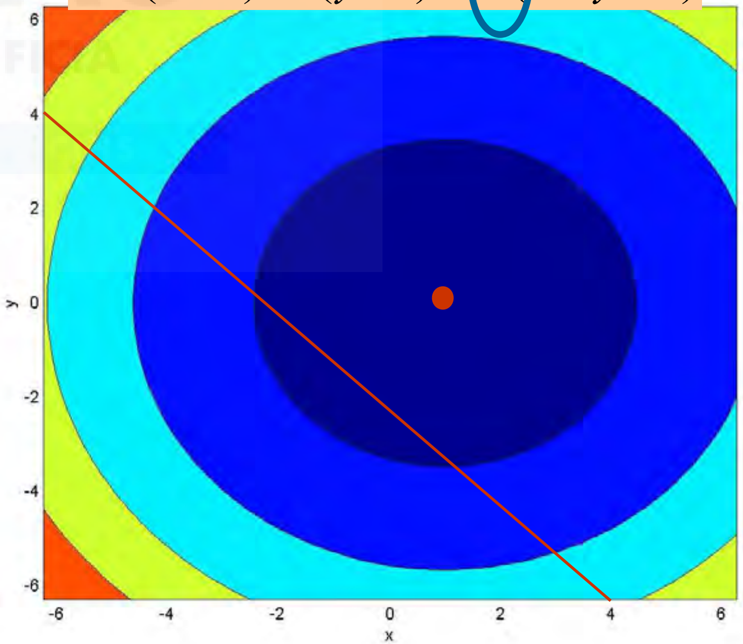
$$\min (x-2)^2 + (y-1)^2 + 5(x+y+2)$$



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

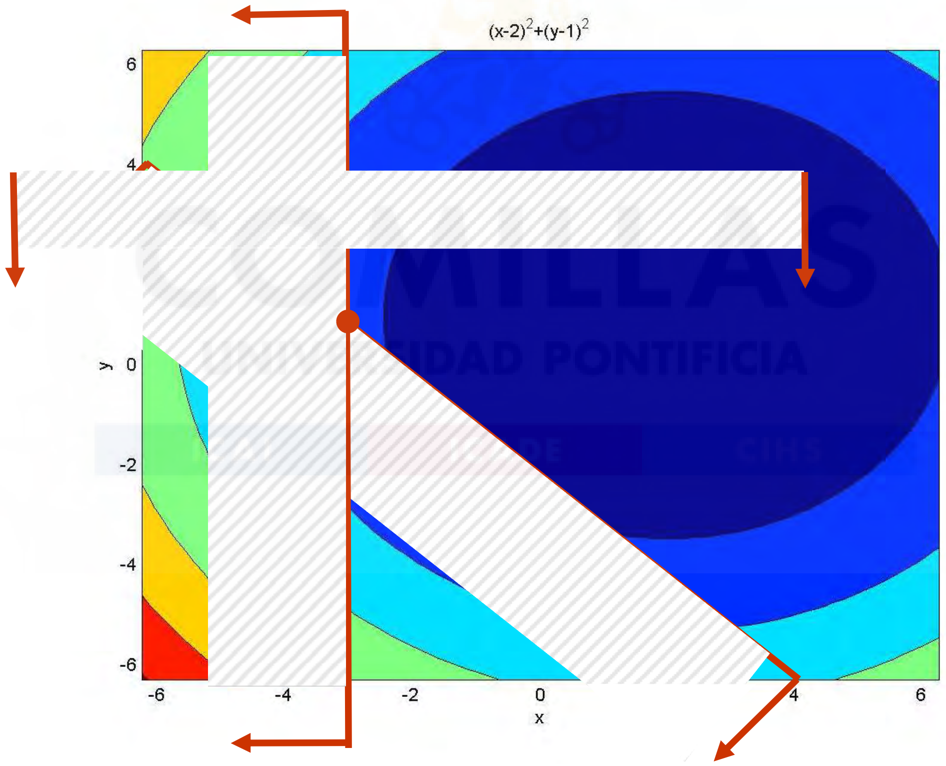


$$\min (x-2)^2 + (y-1)^2 + 2(x+y+2)$$



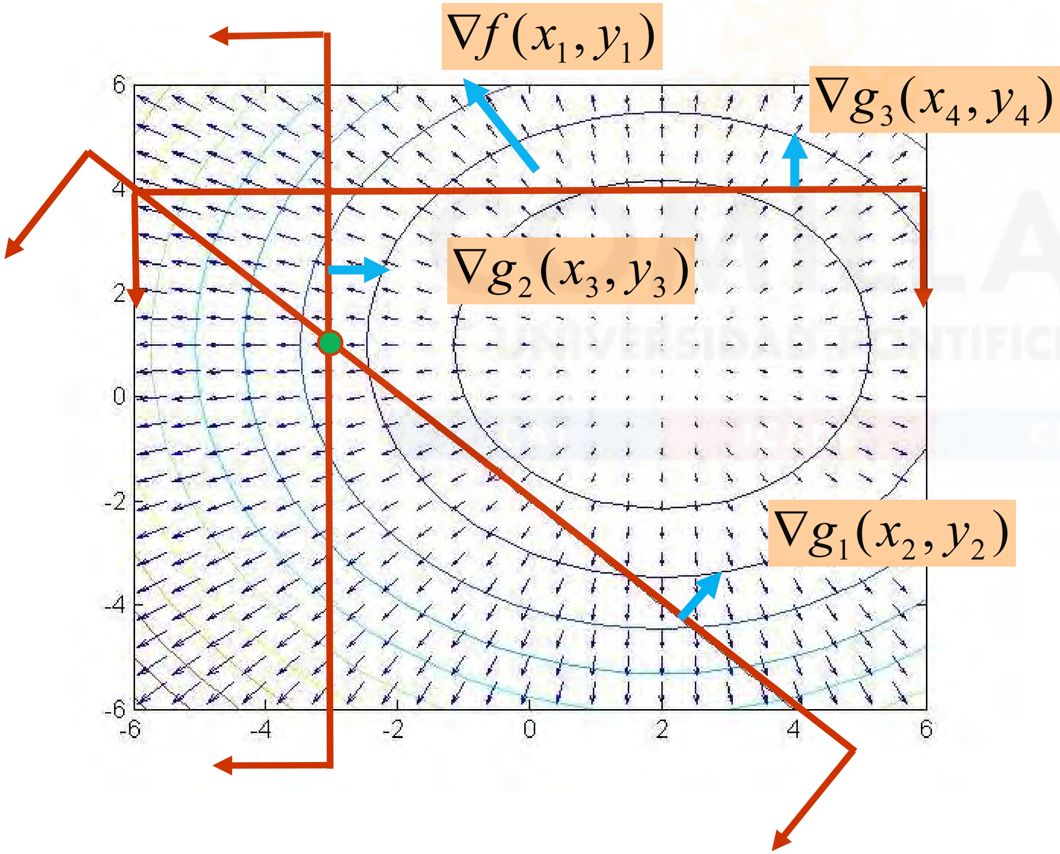
Example 4 (i)

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



Example 4 (ii)

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



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

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

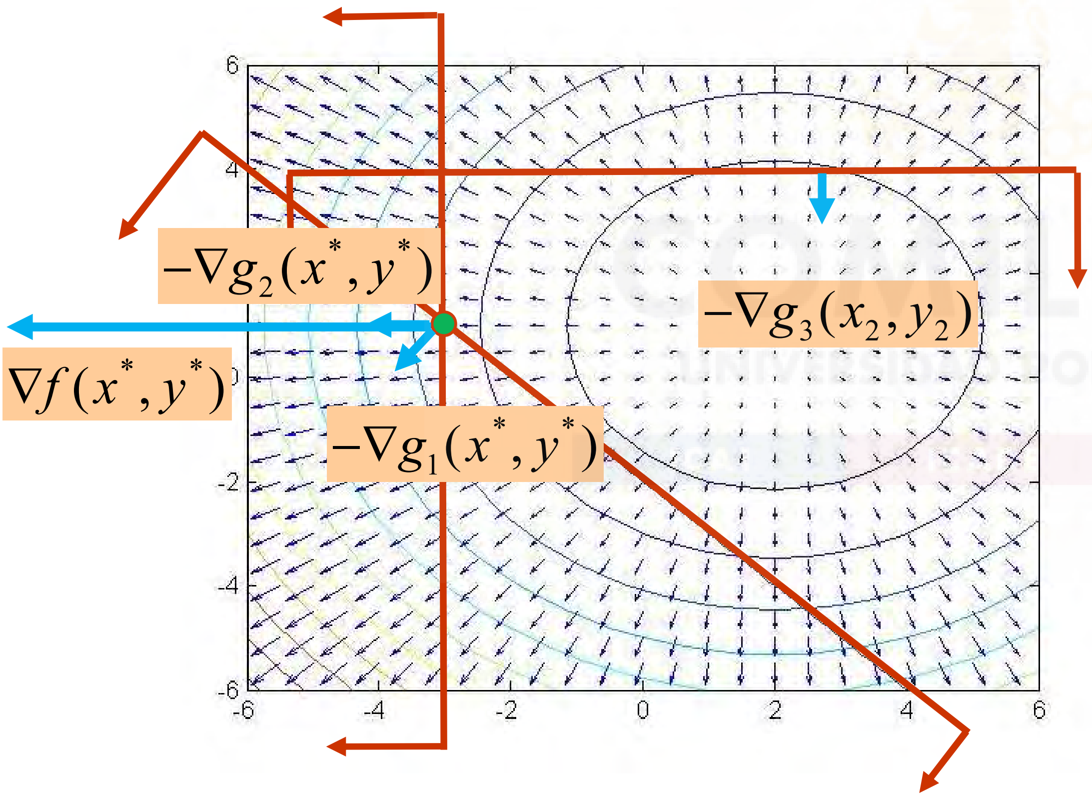
$$\nabla g_2(x,y) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

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

Example 4 (iii)

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

- In the optimum (-3,1), the gradient of the o.f. can be expressed as a linear combination of the gradients of the **binding constraints** changed in sign



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

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

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

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

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

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

The first constraint is superfluous. Degenerate solution.

Necessary conditions with inequality constraints (i)

- Let's see this optimization problem

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

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

- Let's be x^* a feasible point

$I = \{i / g_i(x^*) = 0\}$ the set of **binding constraints**

f and $\{g_i, i \in I\}$ differentiable in x^*

$\{g_i, i \notin I\}$ continuous in x^*

$\{\nabla g_i(x^*)\}_{i \in I}$ linearly independent

Necessary conditions with inequality constraints (ii)

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

- If x^* is a local minimum, then there exist a scalars $\{\lambda_i, i \in I\}$ such that

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

- Besides, if functions $\{g_i, i \notin I\}$ are differentiable in x^*, if x^* is a local optimum, then

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

Complementary Slackness Condition

- Nonbinding constraint → multiplier 0.
  Binding constraint → multiplier does not necessarily have to be 0.

Necessary conditions with inequality constraints (iii)

- Necessary conditions with inequality constraints

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

- First-order Karush-Kuhn-Tucker (KKT) necessary conditions for a local optimum

Gradient of the o.f. is a linear combination of the gradients if the constraints changed in sign

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

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

Complementary Slackness Condition
Nonbinding constraint $\lambda=0$
Binding constraint $\lambda \neq 0$

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

Feasible point

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

Sufficient conditions with inequality constraints (iv)

- Second order sufficient conditions for a local minimum
- Hessian matrix $\nabla^2 f(x^*) + \sum_{i \in I} \lambda_i^* \nabla^2 g_i(x^*)$ must be positive definite

or alternatively $\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(x^*)$

Necessary conditions with inequality constraints (v)

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

- The Lagrangian is
$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$
$$\lambda_i \ge 0$$

- Optimality condition for the Lagrangian
$$\nabla L(x^*,\lambda^*) = 0 \quad \Rightarrow \begin{cases} \nabla_x L(x^*,\lambda^*) = \nabla f(x^*) + \sum_{i \in I} \lambda_i^* \nabla g_i(x^*) = 0 \\ \nabla_\lambda L(x^*,\lambda^*) = g_i(x^*) = 0 \quad \forall i \in I \end{cases}$$

  the 2nd one corresponds to the definition of binding constraints $\forall i \in I$

- To consider all the constraints is expressed as
$$\lambda_i^* g_i(x^*) = 0 \quad i = 1,\ldots,m$$

Sufficient conditions with inequality constraints (i)

- If the o.f. is convex or the feasible region is convex there still can be points that satisfy the necessary conditions

- Let x^* be a feasible point

  $I = \{ i / g_i(x^*) = 0 \}$ is the set of binding constraints

  f and $\{g_i, i \in I\}$ are convex and differentiable in **all the feasible region**

- If there exist scalars $\{\lambda_i, i \in I\}$ such that

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

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

then x^* is a **global minimum**

Sufficient conditions with inequality constraints (ii)

- Condition for a strict local minimum

  Alternatively, instead of the conditions that f and $\{g_i, i \in I\}$ are convex and differentiable in x^*, it can also be expressed as the condition that the Lagrangian $L(x) = f(x) + \sum_{i\in I} \lambda_i^* g_i(x)$,

  where λ_i^* represent the Lagrange multipliers of the constraints, must have a Hessian $\nabla^2 L(x^*) = \nabla^2 f(x^*) + \sum_{i\in I} \lambda_i^* \nabla^2 g_i(x^*)$ which is a positive definite matrix in x^*.
- The sufficient conditions for the case of maximization can be translated into the conditions that f must be concave in the point, that the constraints do not change and that the multipliers are lower or equal to 0.

Necessary conditions with equality and inequality constraints (i)

- Consider the problem

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

where $f: \mathbb{R}^n \to \mathbb{R}$ $\quad g_i: \mathbb{R}^n \to \mathbb{R}$

- Let x^* be a feasible point

$I = \{i / g_i(x^*) = 0\}$ is the set of binding constraints

f and $\{g_i, i \in I\}$ are differentiable in x^*

$\{g_i, i \notin I\}$ are continuous in x^*

$\{h_j, j = 1, \ldots, l\}$ are continuously differentiable in x^*

$\{\nabla g_i(x^*), i \in I; \nabla h_j(x^*), j = 1, \ldots, l\}$ are linearly independent

Necessary conditions with equality and inequality constraints (ii)

- If x^* is a local minimum, then there exist scalars $\{\lambda_i, i \in I; \mu_j, j = 1,\ldots,l\}$ such that

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

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

- Moreover, if the functions $\{g_i, i \notin I\}$ are differentiable in x^*, if x^* is a local optimum then

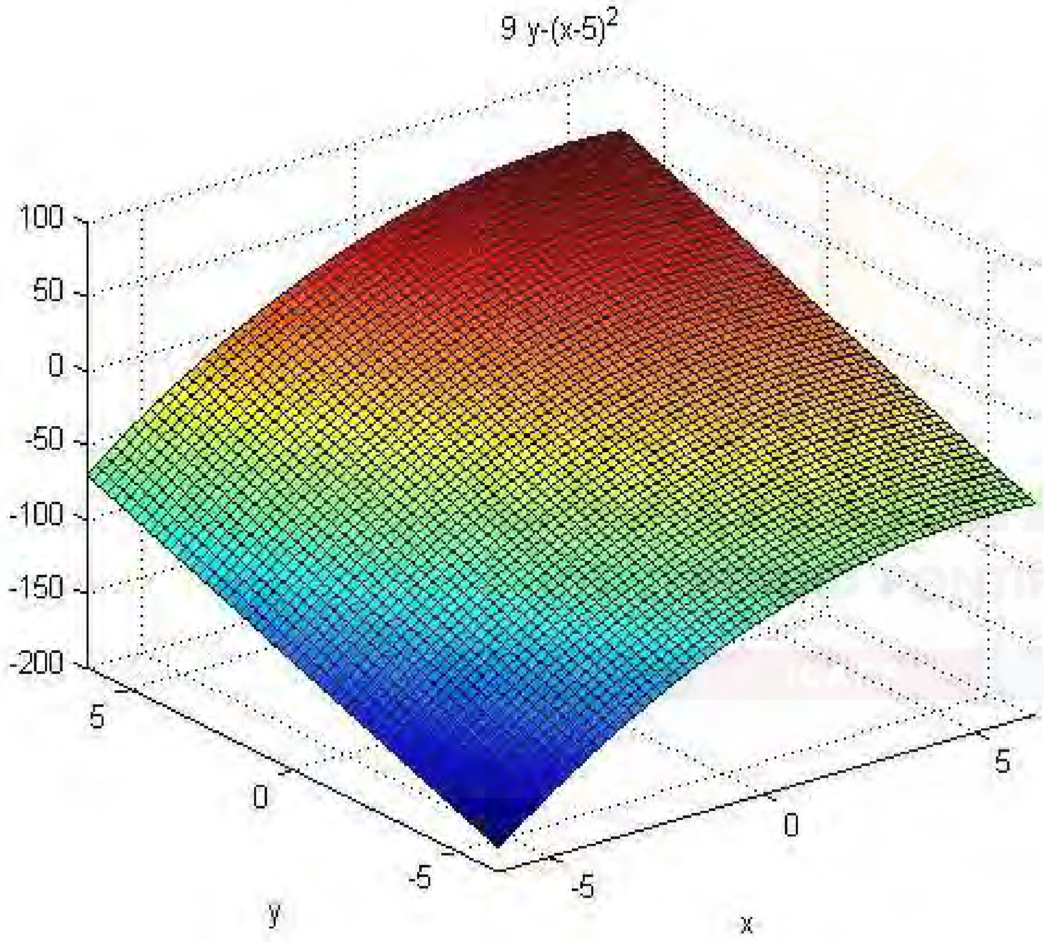
$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{l} \mu_j \nabla h_j(x^*) = 0$$

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

$$\lambda_i \geq 0 \quad i = 1,\ldots,m$$

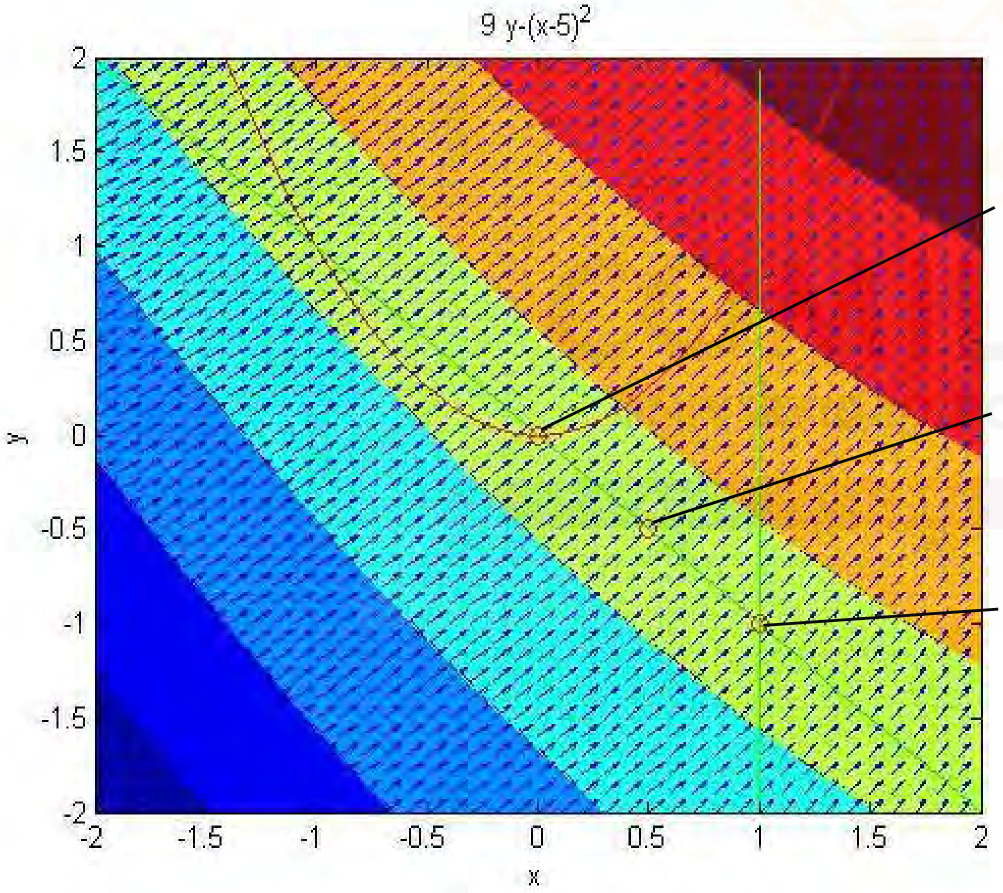
Example 5 (i)

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



Example 5 (ii)

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



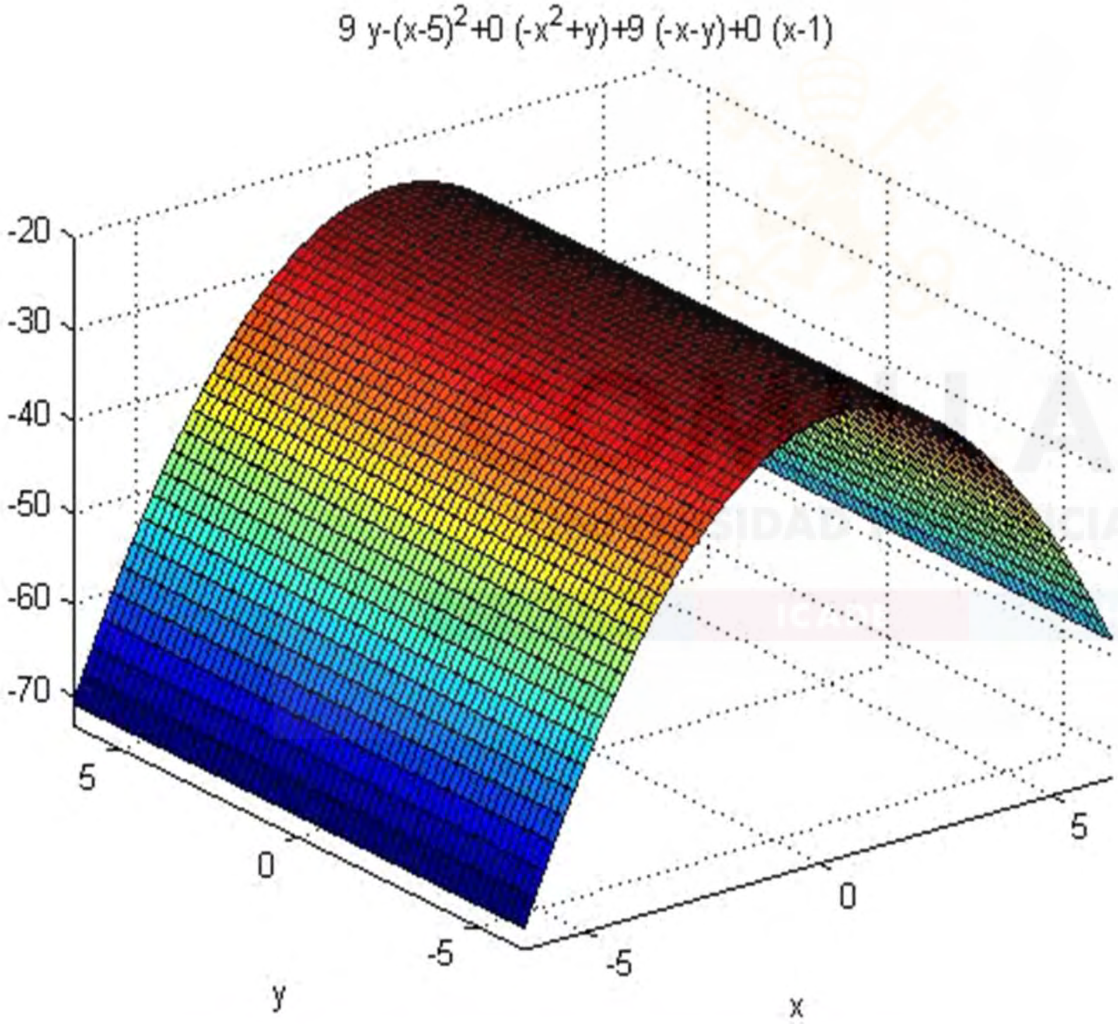
Point C (0,0,1,10,0) o.f.=-25

Point A (1/2,-1/2,0,9,0) o.f.=-24.75

Point B (1,-1,0,9,1) o.f.=-25

Example 5 (iii)

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



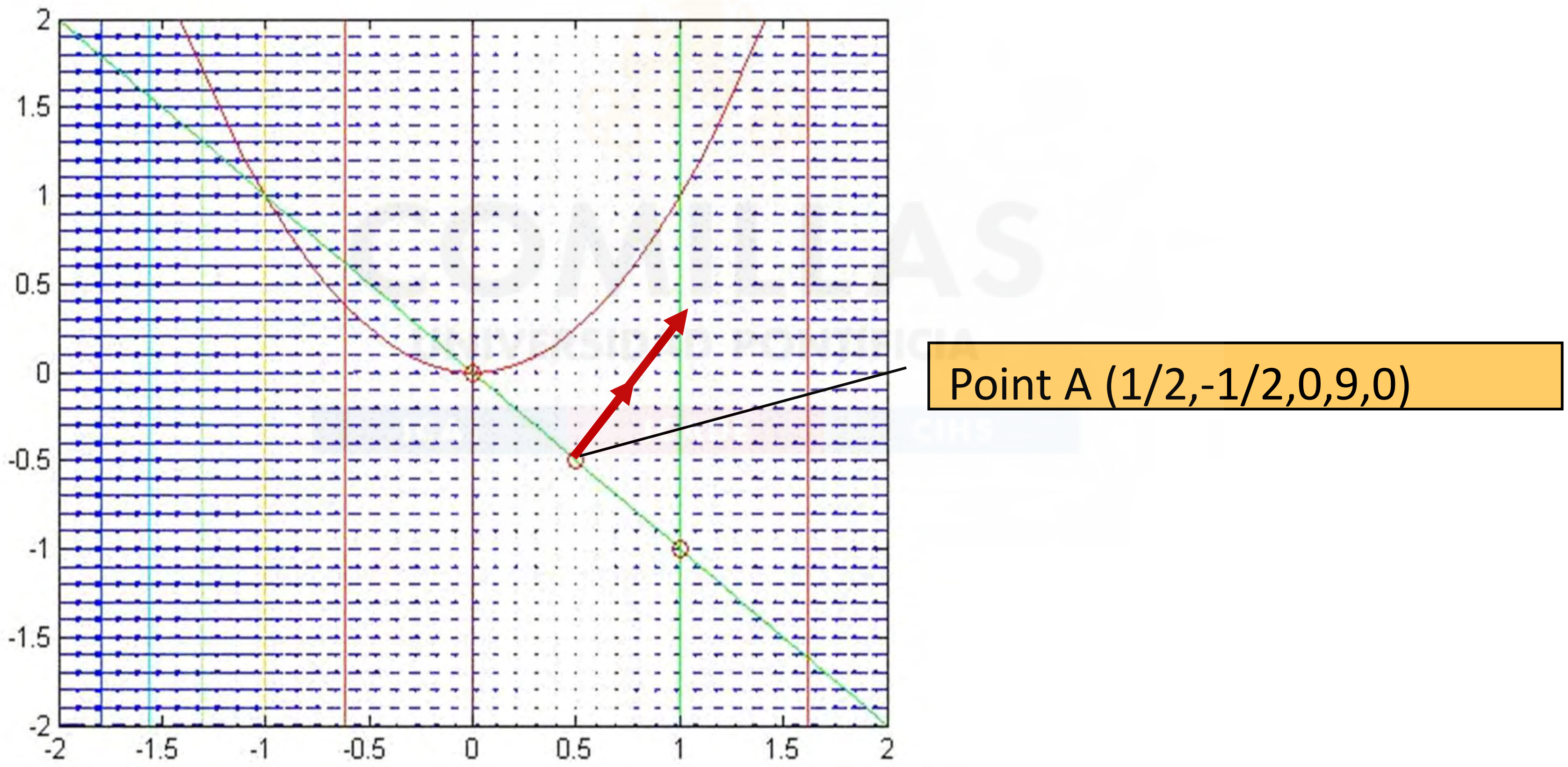
- Lagrangian for the multiplier values corresponding to point A (1/2,-1/2,0,9,0)

Unbounded function

Example 5 (iv)

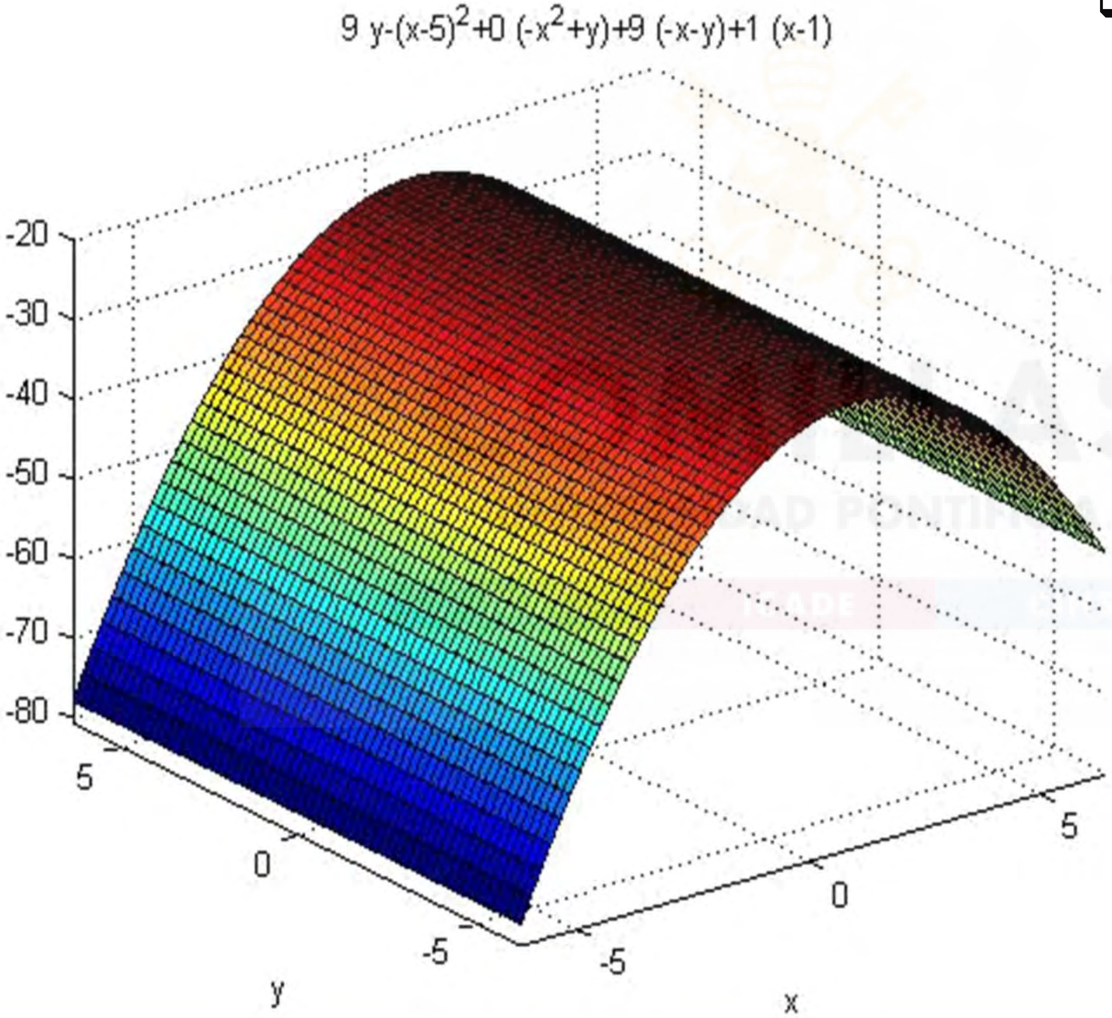
$$\min_{x,y,\lambda_1,\lambda_2,\lambda_3} L(x,y,\lambda_1,\lambda_2,\lambda_3) = 9y - (x-5)^2 + \lambda_1(-x^2+y) + \lambda_2(-x-y) + \lambda_3(x-1)$$

- Lagrangian for multiplier values corresponding to point A (1/2,-1/2,0,9,0)



Example 5 (v)

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



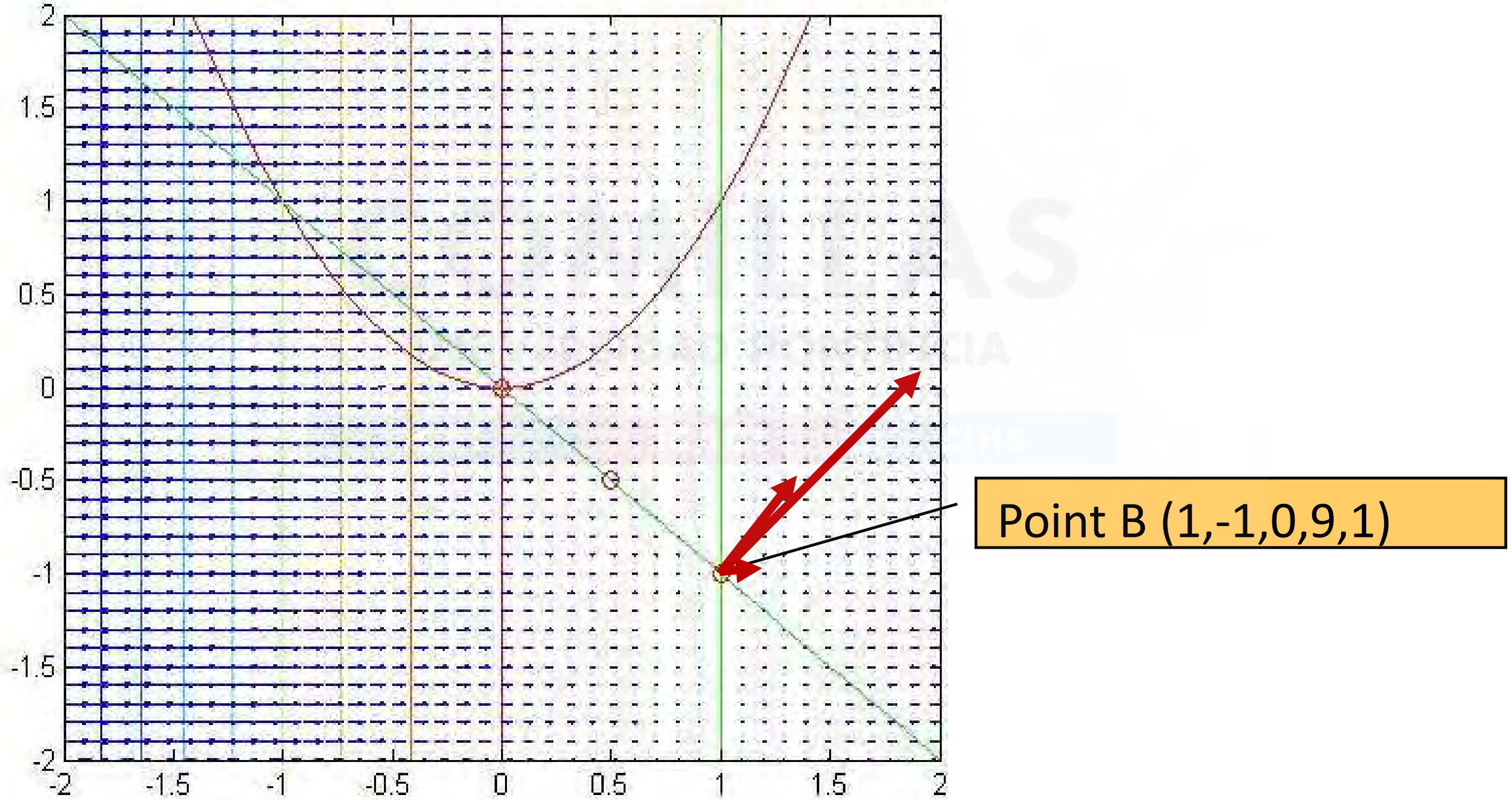
- Lagrangian for the multiplier values corresponding to point B (1,-1,0,9,1)

Unbounded function

Example 5 (vi)

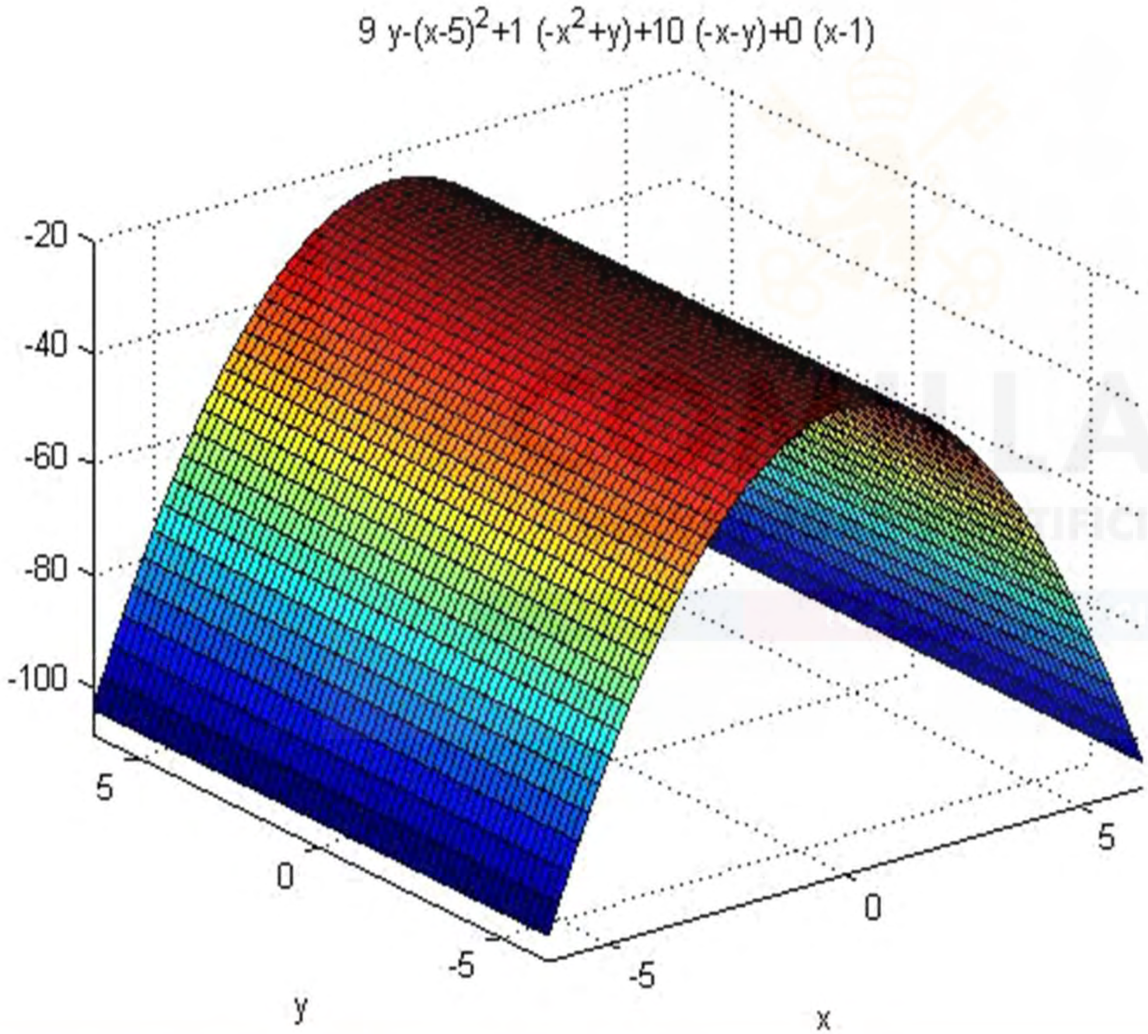
$$\min_{x,y,\lambda_1,\lambda_2,\lambda_3} L(x,y,\lambda_1,\lambda_2,\lambda_3) = 9y - (x-5)^2 + \lambda_1(-x^2+y) + \lambda_2(-x-y) + \lambda_3(x-1)$$

- Lagrangian for multiplier values corresponding to point B (1,-1,0,9,1)



Example 5 (vii)

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



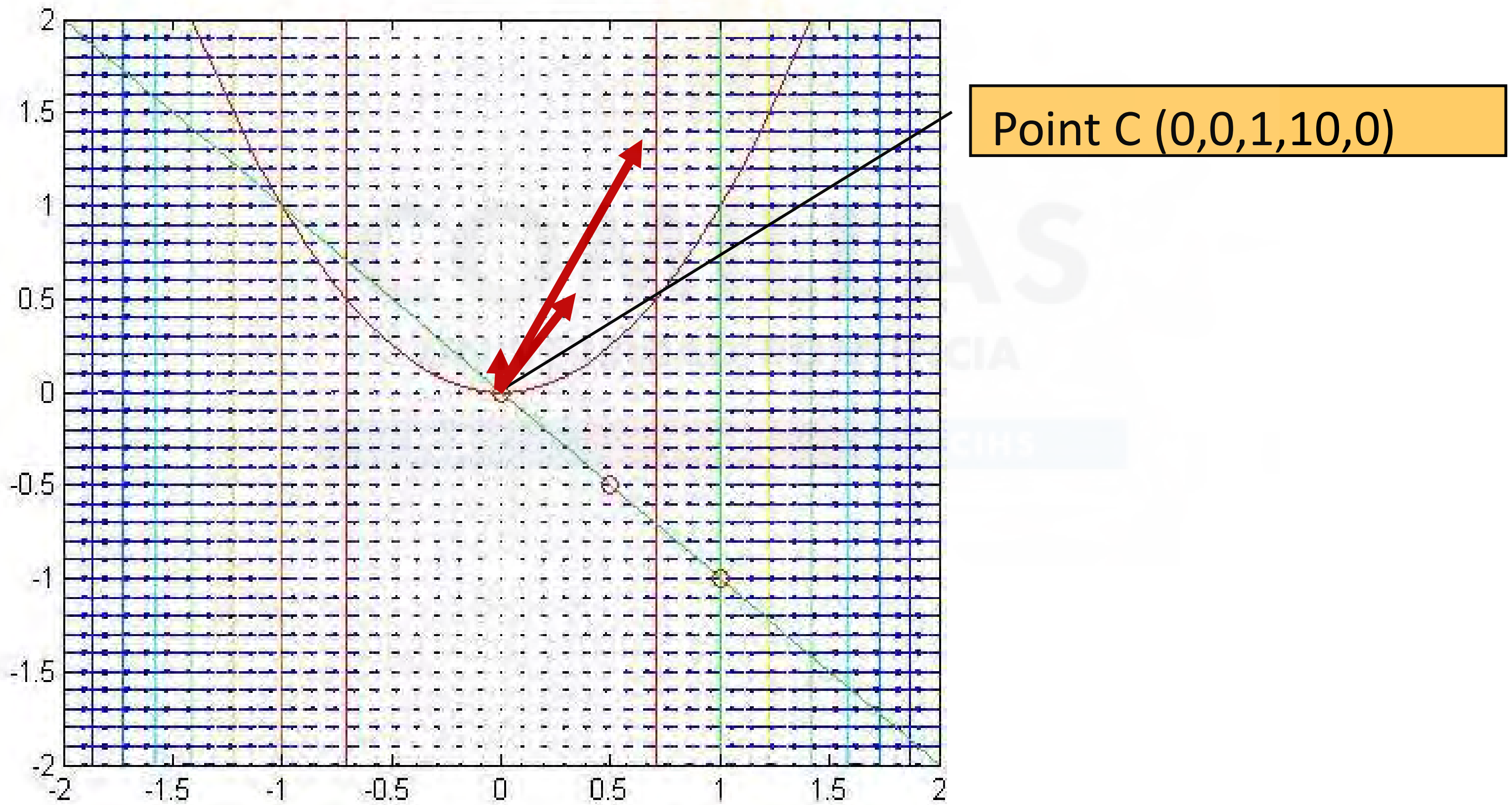
- Lagrangian for the multiplier values corresponding to point C (0,0,1,10,0)

Unbounded function

Example 5 (viii)

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

- Lagrangian for multiplier values corresponding to point C (0,0,1,10,0)



Necessary conditions with equality and inequality constraints (iii)

- Consider the problem

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

- The necessary first-order **Karush-Kuhn-Tucker** (KKT) conditions for a local optimum

Gradient of o.f.: Linear combination of the gradients of the constraints with changed sign

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

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

Complementary slackness conditions
Nonbinding constraint $\lambda = 0$
Binding constraint $\lambda \neq 0$

Feasible point

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

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

Sufficient conditions with equality and inequality constraints (i)

- Let x^* be a feasible point

  $I = \{i / g_i(x^*) = 0\}$ is the set of binding constraints

  f and $\{g_i, i \in I\}$ are convex and differentiable in **all the feasible region**

- If there exist scalars $\{\lambda_i, i \in I; \mu_j, j = 1, \ldots, l\}$ such that

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

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

  so that h_j is convex in all the feasible regions if $\mu_j > 0$ and h_j is concave in all the feasible region if $\mu_j < 0$, then x^* is a global minimum

Sufficient conditions with equality and inequality constraints (ii)

- The second order sufficient conditions for a local minimum
- The Hessian matrix $\nabla^2 f(x^*) + \sum_{i \in I} \lambda_i^* \nabla^2 g_i(x^*) + \sum_{j=1}^{l} \mu_j^* \nabla^2 h_j(x^*)$ must be positive definite

or rather $\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(x^*) + \sum_{j=1}^{l} \mu_j^* \nabla^2 h_j(x^*)$

TYPE OF NLP PROBLEMS AND SOLUTIONS
UNCONSTRAINED NONLINEAR OPTIMIZATION
OPTIMALITY CONDITIONS FOR NLP
METHODS FOR UNCONSTRAINED OPTIMIZATION (master)

4

METHODS FOR UNCONSTRAINED OPTIMIZATION

Classification of optimization methods WITHOUT constraints according to the use of derivatives

- Without derivatives
  - Random search method
  - Pattern search (Hooke-Jeeves method)
  - Rosenbrock method (of rotating or cyclic coordinates)
  - Nelder-Mead method
- First derivatives (gradient)
  - Method of steepest descent
  - Conjugate gradient method (Fletcher y Reeves)
- Second derivatives (Hessian)
  - Newton Method
  - Quasi Newton methods (Broyden-Fletcher-Goldfarb-Shanno BFGS, Davidon-Fletcher-Powell DFP)

Taylor series expansion

- Approximate a function f close to a given point x_0
- It is necessary to know the derivatives of the function

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

$p \in \mathbb{R}^n$ is a vector different from 0

$f(x)$ is the value of the function

$\nabla f(x)$ is the gradient of the function

$\nabla^2 f(x)$ is the Hessian of the function (if f has continuous second derivatives, then this is a symmetric matrix)

- Or alternatively

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

where ξ is a point between x and x_0

Newton method for a one-dimensional function (i)

- The interpolation algorithm approximates the function f in each iteration, in the point x_k considered in this iteration by a second or third-degree polynomial.
- This method fits, in iteration k, a parabola $q(x)$ to $f(x)$ and takes x_{k+1} as the vertex of this parabola.

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

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

- The algorithm has quadratic convergence under certain conditions, but it is very unstable and usually it is necessary to take precautions and to include protections.

Newton method for a one-dimensional function (ii)

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

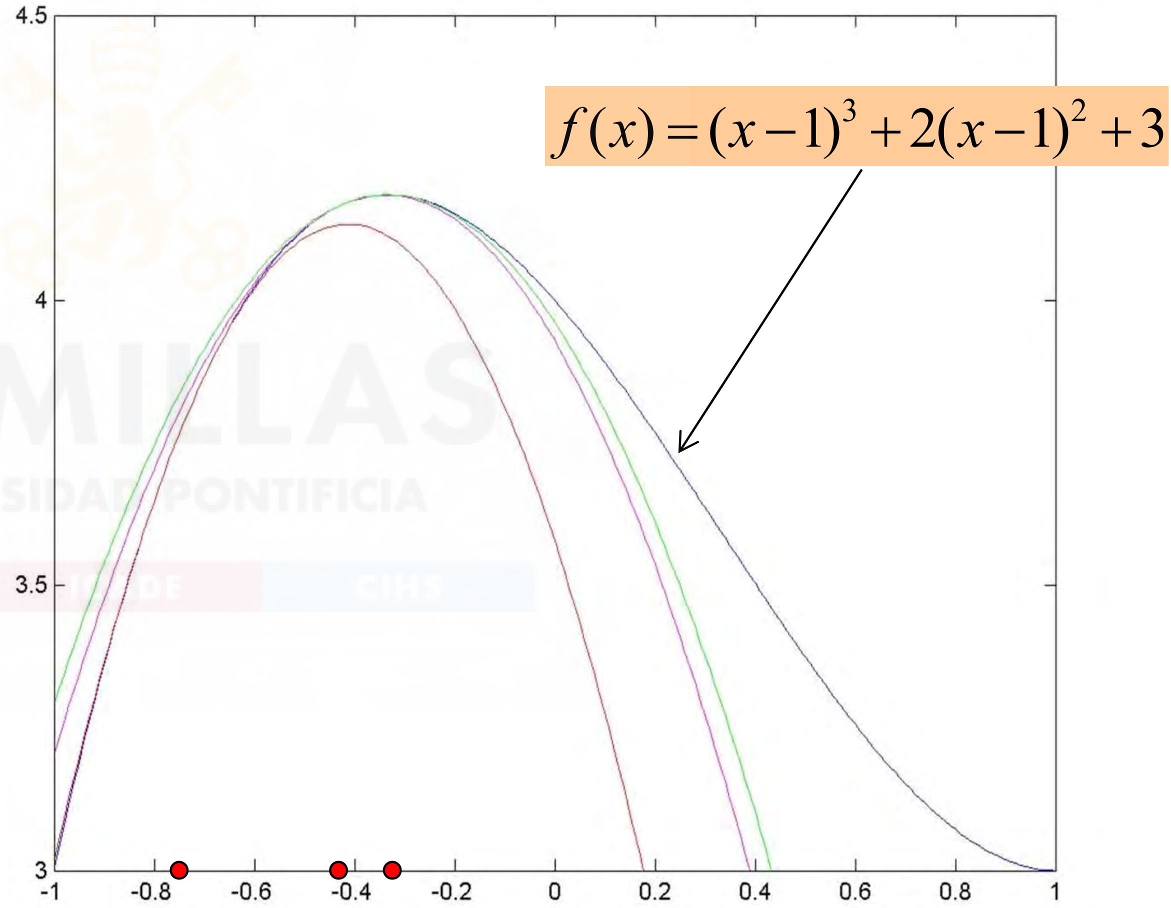
Sequence of points

$x_0 = -0.75$

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

$x_2 = -0.3376$

$x_3 = -0.3333$



General nonlinear optimization procedure

- Generate a sequence of points until convergence
  - Start from an initial point x_k
  - Obtain a search direction p_k
  - Calculate the step size α_k
  - Update the new point x_{k+1}

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

Conditions for each iteration

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

- Choose the new point such that the value of the objective function decreases $f(x_{k+1}) < f(x_k)$
- Conditions for the search direction p_k
  - The direction is descending $p^T \nabla f(x_k) < 0$
  - The descent is sufficient (non orthogonal vectors) $-\dfrac{p^T \nabla f(x_k)}{\|p\| \cdot \|\nabla f(x_k)\|} \geq \varepsilon > 0$
  - The search direction is related to the gradient $\|p\| \geq m \|\nabla f(x_k)\|$
- Conditions for the scalar α_k
  - The descent is sufficient (Armijo condition)

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

  - The descent is not too small

    For example, α_k is defined like a sequence 1, ½, ¼, 1/8, etc. The values of the sequence are used starting with 1. If for $\alpha = 1$ the previous condition is satisfied, we stop, otherwise, the next value of the sequence is used.
  - Minimize the value of the function: one-dimensional search method (*line-search methods*)

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

Newton method for solving a system of nonlinear equations (i)

- Solves a system of nonlinear equations iteratively

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

- Approximates the nonlinear function by a linear function in each point (iteration), using a first-order Taylor expansion

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

- $\nabla f(x)^T = \begin{pmatrix} \nabla f_1(x) & \nabla f_2(x) & \dots & \nabla f_n(x) \end{pmatrix}^T$ Jacobian of the function

Newton method for a system of nonlinear equations (ii)

- Has quadratic convergence if the point is close to the solution
- The Jacobian of the function has to be nonsingular at each point

Newton method for a system of nonlinear equations (iii)

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

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

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

- Consider this as the initial point $x_0=\begin{pmatrix}1\\2\end{pmatrix}$

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

- After 8 iterations, the point takes the value of $x_8=\begin{pmatrix}0\\1.5\end{pmatrix}$ and the functions take value 0.

Newton method for optimization

- The Newton method for a system of nonlinear equations is applied to the first-order necessary optimality conditions.

  $\nabla f(x) = 0$

- The Jacobian of this function $\nabla f(x)$ is the Hessian $\nabla^2 f(x)$
- Iteration $x_{k+1} = x_k + p_k = x_k - \nabla^2 f(x_k)^{-T} \nabla f(x_k)$
- Where p_k (the Newton direction) is obtained by solving a system of linear equations (Newton system) instead of calculating the inverse of the Hessian.

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

Example 1 (i)

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

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

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

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

- Initial point x_k=(0,0)
- Search direction p_k=(2,1)

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

- Calculate α_k

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

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

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

$$10\alpha-10=0$$

$$\alpha=1$$

Example 1 (ii)

- Next point $x_{k+1} = x_k + \alpha_k p_k$ = (2,1)
- Search direction p_k = (0,0)
- We have arrived at the optimum since the gradient is 0

Andrés Ramos

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

Andres.Ramos@comillas.edu

Pedro Sánchez

Pedro.Sanchez@comillas.edu

Sonja Wogrin

httpw://www.iit.comillas.edu/swogrin/

Sonja.Wogrin@comillas.edu