TEHNICI DE OPTIMIZARE
Curs 7

Andrei Pătraşcu

Departament Informatică
Universitatea din Bucureşti

- **Constrângeri de egalitate. Condiţii de optimalitate**
- Algoritmi pentru constrângeri de egalitate
- Metode de penalitate

$$(POCe:) \quad \min_{x\in\mathbb{R}^n} \ f(x)$$
$$\text{s.l.} \ \ g_i(x) = 0 \qquad \forall i = 1, \cdots, m$$

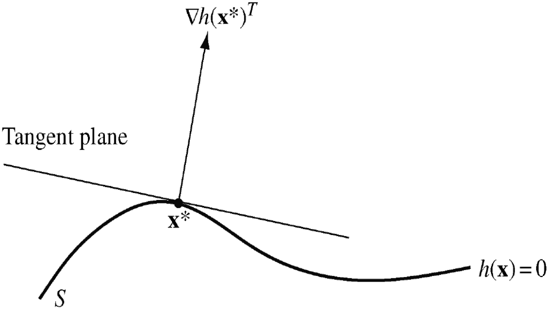
$$(POCe:) \quad \min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.l. } g_i(x) = 0 \qquad \forall i = 1, \cdots, m$$

- Funcţia Lagrange $\mathcal{L} : \text{dom}(f) \times \mathbb{R}^m$:

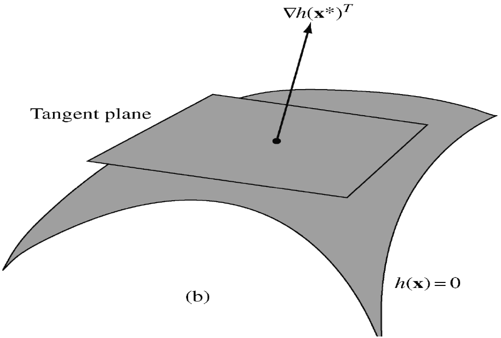
$$\mathcal{L}(x, \mu) = f(x) + \mu^T g(x)$$

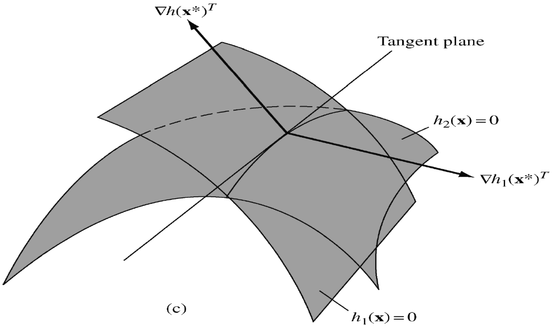
- μ multiplicatori Lagrange (variabile duale)

Condiţii de optimalitate (caz neliniar)



Condiţii de optimalitate (caz neliniar)




$\nabla h(\mathbf{x}^*)^T$
Tangent plane
$h_2(\mathbf{x}) = 0$
$\nabla h_1(\mathbf{x}^*)^T$
$h_1(\mathbf{x}) = 0$

(c)

$$(POCe:)\quad \min_{x\in\mathbb{R}^n} \; f(x)$$
$$\text{s.l.}\;\; g_i(x)=0 \qquad \forall i=1,\cdots,m$$

Definiţie (Plan tangent)

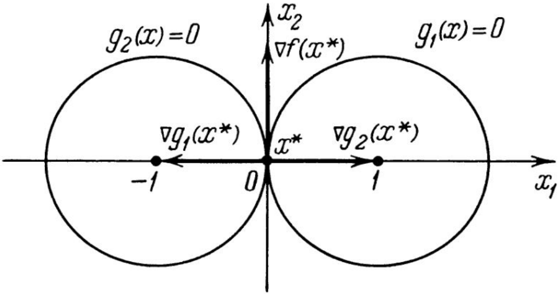
Fie $Q=\{x : g_i(x)=0, i=1,\cdots,m\}$*, unde* g_i *diferenţiabile în jurul lui* y*.*
Mulţimea direcţiilor tangente în punctul y *este descrisă de*
$T(x)=\{s : \nabla g_i(x)^T s=0 \quad \forall i=1,\cdots,m\}$*.*

$$
\begin{aligned}
(POCe:) \quad & \min_{x \in \mathbb{R}^n} \; f(x) \\
& \text{s.l.} \;\; g_i(x) = 0 \qquad \forall i = 1, \cdots, m
\end{aligned}
$$

Definiţie

Punctul x^* este minim regulat dacă $g_i(x)$ sunt continuu diferenţiabile în vecinătatea lui x^* şi $\{\nabla g_i(x^*)\}$ sunt liniar independenţi.

$$\min_{x\in\mathbb{R}^2} \quad x_2$$
$$\text{s.l.} \quad (x_1-1)^2+x_2^2-1=0, \quad (x_1+1)^2+x_2^2-1=0$$



$$(POCe:) \quad \min_{x\in\mathbb{R}^n} \; f(x)$$
$$\text{s.l.} \quad g_i(x) = 0 \qquad \forall i = 1, \cdots, m$$

Teoremă (Multiplicatori Lagrange)

Dacă x^* minim regulat, atunci există μ^* astfel încât:

$$\nabla_x \mathcal{L}(x^*, \mu^*) := \nabla f(x^*) + \mu_1^* \nabla g_1(x^*) + \mu_2^* \nabla g_2(x^*) + \cdots + \mu_m^* \nabla g_m(x^*) = 0 (opt.)$$
$$\nabla_\mu \mathcal{L}(x^*, \mu^*) := g(x^*) = 0 \;\; (fezabilitate)$$

$$(POCe:) \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s.l.} \; g_i(x) = 0 \qquad \forall i = 1, \cdots, m$$

Teoremă (Condiţii necesare ord. 2)

Dacă x^* minim regulat şi f şi g_i dublu diferenţiabile în jurul lui x^*. Atunci există μ^* astfel încât:

$$s^T \nabla^2 \mathcal{L}(x^*, \mu^*) s \geq 0 \qquad \forall s \in T(x^*)$$

unde $T(x^*) = \{s : \nabla g_i(x^*)^T s = 0 \quad \forall i = 1, \cdots, m\}$.

Echivalent, $\nabla_x^2 \mathcal{L}(x^*, \mu^*)$ pozitiv semidefinită pe planul tangent $T(x^*)$.

$$(POCe:) \quad \min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.l. } g_i(x) = 0 \qquad \forall i = 1, \cdots, m$$

Teoremă (Condiţii suficiente ord. 2)

Fie $g_i(x^*) = 0$, f şi g_i dublu diferenţiabile în jurul lui x^*. Mai mult, presupunem $\{\nabla g_i(x^*)\}$ liniar independenţi. Atunci dacă:

$$s^T \nabla^2 \mathcal{L}(x^*, \mu^*) s > 0 \qquad \forall s \in T(x^*)$$

unde $T(x^*) = \{s : \nabla g_i(x^*)^T s = 0 \quad \forall i = 1, \cdots, m\}$, atunci x^* este minim local.

Dacă $\nabla_x^2 \mathcal{L}(x^*, \mu^*)$ este pozitiv definită pe planul tangent $T(x^*)$ atunci x^* este minim local.

Exemplu:

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \ & x_1^2 - x_2^2 \\ \text{s.l.} \ & x_2 = 0. \end{aligned}$$

- $\min_x \sum_{i=1}^{n} x_i^2$ s.l. $\sum_{i=1}^{n} x_i = 1$
- $\min_x \sum_{i=1}^{n} x_i$ s.l. $\sum_{i=1}^{n} x_i^2 = 1$
- $\min_x x^T A x$ s.l. $\sum_{i=1}^{n} x_i^2 = 1$
- $\min_x \|x\|^2$ s.l. $x^T A x = 1$

$$\min_{x} \sum_{i=1}^{n} x_i^2 \text{ s.l. } \sum_{i=1}^{n} x_i = 1$$

Funcţia Lagrangian:

$$\mathcal{L}(x, \mu) = x^T I_n x + \mu(e^T x - 1).$$

Sistem Kuhn-Tucker:

$$\nabla_x \mathcal{L}(x, \mu) = 2x + e\mu = 0$$
$$\nabla_\mu \mathcal{L}(x, \mu) = e^T x = 1.$$

Soluţii? Puncte staţionare sau optime locale?

$$\min_x \sum_{i=1}^{n} x_i \ \text{ s.l. } \ \sum_{i=1}^{n} x_i^2 = 1$$

Funcţia Lagrangian:

$$\mathcal{L}(x, \mu) = e^T x + \mu(\|x\|_2^2 - 1).$$

Sistem Kuhn-Tucker:

$$\nabla_x \mathcal{L}(x, \mu) = e + \mu x = 0$$
$$\nabla_\mu \mathcal{L}(x, \mu) = \|x\|_2 = 1.$$

Soluţii? Indiciu: $\nabla_x^2 \mathcal{L}(x, \mu) = 2\mu I_n$.

$$\min_x \; x^T A x \;\; \text{s.l.} \;\; \sum_{i=1}^{n} x_i^2 = 1, \quad A \succ 0.$$

Funcţia Lagrangian:

$$\mathcal{L}(x,\mu) = x^T A x + \mu(\|x\|_2^2 - 1).$$

Sistem Kuhn-Tucker:

$$\nabla_x \mathcal{L}(x,\mu) = Ax + \mu x = 0$$
$$\nabla_\mu \mathcal{L}(x,\mu) = \|x\|_2 = 1.$$

Soluţii? Indiciu: $\nabla_x^2 \mathcal{L}(x,\mu) = 2A + 2\mu I_n$.

$$Q = \{x : g_i(x) = 0 \quad \forall i = 1, \cdots, m\}$$

$$\min_x f(x) + \iota_Q(x) = \begin{cases} f(x) & x \in Q \\ \infty & x \notin Q \end{cases}$$

Reformulare funcţie indicator (multiplicatori Lagrange):

$$\iota_Q(x) = \max_{\mu \in \mathbb{R}^m} \mu^T g(x) \left(= \sum_i \mu_i g_i(x) \right) = \begin{cases} 0 & g(x) = 0 \\ \infty & g(x) \neq 0. \end{cases}$$

$$\min_x f(x) + \iota_Q(x) = \begin{cases} f(x) & x \in Q \\ \infty & x \notin Q \end{cases}$$

Echivalent (multiplicatori Lagrange):

$$\min_x f(x) + \iota_Q(x) = \min_{x \in \mathbb{R}^n} \max_{\mu \in \mathbb{R}^m} \; f(x) + \mu^T g(x) = \begin{cases} f(x) & x \in Q \\ \infty & x \notin Q \end{cases}$$

$$\min_x f(x) + \iota_Q(x) = \begin{cases} f(x) & x \in Q \\ \infty & x \notin Q \end{cases}$$

Echivalent (multiplicatori Lagrange):

$$\min_x f(x) + \iota_Q(x) = \min_{x \in \mathbb{R}^n} \max_{\mu \in \mathbb{R}^m} \; f(x) + \mu^T g(x) = \begin{cases} f(x) & x \in Q \\ \infty & x \notin Q \end{cases}$$

[Pentru $\mu = \mu^*$ mărginiţi] $\min_x f(x) + \iota_Q(x) = \min_{x \in \mathbb{R}^n} \; f(x) + (\mu^*)^T g(x)$

- Constrângeri de egalitate. Condiţii de optimalitate
- **Algoritmi pentru constrângeri de egalitate**
- Metode de penalitate

Metoda Gradientului Proiectat cu liniarizare:

$$x^{k+1} = \arg\min_z \; f(x^k) + \nabla f(x^k)^T(z - x^k) + \frac{1}{2\alpha_k}\|z - x^k\|^2$$
$$\text{s.l.} \;\; g_i(x^k) + \nabla g_i(x^k)^T(z - x^k) = 0 \quad \forall i = 1, \cdots, m.$$

- Aproximarea pătratică a funcţiei obiectiv în stil MG
- Aproximarea liniară a constrângerilor (în jurul lui x^k)
- Pt constrângeri liniare se reduce la MGP

$$\min_x \; f(x) \;\; \text{s.l.} \;\; \sum_i x_i = 1$$

Metoda Gradientului Proiectat cu liniarizare:

$$x^{k+1} = \arg\min_z \; f(x^k) + \nabla f(x^k)^T (z - x^k) + \frac{1}{2\alpha_k}\|z - x^k\|^2$$
$$\text{s.l.} \;\; \sum_i z_i = 0.$$

Soluţie explicită:

$$x^{k+1} = \left(I_n - \frac{1}{n} e \cdot e^T\right)(x^k - \alpha \nabla f(x^k))$$

$$\min_x \; f(x) \;\text{ s.l. }\; g(x) = 0$$

Metoda Gradientului Proiectat cu liniarizare:

$$x^{k+1} = \arg\min_z \; \|z - (x^k - \alpha \nabla f(x^k))\|^2 \;\text{ s.l. }\; \nabla g(x^k) z = \nabla g(x^k) x^k - g(x^k)$$

Notăm $G^k := \nabla g(x^k), g^k := \nabla g(x^k) x^k - g(x^k)$, atunci

$$x^{k+1} = \left(I_n - [G^k]^\dagger G^k\right)(x^k - \alpha \nabla f(x^k)) + [G^k]^\dagger g^k.$$

Metoda Gradientului Proiectat cu liniarizare:

$$x^{k+1} = \arg\min_z \; f(x^k) + \nabla f(x^k)^T (z - x^k) + \frac{1}{2\alpha_k}\|z - x^k\|^2$$
$$\text{s.l.} \;\; g_i(x^k) + \nabla g_i(x^k)^T (z - x^k) = 0 \quad \forall i = 1, \cdots, m.$$

Teoremă

Fie f, g_i astfel încât $\nabla^2 f(x), \nabla^2 g_i(x)$ sunt continue Lipschitz. MGP cu liniarizare converge local la un minim (nesingular), cu rată liniară.

Metoda Gradient Primal-Dual (Arrow-Hurwicz):

$$
\begin{aligned}
x^{k+1} &:= x^k - \alpha \nabla_x \mathcal{L}(x^k, \mu^k) \quad \left(= x^k - \alpha[\nabla f(x^k) + \nabla g(x^k)^T \mu^k]\right) \\
\mu^{k+1} &:= \mu^k + \alpha \nabla_\mu \mathcal{L}(x^k, \mu^k) \quad \left(= \mu^k + \alpha g(x^k)\right)
\end{aligned}
$$

- Pas de MG pentru minimizarea în x
- Pas de MG pentru maximizarea în μ
- Dacă $\mu^0 = \mu^*$ atunci MGPD se reduce la MG pe $\mathcal{L}(x, \mu^*)$

Metoda Gradient Primal-Dual (Arrow-Hurwicz):

$$x^{k+1} := x^k - \alpha \nabla_x \mathcal{L}(x^k, \mu^k) \quad \left(= x^k - \alpha[\nabla f(x^k) + \nabla g(x^k)^T \mu^k]\right)$$

$$\mu^{k+1} := \mu^k + \alpha \nabla_\mu \mathcal{L}(x^k, \mu^k) \quad \left(= \mu^k + \alpha g(x^k)\right)$$

Teoremă

Fie f, g_i astfel încât $\nabla^2 f(x), \nabla^2 g_i(x)$ sunt continue Lipschitz. Metoda Arrow-Hurwicz converge local la un minim (nesingular), cu rată liniară.

- Constrângeri de egalitate. Condiţii de optimalitate
- Algoritmi pentru constrângeri de egalitate
- **Metode de penalitate**

$$\min_{x} \; f(x) \;\; \text{s.l.} \;\; x \in Q$$

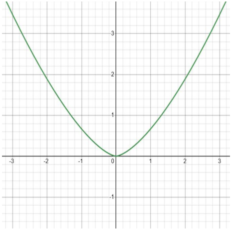
Funcţie penalitate asociată mulţimii fezabile Q:

(*i*) ϕ continuu diferenţiabilă şi strict convexă pe Q

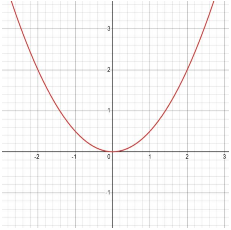
(*ii*) $\phi(0) = 0$ şi $\nabla\phi(0) = 0$

(*iii*) $\lim_{t\to-\infty} \nabla\phi(t) = -\infty$ şi $\lim_{t\to\infty} \nabla\phi(t) = \infty$

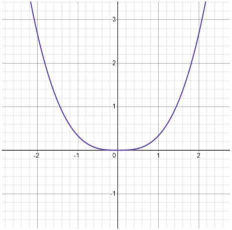
Funcţie penalitate



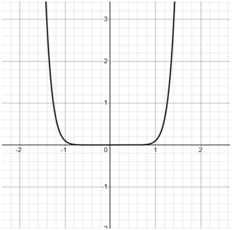
$$\phi(x) = \frac{2}{3}|x|^{3/2}$$



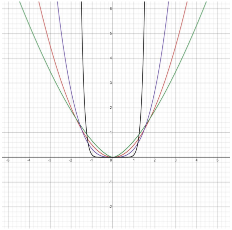
$$\phi(x) = \frac{1}{2}|x|^2$$



$$\phi(x) = \frac{1}{3}|x|^3$$



$$\phi(x) = \frac{1}{10}|x|^{10}$$



$$\min_x \; f(x) \;\; \text{s.l.} \;\; x \in Q$$

Funcţia de penalitate aproximează funcţia indicator asociată mulţimii Q:

$$\min_x \; \phi(\|g(x)\|) = 0$$
$$\arg\min_x \; \phi(\|g(x)\|) = Q$$

Aproximăm (POCe) cu:

$$x_\rho^* := \arg\min_x \;\; f(x) + \rho\phi(\|g(x)\|)$$

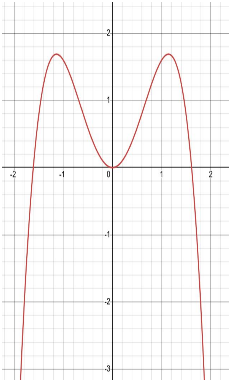
Observăm:

- $\min_x \; \lim_{\rho\to\infty} \; f(x) + \rho\phi(\|g(x)\|) = f^*$
- $\exists \alpha(\rho), \beta(\rho) > 0 : f^* - f(x_\rho^*) \leq \alpha(\rho), \|g(x)\| \leq \beta(\rho)$

Alegerea funcţiei de penalitate:

$$\min_{x} \quad f(x) := -x^4$$
$$\text{s.l. } g(x) := x = 0.$$

$$[\phi \text{ pătratic}] \quad x_\rho^* = \arg\min_{x} \quad f(x) + \rho g(x) = -x^4 + \rho x^2.$$



Alegerea funcţiei de penalitate:

$$\min_{x} \quad f(x) := -x^4$$
$$\text{s.l. } g(x) := x = 0.$$

$$[\phi \text{ polinomial}] \;\; x^*_\rho = \arg\min_x \;\; -x^4 + \rho x^2 + x^\gamma.$$

Când există x^*_ρ?

$$\min_x \; f(x) \;\text{ s.l. }\; g(x) = 0$$

Funcţie penalitate pătratică ($\rho > 0$):

$$\min_x \; f(x) + \rho \sum_i (g_i(x))^2.$$

- Fie $\rho_1 \leq \rho_2$ şi $x^*_\rho = \arg\min_x \; f(x) + \rho \sum_i (g_i(x))^2$, atunci:

$$\sum_i (g_i(x^*_{\rho_2}))^2 \leq \sum_i (g_i(x^*_{\rho_1}))^2.$$

- $\rho \to \infty$ atunci $g_i(x) \to 0$

Considerăm: $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$

- $\min_x \;\; 0 \;\; \text{s.l.} \;\; Ax = b$
- $\min_x \;\; c^T x \;\; \text{s.l.} \;\; Ax = b$
- $\min_x \;\; \|x\|_2^2 \;\; \text{s.l.} \;\; Ax = b$
- $\min_x \;\; \frac{1}{2} x^T H x + q^T x \;\; \text{s.l.} \;\; Ax = b, \quad H \succ 0.$
- $\min_x \;\; \frac{1}{2} \|Ax - b\|_2^2 \;\; \text{s.l.} \;\; \|x\| = 1$

Metoda de Penalitate Pătratică:

$$x^k = \arg\min_{x \in Q_0} f(x) + \frac{\rho_k}{2}\|g(x)\|^2 \qquad \rho_k \to \infty.$$

Teoremă

Considerăm $X^* \neq \emptyset$. Fie f, g_i funcţii continue şi Q_0 mărginită, $Q_0 \cap X^* \neq \emptyset$. Atunci punctele limită ale şirului generat de MP sunt minime globale ale problemei (POCe).

Metoda de Penalitate:

$$x^k = \arg\min_{x \in \mathbb{R}^n} f(x) + \frac{\rho_k}{2}\|g(x)\|^2 \qquad \rho_k \to \infty.$$

- pentru mărginirea domeniului, eventual, înlocuim $\mathbb{R}^n$ cu Q_0
- la fiecare iteraţie se rezolvă subproblema (posibil neconvexă) $\min_x f(x) + \frac{\rho_k}{2}\|g(x)\|^2$
- $\|g(x^k)\| \to 0$ când $\rho_k \to \infty$!

Metoda de Penalitate:

$$x^k = \arg\min_{x \in \mathbb{R}^n} f_{\rho_k}(x) := f(x) + \frac{\rho_k}{2}\|g(x)\|^2 \qquad \rho_k \to \infty.$$

- Scenariu 1: Soluţia x^k nu există
- Exemplu!

Metoda de Penalitate:

$$x^k = \arg\min_{x \in \mathbb{R}^n} f_{\rho_k}(x) := f(x) + \frac{\rho_k}{2}\|g(x)\|^2 \qquad \rho_k \to \infty.$$

- Scenariu 2: Soluţia x^k există, dar $\{x^k\}_{k\geq 0}$ nu converge.
- Exemplu 1: $Q = \emptyset$
- Exemplu 2: $f = 0$ şi $\nabla g(x^*)$ defectiv de rang

$$\min_{x} \ 0$$
$$\text{s.l.} \ \ x_1^2 = 3, \quad -2x_1 + x_2^2 = 0$$

$Q = \{[\sqrt{3} \quad \pm 2\sqrt{3}]^T\}$

Pentru $z = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

$$\nabla f_\rho(z) := \rho \nabla g(z) g(z) = \begin{bmatrix} -2 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Metoda de Penalitate:

$$x^k = \arg\min_{x \in Q_0} f(x) + \frac{\rho_k}{2}\|g(x)\|^2 \qquad \rho_k \to \infty.$$

- la fiecare iteraţie se rezolvă subproblema $\min_x f(x) + \frac{\rho_k}{2}\|g(x)\|^2$
- $\|g(x^k)\| \to 0$ când $\rho_k \to \infty$!
- Pentru ρ mare creşte numărul de condiţionare al subproblemei.
- Rată de convergenţă: $\|x^k - x^*\| = O(1/\rho_k)$

1. Luenberger, David G., and Yinyu Ye. Linear and nonlinear programming. Vol. 2. Reading, MA: Addison-wesley, 1984.