TEHNICI DE OPTIMIZARE
Curs 6

Andrei Pătraşcu

Departament Informatică
Universitatea din Bucureşti

Metoda Newton (pentru constrângeri simple)

$$x^{k+1} = \operatorname*{argmin}_{z \in Q} \; f(x^k) + \nabla f(x^k)^T (z - x^k) + \frac{1}{2\alpha_k}(z - x^k)^T \nabla^2 f(x^k)(z - x^k).$$

Metoda Newton (cu proiecţie)

$$x^{k+1} = \arg\min_{z \in Q} \; f(x^k) + \nabla f(x^k)^T (z - x^k) + \frac{1}{2\alpha_k}(z - x^k)^T \nabla^2 f(x^k)(z - x^k).$$

La iteraţia k, necesită rezolvarea:

$$QP: \; \min_{z \in Q} \; q^T z + \frac{1}{2\alpha_k} z^T H z.$$

Cuprins

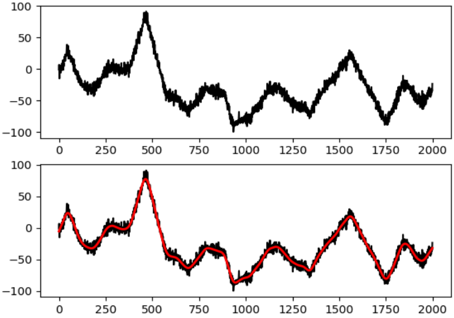
- Constrângeri de egalitate.
- Condiţii de optimalitate

Fie seria de timp $\{y_t\}_{t=1}^n \in \mathbb{R}^n$, presupunem descompunerea:

$$y_t = x_t + \epsilon_t,$$

- x_t reprezintă componenta-trend de variaţie lentă
- ϵ_t componenta zgomot de variaţie rapidă

Filtrare a trend-ului = estimarea componentei ”netede” x_t.


100
50
0
−50
−100
0
250
500
750
1000
1250
1500
1750
2000

$$\min_{x} \ \frac{1}{2}\sum_{t=1}^{n}(x_t - y_t)_2^2 + \rho\sum_{t=2}^{n-1}(x_{t-1} - 2x_t + x_{t+1})^2. \qquad (1)$$

- Funcţia obiectiv urmăreşte reducerea reziduului $\{y_t - x_t\}$ şi, în acelaşi timp, ajustarea "netezimii" lui x_t.
- Diferenţa de ordin 2: $x_{t-1} - 2x_t + x_{t+1}$ este nulă dacă şi numai dacă $\{x_{t-1}, x_t, x_{t+1}\}$ sunt coliniare.

În forma restrânsă:

$$x_{HP}^* := \operatorname*{argmin}_x \frac{1}{2}\|x - y\|_2^2 + \rho\|Dx\|_2^2, \quad \text{unde} \quad D \in \mathbb{R}^{(n-2)\times n}$$

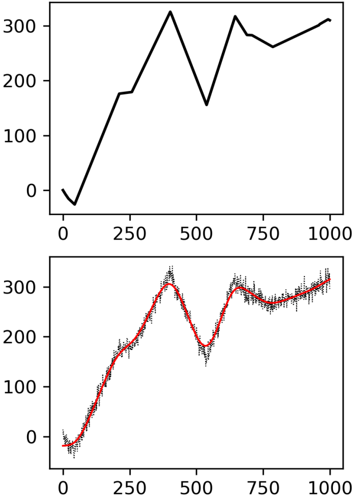
$$D = \begin{bmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \vdots & \vdots & \vdots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1 \end{bmatrix}$$

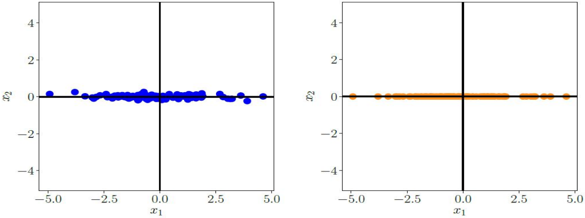
$$\begin{aligned} x_{HP}^* := \operatorname*{argmin}_x \ & \frac{1}{2}\|x - y\|_2^2 + \rho\|z\|_2^2 \\ \text{s.t.} \ & Dx = z. \end{aligned}$$

$$x^*_{\ell_1} := \operatorname*{argmin}_x \ \frac{1}{2}\|x - y\|_2^2 + \rho\|Dx\|_1$$

- $\|\cdot\|_2^2$ devine $\|\cdot\|_1$
- Dacă $D = I_n$ atunci problema este separabilă: $x_t^* = \operatorname{argmin}_{x_t \in \mathbb{R}} \ \frac{1}{2}(x_t - y_t)_2^2 + \lambda|x_t|$

Filtru ℓ_1: rezultate





4
2
0
−2
−4
x_2
−5.0
−2.5
0.0
2.5
5.0
x_1
4
2
0
−2
−4
x_2
−5.0
−2.5
0.0
2.5
5.0
x_1

$$X = \begin{bmatrix} 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \end{bmatrix}$$

- Fiecare X_i se poate recupera complet din baza b şi codul z_i.

$$X = \begin{bmatrix} 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \end{bmatrix}$$

- Fiecare X_i se poate recupera complet din baza b şi codul z_i.
- Dimensiunea lui X (72 octeţi) poate fi redusă la $\text{size}(b) + \text{size}(z) = 9 \cdot 4 =$ **36** octeţi.

$$X = \begin{bmatrix} 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \end{bmatrix}$$

- Fiecare X_i se poate recupera complet din baza b şi codul z_i.
- Dimensiunea lui X (72 octeţi) poate fi redusă la size(b) + size(z) $= 9 \cdot 4 =$ **36** octeţi.
- Dacă numărul de coloane m creşte (coliniar pe b), atunci dimensiunea lui z creşte proporţional **4m** octeţi (dimensiunea bazei este constantă).

$$X = \begin{bmatrix} 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \end{bmatrix}$$

- Fiecare X_i se poate recupera complet din baza b şi codul z_i.
- Dimensiunea lui X (72 octeţi) poate fi redusă la $\text{size}(b) + \text{size}(z) = 9 \cdot 4 =$ **36** octeţi.
- Dacă numărul de coloane m creşte (coliniar pe b), atunci dimensiunea lui z creşte proporţional **4m** octeţi (dimensiunea bazei este constantă).
- În particular, fixând $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ avem reconstruţie perfectă $X_i = z_i b$ unde

$$z_1 = 15, z_2 = 0.1, z_3 = 0.5, z_4 = -100, z_5 = -1.01, z_6 = 0.00001$$

.

Matricea de covarianţă:

$$S = XX^T$$

Calculul componentei principale:

$$PCA: \quad \max_{x \in \mathbb{R}^n} \|Sx\|_2^2$$
$$\text{s.l.} \quad \|x\| = 1$$

Rezultatul pentru $X = \begin{bmatrix} 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \\ 15 & 0.1 & 0.5 & -100 & -1.01 & 0.00001 \end{bmatrix}$ este

$$x^* = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

$$(POCel:) \quad \min_{x \in \mathbb{R}^n} \ f(x)$$
$$\text{s.l.} \ \ Ax = b,$$

unde $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

- f funcţie diferenţiabilă (netedă)
- $Q = \{x \in \mathbb{R}^n : a_i^T x = b_i, \forall i = 1, \cdots, m\} \Rightarrow$ proiecţie costisitoare.
- x^* minim local: $f(x^*) \leq f(x)$ pentru $x \in Q$ în vecinătatea lui x^*

$$(POCel:) \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s.l.} \;\; Ax = b,$$

unde $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

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

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

- $\mu \in \mathbb{R}^m$ multiplicatori Lagrange (variabile duale).

$$(POCel:) \quad \min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s.l.} \;\; Ax = b,$$

unde $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

Teoremă (Multiplicatori Lagrange)

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

$$\nabla_x \mathcal{L}(x^*, \mu^*) := \nabla f(x^*) + \mu_1^* a_1 + \mu_2^* a_2 + \cdots + \mu_m^* a_m = 0 \;\; \textit{(optimalitate)}$$
$$\nabla_\mu \mathcal{L}(x^*, \mu^*) := Ax^* - b = 0. \;\; \textit{(fezabilitate)}$$

Dacă f convexă, atunci condiţiile de mai sus sunt suficiente.

Demonstraţie pe scurt [$rankA = p < n$]:

$$x^* = \pi_Q(x^* - \nabla f(x^*)) = \arg\min_{Az=b} \|z - x^* + \nabla f(x^*)\|$$
$$= x^* + \arg\min_{As=0} \|s + \nabla f(x^*)\|.$$

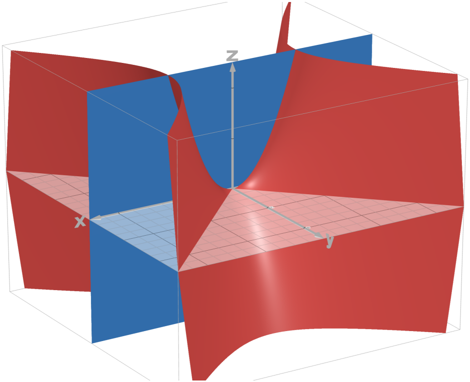
Reamintim: $v = \pi_{Ker(A)}(v) + \pi_{Im(A^T)}(v)$

$$0 = \arg\min_{As=0} \|s + \nabla f(x^*)\| \Rightarrow -\nabla f(x^*) \in Im(A^T)$$

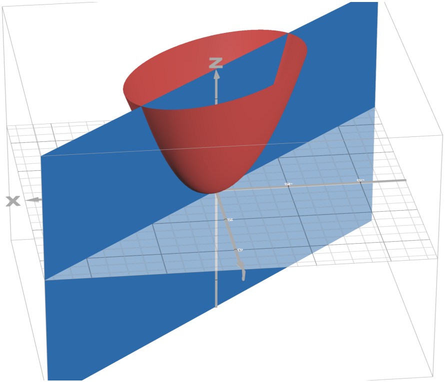
Concluzie:

$$\exists \mu^* \in \mathbb{R}^m : \quad -\nabla f(x^*) = A^T \mu^*.$$

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



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



$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|x - y\|_2^2$$
$$\text{s.l. } a^T x = b,$$

unde $a \in \mathbb{R}^n, b \in \mathbb{R}$.

$$\min_{x \in \mathbb{R}^n} \ \frac{1}{2} x^T H x + q^T x$$
$$\text{s.l.} \ \ Ax = b,$$

unde $H \succeq 0, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

Condiţii de optimalitate:

$$Hx^* + q + A^T \mu^* = 0 \ \ \text{(optimalitate)}$$
$$Ax^* = b. \ \ \text{(fezabilitate)}$$

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^T H x + q^T x \\ \text{s.l.} \ & Ax = b, \end{aligned}$$

unde $H \succeq 0, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

Condiţii de optimalitate (Sistem Kuhn-Tucker):

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \mu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

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

- f, g_i sunt funcţii netede

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

- f, g_i sunt funcţii netede
- $Q = \{x \in \mathbb{R}^n : g_i(x) = 0, \forall i = 1, \cdots, m\} \Rightarrow$ proiecţie dificilă

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

- f, g_i sunt funcţii netede
- $Q = \{x \in \mathbb{R}^n : g_i(x) = 0, \forall i = 1, \cdots, m\} \Rightarrow$ proiecţie dificilă
- x^* minim local: $f(x^*) \leq f(x)$ pentru $x \in Q$ în vecinătatea lui x^*

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

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

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

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

- μ multiplicatori Lagrange (variabile duale)

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

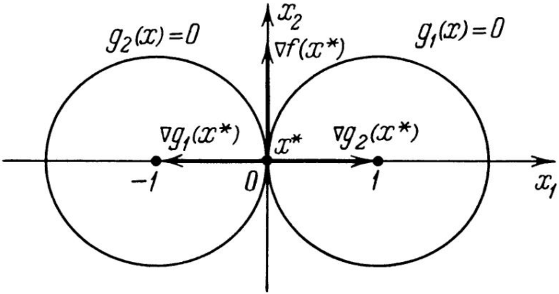
Definiţie

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

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

$$\min_{x \in \mathbb{R}^2} \; x_2$$

$$\text{s.l.} \;\; (x_1 - 1)^2 + x_2^2 - 1 = 0, \;\; (x_1 + 1)^2 + x_2^2 - 1 = 0$$



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

Teoremă (Multiplicatori Lagrange)

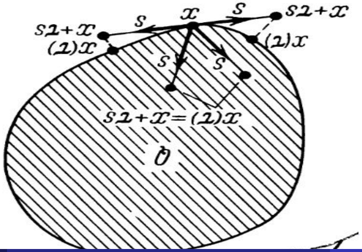
Dacă x^* minim regulat, atunci există μ^* astfel încât:

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

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

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



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

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, $\nabla^2_x \mathcal{L}(x^*, \mu^*)$ 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. $x^T A x = 1$

Cuprins

- Constrângeri de egalitate
- Condiţii de optimalitate
- Filtrare de trend

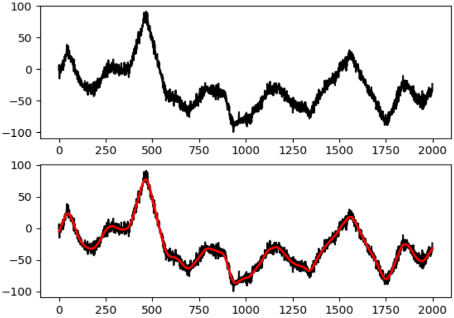
Fie seria de timp $\{y_t\}_{t=1}^n \in \mathbb{R}^n$, presupunem descompunerea:

$$y_t = x_t + \epsilon_t,$$

- x_t reprezintă componenta-trend de variaţie lentă
- ϵ_t componenta zgomot de variaţie rapidă

Filtrare a trend-ului = estimarea componentei "netede" x_t

Serii de timp: Filtrare trend



$$\min_x \ \frac{1}{2}\sum_{t=1}^{n}(x_t - y_t)_2^2 + \rho\sum_{t=2}^{n-1}(x_{t-1} - 2x_t + x_{t+1})^2. \tag{1}$$

- Funcţia obiectiv urmăreşte reducerea reziduului $\{y_t - x_t\}$ şi, în acelaşi timp, ajustarea ”netezimii” lui x_t.
- Diferenţa de ordin 2: $x_{t-1} - 2x_t + x_{t+1}$ este nulă dacă şi numai dacă $\{x_{t-1}, x_t, x_{t+1}\}$ sunt coliniare.

În forma restrânsă:

$$x^*_{HP} := \operatorname*{argmin}_x \frac{1}{2}\|x-y\|_2^2 + \rho\|Dx\|_2^2, \quad \text{unde} \quad D \in \mathbb{R}^{(n-2)\times n}$$

$$D = \begin{bmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \vdots & \vdots & \vdots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1 \end{bmatrix}$$

Soluţia modelului HP are forma:

$$x^*_{HP} := \left(I + 2\rho D^T D\right)^{-1} y.$$

Eroarea relativă de estimare satisface:

$$\|y - x^*_{HP}\|_2 \le \frac{32\rho}{1+32\rho}\|y\|_2$$

$$x^*_{\ell_1} := \operatorname*{argmin}_x \ \frac{1}{2}\|x - y\|_2^2 + \rho\|Dx\|_1$$

- $\|\cdot\|_2^2$ devine $\|\cdot\|_1$
- Dacă $D = I_n$ atunci problema este separabilă:
  $x^*_t = \operatorname{argmin}_{x_t \in \mathbb{R}} \ \frac{1}{2}(x_t - y_t)_2^2 + \lambda|x_t|$

Reformulăm:

$$x^*_{\ell_1} := \operatorname*{argmin}_x \ \frac{1}{2}\|x - y\|_2^2 + \rho\|z\|_1$$
$$\text{s.l. } Dx = z$$

Funcţia Lagrangian:

$$\mathcal{L}(x, z, \lambda) := \frac{1}{2}\|x - y\|_2^2 + \rho\|z\|_1 + \lambda^T(Dx - z)$$

Observăm:

$$[z(x,\lambda)]_i := \begin{cases} 0, & \text{dacă } \ -\rho \le \lambda_i \le \rho \\ -\infty, & \text{dacă } \ |\lambda_i| > \rho. \end{cases}$$

Condiţii de optimalitate:

$$x^*_{\ell_1} = y - D^T \lambda^*$$

Problema duală:

$$\lambda^* = \arg\max_{\lambda} \; -\frac{1}{2}\|D^T\lambda\|_2^2 + \lambda^T D y$$
$$\text{s.l.} \; -\rho \leq \lambda_i \leq \rho \quad \forall i = 1, \cdots, n-2.$$

Filtru ℓ_1: rezultate