How to globally solve non-convex optimization problems involving an approximate ℓ_0 penalization.

A. Marmin, M. Castella, J.C. Pesquet.
ICASSP 2019. IEEE.

Seminario de Optimización y Aprendizaje Automático.

Matías Valdés - IMERL - Fing

15/05/25

Regresión lineal esparsa

Regresión lineal esparsa

Buscamos una solución esparsa de:

$$\arg\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2, \quad A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m.$$

Regresión lineal esparsa

Buscamos una solución esparsa de:

$$\arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2, \quad A\in\mathbb{R}^{m\times n},\ b\in\mathbb{R}^m.$$

Situación ideal (penaliza soluciones poco esparsas):

$$\arg\min_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_0\right\},\ \|x\|_0=\#\{i\in\{1,\dots,n\}\ /\ x_i\neq 0\}.$$

Regresión lineal esparsa

Buscamos una solución esparsa de:

$$\arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2, \quad A\in\mathbb{R}^{m\times n},\ b\in\mathbb{R}^m.$$

Situación ideal (penaliza soluciones poco esparsas):

$$\arg\min_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_0\right\},\ \|x\|_0=\#\{i\in\{1,\ldots,n\}\ /\ x_i\neq 0\}.$$

Difícil de resolver.

Regresión lineal esparsa

Buscamos una solución esparsa de:

$$\arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2, \quad A\in\mathbb{R}^{m\times n},\ b\in\mathbb{R}^m.$$

Situación ideal (penaliza soluciones poco esparsas):

$$\arg\min_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_0\right\},\ \|x\|_0=\#\{i\in\{1,\dots,n\}\ /\ x_i\neq 0\}.$$

Difícil de resolver. Alternativa **convexa** (que fomenta esparsidad):

$$\arg\min_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_1\right\}, \quad \|x\|_1:=\sum_{i=1}^n |x_i| \quad \text{convexa.}$$

Regresión lineal esparsa

Buscamos una solución esparsa de:

$$\arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2, \quad A\in\mathbb{R}^{m\times n},\ b\in\mathbb{R}^m.$$

Situación ideal (penaliza soluciones poco esparsas):

$$\arg\min_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_0\right\},\ \|x\|_0=\#\{i\in\{1,\dots,n\}\ /\ x_i\neq 0\}.$$

Difícil de resolver. Alternativa **convexa** (que fomenta esparsidad):

$$\arg\min_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_1\right\}, \quad \|x\|_1 := \sum_{i=1}^n |x_i| \quad \text{convexa.}$$

No es una buena aproximación de ℓ_0. En particular penaliza soluciones con coordenadas grandes (bias en soluciones).

Alternativa: aproximaciones no convexas de ℓ_0

$$\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{i=1}^{n} R(x_i)\right\}.$$

Alternativa: aproximaciones no convexas de ℓ_0

$$\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{i=1}^{n}R(x_i)\right\}.$$

Buscamos R que se aproxime a ℓ_0.

Alternativa: aproximaciones no convexas de ℓ_0

$$\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{i=1}^{n}R(x_i)\right\}.$$

Buscamos R que se aproxime a ℓ_0. Le pedimos que:

Alternativa: aproximaciones no convexas de ℓ_0

$$\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{i=1}^n R(x_i)\right\}.$$

Buscamos R que se aproxime a ℓ_0. Le pedimos que:

- fomente la esparsidad,

Alternativa: aproximaciones no convexas de ℓ_0

$$\min_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\|Ax-b\|_2^2 + \sum_{i=1}^{n} R(x_i) \right\}.$$

Buscamos R que se aproxime a ℓ_0. Le pedimos que:

- fomente la esparsidad,
- no penalice valores grandes de las coordenadas,

Alternativa: aproximaciones no convexas de ℓ_0

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \sum_{i=1}^{n} R(x_i) \right\}.$$

Buscamos R que se aproxime a ℓ_0. Le pedimos que:

- fomente la esparsidad,
- no penalice valores grandes de las coordenadas,
- sea continua (estabilidad del modelo).

Alternativa: aproximaciones no convexas de ℓ_0

$$\min_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\|Ax-b\|_2^2 + \sum_{i=1}^n R(x_i) \right\}.$$

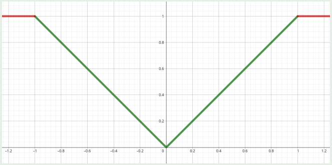
Buscamos R que se aproxime a ℓ_0. Le pedimos que:

- fomente la esparsidad,
- no penalice valores grandes de las coordenadas,
- sea continua (estabilidad del modelo).

Example (Capped ℓ_1)

$$R(x) = \begin{cases} |x|, & |x| \le \lambda \\ \lambda, & |x| > \lambda \end{cases} =$$

$$= |x| 1_{\{|x|\le\lambda\}} + \lambda 1_{\{|x|>\lambda\}}.$$



Ejemplos de funciones de penalización

* Capped ℓ_p:

$$R(x) = |x|^p 1_{\{|x| \leq \lambda\}} + \lambda^p 1_{\{|x| > \lambda\}}, \;\; p > 0$$

Ejemplos de funciones de penalización

* Capped ℓ_p:

$$R(x) = |x|^p 1_{\{|x| \leq \lambda\}} + \lambda^p 1_{\{|x| > \lambda\}}, \;\; p > 0$$

* Smoothly Clipped Absolute Deviation (SCAD) [Fan, Li, 2001]:

$$R(x) = \lambda|x|1_{\{|x|\leq\lambda\}} - \left(\frac{\lambda^2 - 2\gamma\lambda|x| + x^2}{2(\gamma-1)}\right) 1_{\{\lambda<|x|\leq\gamma\lambda\}} + \frac{(\gamma+1)\lambda^2}{2} 1_{\{|x|>\gamma\lambda\}}$$

Ejemplos de funciones de penalización

* Capped ℓ_p:

$$R(x) = |x|^p 1_{\{|x|\leq\lambda\}} + \lambda^p 1_{\{|x|>\lambda\}}, \;\; p > 0$$

* Smoothly Clipped Absolute Deviation (SCAD) [Fan, Li, 2001]:

$$R(x) = \lambda|x|1_{\{|x|\leq\lambda\}} - \left(\frac{\lambda^2 - 2\gamma\lambda|x| + x^2}{2(\gamma-1)}\right) 1_{\{\lambda<|x|\leq\gamma\lambda\}} + \frac{(\gamma+1)\lambda^2}{2} 1_{\{|x|>\gamma\lambda\}}$$

* Minimax Concave Penalty (MCP) [Zhang, 2010]:

$$R(x) = \left(\lambda|x| - \frac{x^2}{2\gamma}\right) 1_{\{|x|\leq\gamma\lambda\}} + \frac{\gamma\lambda^2}{2} 1_{\{|x|>\gamma\lambda\}}, \quad \gamma > 0$$

Ejemplos de funciones de penalización

* Capped ℓ_p:

$$R(x) = |x|^p 1_{\{|x| \leq \lambda\}} + \lambda^p 1_{\{|x| > \lambda\}}, \;\; p > 0$$

* Smoothly Clipped Absolute Deviation (SCAD) [Fan, Li, 2001]:

$$R(x) = \lambda|x| 1_{\{|x| \leq \lambda\}} - \left(\frac{\lambda^2 - 2\gamma\lambda|x| + x^2}{2(\gamma - 1)} \right) 1_{\{\lambda < |x| \leq \gamma\lambda\}} + \frac{(\gamma + 1)\lambda^2}{2} 1_{\{|x| > \gamma\lambda\}}$$

* Minimax Concave Penalty (MCP) [Zhang, 2010]:

$$R(x) = \left(\lambda|x| - \frac{x^2}{2\gamma} \right) 1_{\{|x| \leq \gamma\lambda\}} + \frac{\gamma\lambda^2}{2} 1_{\{|x| > \gamma\lambda\}}, \quad \gamma > 0$$

* Continuous Exact ℓ_0 (CEL0) [Soubies, Blanc, Aubert, 2015]:

$$R(x) = \lambda - \frac{\gamma^2}{2} \left(|x| - \frac{\sqrt{2\lambda}}{\gamma} \right)^2 1_{\left\{|x| \leq \frac{\sqrt{2\lambda}}{\gamma}\right\}}, \quad \gamma > 0$$

Ejemplos de funciones de penalización

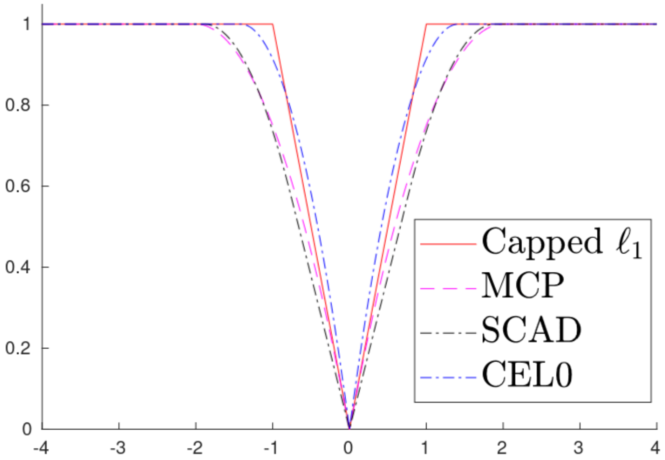


Fig. 1. Examples of continuous relaxation of ℓ_0 penalization ($\lambda = 1$, $\gamma_{\text{SCAD}} = 2.5$, $\gamma_{\text{MCP}} = 2$, $\gamma_{\text{CEL0}} = 1$).

Funciones de penalización

Los ejemplos anteriores son funciones polinomiales a trozos.

Funciones de penalización

Los ejemplos anteriores son funciones polinomiales a trozos.

Idea del artículo:

Funciones de penalización

Los ejemplos anteriores son funciones polinomiales a trozos.

Idea del artículo:

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \sum_{i=1}^{n} R(x_i) \right\}.$$

Funciones de penalización

Los ejemplos anteriores son funciones polinomiales a trozos.

Idea del artículo:

$$\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{i=1}^{n} R(x_i)\right\}.$$

1. Reemplazar la penalización polinomial a trozos R por un polinomio y ciertas restricciones polinomiales.

Funciones de penalización

Los ejemplos anteriores son funciones polinomiales a trozos.

Idea del artículo:

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \sum_{i=1}^{n} R(x_i) \right\}.$$

1. Reemplazar la penalización polinomial a trozos R por un polinomio y ciertas restricciones polinomiales.
2. Esto da un problema de optimización polinomial, para el que existen métodos de optimización **global**.

Funciones de penalización

Los ejemplos anteriores son funciones polinomiales a trozos.

Idea del artículo:

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \sum_{i=1}^{n} R(x_i) \right\}.$$

1. Reemplazar la penalización polinomial a trozos R por un polinomio y ciertas restricciones polinomiales.
2. Esto da un problema de optimización polinomial, para el que existen métodos de optimización **global**.
3. Resolver el problema de optimización polinomial.

Penalización polinomial a trozos

Una función polinomial a trozos es de la forma:

$$R(x) = \sum_{i=1}^{N} g_i(x) 1_{\{\sigma_{i-1} \leq x < \sigma_i\}}, \quad x \in \mathbb{R};$$

con g_i polinomios, y σ_i una secuencia creciente de escalares:

$$-\infty \leq \sigma_0 < \sigma_1 < \ldots < \sigma_{N-1} < \sigma_N \leq \infty.$$

Penalización polinomial a trozos

Una función polinomial a trozos es de la forma:

$$R(x) = \sum_{i=1}^{N} g_i(x) 1_{\{\sigma_{i-1} \leq x < \sigma_i\}}, \quad x \in \mathbb{R};$$

con g_i polinomios, y σ_i una secuencia creciente de escalares:

$$-\infty \leq \sigma_0 < \sigma_1 < \ldots < \sigma_{N-1} < \sigma_N \leq \infty.$$

Example (Capped ℓ_p con $\lambda = 1$: $R(x) = |x|^p 1_{\{|x| \leq 1\}} + 1_{\{|x| > 1\}}$)

Penalización polinomial a trozos

Una función polinomial a trozos es de la forma:

$$R(x) = \sum_{i=1}^{N} g_i(x) 1_{\{\sigma_{i-1} \leq x < \sigma_i\}}, \quad x \in \mathbb{R};$$

con g_i polinomios, y σ_i una secuencia creciente de escalares:

$$-\infty \leq \sigma_0 < \sigma_1 < \ldots < \sigma_{N-1} < \sigma_N \leq \infty.$$

Example (Capped ℓ_p con $\lambda = 1$: $R(x) = |x|^p 1_{\{|x| \leq 1\}} + 1_{\{|x|>1\}}$)

$$\sigma_0 = -\infty,\ \sigma_1 = -1,\ \sigma_2 = 0,\ \sigma_3 = 1,\ \sigma_4 = \infty.$$

Penalización polinomial a trozos

Una función polinomial a trozos es de la forma:

$$R(x) = \sum_{i=1}^{N} g_i(x) 1_{\{\sigma_{i-1} \le x < \sigma_i\}}, \quad x \in \mathbb{R};$$

con g_i polinomios, y σ_i una secuencia creciente de escalares:

$$-\infty \le \sigma_0 < \sigma_1 < \ldots < \sigma_{N-1} < \sigma_N \le \infty.$$

Example (Capped ℓ_p con $\lambda = 1$: $R(x) = |x|^p 1_{\{|x| \le 1\}} + 1_{\{|x| > 1\}}$)

$$\sigma_0 = -\infty,\ \sigma_1 = -1,\ \sigma_2 = 0,\ \sigma_3 = 1,\ \sigma_4 = \infty.$$

$$g_0(x) = 1,\ g_1(x) = (-x)^p,\ g_2(x) = x^p,\ g_3(x) = 1.$$

Penalización polinomial a trozos

Una función polinomial a trozos es de la forma:

$$R(x) = \sum_{i=1}^{N} g_i(x) 1_{\{\sigma_{i-1} \leq x < \sigma_i\}}, \quad x \in \mathbb{R};$$

con g_i polinomios, y σ_i una secuencia creciente de escalares:

$$-\infty \leq \sigma_0 < \sigma_1 < \ldots < \sigma_{N-1} < \sigma_N \leq \infty.$$

Example (Capped ℓ_p con $\lambda = 1$: $R(x) = |x|^p 1_{\{|x| \leq 1\}} + 1_{\{|x|>1\}}$)

$$\sigma_0 = -\infty,\ \sigma_1 = -1,\ \sigma_2 = 0,\ \sigma_3 = 1,\ \sigma_4 = \infty.$$

$$g_0(x) = 1,\ g_1(x) = (-x)^p,\ g_2(x) = x^p,\ g_3(x) = 1.$$

$$R(x) = 1_{\{-\infty \leq x < -1\}} + (-x)^p 1_{\{-1 \leq x < 0\}} + x^p 1_{\{0 \leq x < 1\}} + 1_{\{1 \leq x < \infty\}}.$$

Forma polinomial equivalente

Proposición

Polinomial a trozos $\Leftrightarrow$ *polinomio y restricciones polinomiales.*

Forma polinomial equivalente

Proposición

Polinomial a trozos $\Leftrightarrow$ *polinomio y restricciones polinomiales.*

Se definen $z^{(i)}$, tales que:

$$z^{(i)} = \begin{cases} 1, & \sigma_i \leq x \\ 0, & \sigma_i > x \end{cases} = 1_{\{\sigma_i \leq x\}}, \quad \forall\ i = 1, \ldots, N.$$

Forma polinomial equivalente

Proposición

Polinomial a trozos ⇔ polinomio y restricciones polinomiales.

Se definen $z^{(i)}$, tales que:

$$z^{(i)} = \begin{cases} 1, & \sigma_i \leq x \\ 0, & \sigma_i > x \end{cases} = 1_{\{\sigma_i \leq x\}}, \quad \forall\ i = 1, \ldots, N.$$

Se cumple: $z^{(i)} = 1_{\{\sigma_i \leq x\}} \Leftrightarrow \begin{cases} z^{(i)} \left(1 - z^{(i)}\right) = 0, \\ \left(z^{(i)} - \frac{1}{2}\right)(x - \sigma_i) \geq 0 \end{cases}.$

Forma polinomial equivalente

Proposición

Polinomial a trozos $\Leftrightarrow$ polinomio y restricciones polinomiales.

Se definen $z^{(i)}$, tales que:

$$z^{(i)} = \begin{cases} 1, & \sigma_i \leq x \\ 0, & \sigma_i > x \end{cases} = 1_{\{\sigma_i \leq x\}}, \quad \forall\ i = 1, \ldots, N.$$

Se cumple: $z^{(i)} = 1_{\{\sigma_i \leq x\}} \Leftrightarrow \begin{cases} z^{(i)} \left(1 - z^{(i)}\right) = 0, \\ \left(z^{(i)} - \frac{1}{2}\right)(x - \sigma_i) \geq 0 \end{cases}.$

Por otro lado: $1_{\{\sigma_{i-1} \leq x < \sigma_i\}} = z^{(i-1)} \left(1 - z^{(i)}\right)$.

Forma polinomial equivalente

Proposición

Polinomial a trozos $\Leftrightarrow$ polinomio y restricciones polinomiales.

Se definen $z^{(i)}$, tales que:

$$z^{(i)} = \begin{cases} 1, & \sigma_i \leq x \\ 0, & \sigma_i > x \end{cases} = 1_{\{\sigma_i \leq x\}}, \quad \forall\ i = 1, \ldots, N.$$

Se cumple: $z^{(i)} = 1_{\{\sigma_i \leq x\}} \Leftrightarrow \begin{cases} z^{(i)} \left(1 - z^{(i)}\right) = 0, \\ \left(z^{(i)} - \frac{1}{2}\right)(x - \sigma_i) \geq 0 \end{cases}.$

Por otro lado: $1_{\{\sigma_{i-1} \leq x < \sigma_i\}} = z^{(i-1)} \left(1 - z^{(i)}\right)$. Por lo tanto:

$$R(x) = \sum_{i=1}^{N} g_i(x) 1_{\{\sigma_{i-1} \leq x < \sigma_i\}} = \sum_{i=1}^{N} g_i(x) z^{(i-1)} \left(1 - z^{(i)}\right);$$

con $z^{(i)}$ tal que: $z^{(i)} \left(1 - z^{(i)}\right) = 0$, y $\left(z^{(i)} - \frac{1}{2}\right)(x - \sigma_i) \geq 0$, $\forall\ i$.

Ejemplo: Capped ℓ_p con $\lambda = 1$

Ejemplo: Capped ℓ_p con $\lambda = 1$

Recordar:

$$z^{(i)} = \begin{cases} 1, & \sigma_i \leq x \\ 0, & \sigma_i > x \end{cases} = 1_{\{\sigma_i \leq x\}}.$$

Ejemplo: Capped ℓ_p con $\lambda = 1$

Recordar:

$$z^{(i)} = \begin{cases} 1, & \sigma_i \leq x \\ 0, & \sigma_i > x \end{cases} = 1_{\{\sigma_i \leq x\}}.$$

En el caso general, esto implica:

$$\sigma_0 = -\infty \Rightarrow z^{(0)}(x) = 1, \ \forall\ x, \quad \sigma_N = \infty \Rightarrow z^{(N)}(x) = 0, \ \forall\ x.$$

Ejemplo: Capped ℓ_p con $\lambda = 1$

Recordar:

$$z^{(i)} = \begin{cases} 1, & \sigma_i \leq x \\ 0, & \sigma_i > x \end{cases} = 1_{\{\sigma_i \leq x\}}.$$

En el caso general, esto implica:

$$\sigma_0 = -\infty \Rightarrow z^{(0)}(x) = 1, \ \forall \ x, \quad \sigma_N = \infty \Rightarrow z^{(N)}(x) = 0, \ \forall \ x.$$

Para Capped ℓ_p, con $\lambda = 1$: $R(x) = |x|^p 1_{\{|x| \leq 1\}} + 1_{\{|x|>1\}} \Leftrightarrow$

$$R(x) = (1 - z^{(1)}) + (-x)^p z^{(1)}(1 - z^{(2)}) + x^p z^{(2)}(1 - z^{(3)}) + z^{(3)}$$

$$z^{(i)}(1 - z^{(i)}) = 0, \quad \left(z^{(i)} - 0.5\right)(x - \sigma_i) \geq 0, \quad \forall \ i = 1, 2, 3$$

Optimización polinomial

Usando lo anterior (en cada coordenada x_t), el problema original

$$\arg\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{t=1}^{n}R(x_t)\right\},$$

equivale a un problema de optimización polinomial:

Optimización polinomial

Usando lo anterior (en cada coordenada x_t), el problema original

$$\arg\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{t=1}^{n}R(x_t)\right\},$$

equivale a un problema de optimización polinomial:

$$\arg\min_{x,z}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{t=1}^{n}\sum_{i=1}^{N}g_i(x_t)z_t^{(i-1)}\left(1-z_t^{(i)}\right)\right\},\quad s.t.$$

$$z_t^{(i)}\left(1-z_t^{(i)}\right)=0,\quad \left(z_t^{(i)}-\frac{1}{2}\right)(x_t-\sigma_i)\geq 0,\quad \forall\, i=1:N, t=1:n.$$

Optimización polinomial

Usando lo anterior (en cada coordenada x_t), el problema original

$$\arg\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{t=1}^{n}R(x_t)\right\},$$

equivale a un problema de optimización polinomial:

$$\arg\min_{x,z}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{t=1}^{n}\sum_{i=1}^{N}g_i(x_t)z_t^{(i-1)}\left(1-z_t^{(i)}\right)\right\},\quad s.t.$$

$$z_t^{(i)}\left(1-z_t^{(i)}\right)=0,\quad \left(z_t^{(i)}-\frac{1}{2}\right)(x_t-\sigma_i)\geq 0,\quad \forall\, i=1:N, t=1:n.$$

Existen métodos para estimar un mínimo global:

Optimización polinomial

Usando lo anterior (en cada coordenada x_t), el problema original

$$\arg\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{t=1}^{n}R(x_t)\right\},$$

equivale a un problema de optimización polinomial:

$$\arg\min_{x,z}\left\{\frac{1}{2}\|Ax-b\|_2^2+\sum_{t=1}^{n}\sum_{i=1}^{N}g_i(x_t)z_t^{(i-1)}\left(1-z_t^{(i)}\right)\right\},\quad s.t.$$

$$z_t^{(i)}\left(1-z_t^{(i)}\right)=0,\quad \left(z_t^{(i)}-\frac{1}{2}\right)(x_t-\sigma_i)\geq 0,\quad \forall\, i=1:N, t=1:n.$$

Existen métodos para estimar un mínimo global:

- RLT: Reformulation Linearization Technique (LPs con B.B.).

Optimización polinomial

Usando lo anterior (en cada coordenada x_t), el problema original

$$\arg\min_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\|Ax-b\|_2^2 + \sum_{t=1}^{n} R(x_t) \right\},$$

equivale a un problema de optimización polinomial:

$$\arg\min_{x,z} \left\{ \frac{1}{2}\|Ax-b\|_2^2 + \sum_{t=1}^{n}\sum_{i=1}^{N} g_i(x_t) z_t^{(i-1)} \left(1 - z_t^{(i)}\right) \right\}, \quad s.t.$$

$$z_t^{(i)}\left(1 - z_t^{(i)}\right) = 0, \quad \left(z_t^{(i)} - \frac{1}{2}\right)(x_t - \sigma_i) \geq 0, \quad \forall\, i = 1:N, t = 1:n.$$

Existen métodos para estimar un mínimo global:

- RLT: Reformulation Linearization Technique (LPs con B.B.).
- Lasserre-Parrilo (sucesión de SDPs, asociadas a SOS).

Simulaciones: modelo

Se generan datos usando el modelo $y = Ax + w$, con:

Simulaciones: modelo

Se generan datos usando el modelo $y = Ax + w$, con:

- w ruido gaussiano blanco,

Simulaciones: modelo

Se generan datos usando el modelo $y = Ax + w$, con:

- w ruido gaussiano blanco,
- A matriz de tipo banda Toeplitz: $A = \begin{pmatrix} a_2 & a_1 & 0 & 0 \\ a_3 & a_2 & a_1 & 0 \\ 0 & a_3 & a_2 & a_1 \\ 0 & 0 & a_3 & a_2 \\ 0 & 0 & 0 & a_3 \end{pmatrix}$

Simulaciones: modelo

Se generan datos usando el modelo $y = Ax + w$, con:

- w ruido gaussiano blanco,
- A matriz de tipo banda Toeplitz: $A = \begin{pmatrix} a_2 & a_1 & 0 & 0 \\ a_3 & a_2 & a_1 & 0 \\ 0 & a_3 & a_2 & a_1 \\ 0 & 0 & a_3 & a_2 \\ 0 & 0 & 0 & a_3 \end{pmatrix}$

  (correspondiente a un filtro de convolución gaussiana),

Simulaciones: modelo

Se generan datos usando el modelo $y = Ax + w$, con:

- w ruido gaussiano blanco,
- A matriz de tipo banda Toeplitz: $A = \begin{pmatrix} a_2 & a_1 & 0 & 0 \\ a_3 & a_2 & a_1 & 0 \\ 0 & a_3 & a_2 & a_1 \\ 0 & 0 & a_3 & a_2 \\ 0 & 0 & 0 & a_3 \end{pmatrix}$

  (correspondiente a un filtro de convolución gaussiana),
- $x \in \mathbb{R}^n$, $n = 200$, con coordenadas $x_t \sim U[6/10, 1]$.

Simulaciones: modelo

Se generan datos usando el modelo $y = Ax + w$, con:

- w ruido gaussiano blanco,
- A matriz de tipo banda Toeplitz: $A = \begin{pmatrix} a_2 & a_1 & 0 & 0 \\ a_3 & a_2 & a_1 & 0 \\ 0 & a_3 & a_2 & a_1 \\ 0 & 0 & a_3 & a_2 \\ 0 & 0 & 0 & a_3 \end{pmatrix}$ (correspondiente a un filtro de convolución gaussiana),
- $x \in \mathbb{R}^n$, $n = 200$, con coordenadas $x_t \sim U[6/10, 1]$.

Parámetro γ de las funciones de penalización $R(x_t)$:

- MCP: $\gamma = 0.5$; SCAD: $\gamma = 2.1$;
- CEL0: norma de la columna t de A (para cada x_t).

Compara con tres algoritmos de optimización local:

Compara con tres algoritmos de optimización local:

- Forward-Backward (FB): criterio "greedy" para incorporar (forward) o quitar (backward) coordenadas no nulas al modelo, según variación de la función objetivo.

Compara con tres algoritmos de optimización local:

- Forward-Backward (FB): criterio “greedy” para incorporar (forward) o quitar (backward) coordenadas no nulas al modelo, según variación de la función objetivo.
- Iteratively Re-weighted ℓ_1 (IRL1):

$$x^{(k)} = \arg\min_{x} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|W^{(k)}x\|_1 \right\},$$

$$w_t^{(k+1)} = \frac{1}{|x_t^{(k)}| + \epsilon}, \quad \forall\ t = 1, \ldots, n.$$

Compara con tres algoritmos de optimización local:

- Forward-Backward (FB): criterio "greedy" para incorporar (forward) o quitar (backward) coordenadas no nulas al modelo, según variación de la función objetivo.
- Iteratively Re-weighted ℓ_1 (IRL1):

$$x^{(k)} = \arg\min_{x} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|W^{(k)}x\|_1 \right\},$$

$$w_t^{(k+1)} = \frac{1}{|x_t^{(k)}| + \epsilon}, \quad \forall\ t = 1, \ldots, n.$$

- Descenso por Coordenadas (CD).

Compara con tres algoritmos de optimización local:

- Forward-Backward (FB): criterio "greedy" para incorporar (forward) o quitar (backward) coordenadas no nulas al modelo, según variación de la función objetivo.
- Iteratively Re-weighted ℓ_1 (IRL1):

$$x^{(k)} = \arg\min_{x} \left\{ \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|W^{(k)} x\|_1 \right\},$$

$$w_t^{(k+1)} = \frac{1}{|x_t^{(k)}| + \epsilon}, \quad \forall\ t = 1, \ldots, n.$$

- Descenso por Coordenadas (CD).

El algoritmo propuesto utiliza orden de jerarquía SDP $k = 3$.

Resultados experimentales

- Tres condiciones iniciales: aleatoria, cero, y el valor real $\bar{x}$.

Resultados experimentales

- Tres condiciones iniciales: aleatoria, cero, y el valor real $\bar{x}$.
- Compara métodos evaluando sol. en: $\frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{10}\|x\|_0$.

Resultados experimentales

- Tres condiciones iniciales: aleatoria, cero, y el valor real $\bar{x}$.
- Compara métodos evaluando sol. en: $\frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{10}\|x\|_0$.

Table 1. Optimal criterion value depending on initial point.

Alg. \ Ψ_λ	Capped ℓ_1	SCAD	MCP	CEL0
Proposed ($k = 3$)	**3.725**	**3.506**	**2.749**	**4.425**
Random $\mathbf{x}_{\text{init}}$				
FB	4.372	3.759	3.194	4.638
IRL1	7.052	4.632	3.381	6.926
CD	4.099	3.927	3.520	4.638
$\mathbf{x}_{\text{init}} = 0$				
FB	4.208	3.763	3.149	4.638
IRL1	4.839	4.566	3.013	6.879
CD	5.455	4.371	3.717	5.316
$\mathbf{x}_{\text{init}} = \overline{\mathbf{x}}$				
FB	3.867	3.639	2.871	4.637
IRL1	4.766	4.567	2.914	6.874
CD	4.093	3.639	3.015	5.365

Resultados experimentales

- Tres condiciones iniciales: aleatoria, cero, y el valor real $\bar{x}$.
- Compara métodos evaluando sol. en: $\frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{10}\|x\|_0$.

Table 1. Optimal criterion value depending on initial point.

Alg. \ Ψ_λ	Capped ℓ_1	SCAD	MCP	CEL0
Proposed ($k = 3$)	**3.725**	**3.506**	**2.749**	**4.425**
Random $\mathbf{x}_{\text{init}}$				
FB	4.372	3.759	3.194	4.638
IRL1	7.052	4.632	3.381	6.926
CD	4.099	3.927	3.520	4.638
$\mathbf{x}_{\text{init}} = 0$				
FB	4.208	3.763	3.149	4.638
IRL1	4.839	4.566	3.013	6.879
CD	5.455	4.371	3.717	5.316
$\mathbf{x}_{\text{init}} = \overline{\mathbf{x}}$				
FB	3.867	3.639	2.871	4.637
IRL1	4.766	4.567	2.914	6.874
CD	4.093	3.639	3.015	5.365

- Resultado de métodos locales varía con la condición inicial.

Resultados experimentales

- Tres condiciones iniciales: aleatoria, cero, y el valor real $\bar{x}$.
- Compara métodos evaluando sol. en: $\frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{10}\|x\|_0$.

Table 1. Optimal criterion value depending on initial point.

Alg. \ Ψ_λ	Capped ℓ_1	SCAD	MCP	CEL0
Proposed ($k = 3$)	**3.725**	**3.506**	**2.749**	**4.425**
Random $\mathbf{x}_{\text{init}}$				
FB	4.372	3.759	3.194	4.638
IRL1	7.052	4.632	3.381	6.926
CD	4.099	3.927	3.520	4.638
$\mathbf{x}_{\text{init}} = 0$				
FB	4.208	3.763	3.149	4.638
IRL1	4.839	4.566	3.013	6.879
CD	5.455	4.371	3.717	5.316
$\mathbf{x}_{\text{init}} = \overline{\mathbf{x}}$				
FB	3.867	3.639	2.871	4.637
IRL1	4.766	4.567	2.914	6.874
CD	4.093	3.639	3.015	5.365

- Resultado de métodos locales varía con la condición inicial.
- El método propuesto es siempre el de menor valor.

Optimización polinomial con Lasserre-Parrilo (2001)

Optimización polinomial con Lasserre-Parrilo (2001)

En todo problema de optimización, el valor mínimo **global** cumple:

$$p^* := \min_{x \in K} p(x) = \left\{ \max_{x,\lambda} \lambda, \quad \text{s.a: } p(x) - \lambda \geq 0, \; x \in K \right\}.$$

Optimización polinomial con Lasserre-Parrilo (2001)

En todo problema de optimización, el valor mínimo **global** cumple:

$$p^* := \min_{x \in K} p(x) = \left\{ \max_{x,\lambda} \lambda, \quad \text{s.a: } p(x) - \lambda \geq 0, \; x \in K \right\}.$$

Problema polinomial: minimizar polinomio p, en conjunto factible

$$K := \{x \in \mathbb{R}^n \;/\; p_i(x) \geq 0\}, \quad \text{con } p_i \text{ polinomios.}$$

Optimización polinomial con Lasserre-Parrilo (2001)

En todo problema de optimización, el valor mínimo **global** cumple:

$$p^* := \min_{x \in K} p(x) = \left\{ \max_{x,\lambda} \lambda, \quad \text{s.a: } p(x) - \lambda \geq 0,\ x \in K \right\}.$$

Problema polinomial: minimizar polinomio p, en conjunto factible

$$K := \{x \in \mathbb{R}^n \ / \ p_i(x) \geq 0\}, \quad \text{con } p_i \text{ polinomios.}$$

Theorem (Putinar, 1993)

Si q es polinomio positivo en $K = \{x \in \mathbb{R}^n \ / \ p_i(x) \geq 0\}$, entonces:

$$q = s_0 + s_1 p_1 + \ldots + s_m p_m, \quad \textit{con } s_i \textit{ suma de cuadrados (SOS).}$$

Optimización polinomial con Lasserre-Parrilo (2001)

En todo problema de optimización, el valor mínimo **global** cumple:

$$p^* := \min_{x \in K} p(x) = \left\{ \max_{x,\lambda} \lambda, \quad \text{s.a: } p(x) - \lambda \geq 0, \ x \in K \right\}.$$

Problema polinomial: minimizar polinomio p, en conjunto factible

$$K := \{x \in \mathbb{R}^n \ / \ p_i(x) \geq 0\}, \quad \text{con } p_i \text{ polinomios.}$$

Theorem (Putinar, 1993)

Si q es polinomio positivo en $K = \{x \in \mathbb{R}^n \ / \ p_i(x) \geq 0\}$, entonces:

$$q = s_0 + s_1 p_1 + \ldots + s_m p_m, \quad \text{con } s_i \text{ suma de cuadrados (SOS).}$$

Relajación en SOS:

$$p_k := \max_{x,\lambda} \lambda, \ \text{s.a: } p(x) - \lambda = s_0 + s_1 p_1(x) + \ldots + s_m p_m(x);$$

con s_i suma de cuadrados, y $\deg(s_i p_i) \leq 2k$.

Optimización polinomial con Lasserre-Parrilo (2001)

Relajación SOS (de orden $2k$):

$$p_k := \max_{x,\lambda} \ \lambda, \quad \text{s.a: } p(x) - \lambda = s_0 + s_1 p_1(x) + \ldots + s_m p_m(x);$$

con s_i suma de cuadrados, y $\deg(s_i p_i) \leq 2k$.

Optimización polinomial con Lasserre-Parrilo (2001)

Relajación SOS (de orden $2k$):

$$p_k := \max_{x,\lambda} \ \lambda, \quad \text{s.a: } p(x) - \lambda = s_0 + s_1 p_1(x) + \ldots + s_m p_m(x);$$

con s_i suma de cuadrados, y $\deg(s_i p_i) \leq 2k$.

- p_k es cota inferior de p^* (por ser relajación),

Optimización polinomial con Lasserre-Parrilo (2001)

Relajación SOS (de orden $2k$):

$$p_k := \max_{x,\lambda} \ \lambda, \quad \text{s.a: } p(x) - \lambda = s_0 + s_1 p_1(x) + \ldots + s_m p_m(x);$$

con s_i suma de cuadrados, y $\deg(s_i p_i) \leq 2k$.

- p_k es cota inferior de p^* (por ser relajación),
- p_k es creciente con el orden k de la relajación.

Optimización polinomial con Lasserre-Parrilo (2001)

Relajación SOS (de orden $2k$):

$$p_k := \max_{x,\lambda} \ \lambda, \quad \text{s.a: } p(x) - \lambda = s_0 + s_1 p_1(x) + \ldots + s_m p_m(x);$$

con s_i suma de cuadrados, y $\deg(s_i p_i) \leq 2k$.

- p_k es cota inferior de p^* (por ser relajación),
- p_k es creciente con el orden k de la relajación.
- Problema SDP surge de la condición de suma de cuadrados.

SOS expresada como Programación Semidefinida (SDP)

SOS expresada como Programación Semidefinida (SDP)

- Optimización lineal (LP):

$$\min_{x \in \mathbb{R}^n} c^T x, \quad \text{s.a: } a_j^T x = b_j,\ j = 1, \ldots, m,\ x_i \geq 0,\ \forall\ i.$$

SOS expresada como Programación Semidefinida (SDP)

- Optimización lineal (LP):

$$\min_{x\in\mathbb{R}^n} c^T x, \quad \text{s.a: } a_j^T x = b_j,\ j = 1,\dots,m,\ x_i \geq 0,\ \forall\ i.$$

- Optimización Semidefinida (SDP):

$$\min_{X\in\mathbb{R}^{n\times n}} tr(C^T X), \quad \text{s.a: } tr(A_j^T X) = b_j,\ j = 1,\dots,m,\ X \succcurlyeq 0.$$

SOS expresada como Programación Semidefinida (SDP)

- Optimización lineal (LP):

$$\min_{x\in\mathbb{R}^n} c^T x, \quad \text{s.a: } a_j^T x = b_j,\ j=1,\dots,m,\ x_i \geq 0,\ \forall\ i.$$

- Optimización Semidefinida (SDP):

$$\min_{X\in\mathbb{R}^{n\times n}} tr(C^T X), \quad \text{s.a: } tr(A_j^T X) = b_j,\ j=1,\dots,m,\ X \succcurlyeq 0.$$

- Ejemplo de SDP: $n=3$, $m=1$:

$$\min_{X\in\mathbb{R}^{3\times 3}} x_{11} + 4x_{12} + 6x_{13} + 9x_{22} + 7x_{33},\ s.a.$$

$$x_{11} + 2x_{13} + 3x_{22} + 14x_{23} + 5x_{33} = 11,\ X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix} \succcurlyeq 0.$$

SOS expresada como Programación Semidefinida (SDP)

- Optimización lineal (LP):

$$\min_{x\in\mathbb{R}^n} c^T x, \quad \text{s.a: } a_j^T x = b_j,\ j=1,\ldots,m,\ x_i \geq 0,\ \forall\ i.$$

- Optimización Semidefinida (SDP):

$$\min_{X\in\mathbb{R}^{n\times n}} tr(C^T X), \quad \text{s.a: } tr(A_j^T X) = b_j,\ j=1,\ldots,m,\ X \succcurlyeq 0.$$

- Ejemplo de SDP: $n=3$, $m=1$:

$$\min_{X\in\mathbb{R}^{3\times 3}} x_{11} + 4x_{12} + 6x_{13} + 9x_{22} + 7x_{33},\ s.a.$$

$$x_{11} + 2x_{13} + 3x_{22} + 14x_{23} + 5x_{33} = 11,\ X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{pmatrix} \succcurlyeq 0.$$

SDP es **convexo**. Solución global con métodos de punto interior.

SOS expresada como Programación Semidefinida (SDP)

SOS expresada como Programación Semidefinida (SDP)

- Sea $[x]_k$ vector con monomios de n variables y grado $\leq k$.

SOS expresada como Programación Semidefinida (SDP)

- Sea $[x]_k$ vector con monomios de n variables y grado $\leq k$.
- Por ejemplo, para $n = 2$ variables, y grado $\leq k = 3$:

$$[x]_3 = [1, x_1, x_2, x_1^2, x_1 x_2, x_2^2, x_1^3, x_1^2 x_2, x_1 x_2^2, x_2^3]^T.$$

SOS expresada como Programación Semidefinida (SDP)

- Sea $[x]_k$ vector con monomios de n variables y grado $\leq k$.
- Por ejemplo, para $n = 2$ variables, y grado $\leq k = 3$:

$$[x]_3 = [1, x_1, x_2, x_1^2, x_1 x_2, x_2^2, x_1^3, x_1^2 x_2, x_1 x_2^2, x_2^3]^T.$$

Theorem

Polinomio p de n variables y grado $\leq 2k$ es suma de cuadrados sii

$$p(x) = [x]_k^T M [x]_k;$$

para alguna matriz $M \in \mathbb{R}^{N \times N}$ semi-definida positiva, $N := C_k^{n+k}$.

SOS expresada como Programación Semidefinida (SDP)

- Sea $[x]_k$ vector con monomios de n variables y grado $\leq k$.
- Por ejemplo, para $n = 2$ variables, y grado $\leq k = 3$:

$$[x]_3 = [1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3]^T.$$

Theorem

Polinomio p de n variables y grado $\leq 2k$ es suma de cuadrados sii

$$p(x) = [x]_k^T M[x]_k;$$

para alguna matriz $M \in \mathbb{R}^{N\times N}$ semi-definida positiva, $N := C_k^{n+k}$.

- Igualando coeficientes de cada monomio, se obtienen restricciones lineales en M_{ij}.

SOS expresada como Programación Semidefinida (SDP)

- Sea $[x]_k$ vector con monomios de n variables y grado $\leq k$.
- Por ejemplo, para $n = 2$ variables, y grado $\leq k = 3$:

$$[x]_3 = [1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3]^T.$$

Theorem

Polinomio p de n variables y grado $\leq 2k$ es suma de cuadrados sii

$$p(x) = [x]_k^T M[x]_k;$$

para alguna matriz $M \in \mathbb{R}^{N\times N}$ semi-definida positiva, $N := C_k^{n+k}$.

- Igualando coeficientes de cada monomio, se obtienen restricciones lineales en M_{ij}.
- Recuperar p como SOS usando $M = LL^T$ (Choleski): $p(x) = [x]_k^T \left(LL^T\right)[x]_k = \|L^T[x]_k\|_2^2 = \sum_i \left(L^T[x]_k\right)_i^2$.

Referencias

1. How to globally solve non-convex optimization problems involving an approximate ℓ_0 penalization. Marmin, Castella, Pesquet (2019). ICASSP 2019.
2. Adaptive Forward-Backward Greedy Algorithm for Learning Sparse Representations. Zhang (2011). IEEE Transactions on Information Theory.
3. Enhancing sparsity by reweighted l_1 minimization. Candès, Wakin, Boyd. (2008). Journal of Fourier Anal. Appl.
4. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. Breheny & Huang (2011). Annals of applied statistics.
5. Global optimization with polynomials and the problem of moments. Lasserre (2001).
6. Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Parrilo (2000). California Institute of Technology.
7. The Lasserre hierarchy for polynomial optimization: A tutorial. Etienne de Klerk (2016).
8. SOS and SDP relaxation of polynomial otimization problems. Kojima (2010).
9. Sum of Squares. Parrilo & Lall (2003). IEEE Conference on Decision and Control.
10. The moment-SOS hierarchy. Lasserre. ICM2018.