Compressed Sensing

Algoritmo Re-Weighted l_1 con pesos actualizados resolviendo un problema dual

Matías Valdés
Director de tesis: Marcelo Fiori

12 de abril de 2019

Problema a resolver

- Sistema lineal $\Phi x = b$ con infinitas soluciones
- Hallar solución mayor cantidad coordenadas nulas: más "esparsa"
- $\Phi \in \mathbb{R}^{m \times n}$, $m < n \Rightarrow \Phi = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$
- Problema de optimización equivalente:

$$\begin{array}{ll} \operatorname{argmin} & ||x||_0 \\ \Phi x = b & \\ x \in \mathbb{R}^n & \end{array} \qquad (P_0)$$

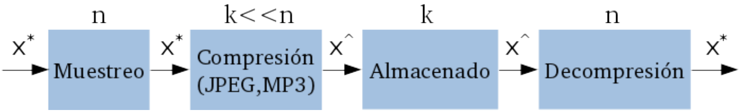
- (Pseudo)norma l_0 cuenta coordenadas no nulas:

$$||x||_0 := \#\{i \,/\, x_i \neq 0\}$$

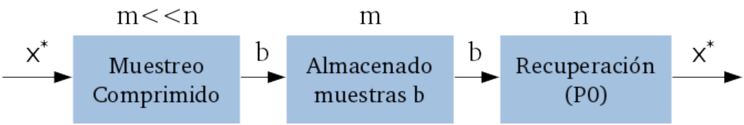
- Señales k-esparsas: $\Sigma_k = \{x \in \mathbb{R}^n \,/\, ||x||_0 \leq k\}$

Motivación: adquisición señales esparsas

Proceso usual realiza n medidas $x^* \in \mathbb{R}^n$ y luego comprime



CS sensa y comprime en único paso: m medidas $b = \Phi x^*$



Recuperar x^* a partir de $b = \Phi x^*$ requiere resolver (P_0)

Surgimiento del área

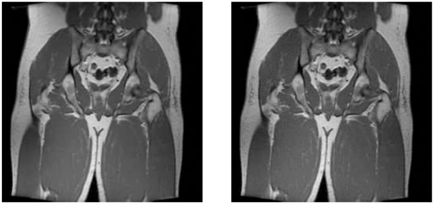
- CS surge como tal con trabajos de Donoho [2006] y de Candès, Romberg y Tao [2005, 2006]

- Recuperación con minimización l_1:
  - Década 1970 recuperar señales esparsas en Sismología
  - 1990 primeros resultados teóricos sobre desempeño
- Recuperación con algoritmos "Greedy" (1993):
  - Matching Pursuit (MP) y
  - Orthogonal Matching Pursuit (OMP) utilizado en actualidad (con sus variantes)

Aplicaciones

Desarrollo impulsado por gran cantidad de aplicaciones:

- Imagenología médica
- Conversión A/D
- Detección por RADAR
- Corrección errores
- Cámaras digitales
- Radioastronomía
- Comunicación inalámbrica



(a) full sampling

(b) 39% sampling, SNR=32.2

Complejidad del problema (P_0)

$$\begin{array}{ll} \text{argmin} & ||x||_0 \\ \Phi x = b & \end{array} \qquad (P_0)$$

- ¿Cómo hallar solución x^* de (P_0)?
  - Si conocemos soporte S, basta resolver $\Phi_S x = b$
- Si no conocemos soporte S:
  - probar todos los soportes, de todos los tamaños k, comenzando con $k = 1$
- Pero hay C_k^n posibles soportes de tamaño k
- Complejidad combinatoria y podría tardar demasiado:
  - $n = 100$ y $k = 15$: $C_{15}^{100} \simeq 2.53 \times 10^{17}$ soportes
  - cada prueba $1\mu s \Rightarrow$ hasta 8000 años en hallar solución
- $\nexists$ algoritmo resolver (P_0) tiempo polinomial: NP-difícil

Minimización l_1 ("Basis Pursuit")

- Se reemplaza (pseudo)norma l_0 por norma l_1:

$$\begin{array}{c} \operatorname{argmin} \\ \Phi x = b \\ x \in \mathbb{R}^n \end{array} \|x\|_1 = \begin{array}{c} \operatorname{argmin} \\ \Phi x = b \\ x \in \mathbb{R}^n \end{array} \sum_{i=1}^{n} |x_i| \qquad (P_1)$$

- Problema convexo $\Rightarrow$ solución en tiempo polinomial

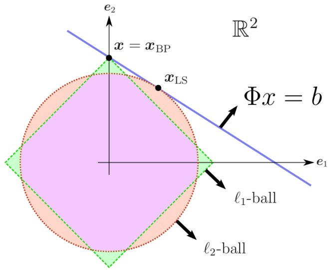


Figure: Norma l_2 solución menos esparsa

Equivalencia (P_1) y (P_0): condición suficiente

Definition (Propiedad de Isometría Restringida (RIP))

$\Phi \in \mathbb{R}^{m \times n}$ cumple RIP de orden k si $\exists\, \delta_k \in [0, 1]$ tal que:

$$(1 - \delta_k)\, ||x||_2^2 \leq ||\Phi x||_2^2 \leq (1 + \delta_k)\, ||x||_2^2\ ,\ \forall x \in \Sigma_k$$

Theorem (Candès y Tao)

Si Φ cumple RIP $2k$, con $\delta_{2k} < \sqrt{2} - 1$ y $b = \Phi x^*$, $x^* \in \Sigma_k$:

1. *(P_0) tiene solución k-esparsa única x^**
2. *Si $\hat{x}$ solución de $(P_1) \Rightarrow \hat{x} = x^*$*

Equivalencia (P_1) y (P_0): Φ Gaussiana

Theorem (Condición necesaria)

Si Φ *RIP orden* k*, con* $\delta_k \in (0,1) \Rightarrow m \geq C_\delta k \log\left(\frac{n}{k}\right)$

Theorem (Candès y Tao)

$\Phi \in \mathbb{R}^{m \times n}$*, con* $\Phi_{ij} \sim N(0,1)$ *iid. Dados* $\epsilon, \gamma \in (0,1)$ *y*

$$m \geq \frac{2}{\gamma^2}\left(k \ln\left(e\frac{n}{k}\right) + \log\left(\frac{2}{\epsilon}\right)\right)$$

Con probabilidad $\geq 1-\epsilon$, $\bar{\Phi} = \frac{1}{\sqrt{m}}\Phi$ *cumple RIP con:*

$$\delta_k \leq 2c\gamma + c^2\gamma^2, \quad c = 1 + \frac{1}{\sqrt{2\ln\left(e\frac{n}{k}\right)}}$$

Problema Weighted l_1 (l_1 con pesos)

- n pesos $w_i \geq 0$, problema Weighted l_1 asociado:

$$\hat{x} \in \underset{\substack{\Phi x = b \\ x \in \mathbb{R}^n}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i |x_i| \qquad (P_1 W)$$

- Objetivo: pesos w_i tales que $\hat{x} = x^*$, solución de (P_0)
- Caso ideal: si se conoce solución x^* de (P_0), tomar:

$$w_i = \frac{1}{|x_i^*|} \in [0, \infty]$$

- Con esto: $x_i^* = 0 \Rightarrow$ "costo" $w_i = +\infty \Rightarrow \hat{x}_i = 0$

Ejemplo en que $P_1 \not\Rightarrow P_0$ pero $P_1 W \Rightarrow P_0$

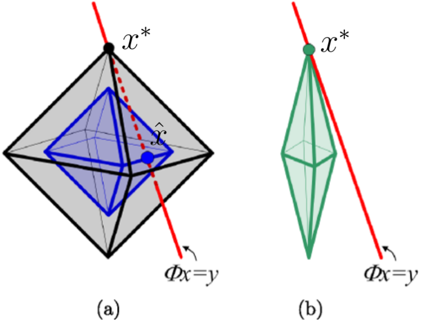


Figure: (a) x^* no es solución P_1; (b) x^* solución $P_1 W$

Algoritmo RWl_1 de Candès, Wakin y Boyd [2008]

- Caso no ideal en que no se conoce x^*
- Caso ideal $w_i = \frac{1}{|x_i^*|}$ sugiere:

$$w_i^{k+1} = \frac{1}{|x_i^{k+1}|}$$

- Si w^k estimación pesos $\Rightarrow$ candidato a x^*:

$$x^{k+1} \in \underset{\substack{\Phi x = b \\ x \in \mathbb{R}^n}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i^k |x_i|, \quad \forall k \geq 0$$

- En la práctica $w_i^{k+1} = \frac{1}{|x_i^{k+1}|+\epsilon}$, $\epsilon > 0$ (algoritmo CWB)

Metodología RWl_1 propuesta

- Basada en relajación Lagrangeana
- Dada $x^* \in \Sigma_k$, solución de (P_0), se define primal ideal:

$$\hat{x} \in \begin{array}{c} \text{argmin} \\ \Phi x = b \\ |x_i| \leq |x_i^*|, \ \forall i \end{array} \quad 0 \qquad (P)$$

- Problema de factibilidad: se busca cumplir restricciones
- $\hat{x}$ solución de $(P) \Rightarrow \|\hat{x}\|_0 \leq \|x^*\|_0 \Rightarrow \hat{x}$ solución (P_0)
- Problema convexo pero ideal

Relajación restricciones ideales

- Relajando restricciones ideales $|x_i| \leq |x_i^*|$, lagrangeano:

$$L(x,w) := 0 + \sum_{i=1}^{n} w_i \left(|x_i| - |x_i^*|\right) = \sum_{i=1}^{n} w_i |x_i| - \sum_{i=1}^{n} w_i |x_i^*|$$

- Función dual asociada:

$$d(w) := \min_{\substack{\Phi x = b \\ x \in \mathbb{R}^n}} L(w,x) =$$

$$= \left(\min_{\substack{\Phi x = b \\ x \in \mathbb{R}^n}} \sum_{i=1}^{n} w_i |x_i| \right) - \sum_{i=1}^{n} w_i |x_i^*|$$

- Problema "Weighted l_1", w_i multiplicadores Lagrange

Metodología propuesta para actualizar pesos

- Pesos $w_i \geq 0$ son multiplicadores Lagrange $\Rightarrow$
- Estimar w_i con algoritmos para estimar multiplicadores
- Alternativamente, estimar w_i como soluciones del dual:

$$\begin{aligned} &\text{argmax} \quad d(w) \\ &w \geq 0 \\ &w \in \mathbb{R}^n \end{aligned} \qquad (D)$$

- Maximización cóncava ($\Leftrightarrow$ minimización convexa)
- Función objetivo $d(w)$ no siempre diferenciable
- Algoritmo subgradiente para resolver (D)

Soluciones del dual (ideal) D

Proposición

Hay dualidad fuerte: $d^* = f^* = 0$

- $w = \vec{0}$ solución dual pero no devuelve solución de (P_0):

$$\hat{x} \in \underset{\Phi x = b}{\operatorname{argmin}} \sum_{i=1}^{n} 0|x_i| = \underset{\Phi x = b}{\operatorname{argmin}} \; 0 = \{\Phi x = b\}$$

Proposición

Sea $\hat{w} \geq 0 \; / \; \hat{w}_i = 0 \Leftrightarrow x_i^* \neq 0$ *(soporte disjunto). Entonces:*

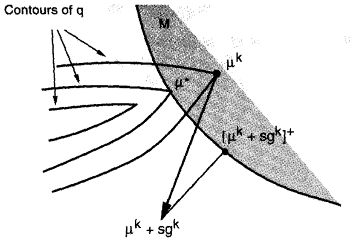
- $\hat{w}$ *solución del dual y*
- $\hat{x} \in \operatorname{argmin}_{\Phi x = b} \sum_{i=1}^{n} \hat{w}_i |x_i|$ *solución de* (P_0)

Algoritmo subgradiente (busca solución dual)

- "Ascenso" por (sub)gradiente en función dual cóncava:

$$\begin{cases} w^{k+1} = w^k + \alpha_k g^k, \alpha_k > 0,\ g^k \in \partial d(w^k) \\ w^{k+1} = \max\{0, w^{k+1}\} \end{cases}$$

- Puede no ser de ascenso: $d(w^{k+1}) < d(w^k)$, $\forall\, \alpha_k > 0$



- Siempre es posible elegir paso tal que $\forall$ solución w^*:

$$\forall\, k \geq 0,\ \exists\, \alpha_k > 0 \,/\, ||w^{k+1} - w^*|| < ||w^k - w^*||$$

Algoritmo subgradiente (busca solución dual)

- ¿Cómo hallar un subgradiente de $d(w)$ en cada paso?

Theorem (Subgradiente de función dual en $\hat{w}$)

$$\hat{x} \in \underset{\Phi x = b}{\operatorname{argmin}} \sum_{i=1}^{n} \hat{w}_i |x_i| \Rightarrow g(\hat{x}) = |\hat{x}| - |x^*| \in \partial d(\hat{w})$$

Metodología + Subgradiente:

- $x^{k+1} \in \operatorname{argmin}_{\Phi x = b} \sum_{i=1}^{n} w_i^k |x_i|$
- $g^k = |x^{k+1}| - |x^*|$ subgradiente dual en w^k
- $w^{k+1} = \max\{w^k + \alpha_k g^k, 0\}, \alpha_k > 0$

Restricciones no ideales

- Restricciones ideales: $g_i(x) = |x_i| - |x_i^*|$
- En la práctica no se conoce solución x^* de (P_0)
- Reemplazar x^* por estimación x^{k+1} ("amplificada"):

$$g_i^k(x) = |x_i| - (1+\epsilon_k)\,|x_i^{k+1}|, \quad \epsilon_k > 0$$

Donde:

$$x^{k+1} \in \underset{\Phi x = b}{\operatorname{argmin}} \sum_{i=1}^{n} w_i^k |x_i|$$

- Cada paso tiene asociado problema primal no ideal:

$$\begin{array}{lcr} \operatorname{argmin} & 0 & (P^k) \\ \Phi x = b & & \\ |x_i| \leq (1+\epsilon_k)\,|x_i^{k+1}|,\ \forall i & & \end{array}$$

Modelo utilizado

- $\Phi \in \mathbb{R}^{m \times n}$ matriz de medida aleatoria
- $n = 256, \; m = 100$ y entradas independientes:

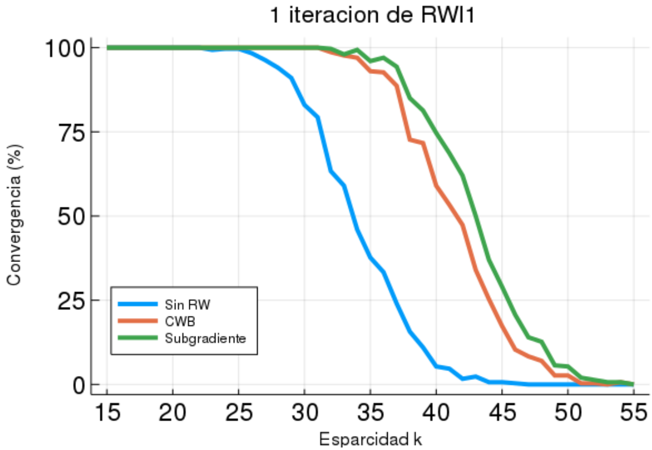
$$\Phi_{ij} \sim N\left(0, \sigma = \frac{1}{\sqrt{m}}\right), \; \forall \, i, j$$

- Soporte x^*: sorteo k indices S_k
- Valores soporte:

$$x_i^* \sim N\left(0, \sigma = \frac{1}{\sqrt{k}}\right), \forall i \in S_k$$

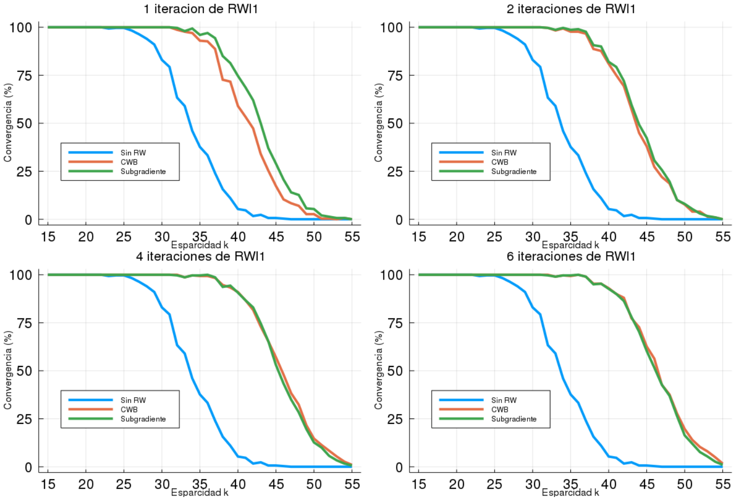
- Término independiente: $b = \Phi x^* \in \mathbb{R}^m$
- Convergencia si: $\|x^{\text{maxIter}} - x^*\|_\infty \leq 1 \times 10^{-3}$

Resultados ($m = 100$ fijo, $k \in [15, 55]$)



- Existe región de transición de fase
- En minimización l_1: si $n \to \infty \Rightarrow$ ancho región $\to 0$

Resultados ($m = 100$ fijo, $k \in [15, 55]$)



Ruido en las medidas

Medidas b afectadas por ruido Gaussiano aditivo:

$$b = \Phi x^* + z, \quad z_i \sim N(0, \sigma^2) \text{ independientes}$$

Se busca resolver:

$$\underset{||\Phi x - b||_2 \leq \eta}{\text{argmin}} \quad ||x||_0 \qquad (P_0^\eta)$$

Problema Weighted l_1 asociado:

$$\underset{||\Phi x - b||_2 \leq \eta}{\text{argmin}} \quad \sum_{i=1}^{n} w_i |x_i| \qquad (P_1^\eta W)$$

Problema W-LASSO (sustituye a Weighted l_1)

- Problema primal ideal:

$$\begin{array}{cc} \underset{\substack{\frac{1}{2}\|\Phi x - b\|_2^2 \leq \frac{\eta^2}{2} \\ |x_i| \leq |x_i^*|, \ \forall i}}{\operatorname{argmin}} & 0 \end{array}$$

- Relajando restricciones ideales:

$$d(w) = \left(\min_{\frac{1}{2}\|\Phi x - b\|_2^2 \leq \frac{\eta^2}{2}} \sum_{i=1}^{n} w_i |x_i| \right) - \sum_{i=1}^{n} w_i |x_i^*|$$

- Problema Weighted l_1 restricciones cuadráticas
- w_i multiplicadores de Lagrange como antes

Problema W-LASSO (sustituye a Weighted l_1)

- Relajando ambas restricciones:

$$d(w,\lambda) = \left(\min_{x \in \mathbb{R}^n} \frac{\lambda}{2}\|\Phi x - b\|_2^2 + \sum_{i=1}^{n} w_i |x_i| \right) +$$

$$- \left(\frac{\lambda}{2}\eta^2 + \sum_{i=1}^{n} w_i |x_i^*| \right) \text{ Weighted LASSO}$$

- Restricciones ideales $g_i(x) = |x_i| - |x_i^*|$ por:

$$g_i^k(x) = |x_i| - (1+\epsilon_k)\,|x_i^{k+1}|,\ \epsilon_k > 0$$

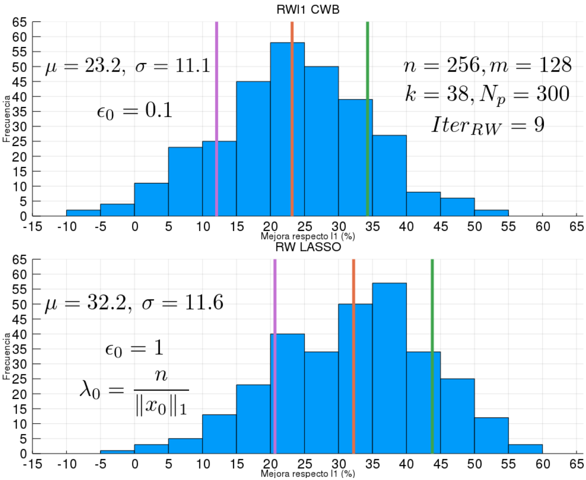
- $x^{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \ \frac{\lambda}{2}||\Phi x - b||_2^2 + \sum_{i=1}^{n} w_i^k |x_i|$
- Resolver W-LASSO en cada paso (antes Weighted l_1)

Modelo utilizado y medida desempeño

- Φ y x^* igual que sin ruido ($m = 128$, $n = 256$)
- Medidas $b = \Phi x^* + z$, $z_i \sim N(0, \sigma^2)$ iid
- $N_p = 300$ pruebas con $k = 38$ fijo
- Mejora respecto solución de l_1 con ruido:

$$a = 100 \times \left(1 - \frac{||x^{RW} - x^*||_2}{||x_{l_1} - x^*||_2}\right) \%$$

Resultados con ruido



- RW-LASSO: +32% minimización l_1 y +40% RWl_1

Conclusiones

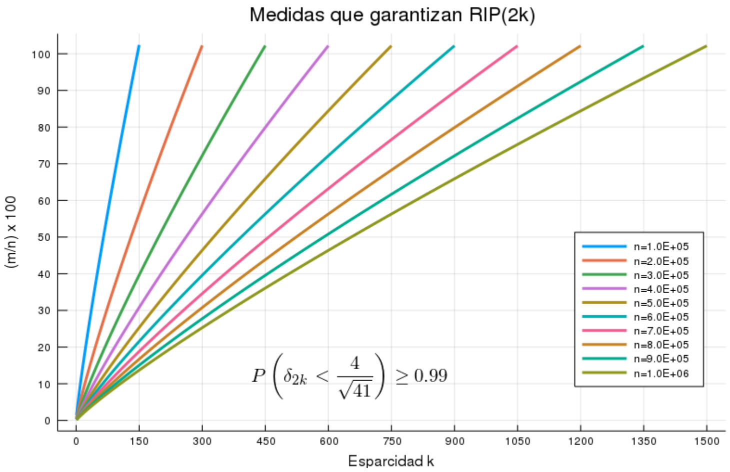
- CS y minimización l_1 con resultados más relevantes
- Problema Weighted l_1 y algoritmo RWl_1 de CWB

- Nueva **metodología** actualizar pesos RWl_1 Lagrange
- **Algoritmo** RWl_1 subgradiente alternativo a CWB
- Misma metodología con ruido: algoritmo RW-LASSO

Conclusiones

- Desempeño RW problema CS aleatorio Φ Gaussiana
- Sin ruido:
  - RWl_1 de CWB y subgradiente desempeño similar
  - Interpretación pesos multiplicadores Lagrange
- Con ruido Gaussiano:
  - RW-LASSO propuesto mejora considerable a RWl_1

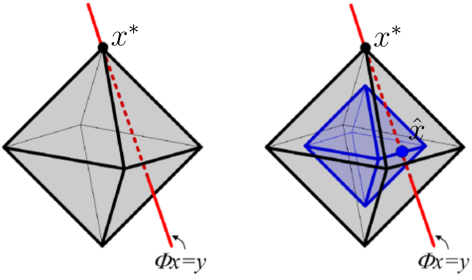
Equivalencia (P_1) y (P_0): Φ Gaussiana



Ejemplo en que $P_1 \not\Rightarrow P_0$

$$\Phi = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix},\ b = \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ x^* = (0,1,0)$$

- Otra solución de $\Phi x = b$: $\hat{x} = \left(\frac{1}{3}, 0, \frac{1}{3}\right)$
- $||\hat{x}||_1 = \frac{2}{3} < 1 = ||x^*||_1 \Rightarrow x^*$ no es solución de (P_1)



(a) (b)

Algoritmo Gradiente Proximal (GP)

- Problema convexo: $\operatorname{argmin}_{x \in \mathbb{R}^n} \quad f(x) + g(x)$
- g continua, no necesariamente diferenciable
- Gradiente Proximal: $x^{k+1} = Prox_{\alpha g}\left(x^k - \alpha \nabla f(x^k)\right)$
- Operador proximal de $h : \mathbb{R}^n \to \mathbb{R}$:

$$Prox_{\alpha h}(v) := \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ h(x) + \frac{1}{2\alpha} ||x - v||_2^2 \right\} : \mathbb{R}^n \to \mathbb{R}^n$$

- GP Acelerado (Nesterov):
  1. Punto intermedio:

$$y^k = x^k + r_k(x^k - x^{k-1}),\ \forall\, k \geq 1,\ r_k = \frac{k}{k+3} \in [0, 1)$$

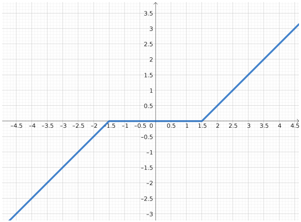
  2. GP a punto intermedio:

$$x^{k+1} = prox_{\alpha g}\left(y^k - \alpha \nabla f(y^k)\right),\ \forall\, k \geq 1$$

GPA para W-LASSO: algoritmo W-FISTA

$$g(x) = ||Wx||_1 = \sum_{i=1}^{n} w_i|x_i| = \sum_{i=1}^{n} g_i(x_i)$$

$$(prox_{\alpha g}(x))_i = prox_{\alpha g_i}(x_i) = \begin{cases} x_i - \alpha w_i & x_i \geq \alpha w_i \\ 0 & |x_i| \leq \alpha w_i \\ x_i + \alpha w_i & x_i \leq -\alpha w_i \end{cases}$$



$$f(x) = \frac{\lambda}{2}\|\Phi x - b\|_2^2 \Rightarrow \nabla f(x) = \lambda \Phi^T (\Phi x - b)$$

Algorithm 1 RW-LASSO subgradiente (no ideal y con ruido)

$x^1 \in \operatorname{argmin}_{x\in\mathbb{R}^n} \;\frac{\lambda^0}{2}\|\Phi x - b\|_2^2 + \sum_{i=1}^{n} w_i^0|x_i|$ {W-LASSO}

$k = 0$

while $k <$ maxIter **do**

    $g_i^k = g_i^k(x^{k+1}) = |x_i^{k+1}| - (1+\epsilon_k)\,|x_i^{k+1}|$

    $w_i^{k+1} = \max\left(0, w_i^k + \alpha_k g_i^k\right)$

    $g_\lambda^k = g_\lambda(x^{k+1}) = \frac{1}{2}\left(\|\Phi x^{k+1} - b\|_2^2 - \eta^2\right)$

    $\lambda^{k+1} = \max\left(0, \lambda^k + \alpha_k g_\lambda^k\right)$

    $k = k+1$

    $x^0 = x^k$

    $x^{k+1} \in \operatorname{argmin}_{x\in\mathbb{R}^n} \;\frac{\lambda^k}{2}||\Phi x - b||_2^2 + \sum_{i=1}^{n} w_i^k|x_i|$

end while

return x^{k+1}

Unicidad de soluciones de (P_0)

Theorem

k-unicidad $\Leftrightarrow$ spark$(\Phi) > 2k$

- Cálculo del spark es un problema NP-difícil
- Como $m \leq n \Rightarrow rg\,(\Phi) \leq m \Rightarrow spark\,(\Phi) \leq m + 1$

Corollary

Se requieren $m \geq 2k$ medidas para tener k-unicidad

Theorem

Si Φ satisface RIP $2k$, con $\delta_{2k} < 1 \Rightarrow$ se tiene k-unicidad

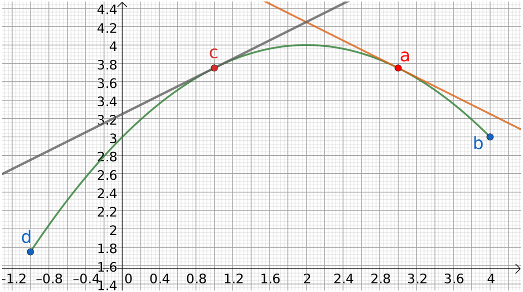
No es necesario RIP con $\delta_{2k} < 1$ para tener k-unicidad:

$$\Phi = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

Marco teórico CWB (algoritmo MM)

Proviene de metodología Mayorización-Minimización (MM):

- Si g función cóncava $\Rightarrow$ "mayorizar" con linealización h



- Minimizar h en lugar de g:

$$v^{k+1} \in \underset{v \in C}{\operatorname{argmin}} \left(g(v^k) + \left\langle \nabla g(v^k), v - v^k \right\rangle \right)$$

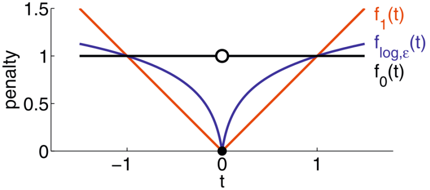
- Puede converger a mínimo no global (condición inicial)

Marco teórico CWB (algoritmo MM)

- CWB surge de aplicar MM a buscar soluciones de:

$$\underset{\substack{x \in \mathbb{R}^n \\ \Phi x = b}}{\operatorname{argmin}} \sum_{i=1}^{n} \ln\left(|x_i| + \epsilon\right) = \underset{\substack{x, u \in \mathbb{R}^n \\ \Phi x = b \\ |x_i| \leq u_i,\ \forall i}}{\operatorname{argmin}} \sum_{i=1}^{n} \ln\left(u_i + \epsilon\right)$$

- $\ln\left(\frac{|x_i|}{\epsilon} + 1\right)$ se parece a (pseudo)norma l_0:



Marco teórico CWB (algoritmo MM)

- Función objetivo cóncava $g(x,u) = \sum_{i=1}^{n} \ln(u_i + \epsilon)$
- Algoritmo MM:

$$(x^{k+1}, u^{k+1}) \in \underset{\substack{x,\, u \in \mathbb{R}^n \\ \Phi x = b \\ |x_i| \leq u_i,\ \forall i}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\frac{1}{u_i^k + \epsilon} \right) u_i \qquad (1)$$

- Eliminando variable auxiliar:

$$x^{k+1} \in \underset{\substack{x \in \mathbb{R}^n \\ \Phi x = b}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\frac{1}{|x_i^k| + \epsilon} \right) |x_i| \qquad (2)$$