# Alternating direction method of multipliers (ADMM)

Oscar S. Dalmau Cedeño
dalmau@cimat.mx

13 de noviembre de 2018

- El metodo del Lagrangiano aumentado este relacionado con los metodos de penalizacion.
- Este metodo intenta resolver el problema del mal condicionamiento que se presenta en los algoritmos de penalizacion: cuadraticos y de barrera

**Motivacion**: Problema de optimizacion con restricciones de igualdad

$$\begin{array}{rl} & \min f(\boldsymbol{x}) \\ s.a. & c_i(\boldsymbol{x}) = 0,\ i \in \mathcal{E} \end{array}$$

Lo anterior se puede lograr mediante la introduccion de una nueva funcion, conocida como **Lagrangiano aumentado** (Hestenes y Powel 1969 de forma independiente) que considera o es una combinacion del Lagrangiano y el termino de penalizacion, y se define como sigue:

$$\mathcal{L}_A(\boldsymbol{x}, \boldsymbol{\lambda}; \mu) \quad \stackrel{def}{=} \quad f(\boldsymbol{x}) - \sum_{i \in \mathcal{E}} \lambda_i c_i(\boldsymbol{x}) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} (c_i(\boldsymbol{x}))^2$$

que es muy parecido al Lagrangiano, y solo se diferencia en el termino de penalizacion.

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \quad = \quad f(\boldsymbol{x}) - \sum_{i \in \mathcal{E}} \lambda_i c_i(\boldsymbol{x})$$

La idea ahora es desarrollar una algoritmo, similar a los de penalizacion en los que se resuelve una secuencia de subproblemas

$$
\begin{aligned}
\boldsymbol{x}_{k+1} &= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) - \sum_{i\in\mathcal{E}} \lambda_i^k c_i(\boldsymbol{x}) + \frac{\mu_k}{2}\sum_{i\in\mathcal{E}} (c_i(\boldsymbol{x}))^2 \\
\lambda_i^{k+1} &= \lambda_i^k - \mu_k c_i(\boldsymbol{x}_{k+1}) \\
\mu_{k+1} &> \mu_k
\end{aligned}
$$

donde los parametros $\boldsymbol{\lambda}^k$ y $\mu_k > 0$ son fijos en cada iteracion. En la practica, $\boldsymbol{x}_{k+1}$ es solo un minimizador aproximado de $\mathcal{L}_A(\boldsymbol{x}, \boldsymbol{\lambda}^k; \mu_k)$.

# Algoritmo Lagrangiano Aumentado

**Algorithm 1** Algoritmo LA (restricciones de igualdad)

Dado un punto inicial $x_0^s$, $\boldsymbol{\lambda}^0$, $\tau_0, \mu_0 > 0$
**for** $k = 0, 1, 2, \cdots$ **do**
  Encontrar un minimizador aproximado $\boldsymbol{x}_k$ de $\mathcal{L}_A(\boldsymbol{x}, \boldsymbol{\lambda}^k; \mu_k)$ iniciando en $x_k^s$ y terminar cuando $\|\nabla_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x}_k, \boldsymbol{\lambda}; \mu_k)\| \leq \tau_k$
  **if** converge **then**
    Parar el algoritmo con solucion $\boldsymbol{x}_k$
  **end if**
  Actulizar $\lambda_i^{k+1} = \lambda_i^k - \mu_k c_i(\boldsymbol{x}_k)$
  Seleccionar $\mu_{k+1} > \mu_k$ y $\tau_{k+1}$
  Seleccionar un nuevo punto inicial $\boldsymbol{x}_{k+1}^s$
**end for**

El problema anterior puede ser reescrito como sigue

$$
\begin{aligned}
\boldsymbol{x}_{k+1} &= \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\mu_k}{2}\|c(\boldsymbol{x}) - \frac{1}{\mu_k}\boldsymbol{\lambda}^k\|^2 \\
\lambda_i^{k+1} &= \lambda_i^k - \mu_k c_i(\boldsymbol{x}_{k+1}) \\
\mu_{k+1} &> \mu_k
\end{aligned}
$$

donde $c(\boldsymbol{x}) = [c_1(\boldsymbol{x}), c_2(\boldsymbol{x}), \cdots]^T$.

# Forma dual escalada

Definifiendo $v_i^k := \frac{1}{\mu_k}\lambda_i^k$ entonces tenemos la Forma dual escalada (scaled dual form)

$$\begin{aligned}
\boldsymbol{x}_{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x}, v^k) := f(\boldsymbol{x}) + \frac{\mu_k}{2}\|c(\boldsymbol{x}) - \boldsymbol{v}^k\|^2 \\
v_i^{k+1} &= v_i^k - c_i(\boldsymbol{x}_{k+1}) \\
\mu_{k+1} &> \mu_k
\end{aligned}$$

ademas, $\nabla \mathcal{L}_A(\boldsymbol{x}, \boldsymbol{v}^k) = \nabla f(\boldsymbol{x}) - \mu_k \nabla c(\boldsymbol{x})(c(\boldsymbol{x}) - \boldsymbol{v}^k)$ y $\nabla c(\boldsymbol{x}) = [\nabla c_1(\boldsymbol{x}), \nabla c_2(\boldsymbol{x}), \cdots]$ es la Traspuesta del Jacobiano de $c(\boldsymbol{x})$

# Ejemplo 1

Consideremos el caso Particular

$$\begin{aligned} & \min f(\boldsymbol{x}) \\ s.a. \quad & \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Forma dual escalada (Algoritmo)

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x}, \boldsymbol{v}^k) := f(\boldsymbol{x}) + \frac{\mu_k}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b} - \boldsymbol{v}^k\|^2 \\ \boldsymbol{v}^{k+1} &= \boldsymbol{v}^k - (\mathbf{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}) \\ \mu_{k+1} &> \mu_k \end{aligned}$$

# Ejemplo 2: Funcion cuadratica

$$\begin{aligned}\min_{\boldsymbol{x},\boldsymbol{z}} \quad & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{c} \\ s.t: \quad & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}\end{aligned}$$

$$\mathcal{L}_A(\boldsymbol{x},\boldsymbol{v}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{c} + \frac{\mu}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b} - \boldsymbol{v}\|^2 + cte$$

# Ejemplo 2: Funcion cuadratica

Algoritmo (Forma dual escalada)

$$\begin{aligned}\boldsymbol{x}_{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x}, \boldsymbol{v}^k) \\ \boldsymbol{v}^{k+1} &= \boldsymbol{v}^k - (\mathbf{A}\boldsymbol{x}_{k+1} - \boldsymbol{b})\end{aligned}$$

con $\boldsymbol{x}_{k+1} = (\mathbf{Q} + \mu_k \mathbf{A}^T\mathbf{A})^{-1}(\mu_k \mathbf{A}^T(\boldsymbol{b} + \boldsymbol{v}^k) + \boldsymbol{c})$

# Problema Dual

Consideremos el caso Particular

$$\begin{aligned} & \min f(\boldsymbol{x}) \\ s.a. \qquad & \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

- Lagrangiano: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \boldsymbol{\lambda}^T(\mathbf{A}\boldsymbol{x} - \boldsymbol{b})$
- Funcion dual: $g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$
- Problema dual: $\max g(\boldsymbol{\lambda})$

# Gradiente ascendente Dual

Sea

$$\boldsymbol{x}_{k+1} \quad := \quad \arg\min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

Entonces

$$\begin{aligned} g(\boldsymbol{\lambda}) &= \mathcal{L}(\boldsymbol{x}_{k+1}, \boldsymbol{\lambda}) = f(\boldsymbol{x}_{k+1}) - \boldsymbol{\lambda}^T (A\boldsymbol{x}_{k+1} - b) \\ \nabla g(\boldsymbol{\lambda}) &= \nabla \mathcal{L}(\boldsymbol{x}_{k+1}, \boldsymbol{\lambda}) \\ &= -(A\boldsymbol{x}_{k+1} - b) \end{aligned}$$

Y el **gradiente ascendente** para $\boldsymbol{\lambda}$ es

$$\begin{aligned} \boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k + \alpha_k \nabla g(\boldsymbol{\lambda}^k) \\ &= \boldsymbol{\lambda}^k - \alpha_k (A\boldsymbol{x}_{k+1} - b) \end{aligned}$$

# Gradiente ascendente Dual

En resumen

$$
\begin{aligned}
\boldsymbol{x}_{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^k) \\
\boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k - \alpha_k (A\boldsymbol{x}_{k+1} - b)
\end{aligned}
$$

# Descomposicion Dual

Consideremos el caso separable con $\boldsymbol{x} = [\boldsymbol{x}_1; \boldsymbol{x}_2; \cdots ; \boldsymbol{x}_N]$

$$\begin{aligned} \min f(\boldsymbol{x}) &= f(\boldsymbol{x}_1) + f(\boldsymbol{x}_2) + \cdots + f(\boldsymbol{x}_N) \\ \mathbf{A}\boldsymbol{x} &= \boldsymbol{b} \end{aligned}$$

Luego

$$\begin{aligned} \mathbf{A}\boldsymbol{x} &= \boldsymbol{b} \\ \sum_{i=1}^{N} \mathbf{A}_i \boldsymbol{x}_i &= \boldsymbol{b} \end{aligned}$$

donde

$$\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots , \mathbf{A}_N]$$

# Descomposicion Dual

Luego el Lagrangiano es

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) &= \min f(\boldsymbol{x}) - \boldsymbol{\lambda}^T(A\boldsymbol{x} - b) \\
&= \sum_{i=1}^{N} [f(\boldsymbol{x}_i) - \boldsymbol{\lambda}^T \mathbf{A}_i \boldsymbol{x}_i] + \boldsymbol{\lambda}^T b = \sum_{i=1}^{N} \mathcal{L}_i(\boldsymbol{x}_i, \boldsymbol{\lambda}) + \boldsymbol{\lambda}^T \boldsymbol{b}
\end{aligned}
$$

donde $\mathcal{L}_i(\boldsymbol{x}_i, \boldsymbol{\lambda}) := f(\boldsymbol{x}_i) - \boldsymbol{\lambda}^T \mathbf{A}_i \boldsymbol{x}_i$

# Descomposicion Dual: Algoritmo

En resumen

$$
\begin{aligned}
\boldsymbol{x}_i^{k+1} &= \arg\min_{\boldsymbol{x}_i} \mathcal{L}_i(\boldsymbol{x}_i, \boldsymbol{\lambda}^k);\; i = 1, 2, \cdots, N \\
\boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k - \alpha_k\Big(\sum_{i=1}^{N} \mathbf{A}_i \boldsymbol{x}_i^{k+1} - \boldsymbol{b}\Big)
\end{aligned}
$$

- Los subproblemas pueden resolverse en paralelo
- La actualizacion de la variable dual nos proporciona el acoplamiento/coordinacion
- El algoritmo trabaja bien, sin muchas consideciones o supuestos, aunque puede ser lento.

# Alternating direction method of multipliers

- Problema: Sean $f, g$ convexas

$$\begin{array}{ll} \min\limits_{\boldsymbol{x},\boldsymbol{z}} & f(\boldsymbol{x}) + g(\boldsymbol{z}) \\ s.t: & A\boldsymbol{x} + B\boldsymbol{z} = c \end{array}$$

- Lagrangiano aumentado

$$\mathcal{L}_A(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) - \boldsymbol{\lambda}^T(A\boldsymbol{x} + B\boldsymbol{z} - c) + \frac{\mu}{2}\|A\boldsymbol{x} + B\boldsymbol{z} - c\|^2$$

# Alternating direction method of multipliers

Algoritmo ADMM: Forma dual escalada

$$\mathcal{L}_A(\boldsymbol{x},\boldsymbol{z},\boldsymbol{v})=f(\boldsymbol{x})+g(\boldsymbol{z})+\frac{\mu}{2}\|A\boldsymbol{x}+B\boldsymbol{z}-c-\boldsymbol{v}\|^2+cte$$

con $\boldsymbol{v}=\frac{1}{\mu_k}\boldsymbol{\lambda}$. Entonces

$$\begin{aligned}
\boldsymbol{x}^{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x},\boldsymbol{z}^k,\boldsymbol{v}^k) \text{ // minimizacion-x}\\
\boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \mathcal{L}_A(\boldsymbol{x}^{k+1},\boldsymbol{z},\boldsymbol{v}^k) \text{ // minimizacion-z}\\
\boldsymbol{v}^{k+1} &= \boldsymbol{v}^k-(A\boldsymbol{x}^{k+1}+B\boldsymbol{z}^{k+1}-c) \text{ // actulizacion dual}
\end{aligned}$$

# Funcion convexa compuesta (I)

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) + g(\boldsymbol{x})$$

Se puede transformar en

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}} \quad & f(\boldsymbol{x}) + g(\boldsymbol{z}) \\ s.t: \quad & \boldsymbol{x} - \boldsymbol{z} = 0 \end{aligned}$$

Luego $\mathcal{L}_A(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{z} - \boldsymbol{v}\|^2 + cte$

$$\begin{aligned} \boldsymbol{x}^{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x}, \boldsymbol{z}^k, \boldsymbol{v}^k) \text{ // minimizacion-x} \\ \boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \mathcal{L}_A(\boldsymbol{x}^{k+1}, \boldsymbol{z}, \boldsymbol{v}^k) \text{ // minimizacion-z} \\ \boldsymbol{v}^{k+1} &= \boldsymbol{v}^k - (\boldsymbol{x}^{k+1} - \boldsymbol{z}^{k+1}) \text{ // actulizacion dual} \end{aligned}$$

con $\boldsymbol{v} = \frac{1}{\mu_k}\boldsymbol{\lambda}$.

# Funcion convexa compuesta (II)

$$\min_{\boldsymbol{x}} \quad f(\mathbf{A}\boldsymbol{x}) + g(\mathbf{B}\boldsymbol{x})$$

Se puede transformar en

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}} \quad & f(\boldsymbol{y}) + g(\boldsymbol{z}) \\ s.t: \quad & \mathbf{A}\boldsymbol{x} - \boldsymbol{y} = 0;\ \mathbf{B}\boldsymbol{x} - \boldsymbol{z} = 0 \end{aligned}$$

Luego

$$\begin{aligned} \mathcal{L}_A(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}_1, \boldsymbol{v}_2) \quad = \quad & f(\boldsymbol{y}) + g(\boldsymbol{z}) + \frac{\mu}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{y} - \boldsymbol{v}_1\|^2 \\ & + \frac{\mu}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{z} - \boldsymbol{v}_2\|^2 + cte \end{aligned}$$

# Funcion convexa compuesta (III)

Luego

$$
\begin{aligned}
\boldsymbol{x}^{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{z}^k, \boldsymbol{v}_1^k, \boldsymbol{v}_2^k) \text{ // minimizacion-x} \\
\boldsymbol{y}^{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x}^{k+1}, \boldsymbol{y}, \boldsymbol{z}^k, \boldsymbol{v}_1^k, \boldsymbol{v}_2^k) \text{ // minimizacion-y} \\
\boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \mathcal{L}_A(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{z}, \boldsymbol{v}_1^k, \boldsymbol{v}_2^k) \text{ // minimizacion-z} \\
\boldsymbol{v}_1^{k+1} &= \boldsymbol{v}_1^k - (\mathbf{A}\boldsymbol{x}^{k+1} - \boldsymbol{y}^{k+1}) \text{ // actulizacion dual } \boldsymbol{v}_1 \\
\boldsymbol{v}_2^{k+1} &= \boldsymbol{v}_2^k - (\mathbf{B}\boldsymbol{x}^{k+1} - \boldsymbol{z}^{k+1}) \text{ // actulizacion dual } \boldsymbol{v}_2
\end{aligned}
$$

donde $\boldsymbol{x}^{k+1}$ tiene formula cerrada

$$
\boldsymbol{x}^{k+1} = (\mathbf{A}^T\mathbf{A} + \mathbf{B}^T\mathbf{B})^{-1}(\mathbf{A}^T(\boldsymbol{y}^k + \boldsymbol{v}_1^k) + \mathbf{B}^T(\boldsymbol{z}^k + \boldsymbol{v}_2^k))
$$

# Funcion convexa compuesta (IV)

Si ademas (no es interesante, solucion formula cerrada)

$$f(\mathbf{A}\boldsymbol{x}) = \frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2; \quad g(\mathbf{B}\boldsymbol{x}) = \frac{1}{2}\|\mathbf{B}\boldsymbol{x}\|^2$$

Entonces $\boldsymbol{x}^* = (\mathbf{A}^T\mathbf{A} + \mathbf{B}^T\mathbf{B})^{-1}\mathbf{A}^T\boldsymbol{b}$. Por otro lado,

$$\begin{aligned} f_1(\boldsymbol{y}) &= \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{b}\|^2 + \frac{\mu_k}{2}\|\mathbf{A}\boldsymbol{x}^{k+1} - \boldsymbol{y} - \boldsymbol{v}_1^k\|^2 \\ f_2(\boldsymbol{z}) &= \frac{1}{2}\|\boldsymbol{z}\|^2 + \frac{\mu_k}{2}\|\mathbf{B}\boldsymbol{x}^{k+1} - \boldsymbol{z} - \boldsymbol{v}_2^k\|^2 \end{aligned}$$

Luego

$$\begin{aligned} \boldsymbol{y}^{k+1} &= \frac{1}{1+\mu_k}(\boldsymbol{b} + \mu_k(\mathbf{A}\boldsymbol{x}^{k+1} - \boldsymbol{v}_1^k)) \\ \boldsymbol{z}^{k+1} &= \frac{\mu_k}{1+\mu_k}(\mathbf{B}\boldsymbol{x}^{k+1} - \boldsymbol{v}_2^k) \end{aligned}$$

# Lasso con ADMM (I)

Lasso

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 + \lambda\|\boldsymbol{x}\|_1$$

Se puede transformar en el problema

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}} \quad & f(\boldsymbol{x}) + g(\boldsymbol{z}) \\ s.t: \quad & \boldsymbol{x} - \boldsymbol{z} = 0 \end{aligned}$$

con

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \\ g(\boldsymbol{z}) &= \lambda\|\boldsymbol{z}\|_1 \end{aligned}$$

# Lasso con ADMM (II)

El Lagrangiano Aumentado es

$$\mathcal{L}_A(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{z} - \boldsymbol{v}\|^2 + cte$$

Luego

$$\begin{aligned}
\boldsymbol{x}^{k+1} &= \arg\min_{\boldsymbol{x}} \mathcal{L}_A(\boldsymbol{x}, \boldsymbol{z}^k, \boldsymbol{v}^k) \text{ // minimizacion-x} \\
\boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \mathcal{L}_A(\boldsymbol{x}^{k+1}, \boldsymbol{z}, \boldsymbol{v}^k) \text{ // minimizacion-z} \\
\boldsymbol{v}^{k+1} &= \boldsymbol{v}^k - (\boldsymbol{x}^{k+1} - \boldsymbol{z}^{k+1}) \text{ // actualizacion dual}
\end{aligned}$$

# Lasso con ADMM (III)

Dado

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\|\mathbf{A}\boldsymbol{x}-\boldsymbol{b}\|^2+\lambda\|\boldsymbol{x}\|_1$$

Se obtiene

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}} \quad & f(\boldsymbol{x})+g(\boldsymbol{z}) \\ s.t: \quad & \boldsymbol{x}-\boldsymbol{z}=0 \end{aligned}$$

El Lagrangiano Aumentado es

$$\mathcal{L}_A(\boldsymbol{x},\boldsymbol{z},\boldsymbol{v}) \;=\; f(\boldsymbol{x})+g(\boldsymbol{z})+\frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{z}-\boldsymbol{v}\|^2+cte$$

# Lasso con ADMM (IV)

Se obtienen las funciones alternas para $\boldsymbol{x}$ y $\boldsymbol{z}$

$$
\begin{aligned}
f_1(\boldsymbol{x}) &= \frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 + \frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{z}^k - \boldsymbol{v}^k\|^2 \\
f_2(\boldsymbol{z}) &= \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{x}^{k+1} + \boldsymbol{v}^k\|^2 + \frac{\lambda}{\mu}\|\boldsymbol{z}\|_1
\end{aligned}
$$

Algoritmo Lasso con ADMM

$$
\begin{aligned}
\boldsymbol{x}^{k+1} &= (\mathbf{A}^T\mathbf{A} + \mu_k I)^{-1}(\mathbf{A}^T\boldsymbol{b} + \mu_k(\boldsymbol{z}^k + \boldsymbol{v}^k)) \\
\boldsymbol{z}^{k+1} &= S(\boldsymbol{x}^{k+1} - \boldsymbol{v}^k, \frac{\lambda}{\mu}), \text{ operador soft-thresholding} \\
\boldsymbol{v}^{k+1} &= \boldsymbol{v}^k - (\boldsymbol{x}^{k+1} - \boldsymbol{z}^{k+1}) \text{ // actualizacion dual}
\end{aligned}
$$

# Consensus ADMM (I)

Consideremos el problema

$$\min_{\boldsymbol{x}} \quad \sum_{i=1}^{N} f_i(\boldsymbol{x})$$

Reparametricemos el problema

$$\begin{aligned} \min_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N,\boldsymbol{x}} \quad & \sum_{i=1}^{N} f_i(\boldsymbol{x}_i) \\ s.t: \quad & \boldsymbol{x}_i = \boldsymbol{x},\ i = 1,\cdots,N \end{aligned}$$

# Consensus ADMM (II)

$$
\begin{aligned}
\mathcal{L}_A(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N, \boldsymbol{x}) &= \sum_{i=1}^{N} f_i(\boldsymbol{x}_i) - \boldsymbol{\lambda}_i^T(\boldsymbol{x}_i - \boldsymbol{x}) + \frac{\mu}{2}\|\boldsymbol{x}_i - \boldsymbol{x}\|^2 \\
&= \sum_{i=1}^{N} f_i(\boldsymbol{x}_i) + \frac{\mu}{2}\|\boldsymbol{x}_i - \boldsymbol{x} - \frac{\boldsymbol{\lambda}_i}{\mu}\|^2 + cte \\
&= \sum_{i=1}^{N} f_i(\boldsymbol{x}_i) + \frac{\mu}{2}\|\boldsymbol{x}_i - \boldsymbol{x} - \boldsymbol{v}_i\|^2 + cte
\end{aligned}
$$

con $\boldsymbol{v}_i = \frac{\boldsymbol{\lambda}_i}{\mu}$

# Consensus ADMM (III)

Luego

$$
\begin{aligned}
\boldsymbol{x}_i^{k+1} &= \arg\min f_i(\boldsymbol{x}_i) + \frac{\mu_k}{2}\|\boldsymbol{x}_i - \boldsymbol{x}^k - \boldsymbol{v}_i^k\|^2 \\
\boldsymbol{x}^{k+1} &= \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i^{k+1} - \boldsymbol{v}_i^k) \\
\boldsymbol{v}_i^{k+1} &= \boldsymbol{v}_i^k - (\boldsymbol{x}_i^{k+1} - \boldsymbol{x}^{k+1}) \text{ // actualizacion dual}
\end{aligned}
$$

# Consensus ADMM (IV)

A partir de la relacion

$$\boldsymbol{v}_i^{k+1} = \boldsymbol{v}_i^k - (\boldsymbol{x}_i^{k+1} - \boldsymbol{x}^{k+1})$$

y usando ademas $\boldsymbol{x}^{k+1} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i^{k+1} - \boldsymbol{v}_i^k)$, se tiene para toda $k$

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{v}_i^{k+1} &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{v}_i^k - (\boldsymbol{x}_i^{k+1} - \boldsymbol{x}^{k+1}) \\ &= -\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i^{k+1} - \boldsymbol{v}_i^k) + \boldsymbol{x}^{k+1} = 0 \end{aligned}$$

# Consensus ADMM (V)

Luego

$$\begin{aligned}
\boldsymbol{x}^{k+1} &= \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i^{k+1} - \boldsymbol{v}_i^{k}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i^{k+1} \\
\bar{\boldsymbol{x}}^{k+1} &:= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i^{k+1}
\end{aligned}$$

# Consensus ADMM (V)

Luego, el algoritmo simplificado quedaria

$$
\begin{aligned}
\boldsymbol{x}_i^{k+1} &= \arg\min f_i(\boldsymbol{x}_i) + \frac{\mu_k}{2}\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}^k - \boldsymbol{v}_i^k\|^2 \\
\bar{\boldsymbol{x}}^{k+1} &= \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i^{k+1} \\
\boldsymbol{v}_i^{k+1} &= \boldsymbol{v}_i^k - (\boldsymbol{x}_i^{k+1} - \bar{\boldsymbol{x}}^{k+1}) \text{ // actualizacion dual}
\end{aligned}
$$

# Consensus ADMM y regularizacion (I)

Consideremos el problema de consenso con regularizacion

$$\min_{\boldsymbol{x}} \quad \sum_{i=1}^{N} f_i(a_i^T \boldsymbol{x} + \boldsymbol{b}) + g(\boldsymbol{x})$$

Reparametrizando

$$\begin{aligned} \min_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N,\boldsymbol{x}} \quad & \sum_{i=1}^{N} f_i(a_i^T \boldsymbol{x}_i + \boldsymbol{b}) + g(\boldsymbol{x}) \\ s.t: \quad & \boldsymbol{x}_i = \boldsymbol{x},\ i = 1, \cdots, N \end{aligned}$$

# Consensus ADMM y regularizacion(II)

Definiendo $\boldsymbol{v}_i := \frac{\boldsymbol{\lambda}_i}{\mu}$

$$\begin{aligned}
\mathcal{L}_A(\{\boldsymbol{x}_i\}, \boldsymbol{x}) &= \sum_{i=1}^{N} f_i(a_i^T \boldsymbol{x}_i + \boldsymbol{b}) - \boldsymbol{\lambda}_i^T(\boldsymbol{x}_i - \boldsymbol{x}) + \frac{\mu}{2}\|\boldsymbol{x}_i - \boldsymbol{x}\|^2 + g(\boldsymbol{x}) \\
&= \sum_{i=1}^{N} f_i(a_i^T \boldsymbol{x}_i + \boldsymbol{b}) + \frac{\mu}{2}\|\boldsymbol{x}_i - \boldsymbol{x} - \frac{\boldsymbol{\lambda}_i}{\mu}\|^2 + g(\boldsymbol{x}) + cte \\
&= \sum_{i=1}^{N} f_i(a_i^T \boldsymbol{x}_i + \boldsymbol{b}) + \frac{\mu}{2}\|\boldsymbol{x}_i - \boldsymbol{x} - \boldsymbol{v}_i\|^2 + g(\boldsymbol{x}) + cte
\end{aligned}$$

# Consensus ADMM y regularizacion (III)

Luego

$$
\begin{aligned}
\boldsymbol{x}_i^{k+1} &= \arg\min_{\boldsymbol{x}_i} f_i(a_i^T \boldsymbol{x}_i + \boldsymbol{b}) + \frac{\mu_k}{2}\|\boldsymbol{x}_i - \boldsymbol{x}^k - \boldsymbol{v}_i^k\|^2;\ i = 1, 2, \cdots, N \\
\boldsymbol{x}^{k+1} &= \arg\min_{\boldsymbol{x}} \frac{\mu_k N}{2}\|\boldsymbol{x} - \boldsymbol{x}_i^{k+1} + \boldsymbol{v}_i^k\|^2 + g(\boldsymbol{x}) \\
\boldsymbol{v}_i^{k+1} &= \boldsymbol{v}_i^k - (\boldsymbol{x}_i^{k+1} - \boldsymbol{x}^{k+1});\ i = 1, 2, \cdots, N
\end{aligned}
$$

# Ejemplo: SVM regularizado con ADMM

$$\arg\min_{\omega,\omega_0} \; J(\omega,\omega_0) = \frac{1}{2}\|\omega\|^2,$$

s.a: $y_i(\omega^T \boldsymbol{x}_i + \omega_0) \geq 1,$ para $i = 1, 2, \cdots, N$

# Ejemplo: SVM regularizado con ADMM

Se propone una funcion que penaliza las restricciones

$$h(\omega, \omega_0) = \frac{1}{N}\sum_{i=1}^{N} \ell(y_i(\boldsymbol{x}_i^T \omega + \omega_0)) + \frac{1}{2}\|\omega\|_2^2$$

donde $\ell(\cdot)$ es la funcion Hinge (Hinge Loss), ie,

$$\ell(x) = (1-x)_+ = \max(0, 1-x)$$

Nota: Probar con logistic loss, $\ell(x) = \log(1 + e^{-x})$ y regularizacion $\ell_1$

$$
\begin{aligned}
h(\{\omega_i\}, \omega, \omega_0) &= \frac{1}{N}\sum_{i=1}^{N} \ell(y_i(\boldsymbol{x}_i^T \omega_i + \omega_0)) + \frac{1}{2}\|\omega\|_2^2 \\
s.t: &\quad \omega_i = \omega,\ i = 1, 2, \cdots, N
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}_A(\{\omega_i\}, \omega, \omega_0) &= \frac{1}{N}\sum_{i=1}^{N} \ell(y_i(\boldsymbol{x}_i^T \omega_i + \omega_0)) + \frac{\mu}{2}\|\omega_i - \omega - \boldsymbol{v}_i\|^2 \\
&\quad + \frac{1}{2}\|\omega\|_2^2
\end{aligned}
$$

# Sobre el operador soft-thresholding

Consideremos el problema en una dimension, con $\lambda \geq 0$,

$$S(a,\lambda) := x^* \quad = \quad \arg\min_x f(x) = \frac{1}{2}(x-a)^2 + \lambda|x|$$

Podemos escribir la funcion $f(x)$ como sigue

$$f(x) \quad = \quad \begin{cases} \frac{1}{2}(x-a)^2 + \lambda x & \text{Si } x \geq 0 \\ \frac{1}{2}(x-a)^2 - \lambda x & \text{Si } x < 0 \end{cases}$$

$$f'(x) \quad = \quad \begin{cases} x - a + \lambda & \text{Si } x \geq 0 \\ x - a - \lambda & \text{Si } x < 0 \end{cases}$$

- Si $a \geq \lambda$ entonces $x^* = a - \lambda \geq 0$
- Si $a < -\lambda$ entonces $x^* = a + \lambda < 0$
- Si $\lambda \leq |a|$ entonces $x^* = 0$

Luego

$$\begin{aligned} S(a,\lambda) &= \begin{cases} a-\lambda & \text{Si } a \geq \lambda \\ 0 & \text{Si } |a| < \lambda \\ a+\lambda & \text{Si } a < -\lambda \end{cases} = \begin{cases} a - sgn(a)\lambda & \text{Si } |a| \geq \lambda \\ 0 & \text{Si } |a| < \lambda \end{cases} \\ &= sgn(a) \begin{cases} |a| - \lambda & \text{Si } |a| \geq \lambda \\ 0 & \text{Si } |a| < \lambda \end{cases} \\ &= sgn(a) \max(0, |a| - \lambda) = sgn(a)(|a| - \lambda)_+ \end{aligned}$$