Programacion Cuadratica: Conjuntos activos

Oscar S. Dalmau Cedeño
dalmau@cimat.mx

18 de octubre de 2018

Un problema de programación cuadrática se puede formular como sigue:

$$\begin{aligned} \min q(\boldsymbol{x}) &= \frac{1}{2}\boldsymbol{x}^T \mathbf{G}\boldsymbol{x} - \boldsymbol{c}^T\boldsymbol{x} && (1)\\ s.a.: \ \boldsymbol{a}_i^T\boldsymbol{x} &= b_i,\ i \in \mathcal{E} && (2)\\ \boldsymbol{a}_i^T\boldsymbol{x} &\geq b_i,\ i \in \mathcal{I} && (3) \end{aligned}$$

donde $\mathbf{G} \in \mathbb{R}^{n\times n}$ es una matriz simetrica semi positiva definida (caso convexo) y $\boldsymbol{c}, \boldsymbol{x}, \{\boldsymbol{a}_i,\ i \in \mathcal{E}\cup\mathcal{I}\} \in \mathbb{R}^n$, $\{b_i,\ i \in \mathcal{E}\cup\mathcal{I}\} \in \mathbb{R}$.

El *conjuntos activo* consiste en los índices donde se activan las restricciones:

$$\mathcal{A}(\boldsymbol{x}^*) = \{i \in \mathcal{E} \cup \mathcal{I} |\ \boldsymbol{a}_i^T \boldsymbol{x}^* = b_i\} \tag{4}$$

Las condiciones de KKT para $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ son

$$\begin{aligned}
\nabla\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= \mathbf{G}\boldsymbol{x}^* - \boldsymbol{c} - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i^* \boldsymbol{a}_i = 0 && (5)\\
\boldsymbol{a}_i^T \boldsymbol{x}^* &= \boldsymbol{b}_i,\ i \in \mathcal{E} && (6)\\
\boldsymbol{a}_i^T \boldsymbol{x}^* &\geq \boldsymbol{b}_i,\ i \in \mathcal{I} && (7)\\
\lambda_i^* &\geq 0,\ i \in \mathcal{I} && (8)\\
\lambda_i^*(\boldsymbol{a}_i^T \boldsymbol{x}^* - \boldsymbol{b}_i) &= 0, i \in \mathcal{E}\cup\mathcal{I} && (9)
\end{aligned}$$

Si se cumple la condicion LICQ (checar esto) y se conoce el conjunto de indices donde se activan las restricciones $\mathcal{A}(\boldsymbol{x}^*)$ entonces podemos reescribir las KKT como sigue

$$
\begin{aligned}
\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= \mathbf{G}\boldsymbol{x}^* - \boldsymbol{c} - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \boldsymbol{a}_i = 0 && (10)\\
\boldsymbol{a}_i^T \boldsymbol{x}^* &= \boldsymbol{b}_i, \; i \in \mathcal{E} && (11)\\
\boldsymbol{a}_i^T \boldsymbol{x}^* &= \boldsymbol{b}_i, \; i \in \mathcal{I} \cap \mathcal{A}(\boldsymbol{x}^*) && (12)\\
\lambda_i^* &\geq 0, \; i \in \mathcal{I} \cap \mathcal{A}(\boldsymbol{x}^*) && (13)\\
\boldsymbol{a}_i^T \boldsymbol{x}^* &> \boldsymbol{b}_i, \; i \in \mathcal{I} \setminus \mathcal{A}(\boldsymbol{x}^*) && (14)\\
\lambda_i^* &= 0, \; i \in \mathcal{I} \setminus \mathcal{A}(\boldsymbol{x}^*) && (15)
\end{aligned}
$$

En caso de conocer el conjunto activo $\mathcal{A}(\boldsymbol{x}^*)$, el problema anterior es equivalente a eliminar las restricciones que no se activan es decir,

$$\begin{aligned}
\nabla\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) &= \mathbf{G}\boldsymbol{x}^* - \boldsymbol{c} - \sum_{i\in\mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \boldsymbol{a}_i = 0 && (16)\\
\boldsymbol{a}_i^T\boldsymbol{x}^* &= \boldsymbol{b}_i,\ i\in\mathcal{A}(\boldsymbol{x}^*) && (17)\\
\boldsymbol{a}_i^T\boldsymbol{x}^* &\geq \boldsymbol{b}_i,\ i\in\mathcal{I}\setminus\mathcal{A}(\boldsymbol{x}^*) && (18)\\
\lambda_i^* &\geq 0,\ i\in\mathcal{I}\cap\mathcal{A}(\boldsymbol{x}^*) && (19)
\end{aligned}$$

o simplemente

$$
\begin{aligned}
\mathbf{G}\boldsymbol{x}^* - \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \boldsymbol{a}_i &= \boldsymbol{c} && (20)\\
\boldsymbol{a}_i^T \boldsymbol{x}^* &= \boldsymbol{b}_i, \ i \in \mathcal{A}(\boldsymbol{x}^*) && (21)\\
\boldsymbol{a}_i^T \boldsymbol{x}^* &\geq \boldsymbol{b}_i, \ i \in \mathcal{I} \setminus \mathcal{A}(\boldsymbol{x}^*) && (22)\\
\lambda_i^* &\geq 0, \ i \in \mathcal{I} \cap \mathcal{A}(\boldsymbol{x}^*) && (23)
\end{aligned}
$$

donde $\lambda_i^* \in \mathbb{R}, \ i \in \mathcal{E}$, por lo que no es una restriccion y simplemente se puede eliminar del problema.

Teorema

Si $\boldsymbol{x}^*$ satisface las condiciones anteriores para algunos $\lambda_i^*,\ i \in \mathcal{A}(\boldsymbol{x}^*)$ y ademas $\mathbf{G}$ es semidefinida positiva, entonces $\boldsymbol{x}^*$ es un optimo global del problema original

Teorema

Sea x un punto factible y x^* satisfaciendo las condiciones del teorema

- Para $i \in \mathcal{E}$

$$\begin{aligned} \boldsymbol{a}_i^T \boldsymbol{x} &= \boldsymbol{a}_i^T \boldsymbol{x}^* = \boldsymbol{b} \\ \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{x}^*) &= 0 \end{aligned}$$

- Para $i \in \mathcal{I} \cap \mathcal{A}(\boldsymbol{x}^*)$

$$\begin{aligned} \boldsymbol{a}_i^T \boldsymbol{x}^* &= \boldsymbol{b} \\ \boldsymbol{a}_i^T \boldsymbol{x} &\geq \boldsymbol{b} = \boldsymbol{a}_i^T \boldsymbol{x}^* \\ \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{x}^*) &\geq 0 \end{aligned}$$

Teorema

- Como para toda $i \in \mathcal{A}(\boldsymbol{x}^*)$ se cumple $\lambda_i^* \geq 0$ entonces:

$$\lambda_i^* \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{x}^*) \quad \geq \quad 0$$

- Por otro lado, premultiplicando la primera KKT por $(\boldsymbol{x} - \boldsymbol{x}^*)^T$

$$(\boldsymbol{x} - \boldsymbol{x}^*)^T (\mathbf{G}\boldsymbol{x}^* - \boldsymbol{c}) \quad = \quad \sum_{i \in \mathcal{A}(\boldsymbol{x}^*)} \lambda_i^* \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0$$

- Reescribiendo la funcion objetivo

$$\begin{aligned} q(\boldsymbol{x}) \quad &= \quad q(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T \mathbf{G} (\boldsymbol{x} - \boldsymbol{x}^*) + (\boldsymbol{x} - \boldsymbol{x}^*)^T (\mathbf{G}\boldsymbol{x}^* - \boldsymbol{c}) \\ &\geq \quad q(\boldsymbol{x}^*) \end{aligned}$$

Luego $\boldsymbol{x}^*$ es una solucion global

Consideremos el caso convexo del problema cuadratico. Si el conjunto activo $\mathcal{A}(\boldsymbol{x}^*)$ se conociera de antemano, entonces se podría resolver el sistema (20)-(23) el cual es equivalente al problema original, o se podria resolver de forma equivalente, el siguiente problema cuadratico:

$$\min q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{G}\boldsymbol{x} - \boldsymbol{c}^T\boldsymbol{x} \quad (24)$$

$$s.a.: \boldsymbol{a}_i^T\boldsymbol{x} = b_i,\ i \in \mathcal{A}(\boldsymbol{x}^*) \quad (25)$$

Nota: verificando las condiciones (22)-(23) para que sea equivalente a las KKTs originales.

- Un problema del enfoque anterior es que no conocemos $\mathcal{A}(\boldsymbol{x}^*)$ de antemano.
- Sin embargo, se puede usar una estrategia basada en *suponer conocido el conjunto activo*, y *resolver el problema cuadratico anterior, para dicho conjunto activo*.

- El metodo consiste en resolver una secuencia de subproblemas cuadraticos con restricciones de igualdad o restricciones activas $\mathcal{W}_k$, *conocido como conjunto trabajo*, de modo que

$$\begin{aligned} \mathcal{W}_0 \to \mathcal{W}_1 \to \cdots \to \mathcal{W}_n = \mathcal{A}(\boldsymbol{x}^*) \\ \boldsymbol{x}^0 \to \boldsymbol{x}^1 \to \cdots \to \boldsymbol{x}^n = \boldsymbol{x}^* \end{aligned}$$

  Garantizando en cada caso que $\boldsymbol{x}^k$ es factitible, ie $\boldsymbol{a}_i^T \boldsymbol{x}^k \geq b_i$, $i \in \mathcal{I} \bigcap \mathcal{W}_k$
- El reto es encontrar un mecanismo que nos permita pasar de $\mathcal{W}_k$ a $\mathcal{W}_{k+1}$ y que al final encontremos el conjunto activo $\mathcal{A}(\boldsymbol{x}^*)$.

- El algoritmo consiste en resolver la secuencia de subproblemas

$$\begin{aligned} \boldsymbol{x}_{k+1} = arg\,\text{mín}\, q(\boldsymbol{x}) &= \frac{1}{2}\boldsymbol{x}^T\mathbf{G}\boldsymbol{x} - \boldsymbol{c}^T\boldsymbol{x} && (26) \\ s.a.:\ \boldsymbol{a}_i^T\boldsymbol{x} &= b_i,\ i \in \mathcal{W}_k && (27) \end{aligned}$$

dados $\mathcal{W}_k$ y $\boldsymbol{x}_k$ factible, ie, $\boldsymbol{a}_i^T\boldsymbol{x}_k \geq b_i$, $i \in \mathcal{I} \bigcap \mathcal{W}_k$

- El punto $\boldsymbol{x}_k$ se usa como punto inicial para un algoritmo iterativo y $\mathcal{W}_k$ es el *conjunto de trabajo* actual.

- Inicialmente se comprueba si $\boldsymbol{x}_k$ minimiza el problema anterior definido sobre el espacio de trabajo $\mathcal{W}_k$.
- En caso de que no sea el minimo, podemos reescribir el problema anterior, considerando que $\boldsymbol{x}_k$ es una buena aproximacion del optimo.
- Para ello se define

$$\boldsymbol{x} = \boldsymbol{x}_k + \boldsymbol{p} \tag{28}$$

y el problema (26)-(27) puede ser reescrito en terminos del del paso $\boldsymbol{p}$.

- Sustituyendo $\boldsymbol{x}$ en (26)-(27) se obtiene

$$\boldsymbol{p}_k = \arg\min_{\boldsymbol{p}} q(\boldsymbol{p}) \quad = \quad \frac{1}{2}\boldsymbol{p}^T \mathbf{G}\boldsymbol{p} - \mathbf{g}_k^T \boldsymbol{p} \tag{29}$$

$$s.a.: \ \boldsymbol{a}_i^T \boldsymbol{p} \quad = \quad 0, \ i \in \mathcal{W}_k \tag{30}$$

donde $\mathbf{g}_k \stackrel{def}{=} \mathbf{G}\boldsymbol{x}_k - \boldsymbol{c}$

- Si $\boldsymbol{p}_k = 0$ y basado en las KKTs, las λ_i se calculan solo en el conjunto de trabajo, ie

$$\mathbf{G}\boldsymbol{x} - \boldsymbol{c} - \sum_{i \in \mathcal{W}_k} \lambda_i \boldsymbol{a}_i = 0 \tag{31}$$

con $\boldsymbol{x} = \boldsymbol{x}_k + \boldsymbol{p}_k$ y en el optimo se debe verificar que $\lambda_i \geq 0, \ i \in \mathcal{I} \cap \mathcal{W}_k$, pues ya se cumple $\boldsymbol{a}_i^T \boldsymbol{x}_k \geq b_i$, $i \in \mathcal{I} \bigcap \mathcal{W}_k$, asumiendo $\lambda_i = 0$ se cumplen las condiciones de complementareidad para $i \in \mathcal{I} \bigcap \mathcal{W}_k$.

Consideremos el caso $\boldsymbol{p}_k \neq 0$

- Como $\boldsymbol{p}_k$ es ortogonal a $\boldsymbol{a}_i$ entonces el valor de $\boldsymbol{a}_i^T \boldsymbol{x}$ es constante para toda $\boldsymbol{x}$ en la linea que pasa por $\boldsymbol{x}_k$ en la direccion $\boldsymbol{p}_k$, es decir, si $\boldsymbol{x} = \boldsymbol{x}_k + \alpha \boldsymbol{p}_k$ entonces

$$\boldsymbol{a}_i^T \boldsymbol{x} \;=\; \boldsymbol{a}_i^T (\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) = \boldsymbol{a}_i^T \boldsymbol{x}_k = b_i \tag{32}$$

lo anterior garantiza que se satisfagan las restricciones de igualdad del problema original para la linea recta $\boldsymbol{x}_k + \alpha \boldsymbol{p}_k$ con $\alpha \in \mathbb{R}$.

- Basado en el comentario anterior, si $\boldsymbol{p}_k$ es el optimo del problema (29)-(30), podemos seleccionar $\boldsymbol{x}_{k+1}$ a lo largo de la recta, es decir,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha \boldsymbol{p}_k \tag{33}$$

  donde $\alpha_k > 0$ es el tamaño de paso.
- Para la seleccion de α_k se toma el valor mas grande tal que $\alpha_k \in [0, 1]$ y se satisfagan las restricciones del problema original.

- Teniendo en cuenta el primer punto: las restricciones de igualdad $i \in \mathcal{W}_k$ en el conjunto de trabajo siempre se satisfacen a lo largo de la recta, es decir, para todo $\alpha \in \mathbb{R}$.
- Por lo tanto, para garantizar que se satisfagan las restricciones solo debemos analizar las restricciones $i \notin \mathcal{W}_k$ o lo que es lo mismo $i \in \mathcal{I} \setminus \mathcal{W}_k$. Luego se debe cumplir,

$$\begin{aligned} \boldsymbol{a}_i^T \boldsymbol{x}_{k+1} &\geq b_i \; i \notin \mathcal{W}_k \\ \boldsymbol{a}_i^T(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) &\geq b_i \\ \boldsymbol{a}_i^T \boldsymbol{p}_k \alpha &\geq b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k \end{aligned}$$

Ahora podemos considerar 2 casos:

- Si $\boldsymbol{a}_i^T \boldsymbol{p}_k \geq 0$: Dado que $\boldsymbol{x}_k$ debe satisfacer todas las restricciones (en particular las de desigualdad) y $\alpha > 0$ entonces:

$$\begin{aligned} \boldsymbol{a}_i^T \boldsymbol{x}_k &\geq b_i \;\; i \in \mathcal{I} \\ \boldsymbol{a}_i^T \boldsymbol{x}_k + \alpha \boldsymbol{a}_i^T \boldsymbol{p}_k &\geq b_i \\ \boldsymbol{a}_i^T \boldsymbol{x}_{k+1} &\geq b_i \end{aligned}$$

  Por lo que en este caso, el nuevo punto $\boldsymbol{x}_{k+1}$ tambien satisface las restricciones de desigualdad

- Si $\boldsymbol{a}_i^T \boldsymbol{p}_k < 0$:

$$\begin{aligned} \boldsymbol{a}_i^T \boldsymbol{p}_k \alpha &\geq b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k \\ \alpha &\leq \frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k}{\boldsymbol{a}_i^T \boldsymbol{p}_k} \end{aligned}$$

luego α_k es menor o igual que el minimo $\frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k}{\boldsymbol{a}_i^T \boldsymbol{p}_k}$, ie

$$\alpha_k \leq \alpha_{min} \stackrel{def}{=} \min_{i \notin \mathcal{W}_k | \boldsymbol{a}_i^T \boldsymbol{p}_k < 0} \frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k}{\boldsymbol{a}_i^T \boldsymbol{p}_k}$$

Como $0 < \alpha_k \leq 1$ entonces

$$\begin{aligned} \alpha_k &\stackrel{def}{=} \min\left(1, \alpha_{min}\right) \\ &= \min\left(1, \min_{i \notin \mathcal{W}_k | \boldsymbol{a}_i^T \boldsymbol{p}_k < 0} \frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k}{\boldsymbol{a}_i^T \boldsymbol{p}_k}\right) \end{aligned} \tag{34}$$

Al conjunto de restricciones $i \notin \mathcal{W}_k | \boldsymbol{a}_i^T \boldsymbol{p}_k < 0$ que satisfacen $\alpha_k = \alpha_{min}$ se les llama *restricciones de bloqueo*, al cual denotaremos por $\mathcal{B}_k$, luego

$$\mathcal{B}_k = \{i \notin \mathcal{W}_k | \boldsymbol{a}_i^T \boldsymbol{p}_k < 0 \text{ y } \alpha_k = \alpha_{min}\} \tag{35}$$

1. Si $\alpha_{min} > 1$ entonces el *conjunto de restricciones de bloqueo* $\mathcal{B}_k$ es **vacio** y $\alpha_k = 1$.
2. Si $\alpha_{min} = 1$ entonces el *conjunto de restricciones de bloqueo* $\mathcal{B}_k$ es **no vacio** y $\alpha_k = 1$
3. Si $\alpha_{min} < 1$ entonces el *conjunto de restricciones de bloqueo* $\mathcal{B}_k$ es **no vacio** y $\alpha_k = \alpha_{min}$.

Notar que se puede dar el caso en que una restriccion $j \in \{i \notin \mathcal{W}_k | \boldsymbol{a}_i^T \boldsymbol{p}_k < 0\}$ (no considerada en el conjunto de trabajo actual $\mathcal{W}_k$) sea activa, ie, $b_j - \boldsymbol{a}_j^T \boldsymbol{x}_k = 0$ y en este caso $\alpha_k = \alpha_{min} = 0$.

Una forma de actualizar el conjunto de trabajo en la proxima iteracion es añadir al conjunto de trabajo actual el conjunto de las restricciones de bloqueo:

$$\mathcal{W}_{k+1} = \mathcal{W}_k \cup \mathcal{B}_k \tag{36}$$

o tambien

$$\mathcal{W}_{k+1} = \mathcal{W}_k \cup \{j\},\ j \in \mathcal{B}_k \tag{37}$$

1. El proceso de obtencion de nuevos conjuntos de trabajo puede seguir sucesivamente como se explico anteriormente, ie, añadiendo restricciones de bloqueo en cada iteracion.
2. Luego se determina una dirección de descenso $\boldsymbol{p}_k$ mediante la solucion del subproblema (29)-(30) para el nuevo conjunto de trabajo.

- La pregunta es, cuando termina el algoritmo?.
- Si $\boldsymbol{p}_k = \mathbf{0}$ es el optimo del subproblema (29)-(30) con conjunto de trabajo $\mathcal{W}_k$, entonces $\boldsymbol{x}_k$ es optimo del problema equivalente (26)-(27).
- Sin embargo, como se ha mencionado anteriormente, para que se cumplan las condiciones de optimalidad del problema original se debe verificar que $\lambda_i \geq 0, \ i \in \mathcal{I} \cap \mathcal{W}_k$.

Para ello, se calculan los multiplicadores de Lagrange, usando las primeras KKT de subproblema (26)-(27).

$$\begin{aligned} [\nabla q(\boldsymbol{x}) - \sum_{i \in \mathcal{W}_k} \lambda_i \nabla(\boldsymbol{a}_i^T \boldsymbol{x} - b_i)]|_{\boldsymbol{x}=\boldsymbol{x}_k} &= 0 \\ \mathbf{G}\boldsymbol{x}_k - \boldsymbol{c} - \sum_{i \in \mathcal{W}_k} \lambda_i \boldsymbol{a}_i &= 0 \end{aligned}$$

o en forma matricial

$$\mathbf{A}_k \boldsymbol{\lambda}^k = \mathbf{g}_k \tag{38}$$

donde

$$\mathbf{A}_k = [\boldsymbol{a}_i]_{i\in\mathcal{W}_k} \tag{39}$$
$$\boldsymbol{\lambda}^k = [\lambda_i]_{i\in\mathcal{W}_k} \tag{40}$$
$$\mathbf{g}_k = \mathbf{G}\boldsymbol{x}_k - \boldsymbol{c} \tag{41}$$

como se ha supuesto que $\{\boldsymbol{a}_i\}$ son l.i. entonces

$$\boldsymbol{\lambda}^k = (\mathbf{A}_k^T\mathbf{A}_k)^{-1}\mathbf{A}_k^T\mathbf{g}_k \tag{42}$$

- Si $\lambda_i \geq 0, \; \forall i \in \mathcal{I} \cap \mathcal{W}_k$ entonces $\boldsymbol{x}^* = \boldsymbol{x}_k$
- En caso contrario, definamos el conjunto

$$\mathcal{C}_k \;\; = \;\; \{i \in \mathcal{I} \cap \mathcal{W}_k | \; \lambda_i < 0\} \tag{43}$$

  que no cumplen las KKT's.
- Ahora podemos actualizar el conjunto de trabajo, eliminando una o varias restricciones que no cumplen las KKTs

- Es decir

$$\mathcal{W}_{k+1} = \mathcal{W}_k \setminus \mathcal{C}_k \tag{44}$$

o tambien eliminar una sola restricción

$$\mathcal{W}_{k+1} = \mathcal{W}_k \setminus \{j\}, \; j \in \mathcal{C}_k \tag{45}$$

- Por lo general se selecciona el indice que corresponde al menor λ_i, ie

$$\mathcal{W}_{k+1} = \mathcal{W}_k \setminus \{j^*\} \tag{46}$$

donde $j^* = \arg\min_{j \in \mathcal{I} \cap \mathcal{W}_k} \boldsymbol{\lambda}_j$.

Algoritmo de Conjuntos Activos para QP convexa

Dado $\boldsymbol{x}_0$ y $\mathcal{W}_0$ por ejemplo $\mathcal{W}_0 = \mathcal{E}$
for $k = 0, 1, 2, \cdots$ **do**
  Hallar $\boldsymbol{p}_k$ resolviendo el subproblema (29)-(30)
  if $\boldsymbol{p}_k = \boldsymbol{0}$ **then**
    Calcular $\boldsymbol{\lambda}^k$ usando Ecs. (39)-(42) y Hallar $\mathcal{C}_k$, Ec. (43)
    if $\mathcal{C}_k = \emptyset$ **then**
      return $\boldsymbol{x}^* = \boldsymbol{x}_k$
    else
      $\mathcal{W}_{k+1} = \mathcal{W}_k \setminus \{j^*\}$, donde $j^* = \arg\min_{j \in \mathcal{I} \cap \mathcal{W}_k} \boldsymbol{\lambda}_j$.
    end if
    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k$
  else
    Hallar α_k, Ec. (34)
    Actualizar $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$
    Hallar $\mathcal{B}_k$, Ec. (35)
    if $\mathcal{B}_k = \emptyset$ **then**
      $\mathcal{W}_{k+1} = \mathcal{W}_k$
    else
      $\mathcal{W}_{k+1} = \mathcal{W}_k \cup \{j\},\ j \in \mathcal{B}_k$
    end if
  end if
end for

- Punto inicial x_0: Usar algoritmo Fase I, similar al caso de Programacion Lineal

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}} \mathbf{1}^T \boldsymbol{z} & \\ s.a: \quad & \boldsymbol{a}_i^T \boldsymbol{x} + \gamma_i z_i = b_i \; i \in \mathcal{E} \\ & \boldsymbol{a}_i^T \boldsymbol{x} + \gamma_i z_i \geq b_i \; i \in \mathcal{I} \\ & \boldsymbol{z} \succeq \mathbf{0} \end{aligned}$$

donde $\gamma_i = -signo(\boldsymbol{a}_i^T \tilde{\boldsymbol{x}} - b_i)$ para $i \in \mathcal{E}$, $\gamma_i = 1$ para $i \in \mathcal{I}$. El punto inicial del problema anterior es:

$$\boldsymbol{x} = \tilde{\boldsymbol{x}}, \; z_i = |\boldsymbol{a}_i^T \tilde{\boldsymbol{x}} - b_i|, \; i \in \mathcal{E}; \; z_i = max(b_i - \boldsymbol{a}_i^T \tilde{\boldsymbol{x}}, 0), \; i \in \mathcal{I}$$

Se puede ver que si $\tilde{x}$ es factible para el problema original $(\tilde{\boldsymbol{x}}, 0)$ es optimo para el problema anterior

- Sobre Degenaracion: en principio en un problema actual, aunque hay varias estrategias (ver tesis doctoral en los documentos: una en Canada y la otra en Stanford).
- Una variante puede ser usar un metodo de perturbacion,

$$\boldsymbol{a}_i^T \boldsymbol{x} = b_i + \epsilon^i$$

  con $\epsilon > 0$ un valor pequeño. Ver variante en el Nocedal (seccion **degenerate steps and cycling**, pagina 381).

Tarea

- Implementar el Algoritmo del Conjuntos activos
- Realizar experimentos con el modelo SVM
  - Generar datos linealmente separables. Ejecutar el algoritmo para SVM y mostrar datos, vectores de soporte y el hiperplano separador (SVM)
  - Generar datos no linealmente separables. Ejecutar el algoritmo para SVM no linealmente separable y mostrar datos, vectores de soporte y el hiperplano separador (SVM)
  - Usar datos Mnist con dos clases para entrenar el SVM. Considerar datos de entrenamiento y datos de prueba. Reportar el accuracy