

PROGRAMACIÓN CUADRÁTICA

Los problemas de programación cuadrática (QP) son aquellos que buscan minimizar una función objetivo cuadrática sujeta a restricciones lineales. El problema general de programación cuadrática se puede plantear de la siguiente forma (la numeración de las fórmulas corresponde a la del libro "Numerical Optimization" de J. Nocedal y S. Wright, disponible en la biblioteca CIMAT):

$$\min_x q(x) = \frac{1}{2}x^T Gx + x^T d \quad (16.1a)$$

$$\text{sujeto a } a_i^T x = b_i, \quad i \in \varepsilon, \quad (16.1b)$$

$$a_i^T x \geq b_i, \quad i \in I, \quad (16.1c)$$

donde G es una matriz simétrica de $n \times n$, ε e I son conjuntos finitos de índices, $d, x, \{a_i\}$, $i \in \varepsilon \cup I$, son vectores con n elementos. Si la matriz G es definida positiva, el problema es convexo y aproximadamente tan fácil de resolver como un problema de programación cuadrática. De lo contrario el problema puede presentar mínimos locales y ser más difícil de resolver.

PROGRAMACIÓN CUADRÁTICA CON RESTRICCIONES DE IGUALDAD (QPS)

Primero se revisará el caso de (QP) cuando solo hay restricciones de igualdad presentes. Este caso particular es útil porque funciona como auxiliar en casos más generales donde para resolver un problema más grande, se resuelven subproblemas con restricciones de igualdad.

Se supondrá que el número de restricciones es m , y se asumirá que $m \leq n$, y el problema de programación cuadrática se establecerá como sigue:

$$\min_x q(x) = \frac{1}{2}x^T Gx + x^T d \quad (16.3a)$$

$$\text{sujeto a } Ax = b, \quad (16.3b)$$

donde A es la matriz Jacobiana de $m \times n$ definida como:

$$A = [a_i]_{i \in \varepsilon}^T.$$

Se asumirá que A es de rango completo y que las restricciones son consistentes. El Lagrangiano y su gradiente serán, respectivamente:

$$\mathcal{L}(x) = \frac{1}{2}x^T Gx + x^T d - \lambda^T (Ax - b)$$

$$\nabla \mathcal{L}(x) = Gx + d - A^T \lambda$$

donde λ es un vector de multiplicadores de Lagrange. Aplicando las condiciones de Karush–Kuhn–Tucker (KKT) obtenemos el siguiente sistema:

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \quad (16.4)$$

donde x^* es una solución del sistema y λ^* es el vector de multiplicadores de Lagrange en la solución. Ahora, se descompone el vector $x^* = x + p$, donde x es alguna estimación de la solución, y p es el paso que se está buscando para llegar a la solución. Introduciendo esta notación en (16.4) tenemos:

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix} \quad (16.5)$$

donde:

$$c = Ax - b, \quad g = d + Gx, \quad p = x^* - x. \quad (16.6)$$

La matriz de (16.5) se conoce como la matriz de Karush–Kuhn–Tucker (KKT). A continuación se mencionan dos lemas que son importantes para los cuales la matriz es no singular y la solución x^* es global. Z denota la matriz de $n \times (n-m)$ cuyas columnas son una base para el espacio nulo de A .

Lema 16.1.

Sea A una matriz de rango por renglones y asumiendo que el Hessiano reducido $Z^T G Z$ es positivo definido. Entonces la matriz KKT:

$$K = \begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \quad (16.7)$$

es no singular, y hay un único par de vectores (x^*, λ^*) satisfaciendo (16.4).

Lema 16.2.

Supóngase que las condiciones del lema 16.1 son satisfechas. Entonces el vector x^* que satisface (16.4) es la única solución global de (16.3).

Se puede encontrar las demostraciones a estos lemas en el libro "Numerical Optimization" en el capítulo 16.1.

RESOLVIENDO EL SISTEMA KKT.

Existen varios métodos para resolver el sistema KKT (16.4) o el (16.5). Primero mencionaremos un lema que es útil para saber que método conviene usar.

Lema 16.3.

Supóngase que A tiene rango completo en los renglones y que el Hessiano reducido $Z^T G Z$ es positivo definido. Entonces la matriz KKT (16.7) tiene n eigen valores positivos, m negativos y ninguno igual a cero.

Solución Directa

Una opción es realizar una factorización y hacer sustitución hacia adelante y hacia atrás. La factorización de Cholesky no funciona ya que la matriz del sistema (16.4-5) es indefinida. Por otro lado, la eliminación Gaussiana con pivoteo no toma en cuenta la simetría del sistema. Lo mas efectivo es usar una factorización simétrica indefinida como se describen en el capítulo 6 del libro "Numerical Optimization". Para una matriz simétrica K , esta factorización tiene la forma:

$$P^T K P = L B L^T \quad (16.12)$$

donde P es una matriz de permutación, L es una matriz unitaria inferior triangular, y B es una matriz diagonal a bloques, con bloques de 1×1 o 2×2 . Las permutaciones simétricas definidas por la matriz P son para mantener la estabilidad numérica y para mantener la dispersidad. Este tipo de factorización es aproximadamente la mitad de una eliminación Gaussiana. Para resolver (16.5) calculamos la factorización (16.12), sustituyendo la matriz KKT por K . Entonces realizamos la siguiente secuencia de operaciones para llegar a la solución:

$$\text{resolver } Ly = P^T \begin{bmatrix} g \\ c \end{bmatrix} \text{ para obtener } y;$$

$$\text{resolver } By^{\hat{}} = y \text{ para obtener } y^{\hat{}};$$

$$\text{resolver } L^T y^- = y^{\hat{}} \text{ para obtener } y^-;$$

$$\text{hacer } \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = P y^-.$$

La mayoría de las operaciones que implica este método pueden implementarse de manera muy eficiente. La multiplicación por la matriz de permutación P y su transpuesta, por ejemplo, implica solamente el reacomodo de renglones o columnas. Sin embargo, puede ser difícil definir la matriz de permutación P .

Método del Espacio de Rangos.

En el método del espacio de rangos, se usa la matriz G para realizar una eliminación por bloques en el sistema (16.5). Asumiendo que G es positiva definida, se multiplica la primera ecuación en (16.5) por AG^{-1} y entonces se extrae la segunda ecuación para obtener un sistema lineal en el vector λ^* :

$$(AG^{-1}A^T)\lambda^* = (AG^{-1}g - c). \quad (16.13)$$

Se resuelve este sistema simétrico positivo definido para λ^* , y se recupera p de la primera ecuación en (16.5) resolviendo

$$Gp = A^T \lambda^* - g \quad (16.4)$$

Este método requiere realizar operaciones con G^{-1} , así como calcular la factorización de la matriz de $m \times m$, $AG^{-1}A^T$. Por lo tanto es mas útil cuando:

- G es bien condicionada y fácil de invertir, (cuando G es diagonal o diagonal por bloques).
- G^{-1} es conocida explícitamente a través de una fórmula de actualización de cuasi-Newton o
- el número de ecuaciones de igualdad m es pequeño, de manera que el número de resoluciones hacia atrás necesario para formar la matriz $AG^{-1}A^T$ no es muy grande.

PROBLEMAS CON RESTRICCIONES DE DESIGUALDAD

En el resto del capítulo se discutirán varios algoritmos para resolver problemas de programación cuadrática que contienen restricciones de desigualdad. El método de conjuntos activos es ya clásico y puede ser aplicado tanto en métodos convexos como no convexos. El método de proyección de gradiente trata de acelerar la solución, permitiendo cambios rápidos en el conjunto activo. Los métodos de punto interior han demostrado ser efectivos para resolver problemas cuadráticos convexos grandes.

CONDICIONES DE OPTIMALIDAD PARA PROBLEMAS CON RESTRICCIONES DE DESIGUALDAD

Se discutirán algunas de las propiedades de los problemas de programación cuadrática con restricciones de desigualdad. Aplicando el Lagrangiano a (16.1), tenemos:

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^T Gx + x^T d - \sum_{i \in I \cup \varepsilon} \lambda_i (a_i^T x - b_i) \quad (16.24)$$

Además, definimos el conjunto activo $A(x^*)$ en un punto óptimo x^* como los índices de las restricciones en las cuales la igualdad se cumple, es decir,

$$A(x^*) = \{i \in \varepsilon \cup I : a_i^T x^* = b_i\} \quad (16.25)$$

Simplificando las condiciones necesarias de primer orden, incluimos que cualquier solución x^* de (16.1) satisfacen las siguientes condiciones de primer orden:

$$Gx^* + d - \sum_{i \in A(x^*)} \lambda_i^* a_i = 0, \quad (16.26a)$$

$$a_i^T x^* = b_i, \text{ para toda } i \in A(x^*), \quad (16.26b)$$

$$a_i^T x^* \geq b_i, \text{ para toda } i \in I \setminus A(x^*), \quad (16.26c)$$

$$\lambda_i^* \geq 0, \text{ para toda } i \in I \cap A(x^*), \quad (16.26d)$$

Las condiciones suficientes de segundo orden para que x^* sea un minimizador local son satisfechas si $Z^T G Z$ es positiva definida, donde Z es una matriz base del espacio nulo para la matriz Jacobiana de restricciones activas

$$[a_i]_{i \in A(x^*)}^T.$$

x^* es la solución global cuando las restricciones son de igualdad. Cuando G no es positiva definida, pueden existir varios mínimos locales estrictos en los cuales las condiciones necesarias de segundo orden son satisfechas. Estos problemas son conocidos como como no "convexos" o "indefinidos" y pueden causar complicaciones en los algoritmos.

DEGENERACIÓN

Otra propiedad que puede causar problemas para algunos algoritmos es la degeneración. Esto se refiere a situaciones en las que pueden ocurrir alguna de estas cosas

- los gradientes de las restricciones activas a_i , $i \in A(x^*)$, son linealmente dependientes en la solución x^* , o
- la condición de complementariedad estricta no se cumple, o sea que el vector de multiplicadores de Lagrange en el óptimo contiene elementos $\lambda_i = 0$ para algunos índices $i \in A(x^*)$.

La degeneración puede traer como consecuencia, dificultad en el cálculo de la matriz de espacio nulo Z , lo que a su vez provoca que la matriz $AG^{-1}A^T$ del método del espacio de rangos sea singular. Cuando las restricciones son débilmente activas, puede ser difícil para un algoritmo decidir si esas restricciones son activas o no.

MÉTODO DE LOS CONJUNTOS ACTIVOS PARA PROGRAMACIÓN CUADRÁTICA

Estos métodos son efectivos para problemas pequeños y medianos. Asumen inicialmente que G es semidefinida positiva. Escencialmente trabajan con subconjuntos de restricciones que están activas, y actualizan este subconjunto hasta dar con la solución óptima. El método simplex es un método de conjunto activos, que entre sus características esta que la solución siempre se encuentra en uno de los vértices del poliotipo que forman las restricciones.

Los tres métodos de conjuntos activos que se manejan son:

- Primal dual- (x, λ)
- Primal- (x)
- Duales- (λ)

Usualmente se asume un punto factible x_0 y se continúa con una secuencia que permanece factible. Cada paso consiste en resolver un subproblema cuadrático con restricciones de igualdad. A cada conjunto de restricciones se le denomina "working set" o conjunto activo y es denotado en la k -ésima iteración como W_k y contiene a todas las restricciones de igualdad y algunas de desigualdad (activas).

Dada x_k y W_k revisamos si x_k minimiza $q(x) = x^T Gx + x^T d$ en el subespacio definido por W_k , si no, calculamos un paso p resolviendo el subproblema con restricciones de igualdad. Definimos:

$$p = x - x_k, \quad q_k = Gx_k + d$$

luego

$$q(x) = q(x_k + p) = \frac{1}{2}p^T Gp + g_k^T p + c$$

donde $c = \frac{1}{2}x_k^T Gx_k + d^T x_k$ es una constante, y por lo tanto podemos desecharla sin alterar el mínimo del problema. Se escribe el subproblema de programación cuadrática a resolver en la k -ésima iteración como:

$$\begin{aligned} \min_p \quad & \frac{1}{2}p^T Gp + g_k^T p \\ \text{sujeto a } & a_i^T p = 0 \text{ para toda } i \in W_k \end{aligned}$$

Y se le llama p_k a la solución de este subproblema. Supóngase que $p_k \neq 0$. Debe decidirse que tan lejos se avanza en esta dirección. Si $x_k + p_k$ es factible para todas las restricciones, entonces la variable x_{k+1} será igual a esta suma. De lo contrario, el valor de x_{k+1} será igual a:

$$x_{k+1} = x_k + \alpha_k p_k$$

donde α es el valor mas alto en $[0, 1)$ que no viola ninguna restricción. Para cualquier valor de alfa, las restricciones del conjunto activo se cumplen, por lo tanto, para todas las restricciones que no estan en el conjunto activo se debe cumplir que $a_i^T(x_k + \alpha_k p_k) \geq a_i^T x_k \geq b_i$. De esto se deriva que para todas las restricciones fuera del conjunto activo, se debe cumplir que:

$$\alpha_k \leq \frac{b_i - a_i^T x_k}{a_i^T p_k}$$

Como se busca hacer α_k lo más grande posible en el intervalo $[0, 1]$ sujeto a mantener la factibilidad, se tiene la siguiente definición:

$$\alpha_k = \min \left(1, \min_{i \notin W_k, a_i^T p_k < 0} \frac{b_i - a_i^T x_k}{a_i^T p_k} \right) \quad (16.29)$$

Se le llama restricción bloqueadora a aquella para la cual el mínimo en (16.29) es alcanzado. Si $\alpha = 1$, entonces no hay restricciones bloqueadoras y no se agrega ningún nuevo elemento al conjunto activo. Si $\alpha < 1$, el conjunto activo es actualizado agregando una de las restricciones bloqueadoras. Seguimos iterando de esta manera hasta encontrar un punto que minimiza la función objetivo para el conjunto activo actual. Esto se detecta fácilmente, ya que el vector p_k será igual a cero.

Lo que sigue es examinar los multiplicadores de Lagrange del conjunto activo actual. Si alguno de los multiplicadores de las restricciones de desigualdad es menor a cero, entonces la condición (16.26d) no se cumple. Se debe retirar tal restricción del conjunto activo y seguir con el algoritmo.

A continuación se presenta un pseudocódigo del método de los conjuntos activos:

CONJUNTOS ACTIVOS

Calcular un punto inicial factible x_0 ;

Hacer W_0 un subconjunto de las restricciones activas en x_0

para $k = 0, 1, 2, \dots$

 Resolver (16.27) y encontrar p_k ;

si $p_k = 0$

 Calcular los multiplicadores de Lagrange $\hat{\lambda}_i$ que satisface (16.30),

 Hacer $\hat{W} = W_k$;

si $\hat{\lambda}_i \geq 0$ para todo $i \in W_k \cap I$;

 DETENER con las solución $x^* = x_k$;

De lo contrario

 Hacer $j = \arg \min_{j \in W_k \cap I} \hat{\lambda}_j$;

$x_{k+1} = x_k$; $W_{k+1} \leftarrow W_k \setminus \{j\}$;

De lo contrario (* $p_k \neq 0$ *)

 Calcular α_k de (16.29);

$x_{k+1} \leftarrow x_k + \alpha_k p_k$;

Si existen restricciones bloqueadoras

 Obtener W_{k+1} agregando una de las restricciones bloqueadoras

a W_{k+1} ;

De lo contrario

$W_{k+1} \leftarrow W_k$

fin (*para*)

MÉTODO DE PROYECCIÓN DE GRADIENTE

Este método está diseñado para realizar cambios rápidos en el conjunto de restricciones activas. Es muy eficiente en problemas de restricciones de cota y puede aplicarse por igual a problemas convexos y no convexos. Se mantendrá el enfoque en problemas del tipo siguiente:

$$\min_x q(x) = \frac{1}{2}x^T Gx + x^T d \quad (16.43a)$$

$$\text{sujeto a } l \leq x \leq u, \quad (16.43b)$$

donde G es simétrica y l y u son vectores de límite superior e inferior en los componentes de x , la zona factible tiene forma de caja. Cada iteración del método de proyección de gradiente consiste en dos etapas. En la primera se mueve en la dirección de máximo descenso a partir del punto actual x . Cuando un límite es alcanzado, la dirección de búsqueda es "doblada" de forma que se mantiene factible. Buscamos a lo largo de este camino a pedazos y localizamos el primer minimizador local de q al que se llamará x^c , también conocido como punto de Cauchy. En la segunda etapa se "explora" la cara de la caja donde se encuentra el punto de Cauchy resolviendo un subproblema en el cual los componentes activos x_i para cada $i \in A(x^c)$ están fijos a los valores x_i^c . Se usarán super índices para indicar el número de iteración y subíndices para indicar elementos de un vector.

CÁLCULO DEL PUNTO DE CAUCHY

La proyección de un punto arbitrario x dentro de la región factible Ω (16.43b) esta dada por

$$P(x, l, u) = \begin{cases} l_i & x_i < l_i \\ x_i & l_i \leq x_i \leq u_i \\ u_i & x_i \geq u_i \end{cases}$$

para el i -ésimo componente de x_i .

Luego el camino recto a pedazos, que inicia en x_0 , a lo largo de $-g$, donde $g = Gx + d$, está dado por:

$$x(t) = P(x^0 - tg, l, u)$$

Nótese que el problema de cota puede tener a lo más n restricciones activas.

Ahora se calcula el punto de Cauchy x^c , el cual se define como el primer minimizador local de la función cuadrática a pedazos $q(x(t))$, para $t \geq 0$. Este minimizador se obtiene examinando cada uno de los segmentos de línea que forman $x(t)$. Para realizar esta búsqueda, se determinan los valores t para los cuales ocurren los dobleces o quiebres. Primero se identifica los valores de t para los cuales cada componente alcanza su límite a lo largo de la dirección $-g$. Estos valores \hat{t}_i se obtienen por las siguientes fórmulas explícitas:

$$\hat{t}_i = \begin{cases} (x_i^0 - u_i)/g_i & \text{si } g_i < 0 \text{ y } u_i < \infty \\ (x_i^0 - l_i)/g_i & \text{si } g_i > 0 \text{ y } u_i > -\infty \\ \infty & \text{si no ocurre lo anterior} \end{cases} \quad (16.46)$$

Los componentes de $x(t)$ para cualquier t son:

$$x_i(t) = \begin{cases} x_i^0 - tg_i & \text{si } t \leq \hat{t}_i \\ x_i^0 - \hat{t}g_i & \text{de lo contrario} \end{cases}$$

Para buscar el primer minimizador local a lo largo de $P(x^0 - tg, l, u)$, se eliminan los valores duplicados y cero de \hat{t}_i del conjunto $\{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_n\}$ y se ordenan los elementos restantes en una secuencia ordenada t_1, t_2, \dots tal que $0 \leq t_1 \leq t_2 \dots$, luego se examinan los intervalos $[0, t_1], [t_1, t_2], \dots$ en turno. Suponiendo que se han examinado los intervalos hasta $[t_{j-2}, t_{j-1}]$ para algún j , y determinado que el minimizador local esta en algún valor $t \geq t_{j-1}$. Para el intervalo $[t_{j-1}, t_j]$ entre el $(j-1)$ avo punto de quiebre, tenemos que

$$x(t) = x(t_{j-1}) + \Delta t p^{j-1}$$

donde

$$\Delta t = t - t_{j-1}, \quad \Delta t \in [0, t_j - t_{j-1}],$$

y

$$p^{j-1} = \begin{cases} -g_i & \text{si } t_{j-1} \hat{t}_i, \\ 0 & \text{de lo contrario} \end{cases} \quad (16.47)$$

Usando esta notación, escribimos la cuadrática (16.43a) en el segmento de línea $[x(t_{j-1}), x(t_j)]$ como

$$q(x(t)) = d^T(x(t_{j-1}) + \Delta t p^{j-1}) + \frac{1}{2}(x(t_{j-1}) + \Delta t p^{j-1})^T G(x(t_{j-1}) + \Delta t p^{j-1}),$$

donde $\Delta t \in [0, t_j - t_{j-1}]$.

Expandiendo y agrupando los coeficientes de 1, Δt , y $(\Delta t)^2$, encontramos que

$$q(x(t)) = f_{j-1} + f'_{j-1} \Delta t + \frac{1}{2} f''_{j-1} (\Delta t)^2, \quad \Delta t \in [0, t_j - t_{j-1}], \quad (16.48)$$

donde los coeficientes f_{j-1} , f'_{j-1} , y f''_{j-1} se definen como

$$\begin{aligned} f_{j-1} &= d^T x(t_{j-1}) + \frac{1}{2} x(t_{j-1})^T G x(t_{j-1}), \\ f'_{j-1} &= d^T p^{j-1} + x(t_{j-1})^T G p^{j-1}, \\ f''_{j-1} &= (p^{j-1})^T G p^{j-1}. \end{aligned}$$

diferenciando (16.48) e igualando a cero

$$\Delta t^* = -f_{j-1} / f''_{j-1}$$

si $\Delta t^* \in [0, t_j - t_{j-1}]$ tenemos el primer mínimo local x^c . Si Δt^* no cae en el intervalo, se busca en el intervalo siguiente $[t_j, t_{j+1}]$ hasta encontrar un x^c .

Ahora se continúa con el segundo paso del método. Dado x^c y su conjunto activo $A(x^c)$ se resuelve aproximadamente el siguiente problema de programación cuadrática:

$$\min_x q(x) = \frac{1}{2}x^T Gx + x^T d \quad (16.49a)$$

$$\text{sujeto a } x_i = x_i^c, \quad i \in A(x^c) \quad (16.49b)$$

$$l_i \leq x_i \leq u_i, \quad i \notin A(x^c) \quad (16.49c)$$

A continuación se presenta el método de proyección de gradiente para programación cuadrática.

Calcular un punto inicial factible x^0 ;

para $k = 0, 1, 2, \dots$

si x^k satisface las condiciones KKT para (16.43)

 PARAR con la solución $x^* = x^k$;

 Hacer $x = x^k$ y encontrar el punto de Cauchy x^c ;

 Encontrar una solución aproximada x^+ de (16.49) tal que $q(x^+) < q(x^c)$

y x^+ es factible;

$x^{k+1} \leftarrow x^+$;

fin (para)

MÉTODOS DE PUNTO INTERIOR

Se pueden usar los métodos de punto interior primales duales para resolver los problemas de programación cuadrática. Se considerará el caso convexo (G es semipositiva definida). Planteamos el problema de la siguiente forma:

$$\min_x q(x) = \frac{1}{2}x^T Gx + d^T x \quad (16.50a)$$

$$\text{sujeto a } Ax \geq b \quad (16.50b)$$

G es positiva definida, A es una matriz de $m \times n$ donde $m \leq n$, $x \in R^n$, $b \in R^m$.

Si x^* es solución del problema, entonces existe un vector λ tal que se satisfacen las condiciones de KKT. Calculando el Lagrangiano e igualando a cero

$$\begin{aligned}
Gx + d - A\lambda &= 0, \\
Ax - b &\geq 0, \\
\lambda, x &\geq 0
\end{aligned}$$

Introduciendo variables de holgura $y = Ax - b$.

$$\begin{aligned}
Gx - A^T\lambda + d &= 0, \\
Ax - y - b &= 0, \\
y_i\lambda_i &= 0, \\
(\lambda, y) &\geq 0
\end{aligned}$$

El problema obtenido equivale a resolver:

$$F(x, y, \lambda) = \begin{cases} Gx - A^T\lambda + d \\ Ax - y - b \\ Y\Lambda e \end{cases} = 0$$

sujeto a $(y, \lambda) \geq 0$

$$Y = \text{diag}(y_1, y_2, \dots, y_m)$$

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$$

Usando el método del camino central, el sistema KKT queda:

$$\begin{bmatrix} G & -A^T & 0 \\ A & -I & 0 \\ 0 & \Lambda & Y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -rd \\ -rb \\ -Y\Lambda e + \sigma\mu e \end{bmatrix}$$

donde

$$\begin{aligned}
rd &= Gx - A^T\lambda + d, \\
rb &= Ax - y - b, \\
(y, \lambda) &\geq 0, \\
\mu &= \frac{1}{m} \sum_{i=1}^m y_i\lambda_i, \\
\sigma &\in [0, 1]
\end{aligned}$$