

Optimización Numérica

Mario Becerra 124362

Proyecto 2: Programación Cuadrática Sucesiva
Noviembre de 2015

Introducción

En este proyecto se construye un método numérico iterativo para encontrar una solución al problema

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeta a} && c(x) = 0, \end{aligned} \tag{1}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ son funciones doblemente diferenciables y continuas. Este problema puede considerarse un clásico de la programación no lineal, y se pretende que, con base en la teoría de optimización con restricciones, se derive un método para resolver el problema.

En la primera sección, se escribe la teoría de optimización con restricciones y se enuncian distintas definiciones y teoremas necesarios para entender el método.

Después se muestra la teoría de programación cuadrática sucesiva (PCS) para llevar a cabo la construcción del método. Posteriormente se presenta la noción de función de mérito y la relación que esta tiene con la convergencia global. En esa misma sección se discuten los resultados de los primeros experimentos del método.

Finalmente, se enuncia el concepto de regularización, el cual se usa para que el método sea más robusto y eficiente, y también se discuten los resultados del método con la regularización.

Teoría de optimización con restricciones

Para poder entender el método que se está construyendo, primero se debe de empezar con la teoría de optimización con restricciones. A continuación se enunciarán varias definiciones y teoremas para los cuales las pruebas pueden ser consultadas en [1].

Definición 1 (Conjunto factible). *Para el problema (1), el conjunto factible Ω es $\Omega = \{x \in \mathbb{R}^n \mid c(x) = 0\}$.*

Definición 2 (Conjunto activo en x). *El conjunto activo $\mathcal{A}(x)$ para cualquier x factible consiste en los índices de la función $c(x) = [c_1(x), \dots, c_m(x)]^T$. Esto es, $\mathcal{A}(x) = \{i \mid c_i(x) = 0\}$. O sea, $\mathcal{A}(x) = \{1, \dots, m\}$.*

Definición 3 (Conjunto de direcciones factibles linealizadas). *Dado un punto factible x y el conjunto de restricciones activas $\mathcal{A}(x)$, el conjunto de direcciones factibles linealizadas $\mathcal{F}(x)$ es*

$$\mathcal{F}(x) = \{d \mid d^t \nabla c_i(x) = 0 \text{ para todo } i = 1, \dots, m\}.$$

Definición 4 (Condición de regularidad de independencia lineal). *Dado el punto x y el conjunto activo $\mathcal{A}(x)$, se dice que se cumple la condición de regularidad de independencia lineal si el conjunto de gradientes de restricciones activas $\{\nabla c_i(x) \mid i \in \mathcal{A}(x)\}$ es linealmente independiente.*

Las condiciones mencionadas anteriormente son condiciones bajo las cuales $\mathcal{F}(x)$ representa las características de Ω en una vecindad de x .

Teorema 1 (Condiciones necesarias de primer orden o condiciones de Karush-Kuhn-Tucker). *Suponer que x^* es un minimizador local de (1), y que se cumple la condición de regularidad de independencia lineal en x^* . Entonces existe un vector $\lambda^* \in \mathbb{R}^m$, llamado multiplicador de Lagrange, con componentes λ_i^* tales que se cumplen las siguientes condiciones en (x^*, λ^*) :*

$$\blacksquare \nabla_x \mathcal{L}(x^*, \lambda^*) = 0$$

$$\blacksquare \quad c(x^*) = 0,$$

donde $\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i c_i(x)$ es llamada la función Lagrangiana.

Demostración. Disponible en [1, p. 321]. Teorema 12.1. \square

Las condiciones de KKT dicen cómo se relacionan las primeras derivadas de f con las restricciones en la solución x^* . Cuando se cumplen las condiciones, moverse a lo largo de cualquier vector w en $\mathcal{F}(x^*)$ o aumenta la aproximación de primer orden a la función objetivo ($w^T \nabla f(x^*) > 0$) o mantiene el valor como está ($w^T \nabla f(x^*) = 0$).

Con esta información de primeras derivadas, en direcciones $w \in \mathcal{F}(x^*)$ en las que $w^T \nabla f(x^*) = 0$ no se puede determinar si al moverse en la dirección w aumenta o disminuye el valor de la función objetivo f .

Definición 5 (Cono crítico). *Dado el conjunto $\mathcal{F}(x^*)$ y un vector de multiplicadores de Lagrange λ^* que satisface las condiciones de KKT, se define el cono crítico como*

$$\mathcal{C}(x^*, \lambda^*) = \{w \in \mathcal{F}(x^*) \mid \nabla c_i(x^*)^T w = 0 \text{ para todo } i \in \mathcal{A}(x^*), \lambda_i^* > 0\}.$$

En el caso en que el vector de multiplicadores de Lagrange es único, o sea, en el caso en que se cumple la condición de regularidad de independencia lineal, entonces el cono crítico se puede definir como

$$\mathcal{C}(x^*, \lambda^*) = \text{Nulo}(\nabla c_i(x^*)^T)_{i \in \mathcal{A}(x^*)} = \text{Nulo}(A(x^*))$$

donde $A(x^*)$ es una matriz en la que cada fila es $\nabla c_i(x^*)^T$. Entonces, si se define la matriz de rango completo Z cuyas columnas forman una base del espacio nulo de $A(x^*)$, entonces

$$\mathcal{C}(x^*, \lambda^*) = \{Zu \mid u \in \mathbb{R}^m\}$$

Teorema 2 (Condiciones suficientes de segundo orden). *Suponer que para algún punto factible $x^* \in \mathbb{R}^n$ existe un vector de multiplicadores de Lagrange λ^* tal que se cumplen las condiciones de KKT. Suponer también que se cumple la condición de regularidad de independencia lineal en x^* , que Z es una matriz de rango completo cuyas columnas forman una base del espacio nulo de $A(x^*)$ con $A^T(x^*) = [\nabla c_1(x^*), \dots, \nabla c_m(x^*)]$, y que $u^T W u > 0$ para todo $u \in \mathbb{R}^m$, i.e., W es positiva definida, con $W = Z^T \mathcal{L}(x^*, \lambda^*) Z$. Entonces x^* es solución local estricta de (1).*

Demostración. Disponible en [1, p. 333]. Teorema 12.6. \square

Construcción del método

Programación cuadrática sucesiva (PCS) local

El método que se va a construir genera pasos al resolver subproblemas cuadráticos. La idea de la PCS local es modelar (1) en cada iteración como un subproblema cuadrático fácil de resolver, y usar el minimizador del subproblema para definir una nueva iteración. El reto es diseñar el subproblema cuadrático de tal forma que dé un buen paso para el problema de optimización (1). En este proyecto, se presenta un método que aplica el método de Newton a las condiciones de KKT definidas anteriormente.

Las condiciones de KKT pueden ser reescritas como un sistema de $n + m$ ecuaciones con $n + m$ incógnitas de la siguiente forma:

$$F(x, \lambda) = \begin{bmatrix} \nabla f(x) - A(x)^T \lambda \\ c(x) \end{bmatrix} = 0. \quad (2)$$

Cualquier solución (x^*, λ^*) de (1) para la cual $A(x^*)$ es de rango completo satisface (2). El problema no lineal (2) se puede resolver usando el método de Newton.

La Jacobiana de (2) respecto a x y λ es:

$$F'(x, \lambda) = \begin{bmatrix} \nabla_x^2 \mathcal{L}(x, \lambda) & -A(x)^T \\ A(x) & 0 \end{bmatrix}.$$

Entonces el paso de Newton de la iteración (x_k, λ_k) es

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} + \begin{bmatrix} p_k \\ p_\lambda \end{bmatrix} \quad (3)$$

donde p_k y p_λ resuelven el sistema

$$F'(x_k, \lambda_k) \begin{bmatrix} p_k \\ p_\lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) + A(x_k)^T \lambda_k \\ -c(x_k) \end{bmatrix}.$$

Esta iteración de Newton está definida cuando $F'(x_k, \lambda_k)$ es no singular, y esto se cumple cuando se cumple lo siguiente:

- (a) La Jacobiana de restricciones $A(x)$ es de rango completo.
- (b) La matriz $W = \nabla_{xx}^2 \mathcal{L}(x, \lambda)$ cumple que $d^T W d > 0$ para todo $d \neq 0$ tal que $A(x)d = 0$.

Una forma distinta de ver la iteración (3) es suponiendo que el problema (1) se modela mediante el problema cuadrático

$$\begin{aligned} \underset{p}{\text{minimizar}} \quad & f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{sujeta a} \quad & A_k p + c_k = 0. \end{aligned} \tag{4}$$

Si los supuestos (a) y (b) se cumplen, entonces este problema tiene solución única (p_k, l_k) que satisface

$$\begin{aligned} \nabla_{xx}^2 \mathcal{L}_k p_k + \nabla f_k - A_k^T p_k &= 0 \\ A_k p_k + c_k &= 0. \end{aligned}$$

El vector (p_k, l_k) se puede obtener de la solución (p_k, λ_k) de (3).

Propiedades de los problemas cuadráticos

El problema cuadrático en forma general se puede expresar como

$$\begin{aligned} \underset{x}{\text{minimizar}} \quad & q(x) = \frac{1}{2} x^T G x + x^T c \\ \text{sujeta a} \quad & A x = b. \end{aligned} \tag{5}$$

Las condiciones de KKT para el punto x^* del problema (5) establecen que debe existir un vector λ^* tal que

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}. \tag{6}$$

Definiendo $x^* = x + p$ se puede reescribir el sistema como

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} c + Gx \\ Ax - b \end{bmatrix}. \tag{7}$$

La matriz $K = \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}$ es llamada la matriz de KKT. A continuación se enuncia un resultado que da las condiciones para las cuales es no-singular.

Teorema 3. Considere el problema (5), y defina a Z como la matriz de $n \times (n-m)$ cuyas columnas son base del espacio nulo de A (i.e., Z de rango completo tal que $AZ = 0$). Si A es de rango completo y la Hessiana reducida, definida como $Z^T G Z$ es positiva definida, entonces la matriz de KKT, K , es no-singular, por lo que existe un único vector λ^* que satisface (6), y el vector x^* que satisface (6) es la única solución global de (5).

Demostración. Disponible en [1, p. 452]. Lema 16.1. □

Teorema 4. Considere el problema (5), la matriz de KKT, K , con A de rango m y la matriz Z de rango completo tal que $AZ = 0$. Entonces

$$\text{inercia}(K) = \text{inercia}(Z^T G Z) + (m, m, 0),$$

donde $\text{inercia}(B) = (n_+, n_-, n_0)$ es la terna que contiene el número de eigenvalores positivos, negativos y cero de la matriz B . Por lo tanto, si $Z^T G Z$ es simétrica positiva definida, entonces

$$\text{inercia}(K) = (n, m, 0).$$

Corolario 1. *La matriz de KKT es indefinida (pues $m \geq 1$).*

Como K es indefinida, no se puede utilizar Cholesky para resolver el sistema de KKT (7), pero se puede utilizar la factorización LDL de tal forma que $P^T K P = LDL^T$, donde P es una matriz de permutaciones, L es una matriz triangular inferior y D es una matriz diagonal por bloques. Se puede utilizar además una matriz de escala S para problemas muy grandes y malos, de tal forma que con S se mantiene la validez del sistema. Así, se tiene el nuevo sistema $P^T S K S P = LDL^T$. Para resolver el sistema (7) se debe de hacer lo siguiente:

1. Resolver $Lu = P^T S \begin{bmatrix} c + Gx \\ Ax - b \end{bmatrix}$.
2. Resolver $Du' = u$.
3. Resolver $L^T u'' = u'$.
4. Resolver $SPu''' = u''$.

Así u''' es la solución al sistema de KKT (7), p^* son las primeras n entradas de u''' y λ^* las últimas m .

Convergencia global

Una forma de lograr solucionar el problema de optimización (1) es reemplazar el problema original por una sucesión de subproblemas en los cuales las restricciones están representadas por términos añadidos a la función objetivo. En este proyecto se usa esto para lograr una convergencia global del método utilizando una función de pérdida ℓ_1 , la cual tiene la deseable propiedad de que es *exacta*, o sea, que minimizando la función de mérito se puede llegar a la solución de un problema de programación no lineal.

Definición 6 (Función de mérito ℓ_1). *La función de mérito ℓ_1 para el problema de optimización (1) es*

$$\phi_1(x; \mu) = f(x) + \mu \|c(x)\|_1.$$

La idea de utilizar una función de mérito es que al usar el paso de Newton p , puede llegar a pasar que el valor de la función objetivo se reduzca, pero que se viole la restricción; con la función de mérito se puede lograr un balance entre avance y factibilidad mediante el parámetro de penalización μ . En otras palabras, se puede descender al mismo tiempo que se mantiene la factibilidad del problema.

Definición 7 (Exactitud de una función de mérito). *Una función de mérito se dice que es exacta si existe un escalar μ^* tal que para todo $\mu > \mu^*$ se cumple que cualquier solución local del problema de programación no lineal (1) es un minimizador local de $\phi_1(x; \mu)$.*

Un problema puede conocer el valor del parámetro μ^* el cual determina el conjunto de μ 's posible para las cuales la solución del problema nos lleva a tener un minimizador local de la función de mérito. Pero el siguiente teorema establece el valor exacto de μ^* .

Teorema 5. *Sea x^* minimizador local fuerte del problema de optimización (1) en el que se cumplen las condiciones de KKT, con λ^* los multiplicadores de Lagrange correspondientes. Entonces x^* es minimizador local de $\phi_1(x; \mu)$ para todo $\mu > \mu^*$, donde $\mu^* = \|\lambda^*\|_\infty$.*

Si además se cumplen las condiciones suficientes de segundo orden y $\mu > \mu^$, entonces x^* es un minimizador local estricto de $\phi_1(x; \mu)$.*

A grandes rasgos, este teorema dice que estando en una solución, x^* , del problema de optimización (1), moverse hacia la región no factible es penalizado suficientemente para producir un aumento en la función de mérito en un valor más grande que $\phi_1(x^*; \mu) = f(x^*)$, forzando así al minimizador local de $\phi_1(\cdot; \mu)$.

Otro resultado que ayuda a definir el método que se construye en el presente trabajo es el siguiente, en el que se toma en cuenta el modelo cuadrático del problema de optimización (1).

A partir de la desigualdad

$$q_\mu(0) - q_\mu(p_k) \geq \mu_k \rho(m(0) - m(p_k))$$

con

$$q_\mu(p_k) = f(x_k) + \nabla f(x_k)^T p_k + \frac{\sigma}{2} p_k^T W_k p_k + \mu \|c + A_k p_k\|_1$$

$$m(p_k) = \|c + A_k p_k\|_1,$$

se llega a que

$$\mu_k \geq \frac{\nabla f(x_k)^T p_k + \frac{\sigma}{2} p_k^T W_k p_k}{(1 - \rho) \|c(x_k)\|_1}, \quad (8)$$

para algún $\rho \in (0, 1)$, $\sigma > 0$.

Así que el método aquí presentado hará PCS utilizando la función de mérito. En cada iteración se plantea el subproblema cuadrático y se hace un recorte del paso de Newton con la condición de Wolfe de descenso suficiente y con una μ_k tal que se cumple (8), de esta forma la dirección de descenso es suficientemente negativa. Como la función de mérito con norma ℓ_1 no es diferenciable en todo el dominio, para este caso se usa la derivada direccional, definida como

$$D(\phi_1(x; \mu), p) = \nabla f(x)^T p - \mu \|c(x)\|_1.$$

Experimentos con y sin función de mérito

El objetivo de esta sección es evaluar el desempeño del método de programación cuadrática sucesiva con función de mérito, se compara el resultado con un método de programación cuadrática sucesiva sin función de mérito. El conjunto de prueba consta de 59 problemas. En el cuadro (1) se pueden ver los resultados de los problemas que se probaron usando la función de mérito, y en el cuadro (2) los resultados sin función de mérito. Se muestra el nombre del problema, n y m las dimensiones del espacio de la función objetivo y restricciones como se define en el problema (1); $f(\hat{x}^*)$, el valor de la función objetivo en la aproximación a la solución; $\|c(x^*)\|_2$, la norma del vector de restricciones en la aproximación a la solución; $\|\nabla f(x^*)\|_2$, la norma del gradiente de la lagrangiana en la aproximación a la solución; **iter**, el número de iteraciones que hizo el método; **feval** el número de veces que se evaluó la función objetivo; el tiempo en segundos que tardó el método en llegar a \hat{x}^* , la aproximación de x^* ; la inercia del problema, donde incorrecta se refiere a que $n_- \neq m$ o $\text{rango}(K) \neq m + n$, con n_- el número de eigenvalores negativos de la matriz de KKT K .

Se puede ver que sin la función de mérito, los problemas **dtoc2**, **orthrdm2**, **orthregc**, **orthregd** y **orthrgdm** no pudieron ser resueltos por tener inercia incorrecta, mientras que con la función de mérito sí se pudo. Otra cosa que llama la atención es que en el problema **hs047** sucede lo contrario: sin la función de mérito sí se puede resolver mientras que con ella no se puede.

Perfiles de rendimiento

Una forma sencilla y visual de comparar dos métodos es usar perfiles de rendimiento. En este caso se usan para evaluar cada método tomando en cuenta cuánto tiempo tarda y cuántas veces evalúa la función objetivo en cada problema. Se usan solo los 25 problemas en los cuales los dos métodos tienen inercia correcta. Se define el perfil de tiempo y de número de evaluaciones del problema i como

$$p_t^i = \log\left(\frac{t_{local}^i}{t_{global}^i}\right) \quad (9)$$

$$p_n^i = \log\left(\frac{n_{local}^i}{n_{global}^i}\right) \quad (10)$$

con $i = 1, \dots, 25$ el índice que corresponde a cada problema, t_{global}^i el tiempo que tardó el método global (con función de mérito) en el problema i , n_{global}^i el número de veces que se evaluó la función objetivo con el método global en el problema i , t_{local}^i y n_{local}^i definidos de forma equivalente.

De esta forma, p_t^i será positivo cuando $t_{global}^i < t_{local}^i$, por lo que si es positivo, es a favor del método global; además, si el tiempo del método global es mucho más chico que el del método local, entonces p_t^i será grande, o sea que p_t^i será mayor de forma positiva cuanto mejor sea el método global que el método local para el problema i . Por el contrario, p_t^i será mayor de forma negativa cuanto peor sea el método global que el método local. Este análisis es totalmente equivalente para el número de iteraciones.

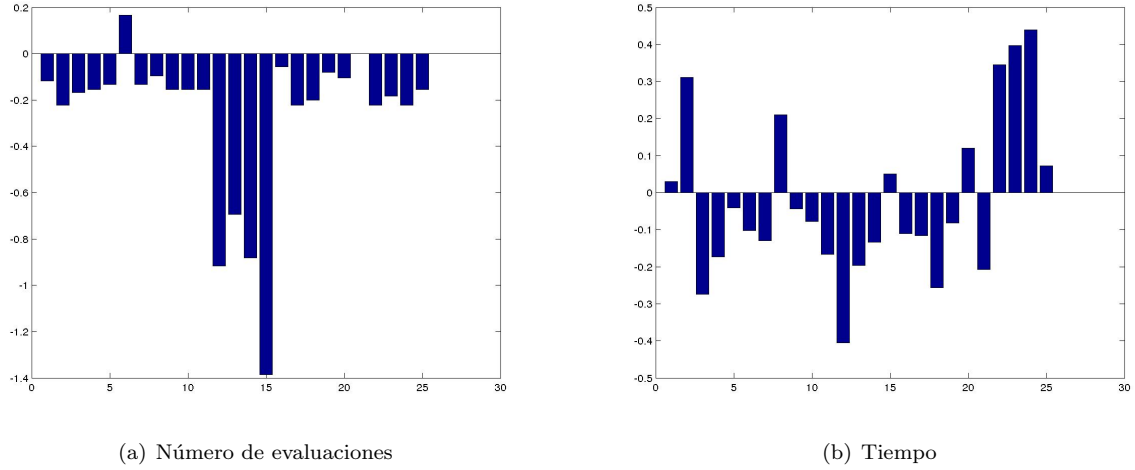


Figura 1: Perfil de rendimiento de tiempo y número de evaluaciones de la función objetivo del método con y sin función de mérito

En la figura (1) se pueden ver los perfiles de rendimiento para el tiempo y para el número de evaluaciones. Se puede ver que en número de evaluaciones, el método local es mejor en todos los problemas menos en uno, más aún, la barra de este problema es pequeña en comparación a las otras, por lo que no significa que sea mucho mejor para ese problema. El rendimiento en cuanto el tiempo también es mejor en el método local para la mayoría de los problemas.

Comparaciones

Un punto muy importante que resalta de los resultados es la inercia: de los 59 problemas, solo 30 tuvieron la inercia correcta. Esto se debe a que el método llega a un punto localmente no convexo, por lo que la inercia en ese punto es incorrecta y el método termina. Este no es un problema que se pueda ignorar fácilmente, pues muchas funciones tienen puntos localmente no convexos; si hacemos el ingenuo supuesto de que nuestro conjunto de prueba es representativo de todos los problemas de optimización con restricciones, entonces cerca del 50% de los problemas tienen inercia incorrecta. Esta aproximación no es correcta, pero da a entender el punto de que el problema de la inercia se debe de corregir de alguna forma. Más adelante se presenta un método heurístico de regularización que permite que se puedan resolver muchos de los problemas con inercia incorrecta.

Es necesario hacer notar que al comparar los perfiles de rendimiento, aunque el método local puede parecer mejor que el global en cuanto al tiempo de ejecución y número de evaluaciones de la función objetivo; el hecho de agregar la función de mérito ayuda a resolver problemas que con el método local no se podía. Este es un *trade-off* que se debe tomar en cuenta al usar cualquiera de los dos métodos.

Regularización

Se vio en los resultados anteriores que muchos de los problemas llegaban a un punto con inercia incorrecta. Esto se puede solucionar con un procedimiento de regularización de la matriz de KKT sumándole elementos a la diagonal y ayudar a la inercia. Así, al algoritmo se le agrega un paso en caso de que la inercia resulte incorrecta y la matriz de KKT K es:

$$K = \begin{bmatrix} \nabla_x^2 \mathcal{L}(x_k, \lambda) + \gamma_k I & -A(x_k)^T \\ A(x_k) & -\delta I \end{bmatrix},$$

donde γ_k cambia en cada iteración del método y $\delta = 10^{-2}$. Para escoger γ_k se prueban distintos valores hasta que la matriz K resulte con inercia correcta o hasta que se haga un número de máximo de intentos, en cuyo caso se concluye que la inercia no es correcta. El parámetro γ_k no debe ser muy grande porque en ese caso, este tendría más peso en el método y la aproximación lineal del problema no sería buena, es por esto que se le pone un límite al número de intentos que se hace para escoger el parámetro. En el caso de

este proyecto, se hace un máximo de 10 intentos, desde un primer valor de $\gamma_k = 10^{-3}$ y multiplicando por 10 cada intento. De esta forma, se tiene que en cada iteración se hace lo siguiente:

- Si la inercia es correcta, seguir con el método. Si no lo es, tomar $\gamma_k = 10^{-3}$ y $\delta = 10^{-2}$. Mientras la inercia sea incorrecta o el número de intentos sea menor a 10:
 - Factorizar con LDL la matriz $K = \begin{bmatrix} \nabla_x^2 \mathcal{L}(x_k, \lambda) + \gamma_k I & -A(x_k)^T \\ A(x_k) & -\delta I \end{bmatrix}$.
 - Revisar inercia
 - Número de intentos = Número de intentos + 1

Experimentos con función de mérito y con regularización

En esta sección se comparan los métodos de PCS regularizada y sin regularizar. En la tabla (3) se pueden ver los resultados de cada uno de los problemas. Con la regularización, todos los problemas tienen inercia correcta, sin embargo, se puede ver que los problemas **bt8**, **eigenb2**, **eigenbco**, **eigenc2**, **hs007**, **lch**, **orthrds2** y **orthrega** el método hace 401 iteraciones, el máximo permitido por el algoritmo. Además, en estos problemas la norma del vector c y/o del gradiente de la Lagrangiana, $\|\nabla f(x)\|$ está lejos de ser cero. Esto significa que el método no está llegando a una solución del problema. Si se analizan los resultados detallados de estos problemas (no incluidos), se puede notar que la función objetivo y las normas descienden, sin embargo descienden muy lentamente; esto puede deberse a que el problema sea muy plano en el punto inicial y en cada iteración el avance sea muy poco.

Aún así, se puede ver que la regularización mejoró notablemente el método, pues se pasó de resolver cerca de un 50 % de los problemas a resolver un 86 %.

Perfiles de rendimiento

En este caso también se usan perfiles de rendimiento para evaluar los métodos tomando en cuenta cuánto tiempo tarda y cuántas veces evalúa la función objetivo en cada problema. Ahora se utilizan los 30 problemas en los cuales el método con regularización y sin regularización tienen inercia correcta. Así, los perfiles de tiempo y de número de evaluaciones del problema i son

$$p_t^i = \log\left(\frac{t_{global}^i}{t_{globalReg}^i}\right) \quad (11)$$

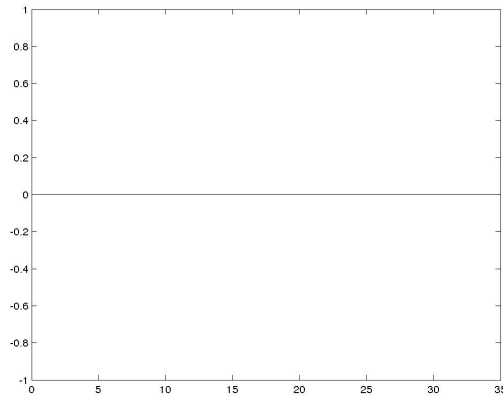
$$p_n^i = \log\left(\frac{n_{global}^i}{n_{globalReg}^i}\right) \quad (12)$$

con $i = 1, \dots, 30$ el índice que corresponde a cada problema, t_{global}^i el tiempo que tardó el método global sin regularización en el problema i , n_{global}^i el número de veces que se evaluó la función objetivo con el método global sin regularización en el problema i , $t_{globalReg}^i$ y $n_{globalReg}^i$ definidos de forma equivalente solo que con el método regularizado.

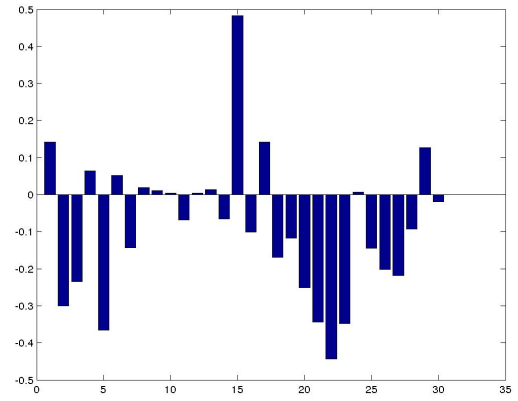
En la figura () se pueden ver los perfiles de rendimiento para el tiempo y para el número de evaluaciones de los métodos regularizado y no regularizado. Se puede ver que en número de evaluaciones, ambos métodos toman el mismo número de evaluaciones en todos los problemas, por lo que $p_n^i = 0$ para toda $i = 1, \dots, 30$. El rendimiento en cuanto el tiempo es mejor en el método sin regularización para 11 de los 30 problemas, sin embargo, en estos casos, la diferencia no fue mucho mayor en ambos métodos, por lo que se puede decir que el agregar la regularización no afecta notablemente el tiempo de ejecución, pero mejora el número de problemas resueltos.

Conclusiones

Viendo los resultados de los cuadros (1), (2) y (3) se puede concluir que para estos problemas, el método hace pocas iteraciones para llegar a la aproximación de la solución; y también que el tiempo en que las hace es relativamente poco, aunque en los problemas con inercia incorrecta el método puede hacer muchas iteraciones y evaluar muchas veces la función objetivo, lo cual hace que en estos problemas se tarde mucho.



(a) Número de evaluaciones



(b) Tiempo

Figura 2: Perfil de rendimiento de tiempo y número de evaluaciones de la función objetivo del método con regularización y sin regularización

Se debe hacer notar que los problemas con los que se probó el método son de escala chica y mediana, máximo en el orden de los miles. Este método en escala grande podría ya no ser eficiente en tiempo pues en cada iteración hace operaciones que pueden ser caras en escala grande.

Referencias

- [1] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization, 2nd edition*. World Scientific, 2006.

i	problema	n	m	$f(\hat{x}^*)$	$\ c(x^*)\ _2$	$\ \nabla f(x^*)\ _2$	iter	feval	tiempo (segundos)	inercia
1	bt11	5	3	8.25e-01	1.11e-16	2.39e-15	8	9	3.53e-02	correcta
2	bt12	5	3	6.19e+00	3.43e-13	9.10e-14	4	5	2.91e-02	correcta
3	bt1	2	1	-1.00e+00	0.00e+00	2.10e-17	11	13	3.24e-02	correcta
4	bt2	3	1	3.26e-02	1.78e-15	1.24e-16	12	15	4.15e-02	correcta
5	bt4	3	2	-	-	-	-	-	-	incorrecta
6	bt5	3	2	9.62e+02	3.55e-15	2.66e-15	7	8	3.70e-02	correcta
7	bt6	5	2	2.77e-01	4.44e-16	2.71e-16	10	12	3.53e-02	correcta
8	bt7	5	3	-	-	-	-	-	-	incorrecta
9	bt8	5	2	-	-	-	-	-	-	incorrecta
10	bt9	4	2	-	-	-	-	-	-	incorrecta
11	byrdsphr	3	2	-	-	-	-	-	-	incorrecta
12	catena	32	11	-2.31e+04	1.62e-15	5.12e-12	7	8	3.47e-02	correcta
13	catenary	496	166	-	-	-	-	-	-	incorrecta
14	dixchlng	10	5	2.47e+03	3.85e-16	1.22e-12	10	12	3.47e-02	correcta
15	dtoc1l	14985	9990	1.25e+02	1.59e-14	6.39e-11	7	11	2.28e+00	correcta
16	dtoc1na	1485	990	1.27e+01	1.65e-15	4.32e-15	7	8	4.62e-01	correcta
17	dtoc1nb	1485	990	1.59e+01	3.37e-15	6.07e-15	6	7	4.01e-01	correcta
18	dtoc1nc	1485	990	-	-	-	-	-	-	incorrecta
19	dtoc1nd	735	490	-	-	-	-	-	-	incorrecta
20	dtoc2	5994	3996	5.09e-01	1.34e-13	3.83e-14	8	15	7.78e-01	correcta
21	dtoc4	14996	9997	2.87e+00	2.89e-13	3.24e-13	6	10	2.17e+00	correcta
22	dtoc5	9998	4999	1.54e+00	7.07e-10	9.16e-10	5	8	5.17e-01	correcta
23	dtoc6	10000	5000	1.35e+05	2.74e-11	6.86e-08	17	29	1.87e+00	correcta
24	eigena2	110	55	-	-	-	-	-	-	incorrecta
25	eigenaco	110	55	-	-	-	-	-	-	incorrecta
26	eigenb2	110	55	-	-	-	-	-	-	incorrecta
27	eigenbco	110	55	-	-	-	-	-	-	incorrecta
28	eigenc2	462	231	-	-	-	-	-	-	incorrecta
29	eigencco	30	15	-	-	-	-	-	-	incorrecta
30	gilbert	10	1	-	-	-	-	-	-	incorrecta
31	hs006	2	1	0.00e+00	0.00e+00	0.00e+00	6	16	3.13e-02	correcta
32	hs007	2	1	-	-	-	-	-	-	incorrecta
33	hs009	2	1	-	-	-	-	-	-	incorrecta
34	hs026	3	1	4.29e-13	3.69e-07	3.00e-09	20	21	3.98e-02	correcta
35	hs027	3	1	-	-	-	-	-	-	incorrecta
36	hs039	4	2	-	-	-	-	-	-	incorrecta
37	hs040	4	3	-2.50e-01	2.48e-16	2.94e-16	4	5	2.91e-02	correcta
38	hs046	5	2	2.37e-12	3.85e-07	7.64e-09	20	34	5.62e-02	correcta
39	hs047	5	3	-	-	-	-	-	-	incorrecta
40	hs049	5	2	3.52e-08	0.00e+00	1.03e-05	14	15	2.80e-02	correcta
41	hs050	5	3	6.16e-32	8.88e-16	2.53e-16	9	10	2.92e-02	correcta
42	hs061	3	2	-	-	-	-	-	-	incorrecta
43	hs077	5	2	2.42e-01	1.05e-13	1.31e-12	9	11	2.70e-02	correcta
44	hs078	5	3	-2.92e+00	1.99e-15	2.06e-15	5	6	3.14e-02	correcta
45	hs079	5	3	7.88e-02	2.22e-16	2.61e-16	5	6	2.77e-02	correcta
46	hs100lnp	7	2	-	-	-	-	-	-	incorrecta
47	hs111lnp	10	3	-	-	-	-	-	-	incorrecta
48	lch	600	1	-	-	-	-	-	-	incorrecta
49	maratos2	2	1	-1.00e+00	0.00e+00	1.16e-16	12	15	2.69e-02	correcta
50	maratos	2	1	-1.00e+00	5.33e-13	4.95e-13	12	14	3.21e-02	correcta
51	mwright	5	3	-	-	-	-	-	-	incorrecta
52	orthrdm2	4003	2000	1.56e+02	2.20e-12	6.10e-13	7	10	9.84e-01	correcta
53	orthrds2	203	100	-	-	-	-	-	-	incorrecta
54	orthrega	517	256	-	-	-	-	-	-	incorrecta
55	orthregb	27	6	-	-	-	-	-	-	incorrecta
56	orthregc	10005	5000	1.90e+02	1.83e-14	1.93e-12	13	19	9.18e+00	correcta
57	orthregd	10003	5000	1.52e+03	3.66e-12	4.17e-13	7	10	6.20e+00	correcta
58	orthrgdm	10003	5000	1.51e+03	3.89e-12	7.94e-13	8	11	6.04e+00	correcta
59	orthrgds	10003	5000	-	-	-	-	-	-	incorrecta

Cuadro 1: Resultados de los problemas con función de mérito

i	problema	n	m	$f(\hat{x}^*)$	$\ c(x^*)\ _2$	$\ \nabla f(x^*)\ _2$	iter	feval	tiempo (segundos)	inercia
1	bt11	5	3	8.25e-01	1.11e-16	2.39e-15	8	8	2.73e-02	correcta
2	bt12	5	3	6.19e+00	3.43e-13	9.10e-14	4	4	2.87e-02	correcta
3	bt1	2	1	-1.00e+00	0.00e+00	5.68e-14	11	11	2.78e-02	correcta
4	bt2	3	1	3.26e-02	2.18e-12	5.67e-12	12	12	3.23e-02	correcta
5	bt4	3	2	-	-	-	-	-	-	incorrecta
6	bt5	3	2	9.62e+02	3.55e-15	2.66e-15	7	7	3.02e-02	correcta
7	bt6	5	2	2.77e-01	5.80e-14	1.13e-12	13	13	2.94e-02	correcta
8	bt7	5	3	-	-	-	-	-	-	incorrecta
9	bt8	5	2	-	-	-	-	-	-	incorrecta
10	bt9	4	2	-	-	-	-	-	-	incorrecta
11	byrdsphr	3	2	-	-	-	-	-	-	incorrecta
12	catena	32	11	-2.31e+04	9.29e-16	3.13e-12	7	7	2.61e-02	correcta
13	catenary	496	166	-	-	-	-	-	-	incorrecta
14	dixchlng	10	5	2.47e+03	6.28e-16	2.06e-12	10	10	3.09e-02	correcta
15	dtoc1l	14985	9990	1.25e+02	2.06e-14	4.64e-15	7	7	2.02e+00	correcta
16	dtoc1na	1485	990	1.27e+01	1.66e-15	4.18e-15	7	7	4.03e-01	correcta
17	dtoc1nb	1485	990	1.59e+01	3.25e-15	6.07e-15	6	6	3.68e-01	correcta
18	dtoc1nc	1485	990	-	-	-	-	-	-	incorrecta
19	dtoc1nd	735	490	-	-	-	-	-	-	incorrecta
20	dtoc2	5994	3996	-	-	-	-	-	-	incorrecta
21	dtoc4	14996	9997	2.87e+00	2.55e-13	3.47e-14	4	4	1.47e+00	correcta
22	dtoc5	9998	4999	1.54e+00	4.93e-12	3.42e-12	4	4	3.12e-01	correcta
23	dtoc6	10000	5000	1.35e+05	3.65e-11	6.25e-11	12	12	1.17e+00	correcta
24	eigena2	110	55	-	-	-	-	-	-	incorrecta
25	eigenaco	110	55	-	-	-	-	-	-	incorrecta
26	eigenb2	110	55	-	-	-	-	-	-	incorrecta
27	eigenbco	110	55	-	-	-	-	-	-	incorrecta
28	eigenc2	462	231	-	-	-	-	-	-	incorrecta
29	eigencco	30	15	-	-	-	-	-	-	incorrecta
30	gilbert	10	1	-	-	-	-	-	-	incorrecta
31	hs006	2	1	0.00e+00	0.00e+00	0.00e+00	4	4	3.04e-02	correcta
32	hs007	2	1	-	-	-	-	-	-	incorrecta
33	hs009	2	1	-	-	-	-	-	-	incorrecta
34	hs026	3	1	4.29e-13	3.69e-07	3.00e-09	20	20	3.59e-02	correcta
35	hs027	3	1	-	-	-	-	-	-	incorrecta
36	hs039	4	2	-	-	-	-	-	-	incorrecta
37	hs040	4	3	-2.50e-01	2.48e-16	2.94e-16	4	4	3.18e-02	correcta
38	hs046	5	2	2.84e-12	4.21e-07	8.75e-09	14	14	2.62e-02	correcta
39	hs047	5	3	1.07e-10	2.37e-07	9.47e-07	16	16	2.39e-02	correcta
40	hs049	5	2	3.52e-08	0.00e+00	1.03e-05	14	14	2.41e-02	correcta
41	hs050	5	3	6.16e-32	8.88e-16	2.53e-16	9	9	2.87e-02	correcta
42	hs061	3	2	-	-	-	-	-	-	incorrecta
43	hs077	5	2	2.42e-01	1.83e-15	3.59e-16	12	12	2.75e-02	correcta
44	hs078	5	3	-2.92e+00	1.99e-15	2.06e-15	5	5	2.30e-02	correcta
45	hs079	5	3	7.88e-02	2.22e-16	2.61e-16	5	5	2.70e-02	correcta
46	hs100lnp	7	2	-	-	-	-	-	-	incorrecta
47	hs111lnp	10	3	-	-	-	-	-	-	incorrecta
48	lch	600	1	-	-	-	-	-	-	incorrecta
49	maratos2	2	1	-1.00e+00	1.08e-12	9.91e-13	12	12	2.82e-02	correcta
50	maratos	2	1	-1.00e+00	1.08e-12	9.90e-13	12	12	2.24e-02	correcta
51	mwright	5	3	-	-	-	-	-	-	incorrecta
52	orthrdm2	4003	2000	-	-	-	-	-	-	incorrecta
53	orthrds2	203	100	-	-	-	-	-	-	incorrecta
54	orthrega	517	256	-	-	-	-	-	-	incorrecta
55	orthregb	27	6	-	-	-	-	-	-	incorrecta
56	orthregc	10005	5000	-	-	-	-	-	-	incorrecta
57	orthregd	10003	5000	-	-	-	-	-	-	incorrecta
58	orthrgdm	10003	5000	-	-	-	-	-	-	incorrecta
59	orthrgds	10003	5000	-	-	-	-	-	-	incorrecta

Cuadro 2: Resultados de los problemas sin función de mérito

i	problema	n	m	$f(\hat{x}^*)$	$\ c(x^*)\ _2$	$\ \nabla f(x^*)\ _2$	iter	feval	tiempo (segundos)	inercia
1	bt11	5	3	8.25e-01	1.11e-16	2.39e-15	8	9	4.70e-02	correcta
2	bt12	5	3	6.19e+00	3.43e-13	9.10e-14	4	5	2.90e-02	correcta
3	bt1	2	1	-1.00e+00	0.00e+00	2.10e-17	11	13	3.03e-02	correcta
4	bt2	3	1	3.26e-02	1.78e-15	1.24e-16	12	15	3.35e-02	correcta
5	bt4	3	2	-4.55e+01	0.00e+00	7.11e-15	12	14	3.94e-02	correcta
6	bt5	3	2	9.62e+02	3.55e-15	2.66e-15	7	8	2.93e-02	correcta
7	bt6	5	2	2.77e-01	4.44e-16	2.71e-16	10	12	3.14e-02	correcta
8	bt7	5	3	3.60e+02	5.46e-13	6.64e-11	14	28	4.73e-02	correcta
9	bt8	5	2	9.90e-01	1.00e-02	5.79e-06	401	8302	4.02e-01	correcta
10	bt9	4	2	-1.00e+00	0.00e+00	2.46e-21	16	29	4.29e-02	correcta
11	byrdsphr	3	2	-4.68e+00	2.51e-15	6.28e-16	25	46	3.53e-02	correcta
12	catena	32	11	-2.31e+04	1.62e-15	5.12e-12	7	8	2.75e-02	correcta
13	catenary	496	166	-3.48e+05	1.18e-13	3.00e-10	40	45	2.33e-01	correcta
14	dixchlng	10	5	2.47e+03	3.85e-16	1.22e-12	10	12	4.52e-02	correcta
15	dtoc1l	14985	9990	1.25e+02	1.59e-14	6.39e-11	7	11	2.14e+00	correcta
16	dtoc1na	1485	990	1.27e+01	1.65e-15	4.32e-15	7	8	3.99e-01	correcta
17	dtoc1nb	1485	990	1.59e+01	3.37e-15	6.07e-15	6	7	4.41e-01	correcta
18	dtoc1nc	1485	990	2.50e+01	5.54e-15	4.57e-15	11	21	8.63e-01	correcta
19	dtoc1nd	735	490	1.26e+01	3.96e-14	1.42e-14	26	76	9.11e-01	correcta
20	dtoc2	5994	3996	5.09e-01	1.34e-13	3.83e-14	8	15	7.49e-01	correcta
21	dtoc4	14996	9997	2.87e+00	2.89e-13	3.24e-13	6	10	2.09e+00	correcta
22	dtoc5	9998	4999	1.54e+00	7.07e-10	9.16e-10	5	8	4.12e-01	correcta
23	dtoc6	10000	5000	1.35e+05	2.74e-11	6.86e-08	17	29	1.56e+00	correcta
24	eigena2	110	55	3.36e-30	3.59e-43	3.67e-14	6	7	4.58e-02	correcta
25	eigenaco	110	55	2.67e-81	5.74e-42	1.57e-17	6	10	4.89e-02	correcta
26	eigenb2	110	55	1.74e+01	6.93e-02	2.62e+00	401	8362	4.35e+00	correcta
27	eigenbco	110	55	1.88e-01	1.82e-02	1.76e-03	401	6944	7.61e+00	correcta
28	eigenc2	462	231	1.01e+04	2.94e-04	4.21e+03	401	8422	1.08e+02	correcta
29	eigencco	30	15	3.01e-27	2.21e-14	2.80e-13	16	25	8.42e-02	correcta
30	gilbert	10	1	3.35e+00	5.55e-17	2.14e-16	14	17	5.17e-02	correcta
31	hs006	2	1	0.00e+00	0.00e+00	0.00e+00	6	16	4.14e-02	correcta
32	hs007	2	1	-1.92e+01	3.71e+02	4.15e+00	401	8402	3.98e-01	correcta
33	hs009	2	1	-5.00e-01	7.11e-15	3.93e-17	6	12	2.91e-02	correcta
34	hs026	3	1	4.29e-13	3.69e-07	3.00e-09	20	21	3.21e-02	correcta
35	hs027	3	1	4.00e-02	0.00e+00	5.91e-16	325	2438	1.71e-01	correcta
36	hs039	4	2	-1.00e+00	0.00e+00	2.46e-21	16	29	4.45e-02	correcta
37	hs040	4	3	-2.50e-01	2.48e-16	2.94e-16	4	5	4.71e-02	correcta
38	hs046	5	2	2.37e-12	3.85e-07	7.64e-09	20	34	5.30e-02	correcta
39	hs047	5	3	2.64e-11	9.36e-08	3.74e-07	17	20	5.39e-02	correcta
40	hs049	5	2	3.52e-08	0.00e+00	1.03e-05	14	15	5.69e-02	correcta
41	hs050	5	3	6.16e-32	8.88e-16	2.53e-16	9	10	4.92e-02	correcta
42	hs061	3	2	-1.44e+02	0.00e+00	0.00e+00	10	16	5.45e-02	correcta
43	hs077	5	2	2.42e-01	1.05e-13	1.31e-12	9	11	5.49e-02	correcta
44	hs078	5	3	-2.92e+00	1.99e-15	2.06e-15	5	6	3.98e-02	correcta
45	hs079	5	3	7.88e-02	2.22e-16	2.61e-16	5	6	5.44e-02	correcta
46	hs100lnp	7	2	6.81e+02	2.84e-14	7.15e-14	7	8	4.98e-02	correcta
47	hs111lnp	10	3	-4.78e+01	4.04e-12	5.94e-12	11	14	5.05e-02	correcta
48	lch	600	1	-4.68e+00	6.36e-02	5.87e-03	401	7905	1.04e+01	correcta
49	maratos2	2	1	-1.00e+00	0.00e+00	1.16e-16	12	15	3.67e-02	correcta
50	maratos	2	1	-1.00e+00	5.33e-13	4.95e-13	12	14	3.46e-02	correcta
51	mwright	5	3	2.50e+01	8.95e-16	1.03e-14	10	25	2.62e-02	correcta
52	orthrdm2	4003	2000	1.56e+02	2.20e-12	6.10e-13	7	10	1.01e+00	correcta
53	orthrds2	203	100	3.05e+01	2.78e-13	5.94e-01	401	7159	1.71e+00	correcta
54	orthrega	517	256	1.57e+03	2.28e-01	1.05e+04	401	8207	1.00e+01	correcta
55	orthregb	27	6	1.16e-21	8.88e-16	6.81e-11	3	4	3.73e-02	correcta
56	orthregc	10005	5000	1.90e+02	1.83e-14	1.93e-12	13	19	1.07e+01	correcta
57	orthregd	10003	5000	1.52e+03	3.66e-12	4.17e-13	7	10	6.46e+00	correcta
58	orthrgdm	10003	5000	1.51e+03	3.89e-12	7.94e-13	8	11	6.62e+00	correcta
59	orthrgds	10003	5000	1.52e+03	2.96e-12	2.01e-12	30	122	1.24e+02	correcta

Cuadro 3: Resultados de los problemas con función de mérito y con regularización