

## APLICACIÓN DE UN ALGORITMO DE LAGRANGIANO AUMENTADO A LA RESOLUCIÓN DEL PROBLEMA DE CUADRADOS MÍNIMOS NO LINEALES

**Graciela M. Croceri y Graciela N. Sottosanto**

*Departamento de Matemática, Universidad Nacional del Comahue, Buenos Aires 1400, 8300 Neuquén,  
Argentina, [graciela.croceri@faea.uncoma.edu.ar](mailto:graciela.croceri@faea.uncoma.edu.ar), [graciela.sottosanto@faea.uncoma.edu.ar](mailto:graciela.sottosanto@faea.uncoma.edu.ar)*

**Palabras Clave:** Lagrangiano aumentado, matriz Hessiana estructurada, problemas de cuadrados mínimos.

**Resumen.** En este trabajo se presenta la aplicación de un algoritmo basado en la minimización secuencial del Lagrangiano aumentado para resolver el problema de cuadrados mínimos no lineales con restricciones. En el caso de restricciones de igualdad, el algoritmo combina una técnica de gradiente conjugado y región de confianza. Para el caso de restricciones de desigualdad, el problema se transforma en uno con restricciones de igualdad y cotas sobre las variables, mediante el agregado de variables de holgura. La estrategia de resolución está basada en un método de gradiente proyectado y la convergencia desde cualquier punto inicial se obtiene por medio de un esquema tipo Armijo. El algoritmo toma ventaja de la forma especial del problema de cuadrados mínimos para usar la información de segundo orden en una forma estructurada del tipo BFGS (Broyden, Fletcher, Goldfard y Shanno), ya sea sobre la matriz Hessiana o su inversa. Se presentan resultados numéricos preliminares usando una implementación computacional en MATLAB.

## 1. INTRODUCCIÓN

En este trabajo consideramos hallar una solución  $x^* \in \mathbb{R}^n$  del problema general de cuadrados mínimos con restricciones de igualdad, desigualdad y cotas sobre las variables

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|F(x)\|_2^2 = \frac{1}{2} \sum_{i=1}^m f_i(x)^2 \\ \text{s.a} \quad & h_j(x) = 0, \quad j = 1, \dots, p, \\ & g_k(x) \leq 0, \quad k = 1, \dots, r, \\ & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n. \end{aligned} \quad (1)$$

La función residual  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , en general, es no lineal,  $f_i(x)$  denota la  $i$ -ésima componente de la función  $F$  y la norma es la estándar  $l_2$ . Las funciones que definen las restricciones  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  y  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ , así como la función objetivo se asumen dos veces continuamente diferenciables.

Los problemas del tipo (1) tienen muchas aplicaciones prácticas en ingeniería, física, programación matemática y otros campos como el análisis estadístico en las ciencias sociales.

En particular, problemas con restricciones de igualdad pueden aparecer en ajuste de curvas en forma implícita, donde un modelo  $h(t, \omega, x) = 0$  con parámetros  $x \in \mathbb{R}^n$ , se usa para ajustar un conjunto de datos  $(t_i, \bar{\omega}_i)$ ,  $i = 1, 2, \dots, m$ . En este caso, toma la forma

$$\begin{aligned} \min_{x, \omega_1, \dots, \omega_m} \quad & \frac{1}{2} \sum_{i=1}^m \|\omega_i - \bar{\omega}_i\|_2^2 \\ \text{s.a} \quad & h(t_i, \omega_i, x) = 0, \quad i = 1, 2, \dots, m. \end{aligned} \quad (2)$$

Este problema pertenece a la clase más general de problemas de cuadrados mínimos no lineales de regresión ortogonal cuya forma general es

$$\begin{aligned} \min_{r \in \mathbb{R}^m, x \in \mathbb{R}^n} \quad & \frac{1}{2} \|r\|_2^2 \\ \text{s.a} \quad & h(r, x) = 0. \end{aligned} \quad (3)$$

Otro ejemplo que se presenta en muchas áreas es el de la estimación de parámetros. Puede ocurrir que algunos de los parámetros a ser estimados estén restringidos a pertenecer a alguna región a fin de que el modelo esté bien definido y tenga significado físico. En este caso, el problema de cuadrados mínimos podría tener sólo restricciones de desigualdad y cotas sobre las variables.

Se han desarrollado muchos algoritmos para resolver estos problemas, la mayoría de los cuales explota su estructura. El método de Gauss-Newton y el de Levenberg-Marquardt consideran el mismo modelo cuadrático y tienen la misma forma de elegir una aproximación de la matriz Hessiana. Su diferencia está en que mientras Gauss-Newton usa una estrategia de búsqueda lineal, Levenberg-Marquardt utiliza región de confianza.

Si bien estos métodos tienen buen desempeño cuando la función residual toma valores pequeños en la solución, para problemas de residuos grandes sus tasas de convergencia tienden sólo a ser lineales. La razón de este comportamiento es que la parte de la matriz Hessiana que se ignora puede resultar muy significativa.

Como en general, el usuario no sabe antes de resolver, si estará ante un caso de residuos pequeños, medianos o grandes, una alternativa es desarrollar algoritmos que hagan una mejor

aproximación de la información de segundo orden.

La estructura especial del problema de cuadrados mínimos para el caso sin restricciones fue estudiada por J.E. Dennis, D.M. Gay y R.E. Welsch (Dennis et al. (1981a)) para establecer una actualización secante de la matriz Hessiana. Esta idea ha sido bien implementada por ellos en el código NL2SOL (Dennis et al. (1981b)). La actualización secante estructurada DFP estudiada por J.E. Dennis y H.F. Walker (Dennis y Walker (1981)) fue incluida en el código pero después se reportaron mejores resultados usando una actualización secante estructurada BFGS (Dennis et al. (1985)) sugerida por M. Al-Baali y R. Fletcher (Al-Baali y Fletcher (1985)). J. E. Dennis, H. J. Martinez y R. A. Tapia (Dennis et al. (1989)) establecieron que bajo adecuadas hipótesis, el método es localmente  $q$ -superlinealmente convergente.

En este trabajo presentamos un algoritmo de Lagrangiano aumentado adaptado para resolver el problema de cuadrados mínimos. El método, desarrollado originalmente para resolver el problema de minimización con restricciones de igualdad (Maciel y Sottosanto (2002)), fue aplicado después a la resolución de problemas de cuadrados mínimos con restricciones de igualdad (Croceri et al. (2007)) donde se incorporó una aproximación secante del tipo BFGS (Fletcher (1987)) que aprovecha la estructura típica del problema. Para el caso de restricciones de desigualdad, el problema se transforma en uno con restricciones de igualdad y cotas sobre las variables, mediante el agregado de variables de holgura. La estrategia de resolución está basada en un método de gradiente proyectado y la convergencia desde cualquier punto inicial se obtiene por medio de un esquema tipo Armijo.

El trabajo está organizado de la siguiente manera. Las características del problema se presentan en la sección 2. La sección 3 se dedica al algoritmo. En la sección 4 se comentan los experimentos numéricos y finalmente, las conclusiones en la sección 5.

## 2. EL PROBLEMA

Dado el problema general de cuadrados mínimos con restricciones y cotas (1) podemos convertirlo en un problema con restricciones de igualdad y cotas sobre las variables introduciendo variables de holgura  $s_k$  y reemplazando las desigualdades generales  $g_k(x) \leq 0$ ,  $k = 1, \dots, r$  por  $g_k(x) + s_k = 0$ ,  $k = 1, \dots, r$ , con  $s_k \geq 0$ ,  $k = 1, \dots, r$ .

Las restricciones de cotas sobre las variables no necesitan ser modificadas y el problema (1) se transforma en

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|F(x)\|_2^2 = \frac{1}{2} \sum_{i=1}^m f_i(x)^2 \\ \text{s.a} \quad & h_j(x) = 0, \quad j = 1, \dots, p + r, \\ & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n + r, \end{aligned} \quad (4)$$

donde algunas de las componentes del vector  $l$  de cotas inferiores sobre las variables pueden fijarse en  $-\infty$ , significando que no existen cotas inferiores sobre esas componentes, similarmente para el vector  $u$ .

Sin pérdida de generalidad y con el objetivo de simplificar la notación, en lo que sigue, consideraremos el siguiente problema

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|F(x)\|_2^2 = \frac{1}{2} \sum_{i=1}^m f_i(x)^2 \\ \text{s.a} \quad & c_j(x) = 0, \quad j = 1, \dots, p, \\ & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n. \end{aligned} \quad (5)$$

### 3. EL ALGORITMO

Asociado al problema (5) podemos definir la siguiente función aumentada de Lagrange (Bertsekas (1996)),

$$L_\mu(x, \lambda) = \|F(x)\|_2^2 + \lambda^T c(x) + \frac{1}{2\mu} \|c(x)\|_2^2, \quad (6)$$

donde  $\lambda$  es el vector de multiplicadores y  $\mu > 0$  es el parámetro de penalización. El método de resolución que proponemos en este trabajo está basado en la minimización secuencial de la función (6). Si las variables del problema original no están obligadas a satisfacer cotas, seguimos el esquema desarrollado en Croceri et al. (2007) y en cada iteración externa se halla una solución aproximada del problema

$$\min_x L_\mu(x, \lambda). \quad (7)$$

Para el caso en que el problema original posea cotas sobre las variables o bien, se hayan impuesto con motivo de la transformación del problema con restricciones de desigualdad, en la iteración externa, proponemos la resolución de subproblemas de la forma

$$\begin{aligned} \min_x \quad & L_\mu(x, \lambda) \\ \text{s.a} \quad & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n. \end{aligned} \quad (8)$$

Al final de cada iteración externa se actualizan tanto el vector de multiplicadores como el parámetro de penalización.

Claramente, el algoritmo consiste en una iteración externa y una interna en la que se resuelve el subproblema de minimizar la función aumentada de Lagrange. Se usan diferentes técnicas para el caso irrestricto y para el caso de minimización con cotas.

Si el problema sólo tiene restricciones de igualdad, la iteración externa termina cuando la norma del gradiente de la función de Lagrange se mantiene menor a cierta tolerancia y se alcanza la factibilidad, es decir,  $\|\nabla l(x_k), \lambda_k\|_2 \leq \epsilon_1$  y  $\|c(x_k)\|_2 \leq \epsilon_2$ , donde

$$l(x, \lambda) = \|F(x)\|_2^2 + \lambda^T c(x).$$

Si además, el problema tiene restricciones de cota, entonces las condiciones para terminar la iteración externa están dadas por  $\|x_k - \mathcal{P}(x_k - \nabla l(x_k), \lambda_k)\|_2 \leq \epsilon_1$  and  $\|c(x_k)\|_2 \leq \epsilon_2$ , donde  $\mathcal{P}$  representa la proyección en la caja determinada por las cotas del problema.

#### 3.1. Descripción del algoritmo

Supongamos que están dados  $x_k \in \mathbb{R}^n$ ,  $\lambda_k \in \mathbb{R}^m$ ,  $\|\lambda_k\|_2 \leq M$ ,  $M > 0$ , constante y  $0 < \mu_{min} \leq \mu_k$ . Dadas, además, las tolerancias  $\epsilon_1$  y  $\epsilon_2 > 0$ , los siguientes pasos permiten hallar  $x_*$  y su multiplicador asociado  $\lambda_*$ .

#### Algorithm 3.1 (Algoritmo básico de la iteración externa. Restricciones de igualdad)

**Paso I.** Hallar  $(x_\star)_{k+1}$ , solución aproximada del subproblema de minimización

$$\min_x L_{\mu_k}(x, \lambda_k)$$

usando el algoritmo que se describe en la sección 3.2. de gradiente conjugado/región de confianza (Steihaug (1983)).

**Paso II. (Test de convergencia)**

- Si  $\|\nabla l((x_\star)_{k+1}, \lambda_k)\|_2 \leq \epsilon_1$  y  $\|c((x_\star)_{k+1})\|_2 \leq \epsilon_2$  entonces TERMINAR,  $(x_\star)_{k+1} = x_\star$  es la solución aproximada del problema de minimización sin restricciones,
- de lo contrario, actualizar el parámetro de penalización tal que  $0 < \mu_{k+1} < \mu_k$ .
  - Si  $\mu_{k+1} \geq \mu_{min}$  actualizar el estimado del multiplicador e ir al paso I,
  - de lo contrario terminar con la solución aproximada  $(x_\star)_{k+1} \cong x_\star$ .

En el caso en que las variables del problema estén acotadas, el algoritmo tiene la siguiente forma.

**Algorithm 3.2 (Algoritmo básico de la iteración externa. Cotas sobre las variables)**

**Paso I.** Hallar  $(x_\star)_{k+1}$ , solución aproximada del subproblema de minimización

$$\begin{array}{ll} \min_x & L_\mu(x, \lambda) \\ \text{s.a} & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n. \end{array}$$

usando el algoritmo que se describe en la sección 3.4. de gradiente proyectado con búsqueda lineal tipo Armijo y actualización BFGS de la inversa de la matriz Hessiana (Kelley (1999)).

**Paso II. (Test de convergencia)**

- Si  $\|(x_\star)_{k+1} - \mathcal{P}((x_\star)_{k+1} - \nabla l((x_\star)_{k+1}, \lambda_k))\|_2 \leq \epsilon_1$  and  $\|c((x_\star)_{k+1})\|_2 \leq \epsilon_2$  entonces TERMINAR,  $(x_\star)_{k+1} = x_\star$  es la solución aproximada del problema de minimización sin restricciones,
- de lo contrario, actualizar el parámetro de penalización tal que  $0 < \mu_{k+1} < \mu_k$ .
  - Si  $\mu_{k+1} \geq \mu_{min}$  actualizar el estimado del multiplicador e ir al paso I,
  - de lo contrario terminar con la solución aproximada  $(x_\star)_{k+1} \cong x_\star$ .

### 3.2. El algoritmo de gradiente conjugado y región de confianza

En el caso irrestricto, el subproblema se resuelve usando un método de gradiente conjugado el cual tiene incorporada una estrategia de región de confianza (Steihaug (1983)). Por lo tanto, el paso se obtiene como solución del subproblema de región de confianza

$$\min_{\|s\|_2 \leq \delta_k} Q_k(s, \lambda_k, \mu_k),$$

donde  $Q_k$  es un modelo cuadrático del Lagrangiano aumentado alrededor de  $x_k$  y  $\delta_k > 0$  es el radio de la región de confianza. Si se consigue suficiente reducción de la función  $L_{\mu_k}(x, \lambda_k)$  en el minimizador aproximado del modelo cuadrático, se acepta el correspondiente paso de prueba y se obtiene el nuevo iterado  $x_{k+1} = x_k + s_k$ , de lo contrario el radio de la región de confianza se reduce y actualiza siguiendo esquemas estandar (Fletcher (1987), Moré (1984)).

Cuando  $\|\nabla L_{\mu_k}(x_{k+1}, \lambda_k)\|_2 \leq \epsilon_k$ , para alguna sucesión dada  $\{\epsilon_k\}$ ,  $\epsilon_k \rightarrow 0$ , la iteración interna termina con  $x_{k+1} = (x_*)_{k+1}$  y el método regresa a la iteración externa. El método de gradiente conjugado nos permite aproximar la solución del subproblema cuadrático vía una técnica *dogleg* de modo que la curva solución  $s(\delta)$  resulta aproximada por una poligonal. La información de segundo orden se aproxima utilizando estrategias secantes en una forma estructurada que aprovecha la ventaja de la forma especial que tienen las segundas derivadas.

Aún cuando se involucren matrices arbitrarias, el método resulta bien definido y en caso que la aproximación de la matriz Hessiana sea definida positiva se obtiene convergencia local superlineal en el minimizador local del subproblema irrestricto.

### 3.3. Aproximación secante estructurada

El gradiente de la función (6) es

$$\nabla L_{\mu}(x, \lambda) = \nabla F(x)F(x) + \nabla c(x) \left( \frac{1}{\mu} c(x) + \lambda \right),$$

donde  $\nabla F \in \mathbb{R}^{n \times m}$  denota la matriz Jacobiana de  $F(x)$  y  $\nabla c \in \mathbb{R}^{n \times p}$  la matriz Jacobiana de las restricciones. Se definen por columnas como sigue: la  $i$ -ésima columna de  $\nabla F(x)$  es

$$\nabla F(x)_{.i} = \left( \frac{\partial f_i(x)}{\partial x_1}, \dots, \frac{\partial f_i(x)}{\partial x_n} \right)^T, \quad i = 1, \dots, m,$$

y la  $j$ -ésima columna de  $\nabla c(x)$ ,

$$\nabla c(x)_{.j} = \left( \frac{\partial c_j(x)}{\partial x_1}, \dots, \frac{\partial c_j(x)}{\partial x_n} \right)^T, \quad j = 1, \dots, p.$$

La matriz Hessiana del Lagrangiano aumentado es

$$\begin{aligned} \nabla^2 L_{\mu}(x, \lambda) &= \nabla F(x) \nabla F(x)^T + \sum_{i=1}^m f_i(x) \nabla^2 f_i(x) + \frac{1}{\mu} \nabla c(x) \nabla c(x)^T \\ &\quad + \frac{1}{\mu} \sum_{i=1}^p c_i(x) \nabla^2 c_i(x) + \sum_{i=1}^p \lambda \nabla^2 c_i(x). \end{aligned}$$

La estructura de esta matriz es importante porque juega un rol fundamental en la formulación de cualquier problema de cuadrados mínimos. Vamos a considerar

$$G_{\mu}(x) = \nabla F(x) \nabla F(x)^T + \frac{1}{\mu} \nabla c(x) \nabla c(x)^T,$$

que es la parte que contiene información de primer orden y llamaremos “disponible” y

$$S_{\mu}(x, \lambda) = \sum_{i=1}^m f_i(x) \nabla^2 f_i(x) + \frac{1}{\mu} \sum_{i=1}^p c_i(x) \nabla^2 c_i(x) + \sum_{i=1}^p \lambda_i \nabla^2 c_i(x),$$

es la matriz que contiene la información de segundo orden, de manera que

$$\nabla^2 L_\mu(x, \lambda) = G_\mu(x) + S_\mu(x, \lambda).$$

Como  $\nabla F(x)$  y  $\nabla c(x)$  están disponibles ya sea analíticamente o mediante el cálculo de diferencias finitas, el término  $G_\mu(x)$  también está disponible al momento de calcular la matriz Hessiana. El objetivo es aproximar el término  $S_\mu(x)$ . En el proceso de minimizar la función  $L_\mu(x, \lambda)$ , dado un iterado  $x$ , la matriz Hessiana se actualiza por un método secante tal que

$$B_+ s = y,$$

donde  $B_+$  es una aproximación de  $\nabla^2 L_\mu(x_+, \lambda)$ ,  $s = x_+ - x$  es el paso de prueba y

$$y = \nabla_x L_\mu(x_+, \lambda) - \nabla_x L_\mu(x, \lambda)$$

es una aproximación de  $\nabla^2 L_\mu(x_+, \lambda)s$ .

Nuestro objetivo es tomar ventaja de la estructura de la matriz Hessiana manteniendo la parte que incluye la información de primer orden y aproximando solamente la parte que contiene la información de segundo orden.

Una aproximación secante estructurada de  $\nabla^2 L_\mu(x, \lambda)$  se define como

$$B = G_\mu(x) + A_\mu,$$

donde  $A_\mu$  es una aproximación de  $S_\mu(x, \lambda)$ . La matriz  $B$  se actualiza de acuerdo a la siguiente fórmula

$$B_+ = G_\mu(x_+) + A_{\mu+},$$

donde  $A_{\mu+}$  es una aproximación secante de  $A$ .

Siguiendo las ideas de J.E. Dennis, H. Martinez and R.A. Tapia ([Dennis et al. \(1989\)](#)) consideramos

$$\bar{y} = y^s - G_\mu(x_+)s,$$

donde  $\bar{y}$  es una aproximación de  $S(x_+, \lambda)s$  y colocamos el superíndice  $s$  para denotar que es estructurada.

Escribimos

$$B^s = G(x_+) + A.$$

La actualización secante  $B^s$  es

$$B_+ = B^s + \Delta(s, y^s, B^s, v)$$

donde

$$\Delta(s, y, B, v) = \frac{(y - Bs)v^T + v(y - Bs)^T}{v^T s} - \frac{(y - Bs)^T s}{(v^T s)^2} v v^T$$

define la clase de actualizaciones secantes y  $v$ , conocido como la escala, en la actualización del tipo BFGS tiene la forma

$$v = y + \left( \frac{y^T s}{s^T B s} \right)^{\frac{1}{2}} B s.$$

Como

$$\Delta(s, y^s, B^s, v) = \Delta(s, \bar{y}, A, v),$$

es posible escribir

$$B_+ = A + G_{\mu_k}(x_+) + \Delta(s, \bar{y}, A, v(s, y^s, B^s)) \quad (9)$$

y llamamos

$$A_+ = A + \Delta(s, \bar{y}, A, v(s, y^s, B^s))$$

aproximación secante estructurada de  $A$ . Claramente, la matriz  $A_+$  satisface la ecuación secante. Para obtener el vector  $y$  vemos que

$$\begin{aligned} S_\mu(x_+, \lambda)s &= \sum_{i=1}^m f_i(x_+) \nabla^2 f_i(x_+)s + \sum_{i=1}^p \left( \frac{1}{\mu} c_i(x_+) + \lambda_i \right) \nabla^2 c_i(x_+)s \\ &\doteq \sum_{i=1}^m f_i(x_+) (\nabla f_i(x_+) - \nabla f_i(x)) \\ &\quad + \sum_{i=1}^p \left( \frac{1}{\mu} c_i(x_+) + \lambda_i \right) (\nabla c_i(x_+) - \nabla c_i(x)) \\ &= (\nabla F(x_+) - \nabla F(x)) F(x_+) + \\ &\quad (\nabla c(x_+) - \nabla c(x)) \left( \frac{1}{\mu} c(x_+) + \lambda \right). \end{aligned}$$

Finalmente definimos

$$\begin{aligned} \bar{y} &\doteq (\nabla F(x_+) - \nabla F(x)) F(x_+) \\ &\quad + (\nabla c(x_+) - \nabla c(x)) \left( \frac{1}{\mu} c(x_+) + \lambda \right) \end{aligned}$$

tal que

$$A_+ \in Q = \{A : A = A^T, As = \bar{y}\}.$$

Observamos que sólo en la definición de  $\bar{y}$  y en el cálculo de  $\nabla L_\mu(x, \lambda)$  se usa la información de primer orden.

Entonces, llamando

$$y^s = \nabla L(x_+, \lambda, \mu) - \nabla L(x, \lambda, \mu)$$

la actualización estructurada  $B_+$  de la matriz  $\nabla^2 L$  dada en (9) está completamente definida.

### 3.4. El algoritmo del gradiente proyectado

Para resolver los subproblemas con cotas sobre las variables

$$\begin{aligned} \text{mín}_x & L_{\mu_k}(x, \lambda_k) \\ \text{s.a.} & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n, \end{aligned} \quad (10)$$



utilizamos un algoritmo que combina un método de gradiente proyectado con una regla tipo Armijo (Bertsekas (1976)).

En este trabajo seguimos una implementación diseñada por Kelley (1999) que usa una estrategia BFGS recursiva para actualizar la inversa generalizada de una parte de la matriz Hessiana.

El conjunto factible de este problema es el conjunto

$$\Omega = \{x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, i = 1, \dots, n\}.$$

Decimos que la  $i$ -ésima restricción es activa en  $x$  si  $(x)_i = l_i$  ó  $(x)_i = u_i$ . En cada  $x \in \Omega$  los conjuntos de índices  $\mathcal{A}(x)$  e  $\mathcal{I}(x)$  son respectivamente los conjuntos de índices activos e inactivos.

El algoritmo del gradiente proyectado es una extensión natural del algoritmo de dirección de máximo descenso para el caso de problemas con restricciones de cotas.

Recordemos que las condiciones de Karush Kuhn y Tucker para el problema (10) establece que un punto  $x_* \in \Omega$  es estacionario para el subproblema si y sólo si

$$x_* = \mathcal{P}(x_* - \alpha \nabla L(x_*))$$

para todo  $\alpha \geq 0$ , donde  $\mathcal{P}$  es la proyección sobre  $\Omega$

$$\mathcal{P}(x)_i = \begin{cases} l_i & \text{si } (x)_i \leq l_i \\ (x)_i & \text{si } l_i < (x)_i < u_i \\ u_i & \text{si } (x)_i \geq u_i. \end{cases}$$

El método del gradiente proyectado tiene una propiedad destacada que consiste en que bajo determinadas condiciones el conjunto de índices de restricciones activas  $\mathcal{A}(x_k)$  a partir de un número determinado  $k$  de iteraciones es el mismo que el que se tiene en la solución  $\mathcal{A}(x_*)$ . Para asegurar el cumplimiento de esta propiedad, el método tiene que identificar, en cada iteración, el conjunto de restricciones activas.

Dado un iterado actual  $x_k$  el nuevo iterado se obtiene como

$$x_{k+1} = \mathcal{P}(x_k - \alpha \nabla L_\mu(x_k))$$

donde  $\alpha$  es un parámetro de longitud de paso correspondiente a una regla tipo Armijo, de manera tal que asegure suficiente decrecimiento en la función objetivo. Para  $\alpha > 0$  se define

$$x(\alpha) = \mathcal{P}(x - \alpha \nabla L_\mu(x)).$$

Para el caso de restricciones de cotas sobre las variables, la condición de suficiente decrecimiento se expresa como

$$L_\mu(x(\alpha)) - L_\mu(x) \leq \frac{-\beta}{\alpha} \|x - x(\alpha)\|^2,$$

donde  $\beta$  es un parámetro cuyo valor típico es  $10^{-4}$ .

Como criterio de terminación para nuestro algoritmo elegimos  $\|x_k - \mathcal{P}(x_k - \nabla L_\mu(x_k))\| < \epsilon$  para cierta tolerancia  $\epsilon$ .

En la implementación que usamos en este trabajo, a efectos de mejorar la convergencia del método, en lugar de elegir como dirección de descenso, la opuesta a la del gradiente, se usa un vector  $d$  que se actualiza recursivamente premultiplicando por una aproximación secante estructurada de la inversa generalizada de la matriz Hessiana.

### 3.5. Actualización del parámetro de penalización y del multiplicador

Un punto crucial del algoritmo es la elección del parámetro de penalización. Para asegurar convergencia global se debe generar una sucesión no decreciente  $\{\mu_k\}$ . Aunque los métodos pueden ser formulados con  $p$  diferentes parámetros de penalización, uno para cada componente de  $c(x)$ , en la práctica, esto no presenta ventajas numéricas.

Nosotros empleamos el siguiente esquema: Si  $\|c(x_{k+1})\|_2$  no es suficientemente menor que  $\|c(x_k)\|_2$  entonces el parámetro de penalización es disminuido por la regla  $\mu_{k+1} = \varphi \mu_k$  donde  $\varphi \in (0, 1)$ . Aquí consideramos que  $\|c(x_{k+1})\|_2$  debe ser menor que  $0,05\|c(x_k)\|_2$ .

El comportamiento del parámetro de penalización es independiente del vector de multiplicadores. Nuestra filosofía para actualizar el vector de multiplicadores es mantenerlo constante a lo largo de la minimización cuasi Newton y actualizarlo al final del mismo, por medio de alguna fórmula. En este trabajo, elegimos una fórmula de primer orden para actualizar el vector de multiplicadores de Lagrange, comenzando con  $\lambda_0 = 0$ ,

$$\lambda_{k+1} = \lambda_k + \frac{1}{\mu_k} c_k.$$

### 3.6. Propiedades del algoritmo

A fin de establecer las propiedades de convergencia global del algoritmo de penalización, es necesario primero, asegurar la convergencia del algoritmo que resuelve la iteración interna. En [Steihaug \(1983\)](#) se probó que el algoritmo de gradiente conjugado y región de confianza está bien definido y converge globalmente a un punto estacionario del problema sin restricciones (nuestro subproblema para el caso en que solamente estén presentes restricciones de igualdad). Si, además, se asume que la actualización secante de la matriz Hessiana satisface la condición general

$$\|B_k\|_2 \leq a_1 + a_2 \sum_{i=0}^k \|s_i\|_2, \quad (11)$$

donde  $a_1$  y  $a_2$  son constantes positivas que dependen de la forma de actualización de  $B_k$  y  $s_i$  es la sucesión de pasos generada por el algoritmo, es posible obtener convergencia q-superlineal a un minimizador.

La condición (11) se satisface con cualquier fórmula basada en las actualizaciones del tipo BFGS (ver [Steihaug \(1983\)](#) y [Toint \(1979\)](#)).

Respecto del algoritmo de gradiente proyectado y actualización BFGS con esquema de Armijo, bajo hipótesis estándar, en [Kelley \(1999\)](#) se probó convergencia local q-superlineal a un minimizador no degenerado del problema de minimización del Lagrangiano aumentado con cotas sobre las variables. Las propiedades de convergencia del algoritmo de Lagrangiano aumentado fueron probadas en [Maciel y Sottosanto \(2002\)](#) y [Croceri et al. \(2007\)](#). Un resultado más general puede encontrarse en [Diniz-Ehrhardt et al. \(2004\)](#).

## 4. RESULTADOS NUMERICOS

El algoritmo fue codificado en Matlab. A fin de evaluar la performance, se han resuelto 86 problemas test tomados de [Hock y Schittkowski \(1981\)](#), [Schittkowski \(1987\)](#), [Betts \(1977\)](#) y [Conn et al. \(1988\)](#). Se ha trabajado tanto con problemas de cuadrados mínimos con restricciones de igualdad y desigualdad y problemas que sólo tienen restricciones de cota.

En el caso de restricciones de igualdad, a fin de validar nuestra actualización secante estructurada (SBFGS), los problemas test fueron también resueltos a modo de comparación, con

implementaciones basadas en los trabajos de Eriksson (1999) y Z. Wei, G. Li y L. Qi (Wei et al. (2006)) que denominamos Broyden con interpolación (BwI) y cuasi Newton modificadas (MqN), respectivamente, que usan diferentes actualizaciones de la información de segundo orden.

### Actualización de Broyden con interpolación (BwI)

En Eriksson (1999) se propone una familia de actualizaciones de Broyden con estrategias de interpolación que aprovecha las buenas propiedades geométricas del problema de cuadrados mínimos sin restricciones y los pasos previos. Cuando estos elementos son incorporados a un algoritmo de la familia de Broyden, se consigue acelerar su convergencia. En este trabajo, la estrategia BwI se adapta para el problema de minimización del Lagrangiano aumentado. La aproximación de la parte  $S_\mu(x, \lambda)$  de la matriz Hessiana se efectúa en dos etapas: en la primera es posible aplicar una actualización tipo BFGS o bien una actualización simétrica de rango uno (SR1) y en la segunda se efectúa una corrección de rango uno basada en información de los dos pasos previos. La corrección modifica el factor de escala  $v$  y permite aproximar la derivada segunda de la función objetivo en la dirección del paso  $s_p = x_k - x_{k-2}$ .

### Actualización cuasi Newton modificada (MqN)

En Wei et al. (2006) los autores propusieron modificar la actualización cuasi Newton original  $B_+s_k = y_k$  por otra  $B_+s = \tilde{y}$  en la cual  $\tilde{y}$  se toma como la suma de  $y_k$  más  $A_k s_k$  para alguna matriz  $A_k$  que se construye en base a información de la matriz Hessiana de la función objetivo. En este trabajo aplicamos esa construcción a la parte  $S_\mu(x, \lambda)$  de la matriz Hessiana del Lagrangiano aumentado. De esta manera, la matriz  $A$  se construye usando información de la función Lagrangiano aumentado y de su gradiente en el punto anterior y en el actual y la información del paso previo.

## 4.1. Resultados para el caso de restricciones de igualdad

En la figura 1 (a) y (b), se muestran los resultados numéricos obtenidos por medio de la técnica de *performance profiles* (Dolan y Moré (2002)) graficados para diferentes rangos a fin de observar distintas áreas de interés. Se representa la performance basada en el tiempo de ejecución para las tres actualizaciones SBFGS, BwI y MqN. En el caso de la actualización BwI se implementó la primera etapa tanto con la técnica BFGS como SR1 resultando la primera en menor tiempo de ejecución.

En el caso (a) se muestra el desempeño en el intervalo  $[0, 40]$ . Se observa que, SBFGS se impone, en general sobre las otras actualizaciones y para problemas que requieren mayor tiempo de ejecución, es igualmente competitiva con MqN. En la figura (b) el intervalo de tiempo se ha reducido a  $[0, 10]$ , en ella es claro que la actualización SBFGS supera a las otras. En general, el comportamiento se mantiene, cuando se comparan número de evaluaciones de funciones (figura 2).

## 4.2. Resultados para el caso de restricciones de igualdad y cotas sobre las variables

El algoritmo descrito fue testeado sobre 46 problemas de variada complejidad en los cuales se presentan restricciones de igualdad, desigualdad y cotas, de modo tal que en todos ellos, en la iteración interna se aplica el algoritmo de gradiente proyectado con actualización BFGS. En la tabla 1 se presentan resultados para algunos de estos problemas test. En la primera columna está el número de problema, se indica con <sup>(1)</sup> aquellos que pertenecen a Schittkowski (1987),

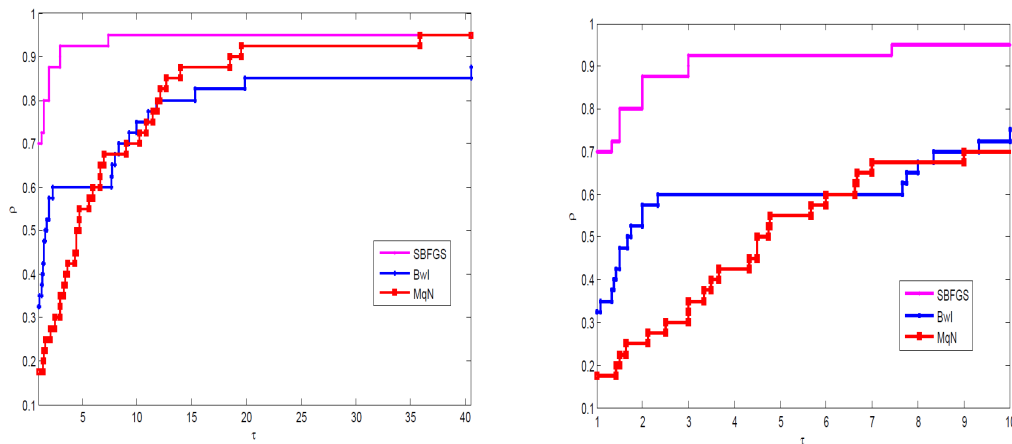


Figura 1: Tiempos de ejecución para (a)  $[0, 40]$ , (b)  $[0, 10]$ .

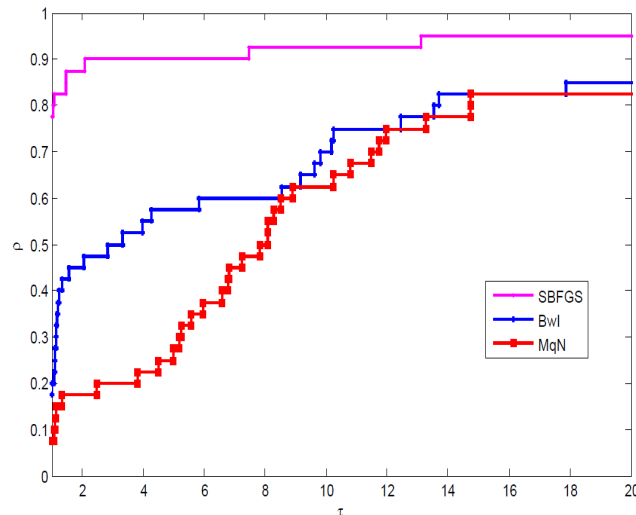


Figura 2: Número de evaluaciones de funciones

con <sup>(2)</sup> a Hock y Schittkowski (1981), con <sup>(3)</sup> a Betts (1977) y con <sup>(2)</sup> a Conn et al. (1988). El número de variables es  $n$ , con  $r$  se representa el número total de restricciones de igualdad, desigualdad y cotas del problema, la cuarta y quinta columna indican el número de iteraciones requeridas externas e internas, para llegar a los valores  $\|Pg\|$  y  $\|c\|$ , gradiente proyectado y norma de las restricciones, respectivamente. Finalmente, en la última columna está el tiempo de CPU necesario para obtener la solución.

## 5. CONCLUSIONES

Hemos introducido un algoritmo de Lagrangiano aumentado para resolver el problema de cuadrados mínimos no lineales con restricciones. El algoritmo combina técnicas como gradiente conjugado y gradiente proyectado con estrategias globalizadoras. Además aprovecha la estructura especial del problema para aproximar la matriz Hessiana mediante técnicas BFGS que resulta muy eficiente cuando se la compara con otras técnicas secantes.

El algoritmo fue testeado con un conjunto de problemas test, que si bien pueden considerarse de porte pequeño o mediano, permiten conjeturar que el método propuesto puede resultar eficiente

Prob.#	$n$	$r$	$I_{tex}$	$I_{tin}$	$\ Pg\ $	$\ c\ $	$nf$	$ng$	$t$
20 <sup>(1)</sup>	2	5	5	199	6.8e-7	3.14e-5	158	57	2.03e-1
23 <sup>(1)</sup>	2	9	4	160653	1.33e-8	1.54e-6	3952	1028	7.02e-1
225 <sup>(2)</sup>	2	5	4	203584	1.89e-6	1.28e-6	4313	1136	7.64e-1
30 <sup>(1)</sup>	3	8	1	55	0	0	27	13	1.09e-1
70 <sup>(1)</sup>	4	9	1	2412	9.88e-5	3.85e-5	180	84	1.87e-1
354 <sup>(2)</sup>	4	5	4	1117	1.41e-8	1.11e-5	310	108	2.18e-1
53 <sup>(1)</sup>	5	10	4	29534	1.68e-7	7.78e-5	1769	416	3.59e-1
11 - 3 <sup>(3)</sup>	5	12	3	496	9.71e-7	1.11e-6	178	638	1.87e-1
11 - 4 <sup>(3)</sup>	5	13	3	393	4.09e-8	1.75e-6	172	62	1.71e-1
11 - 5 <sup>(3)</sup>	3	7	1	193	5.56e-5	3.441e-5	59	26	1.09e-1
Rosenbrock Gen. <sup>(4)</sup>	100	200	1	22831	9.87e-5	0	1177	246	1.25
Rosenbrock Gen. <sup>(4)</sup>	1000	2000	1	22831	9.88e-5	0	1177	246	9.19
Chained Rosenbrock <sup>(4)</sup>	10	20	1	7258	5.47e-5	0	512	156	1.71e-1
Chained Rosenbrock <sup>(4)</sup>	50	100	1	13349	5.07e-5	0	710	205	6.08e-1
Generalized Singular <sup>(4)</sup>	40	800	1	9426	9.61e-5	0	510	196	2.19
Generalized Singular <sup>(4)</sup>	4000	8000	1	25962	9.58e-5	0	714	279	3.11e+2
Generalized Wood <sup>(4)</sup>	40	80	1	4575	8.17e-5	0	358	105	2.03e-1
Generalized Wood <sup>(4)</sup>	400	800	1	4657	6.44e-5	0	347	102	2.028
Generalized Wood <sup>(4)</sup>	4000	8000	1	4671	6.58e-6	0	359	106	1.93e+2

Tabla 1: Algunos problemas test resueltos con el algoritmo de gradiente proyectado.

para problemas más grandes.

## REFERENCIAS

- Al-Baali M. and Fletcher R. Variational methods for non-linear least squares. *J. Oper. Res. Soc.*, 36: 405–421, 1985.
- Bertsekas D. P. On the Goldstein-Levitin-Polyak Gradient Projection Method. *IEEE Transactions on Automatic Control*, 21: 174–184, 1976.
- Bertsekas D. P. *Constrained Optimization and Lagrange Multiplier Methods*. Athenas Scientific, Belmont, MA, 1996.
- Betts J. T. An Accelerated Multiplier Method for Nonlinear Programming. *Journal of Optimization Theory and Applications*, 21(2): 137–174, 1977.
- Conn A. R., Gould N. I. M. and Toint Ph. L. Testing a Class of Methods for Solving Minimization Problems with Simple Bounds on the Variables. *Mathematics of Computation*, 50(182): 399–430, 1988.
- Croceri G. M., Maciel M. C. and Sottosanto G. N. Augmented penalty algorithms based on BFGS secant approximations and trust regions. *Applied Numerical Mathematics*, 57: 320–334, 2007.
- Dennis J. E., Gay D. M. and Welsch R. E. An adaptative nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7: 348–368, 1981.
- Dennis J. E., Gay D. M. and Welsch R. E. Algorithm 573-NL2SOL: An adaptative nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7: 369–383, 1981.
- Dennis J. E., Martínez H. J. and Tapia R. A.. Convergence theory for the structured BFGS secant methods with an application to nonlinear least squares. *Journal of Optimization Theory and Applications*, 61(2): 161–178, 1989.
- Dennis J. E., Songbai S. and Vu P.A. A memoryless augmented Gauss-Newton method for non-linear least squares problems. *Technical Report TR85-1*, Mathematical Sciences Department, Rice University, 1985.

- Dennis J.E. and Walker H. F. Convergence theorems for least change secant update methods. *SIAM Journal on Numerical Analysis*, 18: 949–987, 1981.
- Diniz-Ehrhardt M. A., Gomes-Ruggiero M. A., Martínez J. M. and Santos S. A. Augmented Lagrangian algorithms based on the spectral projected gradient for solving nonlinear programming problems. *Journal of Optimization Theory and Applications*, 123: 497–517, 2004.
- Dolan E., Moré J. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91: 201–213, 2002.
- Eriksson J. Quasi-Newton methods for nonlinear least squares focusing on curvatures. *BIT*, 39(2): 228–254, 1984.
- Fletcher R. *Practical Methods of Optimization*. John Wiley, Chichester and New York, 1987.
- Hock W. and Schittkowski K. Test examples for nonlinear programming codes. *M. Beckmann and H. P. Künzi, eds., Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, Berlin, 1981.
- Kelley C. T. *Iterative Methods for Optimization*. SIAM, Philadelphia, 1999.
- Maciel M. C. and Sottosanto G. N. An augmented penalization algorithm for the equality constrained minimization problem. *Seleto do XXIV CNMAC*, 3(2): 171–180, 2002.
- Moré J. J. Recent developments in algorithms and software for trust region methods. *In A. Bachem, M. Grottschel and B. Korte, eds., Mathematical Programming: The state of the art*, Springer-Verlag, New York, 258–287, 1984.
- Schittkowski K. More test examples for nonlinear programming codes. *M. Beckmann and W. Krelle, eds., Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, Berlin, 1987.
- Steihaug T. The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Numerical Analysis*, 20(3): 626–637, 1983.
- Toint Ph. On the superlinear convergence of an algorithm for solving a sparse minimization problem. *SIAM Journal on Numerical Analysis*, 16(6): 1036–1045, 1979.
- Wei Z., Li G. and Qi L. New quasi-Newton methods for unconstrained optimization problems. *Applied Mathematics and Computation*, 175: 1156–1188, 2006.