Mecánica Computacional, Volumen XIX, pp. 367-372
Compilado por: F. Quintana y S. Felicelli
San Carlos de Bariloche, Noviembre de 2000

UN METODO DE PENALIZACION PARA EL PROBLEMA DE MINIMIZACION CON RESTRICCIONES

María C. Maciel
Departamento de Matemática
Universidad Nacional del Sur, Bahía Blanca, Argentina.
Graciela N. Sottosanto
Departamento de Matemática, Facultad de Economía y Administración
Universidad Nacional del Comahue, Neuquén, Argentina.

Resumen

En esta contribución se presenta un método iterativo para resolver el problema de minimización no lineal con restricciones de igualdad. El método está basado en la minimización secuencial de la función de penalización diferenciable conocida como Lagrangiano Aumentado. Para resolver los subproblemas de minimización irrestricta resultantes se utiliza una técnica de gradientes conjugados combinada con una estrategia de globalización de región de confianza, la cual es especialmente eficiente para problemas de gran porte. Se presenta experimentación numérica preliminar que muestra que con técnicas sencillas el aparente mal condicionamiento de los subproblemas puede ser evitado.

Abstract

In this contribution an iterative method for solving the nonlinear minimization problem with equality constraints is presented. The method is based on the sequential minimization of the differential penalization function known as augmented Lagrangian. In order to solve the unconstrained minimization subproblems a conjugate gradients tecnique combined with a trust region estrategy of globalization is used, which is especially efficient for large scale problems. The preliminary numerical experiments show that the outwardly ill condictioning of the subproblems can be avoid with a simple technique.

1 Descripción del algoritmo

En esta contribución se considera el problema de optimización no lineal con restricciones de igualdad

$$(PRI) \equiv \begin{cases} \text{Minimizar} & f(x) \\ \text{s.a} & C(x) = 0, \end{cases}$$

donde $f : \mathbb{R}^n \to \mathbb{R}$ y $C : \mathbb{R}^n \to \mathbb{R}^m$ son funciones al menos dos veces continuamente diferenciables, $C(x) = (c_1(x), \ldots, c_m(x))^T$ y $m \leq n$.
El algoritmo que se desarrolla para resolver este problema se basa en un método de penalización que involucra la resolución de subproblemas de minimización irrestricta de la función *Lagrangiano aumentado* [1], [2], [4],

$$L(x, \lambda, \mu) = f(x) + \lambda^T C(x) + \frac{1}{2\mu}\|C(x)\|_2^2,$$

donde $\lambda \in \mathbb{R}^m$ es el vector de multiplicadores y $\mu \geq 0$ es el parámetro de penalización.
Cada subproblema se resuelve utilizando *gradientes conjugados* [9], como estrategia de globalización se usa la de *región de confianza* [3] y el parámetro de penalización y el vector de multiplicadores se actualizan siguiendo esquemas estandar.
En cada iteración (iteración externa) se resuelve el problema

$$\min_x L(x, \lambda_k, \mu_k). \tag{1}$$

Este subproblema sin restricciones se resuelve utilizando un método *cuasi-Newton inexacto* al cual se le incorpora la estrategia de región de confianza. El paso cuasi-Newton restringido a la región de confianza se obtiene usando el método de gradientes conjugados [9]. Esto es

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

donde Q es un modelo cuadrático de la función Lagrangiano aumentado alrededor de x_k, λ_k es una estimación del vector de multiplicadores asociado a x_k en la iteración anterior, μ_k el parámetro de penalización y δ_k el radio de la región de confianza.
Un punto clave del algoritmo es la elección del parámetro de penalización: debe generar una sucesión $\{\mu_k\}$ decreciente para poder asegurar convergencia global.
El paso es evaluado, y la región de confianza actualizada en la forma usual de un método que usa esta estrategia de globalización, como se indica en el paso 3 del Algoritmo 2.2.
Claramente, el método usado para resolver el subproblema irrestricto (1) define una iteración interna dentro de cada iteración externa del algoritmo de penalización.
Cada iteración interna termina en $(x_\star)_k$ como solución aproximada de (1) cuando

$$\|\nabla L((x_\star)_k, \lambda_k, \mu_k)\| \leq tol,$$

para alguna tolerancia $tol > 0$.

Un iterado $(x_\star)_k$ y su multiplicador es declarado *próximo* a un punto $(x_\star, \lambda_\star)$ de *Karusch, Kuhn y Tucker* para el problema (PRI) si

$$\|\nabla l((x_\star)_k, \lambda_k)\| \leq \epsilon_1, \quad \|C((x_\star)_k)\| \leq \epsilon_2,$$

donde $l(x, \lambda) = f(x) + \lambda^T C(x)$ es el Lagrangiano asociado al problema (PRI) y ϵ_1, ϵ_2 son constantes positivas.
Si alguna de estas condiciones no se cumple se actualizan el vector de multiplicadores y el parámetro de penalización.
Para el vector de multiplicadores se utilizá la actualización propuesta por Hestenes y Powell [5],

$$\lambda_k = \lambda_{k-1} + \frac{1}{\mu} C(x_k),$$

así como la conocida fórmula de aproximación de cuadrados mínimos

$$\lambda_k = -[\nabla C(x_k)^T \nabla C(x_k)]^{-1} \nabla C(x_k)^T \nabla f(x_k).$$

El parámetro de penalización se actualiza siguiendo un esquema que garantice que la sucesión $\{\mu_k\}$ sea decreciente.

2 Algoritmo propuesto

Algoritmo 2.1 (Algoritmo básico para la iteración externa)
Dado $\mu_{min} > 0$, x_c λ_c no es un punto estacionario de (PRI), los siguientes pasos permiten hallar una nueva aproximación de (x_*, λ_*).

PASO 0. (Inicialización)
Dado $\mu_c \geq \mu_{min}$, calcular $\nabla L(x_c, \lambda_c; \mu_c)$ y $\nabla^2 L(x_c, \lambda_c; \mu_c)$.
Elegir ϵ_1 y $\epsilon_2 > 0$.

PASO I. (Iteración interna)
Hallar x_*^+, solución aproximada del subproblema de minimización sin restricciones

$$\min_x L(x, \lambda_c; \mu_c)$$

usando el algoritmo 2.2.

PASO II. (Test de convergencia)

- **Si** $\|\nabla l(x_*^+, \lambda_c; \mu_c)\| < \epsilon_1$ y $\|C(x_*^+)\| < \epsilon_2$ entonces **PARAR**, $x_*^+ = x_*$ es la solución del problema de minimización con restricciones,
- **caso contrario**, actualizar el parámetro de penalización de manera que $0 < \mu_+ < \mu_c$.
  - **Si** $\mu_+ \geq \mu_{min}$ **actualizar** el estimado del multiplicador e ir al PASO I,
  - **caso contrario PARAR** con la aproximación $x_*^+ \cong x_*$.

Algoritmo 2.2 (Algoritmo para la iteración interna)

PASO 0. (Inicialización)
Dados x_c, λ_c, μ_c, $\nabla L(x_c, \lambda_c; \mu_c)$ *y* $\nabla^2 L(x_c, \lambda_c; \mu_c)$, *construir el modelo cuadrático* Q_c *del Lagrangiano Aumentado alrededor de* x_c.
Elegir $\delta_c \geq \delta_{min} > 0$, η_1, η_2, $\gamma > 0$.

PASO 1. (Cálculo del paso de prueba)
Usar el algoritmo de gradientes conjugados para hallar s_c *solución aproximada de*

$$\min_{\|\delta\| \le \delta_c} Q_c(s; \lambda_c, \mu_c).$$

PASO 2. (Evaluación del paso de prueba)

Calcular

$$Ared_c(s_c; \lambda_c, \mu_c) = L(x_c; \lambda_c, \mu_c) - L(x_c + s_c; \lambda_c, \mu_c),$$
$$Pred_c(s_c; \lambda_c, \mu_c) = -Q_c(s_c; \lambda_c, \mu_c)$$

y

$$\rho_c(s_c; \lambda_c, \mu_c) = \frac{Ared_c(s_c; \lambda_c, \mu_c)}{Pred_c(s_c; \lambda_c, \mu_c)}.$$

PASO 3. (Actualización del punto y del radio de la región de confianza)

- **Si** $\rho_c > \eta_2$ *entonces* $x_+ = x_c + s_c$, *incrementar el radio de la región de confianza* δ_+ *e ir al PASO 4.*
- **Si** $\eta_1 \le \rho_c \le \eta_2$ *entonces* $x_+ = x_c + s_c$, *mantener el radio de la región de confianza* δ_+ *e ir al PASO 4.*
- **Si** $\rho_c < \eta_1$ *entonces* $x_+ = x_c$, *disminuir el radio de la región de confianza* δ_+ *e ir al PASO 1.*

PASO 4. (Test de convergencia)

- **Si** $\|\nabla L(x_+, \lambda_c; \mu_c)\| < \gamma \mu_c$ *entonces PARAR, hacer* $x_+ = x_\star^+$ *e ir al PASO II del algoritmo (2.1) para la iteración externa.*
- **Caso contrario**, *construir el nuevo modelo cuadrático y volver al PASO 1.*

3 Propiedades del algoritmo

Teorema 3.1 *Para el problema (PRI) si las funciones f y C son dos veces continuamente diferenciables, el algoritmo 2.2 está bien definido.*

Demostración: Este teorema es un corolario directo de la teoría desarrollada por Steihaug [9], para el caso sin restricciones. ∎

Teorema 3.2 *Sea $\{(x_\star)_k\} \subset X$ la sucesión generada por el algoritmo 2.1, con $X \subset \mathbb{R}^n$ compacto. Si $x_\star$ es un punto de acumulación de la sucesión tal que $\nabla C(x_\star)$ tiene rango completo y asumiendo que la sucesión de parámetros de penalización $\{\mu_k\}$ tiende a cero cuando k tiende a infinito se cumple*

i) $x_\star$ es un punto estacionario para el problema (PRI) y la sucesión $\{(\lambda_\star)_k\}$ converge al multiplicador de Lagrange $\lambda_\star$ asociado a $x_\star$.

ii) Para todo $k \in K \subset \mathbb{N}$ para los cuales $\{(x_\star)_k\}_{k \in K}$ converge a $x_\star$ se tiene

$$\begin{aligned} (\lambda_\star)_k &= \lambda_\star + o(1) \\ C((x_\star)_k) &= \mu_{k-1}(\lambda_\star - \lambda_{k-1}) + o(\mu_{k-1}). \end{aligned}$$

Demostración: Ver Sottosanto [8]. ∎

4 Experiencia numérica

Para analizar el comportamiento del algoritmo se resolvieron 38 problemas extraidos de los libros de Hock y Schittkowski [6] y Schittkowski [7]. En una implementación preliminar realizada en MATLAB se usaron las constantes $\gamma = 10^{-3}$ y para el paso 3 del algoritmo 2.2, $\eta_1 = 10^{-4}$ y $\eta_2 = 10^{-1}$, y el radio de la región de confianza en cada iteración es al menos $\delta_{min} = 10^{-4}$. Para la sucesión de parámetros de penalización se siguió el esquema $\mu_{k+1} = 0.1\mu_k$, fijando como cota inferior de la sucesión $\mu_{min} = 10^{-10}$. Las tolerancias para convergencia fueron fijadas en $\varepsilon_1 = \varepsilon_2 = 10^{-6}$. La siguiente tabla muestra algunos de los resultados obtenidos. La primer columna indica la referencia y número de problema, las dos siguientes las dimensiones del mismo.
Se presenta la cantidad de iteraciones externas e internas requeridas para dos elecciones diferentes del vector de multiplicadores de Lagrange: la actualización de Hestenes y Powell, λ_{HP} y la de cuadrados mínimos, λ_{CM}. En ambos casos se indica con $\mu_\star$ el valor del parámetro obtenido para obtener convergencia.

Table I: Resultados numéricos.

Problemas			$\lambda = \lambda_{HP}$			$\lambda = \lambda_{CM}$		
# de Prob.	n	m	It. ext.	It. int.	$\mu_\star$	It. ext.	It. int.	$\mu_\star$
HS-40	4	3	5	46	5×10^{-5}	5	46	5×10^{-1}
HS-46	5	2	6	113	5×10^{-6}	6	117	5×10^{-6}
HS-47	5	3	4	27	5×10^{-4}	5	28	5×10^{-5}
HS-56	7	4	6	137	5×10^{-6}	5	104	5×10^{-5}
HS-77	5	2	5	25	5×10^{-5}	4	25	5×10^{-4}
HS-78	5	3	5	46	5×10^{-5}	5	46	5×10^{-5}
HS-79	5	3	5	19	5×10^{-5}	4	18	5×10^{-4}
S-219	4	2	6	32	5×10^{-6}	6	32	5×10^{-6}
S-394	20	1	5	76	5×10^{-5}	5	76	5×10^{-5}
S-395	50	1	5	84	5×10^{-5}	5	84	5×10^{-5}

Más acerca de este algoritmo, sus propiedades y resultados numéricos se pueden ver en Sottosanto [8].

Referencias

[1] D.P. BERTSEKAS. *Constrained Optimization and Lagrange Multiplier Methods.* Athenas Scientific, Belmont, MA, 1996.

[2] D.P. BERTSEKAS. *Nonlinear Programming.* Athenas Scientific, Belmont, MA, 1999. Second Edition.

[3] J.E. DENNIS and R.B. SCHNABEL. *Numerical methods for unconstrained optimization and nonlinear systems.* Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[4] R. FLETCHER. *Practical methods of optimization.* J.Wiley & Sons, New York, 1987.

[5] M.R. HESTENES. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[6] W. HOCK and K.SCHITTKOWSKI. Test example for nonlinear programming codes. In M.Beckmann and W. Krelle, editors, *Lecturs Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 1981.

[7] K. SCHITTKOWSKI. More test example for nonlinear programming codes. In M.Beckmann and W. Krelle, editors, *Lecturs Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 1987.

[8] G. SOTTOSANTO. Métodos de penalización en optimización no lineal. Master's thesis, Departamento de Matemática, Universidad Nacional del Sur, Bahía Blanca, Argentina, 2000. En preparación.

[9] T. STEIHAUG. The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Num. Analysis*, 20(3):626–637, 1983.