TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol. 32, Núm. 1, 1981, pp. 94 a 115

FUNCIONES PENALIDAD Y LAGRANGIANOS AUMENTADOS(*)

Eduardo Ramos Méndez
Departamento de Estadística Matemática
Universidad de Santiago de Compostela

RESUMEN

Por medio de un conjunto de propiedades se caracteriza una amplia familia de funciones que pueden emplearse como penalidad para la resolución numérica de un problema de programación matemática. A partir de ellas se construye un algoritmo de penalizaciones demostrando su convergencia a un punto factible óptimo. Se estudia la situación de los mínimos sin restricciones respecto de la región factible, la monotonía de la sucesión de valores de la función auxiliar y se dan varias cotas de convergencia. Una modificación del término de penalidad convierte a la función objetivo penalizada en un tipo de lagrangiano aumentado con propiedades similares a las del lagrangiano clásico, de las cuales pueden extraerse nuevas técnicas algorítmicas conocidas generalmente como métodos de los multiplicadores.

1. LOS METODOS DE PENALIZACION Y SU CONVERGENCIA

El problema de programación matemática puede formularse de la manera siguiente:

1.1. Problema del mínimo (PM)

Hallar, si existe, un x^* tal que:

(*) Este trabajo es un resumen de la tesis doctoral del autor presentada en la Universidad de Santiago de Compostela.

$$f(x^*) = \min_{x \in X} f(x) \qquad x^* \in X = \{x \mid x \in X^\circ, g(x) \leqslant 0\}$$

siendo $X^\circ \subset \mathbb{R}^n; f: X^\circ \subset \mathbb{R}^n \to \mathbb{R}^1; g: X^\circ \subset \mathbb{R}^n \to \mathbb{R}^m$.

Los métodos de penalización intentan resolver el PM transformándolo en un problema de minimización sin restricciones. Ello se consigue de modo secuencial mediante la consideración de una sucesión de *funciones de penalidad* que se introducen en el objetivo para forzar al óptimo a pertenecer a la región factible.

1.2. Definición

Se llama *función penalidad de índice k asociada al conjunto X* a

$$\Pi_k(x) = \frac{1}{s(k)} p(r(k)\, g(x)) = \sum_{i=1}^{m} \frac{1}{s_i(k)} p_i(r_i(k)\, g_i(x))$$

en donde p_i, r_i, s_i, $i = 1, ..., m$ son funciones reales de variable real verificando: i) p_i es continua; ii) p_i es creciente; iii) $\lim_{y \to -\infty} p_i(y) = 0$; iv) $\lim_{y \to +\infty} p_i(y) = +\infty$; v) $r_i(k) \geqslant s_i(k) > 0$; vi) $r_i(k)$, $s_i(k)$ son crecientes; vii) $\lim_{k \to +\infty} r_i(k) = +\infty$. A las funciones p_i se les llamará simplemente funciones *penalidad* y a r_i, s_i *parámetros admisibles*. Habitualmente se suprimirá el subíndice i en p_i, r_i, s_i bien entendido que no tiene por qué aplicarse las mismas a todas las restricciones.

La definición anterior es más general que las usuales en la literatura (Fiacco-McCormick 68, Polak 70, Bazaraa-Shetty 79) e incluye:

a) Penalizaciones exteriores: $p(0) = 0$

b) Penalizaciones interiores: $p(0) = +\infty$

c) Penalizaciones intermedias: $p(0) = a$; $0 < a < +\infty$

Ejemplos de cada tipo pueden encontrarse en las referencias anteriores y también en Murphy 74, Allran-Johnsen 70. Cada uno de ellos originan una variante del método general, siendo los más conocidos los dos primeros. Por su parte las funciones de tipo intermedio permiten una mejor adaptación a los métodos que emplean lagrangianos aumentados.

1.3. Problema auxiliar de índice k (PAk)

Hallar, si existe, $x^k \in \mathbb{R}^n$ tal que:

$$\psi_k(x^k) = \min_{x \in \mathbb{R}^n} \psi_k(x) \qquad \text{siendo } \psi_k(x) = f(x) + \Pi_k(x)$$

El algoritmo de penalización consiste simplemente en resolver el PAk para una sucesión creciente de valores de k.

La primera cuestión a resolver es la de poder asegurar la existencia de los mínimos x^k. Para las funciones de tipo interior es posible afirmar ésto en condiciones bastantes generales. No ocurre lo mismo para las exteriores e intermedias pudiendo encontrarse ejemplos en los que el PAk no tiene mínimo finito. La solución teórica adoptada por la mayoría de los autores es la de suponer las funciones continuas y limitar la minimización de ψ_k a un compacto $K \subset \mathbb{R}^n$. Sin embargo el mínimo así encontrado puede no ser el mínimo global en todo $\mathbb{R}^n$ por lo que el PAk no sería un verdadero problema sin restricciones. Ello puede evitarse empleando funciones penalidad adecuadas a las características del problema. Otra cuestión interesante es saber si la sucesión $\{x^k\}$ tiene algún punto límite $\bar{x}$ que sea factible y, en este caso, minimice la función objetivo. Eventualmente puede interesar conocer a que valor converge la sucesión $\{\psi_k(x^k)\}$ y los valores del término de penalidad. Estas cuestiones constituyen el análisis de la convergencia que estudian los teoremas siguientes.

1.4. Definición

Una función penalidad se dice *regular* para el PM si se verifica que para cada $\delta > 0$ existen $k^\circ > 0$, $M^\circ > 0$ tales que:

$$0 \leqslant f(x) + M^{\circ} p(r(k^{\circ}) \bar{g}(x)) \quad \forall x \in R^n - X_\delta$$

siendo:

$X_\delta = \{x \mid x \in X^{\circ}, g_i(x) \leqslant \delta \;\; i = 1, ..., m\}$;

$\bar{g}(x) = \text{máx}\ \{g_i(x); i = 1, ..., m\}$

1.5. Teorema

Sean:

a) f y g continuas, p de tipo exterior o intermedio regular,

b) los parámetros tales que $\forall y > 0 \lim\limits_{k \to +\infty} \dfrac{p(r(k)\,y)}{s(k)} = +\infty$, $\lim\limits_{k \to +\infty} \dfrac{r(k)}{s(k)} = +\infty$;

c) $X_\delta \subset X^{\circ}$ compacto no vacío para algún $\delta > 0$;

d) Se verifica que $\forall\, y > 0$, $\forall\, \mu > 1$, $\mu\, p(y) \leqslant p(\mu\, y)$.

Entonces existe un k^* tal que para $k > k^*$ ψ_k tiene un mínimo finito $x^k \in X_\delta$ que es el mínimo global en todo $\mathbb{R}^n$. Además la sucesión $\{x^k\}$ tiene un punto límite $\bar{x}$ factible.

Demostración:

Sea $x^{\circ} \in X$. Entonces $\Pi_k(x^{\circ}) \leqslant m\, a/s(k)$ siendo $a = \text{máx}_i\ \{a_i = p_i(0)\}$. Por ser X_δ compacto existe $x^k \in X_\delta$ tal que $\psi_k(x^k) \leqslant \psi_k(x)\ \forall\, x \in X_\delta$. En particular:

$$\psi_k(x^k) \leqslant \psi_k(x^{\circ}) \leqslant f(x^{\circ}) + \frac{m\, a}{s(k)} \tag{1}$$

Sea ahora $x \in \mathbb{R}^n - X_\delta$. Por a) existen k°, M° tales que $0 \leqslant f(x) + M^{\circ} p(r(k^{\circ})\, \bar{g}(x))$. Por construcción la sucesión $\{1/2\ s(k)\}$ está acotada por un valor $\bar{s}$. Sea $M > \text{máx}\ \{M^{\circ}, \bar{s}\}$. Por b) existe k^1 tal que para $k > k^1$ $2\, M\, r(k^{\circ}) \leqslant r(k)/s(k)$. Entonces:

$$0 \leqslant f(x) + M^{\circ}\, p(r(k^{\circ})\, \bar{g}(x)) \leqslant f(x) + M\, p(r(k^{\circ})\, \bar{g}(x))$$

$$\leqslant f(x) + \frac{2\, M\, s(k)}{2\, s(k)}\, p(r(k^{\circ})\, \bar{g}(x)) \leqslant$$

$$\leqslant f(x) + \frac{1}{2\, s(k)}\, p(2\, M\, s(k)\, r(k^{\circ})\, \bar{g}(x))$$

$$\leqslant f(x) + \frac{1}{2\, s(k)}\, p(r(k)\, \bar{g}(x)) \qquad (2)$$

en donde hemos empleado la hipótesis d). Por b) existe k^2 tal que para $k > k^2$

$$f(x^{\circ}) + \frac{m\, a}{s(k)} \leqslant \frac{1}{2\, s(k)}\, p(r(k)\, \bar{g}(x)) \qquad (3)$$

Sumando (2) y (3) para $k > k^* = \text{máx}\ \{k^1, k^2\}$ se obtiene junto con (1) que:

$$\psi_k(x^k) \leqslant f(x^{\circ}) + \frac{m\, a}{s(k)} \leqslant f(x) + \frac{1}{s(k)}\, p(r(k)\, \bar{g}(x)) \leqslant$$

$$\leqslant \psi_k(x) \quad \forall\, x \in \mathbb{R}^n - X_\delta$$

lo cual prueba que x^k es el mínimo global de ψ_k. La existencia de $\bar{x}$ se deduce de ser X_δ compacto. Supongamos que $\bar{x}$ no es factible. Sea $\ell \in \{1, \ldots, m\}$ tal que $g_\ell(\bar{x}) = \zeta > 0$. Por la continuidad $g_\ell(x^k) = \zeta^k > 0$ para k suficientemente avanzado. Entonces:

$$\lim_{k \to \infty} \psi_k(x^k) = \lim_{k \to \infty} f(x^k) + \lim_{k \to \infty} \sum_{\substack{i=1 \\ i \neq \ell}}^{m} \frac{1}{s(k)}\, p(r(k)\, g_i(x^k)) \ +$$

$$+ \lim_{k \to \infty} \frac{1}{s(k)}\, p(r(k)\, g_\ell(x^k)) = +\infty$$

en donde hemos empleado la continuidad, la propiedad de p de estar acotada inferiormente por cero y la hipótesis b). Por otra parte si x^1 es factible:

$$\psi_k(x^1) \leqslant f(x^1) + \frac{m\,a}{s(k)} \leqslant f(x^1) + m\,a\,\overline{s} < +\infty$$

Como x^k era el mínimo de ψ_k hemos obtenido una contradicción. Así pues $\overline{x}$ es factible.

1.6. Teorema

En las condiciones del teorema anterior si $s(k) \to +\infty$ entonces $f(\overline{x}) \leqslant f(x^*)$, es decir, $\overline{x}$ es un punto factible óptimo.

Demostración:

Sean $I = \{i \mid g_i(x^*) = 0\}$; $J = \{i \mid g_i(x^*) < 0\}$

$$\lim_{k \to \infty} \frac{1}{s(k)}\, p(r(k)\, g_i(x^*)) = \lim_{k \to \infty} \frac{a}{s(k)} = 0 \quad i \in I \tag{4}$$

$$\lim_{k \to \infty} \frac{1}{s(k)}\, p(r(k)\, g_i(x^*)) = 0 \quad i \in J \tag{5}$$

Por ser p definida no negativa y x^k el mínimo de ψ_k

$$f(x^k) \leqslant \psi_k(x^k) \leqslant \psi_k(x^*)$$

de donde:

$$f(\overline{x}) = \lim_{k \to \infty} f(x^k) \leqslant \lim_{k \to \infty} \psi_k(x^k) \leqslant \lim_{k \to \infty} \psi_k(x^*) = f(x^*)$$

en donde hemos empleado (4) y (5).

1.7. Corolario

En las hipótesis del teorema anterior se verifica:

a) $\lim\limits_{k \to \infty} \psi_k(x^k) = f(\overline{x})$

b) $\lim\limits_{k \to \infty} \dfrac{1}{s(k)} p(r(k)\, g(x^k)) = 0$

2. LOS PROBLEMAS AUXILIARES

Garantizada la existencia de la sucesión de mínimos sin restricciones y su convergencia a un punto factible óptimo es interesante estudiar su situación respecto de la región factible. Las funciones interiores generan siempre mínimos interiores a la región factible, mientras que la aproximación al óptimo que se obtiene con las exteriores es siempre por el exterior del conjunto X. Para las funciones de tipo intermedio no puede afirmarse, en general, que los mínimos van a ser interiores o exteriores exclusivamente sino que dependerá de la estructura del problema a resolver. Sin embargo es posible asegurar que, bajo ciertas condiciones, una vez que se obtiene un mínimo x^k factible ya no se abandona la región de programas, con lo cual si es preciso detener el proceso iterativo antes de alcanzar el óptimo se tiene al menos la seguridad de que la solución es realizable. Esta deseable propiedad, similar a la de funciones interiores, se obtiene sin necesidad de emplear una *primera fase* para la determinación de un punto inicial factible.

2.1. Teorema

Sea:

a) X compacto, Int $X \neq \phi$;

b) f, g y p continuamente diferenciables con $p'(0) \neq 0$;

c) g pseudoconvexa;

d) ψ_k cuasiconvexa para cualquier k;

e) r y s tales que $\forall y < 0 \lim\limits_{k \to \infty} \dfrac{r(k)}{s(k)} p'(r(k)\, y) = 0$, $\lim\limits_{k \to \infty} \dfrac{r(k)}{s(k)} = +\infty$.

Entonces existe k^* tal que para $k > k^*$ x^k pertenece a X.

Demostración:

Sea $x^\circ \in \text{Int } X$ y $\hat{x} \in \text{Fr } X$. Se va a probar que $\nabla \psi_k(\hat{x} - x^\circ) > 0$. Como $\hat{x} \in \text{Fr } X$ existe $\ell \in \{1, ..., m\}$ tal que $g_\ell(\hat{x}) = 0$. Como $x^\circ \in \text{Int } X$ $\overline{g}(x^\circ) < 0$. Sean los conjuntos de índices $I(\hat{x}) = \{i \mid g_i(\hat{x}) = 0\}$; $J(\hat{x}, x^\circ) = \{i \mid \overline{g}(x^\circ)/2 < g_i(\hat{x}) < 0\}$; $K(\hat{x}, x^\circ) = \{i \mid g_i(\hat{x}) \leqslant \overline{g}(x^\circ)/2\}$.

i) Sea $i \in I(\hat{x})$. Como g es pseudoconvexa

$$g_i(x^\circ) \leqslant \overline{g}(x^\circ) < \frac{\overline{g}(x^\circ)}{2} < 0 = g_i(\hat{x}) \Rightarrow \nabla g_i(\hat{x})\, (x^\circ - \hat{x}) < 0$$

Entonces:

$$\lim_{k \to \infty} \frac{r(k)}{s(k)} p'(r(k)\, g_i(\hat{x}))\, \nabla g_i(\hat{x})\, (\hat{x} - x^\circ) =$$

$$= \lim_{k \to \infty} \frac{r(k)}{s(k)} p'(0)\, \nabla g_i(\hat{x})\, (\hat{x} - x^\circ) = -\infty$$

en virtud de las hipótesis b) y c).

ii) Sea $i \in J(\hat{x}, x^\circ)$. Como g es pseudoconvexa

$$g_i(x^\circ) < \frac{\overline{g}(x^\circ)}{2} < g_i(\hat{x}) \Rightarrow \nabla g_i(\hat{x})\, (x^\circ - \hat{x}) < 0$$

De ello se deduce que para todo k

$$\frac{r(k)}{s(k)} p'(r(k)\, g_i(\hat{x}))\, \nabla g_i(\hat{x})\, (x^\circ - \hat{x}) \leqslant 0$$

ya que p es creciente.

iii) Sea $i \in K(\hat{x}, x^\circ)$. Se tiene

$$0 \leqslant \lim_{k \to \infty} \left| \frac{r(k)}{s(k)} p'(r(k)\, g_i(\hat{x}))\ \nabla g_i(\hat{x})\ (x^\circ - \hat{x}) \right|$$

$$\leqslant \lim_{k \to \infty} \frac{r(k)}{s(k)} p'(r(k)\, g_i(\hat{x}))\ \|\nabla g_i(\hat{x})\|\ \|x^\circ - \hat{x}\| \qquad \text{(por la desigualdad de Schwarz)}$$

$$\leqslant \lim_{k \to \infty} \frac{r(k)}{s(k)} p'(r(k)\, g_i(\hat{x}))\, M_i \qquad \text{(por las hipótesis a) y b))}$$

$$= 0 \qquad \text{(por la hipótesis e))}$$

Ello quiere decir que para cualquier $\epsilon > 0$ existe k' tal que para $k > k'$

$$\left| \frac{r(k)}{s(k)} p'(r(k)\, g_i(\hat{x}))\ \nabla g_i(\hat{x})\ (x^\circ - \hat{x}) \right| < \epsilon$$

iv) Por las hipótesis a) y b) existe M tal que:

$$|\nabla f(\hat{x})\ (x^\circ - \hat{x})| \leqslant M$$

Reuniendo i), ii), iii) y iv) llegamos a que para $k > k'$

$$\nabla \psi_k(\hat{x})\,(x^\circ - \hat{x}) = \nabla f(\hat{x})\,(x^\circ - \hat{x}) + \sum_{i=1}^{m} \frac{r(k)}{s(k)} p'(r(k)\, g_i(\hat{x}))\ \nabla g_i(\hat{x})\,(x^\circ - \hat{x})$$

$$\leqslant M + \sum_{i \in I(\hat{x})} \frac{r(k)}{s(k)} p'(0)\ \nabla g_i(\hat{x})\ (x^\circ - \hat{x}) + \sum_{i \in K(\hat{x},\, x^\circ)} \epsilon$$

y dado que el segundo miembro tiende a $-\infty$ existe un k^* independiente de $\hat{x}$ en virtud de a) y b) tal que para $k > k^* \geqslant k'$

$$\nabla \psi_k(\hat{x})\,(x^\circ - \hat{x}) < 0 \Leftrightarrow \nabla \psi_k(\hat{x})\,(\hat{x} - x^\circ) > 0$$

Sea ahora x tal que $(x - \hat{x}) = \lambda\,(\hat{x} - x^\circ)$ con $\lambda > 1$. Entonces:

$$0 < \nabla \psi_k(\hat{x})\,(\hat{x} - x^\circ) = \nabla \psi_k(\hat{x})\,\frac{1}{\lambda}\,(x - \hat{x})$$

y por la cuasiconvexidad de ψ_k

$$\psi_k(\hat{x}) \leqslant \psi_k(x)$$

Finalmente para cualquier x fuera de la región factible existe un $\hat{x} \in \operatorname{Fr} X$ que está en la recta que une x° con x. Como k^* no depende de $\hat{x}$ la desigualdad anterior se sigue verificando. Entonces ψ_k alcanza en X su mínimo global.

Para mínimos no factibles se tiene el siguiente resultado:

2.2. Teorema

Sea x^k la solución del PAk. Sea $t \in \mathbb{R}^m$ un vector de perturbación verificando: a) $g(x^k) \leqslant t$; b) $p(r(k)\; g(x^k)) = p(r(k)\; t)$. Entonces x^k es el mínimo exacto de la función f en el conjunto $X_t = \{x \mid x \in X^\circ, g(x) \leqslant t\}$.

Demostración:

Sea $x \in X_t$. Como p es creciente $p(r(k)\,g(x)) \leqslant p(r(k)\; t)$. Por ser x^k la solución del PAk para todo $x \in \mathbb{R}^n$

$$f(x^k) + \frac{1}{s(k)}\,p(r(k)\,g(x^k)) \leqslant f(x) + \frac{1}{s(k)}\,p(r(k)\,g(x)) \leqslant$$

$$\leqslant f(x) + \frac{1}{s(k)}\,p\;(r(k)\;t)$$

pero entonces de b) se sigue que $f(x^k) \leqslant f(x)$ para todo $x \in X_t$.

También se obtiene una cierta monotonía para la sucesión $\{\psi_k(x^k)\}$.

2.3. Teorema

Si x^{k^1} es interior a X entonces para cualquier $k \geqslant k^1$ se verifica $\psi_k(x^k) \leqslant \psi_{k^1}(x^{k^1})$.

Demostración:

Sea $x^{k^1} \in \text{Int } X$. Esto junto con la definición 1.2 nos lleva a

$$\psi_{k^1}(x^{k^1}) = f(x^{k^1}) + \frac{1}{s(k^1)} p(r(k^1)\, g(x^{k^1})) \geqslant$$

$$\geqslant f(x^{k^1}) + \frac{1}{s(k)} p(r(k)\, g(x^{k^1})) \geqslant$$

$$\geqslant f(x^k) + \frac{1}{s(k)} p(r(k)\, g(x^k)) = \psi_k(x^k)$$

en donde la última desigualdad se sigue de la definición de x^k.

Nótese que como corolario de este teorema se deduce el conocido resultado relativo a las funciones interiores de que $\{\psi_k(x^k)\}$ es monótona decreciente, Fiacco-McCormick 68, Polak 70.

Aunque los métodos de penalización constituyen un procedimiento algorítmico teóricamente infinito el proceso iterativo ha de detenerse necesariamente al alcanzar el grado de precisión deseado. Como éste puede fijarse de antemano, si se conoce para que valor del parámetro se alcanza, tan sólo es preciso realizar la iteración correspondiente a dicho valor, ya que, como se desprende del método el conocimiento de los mínimos anteriores no es esencial para la solución del PAk. Se va a dar a continuaolón una cota de la distancia al valor óptimo en cada iteración.

2.4. Teorema

Sea x^k la solución del PAk. Entonces $f(x^k) - f(x^*) \leqslant \dfrac{m\,a}{s(k)}$

Demostración:

Dado que x^* es factible y p creciente y definida no negativa:

$$p(r(k)\,g_i(x^*)) - p(r(k)\,g_i(x^k)) \leqslant a \qquad i = 1, \ldots, m$$

y como x^k es el mínimo de ψ_k resulta:

$$f(x^k) - f(x^*) \leqslant \sum_{i=1}^{m} \frac{1}{s(k)} \{p(r(k)\,g_i(x^*)) -$$

$$- p(r(k)\,g_i(x^k))\} \leqslant \frac{m\,a}{s(k)}$$

2.5. Corolario

a) Si x^k es factible entonces $0 \leqslant f(x^k) - f(x^*) \leqslant \dfrac{m\,a}{s(k)}$

b) Si $a = 0$ (penalidad exterior) el primer punto factible es óptimo.

Si se conocen algunas propiedades de las funciones del problema pueden mejorarse las cotas anteriores. En particular si la función ψ_k es uniformemente cuasiconvexa (Roberts 73) y X° convexo se tienen las cotas:

$$f(x^k) - f(x^*) \leqslant \frac{m\,a}{s(k)} - \delta\,(\|x^k - x^*\|)$$

en donde δ es la función que da la cuasiconvexidad uniforme. Para puntos factibles con δ estrictamente creciente se tiene:

$$\|x^k - x^*\| \leqslant \delta^{-1}\,\frac{m\,a}{s(k)}$$

Los algoritmos de penalización que estudiamos generan simultáneamente con los x^k una sucesión de vectores de multiplicadores de Kuhn-Tucker definidos por:

$$u_i^k = \frac{r(k)}{s(k)} p'(r(k)\, g_i(x^k))$$

Es posible demostrar que en condiciones bastante generales, similares a las utilizadas en la teoría de la programación matemática (Bazaraa-Shetty 79), la sucesión anterior converge a un punto $\bar{u}$ que forma con $\bar{x}$ un punto de silla o un punto estacionario de Kuhn-Tucker según los casos. Junto con el interés teórico de la existencia de los multiplicadores, la consideración de los mismos permite deducir varias cotas de convergencia para el caso en que no puedan obtenerse los mínimos x^k en forma exacta. (Mifflin 75).

3. EL LAGRANGIANO AUMENTADO. METODOS DE LOS MULTIPLICADORES

Los métodos de penalización presentan un comportamiento bastante satisfactorio para la resolución numérica de un problema de programación matemática. Sin embargo no están exentos de ciertos inconvenientes que hacen desaconsejar su empleo en algunos casos especialmente debido a problemas de inestabilidades numéricas. De ahí la necesidad de introducir en ellos alguna modificación que evite estas características negativas. La modificación propuesta convierte a la función auxiliar del método de las penalizaciones en una función del tipo Lagrangiano aumentado al estilo de las propuestas por Roode 69, Arrow-Gould-Howe 73, Mangasarian 75, los cuales presentan una caracterización axiomática de una serie de propiedades que permiten relacionar el lagrangiano aumentado con el clásico, mientras que aquí se obtienen resultados similares partiendo de un contexto de penalizaciones. Distintas formulaciones del lagrangiano aumentado pueden encontrarse en Pierre-Lowe 75.

3.1. Definición

Se llama *lagrangiano aumentado asociado al PM* a la función:

$$L(x, \sigma, r(k), s(k)) = f(x) + \sum_{i=1}^{m} \frac{q_i(\sigma_i)}{s_i(k)} \{p_i(r_i(k)\, g_i(x)) - p_i(0)\}$$

$$= f(x) + P(x, \sigma, r(k), s(k))$$

en donde cada q_i es una función real de variable real dos veces diferenciable continuamente verificando:

$$\text{i) } q(0) = 0; \qquad \text{ii) } q'(0) = 0;\ q''(0) > 0$$

Inmediatamente se deduce que en un entorno de cero E_0 q' es estrictamente creciente, q es estrictamente convexa y en $\sigma = 0$ q tiene un mínimo relativo estricto por lo que se conservará no negativa en E_0. Como en el caso de la función penalidad se suprimirá normalmente el subíndice i.

El lagrangiano aumentado así definido posee propiedades similares a las del clásico de la teoría de Kuhn-Tucker como se ve a continuación.

3.2. Definición

Dado un índice k se llama *punto estacionario del lagrangiano aumentado* (LAP) a un par $(\bar{x}, \bar{\sigma})$ que verifique:

$$\nabla L(\bar{x}, \bar{\sigma}, r(k), s(k)) = 0$$

3.3. Teorema

Sean las funciones f, g y p diferenciables continuamente con $p'(0) > 0$ y $\lim\limits_{k \to \infty} \frac{r(k)}{s(k)} = +\infty$. a) Si $(\bar{x}, \bar{u})$ es un punto estacionario de Kuhn-

Tucker (KTP) entonces para k suficientemente avanzado existe $\overline{\sigma} \in \mathbb{R}^m$ tal que $(\overline{x}, \overline{\sigma})$ forman un LAP. b) Si $(\overline{x}, \overline{\sigma})$ es un LAP para algún k tal que $\overline{\sigma} \in B_\delta(0)$ para δ suficientemente pequeño y $g(\overline{x}) \leqslant 0$ entonces existe $\overline{u}$ tal que $(\overline{x}, \overline{u})$ forman un KTP.

Demostración:

a) De las propiedades de q se puede deducir que existe un intervalo $(0, \xi)$ tal que q es estrictamente creciente en él por lo que puede definirse q^{-1} en $(0, q(\xi))$ tal que para $\eta \in (0, q(\xi))$ sea $q(q^{-1}(\eta)) = \eta$. Sea $(\overline{x}, \overline{u})$ el KTP y denotemos $K = \{i \mid \overline{u}_i = 0\}$; $I = \{i \mid g_i(\overline{x}) = 0\}$; $J = \{i \mid g_i(\overline{x}) < 0\}$. Para $i \notin K$ debido a las hipótesis sobre los parámetros existirá un k suficientemente avanzado tal que:

$$0 < \frac{s(k)}{r(k)} \frac{\overline{u}_i}{p'(0)} < q(\xi)$$

Entonces definimos:

$$\overline{\sigma}_i = q^{-1}\left(\frac{s(k)}{r(k)} \frac{\overline{u}_i}{p'(0)}\right) \qquad i \notin K$$

$$\overline{\sigma}_i = 0 \qquad\qquad i \in K$$

Con estos valores se puede comprobar que para k suficientemente avanzado:

$$\nabla L(\overline{x}, \overline{\sigma}, r(k), s(k)) = 0$$

es decir, $(\overline{x}, \overline{\sigma})$ forman un LAP.

b) Por hipótesis:

$$\frac{\partial L}{\partial \sigma_i} = \frac{q'(\overline{\sigma}_i)}{s(k)} \{p(r(k)\, g_i(\overline{x})) - p(0)\} = 0$$

lo cual por las condiciones impuestas a q y σ es equivalente a:

$$\overline{\sigma}_i = 0 \quad \text{ó} \quad \overline{g}_i(x) = 0 \quad i = 1, \ldots, m$$

Definimos:

$$\overline{u}_i = q(\overline{\sigma}_i)\,\frac{r(k)}{s(k)}\,p'(r(k)\,g_i(\overline{x})) \quad i = 1, \ldots, m$$

Inmediatamente resulta que $\overline{u}_i \geqslant 0$, $i = 1, \ldots, m$. También es fácil ver que:

$$\overline{u}_i\, g_i(\overline{x}) = 0 \quad i = 1, \ldots, m$$

Finalmente:

$$0 = \nabla_x L(\overline{x}, \overline{\sigma}, r(k), s(k)) = \nabla f(\overline{x}) + \sum_{i=1}^{m} \overline{u}_i\, \nabla g_i(\overline{x})$$

lo cual prueba que $(\overline{x}, \overline{u})$ es un KTP.

3.4. Corolario

Si $(\overline{x}, \overline{\sigma})$ es un LAP entonces:

a) $\displaystyle\sum_{i=1}^{m} \frac{q(\overline{\sigma}_i)}{s(k)}\,\{p(r(k)\,g_i(\overline{x})) - p(0)\} = 0$

b) $L(\overline{x}, \overline{\sigma}, r(k), s(k)) = f(\overline{x})$

3.5. Lema

(Arrow-Gould-Howe 73). Sea A una matriz $m \times n$ y B una matriz $n \times n$. Se tiene:

$$\{x \neq 0, A\,x = 0 \Rightarrow x\,B\,x > 0\} \Leftrightarrow \left\{\begin{array}{l} B + \alpha\, {}^t\!A\, A \\ \text{es definida positiva para } \alpha \\ \text{suficientemente grande} \end{array}\right.$$

Basándose en el lema anterior puede demostrarse la equivalencia entre la solución óptima del PM y los mínimos sin restricciones del lagrangiano aumentado para un valor *finito* del parámetro.

3.6. Teorema

Sean f, g y p dos veces diferenciables continuamente, $p'(0) > 0$, $p''(0) \geqslant 0$ y $\lim\limits_{k \to \infty} \dfrac{r(k)}{s(k)} = +\infty$. a) Si $(\overline{x}, \overline{u})$ satisfacen las condiciones suficientes de segundo orden para mínimo local del PM (Fiacco-McCormick 68) y si $\{i \mid g_i(\overline{x}) = 0, \overline{u}_i = 0\} = \phi$ entonces para k suficientemente avanzado existe un $\overline{\sigma}$ tal que $\overline{x}$ satisface las condiciones suficientes de segundo orden para mínimo local sin restricciones del lagrangiano aumentado $L(x, \overline{\sigma}, r(k), s(k))$ $x \in E_0(\overline{x})$. b) Recíprocamente si $(\overline{x}, \overline{\sigma})$ forman un LAP para algún k con $\overline{\sigma} \in B_\delta(0)$, $g(\overline{x}) \leqslant 0$ y $\overline{x}$ satisface las condiciones suficientes de segundo orden para mínimo local sin restricciones de $L(x, \overline{\sigma}, r(k), s(k))$ entonces existe un $\overline{u}$ tal que $(\overline{x}, \overline{u})$ satisfacen las condiciones suficientes de segundo orden para mínimo local del PM.

Demostración:

a) Por el teorema 3.3 existe $\overline{\sigma}$ tal que $\nabla_x L(\overline{x}, \overline{\sigma}, r(k), s(k)) = 0$. Además $\nabla^2_{xx} L = \nabla^2_{xx} L_0(\overline{x}, \overline{u}) + T(\overline{x}, \overline{u}')$ en donde se ha denotado:

$$L_0(x, u) = f(x) + \sum_{i=1}^{m} u_i \, g_i(x)$$

$$T(x, u') = \sum_{i=1}^{m} q(\sigma_i) \frac{r(k)}{s(k)} p''(r(k)\, g_i(x)) \, \nabla g_i(x)^t \nabla g_i(x)$$

Entonces:

$$T(\overline{x}, \overline{u}') = \sum_{i \in I} \frac{p''(0)}{p'(0)} r(k) \, \overline{u}_i \, \nabla g_i(\overline{x})^t \nabla g_i(\overline{x})$$

Si hacemos ahora la identificación:

$$A = {}^t\nabla g_I(\bar{x}) \qquad B = \nabla^2_{xx} L_0(\bar{x}, \bar{u}) \qquad \alpha = \frac{p''(0)}{p'(0)}\, r(k)\, \bar{u}_I$$

al aplicar el lema 3.5 obtenemos:

$$\{y \neq 0,\ {}^t\nabla g_I(\bar{x})\ y = 0 \ \Rightarrow\ y\ \nabla^2_{xx} L_0(\bar{x}, \bar{u})\ y > 0\} \Leftrightarrow$$

$$\Leftrightarrow \{y(\nabla^2_{xx} L_0(\bar{x}, \bar{u}) + \alpha\, \nabla g_I(\bar{x})\ {}^t\nabla g_I(x))\ y > 0\}$$

es decir:

$$y\ \nabla^2_{xx} L(\bar{x}, \bar{\sigma}, r(k), s(k))\, y > 0 \qquad \forall\, y \neq 0, y \in \mathbb{R}^n$$

Se llega entonces a las condiciones suficientes de mínimo local.

b) Similar.

El lagrangiano aumentado que se está considerando permite definir un problema dual asociado al PM. Las relaciones existentes entre ambos son similares a las que se obtienen en la clásica teoría de la dualidad originada a partir del lagrangiano clásico.

3.7. Problema dual (PD)

Hallar, si existe, $\bar{\sigma} \in \mathbb{R}^m$ tal que:

$$D(\bar{\sigma}) = \text{máx}\ D(\sigma) \quad \bar{\sigma} \in Y = \mathrm{E}(\bar{\sigma}) \cap B_\delta(0)$$

siendo $D(\sigma) = L(x(\sigma), \sigma, r(k), s(k)) = \min\limits_{x \in E(x(\sigma))} L(x, \sigma, r(k), s(k))$

3.8. Teorema

Sean f, g y p dos veces diferenciables continuamente, $p'(0) > 0$

$p''(0) \geqslant 0$ y $\lim\limits_{k \to \infty} \dfrac{r(k)}{s(k)} = +\infty$. Sea $\bar{x}$ la solución del PM, $(\bar{x}, \bar{u})$ verificando las condiciones suficientes de segundo orden para mínimo local restringido y $\{i \mid g_i(\bar{x}) = 0,\ \bar{u}_i = 0\} = \phi$ y sean los gradientes de las restricciones activas linealmente independientes. Entonces para k suficientemente avanzado existe $\bar{\sigma} \in \mathbb{R}^m$ tal que $\bar{\sigma}$ resuelve el PD y además $f(\bar{x}) = D(\bar{\sigma})$.

Demostración:

Por los teoremas 3.3 y 3.6 y el corolario 3.4 se llega a que:

$$f(\bar{x}) = L(\bar{x}, \bar{\sigma}, r(k), s(k)) = \min_{x \in E(\bar{x})} L(x, \bar{\sigma}, r(k), s(k)) = D(\bar{\sigma})$$

Nótese en primer lugar que $\nabla D(\bar{\sigma}) = 0$. Además:

$$\nabla^2 D(\bar{\sigma}) = \nabla_{\sigma\sigma} L - \nabla_{\sigma x} L\ \nabla_{xx} L^{-1}\ \nabla_{x\sigma} L =$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & w_J' \end{bmatrix} - \begin{bmatrix} v_I\ {}^t\nabla g_I \\ 0 \end{bmatrix} (\nabla_{xx} L)^{-1}\ (v_I\ \nabla g_I\ 0)$$

$$\begin{bmatrix} -v_I\ {}^t\nabla g_I\ \nabla_{xx} L^{-1}\ v_I \nabla g_I & 0 \\ 0 & w_J' \end{bmatrix}_{(\bar{x}, \bar{\sigma})}$$

en donde hemos denotado:

$$v_i = q'(\sigma_i)\, \frac{r(k)}{s(k)}\, p'(r(k)\, g_i(x))\ {}^t\nabla g_i(x)$$

$$w_i' = \frac{q''(\sigma_i)}{s(k)}\, \{p(r(k)\, g_i(x)) - p(0)\}$$

Al ser $\nabla_{xx} L$ definida positiva y por la independencia de los gradientes activos la matriz $(-v_I\ {}^t\nabla g_I\ \nabla_{xx} L^{-1}\ v_I\ \nabla g_I)$ es definida negativa. También $w_i' < 0$ para $i \in J$. Por tanto $\nabla^2 D(\bar{\sigma})$ es definida negativa lo cual prueba que $\bar{\sigma}$ es un máximo local, es decir, soluciona el PD.

La teoría que acabamos de exponer permite la consideración de nuevas familias de algoritmos de los que vamos a describir dos tipos fundamentales. Algoritmos similares han sido inicialmente propuestos por Hestenes 69, Powell 69 y más tarde estudiados por Rockafellar 73, Bertsekas 76 y Kort-Bertsekas 76.

3.9. Método de Newton-Raphson

Calcular x^* tal que:

$$f(x^*) = \text{mín}\, f(x) \quad x \in \{x \mid g(x) \leqslant 0,\ \nabla L(x, \sigma, r(k), s(k)) = 0\}$$

En definitiva se trata de encontrar los LAP que son los candidatos naturales a mínimo. De este esquema general pueden deducirse varios procedimientos según la técnica empleada para resolver el sistema $\nabla L =$ $= 0$. Si se emplea el método de Newton-Raphson se obtienen para x y σ las fórmulas iterativas

$$\sigma' = \sigma + (M - {}^tGH^{-1}\, G)^{-1}\, ({}^tGH^{-1}\, \nabla L_0 - N)$$

$$x' = x + H^{-1}\, G(M - {}^tGH^{-1}\, G)^{-1}\, N -$$
$$- \{H^{-1}\, G(M - {}^tGH^{-1}\, G)^{-1}\, {}^tGH^{-1} + H^{-1}\}\, \nabla L_0$$

La convergencia del método y su razón de convergencia se deducen de las propiedades del algoritmo de Newton-Raphson (Ortega-Reinhbolt 70).

3.10. Método de los multiplicadores

1) Definir L y seleccionar k°, σ° y hacer $j = 0$.
2) Determinar x^{j+1} tal que $L(x^{j+1}, \sigma^j, r(k^j), s(k^j)) = \text{mín}_x\, L(x, \sigma^j, r(k^j), s(k^j))$
3) Determinar σ^{j+1} tal que $L(x^{j+1}, \sigma^{j+1}, r(k^j), s(k^j)) = \text{máx}_\sigma\, L(x^j, \sigma, r(k^j), s(k^j))$

4) Comprobar las condiciones de convergencia. Ir a STOP o a paso 5.

5) Comprobar las condiciones de modificación de parámetros. Ir a 1 con $j = j + 1$.

Se obtienen distintas variantes del método según las técnicas empleadas en los procesos sin restricciones. Por ejemplo, puede utilizarse $\sigma^{j+1} = \sigma^j + \lambda \nabla D(\sigma^j)$ en cuyo caso se llega a la convergencia y razón de convergencia como consecuencia del teorema del punto de atracción de Ostrowski (Ortega-Reinhbolt 70).

REFERENCIAS

Allran, R.R. y Johnsen, S.E.J. (1970): *An Algorithm for Solving Nonlinear Programming Problems Subject to Nonlinear Inequality Constraints.* Comput. I. vol. 13, núm. 2, pp. 171-177.

Arrow, K.J.; Gould, F.J.; Howe, S.M. (1973): *A General Saddle Point Result for Constrained Optimization.* Mathematical Programming vol. 5, pp. 225-234.

Bazaraa, M.S. y Shetty, C.M. (1979): *Nonlinear Programming. Theory and Algoritms.* John Wiley & Sons.

Bertsekas, D. (1976): *On Penalty and Multiplier Methods for Constrained Maximization.* SIAM J. Control and Optimization vol. 14, núm. 2, pp. 216-235.

Fiacco, A.V. y McCormick, G.P. (1968): *Nonlinear Programming: S.U.M.T.*, John Wiley & Sons.

Hestenes, M.R. (1969): *Multiplier and Gradient Methods.* J. Optimization Theory Appl. vol. 4, núm. 5, pp. 303-320.

Kort, B.W. y Bertsekas, D. (1976): *Combined Primal-Dual and Penalty Methods for Convex Programming.* SIAM J. Control and Optimization vol. 14, núm. 2, pp. 268-294.

Mangasarian, O. (1975): *Unconstrained Lagrangians in Nonlinear Programming.* SIAM J. Control vol. 13, núm. 4, pp. 772-791.

Mifflin, R. (1975): *Convergence Bounds for Nonlinear Programming Algorithms.* Mathematical Programming vol. 8, pp. 251-271.

Murphy, F.H. (1974): *A Class of Exponential Penalty Functions.* SIAM J. Control vol. 13, núm. 1, pp. 679-687.

Ortega, J.M. y Reinhbolt, W.C. (1970): *Iterative Solutions of Nonlinear Equations in Several Variables.* Academic Press.

Pierre, D.A. y Lowe, M.J. (1975): *Nonlinear Programming Via Augmented Lagrangians.* Adisson-Wesley.

Polak, E. (1970): *Computational Methods on Optimization.* Academic Press.

Powell, M.J.D. (1969): *A Method for Nonlinear Constraints in Minimization Problems.* En FLETCHER: *Optimization.* Academic Press, Cap. 19, pp. 283-298.

Roberts, A.W. y Varberg, D.E. (1973): *Convex Functions.* Academic Press.

Rockafellar, R.T. (1973): *A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization.* Mathematical Programming vol. 5, pp. 354-373.