CUBO 11, 81-93 (1995)
Novena Jornada de Matemática de la Zona Sur.

Un algoritmo de puntos interiores con fase-I integrada.

Constande Nicolas B.

Resumen.
En este trabajo se presenta una generalización de cierto tipo de algoritmos Cuasi-Newton de puntos interiores (único en su género), para programación no lineal, desarrollada por Herskovits [3]. Esta familia de algoritmos no requiere de un punto inicial viable. Ellos combinan automáticamente las operaciones de inicialización (fase-I) y de optimización (fase-II). A diferencia de los métodos de direcciones viables desarrollados por Polak [1] con combinación fase-I, fase-II cuya convergencia es de tipo lineal. Estos algoritmos presentan la ventaja de tener convergencia superlineal.

1 Introducción

Los métodos Cuasi-Newton utilizados para resolver el problema de Programación no Lineal (P):

$$(P) \qquad \mathrm{Min}\{f(x)/x \in X\}, \qquad X = \{x \in \mathbb{R}^n / g_i(x) \leq 0, i = 1, \cdots, m\} \quad (1)$$

se fundamentan en encontrar la dirección de búsqueda resolviendo el sistema de ecuaciones no lineales, resultante de aplicar las condiciones de optimalidad de K.K.T. al problema (P), es decir

$$\nabla f(x) + \nabla g(x)\lambda = 0, \qquad g(x)\lambda = 0, \qquad g(x) \leq 0 \quad ; \quad \lambda \geq 0; \qquad g, \lambda \in \mathbb{R}^m \quad (2)$$

y luego resolver el sistema (2) aplicando el método de Newton, con la salvedad que la matriz Hessiana es aproximada por una matriz simétrica definida positiva. Sea

$\mathcal{L}(x,\lambda)$ la función de Lagrange de (1), o sea,

$$\mathcal{L}(x,\lambda) = f(x) + g(x)\lambda$$

El método de Newton linealiza el sistema (2) en una vecindad del punto (x_k, λ_k) en la iteración "k" resultando el sistema:

$$\begin{bmatrix} [B] & \nabla g(x) \\ \Lambda[\nabla g^t(x)] & diag[g(x)] \end{bmatrix} \begin{bmatrix} x_0 - x \\ \lambda_0 - \lambda \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + \nabla g(x)\lambda \\ diag[g(x)\lambda] \end{bmatrix} \tag{3}$$

donde $\Lambda = diag[\lambda]$ y B es una matriz simétrica definida positiva que aproxima a $\nabla_{xx}\mathcal{L}(x,\lambda)$, (x,λ) representa el punto actual y (x_0, λ_0) es la nueva estimación.

Diferentes métodos se encuentran en la literatura usando esta idea, en particular los algoritmos debidos a Herskovits (Cuasi-Newton de puntos Interiores) que se caracterizan por iterar en el interior de la región viable X. La forma en que definen la dirección de búsqueda estos algoritmos es a través de una combinación lineal de la dirección ($d_0 = x_0 - x$) obtenida del sistema lineal (4) con una nueva dirección de deflexión, resultante del sistema homogéneo perturbado de (4) y representado por el sistema (5), es decir

$$Bd_0 + \nabla g(x)\lambda_0 = -\nabla f(x) \tag{4}$$

$$\nabla g^t(x)d_0 + G(x)\lambda_0 = 0$$

$$Bd_1 + \nabla g(x)\lambda_1 = 0 \tag{5}$$

$$\Lambda\nabla g^t(x)d_1 + G(x)\lambda_1 = -\Lambda\omega; \quad \omega > 0, \omega \in \mathbb{R}^m$$

donde $G(x) = diag[\lambda]$.

La dirección en el espacio primal es $d = d_0 + \rho d_1$ y en el espacio dual $\lambda = \lambda_0 + \rho\lambda_1$, $\rho \in \mathbb{R}^+$

En diseño de ingeniería son comunes los problemas de programación no lineal que presentan la siguiente estructura: (OSI-P)

$$\text{Min}\{f(x)/x \in X\},$$

$$X = \{x \in \mathbb{R}^n / g_i(x) \leq 0, i = 1,\cdots,m; \phi_j(x,u) \leq 0; \forall u \in U^j, j = 1,\cdots,r\}.$$

$$f, g_i : \mathbb{R}^n \to \mathbb{R}; \phi_j : \mathbb{R}^n \times U^j \to \mathbb{R}; U^j \subseteq \mathbb{R}^{p_j}, p_j \in \mathbb{N}, j = 1,\cdots,r, i = 1,\cdots,m. \tag{6}$$

Los conjuntos U^j son de cardinalidad infinita y compactos. Si $U^j = \phi$ $\forall j = 1,\cdots,r$ el problema (OSI-P) es de finitas restricciones $(m+r)$, en caso contrario es un problema de infinitas restricciones o problema Semi-Infinito como es conocido en la literatura.

En general el problema (OSI-P) se resuelve por discretizaciones sucesivas de los conjuntos U^j lo cual da origen a una sucesión de problemas $(OSI - P_k)$ de finitas restricciones $\{P_k\}_{k\in\mathbb{N}}$, donde

$$\text{Min}\{f(x)/x \in X\}$$

$$X_k = \{x \in \mathbb{R}^n/g_i(x) \leq 0, i = 1, \cdots, m; \phi_j(x,u) \leq 0; \forall u \in U_k^j, j = 1, \cdots, r.\} \quad (7)$$

con $|U_k^j| < \infty, \quad \forall j = 1, \cdots, r$

2 Preliminares

En la resolución de problemas de Optimización Semi-Infinita, los algoritmos de puntos interiores que resuelven el problema reducido discreto $OSI - P_k)$ requieren de forma continua de un método Fase-I que permita en cada refinamiento de la discretización asegurar el punto de iniciación en el interior de la región viable X_k.

De la literatura se conoce un procedimiento general para hallar un punto viable, reflejado a través de un problema de Programación Matemática. Por simplicidad consideremos el siguiente problema de Programación no Lineal de finitas restricciones, su aplicación al problema $(OSI - P_k)$ resulta inmediata:

$$\text{Min}\{f(x)/g_i(x) \leq 0, \quad i = 1, \cdots, m\} \quad (1)$$

donde $g_i, f : \mathbb{R}^n \to \mathbb{R}, \quad i = 1, \cdots, m$ son funciones continuas diferenciables. Un punto inicial viable para (1) se obtiene resolviendo el problema

$$\text{Min}\{\varphi(x)/x \in \mathbb{R}^n\} \quad (2)$$

donde $\varphi : \mathbb{R}^n \to \mathbb{R}$, está definida por

$$\varphi(x) = \text{Max}\{g_i(x)/ \quad i = 1, \cdots, m\}$$

El problema (2) es equivalente al problema

$$\text{Min}\{w/g_i(x) - w \leq 0, \quad i = 1, \cdots, m \quad , \quad w \in \mathbb{R}\} \quad (3)$$

Para el problema (3) se encuentra un punto inicial viable $z = (x, w_0^0) \in \mathbb{R}^{n+1}$ con $x \in \mathbb{R}^n$ arbitrario y $w_0^0 = max\{g_i(x)/i = 1, \cdots, m\}$. La señal de parada para el problema (3) está condicionada por $w \leq 0$ o alguna otra condición que asegure que en la iteración "k" se genera un punto viable, identificado por (x_k, w_k^0) donde x_k es el punto deseado.

Desgraciadamente el método señalado genera puntos viables sin considerar la función objetivo del problema original, el cual aplicado reiterativamente, como ocurre en los problemas de Optimización Semi infinita, hace que la evolución hacia un punto óptimo local sea degradada por los continuos retrocesos en su descenso. En virtud de esto último, en este capítulo presentamos un método que construye puntos viables considerando la función objetivo. La idea de este método se basa en plantear un problema de Programación no Lineal similar al problema (3) en que la función objetivo del problema (1) es considerada, es decir

$$\text{Min}\left\{f(x) + \frac{1}{2}\gamma w^2 / x \in \mathbb{R}^n, \quad w \in \mathbb{R}\right\} \tag{4}$$

s.a

$$g_i(x) - \gamma w \leq 0, \quad i = 1, \cdots, m$$

donde "γ" es un parámetro Booleano o simplemente un conmutador, $\gamma \in \{0, 1\}$ cuyo objetivo es controlar el sistema completo dependiendo si el punto "x" pertenece o no al conjunto viable. La definición de "γ" es la siguiente

$$\gamma = \begin{cases} 0 & \text{si Max}\{g_i(x)/i = 1, \cdots, m\} \leq 0 \\ 1 & \text{De otra manera} \end{cases}$$

Denotemos por $X \subseteq \mathbb{R}^n$ el conjunto viable para el problema (1), es decir

$$X = \{x \in \mathbb{R}^n / g_i(x) \leq 0, \quad i = 1, \cdots, m\}$$

y denominemos por intX el interior de X, y X^c el complemento de X. Observemos que si $x \in X^c$ en el problema (5) $\gamma = 1$. En este caso la variable $w > 0$ decrecerá hasta que se verifique la condición de parada ($w \leq 0 \Rightarrow \gamma = 0$). Si $x \in X$ entonces el problema (5) es equivalente al problema (1) y en este caso una apropiada dirección "d" generada por el Algoritmo (Quasi Newton de Puntos Interiores) se obtiene de la resolución de los sistemas (6), (7)

$$i) Bd_0 + \nabla g(x)\lambda_0 = -\nabla f(x) \quad ii) \Lambda \nabla g^t(x) d_0 + G(x)\lambda_0 = 0 \tag{6}$$

$$iii) Bd_1 + \nabla g(x)\lambda_1 = -\nabla f(x) \quad iv) \Lambda \nabla g^t(x) d_1 + G(x)\lambda_1 = -\Lambda\omega; \omega > 0, \omega \in \mathbb{R}^m \tag{7}$$

de donde $d = d_0 + \rho d_1; \quad \lambda = \lambda_0 + \rho\lambda_1 \quad \rho \in \mathbb{R}^+$

Con la intención de obtener una dirección de búsqueda de la misma naturaleza que la antes citada cuando $x \in X^c$ y con la misma razón de convergencia, resulta obvio aplicar el Algoritmo al problema (5).

El Lagrangiano queda expresado por

$$L(x,\lambda,\gamma) = f(x) + \frac{1}{2}\gamma w^2 + \lambda^t[g(x) - \gamma we]$$

donde $e^t = (1,\cdots,1)$, $e \in \mathbb{R}^m$.

Las condiciones de optimalidad de primer orden de K.K.T. se expresan por

$$\begin{aligned}
&a)\ \nabla f(x) + \nabla g(x)\lambda = 0\\
&b)\ \gamma(w - \lambda^t e) = 0\\
&c)\ diag[g(x) - \gamma we]\lambda = 0\\
&d)\ g(x) - \gamma we \leq 0\\
&e)\ \lambda \geq 0, \quad \lambda \in \mathbb{R}^m
\end{aligned} \tag{8}$$

Nótese que cuando $\gamma = 0$, (8) representa las condiciones de optimalidad del problema (1) y $L(x,\lambda,\gamma)$ su respectivo Lagrangiano.

Aplicando el método de Newton al sistema de ecuaciones (a, b, c) en (8) y aproximando su Hessiano $H(x,\lambda,\gamma)$ dado por

$$H(x,\lambda,\gamma) = \begin{bmatrix} \nabla^2 f(x) \\ \gamma \end{bmatrix} + [\nabla^2 g(x)\lambda, 0]$$

$H(x,\lambda,\gamma) \approx B \in \mathbb{R}^{(n+1)*(n+1)}$, B matriz simétrica definida positiva (BFGS), obtenemos el siguiente sistema lineal

$$\begin{bmatrix} [B] & \begin{bmatrix}\nabla g(x) \\ -\gamma\end{bmatrix} \\ \Lambda[\nabla g^t(x) - \gamma e] & [(g(x) - \gamma we] \end{bmatrix} \begin{bmatrix} \begin{bmatrix} x_0 - x \\ \vdots \\ w_0 - w \end{bmatrix} \\ \lambda_0 - \lambda \end{bmatrix} = - \begin{bmatrix} \begin{vmatrix} \nabla f(x) + \nabla g(x)\lambda \\ \gamma w - \lambda^t e \end{vmatrix} \\ [g(x) - \gamma we]\lambda \end{bmatrix}$$

denominamos $d_0 = \begin{bmatrix} x_0 - x \\ \vdots \\ w_0 - w \end{bmatrix} \in \mathbb{R}^{n+1}$; $\lambda_0 \in \mathbb{R}^m$. Simplificando y transponiendo términos resulta

$$\begin{aligned}
&1)\quad Bd_0 + \begin{bmatrix}\nabla g(x) \\ -\gamma\end{bmatrix}\lambda_0 = -\begin{bmatrix}\nabla f(x) \\ \gamma w\end{bmatrix}\\
&2)\quad \Lambda[\nabla g(x), -\gamma]d_0 + [g(x) - \gamma we]\lambda_0 = 0
\end{aligned} \tag{9}$$

donde $\begin{bmatrix} \nabla g(x) \\ -\gamma \end{bmatrix} \in \mathbb{R}^{(n+1)*m}$; $\begin{bmatrix} \nabla f(x) \\ \gamma w \end{bmatrix} \in \mathbb{R}^{n+1}$; $\Lambda = diag(\lambda_i) \in \mathbb{R}^{m*n}$, $[\nabla g(x), -\gamma] \in \mathbb{R}^{m*(n+1)}$; $[g(x) - \gamma we] \in \mathbb{R}^{m*m}$ matriz diagonal.

En el sistema (9) es preciso descomponer la solución "d_0" (dirección en el espacio primal) en $d_{0x} \in \mathbb{R}^n$ y $d_{0w} \in \mathbb{R}$ con el objetivo de otorgar la flexibilidad necesaria a la dirección de búsqueda dependiendo del espacio en el que esté operando el Algoritmo $\mathbb{R}^n$ o $\mathbb{R}^{n+1}$.

$$\begin{aligned} &1)\ \hat{B}d_{0x} + \nabla g(x)\lambda_0 = -\nabla f(x) \\ &2)\ cd_{0w} - \gamma e^t\lambda_0 = -\gamma w \\ &3)\ \Lambda\nabla g^t(x)d_{0x} - \gamma\lambda d_{0w} + [g(x) - \gamma we]\lambda_0 = 0 \end{aligned} \tag{10}$$

En (10) $\hat{B} \in \mathbb{R}^{n\times n}$ es la submatriz de $B = (b_{ij}), i, j = 1, \cdots, n+1$ que se obtiene por la eliminación de la fila $n+1$ y columna $n+1$ y

$$c = \sum_{i=1}^{n+1} b_{i(n+1)} \tag{11}$$

Resulta claro que con esta descomposición el sistema (10) calcula la dirección "d_{0x}" en $\mathbb{R}^n$ y (d_{0x}, d_{0w}) en $\mathbb{R}^{n+1}$ de forma combinada. Sin embargo a pesar de ser (d_{0x}, d_{0w}) una dirección de descenso [8] para el problema (5) interesa que cuando $x \in X^c$, la derivada direccional, en la dirección $e_{n+1} = (0, 0, \cdots, 0, 1)$ sea de descenso, es decir, $d_{0w}\gamma w < 0$. Tal situación no está garantizada en la resolución del sistema (10). Para asegurar el descenso en la dirección e_{n+1} optamos por una duplicidad de la dirección d_{0w} definida como sigue

a) Si de la solución de (10) resulta $d_{0w} < 0$ verdadero, tomamos la dirección en $\mathbb{R}^{n+1}$ por (d_{0x}, d_{0w}).

b) Si resulta que $d_{0w} > 0$ hacemos $d_{0w} = -d_{0w}$ y en este caso tomamos la dirección en $\mathbb{R}^{n+1}$ por $(d_{0x}, -d_{0w})$.

En el caso (a) la dirección (d_{0x}, d_{0w}) es de descenso para el problema (5). No es difícil ver que esto también es verdadero en el caso (b). En efecto, como

$$(d_{0x}^t, d_{0w}) \begin{bmatrix} \nabla f(x) \\ \gamma w \end{bmatrix} < 0 \quad cuando \quad d_{0w} > 0$$

además

$$(d_{0x}^t, -d_{0w}) \begin{bmatrix} \nabla f(x) \\ \gamma w \end{bmatrix} \leq (d_{0x}^t, d_{0w}) \begin{bmatrix} \nabla f(x) \\ \gamma w \end{bmatrix} < 0$$

Debe observarse, en general, que cuando $x \in X^c$ la dirección "d_{0x}" no necesariamente será de descenso para f(x), lo cual implica que no siempre decrecerá la función f(x), además w siempre es mayor o igual a cero.

Nótese que $x \in X$ implica $\gamma = 0$. Luego, $d_{0w} = 0$ se obtiene de la ecuación 2 en (10).

3 Algoritmo básico unificado

En esta sección se presenta el Algoritmo Unificado Fase I- Fase II para el problema 2.0 (1), cuya extensión al problema (OSI-P) es inmediata, haciendo $h_j(x) = \phi_j(x, u)$ con $u \in U_k^j$, tal que $U_k^j \subseteq U^j$ y $\mid U_k^j \mid = s < \infty$.

3.1 Algoritmo modelo

Parámetros: $\alpha, \vartheta, \eta, \beta \in (0, 1); \varphi > 0$.
Datos: $x \in \mathbb{R}^n$; $\lambda \in \mathbb{R}^m$, $\lambda > 0$, $B \in \mathbb{R}^{n \times n}$ simétrica definida positiva; $\omega \in \mathbb{R}^m$, $\omega > 0$.

0.- Si $\mathrm{Max}\{g_i(x)/i = 1, \cdots, m\} \leq 0$, haga $\gamma = 0$. En caso contrario haga $\gamma = 1$.

1.- Calcular la dirección de búsqueda.

i) Determine $(d_{0x}, d_{0w}, \lambda_0)$ resolviendo el sistema lineal (1)

$$\begin{aligned} &1) \quad \hat{B}d_{0x} + \nabla g(x)\lambda_0 = -\nabla f(x) \\ &2) \quad cd_{0w} - \gamma e^t \lambda_0 = -\gamma w \\ &3) \quad \Lambda \nabla g^t(x) d_{0x} - \gamma \lambda d_{0w} + [g(x) - \gamma w e]\lambda_0 = 0 \end{aligned} \tag{1}$$

Si $(d_{0x}, d_{0w}) = 0$, pare.
Si $d_{0w} > 0$ haga $d_{0w} = -d_{0w}$

ii) Determine $(d_{1x}, d_{1w}, \lambda_1)$ resolviendo el sistema lineal (2)

$$\begin{aligned} &1) \quad \hat{B}d_{1x} + \nabla g(x)\lambda_1 = 0 \\ &2) \quad cd_{1w} - \gamma e^t \lambda_1 = 0 \\ &3) \quad \Lambda \nabla g^t(x) d_{1x} - \gamma \lambda d_{1w} + [g(x) - \gamma w e]\lambda_1 = -\Lambda\omega \end{aligned} \tag{2}$$

iii) Si $< (d_{1x}, d_{1w}), (\nabla f(x), \gamma w) > \ >0$, haga

$$\rho = \mathrm{Min}\{\varphi \parallel (d_{0x}, d_{0w}) \parallel^2; (\alpha - 1)(d_{0x}, d_{0w}) \begin{bmatrix} \nabla f(x)^t \\ \gamma w \end{bmatrix} / (d_{1x}, d_{1w}) \begin{bmatrix} \nabla f(x)^t \\ \gamma w \end{bmatrix} \}$$

Si no, haga $\rho = \varphi \parallel (d_{0x}, d_{0w}) \parallel^2$

iv) Calcular la dirección de búsqueda (d_x, d_w)

$d_x = d_{0x} + \rho d_{1x}$; $\quad d_w = d_{0w} + \rho d_{1w}$ $\quad$ y $\quad \overline{\lambda} = \lambda_0 + \rho\lambda_1$

2.- Búsqueda Lineal.

Calcular $\quad$ Max $\{t/t = \beta^k, \quad k = 0, 1, 2, \cdots\}$ $\quad$ tal que

$$f(x + td_x) + \frac{1}{2}\gamma(w + td_w)^2 \leq f(x) + \frac{1}{2}\gamma w^2 + t\eta \begin{bmatrix} \nabla f(x)^t \\ \gamma w \end{bmatrix} (d_x, d_w)$$

$[g(x + td_x) - \gamma(w + td_w)e] \leq 0$

o

$[g(x + td_x) - \gamma(w + td_w)e] \leq \theta[g(x) - \gamma we]$

3.- Actualizaciones

i) Haga $x = x + td_x$; $\quad w = w + td_w$ y defina nuevos valores $\omega > 0$; $\lambda > 0$; B simétrica definida positiva.

ii) Vaya al paso (0)

Observación. Diferentes algoritmos se pueden obtener dependiendo de la forma de actualizar λ, ω y B.

En el paso (0) si $\gamma = 1$ haga $w_0 = max\{g_i(x)/i = 1, \cdots, m\}$ y considere el punto inicial $(x, w_0) \in \mathbb{R}^{n+1}$.

3.2 Propiedades del algoritmo modelo

En lo sucesivo asumimos las siguientes hipótesis en adición a las de [3]:

Suposición A: $\quad$ Para todo $x \in X^c$, $\quad \sum_{i=1}^{m} \lambda_0^i \neq w$ y $c \neq 0$, donde "c" está definido como en 2.0 - (11).

El objeto de la hipótesis anterior es asegurar la existencia de una dirección de descenso (d_x, d_w) para el problema expandido 1.0-(4) en el sentido de no alcanzar o evitar un falso punto estacionario (x, w) tal que $x \notin X$.

Sea $\quad S \subseteq X \quad$ el conjunto de puntos estacionarios para el problema 2.0-(1) que satisface las condiciones de K.K.T. Esto es, $\quad \forall x \in S \quad x$ es punto regular y existen multiplicadores $\quad \lambda_i \geq 0 \quad$ i = 1, 2, ..., m $\quad$ tal que

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) &= 0 \\ g_i(x)\lambda_i &= 0 \end{aligned}$$

Lema 3.1 *Asumamos que la suposición A es verdadera. Entonces*
a) Si $\bar{x} \in X$ *es solución óptima para el problema 2.0-(1)* $(d_{0x}, d_{0w}) = 0$
b) Para todo $\bar{x} \in X^c$, $d_{0w} < 0$
c) $(d_{0x}, d_{0w}) = 0$ *si y sólo si* $\bar{x} \in S$

Demostración

a) Como $\bar{x} \in X$ entonces $\gamma = 0$, y de las ecuaciones 1 y 2 del sistema lineal 3.0 (1) en el paso 1i) resulta

1) $\hat{B}d_{0x} + \nabla g(\bar{x})\lambda_0 = -\nabla f(\bar{x})$

2) $cd_{0w} = 0 \rightarrow d_{0w} = 0$ ya que $c \neq 0$

$d_{0x} = -(\nabla f(\bar{x}) + \nabla g(\bar{x})\lambda_0)\hat{B}^{-1}$; $\hat{B}^{-1} \neq 0$ s.d.p. implica $d_{0x} = 0$

b) Como $\bar{x} \in X^c$ entonces $\gamma = 1$. Nuevamente de la ecuación 2 del sistema lineal (1) se tiene $cd_{0w} - \gamma e^t\lambda_0 = -\gamma w \Rightarrow d_{0w} = 1/overc \left(\sum_{i=1}^{m} \lambda_0^i - w\right) \neq 0$, por aplicación de la suposición A. Además del paso 1i) se tiene que $d_{0w} < 0$

c) ($\Rightarrow$) Supongamos $(d_{0x}, d_{0w}) = 0$ del sistema 3.2.-(1) obtenemos

4) $(\nabla f(\bar{x}) + \nabla g(\bar{x})\lambda_0)\hat{B}^{-1} = 0$

5) $\gamma \left(\sum_{i=1}^{m} \lambda_0^i - w\right) = 0$

6) $[g(\bar{x}) - \gamma we]\lambda_0 = 0$

De ecuación (4) como $\hat{B}^{-1} \neq 0$ entonces $\nabla f(\bar{x}) + \nabla g(\bar{x})\lambda_0 = 0$. De la ecuación (5), probaremos que γ no puede tener el valor 1. En efecto, si así fuere implicaría que $\left(\sum_{i=1}^{m} \lambda'_o - w\right) = 0$, lo cual no puede ser por que contradice la suposición A. Luego debe ocurrir que $\gamma = 0$. Esto último implica que Max $\{g_i(\bar{x})/i = 1, \cdots, m\} \leq 0$ y de (6) $g_i(\bar{x})\lambda_0^i = 0 \quad \forall i = 1, \cdots, m$, lo cual prueba $\bar{x} \in S$.

($\Leftarrow$) Supongamos $\bar{x} \in S$, entonces Max $\{g_i(\bar{x})/i = 1, \cdots, m\} \leq 0$ de donde $\gamma = 0$, con lo cual trivialmente se desprende del sistema 3.0-(1) ecuaciones 1), 2) que $(d_{0x}, d_{0w}) = 0$. ∎

Resulta inmediato verificar que las propiedades b) y c) del Lema (3.1) son también propiedades para la dirección deflectada (d_x, d_w) debido a que por construcción de esta misma se tiene

$$d_x = d_{0x} + \rho d_{1x}; \qquad d_w = d_{0w} + \rho d_{1w}$$

en que el parámetro ρ depende de la derivada direccional de (d_{0x}, d_{0w}) respectivamente.

En lo sucesivo denotaremos nuestro Algoritmo Modelo por la multifunción $A : X \to \mathbb{P}(X)$ o por $A_w : X_w \to \mathbb{P}(X_w)$ dependiendo si nos referimos al problema 2.0-(1) o al problema 2.0-(4) respectivamente. Donde X_w es el conjunto definido por

$$X_w = \{(x,w) \in \mathbb{R}^{n+1} / g_i(x) - \gamma w \leq 0 \qquad i = 1, \cdots, m\}$$

Suposición B: $\text{int}X_w$ es no vacío y acotado.

Lema 3.2 *Supongamos que las Suposiciones A y B son verdaderas. Sean* $(x,w) \in (intX_w)$ *y* $(x',w') \in B[(x,w),\epsilon] \cap X_w$ *tales que* $(x",w") \in A_w(x',w')$. *Supongamos que existe* $\delta(x) > 0$ *tal que* $w" - w' \leq -\delta$ *entonces*

$$Max\{g_i(x")/i = 1, \cdots, m\} - Max\{g_i(x')/\ i = 1, \cdots, m\} \leq -\delta$$

Demostración Observemos en primer lugar que $(x",w"),(x',w') \in X_w$. Además por hipótesis $w" - w' \leq -\delta$ lo cual implica que existe $t \in (0,1]$ tal que

$$x" = x' + td_x; \qquad w" = w' + td_w \Rightarrow td_w \leq -\delta$$

Por otro lado $\forall (x,w) \in X_w$ se tiene que :

$$< \begin{bmatrix} \nabla g_i(x) \\ -\gamma \end{bmatrix}, \begin{bmatrix} d_x \\ d_w \end{bmatrix} > \leq 0$$

lo que equivale a
$< \nabla g_i(x), d_x > \leq \gamma d_w \leq \gamma t d_w \leq -\delta, \qquad \forall i = 1, \cdots, m.$
Nótese que $d_w < 0$ luego aplicando el Teorema del Valor Medio a $g_i(x")$ en $t \in (0,1)$ resulta Max $\{g_i(x' + td_x)/i = 1, \cdots, m\}$ − Max$\{g_i(x')/i = 1, \cdots, m\}$ = Max$\{g_i(x') + t < \nabla g_i(x), d_x > /i = 1, \cdots, m\}$ − Max$\{g_i(x')/i = 1, \cdots, m\} \leq$ tMax$\{< \nabla g_i(x), d_x > /i = 1, \cdots, m\} \leq$ Max$\{< \nabla g_i(x), d_x > /i = 1, \cdots, m\} \leq$ $-\delta$ ■

Esto último prueba que el descenso de "w" está estrictamente ligado a un descenso de la función Max $\{g_i(x)/i = 1, \cdots, m\}$.

En el siguiente Teorema (3.3) se dan condiciones que si son satisfechas por el algoritmo A(x) aseguran que todos los puntos de acumulación de la sucesión $\{x_i\}$ pertenecen a S, es decir viables y satisfacen las condiciones necesarias de K.K.T.

Teorema 3.3
a) Sea $x \in X$ *y supongamos* $A(x) \neq 0$ *entonces* $\forall x' \in A(x)$ *se tiene* $x' \in X$, *es*

decir $A(x) \subseteq X$

b) Sea $x \in (\mathbb{R}^n - S)$, *existen escalares* $\delta_i > 0, \epsilon_i > 0 i = 1,2$ *tales que:*

i) $f(x'') - f(x') \leq -\delta_1$; $\quad \forall x' \in B(x,\epsilon_1) \cap X$ *y* $\forall x'' \in A(x')$

ii) $Max\{g_i(x'')/i = 1,\cdots,m\} - Max\{g_i(x')/i = 1,\cdots,m\} \leq -\delta_2$; $\quad \forall (x',w') \in B[(x,w),\epsilon_2] \cap X^c$ *y* $\forall (x'',w'') \in A_w(x',w')$.

c) Todo punto de acumulación $\bar{x}$ *de una sucesión infinita* $\{x_i\}$ *generada por el Algoritmo satisface* $\bar{x} \in S$ *(es decir, es un punto estacionario).*

Observación. La notación $B(x,\epsilon)$ se refiere a una vecindad de centro x y radio ϵ, es decir $B(x,\epsilon) = \{y \in \mathbb{R}^n / \parallel x - y \parallel < \epsilon.\}$

Demostración

a) Si $x \in S$ entonces $\gamma = 0$. Existe $x' \in A(x)$ tal que $x' = x + td_x$, como $t > 0$ y $d_x \neq 0$ están construidos de tal manera que d_x es dirección viable y t es tal que $g_i(x + td_x) \leq 0 \quad i = 1,\cdots,m$. Luego $x' \in X, \forall x' \in A(x)$ por lo tanto $A(x) \subseteq X$.

b) Como $x \in (\mathbb{R}^n - S)$, entonces $x \in (X - S)$ o $x \in X^c$ luego:

i) Si $x \in (X - S)$, como X es compacto, entonces X - S es compacto, por lo tanto existe $\epsilon_1 > 0$ tal que $B(x,\epsilon_1) \subseteq (X - S)$. Sea $x' \in B(x,\epsilon_1)$ y $x'' \in A(x')$ entonces $x'' = x' + \bar{t}d_x$ con $\bar{t} = \beta^k, k = 0,1,2,\cdots$ y $\beta \in (0,1)$. Del sistema (1) en el Algoritmo se desprende que $\quad < \nabla f(x'), d_x > \leq \alpha < \nabla f(x'), d_{0x} > \leq -\alpha\varphi \parallel d_{0x} \parallel^2 < 0$ ([2,3]) luego existe $0 < \delta_1(x) \leq \alpha\varphi \parallel d_{0x} \parallel^2, \alpha > 0, \varphi > 0$. Del paso (2) se tiene que $f(x'') - f(x') \leq \bar{t}\eta < \nabla f(x'), d_x > \leq \bar{t}\eta\alpha < \nabla f(x'), d_{0x} > \leq -\alpha\varphi\eta\bar{t} \parallel d_{0x} \parallel^2$, donde $\delta_1 = \alpha\varphi\eta\bar{t} \parallel d_{0x} \parallel^2 \varphi > 0, \ \alpha, \ \eta, \ \bar{t} \in (0,1)$. Luego $f(x'') - f(x') \leq -\delta_1, \quad \forall x' \in B(x,\epsilon_1)$.

ii) Supongamos que $x \in X^c$ entonces $\gamma = 1$ y existe $w > 0$ tal que $(x,w) \in X_w$, como X_w es compacto entonces existe $\epsilon_2 > 0$ tal que $B[(x,w),\epsilon_2] \subseteq X_w$. Tomemos $(x',w') \in B[(x,w),\epsilon_2]$ y $(x'',w'') \in A_w(x',w')$ entonces del paso (2) se desprende que $f(x'') + \dfrac{1}{2}\gamma(w'')^2 - (f(x') + \dfrac{1}{2}\gamma w'^2) \leq \bar{t}\eta \begin{pmatrix} \nabla f(x)^t \\ \gamma w \end{pmatrix} (d_x, d_w) \leq -\alpha\varphi\eta\bar{t} \parallel (d_{0x}, d_{0w} \parallel^2 < 0$.
Esto último se puede expresar por $w'' - w' \leq -2\alpha\varphi\eta\bar{t} \parallel (d_{0x}, d_{0w}) \parallel^2 / w'' + w' + (f(x') - f(x''))/w'' + w'$. En consecuencia existe $\delta_2 > 0$ tal que $0 < \delta_2 \leq \mid -2\alpha\varphi\eta\bar{t} \parallel (d_{0x}, d_{0w}) \parallel^2 / w'' + w' + (f(x') - f(x''))/w'' + w' \mid$ por lo tanto $w'' - w' \leq -\delta_2$.

Del lema (3.2) se concluye

$$\mathrm{Max}\{g_i(x'')/i = 1,\cdots,m\} - \mathrm{Max}\{g_i(x')/i = 1,\cdots,m\} \leq -\delta_2$$

c) Sea $\{x_i\}_{i\in N}$ una sucesión construida por el algoritmo y supongamos que tiene un punto de acumulación $\bar{x} \notin S$ (es decir, para alguna subsucesión $K \subseteq \mathbb{N}$ $x_i \xrightarrow{K} \bar{x}$ y $\bar{x} \notin S$). Se tienen dos casos:

1) Existe $N_0 > 0$, tal que para todo $i \geq N_0, x_i \in X$ de (bi) se tiene que la sucesión $\{f(x_i)\}_{i=N}^{\infty}$ es monótona decreciente, por la continuidad de f se tiene que $f(x_i) \xrightarrow{K} f(\bar{x})$ cuando $i \to \infty$ o sea $\lim_{i\to\infty}[f(x_{i+1}) - f(x_i)] = 0$ y $f(x_{i+1}) - f(x_i) < -\delta_2$ $\forall i \in K$, lo cual es una contradicción, luego $\bar{x} \in S$

2) Supongamos que $x_i \notin X, i \in \mathbb{N}$ entonces existe una sucesión $\{w_i\}_{i\in\mathbb{N}} \subseteq \mathbb{R}$ tal que la sucesión $\{(x_i, w_i)\}_{i\in\mathbb{N}} \subseteq X_w$ de (b, ii) y Lema (3.2) la sucesión $\{\text{Max}\{g_i(x)/i = 1, \cdots, m\}\}_{i\in\mathbb{N}}$ es una sucesión monótona decreciente, además existe la subsucesión

$$\{\text{Max}\{g_i(x)/i = 1, \cdots, m\}\}_{i\in K}, \qquad K \subseteq \mathbb{N}$$

convergente a $\text{Max}\{g_i(\bar{x})/i = 1, \cdots, m\}$. De aquí que

$$\lim_{i\to\infty} \text{Max}\{g_i(x_{j+1}) - g_i(x_j)/i = 1, \cdots, m\} = 0$$

pero de (b,ii) se tiene que

$$\text{Max}\{g_i(x_{j+1})/i = 1, \cdots, m\} - \text{Max}\{g_i(x_j)/i = 1, \cdots, m\} \leq -\delta_2 < 0$$

$\forall i \in K$, lo cual es una contradicción, concluyendo $\bar{x} \in S$ ■

Observación. La condición a) establece que si x es viable entonces cualquier subsucesión de puntos generada por el algoritmo es viable. La condición b) establece que si x no es un punto estacionario entonces el algoritmo produce una reducción $\delta_1 > 0$ en la función 'f' para todo punto viable $x' \in B(x, \epsilon_1)$ y una reducción $\delta_2 > 0$ en $\text{Max}\{g_i(x')/i = 1, \cdots, m\}$ para todo punto no viable x' tal que $(x', w') \in B[(x, w), \epsilon_2]$.

Referencias

[1] Polak E., Trahan R. *Combined Phase I- Phase II methods of feasible directions*, Mathematical Programming, **17 (1979)**.

[2] Panier E., Tits A., Herskovits J. *A QP-Free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization*. Siam J. Control and optimization, **26**, **Nº 4 (1988)**.

[3] Herskovits J. *An interior point technique for nonlinear optimization* Rapports de Recherche, Le chesnay Cedex (INRIA) France, **(1992)**.

[4] Polak E. *Computational Methods in Optimization*, A Unified Approach, Academic Press, (1971).

Dirección del autor:

Constande Nicolas B.
Instituto de Matemáticas
Universidad Austral de Chile
Casilla 567, Valdivia