Documento de Trabajo

9217

EL TEOREMA DE LA ENVOLVENTE COMO INSTRUMENTO EN TEORIA DE OPTIMIZACION

Emilio Cerdá Tena

FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES.
UNIVERSIDAD COMPLUTENSE DE MADRID.
Campus de Somosaguas. 28223 MADRID.

EL TEOREMA DE LA ENVOLVENTE COMO INSTRUMENTO EN TEORÍA DE OPTIMIZACIÓN.

Emilio Cerdá Tena
Departamento de Análisis Económico

RESUMEN:

En este trabajo se extiende el teorema de la envolvente en dos direcciones: a) Cierto tipo de funciones objetivo compuestas, que aparecen habitualmente en Optimización. b) Enriqueciendo los resultados del teorema con información procedente del programa dual minimax del problema original. Se presentan también aplicaciones de los resultados obtenidos a programación lineal, programación geométrica y programación dinámica, donde se ve que el teorema de la envolvente puede servir para hacer análisis de sensibilidad y también para el cálculo de la solución óptima.

Agradezco a A.Abadía, C.Carrera, M.E. Mera, M. Morán, A. Novales, J.M. Rey, A. Rodrigo y M. Vázquez sus comentarios y sugerencias. La responsabilidad de los errores que puedan persistir es exclusivamente mía.

1. INTRODUCCION

Viner (1932), al analizar el comportamiento de empresas, consideró diferentes curvas de coste a corto plazo, cuyos puntos mínimos decrecian al principio y luego crecian, para factores de capital cada vez mayores. Razonó que si ambos inputs fueran variables, entonces la curva de coste medio a largo plazo sería siempre menor o igual que la de las correspondientes curvas a corto plazo. Concluyó que la curva de coste medio a largo plazo debería ser dibujada como la envolvente de las curvas a corto plazo, aunque le extrañó el hecho de que la envolvente que había considerado no pasara por los puntos mínimos de las curvas a corto plazo. Posteriormente, Viner encargó a su delineante Wong que le representara una curva de costes a largo plazo que fuera envolvente de las de corto plazo y, a la vez, pasara por los puntos mínimos de dichas curvas. Wong respondió que era imposible, a raiz de lo cual Viner decidió representar la curva de coste medio a largo plazo pasando por los puntos mínimos de las curvas de coste a corto plazo y no como la curva envolvente.

Samuelson (1947) analizó el problema matemáticamente, demostrando que la curva a largo plazo era tangente (o sea envolvente) a las curvas a corto plazo. El análisis de Samuelson no se refiere, en particular, a curvas de coste sino que trata con funciones generales que dependen de variables y parámetros. Se trata del teorema de la envolvente.

Silberberg (1971, 1974, 1978, 1990) utiliza la metodologia primal-dual, obteniendo resultados de estática comparativa. Para un problema de optimización con parámetros, define una función auxiliar a la que llama función objetivo primal-dual, cuyas condiciones de primer orden constituyen un sistema formado por las condiciones de primer orden de la función original y las del teorema de la envolvente. Utiliza también las condiciones de segundo orden de la función primal-dual, que le permiten obtener resultados interesantes de estática comparativa. Como señala el mismo Silberberg en el prólogo de su libro de 1990, todos los resultados de dualidad que obtiene son derivaciones o aplicaciones del teorema de la envolvente.

Caputo (1990) extiende la metodologia primal-dual de Silberberg al caso dinámico. Considera un problema de control óptimo en tiempo continuo, donde todas las funciones del problema dependen de un vector de parámetros, y en el que establece un número considerable de hipótesis, a cuya justificación dedica bastante espacio. Enuncia y demuestra un teorema de la envolvente dinámico, fundamento de su metodologia.

La France y Barney (1991) establecen el teorema de la envolvente en optimización dinámica, que es más general que el de Caputo. Consideran un problema de control óptimo en tiempo continuo, con horizonte temporal finito, en el que todas las funciones del problema dependen de un vector de parámetros y en el que, además de la ecuación de estado y condición inicial, considera restricciones de igualdad y de desigualdad. Demuestra el teorema, estudia el caso convexo y otros casos particulares. Finalmente, una aplicación a un problema intertemporal de un consumidor le permite obtener las versiones dinámicas del Lema de Hoteling, la Identidad de Roy y la Ecuación de Slutsky.

En este trabajo se recogen los primeros resultados de una investigación iniciada hace unos meses sobre el teorema de la envolvente, como instrumento en optimización estática y dinámica. El apartado 2 contiene los resultados previos: las versiones del teorema de la envolvente en diferentes programas matemáticos con parámetros. El apartado 3 contiene los resultados y aplicaciones: en primer lugar se estudia el problema de optimización diferenciable de una función compuesta que depende de parámetros; en segundo lugar se estudian conjuntamente los programas primal y dual minimax, dependientes de parámetros, estableciendo un teorema que permite añadir información procedente del dual a la que nos da el teorema de la envolvente en el primal. El teorema demostrado se aplica a continuación a la programación lineal, lo que nos permite obtener unos resultados que tienen clara interpretación intuitiva en análisis de sensibilidad. También se aplica a la programación geométrica y, en particular, a un problema de minimización de costes en donde el teorema demostrado anteriormente nos sirve no sólo para hacer análisis de sensibilidad, sino también para calcular la solución óptima. Por último nos planteamos el problema de control óptimo lineal-cuadrático en tiempo discreto, con horizonte temporal infinito (cuya solución es conocida en la literatura de teoría de control) y lo resolvemos por otro método, utilizando el teorema de la envolvente, lo que nos permite llegar a la solución conocida de una forma más corta y sencilla. El apartado 4 recoge las conclusiones del trabajo.

2. RESULTADOS PREVIOS

El teorema de la envolvente en optimización sin restricciones.

Sea el problema:

$$\text{MIN } f(x_1,x_2,\ldots,x_n,\alpha_1,\alpha_2,\ldots,\alpha_m)$$

, siendo f de clase C^2, en donde

$x = (x_1,\ldots,x_n)$ es un vector de variables de decisión

$\alpha = (\alpha_1,\ldots,\alpha_m)$ es un vector de parámetros.

Supongamos que para todo $\alpha^o \in A$ (abierto), existe x^o, único, que resuelve el problema dado, verificando condiciones necesarias y suficientes de optimalidad local.

Sea $x_1 = x_1^*(\alpha)$, $x_2 = x_2^*(\alpha)$, ..., $x_n = x_n^*(\alpha)$ la solución del problema.

Sea $\phi(\alpha_1,\ldots,\alpha_m) = f(x_1^*(\alpha),\ldots,x_n^*(\alpha),\alpha)$ función objetivo indirecta.

Entonces se verifica que:

$$\frac{\partial \phi}{\partial \alpha_j}(\alpha^o) = \frac{\partial f}{\partial \alpha_j}(x^o;\alpha^o)$$

El teorema de la envolvente en optimización diferenciable con restricciones de igualdad.

Sea el problema:

$$\begin{cases} \text{MIN } f(x,\alpha) \\ \text{, sujeto a: } g_1(x,\alpha) = 0 \\ \qquad\qquad\quad \vdots \qquad\quad \vdots \\ \qquad\qquad\quad g_\ell(x,\alpha) = 0 \end{cases}$$

, siendo f, g_1, ..., g_ℓ funciones de clase C^2, en donde $x = (x_1,...,x_n)$ es un vector de variables de decisión; $\alpha = (\alpha_1,...,\alpha_m)$ es un vector de parámetros.

Supongamos que para todo $\alpha^\circ \in A$ (abierto), existe x°, único, con multiplicador de Lagrange asociado λ°, único, que resuelve el problema dado, verificando condiciones necesarias y suficientes de optimalidad local.

Sea $x = x^*(\alpha)$, $\lambda = \lambda^*(\alpha)$ la solución del problema.

Sea $L(x,\lambda;\alpha) = f(x,\alpha) + \sum_{i=1}^{\ell} \lambda_i\, g_i(x,\alpha)$ la función Lagrangiana.

Sea $\phi(\alpha) = f(x_1^*(\alpha),...,x_n^*(\alpha),\alpha)$ función objetivo indirecta

Entonces se verifica que:

$$\frac{\partial \phi}{\partial \alpha_j}(\alpha^\circ) = \frac{\partial L}{\partial \alpha_j}(x^\circ,\lambda^\circ,\alpha^\circ)$$

<u>Nota</u>:

Una aplicación inmediata del teorema anterior, conocida en cierta literatura económica, la encontramos en la interpretación del valor del multiplicador de Lagrange en el óptimo, asociado a una restricción. En efecto:

Consideramos el programa matemático:

$$\text{MIN } f(x_1,...,x_n)$$
$$\text{sujeto a } g(x_1,...,x_n) = 0$$

que cumple todas las condiciones para que sea aplicable el método de Lagrange.

Consideramos ahora el siguiente programa, dependiente del parámetro $\alpha \in \mathbb{R}$.

$$\text{MIN } f(x_1,\ldots,x_n)$$

$$\text{sujeto a } g(x_1,\ldots,x_n) = \alpha$$

Sea $L(x,\lambda;\alpha) = f(x_1,\ldots,x_n) + \lambda\,[g(x_1,\ldots,x_n) - \alpha]$

Para $\alpha = 0$, sean $x^o = x^*(0)$; $\lambda^o = \lambda^*(0)$ la solución óptima del problema original.

Sea $\phi(\alpha) = f(x_1^*(\alpha),\ldots,x_n^*(\alpha))$ que, en este caso, se llama función de perturbación.

Aplicando el teorema de la envolvente obtenemos que:

$$\phi'(0) = \frac{\partial L}{\partial \alpha}(x^o;\lambda^o;0) = \frac{\partial f}{\partial \alpha}(x^o) = -\lambda^o \Rightarrow \lambda^o = -\frac{\partial f}{\partial \alpha}(x^o)$$

<u>El teorema de la envolvente en optimización diferenciable con restricciones de desigualdad</u>.

Sea el problema:

$$\left[\begin{array}{ll} \text{MIN } f(x,\alpha) & \\ \text{sujeto a:} & g_1(x,\alpha) \le 0 \\ & \vdots \\ & g_\ell(x,\alpha) \le 0 \end{array}\right.$$

, siendo f, g_1, ..., g_ℓ funciones de clase C^2, en donde $x = (x_1,\ldots,x_n)$ es un vector de variables de decisión; $\alpha = (\alpha_1,\ldots,\alpha_m)$ es un vector de parámetros.

Supongamos que para todo $\alpha^o \in A$ (abierto), existe x^o, único, con multiplicador de Lagrange asociado λ^o, único, que resuelve el problema dado, verificando condiciones necesarias y suficientes de optimalidad local.

Sea $x = x^*(\alpha)$, $\lambda = \lambda^*(\alpha)$ la solución del problema.

Sea $L(x,\lambda,\alpha) = f(x,\alpha) + \sum_{i=1}^{\ell} \lambda_i \, g_i(x,\alpha)$ la función Lagrangiana.

Sea $\phi(\alpha) = f(x_1^*(\alpha),\ldots,x_n^*(\alpha),\alpha)$ función objetivo indirecta

Entonces se verifica que:

$$\frac{\partial \phi}{\partial \alpha_j}(\alpha^o) = \frac{\partial L}{\partial \alpha_j}(x^o,\lambda^o,\alpha^o)$$

3. RESULTADOS Y APLICACIONES

3.1. El teorema de la envolvente para una función objetivo compuesta

Un recurso utilizado en Teoría de la Optimización consiste en expresar la función objetivo como composición de funciones más manejables y deducir a partir de ellas la solución óptima del problema original. En otras ocasiones lo que se hace es transformar la función objetivo dada en otra que resulte más manejable (por ejemplo, tomando logaritmos) y deducir a partir de ella la solución al problema original. El problema que abordamos a continuación contempla las dos posibilidades, cuando la función objetivo depende de parámetros. En la siguiente proposición se plantea y resuelve dicho problema, tal como señala el encabezamiento.

Proposición

Sea el programa matemático

(1) $\quad$ OPT $h[f(x_1,\ldots,x_n;\alpha_1,\ldots,\alpha_m)] = g(x_1,\ldots,x_n;\alpha_1,\ldots,\alpha_m)$

, siendo h, f funciones de clase C^2, h estrictamente creciente, en donde $x = (x_1,\ldots,x_n)$ es un vector de variables de decisión; $\alpha = (\alpha_1,\ldots,\alpha_m)$ es un vector de parámetros.

Supongamos que para todo $\alpha^\circ \in A$ (abierto), existe x°, único, que resuelve el programa:

(2) $\quad$ OPT $f(x_1,\ldots,x_n;\alpha_1,\ldots,\alpha_m)$

verificando las condiciones necesarias y suficientes de optimalidad local.

Entonces se verifica:

(i) $x_1 = x_1^*(\alpha)$, ..., $x_n = x_n^*(\alpha)$ es solución del problema (1) si y sólo si es solución del problema (2).

(ii)

$$\frac{\partial \varphi}{\partial \alpha_j}(\alpha^\circ) = \frac{\partial g}{\partial \alpha_j}(x^\circ,\alpha^\circ) = h'[f(x^\circ,\alpha^\circ)]\left\{\frac{\partial f}{\partial \alpha_j}(x^\circ,\alpha^\circ)\right\} =$$

$$= h'[\phi(\alpha^\circ)]\,\frac{\partial \phi}{\partial \alpha_j}(\alpha^\circ)$$

en donde φ es la función objetivo indirecta del programa (1) y ϕ es la función objetivo indirecta del programa (2).

Demostración:

Por ser h estrictamente creciente, se verifica que $h'(c) \neq 0$, para todo c perteneciente al dominio de definición de h.

Las condiciones de primer orden son las mismas para los programas (1) y (2).

En efecto:

$$\frac{\partial g}{\partial x_i} = \underbrace{h'[f(x)]}_{\text{distinto de cero}} \frac{\partial f}{\partial x_i} = 0, \text{ para } i = 1,\ldots,n$$

Como por hipótesis, dado $\alpha \in A$, existe solución única x al problema (2), verificándose las condiciones suficientes de segundo orden de optimalidad local, resulta que $Hf(x,\alpha) \neq 0$, por lo que en el sistema de condiciones de primer orden aplicamos el teorema de la función implícita en un entorno de (x,α), obteniendo que $x_1 = x_1^*(\alpha),\ldots, x_n = x_n^*(\alpha)$, siendo $x_1^*,\ldots,x_n^*$ funciones diferenciables.

También sabemos que x es un punto óptimo de f si y sólo si lo es de g. (Véase Barbolla, Cerdá, Sanz (1991), páginas 229 y siguientes).

Por tanto (i) queda demostrado.

Por definición de función objetivo indirecta del programa (1):

$$\phi(\alpha_1,\ldots,\alpha_m) = g(x_1^*(\alpha),\ldots,x_n^*(\alpha),\alpha_1,\ldots,\alpha_m) =$$

$$= h[f(x_1^*(\alpha),\ldots,x_n^*(\alpha),\alpha_1,\ldots,\alpha_m] = h[\phi(\alpha_1,\ldots,\alpha_m)]$$

Aplicando el teorema de la envolvente al programa (1) obtenemos:

$$\frac{\partial \varphi}{\partial \alpha_j}(\alpha^{\circ}) = \frac{\partial g}{\partial \alpha_j}(x^{\circ},\alpha^{\circ})$$

Por ser $g(x,\alpha) = h[f(x,\alpha)]$, aplicando la regla de la cadena, obtenemos:

$$\frac{\partial g}{\partial \alpha_j}(x^{\circ},\alpha^{\circ}) = h'[f(x^{\circ},\alpha^{\circ})]\,\frac{\partial f}{\partial \alpha_j}(x^{\circ},\alpha^{\circ})$$

Aplicando el teorema de la envolvente al programa (2), obtenemos:

$$\frac{\partial f}{\partial \alpha_j}(x^{\circ},\alpha^{\circ}) = \frac{\partial \phi}{\partial \alpha_j}(\alpha^{\circ})$$

, teniendo en cuenta que $f(x^{\circ},\alpha^{\circ}) = \phi(\alpha^{\circ})$, queda finalmente:

$$\frac{\partial \varphi}{\partial \alpha_j}(\alpha^{\circ}) = \frac{\partial g}{\partial \alpha_j}(x^{\circ},\alpha^{\circ}) = h'[f(x^{\circ},\alpha^{\circ})]\,\frac{\partial f}{\partial \alpha_j}(x^{\circ},\alpha^{\circ}) =$$

$$= h'[\phi(\alpha^{\circ})]\,\frac{\partial \phi}{\partial \alpha_j}(\alpha^{\circ})$$

como queríamos demostrar.

<u>Nota</u>:

Hemos estudiado el caso h estrictamente creciente porque es el que más se utiliza por resultar más práctico. Se pueden estudiar de manera análoga otros casos como h estrictamente decreciente, h creciente, h decreciente, etc.

Ejemplo (procedente del libro de Alpha Chiang)

Supongamos que cierto comerciante posee determinada cantidad de vino que puede vender ahora ($t = 0$) por una suma de K unidades monetarias, o almacenarla por un periodo de tiempo y entonces venderla por un valor superior. El valor V (creciente) del vino se obtiene mediante la siguiente función del tiempo:

$$V = K\, e^{\sqrt{t}} \qquad \text{(definida en } t \geq 0)$$

de modo que si $t = 0$, entonces $V = K$. Se supone que el coste de almacenamiento es nulo y que el tipo de interés anual es r. El problema es determinar cuándo debe vender el vino para maximizar el beneficio.

Se trata de resolver el programa:

$$\text{MAX } A(t) = K\, e^{\sqrt{t}}\, e^{-rt} = K\, e^{\sqrt{t}-rt} \qquad \text{(definido en } t \geq 0)$$

Consideramos $B(t) = h[A(t)] = \ln A(t) = \ln K + \sqrt{t} - rt$

Condiciones de primer orden:

$$B'(t) = 0 = \frac{1}{2\sqrt{t}} - r \Rightarrow \frac{1}{\sqrt{t}} = 2r \Rightarrow t^* = \frac{1}{4r^2} \quad \text{(que es > 0)}$$

Condiciones de segundo orden:

$$B''(t) = -\frac{1}{4t\sqrt{t}} < 0, \quad \forall\, t > 0, \text{ luego es máximo.}$$

La solución óptima al problema es $t^* = \dfrac{1}{4r^2}$

Hasta aquí el ejemplo tal como lo enuncia y resuelve Chiang. Se sirve de la función $B(t) = h[A(t)]$, siendo h monótona creciente, para que los cálculos resulten más sencillos.

Supongamos ahora que $K = 30$ y $r = 0.1$. Queremos saber cuál es en ese caso el momento de venta óptimo y el beneficio óptimo. ¿Cuál es el ritmo de variación del beneficio óptimo al variar K? ¿Y al variar r?.

Si $K = 30$ y $r = 0.1$, entonces $t^* = \dfrac{1}{0.04} = 25$ es el momento de venta óptimo.

$B(t^*) = 5.90$, de donde $A(t^*) = e^{5.90} = 365.04$, es el beneficio óptimo.

Consideremos el problema con función objetivo B(t). La función objetivo indirecta es:

$$\phi(K,r) = \ln K + \frac{1}{4r}$$

Aplicando el teorema de la envolvente a dicho problema:

$$\frac{\partial\phi}{\partial K} = \frac{\partial B}{\partial K} = \frac{1}{K} \Rightarrow \frac{\partial\phi}{\partial K}(30,0.1) = \frac{\partial B}{\partial K}(t^*) = \frac{1}{30}$$

Aplicando la proposición anterior:

$$\frac{\partial\phi}{\partial K}(30,0.1) = \frac{\partial B}{\partial K}(t^*) = \frac{1}{30} = \frac{1}{365.04}\frac{\partial A}{\partial K}(t^*) =$$

$$= \frac{1}{365.04}\frac{\partial\phi}{\partial K}(30,0.1) \Rightarrow$$

$$\Rightarrow \frac{\partial A}{\partial K}(t^*) = \frac{\partial\phi}{\partial K}(30,0.1) = \frac{365.04}{30} = 12.17,$$ es el ritmo de variación del beneficio óptimo al variar K, desde el valor 30.

Análogamente:

$$\frac{\partial \phi}{\partial r} = \frac{\partial B}{\partial r}(t^*) = \frac{-1}{4r^2} \Rightarrow \frac{\partial \phi}{\partial r}(30,0.1) = \frac{\partial B}{\partial r}(t^*) = -25$$

Aplicando la proposición anterior:

$$\frac{\partial \phi}{\partial r}(30,0.1) = \frac{\partial B}{\partial r}(t^*) = -25 = \frac{1}{365.04}\frac{\partial A}{\partial r}(t^*) =$$

$$= \frac{1}{365.04}\frac{\partial \phi}{\partial r}(30,0.1) \qquad \Rightarrow$$

$\Rightarrow \quad \dfrac{\partial A}{\partial r}(t^*) = \dfrac{\partial \phi}{\partial r}(30,0.1) = -9126$, que es el ritmo de variación del beneficio óptimo al variar r desde el valor 0.1.

Nota:

En la segunda parte del ejercicio hemos utilizado el procedimiento que nos da la proposición anterior para simplificar los cálculos, en la línea de lo que hace Chiang en la primera parte del ejemplo. En este caso tan sencillo se pueden hacer todos los cálculos directamente con la función A(t) (tanto en la primera parte como en la segunda), pero en otros casos más complicados estos procedimientos pueden ser imprescindibles.

3.2. El teorema de la envolvente en dualidad minimax.

Consideramos el programa matemático:

$$(P)\quad \begin{cases} \text{MIN } f(x,\alpha) \\ \text{sujeto a:}\quad g_1(x,\alpha) \le 0 \\ \qquad\qquad\quad \vdots \\ \qquad\qquad\quad g_\ell(x,\alpha) \le 0 \end{cases}$$

siendo $S \subset \mathbb{R}^n$ el dominio de las funciones $f, g_1, \ldots, g_\ell$. En dicho programa, al que llamaremos programa primal, $x = (x_1,\ldots,x_n)$ es el vector de las variables de decisión del problema y $\alpha = (\alpha_1,\ldots,\alpha_m)$ es un vector de parámetros.

Para este problema se considera el Lagrangiano:

$$L(x,\lambda,\alpha) = f(x,\alpha) + \sum_{j=1}^{\ell} \lambda_j\, g_j(x,\alpha), \text{ para cada } x \in S,$$

siendo $\lambda_j \ge 0$, para todo $j = 1,\ldots,\ell$.

Su programa dual minimax se define como:

$$(D)\quad \begin{cases} \text{MAX } h(\lambda,\alpha) \\ (\lambda,\alpha) \in D \end{cases}$$

, en donde $h(\lambda,\alpha) = \underset{x\in S}{\text{MIN}}\ L(x,\lambda,\alpha)$

$$D = \left\{ (\lambda,\alpha) \in \mathbb{R}^\ell \times \mathbb{R}^m \mid \lambda \ge 0, \text{ existe } \underset{x\in S}{\text{MIN}}\ L(x;\lambda;\alpha) \right\}$$

Para el programa dual, el Lagrangiano será $L_1(\lambda,u,\alpha)$, en donde u serán las variables duales del dual.

La teoría de dualidad minimax que parte de los programas primal y dual que acabamos de introducir tiene mucho interés en Teoría de Optimización. Hemos considerado en los programas (P) y (D) que las funciones dependen del vector de parámetros α. Nos proponemos, a continuación, enriquecer la información que da el teorema de la envolvente con más información procedente del programa dual. Un resultado interesante, que habrá que completar posteriormente, nos lo da el siguiente teorema.

TEOREMA:

Consideremos los programas primal (P) y dual (D), para $\alpha \in A$, conjunto abierto en $\mathbb{R}^m$. Suponemos que todas las funciones que aparecen en ambos programas son de clase C^2.

Si para cada $\alpha \in A$, existen $x = x^*(\alpha)$, $\lambda = \lambda^*(\alpha)$, siendo x^*, λ^* funciones continuamente diferenciables tales que (x,λ) es punto de silla del Lagrangiano del programa primal dado, entonces se verifica que:

i) La función objetivo indirecta del programa primal $\phi(\alpha)$ es igual a la función objetivo indirecta del programa dual $W(\alpha)$.

ii) Supongamos que $x^\circ = x^*(\alpha^\circ)$, $\lambda^\circ = \lambda^*(\alpha^\circ)$. Entonces se verifica que

$$\frac{\partial \phi}{\partial \alpha_k}(\alpha^\circ) = \frac{\partial W}{\partial \alpha_k}(\alpha^\circ) = \frac{\partial L}{\partial \alpha_k}(x^\circ,\lambda^\circ,\alpha^\circ) = \frac{\partial L_1}{\partial \alpha_k}(\lambda^\circ,u^\circ,\alpha^\circ).$$

Demostración:

i) Sea $\alpha \in A$. Por ser $(x = x^*(\alpha),\ \lambda = \lambda^*(\alpha))$ punto de silla del Lagrangiano del programa dado, sabemos que se cumple que:

$$\begin{cases} x = x^*(\alpha) \text{ es la solución óptima del primal} \\ \lambda = \lambda^*(\alpha) \text{ es la solución óptima del dual} \\ f(x,\alpha) = h(\lambda,\alpha) \end{cases}$$

Entonces:

$\phi(\alpha) = f(x^*(\alpha),\alpha) = f(x,\alpha) = h(\lambda,\alpha) = h\ [\lambda^*(\alpha),\alpha) = W(\alpha)$

con lo que queda demostrado i).

ii) Aplicando el teorema de la envolvente al primal:

$$\frac{\partial \phi}{\partial \alpha_k}(\alpha^o) = \frac{\partial L}{\partial \alpha_k}(x^o,\lambda^o,\alpha^o).$$

Aplicando el teorema de la envolvente al dual:

$$\frac{\partial W}{\partial \alpha_k}(\alpha^o) = \frac{\partial L_1}{\partial \alpha_k}(\lambda^o,u^o,\alpha^o)$$

Pero, por ser $\phi(\alpha) = W(\alpha)$, $\forall\ \alpha \in A$, resulta que

$$\frac{\partial \phi}{\partial \alpha_k} = \frac{\partial W}{\partial \alpha_k}, \quad \forall\ \alpha \in A$$

De donde:

$$\frac{\partial \phi}{\partial \alpha_k}(\alpha^o) = \frac{\partial L}{\partial \alpha_k}(x^o,\lambda^o,\alpha^o) = \frac{\partial W}{\partial \alpha_k}(\alpha^o) = \frac{\partial L_1}{\partial \alpha_k}(\lambda^o,u^o,\alpha^o)$$

como se quería demostrar.

En el teorema anterior partimos de la hipótesis de que para $\alpha \in A$, (x,λ) es punto de silla del Lagrangiano del programa dado, cuya definición no resulta muy manejable, por lo que necesitamos aquí algunas condiciones de punto de silla del Lagrangiano.

Condiciones de punto de silla del Lagrangiano

Sea $\alpha^\circ \in A$. Sean $x^\circ \in S$, $\lambda^\circ \geq 0$.

Consideremos el programa (P), con $\alpha = \alpha^\circ$.

(1) (x°,λ°) es un punto de silla del Lagrangiano del programa dado si y sólo si:

a) x° minimiza $L(x,\lambda^\circ,\alpha^\circ)$, para todo $x \in S$.

b) $g_j(x^\circ,\alpha^\circ) \leq 0$, para $j = 1,\ldots,\ell$.

c) $\lambda_j^\circ.g_j(x^\circ,\alpha^\circ) = 0$, para $j = 1,\ldots,\ell$.

(véase Wismer y Chattergy).

(2) (x°,λ°) es un punto de silla del Lagrangiano del programa dado si y sólo si:

a) x° resuelve el problema primal.

b) λ° resuelve el problema dual.

c) $f(x^\circ,\alpha^\circ) = h(\lambda^\circ,\alpha^\circ)$.

(véase Wismer y Chattergy).

(3) Si el programa primal es convexo se verifica que:

(x°,λ°) es un punto de silla del Lagrangiano del del programa dado si y sólo si verifica las condiciones de Kuhn-Tucker.

(véase Franklin).

3.3. Aplicación a Programación Lineal

Consideremos el programa lineal:

$$\begin{cases} \text{MIN } cx \\ \text{sujeto a:} \quad Ax = b \\ \qquad\qquad\quad x \geq \delta \end{cases}$$

, en donde x es un vector columna n-dimensional.

c es un vector fila n-dimensional.

A es una matriz $m \times n$, con $m < n$.

b es un vector columna m-dimensional, con $b > 0$.

δ es un vector columna de parámetros n-dimensional

Para $\delta = 0$, se trata de un programa lineal en forma estándar.

El programa dual es, en este caso:

$$\begin{cases} \text{MAX } wb + v\delta \\ \text{sujeto a:} \quad wA + v = c \\ \qquad\qquad\quad v \geq 0 \end{cases}$$

Para $\delta = 0$, es decir para el programa lineal en forma estándar, sea x^o la única solución óptima, que suponemos es no degenerada.

Sea $x^o = (x_B{}^o, x_N{}^o) = (B^{-1}b, 0)$

$A = (B, N)$

$c = (c_B, c_N)$

El valor objetivo óptimo es $c_B B^{-1}b = w^o b$, siendo $w^o = c_B B^{-1}$; $v_B{}^o = 0$; $v_N{}^o = c_N - c_B B^{-1}N$ (véase Bazaraa y Jarvis)

Vamos a hacer análisis de sensibilidad del problema lineal estándar, al variar el parámetro δ. Para ello vamos a utilizar el teorema que hemos demostrado en el apartado anterior.

Para $\delta = 0$, tenemos la solución única x^{o}.

Para δ variando alrededor de 0, supongamos que obtenemos como solución:

$$x = x^{*}(\delta), \qquad \text{con } x^{o} = x^{*}(0)$$

$$w = w^{*}(\delta), \qquad \text{con } w^{o} = w^{*}(0)$$

$$v = v^{*}(\delta), \qquad \text{con } v^{o} = v^{*}(0).$$

Sea $\phi(\delta) = c\ x^{*}(\delta) = w^{*}(\delta)\ b + v^{*}(\delta)\ \delta = W(\delta)$

$L(x,w,v,\delta) = cx + w[b - Ax] + v[\delta - x], \quad \text{con } v \geq 0$

$L_1(w,v,x,y,\delta) = wb + v\delta + x[c - wA - v] + yv, \quad \text{con } y \geq 0.$

Aplicamos el teorema de la envolvente ampliando con el dual, tal como hemos demostrado anteriormente:

$$\frac{\partial\phi}{\partial\delta_i} = \sum_{r=1}^{n} c_r \frac{\partial x_r^{*}(\delta)}{\partial\delta_i} = \sum_{m=1}^{m} \frac{\delta w_m^{*}(\delta)}{\partial\delta_i} b_m + v_i^{*}(\delta) +$$

$$+ \sum_{r=1}^{n} \frac{\partial v_r^{*}(\delta)}{\partial\delta_i} \delta_r = \frac{\partial W}{\partial\delta_i} = \frac{\partial L}{\partial\delta_i} = \frac{\partial L_1}{\partial\delta_i} = v_i$$

Particularizando en $\delta = 0$, queda:

$$\frac{\partial\phi}{\partial\delta_i}(0) = \sum_{r=1}^{n} c_r \frac{\delta x_r^{*}(0)}{\partial\delta_i} = v_i^{o} + \sum_{m=1}^{m} \frac{\delta w_m^{*}(0)}{\partial\delta_i} b_m = v_i^{o}$$

De ahí se deduce:

1º) $v_i^o = \frac{\partial \phi}{\partial \delta_i}(0)$, que es la expresión que interpreta el valor en el óptimo del multiplicador asociado a la i-ésima restricción desigualdad del problema dado.

2º) $$0 = \sum_{s=1}^{m} \frac{\partial w_s^*(0)}{\partial \delta_i} b_s = \frac{\partial w^*}{\partial \delta_i}(0)\ b.$$

Como esto se cumple para cualquier vector b > 0, de ahí se deduce que $\frac{\partial w^*}{\partial \delta_i}(0) = 0$ para cada i, por lo que $\frac{\partial w^*}{\partial \delta}(0) = 0$, lo que quiere decir que al modificar ligeramente δ desde 0, las variables duales w, que son los multiplicadores asociados a las restricciones de igualdad no se modifican.

3º) $$v_i^o = \sum_{r=1}^{n} c_r \frac{\partial x_r^*(0)}{\partial \delta_i} = c \frac{\partial x^*}{\partial \delta_i}(0), \quad \text{para cada } i = 1,\dots,n.$$

De donde: $v^o = c \frac{\partial x^*}{\partial \delta}(0)$

Descomponiendo entre partes básicas y no básicas, queda:

$v_B^o = c \frac{\partial x^*}{\partial \delta_B}(0) = 0 \Rightarrow \frac{\partial x^*}{\partial \delta_B}(0) = 0$. Este resultado también resulta intuitivamente claro. Como la solución óptima es no degenerada, los valores óptimos de todas las variables básicas son estrictamente positivos, por lo que una variación infinitesimal alrededor de cero de las exigencias de restricción en las variables básicas no influirá en la solución del problema, por lo que la variación de la solución óptima será cero.

$$v_N^\circ = c \frac{\partial x^*}{\partial \delta_N}(0) = c_B \frac{\partial x_B^*}{\partial \delta_N}(0) + c_N \frac{\partial x_N^*}{\partial \delta_N}(0) = c_N - c_B B^{-1}N,$$

de donde se deduce que:

$$\frac{\partial x_N^*}{\partial \delta_N}(0) = I \quad y \quad \frac{\partial x_B^*}{\partial \delta_N}(0) = - B^{-1}N.$$

Interpretamos dichos resultados:

$\frac{\partial x_N^*}{\partial \delta_N}(0) = I$. La variación en el valor de δ_j alrededor de de 0, para j variable no básica influirá en el cambio de esa variable no básica pero en ninguna más.

Si se modifica ligeramente δ_j, el valor óptimo de dicha variable será dicho valor δ_j, para satisfacer la restricción si $\delta_j > 0$, y para seguir siendo no básica si $\delta_j < 0$.

$\frac{\partial x_B^*}{\partial \delta_N}(0) = - B^{-1}N$. También tiene una clara interpretación.

Veamos:

$Ax = b$, de donde $Bx_B + Nx_N = b \Rightarrow x_B = B^{-1}b - B^{-1}Nx_N$

Sabemos que $x_B^\circ = B^{-1}b$ y que $x_N^\circ = 0$.

Pero si se modifica ligeramente δ_N, correspondiente a las variables no básicas, desde 0, hemos visto que entonces x_N se modificará pasando de 0, al valor que tome δ_N, por lo que entonces x_B se modificará y, será

$$\frac{\partial x_B}{\partial \phi_N} = \frac{\partial x_B}{\partial x_N} = - B^{-1} N.$$

Observación:

De manera análoga al caso estudiado se pueden hacer análisis de sensibilidad para otros casos como: variación del vector de coeficientes de la función objetivo alrededor de c, variación del vector del lado derecho de las restricciones de igualdad alrededor de c, o variación del vector de coeficientes de la restricción s de igualdad alrededor del vector fila s de la matriz A.

3.4. <u>Aplicación a Programación Geométrica</u>

La Programación Geométrica fue introducida por Duffin, Peterson y Zener en 1967.

Un programa geométrico es aquel en el que la función objetivo es una suma de "posinomios" y en el que hay restricciones de desigualdad, cada una de las cuales es de la forma: suma de posinomios ≤ 1, siendo $x > 0$.

Un posinomio es una expresión de la forma:

$$c\ x_1^{a_1}\ x_2^{a_2}\ \ldots\ldots\ x_n^{a_n}\ ,$$ siendo c positivo.

La Programación Geométrica nos interesa porque normalmente trabaja con el programa dual.

Se llama grado de dificultad de un programa geométrico a N - n - 1, en donde N es el nº de posinomios que aparecen en el programa y n es el número de variables. Los programas geométricos con grado de dificultad cero resultan muy sencillos de resolver, ya que en ese caso el programa dual sólo tiene una solución factible que es por tanto su solución óptima, de donde se puede deducir la solución óptima del primal. Algunos problemas de interés tienen grado de dificultad cero.

El teorema fundamental de programación geométrica dice:

Supongamos que algún vector x satisface las restricciones primales $x > 0$, $p_i(x) < 1$, para $i = 1,2,\dots,m$.

Supongamos que el problema primal tiene solución. Entonces el problema dual tiene solución y se verifica que:

MIN valor objetivo óptimo del primal	=	MAX valor objetivo óptimo del dual

(véase Franklin).

<u>Ejemplo</u>:

Consideremos una función de producción de Cobb-Douglas:

$$f(x_1,x_2,\dots,x_n) = A\, x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n} \;, \text{ con } \sum_{i=1}^{n} \alpha_i = 1,\; A > 0$$

Sean w_1, w_2, ..., w_n los precios respectivos de los factores de producción.

Resolver el problema de minimización de costes, si la cantidad producida debe ser mayor o igual que una cantidad dada Q. (siendo $Q > 0$).

El problema es:

$$\begin{cases} \text{MIN} \quad w_1x_1 + w_2x_2 + \ldots + w_nx_n \\ \text{sujeto a:} \quad A\, x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} \geq Q, \text{ con } \sum_{i=1}^{n} \alpha_i = 1 \end{cases}$$

Además por los datos del problema $x > 0$.

El problema lo podemos expresar:

$$\begin{cases} \text{MIN } w_1x_1 + w_2x_2 + \ldots + w_nx_n \\ \text{sujeto a :} \quad \dfrac{Q}{A} x_1^{-\alpha_1} x_2^{-\alpha_2} \ldots x_n^{-\alpha_n} \leq 1, \quad \text{con } \sum_{i=1}^{n} \alpha_i = 1 \\ \qquad\qquad x_1 > 0, \ \ldots\ldots, \ x_n > 0 \end{cases}$$

Es un programa geométrico y su grado de dificultad es cero.

Su dual es:

$$\begin{cases} \text{MAX} \left(\dfrac{w_1}{\lambda_1}\right)^{\lambda_1} \left(\dfrac{w_2}{\lambda_2}\right)^{\lambda_2} \cdots \left(\dfrac{w_n}{\lambda_n}\right)^{\lambda_n} \left(\dfrac{Q}{A\lambda_{n+1}}\right)^{\lambda_{n+1}} (\mu_1)^{\mu_1} \\ \text{sujeto a:} \quad \begin{array}{rcl} \lambda_1 \qquad\qquad\qquad -\alpha_1\lambda_{n+1} & = & 0 \\ \lambda_2 \qquad\qquad -\alpha_2\lambda_{n+1} & = & 0 \\ \vdots & & \\ \lambda_n - \alpha_n\lambda_{n+1} & = & 0 \\ \lambda_1 + \lambda_2 + \ldots + \lambda_n & = & 1 \\ \mu_n & = & \lambda_{n+1} \end{array} \end{cases}$$

$$\lambda_i \geq 0, \text{ para } i = 1,\ldots,n+1$$

Como $\alpha_1 + \ldots + \alpha_n = 1$, sumando las n primeras ecuaciones, obtenemos:

$$\lambda_1 + \ldots\ldots + \lambda_n = \lambda_{n+1}$$

$$\lambda_1 + \ldots\ldots + \lambda_n = 1$$

De donde: $\lambda_{n+1} = 1,\ \lambda_1 = \alpha_1, \ldots,\ \lambda_n = \alpha_n,\ \mu_1 = 1$

Aplicamos el teorema fundamental de Programación Geométrica y a continuación el teorema que hemos establecido en 3.2.

$$\phi(w_1,\ldots,w_n,\alpha_1,\ldots,\alpha_n,A,Q) = W(w_1,\ldots,w_n,\alpha_1,\ldots,\alpha_n,A,Q) =$$

$$= \left(\frac{w_1}{\alpha_1}\right)^{\alpha_1}\left(\frac{w_2}{\alpha_2}\right)^{\alpha_2}\cdots\cdots\left(\frac{w_n}{\alpha_n}\right)^{\alpha_n}\left(\frac{Q}{A}\right)$$

De donde:

$$\frac{\partial\phi}{\partial w_1} = \frac{\partial L}{\partial w_1} = x_1 = \alpha_1\left(\frac{w_1}{\alpha_1}\right)^{\alpha_1-1}\frac{1}{\alpha_1}\left(\frac{w_2}{\alpha_2}\right)^{\alpha_2}\cdots\cdots\left(\frac{w_n}{\alpha_n}\right)^{\alpha_n}\left(\frac{Q}{A}\right) =$$

$$= \left(\frac{Q}{A}\right)\left(\frac{w_1}{\alpha_1}\right)^{\alpha_1-1}\left(\frac{w_2}{\alpha_2}\right)^{\alpha_2}\cdots\cdots\left(\frac{w_n}{\alpha_n}\right)^{\alpha_n}$$

$$\frac{\partial\phi}{\partial w_2} = \frac{\partial L}{\partial w_2} = x_2 = \alpha_2\left(\frac{w_1}{\alpha_1}\right)^{\alpha_1}\left(\frac{w_2}{\alpha_2}\right)^{\alpha_2-1}\frac{1}{\alpha_2}\left(\frac{w_3}{\alpha_3}\right)^{\alpha_3}\cdots\left(\frac{w_n}{\alpha_n}\right)^{\alpha_n}\left(\frac{Q}{A}\right) =$$

$$= \left(\frac{Q}{A}\right)\left(\frac{w_1}{\alpha_1}\right)^{\alpha_1}\left(\frac{w_2}{\alpha_2}\right)^{\alpha_2-1}\left(\frac{w_3}{\alpha_3}\right)^{\alpha_3}\cdots\left(\frac{w_n}{\alpha_n}\right)^{\alpha_n}$$

$$\vdots$$

$$\frac{\partial\phi}{\partial w_n} = \frac{\partial L}{\partial w_n} = x_n = \left(\frac{Q}{A}\right)\left(\frac{w_1}{\alpha_1}\right)^{\alpha_1}\left(\frac{w_2}{\alpha_2}\right)^{\alpha_2}\left(\frac{w_3}{\alpha_3}\right)^{\alpha_3}\cdots\left(\frac{w_{n-1}}{\alpha_{n-1}}\right)^{\alpha_{n-1}}\left(\frac{w_n}{\alpha_n}\right)^{\alpha_n-1}$$

3.5. Aplicación a Programación Dinámica

En el ejemplo anterior el teorema de la envolvente junto con la información procedente del programa dual nos servía para resolver el problema, a diferencia de los casos anteriores en que utilizábamos el teorema para hacer análisis de sensibilidad. De la misma forma en la siguiente aplicación utilizaremos el teorema de la envolvente (sin complemento de programa dual) para resolver un problema básico de Programación Dinámica.

Consideramos el problema de control óptimo lineal cuadrático en tiempo discreto, con horizonte temporal infinito:

$$\begin{cases} \text{MIN} \quad \sum_{k=0}^{\infty} \alpha^k \, (x_k' \, Q \, x_k + u_k' \, R \, u_k) \quad , \text{ siendo } \alpha > 0 \\ \quad , \text{ con } x_{k+1} = Ax_k + Bu_k \; , \quad \text{para } k = 0,1,2,\ldots \end{cases}$$

, en donde, para cada k :

x_k es un vector de estado (n-dimensional)

u_k es un vector de controles (m-dimensional)

Q es una matriz simétrica semidefinida positiva,

R es matriz simétrica definida positiva,

A y B son matrices dadas.

Sabemos por Programación Dinámica que la solución del problema dado se reduce a la de la siguiente ecuación funcional:

$$V(x) = \underset{u}{\text{MIN}} \{x'Q\,x + u'R\,u + \alpha\, V(Ax + Bu)\}$$

Este problema está resuelto y la solución aparece en los libros de Teoría de Control. Nosotros vamos a resolver el problema por otro método, utilizando el teorema de la envolvente.

Consideramos la ecuación funcional dada:

Condición necesaria de mínimo respecto a u: derivada de lo que hay dentro de la llave, igual a cero:

$$2\ Ru + \alpha B'\ \nabla V(Ax + Bu) = 0$$

Condición dada por el teorema de la envolvente:

$$\nabla V(x) = \underbrace{2\ Qx + \alpha A'\ \nabla V(Ax + Bu)}_{\text{(derivada respecto a x de la expresión entre llaves).}}$$

Suponemos ahora que, para cierta matriz simétrica K, es

$$V(x) = x'K\ x, \quad \Rightarrow \quad \nabla V(x) = 2Kx$$

Entonces la primera ecuación queda:

$$2\ Ru + 2\ \alpha B'K(Ax + Bu) = 0$$

$$Ru + \alpha B'KAx + \alpha B'KBu = 0$$

$$(R + \alpha B'KB)u = -\ \alpha B'KAx$$

$$\Rightarrow \quad \mathbf{u = -\ \alpha\ (R + \alpha B'KB)^{-1}\ B'KAx}$$

que nos da el control óptimo.

Operando en la segunda ecuación:

$$2\ Kx = 2\ Qx + 2\ \alpha A'K(Ax + Bu)$$

$$Kx = Qx + \alpha A'KAx + \alpha A'KB\ \{-\alpha(R + \alpha B'KB)^{-1}\ B'KAx\}$$

$$Kx = [Q + \alpha A'KA - \alpha^2 A'KB\ (R + \alpha B'KB)^{-1}\ B'KA]x$$

$$\Rightarrow \quad K = Q + A'[\alpha K - \alpha^2 KB\,(R + \alpha B'KB)^{-1}\,B'K]\,A$$

que es una ecuación de Riccati.

La solución obtenida es la conocida en la literatura (véase Bertsekas).

4. CONCLUSIONES

El teorema de la envolvente se conoce en la literatura económica desde hace cerca de cincuenta años y todavía siguen apareciendo nuevas aplicaciones asi como otras generalizaciones, como hemos señalado en la introducción, lo que da idea de su potencia e interés. En la literatura sobre Matemáticas y sobre Teoría de la Optimización para matemáticos e ingenieros este teorema no es utilizado, probablemente porque sea desconocido.

En este trabajo se extiende el teorema de la envolvente en dos direcciones: a) El caso de cierto tipo bastante utilizado de funciones objetivo compuestas. b) Aprovechando la información que proporciona el programa dual minimax del programa dado. A continuación se aplican los resultados a Programación Lineal y Programación Geométrica, lo que sirve para comprobar que el teorema puede servir tanto para hacer análisis de sensibilidad como para facilitar el cálculo de soluciones óptimas. Esta segunda faceta del teorema de la envolvente se confirma también en un problema de control óptimo en tiempo discreto, con horizonte temporal infinito.

Señalamos finalmente algunas ideas y preguntas referentes a distintas posibilidades sobre la continuación de la investigación comenzada.

- Análisis de sensibilidad en programación no lineal.

- Optimización dinámica en tiempo discreto. Análisis de sensibilidad. ¿Puede desarrollarse una metodologia para el caso discreto similar a la de Caputo y La France-Barney?.

- Profundizar en el caso de problemas con incertidumbre (en la literatura hay ya algunos resultados).

- En Optimización hay muchos algoritmos que utilizan el programa dual para resolver el primal. ¿Puede aportar algo en dichos algoritmos el resultado que hemos obtenido en la sección 3.2.?.

- Profundización y continuación de los trabajos de Caputo y La France-Barney. ¿Qué pasa si se cambian algunos de sus supuestos por otros?.

REFERENCIAS

Barbolla, Cerdá, Sanz (1991). Optimización matemática: teoría, ejemplos y contraejemplos. Ed. Espasa Calpe.

Bazaraa, Jarvis (1977). Linear programming and network flows. J. Wiley.

Bertsekas, D. (1987). Dynamic Programming. Deterministic and stochastic models. Prentice Hall.

Caputo, M.R. (1990). How to do comparative dynamics on the back of an envelope in optimal control theory. Journal of Economic Dynamics and Control. Vol. 14, Nº 3/4, págs. 655-683.

Chiang, A. (1987). Métodos fundamentales de Economía Matemática. 3ª edición. Mc Graw Hill.

Duffin, Peterson y Zener (1967). Geometric Programming. J. Wiley.

Frankling, J. (1980). Methods of Mathematical Economics. Springer Verlag.

La France, Barney (1991). The envelope theorem in dynamic optimization. Journal of Economic Dynamics and Control. Vol. 15, Nº 2, págs. 355-385.

Samuelson, P. (1947). Foundations of Economic Analysis. Harvard University Press.

Silberberg, E. (1971). "The Le Châtelier Principle as a Corollary to a Generalized Envelope Theorem". Journal of Economic Theory, 3; págs. 146-155.

Silberberg, E. (1974). "A Revision of Comparative Statics Methodology in Economics, or, How to Do Economics on the Back of an Envelope". Journal of Economic Theory, 7. págs. 159-172.

Silberberg, E. (1978). The structure of economics: a mathematical analysis. First edition. Mc Graw Hill.

Silberberg, E. (1990). The structure of economics: a mathematical analysis. Second edition. Mc Graw Hill.

Viner, J. (1932). "Cost Curves and Supply Curves". Zeitschrift fur nationalokonomie,3; 1932. Reprinted in Readings in Price Theory (AEA).

Wismer, Chattergy (1978). Introduction to nonlinear optimization. North Holland.