UNIVERSIDAD DE EL SALVADOR

Facultad de Ingeniería y Arquitectura

Departamento de Matemática

OPTIMIZACION NO LINEAL

TRABAJO DE GRADUACION PRESENTADO POR:

JUAN HAROLDO LINARES MARTINEZ

PARA OPTAR AL TITULO DE:

LICENCIADO EN MATEMATICA

ABRIL DE 1991

San Salvador, El Salvador, Centro América.

UNIVERSIDAD DE EL SALVADOR

RECTOR: DR. JOSE BENJAMIN LOPEZ GUILLEN

SECRETARIO GENERAL: DRA. GLORIA ESTELA GOMEZ DE PEREZ

FACULTAD DE INGENIERIA Y ARQUITECTURA

DECANO: ING. JOAQUIN ALBERTO VANEGAS

SECRETARIO: ING. MARIO ARNOLDO MOLINA ARGUETA

DEPARTAMENTO DE MATEMATICA

JEFE EN FUNCIONES: ING. JOAQUIN ALBERTO VANEGAS

****

UNIVERSIDAD DE EL SALVADOR

TRABAJO DE GRADUACION

COORDINADOR: ING. CARLOS MAURICIO CANJURA LINARES

ASESOR: LIC. FRANCISCO RIVERA ZAVALETA

*****

DEDICATORIA

- A DIOS TODOPODEROSO

- A MI ABUELA:	Clementina Martínez
- A MIS PADRES:	Manuel de Jesús Linares Juana Francisca Martínez de Linares
- A MIS HERMANOS:	Manuel de Jesús, Morena Emilia y José Guillermo.
- A MIS PROFESORES:	Por las enseñanzas proporcionadas
- A MIS ALUMNOS:	Porque también de ellos se aprende.

AGRADECIMIENTO

- ING. CARLOS MAURICIO CANJURA LINARES
- LIC. FRANCISCO RIVERA ZAVALETA
- SRA. LETICIA QUIÑONEZ
- RUTH LORENA MARTINEZ

Por la paciencia mostrada hacia mi persona y el tiempo dedicado.

****

I N D I C E

	PAGINA
CONTENIDO:	
GENERALIDADES MATEMATICAS	1
Definiciones Básicas	1
Funciones de una Variable	6
Funciones de N-Variables	8
- El Teorema de Taylor	9
- El Teorema de la Función Implícita	11
- Problemas sin restricciones	17
- Problemas con restricciones	27
- Restricciones de Desigualdad	45
PROGRAMACION CUADRATICA	59
Algoritmo de Wolfe	60
Algoritmo de Beale	68
Algoritmo de Theil y Van der Panne	74
PROGRAMACION CONVEXA	81
Problema Primal	87
Problema Dual	87
Técnica de Minimización Secuencial	90

PAGINA

PROGRAMACION DINAMICA 100

Composición y El Principio de Optimalidad 107

Procesos de Etapas Finitas 112

Procesos de Infinitas Etapas 135

Aspectos Computacionales 140

ANEXOS 152

*****

GENERALIDADES MATEMATICAS

En esta parte, se hará un desarrollo acerca de todas las ideas matemáticas básicas que son necesarias para la resolución de problemas de optimización lineal y no lineal.

Esta parte del trabajo, se ha dividido en tres secciones, las cuales se describen a continuación:

- La primera sección está constituida por un conjunto de definiciones de conceptos matemáticos básicos.

- La segunda sección, está conformada por toda una serie de resultados que se han obtenido para el estudio de extremos en funciones de una variable; en esta sección, se han obviado las demostraciones de los resultados, ya que tales demostraciones, pueden ser estudiadas con mayor facilidad en los libros de cálculo superior.

- La tercera sección está constituida por el desarrollo de las condiciones necesarias y suficientes para extremos de funciones de más de una variable con o sin restricciones; en esta sección, se realizará un desarrollo más profundo ya que ella será la base para nuestro trabajo.

. Definiciones Básicas.

Continuidad de una función: una función de n variables $x_1, x_2, \ldots x_n$ es contínua en el punto $x_1^o, x_2^o, \ldots\ldots, x_n^o$ si para todo $\varepsilon > 0$ existen $\delta_1, \delta_2, \ldots, \delta_n$ tales que para $|h_i| < \delta_i$

$$|f(x_1^o + h_1, x_2^o+h_2 \ldots\ldots x_n^o + h_n) - f(x_1^o, x_2^o, \ldots\ldots, x_n^o)| < \varepsilon$$

donde

$$\delta_i > 0$$

Diferenciabilidad de una función: una función de n variables $x_1, x_2, \ldots\ldots, x_n$ es diferenciable en el punto $X^o = (x_1^o, x_2^o, \ldots x_n^o)$, si existe la transformación lineal $L: R^n \rightarrow R$ y es tal que la función $T: R^n \longrightarrow R$ tq $T(X) = f(X^o)+L(X-X^o)$ donde $X = (x_1, x_2, \ldots, x_n)$, satisface que:

$$\lim_{x\to X^o} \frac{|f(X) - T(X)|}{||X - X^o||} = 0$$

La transformación lineal L se llama "La derivada de f" en X^o y se denota por $D_X{}^o f$.

Si f es una función de R^n en R, la cual es diferenciable en X^o, entonces:

$$D_X{}^o f = L: R^n \rightarrow R \quad \text{t.q.} \quad L(X) = \left.\frac{\partial f}{\partial x_1}\right|_{X^o} x_1 + \left.\frac{\partial f}{\partial x_2}\right|_{X^o} x_2 + \ldots + \left.\frac{\partial f}{\partial x_n}\right|_{X^o} x_n$$

Mínimo y Máximo Absoluto: Una función f de n variables, tiene un mínimo absoluto en $x \in R^n$, si: $f(x) - f(x^*) \geq 0$, para toda x en la cual la función está definida. El mínimo absoluto a menudo es llamado mínimo global.

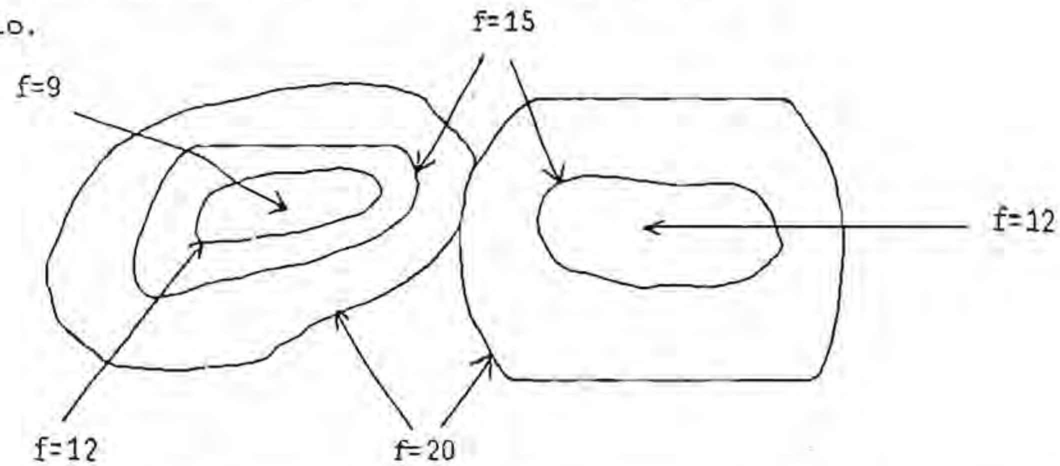
Para obtener la definición de máximo absoluto, sólo se le cambia el sentido a la desigualdad.

<u>Mínimo y Máximo Relativo Propio</u>: una función f de n variables tiene un mínimo relativo propio en $x \in R^n$ si existe un $d > 0$ tal que para toda $h \in R^n$, excepto posiblemente en $h = 0$.

$$f(x^* + h) - f(x^*) > 0 \text{ para } |h_i| < d \quad i: 1,2,\ldots,n$$

para obtener la definición de máximo relativo propio, se le cambia el sentido a la desigualdad.

Ejemplo: A continuación se muestra el gráfico de las curvas de nivel de una función de dos variables, que posee un mínimo relativo propio.



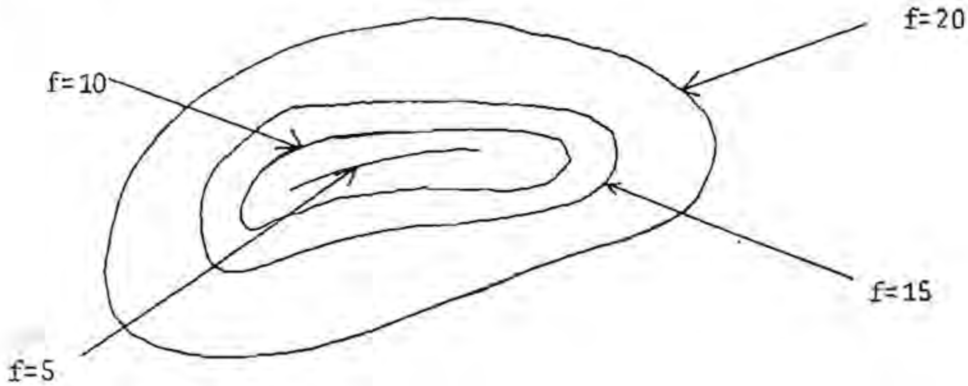
<u>Mínimo y Máximo Relativo Impropio</u>: Una función f de n variables tiene un mínimo relativo impropio en $x^* \in R^n$ si existe un $d > 0$ tal que para toda h

$$f(x^* + h) - f(x^*) \geq 0$$

para $|h_i| < d$ $i = 1,2,\ldots,n$ y existe al menos un $h \neq 0$ para la cual la igualdad se cumple.

La definición de máximo relativo impropio se obtiene cambiando el sentido a la desigualdad.

Ejemplo: A continuación se muestra el gráfico de las curvas de nivel de una función de dos variables que posee un mínimo relativo impropio.



Vectores y Valores Propios de una Matriz: Dada una matriz A de orden nxn, si para algún $\lambda \in R$, existe un $x \in R^n$, $x \neq 0$ tal que:

$$Ax = \lambda x$$

llamamos a λ valor propio de A y a x vector propio de A.

Los autovalores de A son las soluciones de la ecuación $\det(A - \lambda I) = 0$

Si A es una matriz simétrica, se cumple que todos los autovalores son reales.

Forma cuadrática asociada a una matriz:

Dada una matriz simétrica A de orden nxn la forma cuadrática $\psi_A(X)$ asociada a A, se define por:

$$\psi_A(X) = x^t AX \qquad x \in R^n$$

Se dice que:

i) A es definida positiva si $\psi_A(X) > 0$ para todo $x \neq 0$

ii) A es definida negativa si $\psi_A(X) < 0$ para todo $x \neq 0$

iii) A es definida semipositiva si $\psi_A(X) \geq 0$ para todo x

iv) A es definida seminegativa si $\psi_A(X) \leq 0$ para todo X

Otra interpretación viene dada por los valores propios, así:

i) A es definida positiva si, y solo si, todos sus valores propios son mayores que cero.

ii) A es definida negativa si, y solo si, todos sus valores propios son menores que cero.

iii) A es definida semipositiva si, y solo si, todos sus valores propios son mayores o iguales que cero.

iv) A es definida seminegativa si, y solo si, todos sus valores propios son menores o iguales que cero.

Una tercera interpretación es la siguiente;

Sea

$$A_i = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1i} \\ a_{12} & a_{22} & \cdots\cdots & a_{2i} \\ a_{i1} & a_{i2} & \cdots\cdots & a_{ii} \end{bmatrix} \qquad i = 1,2,\ldots\ldots, n$$

y sea $d_i = \det(A_i)$, entonces:

i) A es definida positiva si y solo si, $d_i > 0 \quad i = 1,2,\ldots, n$

ii) A es definida negativa si y solo si $d_1 < 0$, $d_2 > 0$, $d_3 < 0$, $d_4 > 0, \ldots\ldots$

Función Cóncava:

Una función f(x) es cóncava si:

$$f(x + \lambda(z-x)) \geq f(x) + \lambda(f(z)-f(x))$$

para $\lambda \epsilon$ 0, 1; es decir, una función es cóncava si dados dos puntos, x, z; el segmento de recta que une a f(x) con f(z) se encuentra - por debajo del valor de la función f evaluada en todos los puntos de la recta que une a x y z.

Función Convexa:

Una función f(x) es convexa, si:

$$f(x + \lambda(z-x)) \leq f(x) + \lambda(f(z) - f(x))$$

para $\lambda \epsilon$ 0, 1 ; es decir, una función es convexa, si dados dos - puntos cualesquiera x y z, el segmento de recta que une a f(x) con f(z) se encuentra por encima del valor de la función f evaluada en los puntos de la recta que une a x y z.

Funciones de una Variable:

A continuación se proporcionan algunos resultados obtenidos para el estudio de extremos de una función de una variable; las demostraciones de estos resultados se han obviado, pero si, se proporcionan algunos ejemplos ilustrativos.

Condición Necesaria: Sí f(x) es una función contínua que, además, tiene primera derivada continua en el intervalo $a < x < b$. Entonces, la condición necesaria para que f(x) posea un extremo en el punto x^o es que $f'(x) = 0$; es decir, que la primera derivada se anule en el punto x^o

A los puntos donde la primera derivada se anula se les llama puntos críticos o puntos estacionarios.

<u>Condición Suficiente:</u> Si f(x) y sus primeras derivadas son contínuas, entonces f(x) tiene un punto extremo en x^o si, y solo si, el orden de la primera derivada que no se anula en x^o es par. Además, la función f(x) tendrá un máximo, si dicha derivada evaluada en x^o es negativa y tendrá un mínimo si la evaluación es positiva.

Ejemplo:

Buscar los puntos extremos de la función: $f(x) = (x-5)^{10}$

Solución:

Aplicando la condición necesaria, se tiene:

$$\frac{d\ f(x)}{dx} = 10(x-5)^9 = 0$$

ésto es verdadero cuando x=5; es decir, la función tiene un punto crítico en x=5.

Ahora, al utilizar la condición suficiente, se tiene que, la primera derivada que no se anula en x=5, es:

$$\left.\frac{d^{10} f(x)}{dx}\right|_{x=5} = 10! > 0$$

por lo tanto f(x) posee un mínimo en x=5.

Existen algunas funciones en las cuales los criterios anteriores fallan, en tales casos, se deberá considerar la siguiente condición:

<u>Condición suficiente alternativa</u>: Si la función f(x) está definida en un intervalo que contenga al punto x^o, es diferenciable continuamente en cualquier punto del intervalo, excepto posiblemente en x^o mismo, y si f'(x) es nula en un número finito de puntos; entonces, f(x) tiene un extremo propio en el punto x^o si, y solo si, el punto x^o separa al intervalo original en dos subintervalos en los cuales f'(x) tiene diferente signo. En particular, la función tiene un mínimo si la derivada es negativa a la izquierda de x^o y positiva a la derecha; si ocurre la situación inversa, f(x) tiene un máximo.

Ejemplos: Buscar los puntos extremos de la función:

$$f(x) = e^{-|x|}$$

Solución:

Se tiene que $$f(x) = \begin{cases} e^{-x} & \text{si } x \quad 0 \\ e^{x} & \text{si } x \quad 0 \end{cases}$$

resulta fácil verificar que la primera derivada no existe en el punto x = 0; además, se tiene que: a la izquierda de cero, la primera derivada $f'(x) = e^{x}$ resulta ser positiva y a la derecha de cero, la derivada $f'(x) = -e^{-x}$ resulta ser negativa. Por lo tanto, aplicando el criterio anterior, se puede concluir que f(x) tiene un máximo en x=0.

<u>Funciones de N-Variables</u>:

A continuación se muestra el desarrollo para la obtención de las condiciones necesarias y suficientes para extremos de una función de n-va riables sín y con restricciones. Previo a ello, se darán dos resultados que serán de suma importancia en el desarrollo posterior, tales resultados son: El teorema de Taylor y el Teorema de la Función Implícita.

El Teorema de Taylor: La importancia de este teorema radica en que se puede aproximar una función dada f(x) mediante un polinomio de orden n; es decir, se puede aproximar el valor de una función en $x^* + H$, donde $x^* \in R^n$ y $H \in R^n$, a partir de la información de la función y sus primeras n+1 derivadas evaluadas en x^*. El teorema tiene la siguiente forma:

Teorema: Si se tiene una función de n variables f(x) cuyas primeras n+1 derivadas parciales son contínuas. Entonces $f(x^* + h)$ puede ser aproximado mediante la siguiente fórmula:

$$f(x^* + h) = f(x^*) + G_n(x^*) + R_n(x^* + h, x^*)$$

donde:

$G_n(x^*)$ es un polinomio de orden n en la variable h

$R_n(x^*+h, x^*)$ es el término residuo de la fórmula de Lagrange.

Además:

$$G_n(x^*) = \sum_{i=1}^{n} \frac{1}{i} D^{i}_{x^*} f(h)$$

$$R_n(x^*+h, x^*) = \frac{1}{(n+1)} D^{n+1}_{x^*+\theta h} f(h); \; 0 \leq \theta \leq 1$$

<u>Condición suficiente alternativa:</u> Si la función f(x) está definida en un intervalo que contenga al punto x^o, es diferenciable continuamente en cualquier punto del intervalo, excepto posiblemente en x^o mismo, y si f'(x) es nula en un número finito de puntos; entonces, f(x) tiene un extremo propio en el punto x^o si, y solo si, el punto x^o separa al intervalo original en dos subintervalos en los cuales f'(x) tiene diferente signo. En particular, la función tiene un mínimo si la derivada es negativa a la izquierda de x^o y positiva a la derecha; si ocurre la situación inversa, f(x) tiene un máximo.

Ejemplos: Buscar los puntos extremos de la función:

$$f(x) = e^{-|x|}$$

Solución:

Se tiene que $$f(x) = \begin{cases} e^{-x} & \text{si } x \quad 0 \\ e^{x} & \text{si } x \quad 0 \end{cases}$$

resulta fácil verificar que la primera derivada no existe en el punto x = 0; además, se tiene que: a la izquierda de cero, la primera - derivada $f'(x) = e^x$ resulta ser positiva y a la derecha de cero, la derivada $f'(x) = -e^{-x}$ resulta ser negativa. Por lo tanto, aplicando el criterio anterior, se puede concluir que f(x) tiene un máximo en x=0.

<u>Funciones de N-Variables:</u>

A continuación se muestra el desarrollo para la obtención de las condiciones necesarias y suficientes para extremos de una función de n-variables

riables sin y con restricciones. Previo a ello, se darán dos resultados que serán de suma importancia en el desarrollo posterior, tales resultados son: El teorema de Taylor y el Teorema de la Función Implícita.

El Teorema de Taylor: La importancia de este teorema radica en que se puede aproximar una función dada f(x) mediante un polinomio de orden n; es decir, se puede aproximar el valor de una función en $x^* + H$, donde $x^* \in R^n$ y $H \in R^n$, a partir de la información de la función y sus primeras n+1 derivadas evaluadas en x^*. El teorema tiene la siguiente forma:

Teorema: Si se tiene una función de n variables f(x) cuyas primeras n+1 derivadas parciales son contínuas. Entonces $f(x^* + h)$ puede ser aproximado mediante la siguiente fórmula:

$$f(x^* + h) = f(x^*) + G_n(x^*) + R_n(x^* + h, x^*)$$

donde:

$G_n(x^*)$ es un polinomio de orden n en la variable h

$R_n(x^*+h, x^*)$ es el término residuo de la fórmula de Lagrange.

Además:

$$G_n(x^*) = \sum_{i=1}^{n} \frac{1}{i} D^i_{x^*} f(h)$$

$$R_n(x^*+h, x^*) = \frac{1}{(n+1)} D^{n+1}_{x^*+\theta h} f(h); \; 0 \leq \theta \leq 1$$

Prueba: La prueba se realizará construyendo una función $Q: R \to R$ a la cual se le puede aplicar el teorema de Taylor en una variable. El segmento de recta que une a x^* con $x^* + h$ se puede definir paramétricamente por la función $g: R \to R^n$ tal que $g(t) = x^* + t.h$ con $0 \leq t \leq 1$. Sea $Q: R \to R$ la función compuesta $Q = f o g$, es decir, Q es la restricción de f al segmento de línea definida por g, así:

$$Q(t) = f(x^* + th) \quad 0 \leq t \leq 1$$

Aplicándole el teorema de Taylor a $Q(t)$ en el intervalo $[0, 1]$ se obtiene

$$Q(1) = Q(o) + Q'(o)(1-o) + \frac{1}{2} Q''(o)(1-o)^2 + \ldots\ldots$$

$$\ldots\ldots + \frac{1}{n} \quad Q^n(o)(1-o)^n + \frac{Q^{n+1}(\theta)}{(n+1)} \quad (1-o)^{n+1} \quad 0 \leq \theta \leq 1$$

se tiene que:

$$Q(1) = f(x^* + h)$$

$$Q(o) = f(x^*)$$

$$\begin{aligned} Q'(t) &= D_t(f o g) \\ &= D_{g(t)} f \; o \; D_t g \\ &= J_f(g_{(t)}) . \; J_g(t) \\ &= J_f(g_{(t)}) . \; g'(t) \\ &= J_f(g_{(t)}) . \; h \\ &= D_{g(t)} \; f \; (h) \end{aligned}$$

Evaluando $Q'(t)$ en $t = 0$ se obtiene:

$$Q'(o) = D_{g(o)} f(h) = D_{x^*} f(h)$$

Además, se tiene que:

$$Q''(t) = (fog)''(t) = (fog)'(t)' = (f'(g(t)).g'(t))'$$

$$= f''(g(t))\, g'(t) + f'(g(t)).g''(t)$$

$$= D_{g(t)} f .\, g'(t) + f'(g(t)).(0)$$

$$= D^2_{g(t)}\, f(h)$$

Evaluando $Q''(t)$ en $t = 0$ se obtiene:

$$Q''(0) = D^2_{g(o)} f(h) = D^2_{x^*}\, f(h)$$

Se puede llegar a demostrar por inducción que:

$$Q^n(t) = D^n_{g(t)}\, f(h)$$

De donde:

$$Q^n(0) = D^n_{g(0)}\, f(h) = D^n_{x^*} f(h)$$

Sustituyendo estos resultados en el desarrollo de Taylor dado anteriormente se obtiene que:

$$f(x^* + h) = f(x^*) + D_{x^*} f(h) + \frac{1}{2}\, D^2_{x^*}\, f(h) + \ldots\ldots$$

$$\ldots. + \frac{1}{n!}\, D^n_{x^*}\, f(h) + \frac{1}{(n+1)!}.\, D^{n+1}_{x^*+\theta h}\, f(h) \qquad 0 \leq \theta \leq 1$$

$$f(x^* + h) = f(x^*) + G_n(x^*) + R_n(x^* + h,\ x^*)$$

L.Q.Q.D.

<u>El teorema de la Función Implícita:</u>

Si se tiene un sistema de m ecuaciones, así:

$$h_i(x) = 0 \qquad i = 1,2,\ldots\ldots, m$$

donde $x = (x_1, x_2, \ldots\ldots, x_n)$ con $n > m$ y las $h_i(x)$ son diferenciables.

Supóngase, además, que se desea saber si se pueden usar estas m ecuaciones para eliminar m de las n variables en el punto x^o. Es decir, se desea saber si existen m funciones ψ_i, tales que:

$$x_i^o = \psi_i(x_{m+1}^o, x_{m+2}^o, \ldots\ldots, x_n^o) \qquad i = 1, 2, \ldots\ldots, m$$

El teorema de la función implícita nos da una condición que si es satisfecha, nos garantiza la existencia y unicidad de estas funciones; además, asegura que estas funciones son contínuas y diferenciables en algún vecindario de x^o.

TEOREMA: Dado un sistema de ecuaciones de la forma

$$h_i(x) = 0 \qquad i = 1, 2,\ldots\ldots, m$$

con $x = (x_1, x_2,\ldots\ldots, x_n)$ $n > m$. Si $h_i(x)$ son diferenciables una condición necesaria y suficiente para la existencia de funciones ψ_i tales que: puedan representar a m de las n variables en términos de las otras m-n, es que el rango de la siguiente matriz evaluada en x^o sea m

$$P_{x^o}(h_1,h_2..h_m) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots\cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots\cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}$$

Prueba

En primer lugar, se definirá un nuevo sistema de ecuaciones, de la siguiente forma:

$$H_i(x) = h_i(x) \qquad \text{si} \qquad i = 1,2,\ldots, m$$

$$H_i(x) = x_i \qquad \text{si} \qquad i = n\text{-}m, \ldots, n$$

Obsérvese que:

$$H_i(x^0) = x_i^0 \qquad \text{si} \qquad i = n\text{-}m, \ldots, n$$

$$H_i(x^0)=h_i(x^0) = 0 \qquad \text{si} \qquad i = 1,2,\ldots, m$$

Además el Jacobiano asociado al sistema de ecuaciones es:

$$\left|\begin{array}{cccc}
\frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots\cdots & \frac{\partial h_1}{\partial x_n} \\
\frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \cdots\cdots & \frac{\partial h_2}{\partial x_n} \\
\vdots & \vdots & & \vdots \\
\frac{\partial h_m}{\partial x_1} & \frac{\partial h_m}{\partial x_2} & \cdots\cdots & \frac{\partial h_m}{\partial x_n} \\
1 & 0 & \cdots\cdots & 0 \\
0 & 1 & \cdots\cdots & 0 \\
\vdots & \vdots & & \\
0 & 0 & \cdots 1 \cdots & 0
\end{array}\right|$$

Es fácil verificar que el determinante de este Jacobiano es diferente de cero; por lo tanto, se puede aplicar el teorema de la función inversa, de donde H_i tendrá una inversa que se denotará por H_i^{-1} la cual, es diferenciable y con primera derivada contínua.

Así, se tiene que:

$$H_i^{-1} = x_i \qquad \text{si } i = n-m, \ldots, n$$

$$H_i^{-1} = g_i(x) \qquad \text{si } i = 1,2,\ldots, m$$

donde:

$g_i(x)$ también es diferenciable continuamente

Definamos las funciones siguientes:

$$\psi_i\,(x_{n-m}, x_{n-m+1}, \ldots, x_n) = g_i(0_1, 0_2, \ldots, 0_m, x_{n-m}, \ldots, x_n)$$

$$i = 1,2,\ldots,m$$

Obsérvese que:

$$\psi_i(x^o_{n-m}, \ldots, x^o_n) = g_i(0_1, 0_2, \ldots 0_m, x^o_{n-m}, \ldots, x^o_n) = x_i$$

$$i = 1,2,\ldots, m$$

ya que

$$g_i(0_1, 0_2, \ldots, 0_m, x^o_{n-m}, \ldots, x^o_n) = H_i^{-1}(0_i) = x_i \qquad i = 1,\ldots,m$$

se tiene que:

$$x_i = H_o H^{-1}(x_i) \qquad i = 1,2,\ldots,n$$

$$= H(g_1(x), g_2(x), \ldots, g_n(x),\; x_{n-m}, x_{n-m+1}, \ldots, x_n)$$

Si $i = n-m, \ldots, n$

$$x_i = x_i \qquad \text{con} \qquad i = n-m, \ldots, n$$

Si $i = 1,2,\ldots, m$

$$x_i = h_i(g_1(x),\, g_2(x), \ldots,\, g_n(x),\, x_{n-m}, \ldots,\, x_n)$$

Si se hace $x_1, x_2, \ldots, x_n = 0$ se tiene que:

$$0_i = h_i(g_1(0_1,\, 0_2, \ldots,\, 0_n,\, x_{n-m}, \ldots,\, x_n), \ldots, g_n(0_1, 0_2, \ldots 0_n, x_{n-m}, \ldots x_n),\, x_{n-m}, \ldots\, x_n)$$

$$0_i = h_i(\psi_1(x_{n-m}, \ldots, x_n), \ldots,\, \psi_n(x_{n-m}, \ldots x_n), x_{n-m}, \ldots,\, x_n)$$

$$i = 1,2,\ldots, m$$

L.Q.Q.D.

El teorema anterior nos garantiza la existencia de las m funciones ψ_i, pero no nos indica cuáles de las m variables pueden ser eliminadas. Ante este caso, el teorema proporciona una condición adicional que ayuda a seleccionar las m variables que pueden ser eliminadas. Dicha condición, se construye de la siguiente forma:

Definiendo el Jacobiano como:

$$J_x(h_1, \ldots\ldots, h_m) = \begin{vmatrix} \dfrac{\partial h_1}{\partial x_1} & \cdots\cdots & \dfrac{\partial h_1}{\partial x_m} \\ \dfrac{\partial h_m}{\partial x_1} & & \dfrac{\partial h_m}{\partial x_m} \end{vmatrix}$$

Entonces, a partir de la matriz P dada anteriormente, se pueden construir $\binom{n}{m}$ Jacobianos diferentes mediante la selección de m de los siguientes vectores columna:

$$\begin{bmatrix} \frac{h_1}{x_j} \\ \vdots \\ \frac{h_m}{x_j} \end{bmatrix} \qquad j = 1,2,\ldots, m$$

Si el Jacobiano asociado con un conjunto de variables, digamos x_p, x_{p+1}, x_{p+m} no es cero; entonces, existirá un conjunto de funciones ψ_i $(i = p, p+1, \ldots, p+m)$ tales que pueden eliminar a las variables x_p, $x_{p+1}, \ldots, x_p$. Es decir:

$$x_p = \psi_p (x_1, x_2, \ldots, x_{p-1}, x_{p+m+1}, \ldots, x_n)$$

$$x_{p+1} = \psi_{p+1} (x_1, x_2, \ldots, x_{p-1}, x_{p+m+1}, \ldots, x_n)$$

$$\vdots$$

$$x_{p+m} = \psi_{p+m} (x_1, x_2, \ldots, x_{p-1}, x_{p+m+1}, \ldots, x_n)$$

Ejemplos:

Considere el sistema de dos ecuaciones siguientes:

$$x_1 + x_2 + x_3 = 0$$

$$2x_1 + 3x_2 + 2x_3 = 0$$

Se pide utilizar el teorema de la función implícita para determinar que variables pueden ser eliminadas en el punto $x^o = (0, 0, 0)$

Solución:

En primer lugar, construyamos la matriz P así:

$$P_{(0,0,0)}(h_1, h_2) = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 2 \end{bmatrix}$$

Ahora, construyamos los $\binom{3}{2} = 3$ posibles Jacobianos.

El primero sería:

$$J(h_1, h_2) = \begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix} = 3 - 2 = 1 \neq 0$$

Por lo tanto, x_1 y x_2 pueden ser puestos en función de x_3.

El segundo sería:

$$J(h_1, h_2) = \begin{vmatrix} 1 & 1 \\ 2 & 2 \end{vmatrix} = 2 - 2 = 0$$

Por lo tanto, x_1 y x_3 no se pueden expresar en términos de x_2.

Y el último:

$$J(h_1, h_2) = \begin{vmatrix} 1 & 1 \\ 3 & 2 \end{vmatrix} = 2 - 3 = -1 \neq 0$$

Por lo tanto, x_2 y x_3 pueden ser expresados en términos de x_1.

Problemas sin Restricciones.

A continuación se mostrarán las condiciones necesarias y suficientes para el estudio de extremos sin restricciones en una función de n variables.

Condición Necesaria: Una condición necesaria para que una función contínua $f(x)$, con primeras y segundas derivadas parciales contínuas, tenga un punto extremo en x^o es que cada una de las derivadas parciales de $f(x)$, evaluadas en x^o, se anulen.

Prueba: (Por contradicción)

Supóngase que la p-ésima derivada parcial no se anula, entonces, por el teorema de Taylor se tendría que:

$$f(x^o+h) - f(x^o) = \sum_{i=1}^{n} hi \left.\frac{\partial f(x)}{\partial x_i}\right|_{x=x^o} + R_1(x^o+\theta h, x^o)$$

$$= h_p \left.\frac{\partial f(x)}{\partial x_p}\right|_{x=x^o} + R_1(x^o+\theta h, x^o)$$

Como $R_1(x^o+\theta h, x^o)$ es de orden h_1^2 ; así, para h_i pequeños los términos de orden h_i dominan a los términos de orden mayor, y por lo tanto:

$$f(x^o + h) - f(x^o)$$

tendrá el mismo signo que los términos de orden h.

Supóngase que:

$$\left.\frac{\partial f(x)}{\partial x_p}\right|_{x=x^o} > 0$$

Entonces, la diferencia $f(x^0 + h) - f(x^0)$ tendrá el mismo signo que h_p; es decir, cuando $h_p > 0$ la diferencia será positiva, y cuando $h_p < 0$ la diferencia será negativa; consecuentemente, x^0 no puede ser un punto extremo.

El argumento cuando $\left.\frac{\partial f(x)}{\partial x_p}\right|_{x = x^0} < 0$ es de manera similar.

Así, se puede concluir que cuando cualquiera de las derivadas parciales no es exactamente igual a cero en x^0, el punto x^0 no es un punto extremo.

<u>Condición suficiente:</u> Si se tiene una función $f(x)$ con primera y segundas derivadas parciales continuas; entonces, una condición suficiente para que $f(x)$ tenga un mínimo relativo en x^0, es decir, cuando

$$\left.\frac{\partial f(x)}{\partial x_i}\right|_{x = x^0} = 0 \qquad i = 1, 2, \ldots, n$$

es que la matriz de segundas derivadas parciales de $f(x)$, llamada matriz hessiana, sea una matriz definida positiva.

Prueba:

Por el teorema de Taylor se tiene que:

$$f(x^0+h)-f(x^0)= \sum_{i=1}^{n} h_i \left.\frac{\partial f(x)}{\partial x_i}\right|_{x=x^0} + \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left.\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|_{x=x^0+\theta h}$$

$$0 < \theta < 1$$

Ya que se cumple la condición necesaria, se tiene que:

$$f(x^o+h) - f(x^o) = \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{\partial^2 f(x)}{\partial x_j \partial x_j} \right|_{x=x^o+\theta h} \qquad 0<\theta<1$$

Luego, $f(x^o+h)-f(x^o)$ tendrá el mismo signo que:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right|_{x=x^o + \theta h}$$

Sin embargo, ya que $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ es contínua en el vecindario de x^o

$$\left. \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right|_{x=x^o} \quad \text{tendrá el mismo signo que} \quad \left. \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right|_{x=x^o+\theta h}$$

para un h pequeño, $0 < h_i < \delta$

Así, si $$\sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{{}^2 f(x)}{x_i \; x_j} \right|_{x=x^o}$$

es positiva, $f(x^o + h) - f(x^o)$ también será positiva y $f(x)$ tendrá un mínimo en x^o. Sin embargo, el término

$$\sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{{}^2 f(x)}{x_i \; x_j} \right|_{x=x^o}$$

en una forma cuadrática, y puede ser escrita como $h^T Hh$ donde

$h^T = (h_1, h_2 m. . . , h_n)$ y

$$H = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \bigg|_{x=x^0} \right]$$

Una matriz cuyos elementos son las segundas derivadas parciales de f(x) evaluadas en el punto x^0. Esta matriz H, es la matriz Hessiana de una función f(x). Se sabe, del estudio del Algebra Matricial, que una condición necesaria y suficiente para que esta forma cuadrática sea positiva es que la matriz Hessiana sea una matriz definida positiva.

Corolarios:

Si la matriz Hessiana de una función f(x) es indefinida cuando se evalúa en el punto x^*, donde la condición necesaria es satisfecha entonces, el punto x^* no es un punto extremo.

Ejemplos:

Encuentre el mínimo de:

$$f(x) = x_1^2 + x_2^2 + x_3^2 - 4x_1 - 8x_2 - 12x_3 + 56$$

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 - 4 = 0$$

$$\frac{\partial f(x)}{\partial x_2} = 2x_2 - 8 = 0$$

$$\frac{\partial f(x)}{\partial x_3} = 2x_3 - 12 = 0$$

Al resolver el sistema de ecuaciones se obtiene que

$x_1 = 2$, $x_2 = 4$ y $x_3 = 6$

La matriz Hessiana evaluada en (2, 4, 6) es:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

la cual es definida positiva. Por lo tanto, el punto (2,4,6) resulta ser un mínimo de $f(x)$.

Ahora, se procederá a determinar la condición suficiente para el caso cuando la matriz Hessiana de una función dada sea seminegativa o semipositiva.

Respecto a ésto, se han hecho varios intentos por desarrollar las condiciones suficientes para puntos extremos de una función de varias variables. Uno de estos intentos fue hecho por Lagrange, como una extensión del argumento en una dimensión. Lagrange argumentaba que si el término de segundo orden era semidefinido, el término de tercer orden debería ser cero y el término de cuarto orden proporcionaría la información requerida. Si el término de cuarto orden es semidefinido, se realizará una investigación más, en los términos de más alto orden. Sin embargo, Peano desarrolló un contraejemplo a este argumento, el contraejemplo de Peano es:

Ejemplo:

Considere la función:

$$f(x,y) = (y^2-x)^2 = y^4 - 2y^2x + x^2$$

nótese que en cualquier punto donde $y^2 = x$ es un mínimo de $f(x,y)$; la condición necesaria es:

$$\frac{\partial f(x,y)}{\partial x} = 2(x-y^2) = 0$$

$$\frac{\partial f(x,y)}{\partial y} = 4y^3 - 4yx = 0$$

La condición necesaria es satisfecha a lo largo de la curva $y^2 = x$.

Los términos de segundo orden son:

$$\frac{\partial^2 f(x,y)}{\partial x^2} = 2$$

$$\frac{\partial^2 f(x,y)}{\partial x \partial y} = -4y$$

$$\frac{\partial^2 f(x,y)}{\partial y^2} = 12y^2 - 4x = 8y^2$$

La matriz Hessiana está dada por:

$$\begin{bmatrix} 2 & -4y \\ -4y & 8y^2 \end{bmatrix}$$

la cual, es semidefinida positiva para cualquier y. Al aplicarle el argumento de Lagrange se tiene que las terceras derivadas parciales deberán ser todas iguales a cero, ya que la matriz Hessiana es semidefinida. Al investigar en uno de los términos de tercer -

orden, se tiene que:

$$\frac{\partial^3 f(x,y)}{\partial y \partial x^2} = -4$$

ya que este término no es cero, Lagrange podría argumentar que $x=y^2$ no es un mínimo. Sin embargo, se puede verificar que $f(x,y)$ no puede ser negativa, es cero solamente cuando $y^2 = x$, y es positiva para cualquier otro valor de x y y. Por lo tanto, podemos - concluir que la solución obtenida es un mínimo impropio.

Cuando se mostró que el argumento de Lagrange estaba errado, -- Serret modificó el argumento; Serret concluyó que se deberían investigar los términos de más alto orden solamente para aquellos - valores de h_i para los cuales la forma cuadrática es cero. Sin embargo, probó que esta conclusión es falsa.

Ejemplo:

$$\begin{aligned} \text{Sea } f(x,y) &= (y^2 - x)(y^2 - 2x) \\ &= y^4 - 3y^2x + 2x^2 \end{aligned}$$

Solución:

$$\frac{\partial f(x,y)}{\partial x} = -3y^2 + 4x = 0$$

$$\frac{\partial f(x,y)}{\partial y} = 4y^3 - 6yx = 0$$

El punto (0, 0) satisface las condiciones necesarias.

La matriz Hessiana, evaluada en (0, 0) es:

$$H = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

la cual es definida semipositiva. Expandiendo $f(x,y)$ en una serie de Taylor alrededor de (0, 0) se obtiene:

$$f(x+h,\ y+k) = \frac{1}{2}\,h^2 \left.\frac{\partial^2 f(x,y)}{\partial x^2}\right|_{(0,0)} + hk \left.\frac{\partial^2 f(x,y)}{\partial x \partial y}\right|_{(0,0)}$$

$$+ \frac{1}{2!}\,k^2 \left.\frac{\partial^2 f(x,y)}{\partial y^2}\right|_{(0,\ 0)} + R_2(x+h, y+k;\ x,y)$$

$$= 2h^2 + R_2(x+h,\ y+k;\ x,y)$$

El término de segundo orden $2h^2$, es positivo para toda h y k, excepto para h = 0. Por lo tanto, siguiendo el argumento de Gerrat, se observarán los términos de tercer orden, pero sólo para el valor de h donde el término de segundo orden se anula; el término de tercer orden está dado por:

$$\frac{1}{3!} \quad (-18\, h\, k^2)$$

el cual, se anula cuando h = 0. Investigando el término de cuarto orden, se obtiene:

$$f(x+h\ ,\ y+k)-f(x,y)\Big|_{(0,0)} = h^2-3hk^2+k^4+R_4(x+h,y+k;x,y)$$

Aplicando el argumento de Serret se observa que cuando $h = 0$ (es decir, cuando los términos de segundo orden se anulan), la - forma es definida positiva y consecuentemente el punto (0, 0) es un mínimo. Sin embargo, se nota que para $x < y^2 < 2x$ la función es negativa y para $y^2 < x$ ó $y^2 > 2x$ la función es positiva y consecuentemente el punto (0, 0) es un punto de silla, ya que decrece - para toda y en el rango $x < y^2 < 2x$ y crece en el rango $y_2 < x$ y $y^2 > 2x$.

La cuestión de suficiencia, finalmente fue resuelta por Scheffer. Su argumento, el cual es presentado a continuación, incluye transformar el problema en muchos problemas de optimización unidimensional.

Recordemos que la serie de Taylor se puede escribir como:

$$f(x^o+h) = f(x^o) + G_n(x^o) + R_n(x^o + \theta h,\ x^o)$$

Supongamos que x^o es un punto estacionario. Seleccionemos un elemento del vector h, digamos h_p, que sea constante, y dejemos que - los otros elementos de h varíen de tal forma que $|h_i| < h_p$

El mínimo de $G_n(x^o)$ con respecto a estos h_i, es determinado en el espacio de n-1 dimensión, este mínimo es llamado $G_n(x^o)^p$, este -

procedimiento es repetido para cada h_i, $i = 1,2,\ldots, n$. Si $\min_p G_n(x^o)^p$ es positivo, entonces x^o es un mínimo porque $f(x^o+h) - f(x^o) > 0$. Si algunos $G_n(x^o)^p$ son positivos y otros son negativos, entonces x^o no es un punto extremo; si $\min_p G_n(x^o)^p$ es cero, entonces n es aumentado en uno y el proceso completo se repite.

Problemas con Restricciones

Considerese el problema de maximizar una función diferenciable en dos variables $f(x,y)$, sujeta a la restricción de igualdad diferenciable $g(x,y) = 0$. Existen dos formas de resolver el problema, la primera implica resolver la restricción para una de las variables, digamos "y"; y sustituir el resultado en $f(x,y)$; así, dado $g(x,y) = 0$, se obtiene $y = \mu(x)$ (donde $\mu(x)$) es la resolución para "y" que resulta de $g(x,y) = 0$); y sustituimos esto por "y", transformando así al problema a uno de la forma:

$$\max_{x} \quad f(x, \mu(x))$$

Existe una dificultad asociada a este método, y es que el teorema de la función implícita garantiza, bajo ciertas condiciones, la existencia de la función $\mu(x)$, pero no ayuda a obtenerla. Por ejemplo, si la restricción fuera:

$$g(x,y) = e^{xy^2} + x \cos y = 0$$

Aún cuando la existencia de la inversa es garantizada para la mayoría de los puntos en el espacio (x,y), se tiene dificultad para obtener la forma analítica de la función $\mu(x)$; esta dificultad, puede ser obviada al utilizar la técnica del multiplicador de Lagrange, la cual involucra agregar una variable al problema por cada restricción.

El desarrollo de este método se hará inicialmente para una función de dos variables, para hacer uso intuitivo de algunos conceptos de geometría. La extensión a un número arbitrario de variables, se deduce de las pruebas en dos dimensiones.

Considere el problema de maximizar $f(x,y)$ sujeta a la restricción $g(x,y) = 0$, supongamos que podemos obtener una expresión analítica para $y = \mu(x)$; si tal función existe, es diferenciable y posee una primera derivada contínua; además, se puede transformar el problema a uno de una variable de la forma:

$$\max_{x} f(x, \mu(x))$$

Una condición necesaria para que f tenga un máximo en algún punto (x^o, y^o) es que la derivada de f con respecto a x se anule en (x^o, y^o). La diferencial total de $f(x,y)$ puede ser escrita de la forma:

$$d\, f(x,y) = \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy$$

y la derivada total por:

$$\frac{d\ f(x,y)}{dx} = \frac{\partial f(x,y)}{\partial x} + \frac{\partial f(x,y)}{\partial y}\ \frac{dy}{dx}$$

Si se sustituye la definición de "y"; y además, se evalúan las derivadas en x^o; se obtiene:

$$\left.\frac{df(x,y)}{dx}\right|_{x=x^o} = \left.\frac{\partial f(x,y)}{\partial x}\right|_{x=x^o} + \left.\frac{\partial f(x,y)}{\partial y}\ \frac{d\mu(x)}{dx}\right|_{x=x^o}$$

la cual debe ser cero si x^o es un punto estacionario.

Para obtener una expresión para $\frac{d\ \mu(x)}{dx}$ se escribe la derivada total de $g(x,y) = 0$, así:

$$\frac{dg(x,y)}{dx} = \frac{\partial g(x,y)}{\partial x} + \frac{\partial g(x,y)}{\partial y} \cdot \frac{d\ \mu(x)}{dx} = 0$$

Esta expresión puede ser escrita de la siguiente forma:

$$\frac{d\mu(x)}{dx} = - \frac{\partial g(x,y)}{\partial x} \Big/ \frac{\partial g(x,y)}{\partial y}$$

Al sustituirse esta expresión en la condición necesaria, se obtiene:

$$\left.\frac{\partial f}{\partial x}\right|_{x=x^o} - \left.\frac{\partial f}{\partial y}\left(\frac{\partial g}{\partial x} \Big/ \frac{\partial g}{\partial y}\right)\right|_{x=x^o} = 0$$

Definiendo a λ como:

$$\lambda = -\left.\frac{\partial f(x,y)}{\partial y} \Big/ \frac{\partial g(x,y)}{\partial y}\right|_{x=x^{0}}$$

La condición necesaria se puede escribir así:

$$\left.\frac{\partial f(x,y)}{\partial x}\right|_{x=x^{0}} + \lambda \left.\frac{\partial g(x,y)}{\partial x}\right|_{x=x^{0}} = 0$$

de la definición de λ se obtiene, además:

$$\frac{\partial f(x,y)}{\partial y} + \lambda \frac{\partial g(x,y)}{\partial y} = 0$$

y por supuesto la restricción $g(x,y) = 0$

El símbolo λ es llamado multiplicador de Lagrange. Nótese que las condiciones necesarias, las cuales son representadas por las últimas tres ecuaciones, no requieren que $g(x,y) = 0$ sea resuelta para "y"; el requerimiento es que $\mu(x)$ exista, y el teorema de la función implícita da las condiciones bajo las cuales la función $\mu(x)$ puede ser teóricamente obtenida.

En el proceso anterior, se asumió que $\partial g(x,y)/\partial y$ no se anula en el punto estacionario, este supuesto fue necesario para definir el multiplicador de Lagrange λ; sin embargo, resultados equivalentes pueden ser obtenidos si se define a λ en términos de las derivadas parciales con respecto a x. La derivación requiere únicamente que al menos una de las derivadas parciales de $g(x,y)$ no se anule en el punto estacionario.

Existe un método compacto para generar las condiciones necesarias mostradas anteriormente, e implica construir la función de Lagrange la cual, es formada multiplicando la restricción por un multiplicador de Lagrange (λ) y agregando este producto a la función objetivo, así:

$$F(x,y,\lambda) = f(x,y) + \lambda g(x,y)$$

Cuando la función de Lagrange, $F(x,y,\lambda)$, es derivada parcialmente con respecto a sus argumentos y los resultados son igualados a cero, las condiciones necesarias son obtenidas así:

$$\frac{\partial F(x,y,\lambda)}{\partial x} = \frac{\partial f(x,y)}{\partial x} + \lambda \frac{\partial g(x,y)}{\partial x} = 0$$

$$\frac{\partial F(x,y,\lambda)}{\partial y} = \frac{\partial f(x,y)}{\partial y} + \lambda \frac{\partial g(x,y)}{\partial y} = 0$$

$$\frac{\partial F(x,y,\lambda)}{\partial \lambda} = g(x,y) = 0$$

Ejemplos:

Si se tiene el siguiente problema:

minimizar $f(x,y) = 3x^2 + 5xy + 3y^2$

sujeta a $g(x,y) = 5x + 3y - 2 = 0$

Construir la función de Lagrange, y además, obtener las condiciones necesarias a partir de la función de Lagrange.

Solución:

$$F(x,y,\lambda) = 3x^2 + 5xy + 3y^2 + \lambda(5x + 3y - 2)$$

$$\frac{\partial F(x,y,\lambda)}{\partial x} = 6x + 5y + 5\lambda = 0$$

$$\frac{\partial F(x,y,\lambda)}{\partial y} = 5x + 6y + 3\lambda = 0$$

$$\frac{\partial F(x,y,\lambda)}{\partial \lambda} = 5x + 3y - 2$$

El resultado anterior puede ser extendido para incluir funciones de varias variables, así como también restricciones adicionales. Considérese una función de n variables, $f(x)$, sujeta a m restricciones de igualdad de la forma:

$$g_i(x) = 0 \qquad\qquad i = 1,2,\ldots, m < n$$

Supóngase, además, que se desea localizar todos los máximos relativos de $f(x)$ sujeto a estas m restricciones.

En el máximo relativo x^o, se requiere que el diferencial de $f(x)$ se anule, así:

$$df = \sum_{i=1}^{n} \left.\frac{\partial f(x)}{\partial x_i}\right|_{x=x^o} dx_i = 0$$

El razonamiento posterior a esta proposición es similar a los argumentos dados anteriormente.

Si los términos de primer orden de la serie de Taylor no se anulan, se puede mejorar el valor de f mediante una selección adecuada del dx_i en el vecindario de x^o. Sin embargo, no se requiere que cada derivada parcial se anule; en muchos casos, un máximo relativo pue de ocurrir en la frontera de una o más de las restricciones, donde la derivada parcial de la función objetivo no necesariamente se anu le.

El diferencial total de cada una de las restricciones es:

$$dg_i = \sum_{j=1}^{n} \frac{\partial gi(x)}{\partial x_j} \quad dx_j = 0 \qquad i = 1,2,\ldots, m$$

multiplicando cada una de estas ecuaciones por su multiplicador de Lagrange asociado λ_i, y agregándolo a df resulta:

$$df + \sum_{i=1}^{m} \lambda_i dg_i = \sum_{j=1}^{m} \left[\frac{\partial f(x)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial gi(x)}{\partial x_j} \right] dx_j = 0$$

Supóngase que se deciden eliminar las primeras m variables, x_1, x_2, ..., x_m; estas m variables, serán variables dependientes y las restantes n-m variables serán las variables independientes. Ya que cada uno de los diferenciales dx_{m+1}, dx_{m+2},..., dx_n corresponden a una variable independiente, y pueden ser seleccionados para ser diferentes de cero, el coeficiente de cada diferencial deberá ser nulo, es decir:

$$\frac{\partial f(x)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial gi(x)}{\partial x_j} = 0 \qquad j = m+1,\ m+2, \ldots, n$$

Así, se tiene a la izquierda que:

$$\sum_{j=1}^{m}\left[\frac{\partial f(x)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial gi(x)}{\partial x_j}\right] dx_j = 0$$

Los diferenciales dx_j $j = 1,2,\ldots, m$ son dependientes y únicamente determinados por los valores de los diferenciales independientes por lo tanto, para garantizar que

$$df + \sum_{i=1}^{m} \lambda_i dg_i$$

se anule, se deberá requerir que los coeficientes de dx_j $j = 1,2,\ldots, m$ sean cero.

$$\left.\frac{\partial f(x)}{\partial x_j}\right|_{x=x^o} + \sum_{i=1}^{n} \lambda_i \left.\frac{\partial gi(x)}{\mu x_j}\right|_{x=x^o} = 0 \qquad j = 1,2,\ldots,m$$

ó

$$\sum_{i=1}^{m} \lambda_i \frac{\partial gi(x)}{\partial x_j} = -\frac{\partial f(x)}{\partial x_j} \text{ en } x = x^o \qquad j = 1,2,\ldots, m$$

Escribiendo lo anterior en forma matricial, se tiene:

$$\hat{J}_{x_1, x_2,\ldots, x_m}\left[g_1(x),\ldots, g_m(x)\right]\lambda = -\Delta f(x)$$

donde: $\hat{J}$ es la matriz Jacobiana de $g_i(x)$, λ es el vector columna de los multiplicadores de Lagrange y Δf es el vector de las derivadas parciales de $f(x)$.

Si se asume que $\Delta f(x)$ no es nulo, se tienen m ecuaciones no homogéneas, las cuales son lineales en λ. En general, una solución a estas ecuaciones existe si, y solo si, la matriz de coeficientes de λ es no singular; la cual es la misma condición requerida por el teorema de la función implícita para permitir la eliminación de $x_1, x_2, \dots x_m$. Nótese que se pueden tener n-m variables independientes si el Jacobiano formado por las restantes variables es no singular.

Como se mencionó antes, las condiciones necesarias pueden ser generadas a partir de la función de Lagrange, definiendo la función de Lagrange, como:

$$F(x, \lambda) = F(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

Teorema: Una condición necesaria para que f(x) tenga un máximo relativo en x^o es que las derivadas parciales de la función de Lagrange, con respecto a cada uno de sus argumentos sea cero.

Prueba:

Si se toman las derivadas parciales de $F(x, \lambda)$ con respecto a cada uno de sus argumentos e igualando a cero el resultado, se obtiene:

$$\frac{\partial F(x, \lambda)}{\partial x_j} = \frac{\partial f(x)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(x)}{\partial x_j} = 0 \quad j=1,2,\ldots,n$$

$$\frac{\partial F(x, \lambda)}{\partial \lambda_i} = g_i(x) = 0 \qquad i = 1,2,\ldots, m$$

Las cuales son idénticas al conjunto de ecuaciones obtenido antes. Estas ecuaciones son también las condiciones necesarias para un mínimo relativo de $f(x)$.

A continuación se hará el desarrollo para obtener las condiciones suficientes para un máximo relativo de $f(x)$ sujeta a las restricciones $g_i(x) = 0$. Obsérvese que una solución x^o, debe ser factible para ser candidata para un máximo relativo; es decir, debe satisfacer $g_i(x^o) = 0$ $i = 1,\ldots, m$.

Si una solución es factible, la función de Lagrange es dada por:

$$F(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) = f(x)$$

Un punto x^o es un máximo relativo si:

$$F(x^o + h) - f(x^o) < 0$$

Sin embargo, los elementos de h no son todos independientes ya que las restricciones evaluadas en $x^o + h$ deben ser satisfechas, es decir: $g_i(x^o + h) = 0$ $i = 1,2,\ldots, m$, Por lo tanto:

$F(x^{o}+h,\lambda)-F(x^{o},\lambda) = f(x^{o}+h) - f(x^{o})$.-

Expandiendo a F en la serie de Taylor en x alrededor de x^{o} se obtiene:

$$F(x^{o}+h,\lambda)-F(x^{o},\lambda) = \sum_{j=1}^{n} \left.\frac{\partial F(x,\lambda)}{\partial x_j}\right|_{x=x^{o}} h_j + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j \left.\frac{\partial^2 F(x,\lambda)}{\partial x_i \partial x_j}\right|_{x=x^{o}} + R_2(x^{o}+h,\lambda;x^{o},\lambda)$$

Como la condición necesaria requiere que cada primera derivada de F(x) sea nula, entonces, la ecuación se reduce a:

$$F(x^{o}+h,\lambda)-F(x^{o},\lambda) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \left.\frac{\partial^2 F(x,\lambda)}{\partial x_i \partial x_j}\right|_{x=x^{o}} + R_2(x^{o}+h,\lambda;x^{o},\lambda)$$

TEOREMA: Una condición suficiente para que f(x) tenga un máximo relativo propio en x^{o} es que la forma cuadrática

$$Q = \sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j \left.\frac{\partial^2 F(x,)}{\partial x_i \partial x_j}\right|_{x=x^{o}}$$

Sea definida negativa para todos los valores de h para los cuales las restricciones son satisfechas.

PRUEBA: Supongase que la forma cuadrática del Teorema no es definida negativa; entonces, $F(x^o + h) - f(x^o)$ no será negativo para todos los valores admisibles de h y x^o no será un máximo relativo; sin embargo. Obsérvese que la forma cuadrática es requerida que sea definida negativa sólo para los h que satisfacen las restricciones, porque cualquier punto donde las condiciones no sean satisfechas no cumplirá las condiciones necesarias, es decir;

$$g_i(x) = 0$$

Hancock mostró que la condición necesaria para que la forma cuadrática Q sea definida negativa para todas las h admisibles, es que cada raíz U_i del polinomio definido por el siguiente determinante sean negativas:

$$\begin{vmatrix} (F_{11}-U) & \cdots\cdots & F_{1n} & g_{11} & \cdots & g_{m1} \\ \vdots & & & \vdots & & \vdots \\ F_{n1} & \cdots\cdots & (F_{nn}-U) & g_{1n} & \cdots & g_{mn} \\ g_{11} & \cdots\cdots & g_{m1} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ g_{1n} & \cdots\cdots & g_{mn} & 0 & \cdots & 0 \end{vmatrix}$$

donde:

$$F_{ij} = \left.\frac{\partial^2 F(x, \lambda)}{\partial x_i \, \partial x_j}\right|_{x=x^o}$$

$$g_{ij} = \left.\frac{\partial g_i(x)}{\partial x_j}\right|_{x=x^o}$$

El determinante anterior, proporciona un polinomio de orden $(n-m)$ en U. Si cada raíz de esta ecuación es negativa, el punto x^o es un máximo relativo.

Si se está investigando $f(x)$ para un mínimo, entonces las raíces U_i deberán ser todas positivas. Cuando algunas raíces son positivas y otras negativas, el punto x^o no es un punto extremo.

Ejemplo: Considere el siguiente problema:

$$\min x^2 + y^2 \text{ sujeta a } g(x,y) = x-y-5 = 0$$

Solución:

Formando la función de Lagrange, se tiene:

$$F(x,y,\lambda) = x^2 + y^2 + \lambda(x-y-5)$$

de donde:

$$\frac{\partial F(x,y,\lambda)}{\partial x} = 2x + \lambda = 0$$

$$\frac{\partial F(x,y,\lambda)}{\partial y} = 2y - \lambda = 0$$

$$\frac{\partial F(x,y,\lambda)}{\partial \lambda} = x - y - 5 = 0$$

De donde resulta que $x = 2.5$, $y = -2.5$ y $\lambda = -5$ es un punto estacionario. Además:

$F_{11} = 2$ $\quad F_{22} = 2$ $\quad F_{12} = F_{21} = 0$ $\quad g_{12} = -1$ $\quad g_{11} = 1$

Al construir el determinante sugerido por Hancock resulta:

$$\begin{vmatrix} 2-U & 0 & 1 \\ 0 & 2-U & -1 \\ 1 & -1 & 0 \end{vmatrix} = 0$$

de donde:

$$(2-U)(0-1)-0(0+1)+1(0-(2-U)) = 0$$

$$U-2 + U-2 = 0$$

$$2U - 4 = 0$$

$$\underline{\underline{U = 2}}$$

ya que $U = 2 > 0$. Podemos concluir que el punto $x = 2.5$, $y = -2.5$ resulta ser un mínimo

La restante posibilidad, la cual es similar al caso semidefinido discutido antes para problemas sin restricciones será discutido a continuación:

El último resultado obtenido, ha sido restringido al caso donde al menos uno de los Jacobianos de orden m, construido a partir de las m variables y las m restricciones, no se anule.

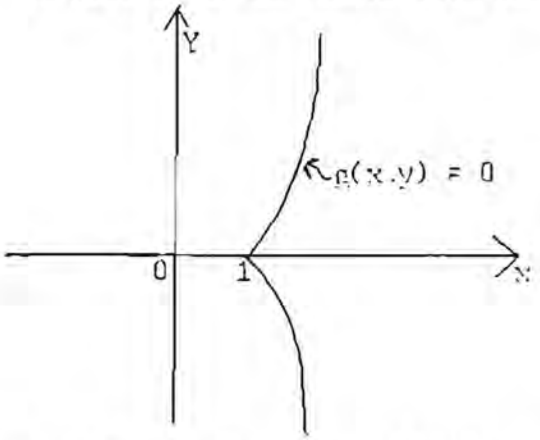
Cuando esta condición no es satisfecha, el resultado anterior no representa la condición necesaria para un punto extremo; ya que, si se considera el siguiente ejemplo mostrado por Courant se observará su invalidez.

Ejemplo:

$$\min f(x,y) = x^2 + y^2 \quad \text{sujeto a } g(x,y) = (x-1)^3 - y^2 = 0$$

Solución:

El gráfico de la restricción resulta ser:



formando la función de Lagrange y diferenciándola con respecto a cada uno de sus argumentos, resulta:

$$F(x,y,\lambda) = x^2 + y^2 + \lambda((x-1)^3 - y^2)$$

$$\frac{\partial F(x,y,\lambda)}{\partial x} = 2x + 3\lambda(x-1)^2 = 0$$

$$\frac{\partial F(x,y,\lambda)}{\partial y} = 2y - 2\lambda y = 0$$

$$\frac{\partial F(x,y,\lambda)}{\partial \lambda} = (x-1)^3 - y^2 = 0$$

A partir del gráfico de la restricción se observa que el mínimo ocurre en $x=1$, $y=0$. Además, este punto satisface las dos últimas ecuaciones, pero no, la primera, ya que:

$$\left.\frac{\partial F(x,y,\lambda)}{\partial x}\right|_{(1,0)} = 2$$

El punto es un mínimo, pero no satisface las condiciones necesarias, ya que, las condiciones bajo las cuales se obtuvo la condi ción necesaria no han sido satisfechas. Por lo tanto, las condiciones

$$\frac{\partial F(x, \lambda)}{\partial x_j} = 0 \qquad j = 1, 2, \ldots, n$$

$$\frac{\partial F(x, \lambda)}{\partial \lambda_i} = 0 \qquad i = 1, 2, \ldots, m$$

son necesarias únicamente cuando al menos uno de los Jacobianos formados por las m restricciones y n variables no se anula. Existe una manera simple de obviar esta dificultad y ésta consiste en lo siguiente:

Se está interesado en generar todos los puntos que pueden ser extremos; la técnica del multiplicador de Lagrange genera un subconjunto de puntos para los cuales uno de los Jacobianos no se anula; se puede demostrar que estos puntos pueden ser obtenidos al resolver el conjunto de ecuaciones siguiente:

$$\hat{J}_{x_1, x_2, \ldots x_m}(g_1, g_2, \ldots, g_m)\, \lambda = \emptyset$$

donde:

$\hat{J}$ = es la matriz Jacobiana asociada a g_i

λ = es el vector columna de multiplicadores de Lagrange

$\emptyset$ = es el vector nulo

Además, se requiere que el vector λ sea diferente del vector nulo.

Una condición necesaria para que exista una solución de estas ecuaciones es que el Jacobiano sea nulo; así, todos los puntos para los cuales la técnica del multiplicador de Lagrange falla son generados.

Estas ecuaciones forman un conjunto de condiciones necesarias para un punto extremo; cuando la técnica del multiplicador de Lagrange falla.

Los dos conjuntos mutuamente excluyentes de condiciones pueden ser generados mediante una simple modificación del método de Lagrange; es decir, construyendo la función de Lagrange como:

$$F(x, \lambda) = \lambda_o f(x) \pm \sum_{i=1}^{m} \lambda_i g_i(x)$$

donde:

λ_o es una constante que puede valer 0 ó 1.

Cuando λ_o vale uno, las condiciones necesarias para el caso cuando el Jacobiano no sea nulo son generadas; y, cuando $\lambda_o = 0$, las condiciones necesarias para cuando el Jacobiano se anule son generadas.

Por lo tanto, las ecuaciones:

$$\frac{\partial F(x, \lambda)}{\partial x_j} = 0 \qquad j = 1, 2, \ldots, n$$

$$\frac{\partial F(x, \lambda)}{\partial \lambda_i} = 0 \qquad i = 1, 2, \ldots, m$$

primero, son evaluadas cuando $\lambda_o = 0$ y luego cuando $\lambda_o = 1$.

Todos los puntos generados por este nuevo método son factibles de ser punto extremo.

Ejemplo: Consideremos nuevamente el problema dado por Courant

$$\min f(x,y) = x^2 + y^2$$

sujeto a

$$g(x,y) = (x-1)^3 - y^2 = 0$$

Formando la función de Lagrange y diferenciándola, se obtiene:

$$F(x, y, \lambda_0, \lambda_1) = \lambda_0(x^2 + y^2) + \lambda_1((x-1)^3 - y^2)$$

$$\frac{\partial F(x,y, \lambda_0, \lambda_1)}{\partial x} = 2\lambda_0 x + 3\lambda_1(x-1)^2 = 0$$

$$\frac{\partial F(x,y, \lambda_0, \lambda_1)}{\partial y} = 2\lambda_0 y - 2\lambda_1 y = 0$$

$$\frac{\partial F(x,y, \lambda_0, \lambda_1)}{\partial \lambda_1} = (x-1)^3 - y^2 = 0$$

Si $\lambda_0 = 1$, se tiene el mismo sistema de ecuaciones que no pudieron generar el punto factible (1, 0) en el ejercicio anterior. Ahora, si $\lambda_0 = 0$ se obtiene:

$$3\lambda_1(x-1)^2 = 0$$

$$2\lambda_1 y = 0$$

$$(x-1)^3 - y^2 = 0$$

el cual tiene por solución trivial x=1, y=0 y este punto es el mínimo deseado.

Restricciones de Desigualdad.

A continuación se desarrollarán las condiciones necesarias y suficientes para problemas de optimización con restricciones de desigualdad. Estas condiciones son llamadas condiciones de Kuhn-Tucker en honor a los hombres que los desarrollaron.

Supóngase que se desea maximizar la función $f(x)$, la cual está sujeta al conjunto de restricciones de desigualdad.

$$g_i(x) \leq b_i \qquad i = 1,2,\ldots, m$$

Las restricciones de desigualdad pueden ser transformadas a igualdades aumentando el número de variables en el problema. Si se agrega una variable de holgura no negativa U_i^2 a cada desigualdad, se obtiene:

$$g_i(x) + U_i^2 = b_i \qquad i = 1,2,\ldots, m$$

restando b_i se obtiene:

$$\bar{g}_i(x,U_i) = g_i(x) + U_i^2 - b_i = 0 \qquad i = 1,2,\ldots, m$$

Ahora, el problema es de la forma en que se puede aplicar la teoría dada anteriormente, es decir, de la forma:

$$\max f(x)$$

$$\text{sujeta a } \bar{g}_i(x, U_i) = 0 \qquad i = 1,2,\ldots, m$$

La solución es obtenida generando todos los puntos estacionarios de la función de Lagrange, la cual es dada por:

$$F(x, \lambda, U) = f(x) + \sum_{i=1}^{m} \lambda_i \bar{g}_i(x, U_i)$$

los puntos estacionarios, son obtenidos resolviendo las ecuaciones

$$\frac{\partial F(x, \lambda, U)}{\partial x_j} = \frac{\partial f(x)}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(x)}{\partial x_j} = 0 \qquad j = 1, 2, \ldots, n$$

$$\frac{\partial F(x, \lambda, U)}{\partial \lambda_i} = g_i(x, U_i) = g_i + U_i^2 - b_i = 0 \qquad i = 1, 2, \ldots, m$$

$$\frac{\partial F(x, \lambda, U)}{\partial U_i} = -2 \lambda_i U_i = 0 \qquad i = 1, 2, \ldots, m$$

Multiplicando la última ecuación por $\frac{U_i}{2}$ resulta:

$$\lambda_i U_i^2 = 0$$

y despejando U_i^2 de $g_i(x, U_i)$ se obtiene:

$$U_i^2 = b_i - g_i(x)$$

Combinando estas dos últimas ecuaciones resulta:

$$\lambda_i \left[b_i - g_i(x) \right] = 0 \qquad i = 1, 2, \ldots, m$$

esta ecuación junto con

$$\frac{\partial F(x, \lambda, b)}{\partial x_j} = 0 \qquad j = 1, 2, \ldots, n$$

y las restricciones $g_i(x) \leq b_i$ representan un subconjunto de las condiciones necesarias para un máximo de $f(x)$.

El requerimiento final es:

$$\lambda_i \geq 0 \qquad i = 1, 2, \ldots, m$$

A continuación se da la prueba de esta última condición.

Supóngase que las condiciones derivadas anteriormente son satisfechas en el punto (x^o, λ^o, U^o); y además, supóngase que este punto es un máximo relativo.

Sea $z^o = f(x^o)$ y se define

$$gi\,(x^o, U_i^o) = b_i \qquad i = 1, 2, \ldots, m$$

TEOREMA:

$$\frac{\partial z^o}{\partial b_i} = -\lambda_i^o$$

Este teorema proporciona una cantidad medible del cambio en el valor óptimo de la función objetivo con respecto a cambios en la restricción. Consecuentemente, los λ_i^o han sido llamados coeficientes de sensibilidad, así también variables duales, funciones adjuntas y espectro precios.

Prueba:

De la regla de la cadena del cálculo diferencial se tiene:

$$\frac{\partial z^o}{\partial b_i} = \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_j} \cdot \frac{\partial x_j}{\partial b_i}\Bigg|_{x=x^o}$$

$$\frac{\partial g_i(x, U_k)}{\partial b_i}\Bigg|_{\substack{x=x^o \\ U_k=U_k^o}} = \sum_{j=1}^{n} \frac{\partial g_i(x, U_i)}{\partial x_j}\Bigg|_{\substack{x=x^o \\ U_j=U_j^o}} \qquad \frac{\partial x_j}{\partial b_j} = \delta_{ik}$$

donde:

$$\delta_{ik} = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}$$

Multiplicando a_{ik} por λ_k^0 y sumando todos los valores de k resulta:

$$\sum_{k=1}^{m} \lambda_k^0 \, a_{ik} = \sum_{k=1}^{m} \lambda_k^0 \sum_{j=1}^{n} \left. \frac{\partial g_k(x, b_k)}{\partial x_j} \right|_{\substack{x=x^0 \\ b_k=b_k^0}} \cdot \frac{\partial x_j}{\partial b_i}$$

La expresión anterior es igual a λ_i^0, ya que $\lambda_{ik} = 0$, excepto para k = i por lo tanto:

$$\lambda_i^0 = \sum_{k=1}^{m} \lambda_k^0 \sum_{j=1}^{n} \left. \frac{\partial g_k(x, b_k)}{\partial x_j} \right|_{\substack{x=x^0 \\ b_k=b_k^0}} \cdot \frac{\partial x_j}{\partial b_i}$$

Sumando este resultado a $\dfrac{\partial z^0}{\partial b_i}$ se obtiene:

$$\frac{\partial z^0}{\partial b_i} + \lambda_i^0 = \sum_{j=1}^{n} \left[\frac{\partial f(x)}{\partial x_j} + \sum_{k=1}^{m} \lambda_k^0 \left. \frac{\partial g_k(x, b_k)}{\partial x_j} \right|_{\substack{x=x^0 \\ b_k=b_k^0}} \right] \frac{\partial x_j}{\partial b_i}$$

Sin embargo, el lado derecho si es cero, ya que el término entre corchetes debería ser igual a cero para satisfacer las condiciones.

necesarias, así:

$$\frac{\partial z^o}{\partial b_i} = -\lambda_i^o$$

De la derivación de las condiciones necesarias dadas anteriormente, se sabe que cualquier $\lambda_j^o = 0$, o, $U_i^o = 0$ o ambos se anulan en la condición óptima. Se investigará ahora, el caso cuando $U_i^o \neq 0$, esto implica que la restricción es satisfecha como una desigualdad estricta en x^o y consecuentemente, si se relaja la restricción (haciendo b_i más grande) el punto extremo no será afectado. Por lo tanto, el cambio en el valor óptimo de la función objetivo con cambios en b_j será cero, es decir:

$$\frac{\partial z^o}{\partial b_i} = -\lambda_i^o = 0$$

Ahora, supóngase que $\lambda_i^o \neq 0$. Esto implica que la variable de holgura U_i^o es cero. Así:

$$g_i(x^o) = b_i$$

Supóngase que $\lambda_i^o > 0$. Entonces:

$$\frac{\partial z^o}{\partial b_i} < 0$$

Esto indica que cuando b_i es aumentado, la función objetivo decrece. Sin embargo, como b_i aumenta, más del espacio será factible, y el valor óptimo de la función objetivo claramente no puede decrecer.

Así, en una solución óptima $\lambda_i \leq 0$

Realizando un resumen de las condiciones desarrolladas se tiene que:

Dado el problema de maximizar $f(x)$ sujeta a las m restricciones de desigualdad

$$g_i(x) \leq b_i \qquad i = 1,2,\ldots,m$$

Las condiciones necesarias para un máximo relativo en x^o son:

1) $$\frac{\partial F(x^o, \lambda^o, U^o)}{\partial x_i} = 0 \qquad i = 1,2,\ldots,n$$

2) $$\lambda_i(g_i(x^o)-b_i) = 0 \qquad i = 1,2,\ldots,m$$

3) $$\lambda_i^o \leq 0 \qquad i = 1,2,\ldots,m$$

4) $$g_i(x^o) \leq b_i \qquad i = 1,2,\ldots,m$$

Se puede demostrar, con un argumento similar, que si se desea un mínimo de $f(x)$, se debe cambiar el sentido de la desigualdad en la condición 3, así:

$$\lambda_i^o \geq 0 \qquad i = 1,2,\ldots,m$$

Ejemplo:

Considere el siguiente problema:

$$\min f(x) = cx + 1/2\ x^T Gx$$

Sujeto a: $Ax \leq b \qquad x \geq 0$

donde:

$$c = (2,4) \qquad G = \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} \qquad A = \begin{bmatrix} -1 & 4 \\ 6 & 2 \\ 1 & -1 \end{bmatrix}$$

$$b = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Formando la función de Lagrange, se tiene:

$$F(x, \lambda, U) = 2x_1 + 4x_2 + x_1^2 + 3x_2^2 + 3x_1x_2 + \lambda_1(-x_1+4x_2+U_1^2 - 4)$$

$$+ \lambda_2(6x_1 + 2x_2 - U_2^2 - 1) + \lambda_3(x_1-x_2+U_3^2 - 3)$$

$$+ \lambda_4(-x_1 + U_4^2) + \lambda_5(-x_2 + U_5^2)$$

Las condiciones necesarias son:

1) $$\frac{\partial F}{\partial x_1} = 2 + 2x_1 + 3x_2 - \lambda_1 + 6\lambda_2 + \lambda_3 - \lambda_4 = 0$$

$$\frac{\partial F}{\partial x_2} = 4 + 3x_1 + 6x_2 + 4\lambda_1 + 2\lambda_2 - \lambda_3 - \lambda_5 = 0$$

2) $$\lambda_1(-x_1 + 4x_2 - 4) = 0 \qquad \lambda_4 x_1 = 0$$

$$\lambda_2(6x_1 + 2x_2 - 1) = 0 \qquad \lambda_5 x_2 = 0$$

$$\lambda_3(x_1 - x_2 - 3) = 0$$

3) $x \geqslant 0$

$Ax \leqslant b$

4) $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 > 0$

Se puede verificar que una solución es:

$x_1 = x_2 = \lambda_1 = \lambda_2 = \lambda_3 = 0 \qquad \lambda_4 = 2 \qquad \lambda_5 = 4$

se demostrará que esta es la única solución cuando $F(x, \lambda, U)$ es una función convexa.

A continuación se mostrará que si $f(x)$ es estrictamente cóncava y $g_i(x)$ $i = 1, 2, \ldots, m$ son convexas; entonces, las condiciones dadas anteriormente son suficientes como también necesarias para un máximo absoluto.

Supóngase que $f(x)$ y $g_i(x)$ satisfacen las condiciones especificadas anteriormente, entonces:

$$F(x, \lambda, U) = f(x) + \sum_{i=1}^{m} \lambda_i (g_i(x) + U_i^2 - b_i)$$

Si $\lambda_i < 0$; entonces, $\lambda_i g_i(x)$ es cóncava si $g_i(x)$ es convexa.

Por lo tanto

$$f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

es estrictamente cóncava. Ya que $\lambda_i U_i = 0$ y $\lambda_i b_i$ es una constante, si:

$$f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

es cóncava, $F(x, \lambda, U)$ es cóncava. Se había demostrado antes que una condición necesaria para que $F(x)$ sea un máximo en x^o es que $F(x, \lambda, U)$ tenga un punto estacionario en x^o. Sin embargo, si - $F(x, \lambda, U)$ es estrictamente cóncava sus derivadas se anularían en un solo punto. Consecuentemente, este punto deberá ser el máximo relativo. Por lo tanto, las condiciones 1,2,3 y 4 son suficientes como también necesarias para un máximo absoluto de $f(x)$ en x^o.

Recuérdese que existe una restricción en el desarrollo dado. las - condiciones necesarias dadas antes, se derivaron basándose en la derivación de las restricciones de desigualdad. Uno de los requerimientos para esta derivación fue que al menos uno de los Jacobianos compuesto de las m restricciones y m de las n+m variables x_i - $i = 1,2,\ldots, n$, U_j, $j = 1,2,\ldots, m$, sea diferente de cero, este requerimiento está incluido en el desarrollo dado.

Hay varios casos especiales de las condiciones de Kuhn-Tucker. Supóngase que p de las restricciones son de la forma $x_i \geqslant 0$ $i = 1,2,\ldots, p$. Se pueden convertir estas restricciones a la forma usada antes, así:

$$-x_i \leqslant 0 \qquad\qquad i = 1,2,\ldots, p$$

donde $b_i = 0$ $i = 1,2,\ldots, p$ y $g_i(x) = -x_i$ el multiplicador de Lagrange asociado a cada una de estas restricciones deberá ser no positivo. Las condiciones de Kuhn-Tucker son:

$$\frac{\partial f(x)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(x)}{\partial xj} \Bigg|_{\underline{x}=\underline{x}^o} = 0 \qquad j = 1,2,\ldots, n$$

$$\lambda_i^o x_i^o = 0 \qquad i = 1,2,\ldots, p$$

$$\lambda_i^o (g_i(x^o)-b_j) = 0 \qquad i = p+1,\ldots, m$$

$$\lambda_i^o \leqslant 0 \qquad i = 1,2,\ldots, m$$

$$x_i \geqslant 0 \qquad i = 1,\ldots, p$$

$$g_i(x^o) \leqslant b_i \qquad i = p+1,\ldots, m$$

<u>Derivación alternativa de las condiciones de Kuhn-Tucker.</u>

Existen alternativas para derivar las condiciones de Kuhn-Tucker a continuación, se muestra la derivación hecha por Bernholtz:

Supóngase que $f(x)$ es diferenciable y que las funciones $g_i(x)$ $i = 1,2,\ldots, m$ son continuamente diferenciables. Si $f(x)$ tiene un máximo relativo en x^o sujeto a $g_i(x) = 0$; entonces, existirán constantes $\lambda_i^o,\ldots, \lambda_m^o$ tales que

$$\frac{\partial f(x)}{\partial xj} + \sum_{i=1}^{m} \lambda_i^o \frac{\partial g_i(x)}{\partial x_j} \Bigg|_{x=x^o} = 0 \qquad j = 1,2,\ldots, n$$

$$g_i(x) = 0 \qquad i = 1,2,\ldots, m$$

Cuando evaluamos en x^o con tal de que se obtenga uno de los dos conjuntos de condiciones mutuamente excluyentes

A) (1) $m \geqslant n$

(2) en x^o, al menos uno de los Jacobianos de orden n no se anula.

ó

B) (1) Existe un entero h, $0 < h < n$ tal que en x^o todos los Jacobianos de orden mayor que h se anulan, pero al menos uno de los Jacobianos de orden h no se anula.

(2) Correspondiente a cada vector diferencial $dX = (dx_1, dx_2, \ldots, dx_n)$ para el cual $dg_i(x) = 0$, existe un arco diferencial $x = a(t)$ definida para $-1 \leqslant t \leqslant 1$ el cual se encuentra en la intersección de las superficies $g_i(x) = 0$ y es tal que $x^o = a(0)$ y $dx = a'(0)dt$.

La prueba es como sigue:

Supóngase que se obtienen las condiciones de A. Entonces las columnas del Jacobiano son linealmente independientes y forman una base para algún espacio vectorial lineal de dimensión n. Así, se pueden introducir m constantes λ_1^o para satisfacer la ecuación lineal no homogénea.

$$\sum_{i=1}^{m} \lambda_i^o \left.\frac{\partial g_i(x)}{\partial x_j}\right|_{x=x^o} = -\left.\frac{\partial f(x)}{\partial x_j}\right|_{x=x^o}$$

Un argumento similar puede ser obtenido si las condiciones de B son satisfechas. Bernholtz deriva a continuación un resultado similar a las relaciones dadas previamente para el coeficiente de sensibilidad.

$$\frac{\partial f(x^o)}{\partial b_i} = -\lambda_i^o \qquad i = 1,2,\ldots, m$$

Sea $z = x^o + \theta (y-x^o)$ donde x^o es un máximo relativo de $f(x)$ sujeto a $g_i(x) = 0 \quad i = 1,2,\ldots, m$ y "y" es seleccionado de tal forma que:

$$\sum_{i=1}^{n} \frac{\partial g_j(x^o)}{\partial x_i} (y_i - x_i^o) \neq 0$$

para algún $J \in I$, donde $I = \{1,2,\ldots, n\}$ si se cumple A y $I = \{1,2,\ldots, h\}$ si se cumple B; y además:

$$\sum_{i=1}^{n} \frac{\partial g_k(x^o)}{\partial xi} (y_i - x_i^o) = 0$$

para toda k en I excepto k = j

Entonces por el teorema de Taylor se tiene:

$$f(z) + \sum_{i \in I} \lambda_i^o g_i(z) = f(x^o) + \sum_{\lambda \in I} \lambda_i^o g_i(x^o)$$

$$+ \theta \sum_{i=1}^{n} \left.\frac{\partial f(x)}{\partial y_i}\right|_{x=x^o} (y_i - x_i^o)$$

$$+ \theta \sum_{i=1}^{n} \lambda_j \left.\frac{\partial g_j(x)}{\partial x_i}\right|_{x=x^o} (y_i - x_i^o) + \theta\ \delta(\theta)$$

$$f(z) + \sum_{i\,\varepsilon I} \lambda_i^o g_i(z) = f(x^o) + \theta\,\delta\,(\theta)$$

donde $\delta(\theta) \longrightarrow 0$ cuando $\theta \longrightarrow 0$. Usando el teorema de Taylor nuevamente se obtiene:

$$g_j(z) = \theta \sum_{i=1}^{n} \left.\frac{\partial g_j(x)}{\partial x_i}\right|_{x=x^o} (y_i - x_i^o) + \delta_j(\theta)\,\theta = (A + \delta_j)\,\theta$$

donde $\delta \longrightarrow 0$ cuando $\theta \longrightarrow 0$ y A es una constante. Por lo tanto

$$\frac{f(z) - f(x^o)}{g_j(z)} + \lambda_j^o = - \frac{\lambda_j^o - (\sum_{i\varepsilon I} \lambda_i g_i(z) + \theta\ \delta(\theta))}{g_j\ (z)}$$

$$= - \lambda_j - ((A + \delta_j)\theta\ \lambda_j + \sum_{\substack{i\varepsilon I\\ i\neq j}} \lambda_i \delta_i\ \theta - \theta\delta(\theta))/(A + \delta_j)\theta$$

$$= -(\sum_{\substack{i\varepsilon I\\ i\neq j}} \lambda_i^o \delta_i\ \theta - \theta\ \delta(\theta))/\ (A + \delta_j)\ \theta$$

$$= -(\sum_{\substack{i\varepsilon I\\ i\neq j}} \lambda_i^o\ \delta_i\ \theta - \theta\ \delta(\theta))/(A + \delta j)\theta$$

$$= 0$$

Tomando el límite cuando $\theta \longrightarrow 0$ el lado derecho es igual a cero y el resultado es:

$$\lim_{\theta \to 0} \frac{f(z) - f(x^o)}{g_j(z)} = - \lambda_j^o$$

Cuando $g_i(x) < 0$, las condiciones necesarias

$$\frac{\partial f(x)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i^o \frac{\partial g_i(x)}{\partial x_j}\Bigg|_{x=x^o} = 0 \qquad j = 1,2,\ldots\ n$$

son satisfechas aún si las condiciones A o B se cumplen.

Esta prueba es similar a la anterior.

Para mostrar que $\lambda_j^o < 0$ se usará la definición de λ_j^o

$$\lim_{\theta \to 0} \frac{f(z) - f(x^o)}{g_i(z)} = - \lambda_j^o$$

seleccionando a θ lo suficientemente pequeño como para que $g_j(z) < 0$ y $f(z) < f(x^o)$ por definición de máximo relativo, se tiene que:

$$\frac{f(z) - f(x^o)}{g_j(z)} \geqslant 0$$

y así $\lambda_j < 0$ para toda λ_j^o no exactamente igual a cero

PROGRAMACION CUADRATICA

En esta sección se tratarán los diferentes algoritmos que existen para resolver un problema de programación cuadrática.

Un problema de programación cuadrática se define como la minimización (o maximización) de una forma cuadrática, como función objetivo sujeta a restricciones lineales.

La representación matemática de dicho problema es:

minimizar $f(X) = C^T X + 1/2\ X^T CX$

sujeta a:

$$A X \leqslant b$$

$$X \geqslant 0$$

donde:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad C = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} \qquad G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Se asumirá que G es simétrica y definida positiva, ya que si G es nula se tendría un problema estándar de programación lineal.

Algoritmo de Wolfe:

La aplicación de las condiciones de Kuhn-Tucker al problema de programación cuadrática, constituye el fundamento del algoritmo de Wolfe para la resolución de problemas de esta naturaleza. De la aplicación resulta un sistema de ecuaciones lineales, las cuales, mediante el uso de técnicas de programación lineal se resuelven, generando así, la solución óptima del problema.

En primer lugar asumamos la siguiente notación:

-) a_i denotará la i-ésima columna de la matriz A.

-) g_i denotará la i-ésima columna de la matriz G.

Introduciendo las variables de holgura q_i^2 y r_j^2 el problema se convierte en:

$$\min f(x)$$

sujeto a:

$$a_i^T X + q_i^2 = b_i \qquad i = 1,2,\dots m$$

$$-x_j + r_j^2 = 0 \qquad j = 1,2,\dots, n$$

Ahora, procedemos a formar la función de Lagrange, teniendo:

$$F(x, \lambda, U) = f(x) + \sum_{i=1}^{m} \lambda_i (a_i^T X + q_i^2 - b_i^2) + \sum_{j=1}^{n} U_j(-xj + r_j^2)$$

formando las condiciones necesarias, se tiene:

$$\frac{\partial F(\ )}{\partial xj} = \frac{f}{xj} + \sum_{i=1}^{m} \lambda_i a_{ij} - U_j = 0 \qquad j = 1,2,\ldots\ n$$

$$\lambda_i (a_i^T x - b_i) = 0 \qquad i = 1,2,\ldots, m$$

$$U_i x_i = 0 \qquad i = 1,2,\ldots, n$$

$$Ax \leqslant b$$

y finalmente x, λ y U deben ser todos no negativos.

Reescribiendo la 1a. ecuación, se tiene:

$$\frac{\partial F}{\partial xj} = c_j + \sum_{i=1}^{n} g_{ij} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} - U_j = 0 \qquad j = 1,2,\ldots, m$$

definiendo $y_i = q_i^2$ $(y_i \geqslant 0)$. Con esta definición, se tiene que $a_i^T x - b_i = -y_i$ y se puede simplificar la segunda condición para - leer $\lambda_i y_i = 0$. Así, el problema queda convertido en:

$$-U_j + C_j + \sum_{i=1}^{n} g_{ij} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} = 0 \qquad j = 1,2,\ldots, n$$

$$Ax + Iy = b$$

$$x \geqslant 0$$

$$y \geqslant 0$$

$$\lambda \geqslant 0$$

$$U \geqslant 0$$

$$\lambda_i y_i = 0 \qquad i = 1,2,\ldots, m$$

$$U_j x_j = 0 \qquad j = 1,2,\ldots, n$$

Nótese un importante resultado, con excepción de las dos últimas igualdades, las condiciones de Kuhn-Tucker son lineales. El problema es encontrar una solución a este conjunto de ecuaciones e inecuaciones lineales donde las variables deben ser no negativas y las m+n condiciones de holgura deben ser satisfechas.

Wolfe sugirió un procedimiento de solución para este problema usando el algoritmo simplex con una pequeña modificación. Su sugerencia, incluye la introducción de n variables artificiales no negativas v_j en las ecuaciones que representan $\partial F/\partial x_j = 0$, así:

$$-U_j + C_j + \sum_{i=1}^{n} g_{ji} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} - v_j = 0$$

comenzando con la solución básica inicial de v = c, y = b y x = 0 $\lambda = 0$ se minimiza

$$\sum_{i=1}^{n} v_i$$

sujeto a: $\quad -U + C + g_i^T x + a_I^T \lambda \equiv v = 0 \qquad i = 1,2,\ldots, n$

$$\Lambda x + y = b$$

$$U \geq 0$$
$$x \geq 0$$
$$y \geq 0$$
$$\lambda \geq 0$$

y las condiciones complementarias de holgura

$$\lambda_i y_i = 0 \qquad i = 1,2,\ldots, m$$
$$U_j x_j = 0 \qquad j = 1,2,\ldots, n$$

Con una excepción (las dos últimas igualdades) este es un problema de programación lineal, que puede ser resuelto por el algoritmo simples.

Se debe modificar el algoritmo, para incluir las condiciones complementarias de holgura.

Así, cuando se decida que y_i introducir en la solución básica, primero se debe estar seguro que λ_i no está en la solución o que λ_i será removido cuando entre y_i; de igual forma se procederá con x_j y U_j.

Obsérvese que el problema tiene $2(n+m)$ variables y $n+m$ restricciones lineales, junto con $m+n$ condiciones complementarias de holgura $\lambda_i y_i = 0$ y $U_j x_j = 0$.

Ejemplo:

Considerese el siguiente problema:

$$\max f(x_1,x_2) = 10x_1 + 25x_2 - 10x_1^2 - x_2^2 - 4x_1x_2$$

sujeto a:

$$x_1 + 2x_2 \leq 10$$

$$x_1 + x_2 \leq 9$$

$$x_1 \,.\, x_2 \geq 0$$

Recordemos que cuando maximizamos f(x), los multiplicadores de Lagrange deben ser todos no positivos. Ya que se desea usar la técnica del simplex, se convertirá el problema en uno de minimización, multiplicando la función objetivo por -1. Esta operación cambia el signo de los multiplicadores de Lagrange. Después de realizar esta transformación y agregar las variables de holgura se tiene:

$$\min -f(x_1,x_2) = -10x_1 - 25x_2 + 10x_1^2 + x_1^2 + 4x_1x_2$$

sujeta a:

$$x_1 + 2x_2 + q_1^2 - 10 = 0$$

$$x_1 + x_2 + q_2^2 - 9 = 0$$

$$-x_1 + r_1^2 = 0$$

$$-x_2 + r_2 = 0$$

al construir la función de Lagrange, se obtiene:

$$\begin{aligned} F(x_1,x_2,U_1,U_2,r_1,r_2,\lambda_1,\lambda_2,q_1,q_2) = & -10x_1-25x_2 +10x_1^2 + x_2^2+4x_1x_2 \\ & +\lambda_1(x_1+2x_2 + q_1^2 - 10) \\ & +\lambda_2(x_1 + x_2 + q_2^2 - 9) \\ & +U_1(-x_1 + r_1^2) +U_2(-x_2 + r_2^2) \end{aligned}$$

Las condiciones necesarias y suficientes son:

$$\frac{\partial F}{\partial x_1} = -10 + 20x_1 + 4x_2 + \lambda_1 + \lambda_2 - U_1 = 0$$

$$\frac{\partial F}{\partial x_2} = -25 + 4x_1 + 2x_2 + 2\lambda_1 + \lambda_2 - U_2 = 0$$

$$\lambda_1 y_1 = 0$$

$$\lambda_2 y_2 = 0$$

$$U_1 x_1 = 0$$

$$U_2 x_2 = 0$$

$$x_1 + 2x_2 + y_1 = 10$$

$$x_1 + x_2 + y_2 = 9$$

$$x_1, x_2, y_1, y_2, U_1, U_2, \lambda_1, \lambda_2 \geq 0 \quad \text{donde } y_1 = q_i^2$$

Introduciendo variables artificiales v_1 y v_2, se puede construir el siguiente problema lineal modificado:

$$\text{máx} \quad g(v_1, v_2) = -v_1 - v_2$$

sujeta a

$$20x_1 + 4x_2 + \lambda_1 + \lambda_2 - U_1 + v_1 = 10$$

$$4x_1 + 2x_2 + 2\lambda_1 + \lambda_2 - U_2 + v_2 = 25$$

$$x_1 + 2x_2 + y_1 = 10$$

$$x_1 + x_2 + y_2 = 9$$

Con todas las variables no negativas y

$$x_1 U_1 = 0 \qquad \lambda_1 y_1 = 0$$

$$x_2 U_2 = 0 \qquad \lambda_2 y_2 = 0$$

La primera tabla sería:

	b	v_1	v_2	x_1	x_2	λ_1	λ_2	μ_1	μ_2	y_1	y_2
v_1	10	1	0	20	4	1	1	-1	0	0	0
v_2	25	0	1	4	2	2	1	0	1	0	0
y_1	10	0	0	1	2	0	0	0	0	1	0
y_2	9	0	0	1	1	0	0	0	0	0	1
		0	0	-24	-6	-3	-2	1	1	0	0

Introduciendo x_1 en la solución básica, se obtiene:

	b	v_1	v_2	x_1	x_2	λ_1	λ_2	μ_1	μ_2	y_1	y_2
x_1	1/2	1/20	0	1	1/5	1/20	1/20	-1/20	0	0	0
v_2	23	-1/5	1	0	6/5	9/5	4/5	1/5	-1	0	0
y_1	9.5	-1/20	0	0	9/5	-1/20	-1/20	1/20	0	1	0
y_2	8.5	-1/20	0	0	4/5	-1/20	-1/20	1/20	0	0	1
		6/5	0	0	-6/5	-9/5	-4/5	-1/5	1	0	0

Ahora x_2 entra a la solución básica. Nótese que ya sea λ_1 o λ_2 no pueden entrar, ya que y_1 y y_2 ambas son variables básicas.

	b	v_1	v_2	x_1	x_2	λ_1	λ_2	U_1	U_2	y_1	y_2
x_2	5/2	1/4	0	5	1	1/4	1/4	-1/4	0	0	0
v_2	20	-1/2	1	-6	0	3/2	1/2	1/2	-1	0	0
y_1	5	-1/2	0	-9	0	-1/2	1/2	1/2	0	1	0
y_2	6.5	-1/4	0	-4	0	-1/4	-1/4	1/4	0	0	1
		3/2	0	6	0	-3/2	-1/2	-1/2	1	0	0

U_1 es introducido en la siguiente solución

	b	v_1	v_2	x_1	x_2	λ_1	λ_2	U_1	U_2	y_1	y_2
x_2	5	0	0	1/2	1	0	0	0	0	1/2	0
v_2	15	0	1	3	0	2	1	0	-1	-1	0
U_1	10	-1	0	-18	0	-1	-1	1	0	2	0
y_2	4	0	0	1/2	0	0	0	0	0	-1/2	1
		1	0	-3	0	-2	-1	0	1	1	0

La regla normal diría que x_1 debe ser introducido en este momento. sin embargo ya que U_1 está en la solución. x_1 no puede ser introducido sin sacar a U_1. Así, se introducirá a λ_1

	b	v_1	v_2	x_1	x_2	λ_1	λ_2	U_1	U_2	y_1	y_2
x_1	5	0	0	1/2	1	0	0	0	0	1/2	0
λ_1	7.5	0	1/2	3/2	0	1	1/2	0	-1/2	-1/2	0
U_1	17.5	-1	1/2	3/2	0	0	-1/2	1	-1/2	3/2	0
y_2	4	0	0	1/2	0	0	0	0	0	-1/2	1
		1	1	0	0	0	0	0	0	0	[illegible]

Ya que v_1 y v_2 están fuera de la solución, el cálculo ha sido completado. La solución óptima es $x_1 = 0$, $x_2 = 5$ $v_1 = 0$, $v_2 = 4$, $\lambda_1 = 7.5$, $\lambda_2 = 0$, $U_1 = 17.5$, $U_2 = 0$ los cuales además, satisfacen las condiciones de holgura complementarias y las restricciones en los signos de los multiplicadores de Lagrange.

Algoritmo de Beale.

Beale ha desarrollado un método que no utiliza las condiciones de Kuhn-Tucker, en el procedimiento para resolver un problema de programación cuadrática. Su método involucra partir las variables en variables básicas y no básicas en cada iteración y escribir la función objetivo en términos de las variables no básicas.

Considérese que se tienen el siguiente conjunto de restricciónes:

$$D_z \leqslant b$$

donde:

D es una matriz de orden mxn

b es un vector de n componentes

z es un vector de n variables

Al introducir variables de holgura y_1 i: 1,2,.... m para las m restricciones de desigualdad dadas, se obtiene el siguiente conjunto - de restricciones.

$$Ax = b$$

donde:

i) A es una matriz de m x (n+m) y tiene la forma:

$$A = (D, I)$$

ii) x es el vector de variables de orden (m+n)x1 de la forma

$$x = \begin{pmatrix} z \\ y \end{pmatrix}$$

Se pueden seleccionar m de las (m+n) variables para ser las variables básicas y las restantes n serán las variables no básicas. Al realizar esta partición, se obtiene:

$$Bx_B + Rx_{NB} = b$$

donde:

x_B denota el vector de las variables básicas

x_{NB} denota el vector de variables no básicas, las cuales son cero.

B es la submatriz de A, asociada con x_B

R es la submatriz de A, asociada con x_{NB}

De la ecuación anterior, se puede obtener que

$$x_B = B^{-1}b - B^{-1}Rx_{NB}$$

Al sustituir este resultado en la función objetivo, se llega a convertir en una función solamente de x_{NB}.

Supongamos que se trata de un problema de maximización, y además que al derivar la función objetivo con respecto a cada una de las variables no básicas se obtiene que:

$$\frac{\partial z}{\partial x_{NBi}} > 0 \qquad \text{para alguna } i = 1,2,\ldots, n$$

(NOTA: en caso de tenerse más de una, se tomará la que sea más positiva).

Este último resultado nos indica que al incrementar el valor de la variable x_{NBi} se aumentará el valor de la función objetivo; pero el aumento de x_{NBi} estará limitado al mínimo de:

i) El mayor valor que puede obtener x_{NBi} sin convertir a una variable básica en negativa.

ii) El valor de x_{NBi} para el cual $\frac{\partial z}{\partial x_{NBi}} = 0$

Si el valor de x_{NBi} se obtiene por la condición i), entonces x_{NBi} entrará a formar parte de la solución básica y saldrá de la solución básica aquella variable que se ha hecho exactamente igual a cero.

Si el valor se obtiene de la condición ii), entonces simplemente se define una nueva variable, y_i, así:

$$y_i = \frac{\partial z}{\partial x_{NBi}}$$

y se agrega una nueva restricción $y_i = 0$; además se introduce x_{NBi} a la solución básica.

El procedimiento es repetido hasta que no se puedan obtener mejoras en la función objetivo mediante incremento en una de las variables no básicas.

Ejemplo:

A continuación se resolverá el mismo problema que se desarrolló con el método de Wolfe.

$$\max f(x_1, x_2) = 10x_1 + 25x_2 - 10x_1^2 - x_2^2 - 4x_1x_2$$

sujeta a:

$$x_1 + 2x_2 \leq 10$$

$$x_1 + x_2 \leq 9$$

$$x_1, x_2 \geq 0$$

Introduciendo las variables de holgura, se obtiene el conjunto de restricciones

$$x_1 + 2x_2 + x_3 = 10$$

$$x_1 + x_2 + x_4 = 9$$

$$x_1, x_2, x_3, x_4 \geq 0$$

Seleccionando a x_1 y x_2 como variables básicas, tenemos:

$$x_1 = 8 + x_3 - 2x_4$$

$$x_2 = 1 - x_3 + x_4$$

donde

$$x_B = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad x_{NB} = \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}$$

Expresando a f( ) en términos de x_{NB} se obtiene:

$$f(x_3, x_4) = 10(8+x_3-2x_1) + 25(1-x_3+x_4)-10(8+x_3-2x_4)^2$$
$$- (1-x_3 + x_4)^2 - 4(8+x_3-2x_4)(1-x_3+x_4)$$

$$\frac{\partial f}{\partial x_3} = 10 - 25 - 20(8+x_3-2x_4) + 2(1-x_3 + x_4)$$
$$-4(1-x_3 + x_4) + 4(8 + x_3 - 2x_4)$$

Evaluando esta derivada parcial en $x_{NB} = 0$, resulta:

$$\frac{\partial f(x_{NB})}{\partial x_3} = -145$$

Luego, la función objetivo decrecerá si x_3 es aumentado.

La derivada parcial con respecto a x_4 resulta ser:

$$\frac{\partial f(x_{NB})}{\partial x_4} = -20 + 25 - 20(-2)(8+x_3-2x_4)-2(1-x_3+x_4)$$
$$+ 8(1-x_3+x_4) - 4(8+x_3 - 2x_4)$$

Evaluando, se tiene:

$$\frac{\partial f(x_{NB})}{\partial x_4} = 299$$

Luego, al incrementar x_4 se mejorará la función objetivo.

Ahora se determinará cuanto puede o debe ser incrementado x_4, sus posibles valores son:

i) Si se aumenta el valor de x_4 mayor que 4, x_1 será negativo, ya que $x_1 = 8 + x_3 - 2x_4$ y $x_3 = 0$

ii) La derivada parcial con respecto a x_4 se anula en $x_4 = 299/66$. Tomando el mínimo de estos dos, es decir, $x_4 = 4$, se tienen las siguientes nuevas variables básicas x_4 y x_2, se iniciará una nueva iteración resolviendo x_2 también x_4 en términos de x_1 y x_3. así:

$$x_2 = 5 - 1/2\,(x_1 + x_3)$$

$$x_4 = 4 + 1/2(x_3 - x_1)$$

en este caso

$$x_B = \begin{pmatrix} x_2 \\ x_4 \end{pmatrix} \qquad\qquad x_{NB} = \begin{pmatrix} x_1 \\ x_3 \end{pmatrix}$$

Expresando a f( ) en términos de x_{NB} resulta:

$$f(x_1, x_3) = 10x_1 + 25(5-1/2(x_1+x_3)) - 10x_1^2$$

$$-(5 - 1/2(x_1 + x_3))^2 - 4x_1(5 - 1/2(x_1 + x_3))$$

$$\left.\frac{\partial f}{\partial x_1}\right|_{x_{NB}=0} = -\ 35/2$$

$$\left.\frac{\partial f}{\partial x_2}\right|_{x_{NB}=0} = -15/2$$

Por lo tanto, ninguna variable no básica puede ser introducida para incrementar a f( ); y así, la solución óptima ha sido obtenida.

La solucion es: $x_1 = 0$; $x_3 = 0$; $x_2 = 5$ y $x_4 = 4$

Algoritmo de Theil y Van der Panne.

Este algoritmo está basado en la noción que para un problema de la forma:

$$\max f(x) = cx + x^T Gx$$

sujeto a:

$$Ax \leq b$$

donde en $Ax \leq b$ se tienen incorporadas las restricciones de la forma $x \geq 0$

Resulta que si x^o es una solución óptima del problema se cumple que:

$$a_i^T x^o = b_i \qquad i = 1,2,\ldots, p$$

$$a_i^T x^o < b_i \qquad i = p+1,\ldots, m$$

Es decir, p restricciones son satisfechas en forma de igualdad y las restantes m-p en forma de desigualdad.

Resulta obvio que x^o será también una solución óptima para el problema

$$\max f(x) = cx + x^T Gx$$

sujeto a

$$a_i^T x = b_i \qquad i = 1,2,\ldots, p$$

A continuación se muestra el procedimiento de este algoritmo.

El primer paso de este algoritmo es resolver el problema sin restricciones.

$$\max f(x) = cx + x^T Gx$$

A la solución óptima de este problema se le llamará x_1^o. Si se cumple que $Ax_1^o \leq b$, entonces, resulta que x_1^o es la solución óptima al problema original de programación cuadrática.

Supóngase, sin embargo, que en x_1^o resulta que

$$a_i^T x_1^o > b_i \qquad i = 1,2,\ldots, p$$

es decir, no se satisfacen p restricciones.

Entonces se resuelven p problemas de la forma

$$\max f(x) = cx + x^T Gx$$

sujeto a

$$a_i^T x = b_i$$

Las soluciones a estos problemas serán denotadas por x_{2i}^o $i = 1,2\ldots,p$

Si alguna de estas p soluciones (dígase x_{2j}^o, $j < p$) es factible con el problema original, es decir, satisface todas las restricciones del problema original, entonces dicha solución(x_{2j}^o) es la solución óptima al problema original.

Si por otro lado, ninguna de estas p soluciones resultó ser la solución óptima, entonces, por cada solución x_{2i}^o un subconjunto de las restantes restricciones es violado, es decir:

$$a_j^T x_{2j}^o > b_i \qquad j \in V(x_{2j}^o) \qquad j \leq p$$

donde:

$V(x)$ es el conjunto de índices asociado con las restricciones violadas en el punto x.

A continuación se resolverán una sucesión de problemas, sujetos a dos restricciones de igualdad. Si una solución óptima de cualquiera de estos problemas resulta ser factible con el problema original y si la solución a los problemas sujetos a una sola restricción activa viola la restricción liberada o inactiva, entonces la solución resulta ser la solución óptima del problema original.

A continuación se describe el proceso en una forma más precisa. Considere problemas sujetos a k restricciones activas.

$$a_i^T x = b_i \qquad\qquad i = 1,2,\ldots k$$

Además, supóngase que al menos una restricción del problema original es violada por cada solución del conjunto anterior de problemas, es decir,

$$V(x_k^O) \neq 0 \qquad\qquad \text{para toda } k$$

Entonces, a cada conjunto de restricciones activas de las cuales se obtuvo x_k^O se le agregará una de las restricciones del conjunto - $V(x_k^O)$ y se resolverá este nuevo problema

La solución a uno de estos nuevos problemas, x_{k+1}^O, será la solución óptima del problema original, si y solo si:

1.) $Ax_{k+1}^O \leqslant b$

2.) La solución óptima del conjunto de k+1 problemas generados al eliminar, una a la vez, una de las restricciones, viola la restricción que ha sido eliminada.

Ejemplo: A continuación se resolverá por este algoritmo, el mismo problema que ha sido resuelto por los otros dos algoritmos.

Considere el problema

$$\max F(x_1, x_2) = 10x_1 + 25x_2 - 10x_1^2 - x_2^2 - 4x_1x_2$$

sujeto a:

$$\begin{array}{ll} (1) & x_1 + x_2 \leq 9 \\ (2) & x_1 + 2x_2 \leq 10 \\ (3) & x_1 \geq 0 \\ (4) & x_2 \geq 0 \end{array}$$

El algoritmo de Theil-Van der Panne primero resuelve el problema sin restricciones.

$$\max f(x_1, x_2) = 10x_1 + 25x_2 - 10x_1^2 - x_2^2 - 4x_1x_2$$

Resolviendo este problema, se tiene:

$$\frac{\partial f}{\partial x_1} = 10 - 20x_1 - 4x_2 = 0$$

$$\frac{\partial f}{\partial x_2} = 25 - 4x_1 - 2x_2 = 0$$

La solución a estas ecuaciones es $x_1 = -\frac{40}{12}$, $x_2 = \frac{115}{6}$

Este punto, viola las restricciones 1, 2 y 3. Así, se tiene que:

$$\{j\} = V(x_1^o) = \{1,2,3\}$$

Siguiendo el algoritmo ahora, tendríamos que resolver tres nuevos problemas. Ellos son:

1. $\max f(x_1, x_2)$ sujeto a $x_1 + x_2 = 9$
2. $\max f(x_1, x_2)$ sujeto a $x_1 + 2x_2 = 10$
3. $\max f(x_1, x_2)$ sujeto a $x_1 = 10$

al resolver el problema 1 resulta que:

$$x_1 = -\frac{33}{14} \qquad x_2 = \frac{159}{14}$$

lo cual viola las restricciones 2 y 3. y por lo tanto

$$V(x_{21}^{o}) = \{2,3\}$$

La solución al problema 2 resulta ser el punto

$$x_1 = -\frac{70}{66} \qquad x_2 = \frac{365}{66}$$

esto viola la restricción 3, teniendo así que:

$$V(x_{22}^{o}) = \{3\}$$

La solucion del problema 3 será:

$$x_1 = 0 \qquad x_2 = \frac{25}{2}$$

la cual viola las restricciones 1 y 2. Por lo tanto

$$V(x_{23}^{o}) = \{1,2\}$$

Ahora, deberemos resolver una sucesión de problemas sujetos a dos restricciones de igualdad.

El problema 1 de la iteración previa, genera dos problemas nuevos.

los cuales son:

1.1 max $f(x_1, x_2)$

sujeto a

$$x_1 + x_2 = 9$$

$$x_1 + 2x_2 = 10$$

y

1.2 max $f(x_1, x_2)$

sujeto a

$$x_1 + x_2 = 9$$

$$x_1 = 0$$

la solución al problema 1.2 ($x_1 = 0$, $x_2 = 9$) viola la restricción 2. Resolviendo el problema 1.1 resulta que $x_1 = 8$, $x_2 = 1$, la cual satisface todas las restricciones del problema original. Ahora se aplicará la segunda parte de la regla que se ha dado. La regla dice que al resolver los problemas generados al liberar una restricción a la vez deben resultar puntos que violen la restricción liberada. Los problemas con las restricciones liberadas una a la vez son simplemente los problemas 1 y 2 respectivamente. La solución al problema 1 viola la restricción 2, pero la solución al problema 2 no viola la restricción 1, y así, nuestra regla nos indica que el punto no es óptimo.

Continuando con el proceso, generaremos a partir del problema 2 el siguiente problema.

2.1 max $f(x_1, x_2)$

sujeto a

$$x_1 + 2x_2 = 10$$

$$x_1 = 0$$

la solución a este problema resulta ser:

$x_1 = 0$ $\qquad\qquad$ $x_2 = 5$

Aplicando la segunda parte de la regla para chequear la optimalidad del punto, resulta que al liberar una de estas restricciones a la vez, se obtienen los siguientes problemas.

max $f(x_1, x_2)$ $\qquad$ y $\qquad$ max $f(x_1, x_2)$

sujeto a $x_1 + 2x_2 = 10$ $\qquad\qquad$ sujeto a $x_1 = 0$

pero, resulta que estos son los problemas 2 y 3 respectivamente. La solución al problema 2 viola la restricción 3 y la solución al problema 3 viola la restricción 2. Así, la regla es satisfecha y el punto $x_1 = 0$; $x_2 = 5$ ha resultado ser la solución óptima del problema original.

PROGRAMACION CONVEXA

En esta sección se estudiará el problema de Programación Convexa, como una primera aproximación podemos decir que un problema de programación convexa, es aquel en el cual se busca minimizar una función convexa; sin embargo, no debe de olvidarse que minimizar una función convexa F(x) es equivalente, a maximizar una función cóncava f(x), siempre que $f(x) = - F(x)$

Formalmente, un problema de programación convexa se define así:

Definición: Sea U un conjunto convexo en R^n y $f: U \to R$ una función cóncava. Un problema de programación convexa es: maximizar f(x) sujeta a $g_i(x) < 0$ $\qquad i = 1,2,\dots, m$

donde cada una de las funciones $g_i: U \to R$ son convexas.

Por conveniencia en la notación, se definirá una función vectorial $G: U \to R^m$, tal que $G(x) = (g_1(x),\dots,g_m(x)$.

El conjunto factible para el problema de programación convexa, se define así:

$$\mathcal{F} = \{x \in U: G(x) < 0\}$$

El problema requiere que se encuentre un punto $x^* \in \mathcal{F}$, tal que, si ese punto existe, $f(x) \leq f(x^*)$ para cualquier otra $x \in \mathcal{F}$

Resulta fácil demostrar que el conjunto $\mathcal{F}$ de soluciones factibles es convexo, ya que:

$$\mathcal{F} = \bigcap_1^m \{ x \in U: g_i(x) \leq 0 \}$$

Es decir, $\mathcal{F}$ es la intersección de un número finito de conjuntos convexos.

Si todas las funciones involucradas en el problema fueran diferenciables, la técnica del multiplicador de Lagrange pudiese ser utilizada; es decir, tendríamos la función:

$$F(x, \lambda) = f(x) + \lambda_1 g_1(x) + \lambda_2 g_2(x) + \ldots + \lambda_m g_m(x)$$

aún cuando no se ha tomado ningún supuesto de diferenciabilidad, resulta útil considerar esta función; obsérvese que cuando $\lambda \leq 0$ se fija, la función $F(x, \lambda)$ es una función cóncava de x en U; y además, cuando se fija $x \in U$, $F(x, \lambda)$ es una función convexa de λ.

Para tal función, el punto $(x^*, \lambda^*) \in U x R_-^m$ es un punto de silla, si para toda $x \in U$, $\lambda \in R_-^m$ se cumple que:

$$F(x, \lambda^*) \leq F(x^*, \lambda^*) \leq F(x^*, \lambda)$$

La cuestión de saber si un problema de programación convexa tiene o no solución, está intimamente relacionado a la cuestión de que la función $F(x, \lambda)$ definida anteriormente, tenga o no un punto de silla. Para establecer precisamente la conexión entre estos dos problemas primero se necesita definir el siguiente término: Un punto $x \in \mathcal{F}$ es llamado solución estrictamente factible del problema de programación convexo si $g_i(x) < 0$ para todo $i = 1,2,\ldots, m$

Teorema:

Nos referimos al problema de programación convexa como P y a la función $f(x, \lambda)$ definida anteriormente:

(a) Si F tiene un punto de silla $(x^*, \lambda^*) \in U \times R^m_-$, entonces x^* es una solución óptima de P.

(p) Supóngase que P tiene una solución estrictamente factible s. Entonces, si P tiene una solución óptima x^*, existe un $\lambda^* \in R^m_-$ tal que (x^*, λ^*) es un punto de silla para $F(x, \lambda)$

Prueba:

(a) Si $(x^*, \lambda^*) \in U \times R^m_-$ es un punto de silla para F; entonces, de la definición de F se tiene:

$$f(x) + \left\langle \lambda^*, G(x) \right\rangle \leqslant f(x^*) + \left\langle \lambda^*, G(x^*) \right\rangle \leqslant f(x^*) + \left\langle \lambda, G(x^*) \right\rangle \quad \ldots.(1)$$

para toda $x \in U$, $\lambda \in R^m_-$

De la desigualdad del lado derecho, se tiene:

$$\left\langle \lambda^*, G(x^*) \right\rangle \leqslant \left\langle \lambda, G(x^*) \right\rangle \quad \ldots\ldots. \quad (2)$$

ya que ésto debe cumplirse para toda $\lambda \leqslant 0$, resulta que $G(x^*)$ debe ser menor o igual que cero, por lo tanto $x^* \in \mathcal{F}$

También, ya que $G(x^*) \leqslant 0$, se tiene que $\left\langle \lambda^*, G(x^*) \right\rangle \geqslant 0$, pero si se selecciona $\lambda = 0$, se obtiene la desigualdad opuesta al sustituir en (2); por lo que podemos concluir $\left\langle \lambda^*, G(x^*) \right\rangle = 0$, así también se puede concluir que el lado izquierdo de (1) se puede reescribir así:

$$f(x) + \langle \lambda^*, G(x) \rangle < f(x^*)$$

para toda $x \in U$. Ahora, si $x \in \mathcal{F}$ donde $G(x) \leq 0$, se tiene que

$$\langle \lambda^*, G(x) \rangle > 0.$$

Por lo tanto, se puede concluir que $f(x) \leq f(x^*)$; con lo cual termina la prueba de la parte (a).

Parte (b): Para probar esta parte, primero se necesita definir dos conjuntos M y N en $R \times R^m$ de la siguiente forma:

$$M = \{ (\lambda_o, \lambda): \text{para algún } x \in U, \lambda_o \leq f(x), \lambda \geq G(x) \}$$

$$N = \{ (\lambda_o, \lambda): \lambda_o > f(x^*), \lambda_i < 0 \text{ para } i \geq 1 \}$$

El conjunto M es convexo, como puede verse al considerar una combinación convexa de puntos en M y haciendo uso del hecho que f es cóncava, y que las $g_i(x)$ son convexas. El conjunto N, siendo la intersección de m+1 semi-espacios, es también convexa. Ambos conjuntos tienen interiores no vacíos y finalmente, se desea mostrar que $M \cap N = \Phi$.

Considere un punto $(\lambda_o, \lambda_1, \lambda_2, \ldots, \lambda_m) \in M$. Si el correspondiente x no está en $\mathcal{F}$; entonces, para algún $i = 1,2,\ldots, m$, se debe cumplir que $g_i(x) > 0$, de aquí que $\lambda_i > 0$, lo cual excluye al punto de N. Y, si el correspondiente $x \in \mathcal{F}$, entonces de la condición de optamilidad de x^*

$$\lambda_0 \leq f(x) \leq f(x^*)$$

lo cual, nuevamente excluye el punto de N.

Ahora, estamos dispuestos a asegurar que existe un hiperplano H que separa a M y N. Es decir, existe un vector no nulo $(a_o, a) \in R \times R^m$ y un número real α tal que, para todo $(\lambda_o, \lambda) \in M$, $(z_o, z) \in N$.

$$a_o\lambda + \langle a, \lambda \rangle \leqslant \alpha \leqslant a_o z_o + \langle a, z \rangle$$

ya que el lado derecho de la desigualdad se cumple para toda $z_o > f(x^*)$ y toda $(z_1, z_2, \ldots z_m)$ con $z_i < 0$, se concluye que $a_o \geqslant 0$ y $a \leqslant \theta$.

Ya que para cualquier $x \in U$, el vector$(f(x), G(x)) \in M$ mientras que el vector

$$Z_\varepsilon = (f(x^*) + \varepsilon_o, -\varepsilon_1, -\varepsilon_2, \ldots, -\varepsilon_m) = (f(x^*) + \varepsilon_o, -\varepsilon)$$

está en N, se puede usar la última desigualdad para escribir.

$$a_o f(x) + \langle a, G(x) \rangle \leqslant a_o(f(x^*) + \varepsilon_o) - \langle a, \varepsilon \rangle$$

al hacer que $\varepsilon_o \to 0$ y $\varepsilon \to 0$, se obtiene:

$$a_o f(x) + \langle a, G(x) \rangle \leqslant a_o f(x^*)$$

esto último se cumple para toda $x \in U$, y en particular para la solución estrictamente factible s. Como $a_o > 0$, ya que si $a_o = 0$, entonces $\langle a, G(s) \rangle \leqslant 0$ lo cual implica que $a = 0$, violando el hecho de que el hiperplano H es determinado por un vector no nulo.

Así, se puede multiplicar la última desigualdad por $1/a_o$, obteniendo:

$$f(x) + 1/a_o \langle a, G(x) \rangle \leqslant f(x^*)$$

$$f(x) + \langle (1/a_o)a, G(x) \rangle \leqslant f(x^*)$$

haciendo $\lambda^* = (1/a_o)a$, se obtiene:

$$f(x) + \langle \lambda^*, G(x) \rangle \leq f(x^*)$$

para toda $x \in U$. Ya que ésto es cierto para $x = x^*$, $\langle \lambda^*, G(x^*) \rangle \leq 0$. La desigualdad contraria se cumple siempre para $x \in \mathcal{F}$, así

$\langle \lambda^*, G(x^*) \rangle = 0$

Pudiendo extender la última desigualdad, así:

$f(x) + \langle \lambda^*, G(x) \rangle \leq f(x^*) + \langle \lambda^*, G(x^*) \rangle \leq f(x^*) + \langle \lambda, G(x^*) \rangle$

lo cual se cumple para toda $x \in U$, $\lambda \in R^m_-$. Concluyendo así la prueba.

Las conclusiones de este teorema pueden, por supuesto, ser establecidas sin hacer referencia específica a la función $F(x, \lambda)$.

En la prueba de la parte (a), se dijo que si (x^*, λ^*), fuese un punto de silla, entonces

$$f(x) + \lambda_1^* g_1(x) + \ldots\ldots + \lambda_m^* g_m(x)$$

tiene un máximo en x^*, y

$$\lambda_1^* g_1(x^*) + \ldots\ldots + \lambda_m^* g_m(x^*) = 0$$

Estas condiciones fueron suficientes para implicar que P tiene una solución óptima. En la parte (b) se dijo que si P tiene una solución estrictamente factible; entonces, la existencia de una solución óptima en x^* implicaba las condiciones anteriores como una consecuencia necesaria. Ya que $\lambda^* \leq 0$, y $G(x^*) \leq 0$, la segunda condición nos dice que si $g_i(x^*) < 0$, entonces $\lambda_i = 0$. El teorema anterior se puede reestablecer de la siguiente forma:

Teorema: Supóngase que el problema de programación convexa tiene una solución estrictamente factible. Entonces para que P tenga una solución óptima en x^*, es necesario y suficiente que exista un $\lambda^* \varepsilon R^m_-$, que satisfaga las condiciones de Kuhn-Tucker; es decir, que $f(x) + \left[\lambda_1^* g_1(x) + \ldots . + \lambda_m^* g_m(x)\right]$ tenga un máximo en el conjunto U en x^*, y que si $g_i(x^*) < 0$ entonces $\lambda_i^* = 0$.

Ahora, se realizará un pequeño estudio acerca de lo que es el teorema de Dualidad para programación convexa.

Primero, reestableceremos lo que es el problema de programación convexa.

Problema Primal (P): Maximice la función cóncava $f: U \to R$, sujeta a las restricciones $g_i(x) \leqslant 0$, donde $g_i: U \to R$ es convexa $(i=1,2\ldots m)$

Definamos:

$$L : R^m_- \to R \text{ por}$$

$$L(\lambda) = \sup_{x \varepsilon U} F(x, \lambda) = \sup_{x \varepsilon U} \left[f(x) + \lambda_1 g_1(x) + \ldots \pm \lambda_m g_m(x) \right]$$

para una x fija, $F(x, \lambda)$ es una función afín de λ, y L es, por lo tanto, el supremo de una familia de funciones convexas (afín), resultando así, que L es convexa. Sea $\mathcal{F}^*$ el conjunto en el cual L es finita, llamemos a $\mathcal{F}^*$ el conjunto factible para la función L.

Problema Dual (P^*): Si $\mathcal{F}^* \neq \phi$, minimice la función convexa $L : R^m_- \to R$ en $\mathcal{F}^*$.

Teorema de Dualidad de Kuhn-Tucker:

Sea P un problema primal de programación convexa, y sea P^* el problema dual.

(i) Si x es una solución factible de P y λ es una solución factible de P^*, entonces $f(x) \leq L(\lambda)$

(ii) Supóngase que P tiene una solución estrictamente factible; entonces, si el problema P tiene una solución óptima x^*, el problema P^* tiene una solución óptima λ^* y $f(x^*) = L(\lambda^*)$.

Para poder demostrar este teorema de dualidad, se hace necesario el siguiente teorema:

Teorema: Sean U y V conjuntos arbitrarios y supóngase una función $k : U \times V \to R$. Entonces:

$$\sup_{x\varepsilon U} \inf_{y\varepsilon V} k(x, y) \leq \inf_{y\varepsilon V} \sup_{x\varepsilon U} k(x, y) \text{ -------} \quad (a)$$

además, si existe un punto $(x^*, y^*) \varepsilon U x_V$ tal que

$$k(x, y^*) \leq k(x^*, y^*) \leq k(x^*, y) \text{---------------} \quad (b)$$

para toda $(x,y) \varepsilon U \times V$, entonces (a) es una igualdad y ambos términos son iguales a $K(x^*, y^*)$.

<u>Prueba</u>: Sea $f(x) = \inf_{y\varepsilon V} k(x,y)$ y $g(y) = \sup_{x\varepsilon U} k(x,y)$

claramente $f(x) \leq k(x,y) \leq g(y)$

para toda $(x,y) \varepsilon U \times V$ y consecuentemente:

$$\sup_{x\varepsilon U} f(x) \leqslant \inf_{y\varepsilon V} g(y)$$

con lo cual, se prueba (a)

Ahora, supóngase que (x^*, y^*) satisfacen (b). Entonces

$$k(x^*, y^*) = \inf_{y\varepsilon V} k(x^*, y) = f(x^*) \leqslant \sup_{x\varepsilon U} f(x) \ldots\text{-----} \quad (c)$$

y

$$k(x^*, y^*) = \sup_{x\varepsilon U} k(x, y^*) = g(y^*) \geqslant \inf_{y\varepsilon V} g(y) \text{ --------} \quad (d)$$

de (c) y (d) resulta

$$\inf_{y\varepsilon V} g(y) \leqslant k(x^*, y^*) \leqslant \sup_{x\varepsilon U} f(x)$$

$$\inf_{y\varepsilon V} g(y) \leqslant \sup_{x\varepsilon U} f(x)$$

De este último resultado y de (a) se obtiene la igualdad deseada.

Ahora, se pasará a probar el Teorema de Dualidad

Prueba: (Teorema de Dualidad)

Si se define $F^* : U \to R$ por

$$F^*(x) = \inf_{\lambda\varepsilon R^m_-} \quad F(x, \lambda) = \begin{cases} f(x) & x \varepsilon \mathcal{F} \\ -\alpha & x \varepsilon U - \mathcal{F} \end{cases}$$

Entonces, de acuerdo al teorema anterior se tiene:

$$F^*(x) \leqslant \sup_{x\varepsilon U} F^*(x) \leqslant \inf_{y\varepsilon R^m_-} L(\lambda) \leqslant L(\lambda)$$

para toda $x \varepsilon U$, $\lambda \varepsilon R^m_-$. Esto nos indica que $f(x) \leqslant L(\lambda)$ para toda

$x \in \mathcal{F}$, y $\in \mathcal{F}^*$, estableciéndose así, la parte (i) del teorema de dualidad.

De los teoremas anteriores, se sabe que la existencia de una solución óptima x^* de P implica la existencia de un $\lambda^* \in R_+^m$ tal que (x^*, λ^*) es un punto de silla para $F(x, \lambda)$, si se utiliza nuevamente el último teorema, resulta que $F^*(x^*) = L(\lambda^*)$ o equivalentemente $f(x^*) = L(\lambda^*)$. Esto combinado con la parte (i) hace que * sea una solución óptima de P^*.

Existen ciertos métodos para transformar un problema con restricciones en un problema sin restricciones. Uno de los métodos más prometedores es la técnica de función de penalización o también - llamada técnica de minimización secuencial sin restricciones.

Consideremos el siguiente problema de programación convexa:

minimizar $f(x)$

sujeto a $g_i(x) \leq 0$ $\quad$ $i = 1,2,\ldots, m$

donde $f(x)$ y $g_i(x)$ son funciones convexas.

Para nuestro propósito, es más conveniente considerar problemas donde las desigualdades estén en sentido inverso, es decir, problemas de la forma:

minimizar $f(x)$

sujeto a $h_i(x) \geq 0$ $\quad$ $i = 1,2,\ldots, m$

donde $h_i(x) = -g_i(x)$. Obsérvese que no se pierde generalidad al con siderar problemas de esta forma.

La técnica de minimización secuencial sin restricciones requiere que se construya la siguiente función:

$$P(x, r_k) = f(x) + r_k \sum_{i=1}^{m} \frac{1}{h_i(x)} \qquad k = 1,2,.... \qquad r_k > 0$$

y que se minimice la función $P(x, r_k)$ para una sucesión decreciente de valores de r_k.

A continuación se demostrará que para un problema de programación convexa con un punto interior factible a las restricciones; los $x(r_k)$ correspondientes a r_k convergen a la solución óptima x^* cuando $k \to \alpha$, si $r_k \to 0$ cuando $k \to \alpha$.

Ya que $f(x)$ y $g_i(x)$ son convexas la función $p(x, r_k)$ es estrictamente convexa para $r_k > 0$. Así el mínimo de $P(x, r_k)$ ocurre en el punto donde

$$\frac{\partial P}{\partial x_j} = 0 \qquad j = 1,2,..., n$$

y además, este punto es único.

Para demostrar que la sucesión de los $x(r_k)$ converge a x^* se usará el argumento dado por Fiacco y McCormick. Seleccionemos un $\varepsilon > 0$ tal que para algún $x^0 \in \mathcal{F}$

$$f(x^0) - f(x^*) < \varepsilon/2$$

ésto es posible, ya que:

i) $f(x) \geq f(x^*)$ para toda $x \in \mathcal{F}$

ii) se asumirá que f es diferenciable y por lo tanto, es contínua.

A continuación, seleccionemos algún k, digamos k^o, tal que

$$r_{k^o} < \min_i \,(1/2\ m) \quad h_i(x^o)\ \varepsilon$$

Por definición de $f(x^*)$, se tiene:

$$f(x^o) < \min_x P(x, r_k) = P\left[x(r_k), r_k\right]$$

la primera desigualdad se cumple porque

$$P(x, r_k) = f(x) + r_k \sum_{i=1}^{m} \frac{1}{h_i(x)}$$

$$y \ r_k \sum_{i=1}^{m} 1/h_i(x) \quad > 0$$

Pero $P\left[x(r_k), r_k\right] < P\left[x(r_{k^o}), r_{k^o}\right]$ porque $x(r_k)$ minimiza a $P\left[x(r_k), r_k\right]$. Con nuestra selección de k^o podemos escribir:

$$f(x^o) \leq P\left[x(r_k), r_k\right] \leq P\left[x(r_k^*), r_k\right] < P\left[x(r_{k^o}), r_{k^o}\right]$$

el último paso se sigue del hecho que $r_k < r_k^*$.

Por lo tanto:

$$P\left[x(r_{k^o}), r_k\right] = f\left[x(r_{k^o})\right] + r_k \sum_{i=1}^{m} \frac{1}{h_i\left[x(r_{k^o})\right]}$$

$$\leq f\left[x(r_{k^o})\right] + r_{k^o} \sum_{i=1}^{m} \frac{1}{h_i\left[x(r_{k^o})\right]}$$

$$= P\left[x(r_{k^o}), r_{k^o}\right]$$

La desigualdad es probada fácilmente por:

i) restando $f\left[x(r_k o)\right]$ de ambos lados y

ii) recordando que $\sum_{i=1}^{m} \frac{1}{h_i\left[x(r_k o)\right]} > 0$

El paso final involucra recordar que:

$$P\left[x(r_k o), r_k o\right] \leq P\left[x^o, r_k o\right] = f(x^o) + r_k^o \sum_{i=1}^{m} \frac{1}{h_i(x^o)}$$

de todas las desigualdades anteriores tenemos:

$$f(x^*) \leq P\left[x(r_k), r_k\right] \leq P\left[x(r_k o), r_k o\right] \leq f(x^*) + r_k^o \sum_{i=1}^{m} \frac{1}{h_i(x^o)}$$

pero, por continuidad y la selección de x^o sabemos que:

i) $f(x^o) < f(x^*) + \varepsilon/2$

ii) $r_k o < \min_j \frac{\varepsilon}{2m} h_j(x^o)$

Usando estos dos resultados en el lado derecho de la desigualdad se tiene:

$$f(x^*) \leq P\left[x(r_k), r_k\right] < f(x^*) + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = f(x^*) + \varepsilon$$

así, cuando $k \to \alpha$, $P\left[x(r_k), r_k\right]$ debe converger a $f(x^*)$

Como un resultado de esta convergencia y el hecho que $r_k \to 0$ cuando

$$P(r) = \frac{\left[(r^{1/2} + 1)^{1/2} + 1\right]^3}{3} + r^{1/2} + \frac{r}{(r^{1/2}+1)^{1/2} - 1} + \frac{r}{r^{1/2}}$$

$$\lim_{r\to 0} P(r) = \frac{2^3}{3} = \frac{8}{3}$$

y los valores de x_1^* y x_2^* son:

$$x_1^* = \lim_{r\to 0} (r^{1/2} + 1)^{1/2} = 1$$

$$x_2^* = \lim_{r\to 0} r^{1/2} = 0$$

Hasta ahora, sólo se ha dicho que $P\left[x(r_k), r_k\right]$ converge a $f(x^*)$ en el límite. Uno puede concebir la gran cantidad de tiempo de computación utilizado para minimizar la sucesión de las funciones P con muy pequeñas mejoras en f(x) durante el curso de un gran número de iteraciones. Sin embargo, usando la teoría de Dualidad, se puede desarrollar una cota inferior a $f(x^*)$, usando esta cota inferior junto con el valor actual de $f\left[x(r_k)\right]$ como una cota superior se pueden obtener límites superior e inferior para $f(x^*)$, si en algún paso del proceso, estas cotas se encuentran lo suficientemente cercanas, el proceso puede ser terminado.

La cota inferior, es desarrollada a continuación:

Dado el siguiente problema primal de programación convexa:

minimizar la función convexa $f(x)$

sujeta a

$$h_i(x) \geq 0 \qquad i = 1,2,\ldots, m$$

donde $h_i(x)$ son funciones cóncavas

Este problema primal tiene su problema dual, el cual será:

$$\text{maximizar } F(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i h_i(x)$$

El problema dual estará sujeto a las siguientes restricciones:

$$\frac{\partial F}{\partial x_j} = 0 \qquad j = 1,2,\ldots, n \qquad y\ \lambda_i \geq 0 \quad i = 1,2,\ldots, m$$

si se define:

$$\lambda_i(r_k) = \frac{r_k}{h_i^2 \quad x(r_k)} \qquad i = 1,2,\ldots, m$$

entonces $\lambda_i(r_k) > 0$

Este punto también satisface las restricciones del problema Dual, porque en $x(r_k)$, se tiene:

$$\left.\frac{\partial P}{\partial x_j}\right|_{x(r_k)} = \frac{\partial f}{\partial x_j} - r_k \sum_{i=1}^{m} \left\{ \left.\frac{\frac{\partial h_i}{\partial x_j}}{h_i^2}\right|_{x(r_k)} \right\} = \left.\frac{\partial F}{\partial x_j}\right|_{x(r_k)}$$

$\partial P/\partial x_j \Big|_{x(r_k)}$ es igual a cero, ya que $x(r_k)$ minimiza a $P(x, r_k)$

Anteriormente, se había probado que para algún k, existe un ε tal que

$$f(x^*) < f\left[x(r_k)\right] < f(x^*) + \varepsilon$$

y además que:

$$0 < r_k \sum_{i=1}^{m} \frac{1}{h_i\left[x(r_k)\right]} < \varepsilon$$

Rearreglando la última expresión resulta:

$$-\varepsilon < -r_k \sum_{i=1}^{m} \frac{1}{h_i\, x(r_k)} < 0$$

Combinando las dos desigualdades resulta:

$$f(x^*) - \varepsilon < f\left[x(r_k)\right] - r_k \sum_{i=1}^{m} \frac{1}{h_i\left[x(r_k)\right]} < f(x^*) + \varepsilon$$

la parte central de esta desigualdad es la función dual, siempre y cuando λ_i se defina apropiadamente, del teorema de dualidad se puede concluir que:

$$f(x^*) - \varepsilon < f\left[x(r_k)\right] - r_k \sum_{i=1}^{m} \frac{1}{h_i\left[x(r_k)\right]} \leq f(x^*) < f(x^*) + \varepsilon$$

luego, una cota inferior de $f(x^*)$ es el valor actual de la función objetivo Dual, la cota superior es el valor actual de f en la iteración k-ésima; si en alguna iteración la diferencia entre la cota superior e inferior, la cual es

$$r_k \sum_{i=1}^{m} \frac{1}{h_i\left[x(r_k)\right]}$$, es suficientemente pequeña, los cálculos pueden ser terminados.

Por último, se hablará acerca del esquema para seleccionar el valor inicial de x para minimizar $P(x, r_k)$; algunos trabajos indican que el procedimiento de extrapolación puede ser usado para estimar $x(r_k)$

Así, dados $x(r_{k-2})$, $x(r_{k-1})$, r_{k-2}, r_{k-1} y el siguiente valor de r, es decir r_k, resulta razonable seleccionar como un valor inicial de $x(r_k)$ el valor obtenido de

$$\hat{x}(r_k) = a + br_k$$

($\hat{x}(r_k)$ indica un estimado de $x(r_k)$), donde a y b son determinados por la información acerca de $x(r_{k-2})$ y $x(r_{k-1})$

En la práctica, un procedimiento ligeramente más general es usar la información de las últimas tres soluciones óptimas $x(r_{k-3})$, $x(r_{k-2})$ y $x(r_{k-1})$ para obtener una ecuación cuadrática para estimar $x(r_k)$.

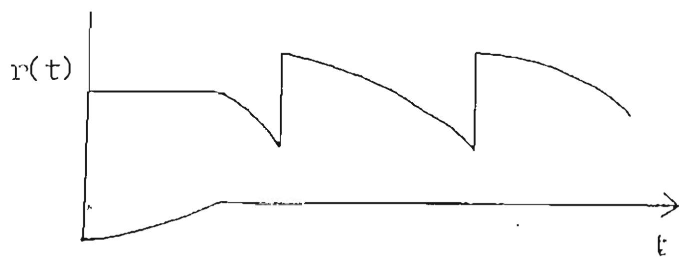
PROGRAMACION DINAMICA

INTRODUCCION:

En esta parte se tratará lo que concierne a la optimización de sistemas en los cuales cada variable puede ser representada como una función de un parámetro (parámetro que en muchos casos es el tiempo). En estos casos, se trata de un problema de optimización dinámico y el problema consiste en determinar la trayectoria óptima.

Los problemas de optimización dinámicos son bastante utilizados en problemas de inventario, problemas de control óptimo, en el diseño de sistemas de procesos de etapas múltiples, etc.

Para ilustrar la diferencia básica entre un sistema dinámico y uno estático consideremos el siguiente ejemplo: supóngase que se tiene el problema de determinar la política de inventario óptima para dos sistemas que difieren únicamente en las propiedades de la demanda del inventario; en el primer sistema, la razón de demanda y, es constante, es decir, no depende del tiempo, mientras que en el segundo caso, la razón de demanda es una función del tiempo, como lo muestra la siguiente figura:



Ya que la razón de demanda en el primer caso es constante, sólo se necesita especificar la política de inventario óptima para un período como nada cambia de período a período la política será óptima para todos los intervalos de tiempo futuros.

El segundo sistema es más complicado, ya que como la razón de demanda varía con el tiempo y, en general, no será la misma de un período a otro. Este sistema de inventario es dinámico porque sus propiedades varían con el tiempo.

Existen tres técnicas para resolver lo que se le ha llamado un problema de optimización dinámico, ellas son: el cálculo de Variaciones, El Principio de Máximo y la Programación Dinámica. De ellas, la última es la más general, un mayor número de problemas pueden - ser resueltos por dicha técnica; además, ésta será la técnica que se desarrollará en esta parte.

Los problemas de optimización dinámicos o también llamados problemas de optimización de etapas múltiples, aparecen en varios contextos; a continuación, se presentan tres ejemplos para ilustrar la necesidad de optimizar un sistema dinámico.

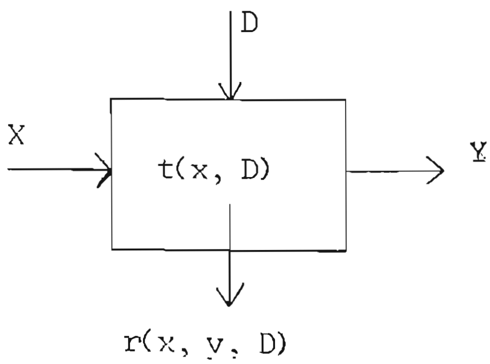
El primer ejemplo que consideraremos es un problema de control óptimo, supóngase que se tiene un cohete en la plataforma de lanzamiento y que el objetivo deseado es hacer que el cohete golpee un blanco móvil en un intervalo de tiempo dado, el blanco tomará acciones evasivas recién se inicie el intento de golpearlo; si se asume que el cohete puede generar cierta velocidad para alcanzar al blanco, entonces el problema será generar un conjunto de comandos para que el cohete esté en disposición de golpear al blanco en el intervalo de tiempo dado; éste es un problema de decisión de múltiples etapas ya que se observará el blanco y de sus acciones se generará, periódicamente, una nueva dirección y velocidad para nuestro cohete.

El segundo ejemplo es el diseño de una sucesión de unidades de proceso que aceptarán una cantidad de material y producirán un producto acabado de calidad aceptable; ya que la salida de una unidad de la sucesión es la entrada de la siguiente, el diseño de una unidad dependerá del diseño de la unidad que le precede; el objetivo deseado es minimizar el costo de la sucesión de unidades de proceso mientras se mantiene una calidad aceptable, las decisiones que pueden ser hechas son las especificaciones de diseño de los componentes individuales del sistema, este es un problema de decisiones de múltiples etapas.

Como ejemplo final, consideremos el problema de precisión, supóngase que se desea maximizar la precisión de un sistema sin exceder una cantidad dada de dinero para su construcción. El sistema es una sucesión de componentes arreglados en una serie de etapas, de tal forma que si una etapa falla, el sistema completo falla; una etapa es -

definida como un grupo de componentes idénticos, arreglados de tal forma que sólo uno está operando y los otros están esperando, en el caso que el componente original falle, uno de los otros componentes entra en operación inmediatamente. La decisión en cada etapa es el número de componentes redundantes que se deben instalar; sin olvidar que, los componentes adicionales en cada etapa mejoran la precisión del sistema, pero también, aumentan el costo.

Todos los procesos de decisión de etapas múltiples tienen ciertas similitudes, considere el proceso de decisión de una etapa representado en la siguiente figura:



Todos los procesos de decisión tienen ciertos parámetros, representados en nuestra caja como entradas, éstos parámetros, x, proporcionan toda la información relevante acerca de las entradas a la caja, y son llamadas variables de estado. El siguiente componente de un proceso de decisión es el conjunto de variables por el vector D y son llamadas variables de decisión. Asociada con cada decisión D y cada variable de estado x está una salida de la caja, la cual es el resultado de la decisión hecha; estas salidas, y, son llamadas las variables de estado de salida y ellas caracterizan completamente la salida de la caja.

Las salidas están relacionadas con las entradas a través de la función de transformación de la etapa, la cual es representada por

$$y = t(x, D)$$

Por último, existe una función objetivo o función de retorno que mide la efectividad de las decisiones que son hechas y las salidas que resultan de estas decisiones. Ya que, las decisiones que se hagan pueden cambiar con cambios en el estado del sistema, la función de retorno está representada por:

$$R = r(x, D, y)$$

como y está completamente especificada por x y D a través de la función de transformación de la etapa, y puede ser eliminada de la función de retorno, resultando:

$$R = r\left[x, D, t(x, D)\right] = r(x, D)$$

En muchos problemas, resulta útil expresar el retorno en términos de las variables de estado de entrada y de salida (x, y), si se puede invertir la función de transformación de la etapa y expresar a D en -

términos de x y y, entonces D puede ser eliminado de la función de retorno, teniendo:

$$D = t_1(x,y)$$

$$R = r\left[x,\ t_1(x,y),y\right] = r(x,y)$$

donde t_1 representa a D en términos de x y y.

Por último, puede ser útil expresar el retorno como una función de las variables de estado de salida en lugar de las de entrada; si la inversa existe.

$$x = t_2(y,\ D)$$

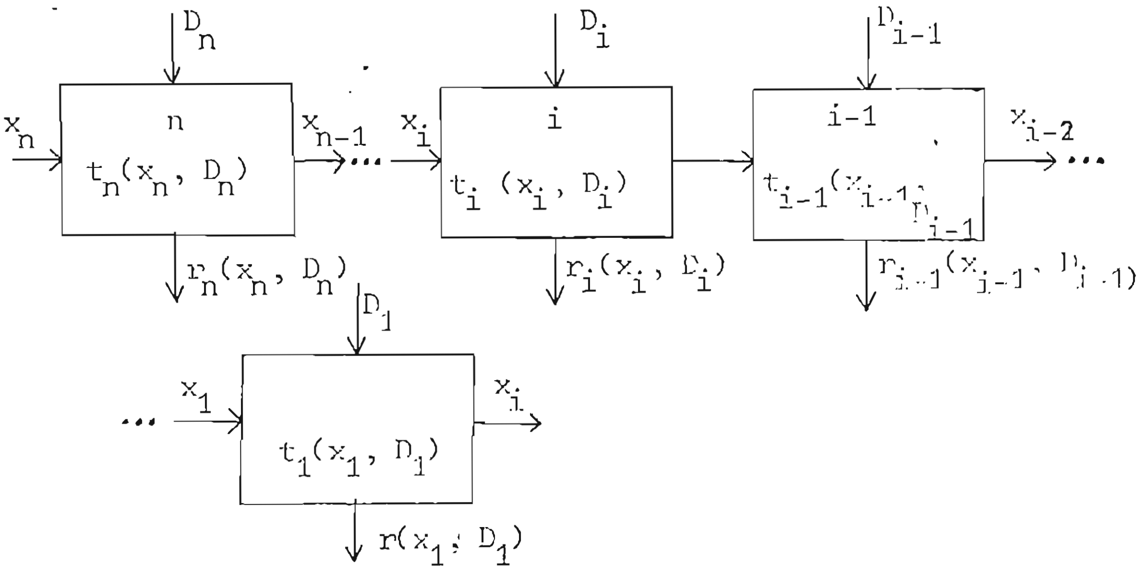
entonces:

$$R = r\left[x,D,y\right] = r\left[t_2(y,D),\ D,\ y\right] = r(y,\ D)$$

donde t_2 representa a x en términos de Y y D.

El teorema de la función implícita, proporciona las condiciones necesarias y suficientes para la existencia de las inversas discutidas anteriormente cuando la función t sea diferenciable.

Un proceso de decisión en serie de etapas múltiples es aquel en el cual un número de procesos de una etapa están conectados en serie, de tal forma que la salida de una etapa es la entrada de la etapa subsiguiente; la figura siguiente representa un proceso serial de decisión de múltiples etapas.



De esta figura, resulta claro que las decisiones de cada etapa no pueden ser hechas independientemente de cualquier otra; como, la entrada de la etapa i, x_i, es una función de la entrada a la etapa i+1, x_{i+1}, y de la decisión D_{i+1}, ya que, en general, la decisión en el estado i depende del estado del sistema x_i, la dependencia de la D_i óptima de D_{i+1} es establecida.

Cada etapa en un proceso de decisión de etapas múltiples tiene una función de retorno asociada con ella; el objetivo de este problema de múltiples etapas es optimizar alguna función de los retornos de cada etapa, es decir, $g(r_1, r_2, \ldots, r_n)$. La composición de las funciones de retorno de las n etapas determina cuando un problema dado puede ser resuelto por programación dinámica.

La programación dinámica es una técnica de descomposición para resolver problemas de decisión de múltiples etapas.

El problema de encontrar un conjunto óptimo de decisiones para un conjunto dado de parámetros de entrada (variables de estado) en un problema de n etapas puede ser resuelto por dos métodos: el primero, involucra la aplicación de la teoría clásica de optimización, cuando sea apropiado, al problema de determinar valores de las variables idependientes (de decisión) que optimice la función de retorno compuesta; el segundo método, la programación dinámica, es una descomposición del problema con n variables de decisión en n problemas con una variable de decisión; en muchos casos, estos n subproblemas son más fáciles de resolver que el problema original, la descomposición es efectuada de manera que, la solución óptima al problema de n variables es obtenida de las soluciones óptimas de los n problemas de una dimensión.

Composición y el Principio de Optimalidad.

La programación dinámica está basada en el principio de optimalidad, el cual dice: "Una política óptima (conjunto de decisiones) tiene la propiedad que sin importar el estado inicial y las decisiones hechas, las restantes decisiones deben constituir una política óptima con respecto al estado resultante de la primera decisión".

Se considerará la implicación de este principio en un problema de decisión de etapas múltiples, supóngase que el objetivo deseado es maximizar la función objetivo de n etapas, la cual es dada por la suma de los retornos de cada etapa, es decir:

$$\max_{Dn,\ldots,D_1} \left[r_n(x_n, D_n) + r_{n-1}(x_{n-1}, D_{n-1}) + \ldots\ldots\ldots + r_1(x_1, D_1) \right]$$

sujeto a

$$x_{j-1} = t_j(x_j, D_j) \qquad j = 1,2,\ldots\ldots, n$$

Si se define

$$f_n(x_n) = \max_{D_n,\ldots D_1} \left[r_n(x_n, D_n) + \ldots + r_1(x_1, D_1) \right] \ldots\ldots \qquad (1)$$

sujeta a:

$$x_{j-1} = t_j(x_j, D_5) \qquad j = 1,2,\ldots, n$$

Se tiene:

$$f_n(x_n) = \left\{ \max_{D_n} \max_{D_{n-1}\ldots D_1} \left[r_n(x_n, D_n) + \ldots + r_1(x_1, D_1) \right] \right\}$$

sujeto a

$$x_{j-1} = t_j(x_j, D_j) \qquad j = 1,2,\ldots, n$$

Pero, $r_n(x_n, D_n)$ no depende de $D_{n-1},\ldots, D_1$. Por lo tanto, la maximización con respecto a estas variables puede ser establecida así:

$$f_n(x_n) = \max_{D_n} \left\{ r_n(x_n, D_n) + \max_{D_{n-1}\ldots D_1} \left[r_{n-1}(x_{n-1}, D_{n-1}) + \ldots + r_1(x_1, D_1) \right] \right\}$$

sujeta a

$$x_{j-1} = t_j(x_j, D_j) \qquad j = 1,2,\ldots, n$$

La ecuación (1) puede ser utilizada para reescribir esto último así:

$$f_n(x_n) = \max_{D_n} \left[r_n(x_n, D_n) + f_{n-1}(x_{n-1}) \right]$$

donde

$$x_{n-1} = t_n(x_n, D_n)$$

Anteriormente, se indicó que el objetivo en programación dinámica es seleccionar un conjunto de decisiones para optimizar alguna función de los retornos de cada etapa; para tratar cierta clase de funciones, Mitten introdujo la noción de composición de retornos de cada etapa, la función de retornos de cada etapa puede ser escrita así:

$$g(r_n, r_{n-1},\ldots,r_1) = r_n \ (+) \ r_{n-1} (+) \ldots\ldots (+) \ r_1$$

donde (+) es llamado el operador composición. La forma del operador composición depende del problema particular que se esté considerando; por ejemplo, en el segundo ejemplo que se dió en la introducción, el objetivo era minimizar el costo total del sistema de procesamiento, donde el costo total puede ser representado como la suma de los costos de cada etapa en el sistema; en este caso, el operador composición será todo +, así:

$g(r_n, r_{n-1}, \ldots, r_1) = r_n + r_{n-1} + \ldots + r_1$

En el tercer ejemplo, se deseaba maximizar la precisión del sistema, en este caso la función objetivo general puede ser escrita de esta forma:

$$g(r_n, r_{n-1}, \ldots, r_1) = r_n \cdot r_{n-1} \cdot \ \ldots \ \cdot r_1$$

resultando que el operador composición para los retornos de cada etapa es el producto.

El operador composición puede variar de etapa en etapa, por ejemplo:

$$g(r_3, r_2, r_1) = r_3 \ (+) \ r_2 \ (+) \ r_1$$

$$= r_3 \cdot r_2 + r_1$$

Como puede observarse, el operador composición entre las etapas 2 y 3 es "." y entre las etapas 2 y 1 es "+".

Desafortunadamente, no todos los problemas de decisión de múltiples etapas pueden ser resueltos por programación dinámica, ya que se requiere que la función objetivo sea separable, es decir, se debe representar la función objetivo como la composición de los retornos de cada etapa, así:

$$g\left[r_n(x_n, D_n), \ldots, r_1(x_1, D_1)\right] = r_n\left[x_n, D_n\right] (+) \ldots (+) r_2\left[x_2, D_2\right] (+) r_1\left[x_1, D_1\right]$$

Resulta fácil construir funciones que no satisfacen este requerimiento, por ejemplo, la siguiente función de retorno no es separable:

$$g\left[r_4(x_4,D_4), r_3(x_3,D_3),\ r_2(x_2,D_2),\ r_1(x_1,D_1)\right]$$

$$=\left[r_4\ (x_4,D_4) + r_2(x_2,D_2)\right] . \left[\ r_3(x_3,D_3) + r_1(x_1,D_1)\right]$$

aún así, existen muchos problemas de aplicación que satisfacen la condición de separabilidad.

Teorema: Una condición suficiente para la descomposición por programación dinámica de un objetivo en una función con retornos separables es que, la función objetivo de n etapas, $g(r_n, r_{n-1}, \dots, r_1)$ sea una función monótona no decreciente de $r_{n-1}, r_{n-2}, \dots, r_1$.

La prueba de este teorema, se encuentra en Mitten L.G. "Composition Principles for Synthesis of Optimal Multistage Processes" Operations Research, Vol. 12 (1964).

Problemas en los cuales el operador composición es "+" satisfacen esta condición; un ejemplo de una función objetivo que no satisface el teorema es la siguiente:

$$g\left[r_3(x_3D_3), r_2(x_2D_2), r_1(x_1,D_1)\right] = r_3(x_3,D_3) .\ r_2(x_2,D_2)\ .\ r_1(x_1,D_1)$$

donde el rango de r_i puede ser positivo y negativo; así cuando -- $r_3\ (x_3,D_3) < 0$ y $r_2(x_2,D_2) > 0$ g es monótona decreciente cuando r_1 aumen

$g(r_n, r_{n-1}, \ldots, r_1) = r_n + r_{n-1} + \ldots + r_1$

En el tercer ejemplo, se deseaba maximizar la precisión del sistema, en este caso la función objetivo general puede ser escrita de esta forma:

$$g(r_n, r_{n-1}, \ldots, r_1) = r_n \cdot r_{n-1} \cdot \ \ldots \ \cdot r_1$$

resultando que el operador composición para los retornos de cada etapa es el producto.

El operador composición puede variar de etapa en etapa, por ejemplo:

$$g(r_3, r_2, r_1) = r_3 \ (+) \ r_2 \ (+) \ r_1$$

$$= r_3 \cdot r_2 + r_1$$

Como puede observarse, el operador composición entre las etapas 2 y 3 es "." y entre las etapas 2 y 1 es "+".

Desafortunadamente, no todos los problemas de decisión de múltiples etapas pueden ser resueltos por programación dinámica, ya que se requiere que la función objetivo sea separable, es decir, se debe representar la función objetivo como la composición de los retornos de cada etapa, así:

$$g\left[r_n(x_n, D_n), \ldots, r_1(x_1, D_1)\right] = r_n\left[x_n, D_n\right] (+) \ldots (+) r_2 \left[x_2, D_2\right] (+) r_1 \left[x_1, D_1\right]$$

Resulta fácil construir funciones que no satisfacen este requerimiento, por ejemplo, la siguiente función de retorno no es separable:

$$g\left[r_4(x_4,D_4), r_3(x_3,D_3),\ r_2(x_2,D_2),\ r_1(x_1,D_1)\right]$$

$$=\left[r_4\ (x_4,D_4) + r_2(x_2,D_2)\right].\left[\ r_3(x_3,D_3) + r_1(x_1,D_1)\right]$$

aún así, existen muchos problemas de aplicación que satisfacen la condición de separabilidad.

Teorema: Una condición suficiente para la descomposición por programación dinámica de un objetivo en una función con retornos separables es que, la función objetivo de n etapas, $g(r_n,\ r_{n-1},\ldots,\ r_1)$ sea una función monótona no decreciente de $r_{n-1}, r_{n-2},\ldots, r_1$.

La prueba de este teorema, se encuentra en Mitten L.G. "Composition Principles for Synthesis of Optimal Multistage Processes" Operations Research, Vol. 12 (1964).

Problemas en los cuales el operador composición es "+" satisfacen esta condición; un ejemplo de una función objetivo que no satisface el teorema es la siguiente:

$$g\left[r_3(x_3D_3), r_2(x_2D_2), r_1(x_1,D_1)\right] = r_3(x_3,D_3)\ .\ r_2(x_2,D_2)\ .\ r_1(x_1,D_1)$$

donde el rango de r_i puede ser positivo y negativo; así cuando -- $r_3\ (x_3,D_3)<0$ y $r_2(x_2,D_2)>0$ g es monótona decreciente cuando r_1 aumen te, y así, en este ejemplo falla el teorema. Sin embargo, esta condición no es necesaria, y es posible descomponer algunos problemas por programación dinámica, aún cuando el teorema anterior falla.

Procesos de Etapas finitas.

Ahora, se centrará la discusión en sistemas con un número finito de etapas, serán consideradas muchos tipos diferentes de ejemplos, problemas donde las variables son discretas, ejemplos con variables continuas y procesos donde algunas de las variables de interés son variables aleatorias; además, se realizarán comparaciones entre los varios tipos de problemas respecto a su forma de solución y se generalizará la estructura de las soluciones.

En las secciones previas, las funciones discutidas eran funciones de un arbitrario, pero finito, número de variables; por ejemplo, la transformación de etapa fue expresada como:

$$x_{n-1} = t_n(x_n, D_n)$$

donde cada variable era un vector de variables.

En esta sección, las dificultades de obtener soluciones para problemas donde el vector de variables de estado tiene un gran número de componentes será evidente, más adelante se discutirán las formas para ayudarnos a esquivar esta dificultad de cálculo.

En la discusión de la programación dinámica, se obviarán, para la mayor parte, discusiones de la forma en que se obtuvieron la formulación de los problemas y se concentrará en obtener la solución a los problemas planteados.

Consideremos el primer ejemplo:

Supóngase que se desea

$$\max_{d_i > 0} \sum_{i=1}^{3} d_i^2$$

sujeto a

$$x_{i-1} = x_i - d_i \qquad i = 1,2,3$$

$$x_i = 0, 1, 2, \ldots, 5 \qquad i = 0, 1, 2$$

$$x_3 = 5$$

usando los resultados de la sección anterior, se tiene:

$$f_1(x_1) = \max d_1^2$$

Investigando los límites de d_i se encuentra que d_i está acotado por abajo por 0; también, nótese que como cada x_{i-1} es requerida que sea no negativa, d_i está acotado por arriba por x_i, finalmente, ya que x_i es requerida que sea entera, los únicos valores de d_i que satisfacen este requerimiento son los enteros no negativos.

$$d_i = 0, 1, 2, \ldots, x_i$$

así

$$f_1(x_1) = \max d_i^2 \qquad d_i = 0, 1, 2, \ldots, x_i$$

Claramente, el máximo es obtenido cuando $d_1 = x_1$; por lo tanto

$f_1(x_1) = x_1^2$

Ahora, se procede a la segunda etapa:

$$f_2(x_2) = \max \left[d_2^2 + f_1(x_1) \right] \qquad d_2 = 0, 1, \ldots, x_2$$

$$= \max \left[d_2^2 + x_1^2 \right] \qquad d_2 = 0, 1, \ldots, x_2$$

pero de la función de transformación de la etapa, se tiene que:

$x_1 = x_2 - d_2$, de donde,,se sigue que:

$$f_2(x_2) = \max \left[d_2^2 + (x_2 - d_2)^2 \right] \qquad d_2 = 0, 1, \ldots, x_2$$

El valor óptimo de d_2 no es tan obvio como lo fue en la etapa anterior; obsérvese que los posibles valores de x_2 son 0, 1,2,3,4, y 5, el valor óptimo de d_2 debe ser determinado para cada uno de los valores de x_2, resultando conveniente escribirlo en forma tabular como se ilustra en la tabla 01 del Anexo 01, $f_2(x_2)$ es tabulado a lo largo de la tabla con los valores óptimos de d_2, d_2^*, para cada x_2.

(Ver Tabla 01 del Anexo 01).

Obsérvese que existe un óptimo alternativo para cada x_2; es decir, cuando x_2 es 4, d_2^* puede ser 0 ó 4. Ahora que se tiene $f_2(x_2)$ determinado, se continúa a $f_3(x_3)$ (Tabla 02) lo cual es dado por;

$$f_3(x_3) = \max_{d_3=0,1,..,x_3} \left[d_3^2 + f_2(x_2) \right]$$

sujeto a $x_2 = x_3 - d_3$

$$= \max_{d_3=0,1,..,x_3} \left[d_3^2 + f_2 (x_3 - d_3) \right]$$

(Ver Tabla 02 del Anexo 01)

Sin embargo, se tiene sólo un valor para x_3, $x_3 = 5$, así:

$$f(x_3=5) = \max_{d_3=0,1,2,...,5} \left[d_3^2 + (x_3 - d_3)^2 \right]$$

Se ha obtenido la solución óptima de un problema de tres variables resolviendo tres problemas de una variable, ésto es descomposición por programación dinámica, la solución a los tres problemas de una variable fue obtenida con mucho menos esfuerzo que si se hubiera resuelto el problema en tres variables, el ahorro puede ser aún más grande si el número de etapas y/o el número de posibles valores de x_1 se aumentan.

Obsérvese que el problema anterior es similar al siguiente:

$$\max \sum_{i=1}^{3} d_i^2$$

sujeta a:

$$\sum_{i=1}^{3} d_i = 5 \qquad d_i = 0,\ 1,2,3,4,5$$

ya que si:

$$x_{i-1} = x_i - d_i$$

$$d_i = x_i - x_{i-1}$$

entonces:

$$\sum_{i=1}^{3} d_i = \sum_{i=1}^{3} (x_i - x_{i-1}) = x_3 - x_o$$

ésto implica que $x_o = 0$ porque $x_3 = 5 = \sum_{i=1}^{3} d_i$, se observa que los dos problemas son idénticos

El siguiente ejemplo que se considerará es un problema con dos variables de estado y una variable de decisión

$$\max_{d_i = 0,1,\ldots} \sum_{i=1}^{3} c_i d_i$$

sujeto a

$$\sum_{i=1}^{3} k_i d_i \leq 10$$

$$\sum_{i=1}^{3} L_i d_i \leq 15$$

donde los parámetros están dados por la siguiente tabla:

i	c_i	k_i	L_i
1	1	2	3
2	3	3	5
3	5	4	7

Se deben obtener las funciones de transformación de etapa para este problema, en general, cuando las restricciones son lineales, la función de transformación de etapa será de la forma

$$x_{i-1} = x_i - a_i d_i$$

A continuación, se muestra la forma de obtener esta última función considere un problema con la siguiente restricción en las variables de decisión

$$\sum_{i=1}^{n} a_i d_i \leq b$$

Definiendo x_i como la cantidad de b sin usar en las (n-1) etapas precedentes, así:

$$x_i = b - \sum_{j=i+1}^{n} a_j d_j$$

entonces:

$$x_{i-1} = b - \sum_{j=1}^{n} a_j d_j$$

restando estas últimas ecuaciones, resulta:

$$x_i - x_{i-1} = a_i d_i$$

Reescribiendo, se obtiene:

$$x_{i-1} = x_i - a_i d_i$$

Se ha obtenido una función de transformación de etapa general, cuando las decisiones están limitadas por restricciones lineales.

Para el ejemplo en consideración, se tiene:

$$x_{i-1} = x_i - k_i d_i$$

$$y_{i-1} = y_i - L_i d_i$$

Recordemos que, por definición, se tiene:

$$f_1(x_1, y_1) = \max_{d_1} \quad c_1 d_1$$

donde

$$d_1 = 0, 1, \ldots \ldots \leq \min \left(\frac{x_1}{k_1}, \frac{y_1}{L_1} \right)$$

los resultados se muestran en la tabla 03

(Ver tabla 03 del Anexo 01)

Ahora, se tiene $f_1(x_1, y_1)$ para todas las combinaciones posibles de x_1 y y_1; los cálculos para $f_2(x_2, y_2)$ se muestran en la Tabla 04.

(Ver Tabla 04 del Anexo 01)

Finalmente, se obtiene $f_3(x_3, y_3)$, es claro que $\sum c_i d_i$ será - máxima cuando $x_3 = 10$ y $y_3 = 15$, por lo tanto, sólo se considerarán estos dos valores para las variables de estado de entrada; - el resultado se muestra en la Tabla 05.

(Ver Tabla 05 del Anexo 01)

El ejemplo anterior, sirve para ilustrar las dificultades encontradas cuando el número de variables de estado es aumentado, el aumento en el número de cálculos en el ejemplo dado fue solamente fastidioso; sin embargo, cuando el número de combinaciones de las variables de estado que deben ser consideradas exceden las 100 000, la cantidad de memoria requerida para almacenarlos alcanza los límites de la mayoría de las computadoras corrientes; el tiempo de - computación también aumenta rapidamente cuando se aumentan las variables de estado.

Ahora, se considerará el caso cuando las funciones son diferenciables; supóngase que en el primer ejemplo las variables son contínuas; específicamente:

$$\max_{d_i \geq 0} \sum_{i=1}^{3} d_i^2$$

sujeta a

$$x_{i-1} = x_i - d_i \qquad i = 1,2,...$$

$$x_3 = 5$$

$$x_i \geqslant 0 \qquad i = 0, 1, 2$$

así:

$$f_1(x_1) = \max d_1^2 \qquad 0 \leqslant d_1 \leqslant x_1$$

empleando las condiciones de Kuhn-Tucker, se obtiene:

$$F(d_1, \lambda_1, \lambda_2, U_1, U_2) = d_1^2 + \lambda_1(d_1 - x_1 + U_1^2) + \lambda_2(-d_1 + U_2^2)$$

$$\frac{\partial F}{\partial d_1} = 2d_1 + \lambda_1 - \lambda_2 = 0$$

$$\frac{\partial F}{\partial \lambda_1} = d_1 - x_1 + U_1^2 = 0$$

$$\frac{\partial F}{\partial \lambda_2} = -d_1 + U_2^2 = 0$$

$$\frac{\partial F}{\partial U_1} = 2\lambda_1 U_1 = 0$$

$$\frac{\partial F}{\partial U_2} = 2\lambda_2 U_2 = 0$$

$$\lambda_1, \lambda_2 < 0$$

Al asumir que λ_1 y U_2 son diferentes de cero, ésto implica que $U_1 = \lambda_2 = 0$, así:

$$d_1 = -\frac{\lambda_1}{2} = x_1$$

Ya que $x_1 \geqslant 0$, $\lambda_1 \leqslant 0$ y las condiciones necesarias son satisfechas.

La otra solución ocurre cuando $\lambda_1 = U_2 = d_1 = 0$, este no es un máximo porque $\lambda_2 \geqslant 0$.

Por lo tanto, $f_1(x_1) = x_1^2$

Ahora, determinaremos a $f_2(x_2)$:

$$f_2(x_2) = \max_{0 \leq d_2 \leq x_2} \left[d_2^2 + f_1(x_1) \right]$$

sujeto a $x_1 = x_2 - d_2$

$$f_2(x_2) = \max_{0 \leq d_2 \leq x_2} \left[d_2^2 + (x_2 - d_2)^2 \right]$$

resolviendo este problema, se obtiene:

$$f_2(x_2) = x_2^2$$

$$d_2^* = x_2$$

Obsérvese que en este caso, existe una solución óptima alternativa con $d_2^* = 0$ y $f_2(x_2) = x_2^2$. Finalmente

$$f_3(x_3 = 5) = \max_{0 \leq d_3 \leq x_3} \left[d_3^2 + f_2(x_2) \right]$$

sujeto a $x_2 = x_3 - d_3$

$$f_3(x_3) = \max_{0 \leq d_3 \leq x_3} \left[d_3^2 + (x_3 - d_3)^2 \right]$$

En la solución, nuevamente resulta un óptimo alternativo $d_3^* = x_3$ ó

$d_3^* = 0$ y $f_3(x_3) = x_3^2 = 25$; la regla de decisión será:

$$d_3 = 5 \qquad d_2 = 0 \qquad d_1 = 0$$

ó

$$d_3 = 0 \qquad d_2 = 5 \qquad d_1 = 0$$

y finalmente:

$$d_3 = 0 \qquad d_2 = 0 \qquad d_1 = 5$$

Los resultados de este ejemplo, pueden ser ampliados a un número arbitrario, pero finito, de etapas; el problema será:

$$\max_{d_i \geq 0} \sum_{i=1}^{n} d_i^2$$

$$x_{i-1} = x_i - d_i \qquad i = 1,2,\ldots, n$$

$$x_i \geq 0 \qquad i = 1,2,\ldots, n$$

$$x_n = c$$

Se tiene mostrado que: $f_1(x_1) = x_1^2$, $f_2(x_2) = x_2^2$ y $f_3(x_3) = x_3^2$; usando la prueba por inducción, se asumirá que si:

$$f_{n-1}(x_{n-1}) = x_{n-1}^2$$

entonces:

$$f_n(x_n) = \max_{0 \leq d_n \leq x_n} \left[d_n^2 + f_{n-1}(x_{n-1}) \right]$$

sujeta a

$$x_{n-1} = x_n - d_n$$

incorporando la función de transformación de etapa, resulta:

$$f_n(x_n) = \max_{0 \leq d_n \leq x_n} \left[d_n^2 + (x_n - d_n)^2 \right] = \max_{0 \leq d_n \leq x_n} q_n(x_n, d_n)$$

Las condiciones de Kuhn-Tucker son:

$$F(x_n, \lambda_1, \lambda_2, U_1, U_2) = q_n(x_n, d_n) + \lambda_1(d_n - x_n - U_1^2) + \lambda_2(-d_n + U_2^2)$$

$$\frac{\partial F}{\partial d_n} = 2d_n - 2(x_n - d_n) + \lambda_1 - \lambda_2 = 0$$

$$\frac{\partial F}{\partial \lambda_1} = d_n - x_n + U_1^2 = 0$$

$$\frac{\partial F}{\partial \lambda_2} = -d_n + U_2^2 = 0$$

$$\frac{\partial F}{\partial U_1} = 2\lambda_1 U_1 = 0$$

$$\frac{\partial F}{\partial U_2} = 2\lambda_2 U_2 = 0$$

$$\lambda_1, \lambda_2 < 0$$

Existen tres soluciones factibles: $d_n = x_n$, $d_n = 0$ y $d_n = x_n/2$; evaluando estas posibilidades se obtiene que $d_n = 0$ ó $d_n = x_n$ y $f_n(x_n) = x_n^2$ así:

$$f_n(x_n) = c^2$$

$$d_i = c\,\delta_{ik} \qquad i = 1, \ldots, n$$

donde

$$\delta_{ik} = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases}$$

este resultado se cumple para cualquier $k = 1,2,\ldots, n$.

Pongamos nuestra atención en otro problema con número arbitrario, pero finito, de etapas. supóngase que se desea:

$$\min_{d_j \geq 0} \sum_{j=1}^{n} d_j^2$$

sujeto a

$$\sum_{j=1}^{n} jd_j = c$$

Anteriormente, se había establecido que cuando las restricciones fueran lineales las funciones de transformación de estado tienen la forma

$$x_{j-1} = x_j - a_j d_j$$

Para el presente ejemplo, se tiene que $a_j = j$ y por lo tanto

$$x_{j-1} = x_j - jd_j$$

siguiendo el método de descomposición por programación dinámica, se tiene:

$$f_1(x_1) = \min_{d_1 = x_1} d_1^2$$

ya que, recordemos que x_1 puede ser considerado como la cantidad de c que no fue utilizada en las etapas $2,3,\ldots, n$ como se tiene una - restricción de igualdad, se debe de requerir que la cantidad de c sin usar en las etapas previas sea utilizada en la etapa 1; por lo tanto, $d_1 = x_1$ y se está en libertad de seleccionar solamente n-1 de los d_n independientemente; obsérvese que este resultado puede ser generalizado a restricciones de igualdad arbitrarias, es decir, siempre que se encuentre una restricción de igualdad, se pierde un grado de libertad en la selección de d,

luego:

$$f_1(x_1) = x_1^2$$

calculando el retorno óptimo de la etapa 2 para una entrada x_2, se tiene:

$$f_2(x_2) = \min_{0 \leq d_2 \leq x_2/2} \left[d_2^2 + f_1(x_1) \right]$$

sujeto a

$$x_1 = x_2 - 2d_2, \text{ luego:}$$

$$f_2(x_2) = \min_{0 \leq d_2 \leq x_2/2} \left[d_2^2 + x_1^2 \right]$$

$$= \min_{0 \leq d_2 \leq x_2/2} \left[d_2^2 + (x_2 - 2d_2)^2 \right]$$

$$= \min_{0 \leq d_2 \leq x_2/2} g_2(x_2, d_2)$$

formando la función de Lagrange, se obtiene:

$$F_2(d_2, \lambda_1, \lambda_2) = d_2^2 + (x_2 - 2d_2)^2 + \lambda_1(d_2 - x_2/2 + U_1^2) + \lambda_2(-d_2 + U_2^2)$$

Las condiciones de Kuhn-Tucker son:

$$\frac{\partial F}{\partial d_2} = 2d_2 - 4(x_2 - 2d_2) + \lambda_1 - \lambda_2 = 0$$

$$\lambda_1 (d_2 - x_2/2) = 0$$

$$\lambda_2 d_2 = 0$$

$$\lambda_1, \lambda_2 > 0$$

La solución a estas ecuaciones es:

$$d_2 = \frac{2x_2}{5} \qquad \lambda_1 = \lambda_2 = 0$$

esta es la solución óptima porque $q_2(x_2, d_2)$ es estrictamente convexa en d_2 y las restricciones son convexas.

Sustituyendo este resultado para d_2 en la definición de $f_2(x_2)$, resulta:

$$f_2(x_2) = \frac{x_2^2}{5}$$

Un modelo parecido se desarrolló cuando las funciones $f_1(x_1)$ y $f_2(x_2)$ fueron funciones lineales de x_1^2 y x_2^2

respectivamente, sin embargo, para estar seguros que este es el modelo correcto, se determinará $f_3(x_3)$:

$$f_3(x_3) = \min_{0 \leq d_3 \leq x_3/3} \left[d_3^2 + f_2(x_2) \right]$$

sujeto a

$$x_2 = x_3 - 3d_3 \text{ por lo tanto}$$

$$f_3(x_3) = \min_{0 \leq d_3 \leq x_3/3} \left[d_3^3 + 1/5(x_3 - 3d_3)^2 \right]$$

Nuevamente, se tiene una función convexa definida en un conjunto convexo de tal forma que existe una solución única para las condiciones necesarias, y además, estas condiciones son suficientes para un mínimo; las condiciones de Kuhn-Tucker son:

$$2d_3 - 6/5\ (x_3 - 3d_3) + \lambda_1 - \lambda_2 = 0$$

$$\lambda_1(d_3 - x_3/3) = 0$$

$$\lambda_2 d_3 = 0$$

$$\lambda_1, \lambda_2 > 0$$

la solución es $d_3^* = 3x_3/14 \quad \lambda_1 = 0,\ \lambda_2 = 0$, usando este resultado para calcular $f_3(x_3)$, se obtiene:

$$f_3(x_3) = \left[(3/14)^2 x_3^2 + (1/5)(5/14)^2 x_3^2\right] = x_3^2/14$$

aparentemente, se ha establecido un modelo; ahora se probará el modelo intuitivo, supóngase que:

$$f_{n-1}(x_{n-1}) = k_{n-1}x_{n-1}^2$$

entonces, si resulta que $f_n(x_n) = k_n x_n^2$ el razonamiento habrá sido justificado

recordemos que:

$$f_n(x_n) = \min_{0 \leq d_n \leq x_n/n} \left[d_n^2 + f_{n-1}(x_{n-1}) \right]$$

$$x_{n-1} = x_n - nd_n$$

usando la supuesta forma de f_{n-1} y usando la función de transformación de la etapa para eliminar a x_{n-1} se obtiene:

$$f_n(x_n) = \min_{0 \leq d_n \leq x_n/n} \left[d_n^2 + k_{n-1}(x_n - nd_n)^2 \right]$$

nuevamente resulta que la expresión entre corchetes es estrictamente convexa y las restricciones son convexas; por lo tanto, la solución a las siguientes ecuaciones resulta ser el valor deseado de d_n.

$$2d_n - 2_n k_{n-1}(x_n - nd_n) + \lambda_1 - \lambda_2 = 0$$

$$\lambda_1 (d_n - x_n/n) = 0$$

$$\lambda_2 d_n = 0$$

$$\lambda_1, \lambda_2 \geq 0$$

la solución es

$$d_n^* = \frac{n\, k_{n-1}\, x_n}{1 + n^2 k_{n-1}} \qquad \lambda_1 = \lambda_2 = 0$$

evaluando $f_n(x_n)$, se obtiene:

$$f_n(x_n) = \frac{n^2 k_{n-1}^2\, x_n^2}{[1 + n^2 k_{n-1}]^2} + \frac{k_{n-1}\, x_n^2}{[1 + n^2 k_{n-1}]^2}$$

$$= \frac{k_{n-1}\, x_n^2}{1 + n^2 k_{n-1}}$$

$$= k_n x_n^2$$

donde:

$$k_n = \frac{k_{n-1}}{1 + n^2 k_{n-1}}$$

Luego, nuestra hipótesis fue correcta y para cualquier etapa p se tiene:

$$f_p(x_p) = k_p x_p^2$$

$$k_p = \frac{k_{p-1}}{1+p^2 k_{p-1}}$$

$$k_1 = 1$$

la decisión óptima en cualquier etapa p, p=1,2,,..., n está dada por:

$$d_p = pk_p x_p$$

con $x_n = c$

Variables aleatorias: el resultado final de esta sección será la extensión de la programación dinámica a ciertos sistemas donde las variables son aleatorias.

Considérese el siguiente problema, el cual tiene su origen la teoría de control óptimo:

$$\min \sum_{i=0}^{n-1} (x_i^2 + d_{i+1}^2)$$

sujeto a $x_{i-1} = x_i - d_i + u_i$, donde u_i es una variable aleatoria de la cual, se conoce su función de densidad de probabilidad que viene dada por:

$$g_i(\mu_i, \sigma_i^2) \qquad -\alpha < \mu_i, \sigma_i^2 < +\alpha$$

el primer resultado será obtenido para el caso cuando las variables aleatorias U_i y U_j, $i \neq j$, sean independientes.

Puede parecer que toda la información necesaria para resolver el problema está a la mano, sin embargo, al inspeccionar más cercanamente se encuentra que las variables de interés son aleatorias, por lo tanto, realmente sólo se tiene un control limitado sobre la realización de x_i; en tales circunstancias, se usará el valor esperado de la función objetivo como una medida de efectividad.

Bajo el criterio del valor esperado, el problema sería:

$$\min_{d_i \; U_i} E\left(\sum_{i=0}^{n-1} x_i^2 + d_{i+1}^2 \right)$$

sujeto a:

$$x_{i-1} = x_i - d_i + U_i$$

Calculando $f_1(x_1)$ resulta:

$$\overline{f}_1(x_1) = \min_{d_1} (x_o^2 + d_1^2)$$

$$= \min_{d_1} \; E_{U_1} \left[(x_1 - d_1 + U_1)^2 + d_1^2 \right]$$

$$= \min_{d_1} \; E_{U_1} \left[g_1(x_1, U_1, d_1) \right]$$

$$\frac{\partial E\left[q_1(x_1,\ U_1,\ d_1)\right]}{\partial\ d_1} = E\left[-2(x_1-d_1+U_1)\right] + 2d_1 = 0$$

de donde:

$$d_1 = \frac{x_1 + \mu_1}{2}$$

Usando esteresultado y la definición de $f_1(x_1)$, resulta:

$$f_1(x_1) = E\left[x_1 + U_1 - \frac{x_1 + \mu_1}{2}\right]^2 + 1/4\ (x_1 + \mu_1)^2$$

$$= \frac{x_1^2 + \mu_1^2}{2} + \sigma_1^2 + x_1\ \mu_1$$

$$\overline{f}_2(x_2) = \min_{d_2}\ \underset{U_2}{E}\ \left[x_1^2 + d_2^2 + \overline{f}_1(x_1)\right]$$

$$= \min_{d_2}\ \underset{U_2}{E}\ \left[(x_2-d_2+U_2)^2 + d_2^2 + \frac{(x_2 - d_2 + U_2)^2}{2} + \frac{\mu_1^2}{2} + \sigma_1^2\right.$$

$$\left. + \mu_1\ (x_2 - d_2 + U_2)\right]$$

$$= \min_{d_2}\ \underset{U_2}{E}\ \left[3/2(x_2-d_2+U_2)^2 + d_2^2 + \frac{\mu_1^2}{2} + \sigma_1^2 + \mu_1(x_2-d_2+U_2)\right]$$

$$= \min_{d_2}\ \underset{U_2}{E}\ \left[q_2(x_2,\ U_2,\ d_2)\right]$$

$$\frac{\partial E\quad q_2}{\partial d_2} = -3(x_2 - d_2 + U_2) + 2d_2 - \mu_1 = 0$$

resolviendo, resulta:

$$d_2 = \frac{3x_2 + 3\mu_2 + \mu_1}{5}$$

luego:

$$\bar{f}_2(x_2) = 0.6x_2^2 + 1.1\,\sigma_2^2 + 0.6\,\mu_2^2 + 0.6\,\mu_1^2 + 1.2\,\mu_2 x_2 + \sigma_1^2 + 0.4x_2\,\mu_1$$

supóngase que la forma general de la solución viene dada por:

$$\bar{f}_{n-1}(x_{n-1}) = a_{n-1}x_{n-1}^2 + \sum_{i=1}^{n-1} b_{n-1}(i)\sigma_i^2 + \sum_{i=1}^{n-1} c_{n-1}(i)\,\mu_i^2 + \sum_{i=1}^{n-1} \mu_i x_{n-1} h_{n-1}(i)$$

entonces:

$$\bar{f}_n(x_n) = \min_{d_n}\ \underset{U_n}{E}\left[x_{n-1}^2 + d_n^2 + \bar{f}_{n-1}(x_{n-1})\right]$$

sujeto a

$$x_{n-1} = x_n - d_n + U_n$$

de donde:

$$\bar{f}_n(x_n) = \min_{d_n}\ \underset{U_n}{E}\left[(x_n - d_n + U_n)^2 + d_n^2 + a_{n-1}(x_n - d_n + U_n)^2 \right.$$

$$+ \sum_{i=1}^{n-1} b_{n-1}(i)\,\sigma_i^2 + \sum_{i=1}^{n-1} c_{n-1}(i)\,\mu_i^2$$

$$\left. + \sum_{i=1}^{n-1} h_{n-1}(i)\,\mu_i(x_n - d_n + U_n)\right]$$

$$\bar{f}_n(x_n) = \min_{d_n}\ \underset{U_n}{E}\left[q_n(x_n, d_n, U_n)\right]$$

$$\frac{\partial E\ q_n}{\partial d_n} = \underset{U_n}{E}\left[-2(x_n - d_n + U_n)(1 + a_{n-1}) + 2d_n - \sum_{i=1}^{n} h_{n-1}(i)\,\mu_i\right]$$

$$d_n = \frac{(1+a_{n-1})(x_n + \mu_n)}{2 + a_{n-1}} + \sum_{i=1}^{n-1} \frac{h_{n-1}(i)\ \mu_i}{2(2+ a_{n-1})}$$

sustituyendo este valor de d_n en σ_n, resulta:

$$\bar{f}_n(x_n) = a_n x_n + \sum_{i=1}^{n} b_n(i)\ \sigma_i^2 + \sum_{i=1}^{n} c_n(i) \mu_i + x_n \sum_{i=1}^{n} h_n(i)\ \mu_i$$

Los coeficientes de los términos de la función objetivo vienen dados por ecuaciones en diferencia no lineales, por ejemplo:

$$a_n = \frac{1 + a_{n-1}}{2 + a_{n-1}}$$

como se ha visto, la introducción de aleatoriedad en este problema particular, solamente causa dificultad algebraica; no se han introducido obstáculos conceptuales; sin embargo, esto solamente es cierto cuando las variables aleatorias no tienen correlación; cuando ellas son correlacionadas, ocurren complicaciones adicionales; de hecho, cuando las variables aleatorias son correlacionadas, se debe - de introducir una variable de estado adicional por cada grado de dependencia de la densidad de $F(U_i)$ con U_{i-1}, U_{i-2},.... .-

El razonamiento anterior se sigue así: en la etapa 1, se tiene la expectación con respecto a U_1. Sin embargo, si U_1 está correlacionada con U_2, la función de densidad de probabilidad de U_1 depende de la observación de U_2; consecuentemente, se debe determinar E( )

para todos los valores de U_2, porque en esta etapa U_2 es desconocido, luego U_2 llega a ser una variable de estado. Similarmente, si la - función de densidad de probabilidad de U_1 depende de $U_2, \ldots, U_i$; cada una de estas variables debe ser llevada como una variable de estado a la etapa 1, ya que E_{U_1} depende de la observación actual de estas variables aleatorias. Cuando la solución puede ser obtenida analíticamente, el álgebra llega a ser incómoda pero aún así es tratable; cuando las soluciones son obtenidas computacionalmente, la introducción de variables aleatorias que están correlacionadas de etapa en etapa origina dudas acerca de la factibilidad computacional - para obtener la solución.

Otro factor que llega a ser importante en la obtención de la regla de decisión cuando existe correlación entre las variables aleatorias en etapas sucesivas es: la regla de decisión es una variable aleatoria, porque en cada etapa, la decisión óptima es una función de las condiciones mencionadas. Luego, hasta que las observaciones de la variable aleatoria están disponibles, lo dicho no puede ser calculado - y la regla de decisión no puede ser implementada. El procedimiento - para obtener la regla de decisión en la etapa i es como sigue:

(1) obtener las observaciones actuales de las variables aleatorias en las etapas $i+1, \ldots, n$;

(2) calcular el valor actual de U_i, basado en estas observaciones, y

(3) calcule la regla de decisión.

<u>Procesos de Infinitas etapas:</u>

Los problemas de infinitas etapas pueden darse en dos formas diferentes; un ejemplo donde pueden darse problemas de decisión de infinitas etapas es en los problemas de planificación; en tales problemas, las decisiones son hechas en intervalos discretos de tiempo; pero la planificación, por razones obvias, puede no necesitar colocarle un límite de tiempo a su efectividad; grandes firmas industriales operan bajo el supuesto que ellas continuarán funcionando indefinidamente y ellas planifican sobre esta base llamada horizonte infinito de planificación.

La otra forma es que se pueden obtener un número infinito de etapas, puede ser ilustrada por el programa para controlar un misil; el programa termina después de un intervalo de tiempo fijo, pero el tiempo entre las etapas, en el sentido de programación dinámica, es pequeño. En efecto, las decisiones son hechas continuamente; luego, un número infinito de decisiones son hechas en un intervalo finito de tiempo; ya que, se definió una etapa como el punto donde se hacen las decisiones, es obtenido un problema de infinitas etapas.

Los problemas surgen con los problemas de infinitas etapas a los cuales no se les puede encontrar una formulación de finitas etapas. Por ejemplo, en uno de localización de recursos de finitas etapas, la -

para todos los valores de U_2, porque en esta etapa U_2 es desconocido, luego U_2 llega a ser una variable de estado. Similarmente, si la - función de densidad de probabilidad de U_1 depende de $U_2, \ldots, U_i$; cada una de estas variables debe ser llevada como una variable de estado a la etapa 1, ya que E_{U_1} depende de la observación actual de estas variables aleatorias. Cuando la solución puede ser obtenida analíticamente, el álgebra llega a ser incómoda pero aún así es tratable; cuando las soluciones son obtenidas computacionalmente, la introducción de variables aleatorias que están correlacionadas de etapa en etapa origina dudas acerca de la factibilidad computacional - para obtener la solución.

Otro factor que llega a ser importante en la obtención de la regla de decisión cuando existe correlación entre las variables aleatorias en etapas sucesivas es: la regla de decisión es una variable aleatoria, porque en cada etapa, la decisión óptima es una función de las condiciones mencionadas. Luego, hasta que las observaciones de la variable aleatoria están disponibles, lo dicho no puede ser calculado - y la regla de decisión no puede ser implementada. El procedimiento - para obtener la regla de decisión en la etapa i es como sigue:

(1) obtener las observaciones actuales de las variables aleatorias en las etapas i+1,...,n;

(2) calcular el valor actual de U_i, basado en estas observaciones, y

(3) calcule la regla de decisión.

Procesos de Infinitas etapas:

Los problemas de infinitas etapas pueden darse en dos formas diferentes; un ejemplo donde pueden darse problemas de decisión de infinitas etapas es en los problemas de planificación; en tales problemas, las decisiones son hechas en intervalos discretos de tiempo; pero la planificación, por razones obvias, puede no necesitar colocarle un límite de tiempo a su efectividad; grandes firmas industriales operan bajo el supuesto que ellas continuarán funcionando indefinidamente y ellas planifican sobre esta base llamada horizonte infinito de planificación.

La otra forma es que se pueden obtener un número infinito de etapas, puede ser ilustrada por el programa para controlar un misil; el programa termina después de un intervalo de tiempo fijo, pero el tiempo entre las etapas, en el sentido de programación dinámica, es pequeño. En efecto, las decisiones son hechas continuamente; luego, un número infinito de decisiones son hechas en un intervalo finito de tiempo; ya que, se definió una etapa como el punto donde se hacen las decisiones, es obtenido un problema de infinitas etapas.

Los problemas surgen con los problemas de infinitas etapas a los cuales no se les puede encontrar una formulación de finitas etapas. Por ejemplo, en uno de localización de recursos de finitas etapas, la -

función de retorno está definida como la suma de los retornos individuales de cada etapa, al intentar extender este tipo de medida de efectividad a un problema de infinitas etapas, la medida de efectividad sería la suma de infinitos términos; para que exista una solución óptima, ésta suma debe de converger.

Una condición necesaria, pero no suficiente para la convergencia es que

$$\lim_{n \to \alpha} r_n(x_n, d_n) = 0$$

Sin embargo, no se puede estar satisfecho con una función de retorno que disminuya de etapa en etapa. Más pragmáticamente, un investigador de operaciones que prescribe una regla de decisión para su uso, se puede encontrar en dificultades al intentar explicar el porqué las ganancias deben ser decrecientes.

Varios esquemas han sido desarrollados para llegar a una solución satisfactoria a este dilema. Un método involucra promediar el tiempo, la función de retorno g, es dividida entre el número de etapas, N. Luego para obtener la solución óptima se resuelve el problema de infinitas etapas por un proceso limitado, es decir:

$$\lim_{N \to \alpha} \max g/N = \lim_{N \to \alpha} \max 1/N \sum_{n=1}^{N} r_n(x_n, d_n)$$

Otro método, involucra descontar los retornos futuros para obtener el valor presente; el factor de descuento, α , entre 0 y 1 es elevado a la potencia n y multiplicado por r_n, entonces los retornos serían:

$$\sum_{n=1}^{\alpha} \alpha^n r_n(x_n, d_n)$$

Uno debe cuidarse de los problemas asociados con la existencia de la solución; lo que puede parecer la medida de efectividad obvia - para un problema de finitas etapas, puede no tener sentido cuando el concepto se extiende a infinitas etapas. Uno puede extrañarse de porqué se está interesado en obtener la solución a problemas de esta naturaleza; acaso no es correcto y efectivo hacer a N suficientemente grande, pero finita, y olvidarse de los problemas de existencia de solución? Una razón es que, en muchos casos, la estructura de la solución puede ser obtenida a partir del estudio de problemas de infinitas etapas, el conocimiento de la estructura puede darnos una mayor comprensión del problema que se tiene. Otra razón es que, en los casos cuando los problemas son de etapas contínuas, el método de infinitas etapas es el único que tiene significado físico. A continuación, se ilustra un ejemplo de solución a problemas contínuos. En este problema, la función objetivo es una integral en lugar de una suma infinita, y la ecuación en diferencia que representa a la función de transformación de la etapa es reemplazada por - una ecuación diferencial, el problema es:

$$\max \int_t^T h(x, d, t)\, dt$$

sujeto a

$$\frac{dx}{dt} = g(x, d, t)$$

La variable t, juega un papel similar a n en los problemas de etapas discretas.

El método clásico para resolver este problema es el cálculo de variaciones; sin embargo, las soluciones analíticas no pueden ser obtenidas excepto para problemas muy simples; ya que, del cálculo de variaciones resultan ecuaciones diferenciales parciales de segundo orden las cuales pueden ser muy dificultosas de resolver numéricamente.

La programación dinámica proporciona un procedimiento de aproximación numérica bastante eficiente.

Defínase:

$$f\left[x(t), t\right] = \max_{d(t)} \int_t^T h(x, d, t)\, dt$$

donde el intervalo de integración es $\left[t, T\right]$, pero

$$\int_t^T h(x, d, t)dt = \int_t^{T+\Delta t} h(x,d,t)dt + \int_{t+\Delta t}^{T} h(x,d,t)\, dt$$

sustituyendo esta identidad en la definición de $f\left[x(t),\ t\right]$ resulta:

$$f\left[x(t),\ t\right] = \max_{\substack{d(t)\\ [t,\ T]}} \int_{t}^{t+\Delta t} h(x,d,t)dt + \int_{t+\Delta t}^{T} h(x,d,t)dt$$

sujeto a

$$\frac{dx}{dt} = g(x,d,t)$$

$$f\left[x(t),t\right] = \max_{\substack{d(t)\\ [t,t+\Delta t]}} \max_{\substack{d(t)\\ [t+\Delta t,T]}} \left[\int_{t}^{t+\Delta t} h(x,d,t)dt + \int_{t+\Delta t}^{T} h(x,d,t)dt\right]$$

sujeto a

$$\frac{dx}{dt} = g(x,d,t)$$

Pero, el primer término en los corchetes depende solamente de $d(t)$ en el intervalo $\left[t,\ t+\Delta t\right]$, resultando que:

$$f\left[x(t),t\right] = \max_{\substack{d(t)\\ [t,\ t+\Delta t]}} \left[\int_{t}^{t+\Delta t} h(x,d,t)dt + \max_{\substack{d(t)\\ [t+\Delta t,T]}} \int_{t+\Delta t}^{T} h(x,d,t)dt\right]$$

sujeto a

$$\frac{dx}{dt} = g(x,d,t)$$

pero, por definición:

$$\max_{\substack{d(t)\\ [t+\Delta t,T]}} \int_{t+\Delta t}^{T} h(x,d,t)dt = f\left[x(t+\Delta t),\ t+\Delta t\right]$$

Por lo tanto,

$$f[x(t),t] = \max_{\substack{d(t)\\ [t,t+\Delta t]}} \left[\int_t^{t+\Delta t} h(x,d,t)dt + f[x(t+\Delta t), t+\Delta t] \right]$$

donde

$$x(t+\Delta t) = x(t) + \int_t^{t+\Delta t} g(x,d,t)dt$$

Para implementar el procedimiento de aproximación numérica recuérdese que se puede aproximar

$$\int_t^{t+\Delta t} h(x,d,t)dt$$

por $h(x,d,t)\,\Delta t$ para valores pequeños de Δt. Usando esta aproximación resulta:

$$f[x(t), t] = \max_{\substack{d(t)\\ [t,t+\Delta t]}} \left[h(x,d,t)\Delta t + f[x(t+\Delta t), t+\Delta t] \right]$$

sujeto a

$$x(t+\Delta t) = x(t) + g(x,d,t)\Delta t$$

Aspectos Computacionales.

En las secciones anteriores, se presentó la teoría de la programación dinámica; en esta sección, se discutirán los problemas computacionales asociados con la solución de programas dinámicos.

El procedimiento iteractivo seguido en la solución de problemas por - programación dinámica se presta por sí mismo a la adaptación a las -

computadoras digitales. De hecho, el desarrollo de la programación dinámica ha ido de la mano con el desarrollo de las computadoras digitales; para la solución de problemas en la computadora, se usa la construcción de flujogramas que ilustran el progreso de los cómputos, para construir un flujograma facilmente legible se usará una notación adicional; recordemos que la ecuación funcional de la programación dinámica tiene la forma

$$f_n(x_n) = \underset{Dn}{\text{op t}} \left[r_n(x_n, D_n) + f_{n-1}(x_{n-1}) \right]$$

sujeta a

$$x_{n-1} = t_n(x_n, D_n)$$

Defínase:

$$f_o(x_o) = 0 \qquad q_n(x_n, D_n) = r_n(x_n, Dn) + f_{n-1}(x_{n-1})$$

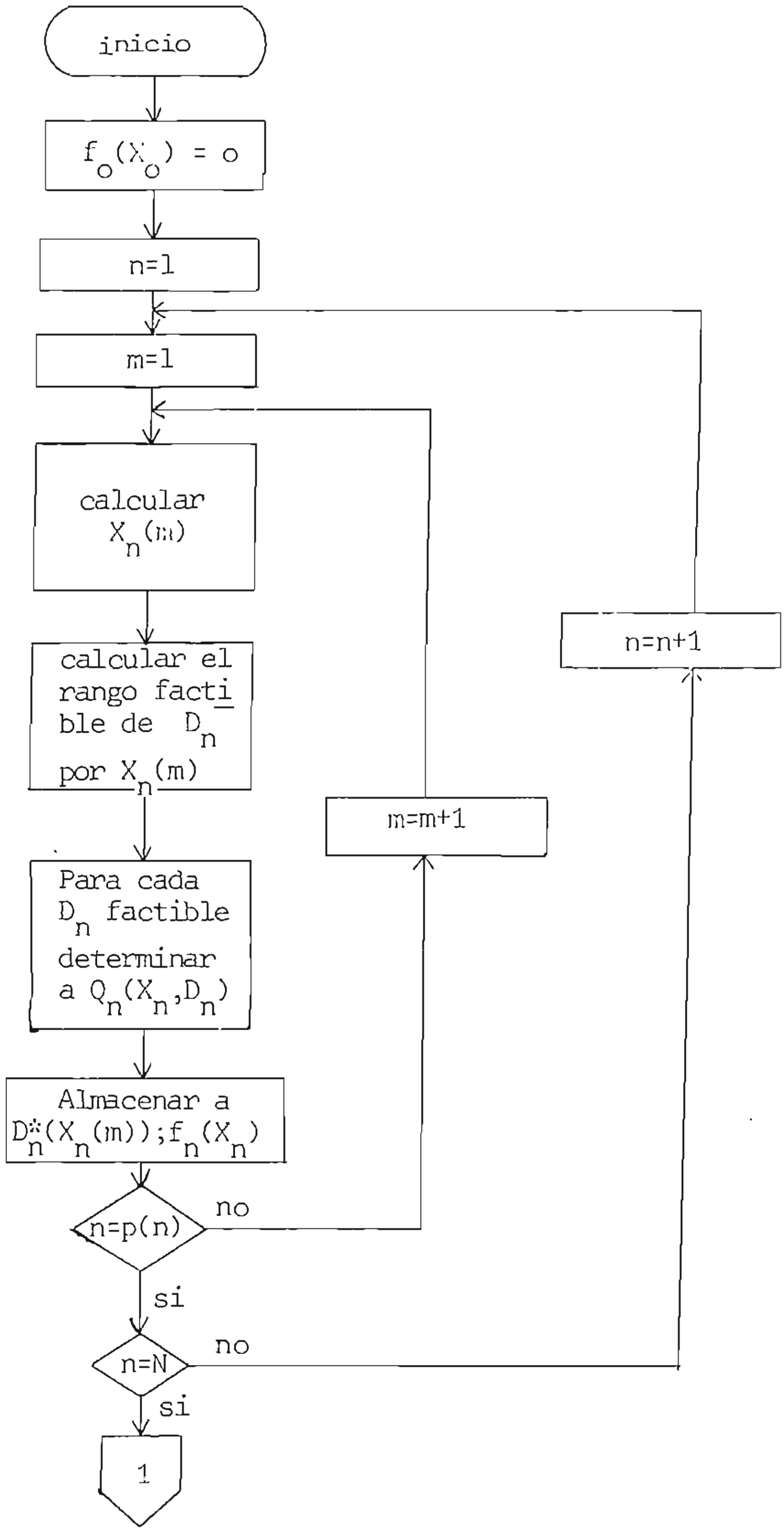
entonces:

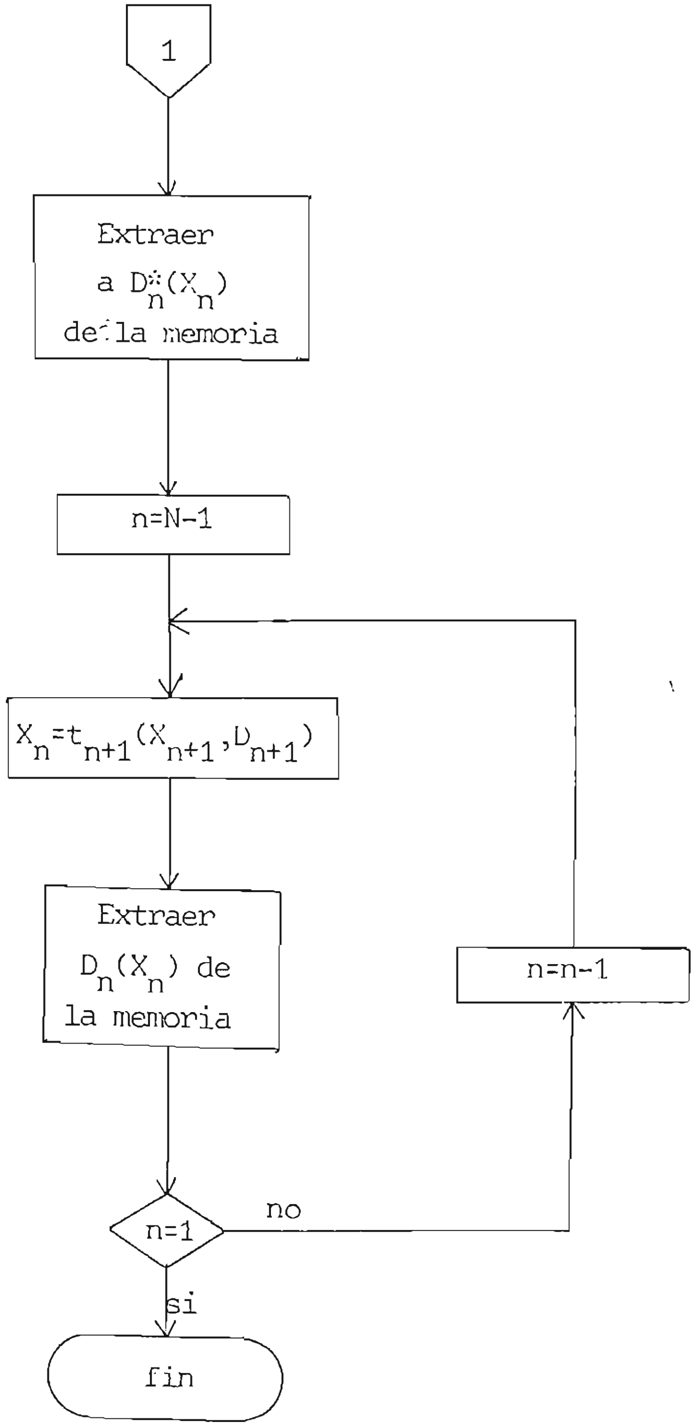
$$f_n(x_n) = \underset{D_n}{\text{op t}} \left[q_n(x_n, D_n) \right] .$$

donde x_{n-1} tiene que ser eliminado usando

$$x_{n-1} = t_n(x_n, D_n)$$

El flujograma se muestra a continuación:


inicio
$f_o(X_o) = o$
n=1
m=1
calcular $X_n(m)$
calcular el rango facti ble de D_n por $X_n(m)$
Para cada D_n factible determinar a $Q_n(X_n,D_n)$
Almacenar a $D^*_n(X_n(m));f_n(X_n)$
n=p(n)
no
si
n=N
no
si
m=m+1
n=n+1
1


1
Extraer
a $D_n^*(X_n)$
de la memoria
n=N-1
$X_n=t_{n+1}(X_{n+1},D_{n+1})$
Extraer
$D_n(X_n)$ de
la memoria
n=n-1
n=1
no
si
fin

A continuación se dará una breve explicación de algunos términos. El cuarto rectángulo especifica el cálculo de $x_n(m)$; cuando se resuelven problemas en computadoras se debe especificar en cada etapa una partición para las variables de estado y de decisión, en los problemas discretos la partición aparece en forma natural; en problemas contínuos se debe imponer una partición como una aproximación a las funciones contínuas, naturalmente se desea una partición fina; sin embargo, al afinar la partición se aumenta el número de combinaciones de variables de estado para las cuales $f_n(x_n)$ debe ser calculada; en este caso, se debe hacer un estudio acerca de la exactitud y el costo computacional, este problema no es exclusivo de la solución de problemas de programación dinámica, dificultades similares son encontradas al resolver numéricamente ecuaciones diferenciales. Los valores de $x_n(m)$, $m = 1,2,,\ldots, p(m)$, simplemente son la partición para x_n y $p(n)$ especifica el número de valores de x_n (tamaño de la partición). Obsérvese que se permite cambiar el tamaño de la partición de etapa en etapa, tal flexibilidad puede ser utilizada en cierto tipo de problemas.

Se ha mencionado anteriormente que los problemas aparecen cuando el número de combinaciones de variables de estado en cada etapa es grande; este problema, la maldición de la dimensionalidad, ha originado varios intentos para reducir el número de combinaciones de variables de estado que deben ser examinadas.

Otra formulación nos indica que solamente una variable de estado es obtenida mediante el uso del método del multiplicador de Lagrange. Supóngase que se selecciona la segunda restricción para eliminarla incorporándola dentro de la función objetivo formando la función de Lagrange,

$$F(\lambda\ d_1,\ldots,d_n) = \sum_{n=1}^{N} r_n(d_n) + \lambda(\sum_{n=1}^{N} b_n d_n - k_2 + U_1^2)$$

sujeta a

$$x_{n-1} = x_n - a_n d_n$$

Reordenándola y recordando que $\lambda_1 U_1^2 = 0$, resulta:

$$F(d_1,\ldots, d_N, \lambda) = \max \sum_{n=1}^{N} \left[r_n(d_n) + \lambda b_n d_n \right] - \lambda k_2$$

sujeta a

$$x_{n-1} = x_n - a_n d_n$$

De las partes anteriores, recordemos que F tiene un punto de silla, el cual es máximo con respecto a $d_1,\ldots, d_n$ y mínimo con respecto a λ ; luego, se puede tratar el problema de maximización mediante la programación dinámica.

$$f_o(x_o, \lambda) = 0$$

$$f_n(x_n, \lambda) = \max_{d_n} \left[r_n(d_n) + \lambda a_n d_n + f_{n-1}(x_{n-1}) \right]$$

sujeto a

$$x_{n-1} = x_n - a_n d_n$$

El término λk_2 ha sido omitido, ya que para un valor fijo de λ resulta ser una constante y por lo tanto, no afecta la política de decisión óptima.

El procedimiento computacional es como sigue:
Seleccione un valor para $\lambda = \lambda^0 \neq 0$ y resuelva el problema, para un λ^0 dado generar la regla de decisión óptima $d_1^0, \ldots, d_N^0$. Ya que se seleccionó $\lambda^0 \neq 0$ la restricción debe cumplirse como una igualdad estricta para algún valor de k_2, digamos k_2^0. Sin embargo, excepto en la situación más fortuita, k_2^0 no será el valor deseado k_2. Por lo tanto, se selecciona otro valor de λ, digamos λ^1, y se resuelve el problema; esto puede parecer un procedimiento de prueba y error, sin embargo, recordemos que k_2 será una función monótona creciente de λ ; luego, se pueden orientar las sucesivas selecciones de λ para converger al valor deseado k_2. Esta técnica está garantizada a converger el valor deseado de k_2 si las restricciones forman un conjunto convexo; en otros casos, puede no existir un valor de λ para el cual resulte el valor deseado de k_2.

Otro método que sirve para reducir el número de combinaciones de variables de estado para las cuales $f_n(x_n)$ debe ser calculado y almacenado es llamado "El método de una a la vez". Supóngase que se considera el siguiente problema con dos variables de decisión en cada -

etapa:

$$\max \sum_{n=1}^{N} r_n(d_n^{(1)}, d_n^{(2)})$$

Además, supóngase que sólo se tienen dos restricciones y que ellas son separables, es decir, cada restricción es una función de d_n^1 ó de d_n^2 pero no de las dos

$$\sum_{n=1}^{N} a_n d_n^{(1)} \leq k_1$$

$$\sum_{n=1}^{N} b_n d_n^{(2)} \leq k_2$$

En este caso, cuando las restricciones formen un conjunto convexo, el procedimiento de aproximación puede ser aplicado; cuando la función objetivo es cóncava se puede aplicar un procedimiento similar a las técnicas de búsqueda de la programación no lineal. Se procederá así: seleccione un conjunto de d_n^i $i = 1$ ó 2, el cual satisfaga las restricciones, para ser más explícitos se puede seleccionar $d_n^2 = 0$ $n = 1,\ldots, N$; ahora, se resuelve el problema

$$\max \sum_{n=1}^{N} r_n(d_n^{(1)}, d_n^{(2)} = 0)$$

sujeto a:

$$\sum_{n=1}^{N} a_n d_n^{(1)} \leq k_1$$

mediante el método estándar de la programación dinámica. Esto nos lleva a una solución para $d_n^{(1)} = d_n^{(1)}$ $(k_2 = 0)$; ahora, manteniendo fijo a $d_n^{(1)}$ en esta solución, se resuelve el siguiente problema:

$$\max \sum_{n=1}^{N} r_n \left[d_n^1 (k_2 = 0),\ d_n^2 \right]$$

sujeto a

$$\sum_{n=1}^{N} b_n d_n^{(2)} \leqslant \Delta > 0$$

Esto nos lleva a un nuevo conjunto de valores para $d_n^{(2)}$ que satisfacen las restricciones anteriores; manteniendo fijos los nuevos valores de $d_n^{(2)}$ se resuelve el siguiente problema:

$$\max \sum_{n=1}^{N} r_n \left[d_n^{(1)},\ d_n^{(2)}(\Delta) \right]$$

sujeto a

$$\sum_{n=1}^{N} a_n d_n^{(1)} \leqslant k_1$$

Este procedimiento iterativo se continúa alternadamente, resolviendo para valores de $d_n^{(1)}$ y $d_n^{(2)}$. En cada iteración, Δ es agregado al lado derecho de la restricción; en particular en la iteración $(p+1)$ se tiene:

$$\sum_{n=1}^{N} b_n d_n \leqslant p\Delta$$

Cuando las variables d_n^i son contínuos, el procedimiento iterativo converge a la solución óptima del problema original.

Se han discutido dos procedimientos para reducir el problema de almacenamiento causado por problemas en los cuales existe un gran número de combinaciones de valores de las variables de estado a investigar.

Otra técnica es aproximar la función $f_{n-1}(x_{n-1})$ mediante un polinomio de bajo orden, si una aproximación apropiada puede ser construida, el único almacenamiento requerido sería para los coeficientes del polinomio; este método no reduce los problemas computacionales introducidos por los problemas de dimensionalidad, pero alivia el problema de almacenamiento asociado con la maldición de la dimensionalidad.

El último procedimiento que se discutirá es llamado la técnica de la partición áspera, en lugar de definir una partición fina, se usa un número relativamente pequeño de puntos en la partición y se resuelve el problema para este conjunto reducido de variables de estado. Una vez que la solución a este problema es obtenida, una partición más - fina es construido alrededor de los puntos, para aquellos valores de las variables de estado de las cuales resultó la regla de decisión -

óptima a la aproximación original; este esquema se continúa hasta que se obtenga una solución precisa aceptable. Debería de darse - cuenta, sin embargo, que cuando existe un óptimo local alternativo, la solución final puede ser cualquiera de estos óptimos locales; además, el óptimo local obtenido puede no ser un óptimo global.

A N E X O S

TABLA 01

X2	D2	X1(=X2-D2)	D2^2	FI(X1)	D2^2+F1(X1)	F2(X2)
0	0*	0	0	0	0	0
1	0*	1	0	1	1	1
1	1*	0	1	0	1	1
2	0*	2	0	4	4	4
2	1	1	1	1	2	
2	2*	0	4	0	4	4
3	0*	3	0	9	9	9
3	1	2	1	4	5	
3	2	1	4	1	5	
3	3*	0	9	0	9	9
4	0*	4	0	16	16	16
4	1	3	1	9	10	
4	2	2	4	4	8	
4	3	1	9	1	10	
4	4*	0	16	0	16	16
5	0*	5	0	25	25	25
5	1	4	1	16	17	
5	2	3	4	9	13	
5	3	2	9	4	13	
5	4	1	16	1	17	
5	5*	0	25	0	25	25

TABLA 02

X3	D3	X2(=X3-D3)	D3^2	F2(X2)	D3^2+F2(X2)	F3(X3)
5	0*	5	0	25	25	25
5	1	4	1	16	17	
5	2	3	4	9	13	
5	3	2	9	4	13	
5	4	1	16	1	17	
5	5*	0	25	0	25	25

TABLA 03

X1	Y1	D1	C1*D1	F1(X1,Y1)
0	0	0*	0	0
0	1	0*	0	0
0	2	0*	0	0
0	3	0*	0	0
0	4	0*	0	0
0	5	0*	0	0
0	6	0*	0	0
0	7	0*	0	0
0	8	0*	0	0
0	9	0*	0	0
0	10	0*	0	0
0	11	0*	0	0
0	12	0*	0	0
0	13	0*	0	0
0	14	0*	0	0
0	15	0*	0	0
1	0	0*	0	0
1	1	0*	0	0
1	2	0*	0	0
1	3	0*	0	0
1	4	0*	0	0
1	5	0*	0	0
1	6	0*	0	0
1	7	0*	0	0
1	8	0*	0	0
1	9	0*	0	0
1	10	0*	0	0
1	11	0*	0	0
1	12	0*	0	0
1	13	0*	0	0
1	14	0*	0	0
1	15	0*	0	0
2	0	0*	0	0
2	1	0*	0	0
2	2	0*	0	0
2	3	0	0	
2	3	1*	1	1
2	4	0	0	
2	4	1*	1	1
2	5	0	0	
2	5	1*	1	1
2	6	0	0	
2	6	1*	1	1
2	7	0	0	

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
2	7	1*	1	1
2	8	0	0	
2	8	1*	1	1
2	9	0	0	
2	9	1*	1	1
2	10	0	0	
2	10	1*	1	1
2	11	0	0	
2	11	1*	1	1
2	12	0	0	
2	12	1*	1	1
2	13	0	0	
2	13	1*	1	1
2	14	0	0	
2	14	1*	1	1
2	15	0	0	
2	15	1*	1	1
3	0	0*	0	0
3	1	0*	0	0
3	2	0*	0	0
3	3	0	0	
3	3	1*	1	1
3	4	0	0	
3	4	1*	1	1
3	5	0	0	
3	5	1*	1	1
3	6	0	0	
3	6	1*	1	1
3	7	0	0	
3	7	1*	1	1
3	8	0	0	
3	8	1*	1	1
3	9	0	0	
3	9	1*	1	1
3	10	0	0	
3	10	1*	1	1
3	11	0	0	
3	11	1*	1	1
3	12	0	0	
3	12	1*	1	1
3	13	0	0	
3	13	1*	1	1
3	14	0	0	
3	14	1*	1	1

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
3	15	0	0	
3	15	1*	1	1
4	0	0*	0	0
4	1	0*	0	0
4	2	0*	0	0
4	3	0	0	
4	3	1*	1	1
4	4	0	0	
4	4	1*	1	1
4	5	0	0	
4	5	1*	1	1
4	6	0	0	
4	6	1	1	
4	6	2*	2	2
4	7	0	0	
4	7	1	1	
4	7	2*	2	2
4	8	0	0	
4	8	1	1	
4	8	2*	2	2
4	9	0	0	
4	9	1	1	
4	9	2*	2	2
4	10	0	0	
4	10	1	1	
4	10	2*	2	2
4	11	0	0	
4	11	1	1	
4	11	2*	2	2
4	12	0	0	
4	12	1	1	
4	12	2*	2	2
4	13	0	0	
4	13	1	1	
4	13	2*	2	2
4	14	0	0	
4	14	1	1	
4	14	2*	2	2
4	15	0	0	
4	15	1	1	
4	15	2*	2	2
5	0	0*	0	0
5	1	0*	0	0
5	2	0*	0	0

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
5	3	0	0	
5	3	1*	1	1
5	4	0	0	
5	4	1*	1	1
5	5	0	0	
5	5	1*	1	1
5	6	0	0	
5	6	1	1	
5	6	2*	2	2
5	7	0	0	
5	7	1	1	
5	7	2*	2	2
5	8	0	0	
5	8	1	1	
5	8	2*	2	2
5	9	0	0	
5	9	1	1	
5	9	2*	2	2
5	10	0	0	
5	10	1	1	
5	10	2*	2	2
5	11	0	0	
5	11	1	1	
5	11	2*	2	2
5	12	0	0	
5	12	1	1	
5	12	2*	2	2
5	13	0	0	
5	13	1	1	
5	13	2*	2	2
5	14	0	0	
5	14	1	1	
5	14	2*	2	2
5	15	0	0	
5	15	1	1	
5	15	2*	2	2
6	0	0*	0	0
6	1	0*	0	0
6	2	0*	0	0
6	3	0	0	
6	3	1*	1	1
6	4	0	0	
6	4	1*	1	1
6	5	0	0	

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
6	5	1*	1	1
6	6	0	0	
6	6	1	1	
6	6	2*	2	2
6	7	0	0	
6	7	1	1	
6	7	2*	2	2
6	8	0	0	
6	8	1	1	
6	8	2*	2	2
6	9	0	0	
6	9	1	1	
6	9	2	2	
6	9	3*	3	3
6	10	0	0	
6	10	1	1	
6	10	2	2	
6	10	3*	3	3
6	11	0	0	
6	11	1	1	
6	11	2	2	
6	11	3*	3	3
6	12	0	0	
6	12	1	1	
6	12	2	2	
6	12	3*	3	3
6	13	0	0	
6	13	1	1	
6	13	2	2	
6	13	3*	3	3
6	14	0	0	
6	14	1	1	
6	14	2	2	
6	14	3*	3	3
6	15	0	0	
6	15	1	1	
6	15	2	2	
6	15	3*	3	3
7	0	0*	0	0
7	1	0*	0	0
7	2	0*	0	0
7	3	0	0	
7	3	1*	1	1
7	4	0	0	

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
7	4	1*	1	1
7	5	0	0	
7	5	1*	1	1
7	6	0	0	
7	6	1	1	
7	6	2*	2	2
7	7	0	0	
7	7	1	1	
7	7	2*	2	2
7	8	0	0	
7	8	1	1	
7	8	2*	2	2
7	9	0	0	
7	9	1	1	
7	9	2	2	
7	9	3*	3	3
7	10	0	0	
7	10	1	1	
7	10	2	2	
7	10	3*	3	3
7	11	0	0	
7	11	1	1	
7	11	2	2	
7	11	3*	3	3
7	12	0	0	
7	12	1	1	
7	12	2	2	
7	12	3*	3	3
7	13	0	0	
7	13	1	1	
7	13	2	2	
7	13	3*	3	3
7	14	0	0	
7	14	1	1	
7	14	2	2	
7	14	3*	3	3
7	15	0	0	
7	15	1	1	
7	15	2	2	
7	15	3*	3	3
8	0	0*	0	0
8	1	0*	0	0
8	2	0*	0	0
8	3	0	0	

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
8	3	1*	1	1
8	4	0	0	
8	4	1*	1	1
8	5	0	0	
8	5	1*	1	1
8	6	0	0	
8	6	1	1	
8	6	2*	2	2
8	7	0	0	
8	7	1	1	
8	7	2*	2	2
8	8	0	0	
8	8	1	1	
8	8	2*	2	2
8	9	0	0	
8	9	1	1	
8	9	2	2	
8	9	3*	3	3
8	10	0	0	
8	10	1	1	
8	10	2	2	
8	10	3*	3	3
8	11	0	0	
8	11	1	1	
8	11	2	2	
8	11	3*	3	3
8	12	0	0	
8	12	1	1	
8	12	2	2	
8	12	3	3	
8	12	4*	4	4
8	13	0	0	
8	13	1	1	
8	13	2	2	
8	13	3	3	
8	13	4*	4	4
8	14	0	0	
8	14	1	1	
8	14	2	2	
8	14	3	3	
8	14	4*	4	4
8	15	0	0	
8	15	1	1	
8	15	2	2	

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
8	15	3	3	
8	15	4*	4	4
9	0	0*	0	0
9	1	0*	0	0
9	2	0*	0	0
9	3	0	0	
9	3	1*	1	1
9	4	0	0	
9	4	1*	1	1
9	5	0	0	
9	5	1*	1	1
9	6	0	0	
9	6	1	1	
9	6	2*	2	2
9	7	0	0	
9	7	1	1	
9	7	2*	2	2
9	8	0	0	
9	8	1	1	
9	8	2*	2	2
9	9	0	0	
9	9	1	1	
9	9	2	2	
9	9	3*	3	3
9	10	0	0	
9	10	1	1	
9	10	2	2	
9	10	3*	3	3
9	11	0	0	
9	11	1	1	
9	11	2	2	
9	11	3*	3	3
9	12	0	0	
9	12	1	1	
9	12	2	2	
9	12	3	3	
9	12	4*	4	4
9	13	0	0	
9	13	1	1	
9	13	2	2	
9	13	3	3	
9	13	4*	4	4
9	14	0	0	
9	14	1	1	

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
9	14	2	2	
9	14	3	3	
9	14	4*	4	4
9	15	0	0	
9	15	1	1	
9	15	2	2	
9	15	3	3	
9	15	4*	4	4
10	0	0*	0	0
10	1	0*	0	0
10	2	0*	0	0
10	3	0	0	
10	3	1*	1	1
10	4	0	0	
10	4	1*	1	1
10	5	0	0	
10	5	1*	1	1
10	6	0	0	
10	6	1	1	
10	6	2*	2	2
10	7	0	0	
10	7	1	1	
10	7	2*	2	2
10	8	0	0	
10	8	1	1	
10	8	2*	2	2
10	9	0	0	
10	9	1	1	
10	9	2	2	
10	9	3*	3	3
10	10	0	0	
10	10	1	1	
10	10	2	2	
10	10	3*	3	3
10	11	0	0	
10	11	1	1	
10	11	2	2	
10	11	3*	3	3
10	12	0	0	
10	12	1	1	
10	12	2	2	
10	12	3	3	
10	12	4*	4	4
10	13	0	0	

TABLA 03 (Continuacion)

X1	Y1	D1	C1*D1	F1(X1,Y1)
10	13	1	1	
10	13	2	2	
10	13	3	3	
10	13	4*	4	4
10	14	0	0	
10	14	1	1	
10	14	2	2	
10	14	3	3	
10	14	4*	4	4
10	15	0	0	
10	15	1	1	
10	15	2	2	
10	15	3	3	
10	15	4	4	
10	15	5*	5	5

TABLA 04

X2	Y2	D2	X1	Y1	C2*D2	F1(X1,Y1)	C2*D2+F1(X1,Y1)	F2(X2,Y2)
0	0	0*	0	0	0	0	0	0
0	1	0*	0	1	0	0	0	0
0	2	0*	0	2	0	0	0	0
0	3	0*	0	3	0	0	0	0
0	4	0*	0	4	0	0	0	0
0	5	0*	0	5	0	0	0	0
0	6	0*	0	6	0	0	0	0
0	7	0*	0	7	0	0	0	0
0	8	0*	0	8	0	0	0	0
0	9	0*	0	9	0	0	0	0
0	10	0*	0	10	0	0	0	0
0	11	0*	0	11	0	0	0	0
0	12	0*	0	12	0	0	0	0
0	13	0*	0	13	0	0	0	0
0	14	0*	0	14	0	0	0	0
0	15	0*	0	15	0	0	0	0
1	0	0*	1	0	0	0	0	0
1	1	0*	1	1	0	0	0	0
1	2	0*	1	2	0	0	0	0
1	3	0*	1	3	0	0	0	0
1	4	0*	1	4	0	0	0	0
1	5	0*	1	5	0	0	0	0
1	6	0*	1	6	0	0	0	0
1	7	0*	1	7	0	0	0	0
1	8	0*	1	8	0	0	0	0
1	9	0*	1	9	0	0	0	0
1	10	0*	1	10	0	0	0	0
1	11	0*	1	11	0	0	0	0
1	12	0*	1	12	0	0	0	0
1	13	0*	1	13	0	0	0	0
1	14	0*	1	14	0	0	0	0
1	15	0*	1	15	0	0	0	0
2	0	0*	2	0	0	0	0	0
2	1	0*	2	1	0	0	0	0
2	2	0*	2	2	0	0	0	0
2	3	0*	2	3	0	1	1	1
2	4	0*	2	4	0	1	1	1
2	5	0*	2	5	0	1	1	1
2	6	0*	2	6	0	1	1	1
2	7	0*	2	7	0	1	1	1
2	8	0*	2	8	0	1	1	1
2	9	0*	2	9	0	1	1	1
2	10	0*	2	10	0	1	1	1
2	11	0*	2	11	0	1	1	1

TABLA 04 (Continuacion)

X2	Y2	D2	X1	Y1	C2*D2	F1(X1,Y1)	C2*D2+F1(X1,Y1)	F2(X2,Y2)
2	12	0*	2	12	0	1	1	1
2	13	0*	2	13	0	1	1	1
2	14	0*	2	14	0	1	1	1
2	15	0*	2	15	0	0	0	0
3	0	0*	3	0	0	0	0	0
3	1	0*	3	1	0	0	0	0
3	2	0*	3	2	0	0	0	0
3	3	0*	3	3	0	1	1	1
3	4	0*	3	4	0	1	1	1
3	5	0	3	5	0	1	1	
3	5	1*	0	0	3	0	3	3
3	6	0	3	6	0	1	1	
3	6	1*	0	1	3	0	3	3
3	7	0	3	7	0	1	1	
3	7	1*	0	2	3	0	3	3
3	8	0	3	8	0	1	1	
3	8	1*	0	3	3	0	3	3
3	9	0	3	9	0	1	1	
3	9	1*	0	4	3	0	3	3
3	10	0	3	10	0	1	1	
3	10	1*	0	5	3	0	3	3
3	11	0	3	11	0	1	1	
3	11	1*	0	6	3	0	3	3
3	12	0	3	12	0	1	1	
3	12	1*	0	7	3	0	3	3
3	13	0	3	13	0	1	1	
3	13	1*	0	8	3	0	3	3
3	14	0	3	14	0	1	1	
3	14	1*	0	9	3	0	3	3
3	15	0	3	15	0	0	0	
3	15	1*	0	10	3	0	3	3
4	0	0*	4	0	0	0	0	0
4	1	0*	4	1	0	0	0	0
4	2	0*	4	2	0	0	0	0
4	3	0*	4	3	0	1	1	1
4	4	0*	4	4	0	1	1	1
4	5	0	4	5	0	1	1	
4	5	1*	1	0	3	0	3	3
4	6	0	4	6	0	2	2	
4	6	1*	1	1	3	0	3	3
4	7	0	4	7	0	2	2	
4	7	1*	1	2	3	0	3	3
4	8	0	4	8	0	2	2	
4	8	1*	1	3	3	0	3	3

TABLA 04 (Continuacion)

X2	Y2	D2	X1	Y1	C2*D2	F1(X1,Y1)	C2*D2+F1(X1,Y1)	F2(X2,Y2)
4	9	0	4	9	0	2	2	
4	9	1*	1	4	3	0	3	3
4	10	0	4	10	0	2	2	
4	10	1*	1	5	3	0	3	3
4	11	0	4	11	0	2	2	
4	11	1*	1	6	3	0	3	3
4	12	0	4	12	0	2	2	
4	12	1*	1	7	3	0	3	3
4	13	0	4	13	0	2	2	
4	13	1*	1	8	3	0	3	3
4	14	0	4	14	0	2	2	
4	14	1*	1	9	3	0	3	3
4	15	0	4	15	0	0	0	
4	15	1*	1	10	3	0	3	3
5	0	0*	5	0	0	0	0	0
5	1	0*	5	1	0	0	0	0
5	2	0*	5	2	0	0	0	0
5	3	0*	5	3	0	1	1	1
5	4	0*	5	4	0	1	1	1
5	5	0	5	5	0	1	1	
5	5	1*	2	0	3	0	3	3
5	6	0	5	6	0	2	2	
5	6	1*	2	1	3	0	3	3
5	7	0	5	7	0	2	2	
5	7	1*	2	2	3	0	3	3
5	8	0	5	8	0	2	2	
5	8	1*	2	3	3	1	4	4
5	9	0	5	9	0	2	2	
5	9	1*	2	4	3	1	4	4
5	10	0	5	10	0	2	2	
5	10	1*	2	5	3	1	4	4
5	11	0	5	11	0	2	2	
5	11	1*	2	6	3	1	4	4
5	12	0	5	12	0	2	2	
5	12	1*	2	7	3	1	4	4
5	13	0	5	13	0	2	2	
5	13	1*	2	8	3	1	4	4
5	14	0	5	14	0	2	2	
5	14	1*	2	9	3	1	4	4
5	15	0	5	15	0	0	0	
5	15	1*	2	10	3	1	4	4
6	0	0*	6	0	0	0	0	0
6	1	0*	6	1	0	0	0	0
6	2	0*	6	2	0	0	0	0

TABLA 04 (Continuacion)

X2	Y2	D2	X1	Y1	C2*D2	F1(X1,Y1)	C2*D2+F1(X1,Y1)	F2(X2,Y2)
6	3	0*	6	3	0	1	1	1
6	4	0*	6	4	0	1	1	1
6	5	0	6	5	0	1	1	
6	5	1*	3	0	3	0	3	3
6	6	0	6	6	0	2	2	
6	6	1*	3	1	3	0	3	3
6	7	0	6	7	0	2	2	
6	7	1*	3	2	3	0	3	3
6	8	0	6	8	0	2	2	
6	8	1*	3	3	3	1	4	4
6	9	0	6	9	0	3	3	
6	9	1*	3	4	3	1	4	4
6	10	0	6	10	0	3	3	
6	10	1	3	5	3	1	4	
6	10	2*	0	0	6	0	6	6
6	11	0	6	11	0	3	3	
6	11	1	3	6	3	1	4	
6	11	2*	0	1	6	0	6	6
6	12	0	6	12	0	3	3	
6	12	1	3	7	3	1	4	
6	12	2*	0	2	6	0	6	6
6	13	0	6	13	0	3	3	
6	13	1	3	8	3	1	4	
6	13	2*	0	3	6	0	6	6
6	14	0	6	14	0	3	3	
6	14	1	3	9	3	1	4	
6	14	2*	0	4	6	0	6	6
6	15	0	6	15	0	0	0	
6	15	1	3	10	3	1	4	
6	15	2*	0	5	6	0	6	6
7	0	0*	7	0	0	0	0	0
7	1	0*	7	1	0	0	0	0
7	2	0*	7	2	0	0	0	0
7	3	0*	7	3	0	1	1	1
7	4	0*	7	4	0	1	1	1
7	5	0	7	5	0	1	1	
7	5	1*	4	0	3	0	3	3
7	6	0	7	6	0	2	2	
7	6	1*	4	1	3	0	3	3
7	7	0	7	7	0	2	2	
7	7	1*	4	2	3	0	3	3
7	8	0	7	8	0	2	2	
7	8	1*	4	3	3	1	4	4
7	9	0	7	9	0	3	3	

TABLA 04 (Continuacion)

X2	Y2	D2	X1	Y1	C2*D2	F1(X1,Y1)	C2*D2+F1(X1,Y1)	F2(X2,Y2)
7	9	1*	4	4	3	1	4	4
7	10	0	7	10	0	3	3	
7	10	1	4	5	3	1	4	
7	10	2*	1	0	6	0	6	6
7	11	0	7	11	0	3	3	
7	11	1	4	6	3	2	5	
7	11	2*	1	1	6	0	6	6
7	12	0	7	12	0	3	3	
7	12	1	4	7	3	2	5	
7	12	2*	1	2	6	0	6	6
7	13	0	7	13	0	3	3	
7	13	1	4	8	3	2	5	
7	13	2*	1	3	6	0	6	6
7	14	0	7	14	0	3	3	
7	14	1	4	9	3	2	5	
7	14	2*	1	4	6	0	6	6
7	15	0	7	15	0	0	0	
7	15	1	4	10	3	2	5	
7	15	2*	1	5	6	0	6	6
8	0	0*	8	0	0	0	0	0
8	1	0*	8	1	0	0	0	0
8	2	0*	8	2	0	0	0	0
8	3	0*	8	3	0	1	1	1
8	4	0*	8	4	0	1	1	1
8	5	0	8	5	0	1	1	
8	5	1*	5	0	3	0	3	3
8	6	0	8	6	0	2	2	
8	6	1*	5	1	3	0	3	3
8	7	0	8	7	0	2	2	
8	7	1*	5	2	3	0	3	3
8	8	0	8	8	0	2	2	
8	8	1*	5	3	3	1	4	4
8	9	0	8	9	0	3	3	
8	9	1*	5	4	3	1	4	4
8	10	0	8	10	0	3	3	
8	10	1	5	5	3	1	4	
8	10	2*	2	0	6	0	6	6
8	11	0	8	11	0	3	3	
8	11	1	5	6	3	2	5	
8	11	2*	2	1	6	0	6	6
8	12	0	8	12	0	4	4	
8	12	1	5	7	3	2	5	
8	12	2*	2	2	6	0	6	6
8	13	0	8	13	0	4	4	

TABLA 04 (Continuacion)

X2	Y2	D2	X1	Y1	C2*D2	F1(X1,Y1)	C2*D2+F1(X1,Y1)	F2(X2,Y2)
8	13	1	5	8	3	2	5	
8	13	2*	2	3	6	1	7	7
8	14	0	8	14	0	4	4	
8	14	1	5	9	3	2	5	
8	14	2*	2	4	6	1	7	7
8	15	0	8	15	0	0	0	
8	15	1	5	10	3	2	5	
8	15	2*	2	5	6	1	7	7
9	0	0*	9	0	0	0	0	0
9	1	0*	9	1	0	0	0	0
9	2	0*	9	2	0	0	0	0
9	3	0*	9	3	0	1	1	1
9	4	0*	9	4	0	1	1	1
9	5	0	9	5	0	1	1	
9	5	1*	6	0	3	0	3	3
9	6	0	9	6	0	2	2	
9	6	1*	6	1	3	0	3	3
9	7	0	9	7	0	2	2	
9	7	1*	6	2	3	0	3	3
9	8	0	9	8	0	2	2	
9	8	1*	6	3	3	1	4	4
9	9	0	9	9	0	3	3	
9	9	1*	6	4	3	1	4	4
9	10	0	9	10	0	3	3	
9	10	1	6	5	3	1	4	
9	10	2*	3	0	6	0	6	6
9	11	0	9	11	0	3	3	
9	11	1	6	6	3	2	5	
9	11	2*	3	1	6	0	6	6
9	12	0	9	12	0	4	4	
9	12	1	6	7	3	2	5	
9	12	2*	3	2	6	0	6	6
9	13	0	9	13	0	4	4	
9	13	1	6	8	3	2	5	
9	13	2*	3	3	6	1	7	7
9	14	0	9	14	0	4	4	
9	14	1	6	9	3	3	6	
9	14	2*	3	4	6	1	7	7
9	15	0	9	15	0	0	0	
9	15	1	6	10	3	3	6	
9	15	2	3	5	6	1	7	
[illegible]	15	3*	0	0	9	0	9	9
	0	0*	10	0	0	0	0	0
	1	0*	10	1	0	0	0	0

TABLA 04 (Continuacion)

X2	Y2	D2	X1	Y1	C2*D2	F1(X1,Y1)	C2*D2+F1(X1,Y1)	F2(X2,Y2)
10	2	0*	10	2	0	0	0	0
10	3	0*	10	3	0	1	1	1
10	4	0*	10	4	0	1	1	1
10	5	0	10	5	0	1	1	
10	5	1*	7	0	3	0	3	3
10	6	0	10	6	0	2	2	
10	6	1*	7	1	3	0	3	3
10	7	0	10	7	0	2	2	
10	7	1*	7	2	3	0	3	3
10	8	0	10	8	0	2	2	
10	8	1*	7	3	3	1	4	4
10	9	0	10	9	0	3	3	
10	9	1*	7	4	3	1	4	4
10	10	0	10	10	0	3	3	
10	10	1	7	5	3	1	4	
10	10	2*	4	0	6	0	6	6
10	11	0	10	11	0	3	3	
10	11	1	7	6	3	2	5	
10	11	2*	4	1	6	0	6	6
10	12	0	10	12	0	4	4	
10	12	1	7	7	3	2	5	
10	12	2*	4	2	6	0	6	6
10	13	0	10	13	0	4	4	
10	13	1	7	8	3	2	5	
10	13	2*	4	3	6	1	7	7
10	14	0	10	14	0	4	4	
10	14	1	7	9	3	3	6	
10	14	2*	4	4	6	1	7	7
10	15	0	10	15	0	5	5	
10	15	1	7	10	3	3	6	
10	15	2	4	5	6	1	7	
10	15	3*	1	0	9	0	9	9

TABLA 05

X3	Y3	D3	X2	Y2	C3*D3	F2(X2,Y2)	C3*D3+F2(X2,Y2)	F3(X3,Y3)
10	15	0	10	15	0	9	9	
10	15	1	6	8	5	4	9	
10	15	2*	2	1	10	0	10	10