CAPÍTULO XI

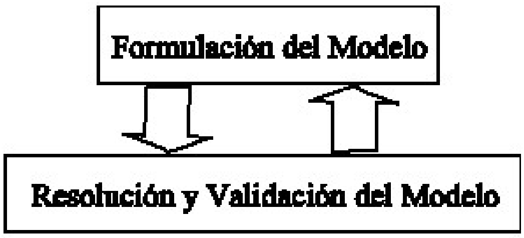
OPTIMIZACIÓN

Por
Marta B. Ferrero y Omar J. A. Chiotti

XI.1 INTRODUCCIÓN

El concepto de optimización data de tiempos inmemorables y fue incluido en la empresa cuando el mercado comprador que caracterizó las primeras décadas de la revolución industrial comenzó a transformarse hasta convertirse en el mercado vendedor fuertemente competitivo de nuestros días.

Se puede definir como optimización al proceso de seleccionar, a partir de un conjunto de alternativas posibles, aquella que mejor satisfaga el o los objetivos propuestos. Para resolver un problema de optimización se requieren dos etapas principales:



La formulación del modelo de optimización no es un procedimiento formal estructurado, sino más bien es un proceso que requiere de experiencia y creatividad. Una vez generado el modelo, la etapa siguiente es resolver y validar dicho modelo. Esta etapa puede considerarse suficientemente formalizada puesto que los modelos de problemas de optimización han sido muy estudiados y se han desarrollado innumerables métodos y estrategias para resolverlos. En este capítulo se describe una estrategia para formular el modelo y un ejemplo de aplicación. Posteriormente se presenta una descripción conceptual de la teoría y de los principales algoritmos asociados a un grupo particular de modelos matemáticos de optimización denominados *programación lineal y programación no lineal.*

XI.2 FORMULACIÓN DEL MODELO

Si bien, como se mencionara anteriormente, el proceso de modelado es esencialmente cualitativo y requiere de la habilidad y la experiencia de quien

ISBN: 950-42-0022-2 - 81999

desarrolla el modelo, en términos generales se pueden definir los siguientes pasos a seguir para la formulación del modelo:

Identificar las Variables de Decisión
Identificar y/o fijar las restricciones
Definición de los Objetivos
Análisis de la Información Disponible

Identificar las Variables de Decisión: Las variables de decisión representan las alternativas de decisión del problema. Pertenecen a la propia naturaleza del problema y no pueden ser establecidas arbitrariamente.
Identificar y/o fijar las restricciones: Las restricciones de un problema de optimización definen el conjunto de valores que pueden tomar las variables de decisión. En el caso de restricciones de igualdad, éstas además generan dependencia entre variables, reduciendo los grados de libertad del problema. El conjunto de todas las variables del problema se divide así en el subconjunto de variables independientes y el subconjunto de las variables dependientes.

Las restricciones pueden pertenecer a la naturaleza del problema, como lo son las restricciones físicas (límites de presión y temperatura, equilibrio liquido-vapor, etc.), pero también puede haber restricciones fijadas arbitrariamente por quien debe decidir, según su propio criterio.
Definición de los Objetivos: Los objetivos no pertenecen a la naturaleza del problema sino que son fijados arbitrariamente por quien debe decidir. El mismo puede definir un único objetivo o varios objetivos a ser considerados simultáneamente. Por ejemplo se suelen definir como objetivos: la rentabilidad del proceso, la calidad del producto, la seguridad del proceso, la satisfacción del cliente, etc. En este capítulo, sólo se considerarán los problemas con objetivo único. Para problemas con múltiples objetivos se puede consultar la bibliografía del final del capítulo.
Análisis de la Información Disponible: La información a cerca de los parámetros del proceso permitirá definir el criterio de decisión a adoptar. Si se conoce con certeza el valor de los parámetros, el criterio seleccionado será el de maximizar o minimizar el objetivo propuesto. En el extremo opuesto es posible encontrar parámetros cuyo valor es incierto. Usualmente en estos casos con algún criterio es posible definir para cada parámetro sujeto a incertidumbre un rango de valores posibles, quedando así definida una región paramétrica. Los criterios de decisión a utilizar en estos problemas son generalmente conservativos, aspirando a asegurar lo mejor para los peores valores que pueden ocurrir. Como ejemplo puede mencionarse el trabajo de Sargent & Grossmann 1983, quienes determinan la existencia de un punto paramétrico crítico (peor caso), tal que el diseño calculado para ese punto es factible para cualquier valor de la región paramétrica que puedan tomar los parámetros.

En el caso que para estos parámetros cuyo valor está sujeto a incertidumbre, se dispusiera de una función de densidad de probabilidad, el tomador de decisión

ISBN: 950-42-0022-2 - ©1999

podría arriesgarse a tomar decisiones en función de esa información probabilística, adoptando como criterio de decisión optimizar el *valor esperado* del objetivo elegido.

Nuevamente, en virtud del carácter introductorio del presente capítulo, se limitará el desarrollo del tema a los problemas de optimización bajo certidumbre.

Adoptado el criterio de decisión, el paso que sigue es expresar las restricciones y el objetivo como funciones matemáticas de las variables de decisión. Se obtiene así un modelo matemático del problema de optimización. El modelo matemático resultante tendrá la siguiente estructura general:

$$\begin{array}{lll} optimizar\ f(x) & x \in R^n & \\ sujeto\ a: & & \\ & g_1(x) \le b_i & i = 1,...,u \\ & g_1(x) \ge b_i & i = 1,...,v \\ & g_1(x) = b_i & i = 1,...,m \end{array}$$

Donde f(x) y $g_i(x)$ $\forall i = 1,...,m$ son funciones definidas en el espacio Real n-dimensional R^n. f(x) es la *función objetivo* del modelo de optimización y $g_i(x)$ ($\le$, $\ge$, $=$) b_i $\forall i=1,...,m$ son las *restricciones* de desigualdad e igualdad del modelo.

Cuando la función objetivo y todas las restricciones son lineales esta estructura matemática se denomina modelo de Programación Lineal, (Linear Programming, LP), mientras que si al menos una de las funciones descriptas es no lineal, se denomina modelo de Programación No Lineal (Nonlinear Programmig, NLP). En el caso en que las variables del modelo son enteras, se denomina modelo de Programación Lineal Entero, (Integer Linear Programming, ILP) y modelo de Programación No Lineal Entero (Integer Nonlinear Programmig, INLP) respectivamente.

Puede ocurrir que el modelo posea variables continuas y enteras mezcladas, en este caso se denomina modelo de Programación Lineal Entero Mixto, (Mixed Integer Linear Programming, MILP) y modelo de Programación No Lineal Entero Mixto (Mixed Integer Nonlinear Programmig, MINLP) respectivamente. Es este el modelo asociado al problema general de diseño, según se vio en el capítulo II, con múltiples objetivos y generalmente con incertidumbre en la información. Cada tipo de modelo de programación matemática tiene asociado elementos teóricos y algoritmos particulares. Este capítulo presenta una breve descripción conceptual de elementos teóricos y algoritmos de los modelos LP y NLP exclusivamente.

Algunos conceptos asociados a los modelos matemáticos de optimización son los siguientes:

Región Factible: es el conjunto RF de puntos $x \in R^n$ que satisfacen simultáneamente todas las restricciones del Programa Matemático.

Punto Factible*:* es todo punto $x \in$ RF.

Punto No Factible*:* es todo punto $x \notin$ RF.

Solución Óptima del Modelo*:* es todo punto factible x^* que brinda el mejor valor de

ISBN: 950-42-0022-2 - ©1999

la función objetivo. x^* es un punto óptimo y $f(x^*)$ el valor óptimo de la función objetivo. Es necesario destacar aquí que el concepto de óptimo está asociado al modelo y no al problema. Es decir, el óptimo corresponde al modelo y podrá aproximar más o menos bien al óptimo del problema dependiendo de cuan bien el modelo representa al problema de optimización en cuestión.

Con el propósito de explicar mejor los conceptos anteriores, se incluye el siguiente ejemplo.

XI.2.1 EJEMPLO: Planificación de Producción de una Refinería.

Una refinería de petróleo produce cuatro cortes intermedios de naftas, los cuales deben posteriormente ser mezclados para producir dos tipos comerciales de combustibles: común y extra. Cada corte posee un índice de performance que define la calidad del mismo, un máximo de disponibilidad y un costo unitario fijo. Los dos combustibles a producir tienen especificado un valor mínimo del índice de performance, un precio de venta conocido, y el mezclado tiene asociado un costo fijo conocido. Las obligaciones contractuales imponen una cantidad mínima a entregar de cada tipo de combustible, mientras que los excedentes de cada producto o de cortes intermedios pueden ser vendidos en el mercado libre a los precios estipulados por el mismo.

Las tablas siguientes muestran los valores de todos los parámetros del problema para un período de tiempo de un mes.

Tabla 1

Corte	Disponibilidad * b_i[bar/m]	Indica Perfor. a_i	Precio de venta $C1_i$**	Costo del corte $C2_i$**	Costo de mezclado $C3_i$**
1	3	73	32	27	1.5
2	3	80	37	31	1.5
3	5	92	42	36.5	2
4	6	98	59	42.5	2

* la disponibilidad está expresada en 10^5 barriles por mes.
** el precio de venta y los costos están expresados en $ por barriles.

ISBN: 950-42-0022-2 - 81999

Tabla 2

Producto	Mínimo a entregar s/contrato d_j[bar/m]*	Mínimo índice Perfor. f_j	Precio de venta s/contrato $v1_j$**	Precio de venta al mercado $V2_j$**
común	3	73	32	27
extra	3	80	37	31

* la disponibilidad está expresada en 10^5 barriles por mes.
** el precio de venta y los costos están expresados en $ por barriles.

XI.2.1.1 Modelado del Ejemplo

Variables de decisión: en el problema en consideración es posible decidir para cada corte intermedio la cantidad a vender directamente y la cantidad a mezclar para producir los combustibles común y extra. También es posible decidir para cada uno de los combustibles la cantidad a producir, la cantidad a vender por contrato y la cantidad a vender al mercado. Se identifican entonces las siguientes variables:
X_{ji}: cantidad del corte i a mezclar para producir combustible j [bar/mes]
Xv_i: cantidad del corte i a vender directamente [bar/mes]
Yc_j: cantidad del combustible j a vender por contrato [bar/mes]
Ym_j: cantidad del combustible j a vender al mercado [bar/mes]

Restricciones: Las restricciones a ser consideradas en este problema son las resultantes de los balances de materiales, las que aseguran el valor mínimo del índice de performance de los productos, las que aseguran los montos mínimos de ventas de los contratos y las que aseguran la satisfacción de las disponibilidades de cada tipo de corte intermedio a ser mezclado.

Balance de materiales para cada corte intermedio i:

$$\Sigma_j X_{ji} + Xv_i \leq b_i \qquad \forall\ i=1,...,4$$

Balance de materiales para cada combustible j a producir:

$$\Sigma_i X_{ji} = Yc_j + Ym_j \qquad \forall\ j=1,2$$

Valor mínimo del índice de performance de los productos:

$$\Sigma_i\ a_i X_{ji} \ \exists\ f_j (Yc_j + Ym_j) \qquad \forall j=1,2$$

Cantidad mínima de combustible a vender según contrato:

ISBN: 950-42-0022-2 - ©1999

$Yc_j \geq d_j \quad \forall\ j=1,2$

Restricción de no negatividad de las variables (un valor negativo de las variables que intervienen en este problema no tienen sentido físico):

$X_{ji} \geq 0 \quad \forall\ j=1,2 \quad \forall\ i=1,...,4$
$Yc_j \geq 0 \quad \forall\ j=1,2$
$Ym_j \geq 0 \quad \forall\ j=1,2$

Función objetivo: para este problema, se definirá como tal a una función que proporcione el Beneficio Neto = (Ventas) - (Costos). Las ventas están conformadas por la venta de cada uno de los combustibles más la venta directa de cortes. Los costos están constituidos por los costos de cada uno de los cortes más los costos de mezclado.

Análisis de la información disponible: en este problema se ha supuesto que todos los parámetros se conocen con certeza, luego el criterio de decisión a utilizar será maximizar la función Beneficio Neto.

La función objetivo expresada en términos de las variables de decisión resultará en:

Beneficio Neto = $\Sigma_j V1_j Yc_j + \Sigma_j V2_j Ym_j + \Sigma_i C1_i Xv_i - \Sigma_i C2_i (\Sigma_j X_{ji} + Xv_i) - \Sigma_i C3_i \Sigma_j X_{ji}$
Finalmente, el modelo matemático de optimización resultante será:

Maximizar $32Yc_1 + 37Yc_2 + 27Ym_1 + 31Ym_2 + 32\ Xv_1 + 37\ Xv_2 + 42\ Xv_3 + 59\ Xv_4 - 28{,}50\ X_{11} - 32{,}50\ X_{12} - 38{,}50\ X_{13} - 44{,}50\ X_{14} - 28{,}50\ X_{21} - 32{,}50\ X_{22} - 38{,}50\ X_{23} - 44{,}50\ X_{24} - 27Xv_1 - 31Xv_2 - 36{,}50\ Xv_3 - 42{,}50\ Xv_4$

Sujeto a:

$X_{11} + Xv_1 \leq 3$
$X_{12} + Xv_2 \leq 3$
$X_{13} + Xv_3 \leq 5$
$X_{14} + Xv_4 \leq 6$
$X_{21} + Xv_1 \leq 3$
$X_{22} + Xv_2 \leq 3$
$X_{23} + Xv_3 \leq 5$
$X_{24} + Xv_4 \leq 6$
$\Sigma_i X_{ji} = Yc_j + Ym_j \quad \forall\ j=1,2$
$3\ X_{11} + 80\ X_{12} + 92\ X_{13} + 98\ X_{14} \exists\ 73\ (Yc_1 + Ym_1)$
$73\ X_{21} + 80\ X_{22} + 92\ X_{23} + 98\ X_{24} \exists\ 73\ (Yc_2 + Ym_2)$

ISBN: 950-42-0022-2 - ©1999

$$Yc_1 \geq 3$$
$$Yc_2 \geq 3$$
$$X_{ji} \geq 0 \qquad \forall\ j=1,2 \qquad \forall\ i=1,...,4$$
$$Ym_j \geq 0 \qquad \forall\ j=1,2$$

El modelo de optimización resultante es un modelo de Programación Lineal. La etapa que sigue es resolver el LP utilizando algún algoritmo apropiado y luego validar el modelo.

En las siguientes secciones se presenta un breve estudio de los modelos de programación matemática lineal (LP) y no lineal (NLP) y los principales algoritmos diseñados para resolver dichos modelos.

XI.3 TEORÍA Y ALGORITMOS DE OPTIMIZACIÓN

La resolución de los modelos de programación matemática fue abordada inicialmente desde una óptica matemática rigurosa, utilizando principalmente elementos de álgebra y análisis matemático. Surgieron así importantes principios, los cuales establecen las condiciones necesarias y suficientes que deben verificarse para poder asegurar que un punto es óptimo de un programa matemático. Estos postulados teóricos no fueron útiles como tales al momento de resolver los programas matemáticos que modelan problemas reales de optimización.

Ante la necesidad de disponer de una herramienta para resolver los problemas reales y con el advenimiento de la computadora, comenzaron a desarrollarse los algoritmos de optimización. Estos son derivados inicialmente más de la heurística que de la matemática. Los algoritmos son hoy poderosas herramientas para la resolución de los modelos de optimización de problemas reales.

Es de destacar que, si bien los elementos teóricos no constituyen una herramienta práctica para la resolución de los modelos de optimización, definen un marco conceptual importante para el desarrollo de los algoritmos de optimización y su aplicación, como así también para el modelado de un problema de optimización y su resolución. Sería casi imposible modelar y resolver un problema de optimización real sin conocer ciertos conceptos teóricos, por más que se disponga de un poderoso software de optimización.

XI.3.1 Programación Lineal

El modelo LP es extensamente utilizado en casi todas las áreas del conocimiento. La relación lineal entre variables le confiere la particularidad de ser un modelo fácil de generar y simple de resolver y analizar. Esto permite automatizar el proceso de generación del modelo, por lo cual es posible generar grandes modelos LP. Publicaciones recientes han reportado trabajos con modelos LP de más de cien mil variables.

A efectos de visualizar la simplicidad de un modelo LP, se realiza una representación gráfica con la cual se analiza el modelo y su solución. En cuanto a la

ISBN: 950-42-0022-2 - ©1999

resolución de un LP, el algoritmo convencional ampliamente utilizado es el método SIMPLEX diseñado por Dantzig en la década del cincuenta. En la década del ochenta se ha publicitado el método de Karmarkar como un nuevo y poderoso algoritmo para resolver LP, el cual reemplazaría al SIMPLEX. Dicho algoritmo por el momento permanece en el olvido, siendo universalmente usado el SIMPLEX.

XI.3.1.1 Representación y Solución Gráfica del LP

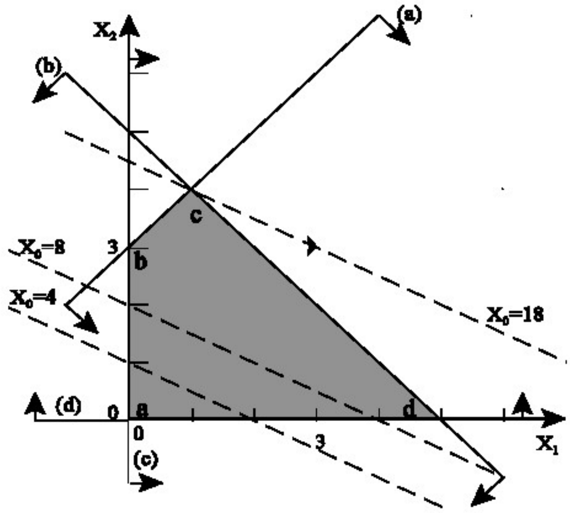
Desde el punto de vista didáctico, resulta conveniente realizar un análisis gráfico del procedimiento de resolución de un LP. Con esta finalidad se considera el siguiente ejemplo:

$$\max \quad x_0 = 2x_1 + 4x_2$$

s.a.:

$$-x_1 + x_2 \leq 3 \qquad (a)$$
$$x_1 + x_2 \leq 5 \qquad (b)$$
$$x_1 \geq 0,\ x_2 \geq 0 \qquad (c), (d)$$

La representación gráfica se muestra en la siguiente figura:



El plano asociado a cada restricción, divide al espacio en un semiespacio factible (conjunto de punto que verifican la restricción) y un semiespacio no factible (conjunto de puntos que no verifican la restricción). La intersección de todos los semiespacios factibles (si es no vacío y acotado) define un poliedro (el conjunto de puntos que satisface simultáneamente todas las restricciones), el cual representa la región factible del LP. Dicho poliedro se representa sombreado en el gráfico.

De todos los puntos del poliedro, habrá al menos uno que asigna a la función

ISBN: 950-42-0022-2 - ©1999

objetivo del LP el máximo valor. Dicho punto es la solución óptima del LP. Para determinar este punto se procede del siguiente modo:

- Se asigna al plano que define la función objetivo un valor arbitrario, en el ejemplo se asignó $x_0 = 4$, esto define en R^2 una recta. Se puede ver en el gráfico que existen aún infinitos puntos factibles que asignan a la función objetivo dicho valor.
- Asignando luego un segundo valor arbitrario mayor que el anterior, en el ejemplo se asignó x0 = 8, se obtiene una segunda recta, paralela a la anterior. Del gráfico se observa que existen infinitos puntos factibles que asignan a la función objetivo este mayor valor. Luego, es fácil deducir que el máximo del problema corresponderá a una recta paralela a las anteriores, desplazada lo máximo posible en el sentido de crecimiento de la función objetivo, la cual incluya al menos un punto factible. Esta es la recta x0=18 en el gráfico, la cual incluye el punto factible x*=(x1;x2)=(1;4). Este es el punto factible que asigna el mejor (máximo) valor a la función objetivo, por lo tanto es el punto óptimo, y x0=18 es el valor óptimo (máximo) de la función objetivo.

Si se analizan distintas funciones objetivos (lineales) manteniendo la región factible, es fácil ver que el óptimo siempre corresponde a un vértice del poliedro. Esto permite inferir la siguiente conclusión: *el óptimo de un LP cuya región factible es no vacía y acotada estará siempre asociado a un punto extremo (vértice del poliedro) de la misma.*

Algunas particularidades que pueden presentarse en un modelo LP son las siguientes:

1. Si la región factible es vacía esto significa que el LP no tiene solución. *Se dice que el LP es infactible.*

- Si la región factible es no acotada en la dirección en la cual la función objetivo es optimizada, lo cual ocurriría si en el ejemplo anterior se elimina la restricción (b), es fácil ver que el máximo valor de la función objetivo es infinito. *En este caso se dice que el LP es no acotado*.
- Si la recta correspondiente a la función objetivo es paralela a la recta correspondiente a una restricción sobre la que se encuentra el extremo óptimo, en este caso todos los puntos extremos localizados sobre dicha restricción serán óptimos. Asimismo serán también óptimos todos los puntos factibles localizados sobre dicha restricción. *En este caso*, *se dice que el LP tiene soluciones óptimas alternativas.* En el ejemplo anterior, esto ocurre si la función objetivo a maximizar fuese $x_0 = 2x_1+2x_2$, en este caso los puntos extremos x_c y x_d (vértices c y d) asignan a la función objetivo el valor máximo $x_0=10$. Del mismo modo, todo punto $x=\theta\ x_c+(1-\theta)\ x_d$ □ θ0[0,1] será también un óptimo local alternativo.

Todo LP puede ser representado por la siguiente estructura general, conocida como *formato estándar*:

$$\max \quad x_0 = c^T x \qquad x \in R^n$$

ISBN: 950-42-0022-2 - 81999

s.a:

$$A\,x = b$$
$$x \geq 0$$

Donde A(mxn) es una matriz de coeficientes, b(mx1) es el vector de términos independientes de las restricciones para el cual se verifica $b_i \geq 0 \quad \forall i=1,...,m$, c(nx1) es el vector de coeficientes de la función objetivo, x(nx1) es el vector de variables y m<n.

Para el ejemplo anterior, el formato estándar es el siguiente:

$$\max \quad x_0 = 2x_1 + 4x_2 + 0x_3 + 0x_4$$

s.a.:

$$\begin{aligned} -x_1 + x_2 + x_3 \qquad &= 3 \\ x_1 + x_2 + \qquad x_4 &= 5 \\ x_1,\ x_2,\ x_3,\ x_4 &\geq 0 \end{aligned}$$

Donde x_3, x_4 son variables que se agregaron para transformar las inecuaciones en ecuaciones. Se denominan *variables de holgura*.

Dado un LP en el formato estandar, si el Rango(A,b)=m es posible deducir:

1 Dado que posee m restricciones de igualdad linealmente independientes, el sistema de ecuaciones Ax=b tiene n-m variables independientes (grados de libertad) que serán identificadas con el vector $x_N 0 R^{n-m}$ y m variables dependientes que se identificarán con el vector $x_B 0 R^m$. Luego $x^T = (x_B^T, x_N^T)$.

2 Desde el punto de vista geométrico (recordar análisis gráfico), dado que las restricciones de un LP son hiperplanos definidos en R^n, es fácil deducir que la región factible resultante de la intersección de dichos hiperplanos si existe y es acotada será un hiperpoliedro en R^n. Luego, dado que la función objetivo es también un hiperplano en R^n, siempre exitirá un vértice del hiperpoliedro factible que maximiza y otro que minimiza el hiperplano de la función objetivo. En otros términos, el óptimo de un LP cuya región factible es no vacía y acotada estará siempre asociado a un punto extremo (vértice del hiperpoliedro) de la misma.

3 Desde el punto de vista algebraico se puede demostrar que un punto de la región factible tiene al menos n-m variables iguales a 0 y todas las restantes positivas (condición de no negatividad de las variables, $x \geq 0$) si y solo si es un punto extremo de la misma.

Como ejemplo se analizan algunos puntos (soluciones factibles) del ejemplo anterior, para el cual (n-m)=2.

$x_a=(0;0;3;5)$ punto extremo

ISBN: 950-42-0022-2 - 81999

x_b=(0;3;0;2) punto extremo
x_c=(1;4;0;0) punto extremo
x_d=(5;0;8;0) punto extremo
x_e=(0;2;1;3) punto de frontera
x_f=(2;2;3;1) punto interior

Es fácil ver que los puntos extremos son los únicos que tienen al menos dos variables iguales a cero.

4 Siendo finito el número de restricciones de un PL, el número de vértices del hiperpoliedro factible será finito.

Luego, de estas cuatro consideraciones es posible deducir:

- La región factible de un LP si es no vacía y acotada tendrá un número finito de puntos extremos (vértices).

- El óptimo de un LP siempre está asociado a un punto extremo de la región factible.

- Dada la región factible Ax=b, x≥0 del LP en su forma estandar, si se verifica que cualesquiera m columnas de la matriz A son linealmente independiente, es posible representar la matriz de coeficientes como A=(B,N) siendo B una matriz (mxm) no singular y N una matriz (mx(n-m)). B se denomina matriz *Básica* y N se denomina matriz *No Básica.* Las variables asociadas a cada columna de la matriz B son las variables dependientes x_B y se denominan *variables básicas*, mientras que las variables asociadas a cada columna de la matriz N son las variables independientes x_N y se denominan *variables no básicas*. Luego, la región factible del LP en forma estandar queda expresada como: $(B,N)(x_B,x_N)^T=b$, $(x_B,x_N)\geq 0$, o bien $Bx_B+Nx_N=b$, $x_N\geq 0$, $x_B\geq 0$. Siendo la matriz B no singular, la B^{-1} existe, luego premultiplicando la ecuación anterior por B^{-1} se tiene: $x_B=B^{-1}b-B^{-1}Nx_N$.

- Todo punto extremo de la región factible puede ser determinado algebraicamente asignando arbitrariamente valor cero a n-m variables (independientes), luego calculando el valor de las restantes variables (dependientes) y descartando finalmente toda solución que no cumpla la condición de no negatividad ($x_B\geq 0$). Este punto se denomina *solución básica factible.* Luego, la solución obtenida asignando $x_N=0$ y calculando $x_B=B^{-1}b$, si $x_B\geq 0$ será una *solución básica factible del LP en el formato estándar*. En particular dicha solución se denomina *no degenerada* si $x_B>0$.

En base a estas consideraciones se ha desarrollado un método de búsqueda el óptimo de LP denominado SIMPLEX, el cual es muy eficiente y robusto (aplicable a todo LP). Este método se basa en dos condiciones fundamentales:

ISBN: 950-42-0022-2 - ©1999

Condición de factibilidad: esta condición asegura que partiendo de una solución básica factible inicial sólo se analicen nuevas soluciones básicas factibles.

Condición de optimicidad: esta condición asegura que sólo se evaluarán soluciones básicas factibles no inferiores. Esto es, soluciones básicas factibles que asignan a la función objetivo un valor no inferior al asignado por la solución actual.

Convergencia del método: Si la región factible de un LP es no vacía y acotada tiene un número finito de puntos extremos (vértices), luego, dado que el método sólo evalúa vértices no inferiores, necesariamente (en ausencia de degeneración) el método alcanzará un óptimo en un número finito de iteraciones.

XI.3.1.2 El algoritmo SIMPLEX

Etapa Inicial

Se puede definir a la región factible del LP en la forma estandar como el conjunto de puntos $S=\{x \,/\, Ax=b,\ x\geq 0\}$. Sea $x \in S$ una *solución básica factible no degenerada inicial* del LP. Luego, el vector no básico asociado a dicha solución será $x_N=0$ y el vector básico asociado será $x_B=B^{-1}b-B^{-1}Nx_N=B^{-1}b>0$, siendo la matriz B la base asociada a dicha solución.

Etapa principal

Dada una solución básica factible $x \in S$, si dicha solución es no óptima, el método genera una nueva solución básica factible eligiendo una variable no básica x_{Nj} (con valor actual cero) y la convierte en variable básica (incrementando su valor), en reemplazo de una variable básica x_{Bi} (con valor actual mayor que cero) la que pasa a no básica (tomando valor cero).

La variable x_{Nj} que pasa a x_{Bi} se denomina *variable ingresante* a la base .

La variable x_{Bi} que pasa a x_{Nj} se denomina *variable saliente* de la base.

Esta operación es realizada en base a la *condición de optimicidad* y a la *condición de factibilidad* definidas en forma conceptual anteriormente, cuya operatoria se describe a continuación.

Condición de optimicidad

El valor actual de la función objetivo vendrá dado por $x_0 = c_Bx_B+c_Nx_N = c_B(B^{-1}b-B^{-1}Nx_N) + c_Nx_N = c_BB^{-1}b-(c_BB^{-1}N-c_N)x_N$ o bien $x_0 = c_BB^{-1}b-\Sigma_{j \in J}(c_BB^{-1}a_j-c_j)x_j$ siendo J el conjunto de subíndices de las variables No básicas y a_j la columna de A asociada a la variable x_j. Un análisis del valor de $(c_BB^{-1}a_j-c_j)$ permite deducir (para problemas de maximización):

- Si $(c_BB^{-1}a_j-c_j) \geq 0 \ \ \forall\ j \in J \rightarrow$ cualquier x_j No básica que ingrese a la base no mejorará el valor de la función objetivo, luego *el punto actual es óptimo* (condición de parada del algoritmo)
- Si $(c_BB^{-1}a_j-c_j) <0$ para algún $j \in J$, x_j será candidata a ingresar a la base puesto que

ISBN: 950-42-0022-2 - ©1999

al aumentar el valor de x_j puede mejorar el valor de la función objetivo. De todas las x_j / $j \in J$ con valor $(c_B B^{-1} a_j - c_j) < 0$ se elige como *variable ingresante* aquella cuyo valor de $(c_B B^{-1} a_j - c_j)$ sea menor.

Si se elige como una solución básica factible inicial el vértice x_a del ejemplo anterior, se tiene que $x_B=(x_3,x_4)$ y $x_N=(x_1,x_2)$. Luego,

$$B = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \quad N = \begin{vmatrix} -1 & 1 \\ 1 & 1 \end{vmatrix}$$

$c_B=(0;0)$, $c_N=(2;4)$, $b=(3;5)$, $x_N=(x_1,x_2)=(0;0)$
Por lo tanto, $x_B=B^{-1}b=(3;5)$ y $x_0=c_B x_B + c_N x_N = 0$.

Determinación de la *variable ingresante* en el ejemplo anterior:
Se calcula $(c_B B^{-1} a_j - c_j)$ para j0{1,2}
Para j=1 $(c_B B^{-1} a_1 - c_1) = -2 \rightarrow x_1$ es candidata a ingresar a la base
Para j=2 $(c_B B^{-1} a_2 - c_2) = -4 \rightarrow x_2$ es candidata a ingresar a la base
No verifica la condición de óptimo.
Ambas variables son candidatas a ingresar a la base. Se elige la que tiene el potencial de producir la mayor mejora, esto es, la que posee el menor valor de $(c_B B^{-1} a_j - c_j)$. Por lo tanto se elige a x_2 como la variable ingresante.

Condición de factibilidad
Selecciona la *variable saliente* en base a satisfacer:
(1) La condición de no negatividad, $x \geq 0$.
(2) La condición de solución básica $\rightarrow$ la variable saliente debe tomar valor cero.

Luego, siendo x_j / $j \in J$ la variable ingresante (su valor aumentará desde el actual cero), con lo cual el nuevo valor del vector de variables básicas vendrá dado por $x_B=B^{-1}b-B^{-1}a_j x_j$. Las restantes variables no básicas se mantendrán con valor cero. Sea x_{Bi} el i-ésimo elemento del vector de variables básicas x_B (mx1). Luego, la condición (1) establece que $x_{Bi} = (B^{-1}b)_i-(B^{-1}a_j)_i\, x_j \geq 0 \;\; \forall \;$ i=1,...,m.
Por otra parte, la condición (2) exige que la variable saliente x_{Br} debe tomar valor cero.
Estas condiciones se satisfacen eligiendo el valor de x_j (variable ingresante) como:
$x_j=\min_i \{(B^{-1}b)_i /(B^{-1}a_j)_i \quad \forall (B^{-1}a_j)_i>0\}=(B^{-1}b)_r /(B^{-1}a_j)_r$
Luego, la variable no básica x_j ingresa a la base con valor $(B^{-1}b)_r /(B^{-1}a_j)_r$ y la variable básica x_{Br} sale de la base y toma valor cero (pasa a no básica).

Se habrá generado así una nueva solución básica factible no inferior a la anterior. El procedimiento vuelve a la *condición de optimicidad*.

Determinación de la *variable saliente* en el ejemplo anterior:
Se calcula $\min_i \{(B^{-1}b)_i /(B^{-1}a_j)_i \quad \forall (B^{-1}a_j)_i >0,\}$ para i=1,2 (i-ésimo elemento de x_B) y j=2 (variable ingresante)
Para i=1 $(B^{-1}b)_1 /(B^{-1}a_2)_1 = 3$
Para i=2 $(B^{-1}b)_2 /(B^{-1}a_2)_2 = 5$

ISBN: 950-42-0022-2 - 81999

Luego $min_i\{3;5\}=3$ que corresponde a $j=3 \rightarrow x_3$ es la *variable saliente*
x_2 ingresa a la base con valor $x_2=min_i\{3;5\}=3$
Se genera así una nueva solución básica factible x para la cual, $x_B=(x_2,x_4)$ y $x_N=(x_1,x_3)$

$$B=\begin{vmatrix}1 & 0\\ 1 & 1\end{vmatrix} \quad N=\begin{vmatrix}-1 & 1\\ 1 & 0\end{vmatrix}$$

$c_B=(4;0)$, $c_N=(2;0)$, $b=(3;5)$, $x_N=(x_1,x_3)=(0;0)$
Por lo tanto, $x_B=(x_2,x_4)=B^{-1}b=(3;2)$ y $x_0=c_Bx_B+c_Nx_N=12$

Se ha generado así una nueva solución básica factible que mejoró el valor de la función objetivo. x_0 incrementó su valor de 0 a 12. Puede verse en el gráfico que el algoritmo pasó del punto extremo x_a al punto extremo x_b. Repitiendo este procedi-miento, desde el vértice x_b pasará al vértice x_c el cual es el óptimo.

Es necesario destacar que aquí se ha descripto la estrategia básica del algoritmo. Existen varias versiones del algoritmo desarrolladas en base a esta estrategia las cuales involucran desde estructuras en forma de tabla de cálculos hasta estructuras vectoriales compactas que explotan la relación de compromiso entre, complejidad del algoritmo, número de operaciones, espacio de operación y almacenamiento de datos y problemas de cálculos numéricos. Esto es muy importante en los LP de gran tamaño, en los cuales la matriz A tiene generalmente un elevado porcentaje de elementos nulos.

El Algoritmo Simplex es el único método masivamente difundido y utilizado en todo el mundo para la resolución de LP. Actualmente hay muchos softwares comerciales que han implementado el mismo. A modo de ejemplo podemos mencionar LINDO, XA, ZOOM y WHAT'S BEST!. También para problemas de menor tamaño es posible utilizar el algoritmo Simplex implementado en planillas de cálculo tales como QuatroPro y Excel.

XI.3.2 Programación No Lineal

A diferencia de lo que ocurre con los LP en los cuales el óptimo siempre corresponde a un vértice de la región factible, lo cual implica un número finito de puntos de búsqueda, en los NLP cualquiera de los infinitos puntos de la región factible puede ser un óptimo. Además, es común la presencia de óptimos locales. Estas características dificultan considerablemente la estrategia de resolución de los NLP. Según se indicó al comienzo del capítulo, inicialmente los NLP fueron estudiados desde la perspectiva de la matemática, dando origen a postulados teóricos que establecen condiciones necesarias y suficientes que verifica un punto óptimo (local o global) del problema. Estos postulados teóricos conocidos como la *Teoría Clásica de los NLP*, si bien son elementos conceptuales muy importantes, por su complejidad, no constituyen herramientas prácticas para resolver los NLP. Por esta razón, con el avance de la informática, surgieron *algoritmos* de optimización, desarrollados en base a razonamiento principalmente heurístico, que se constituyeron en herramientas eficientes y robustas para la resolución de NLP. En las secciones

ISBN: 950-42-0022-2 - ©1999

siguientes, se presenta una descripción conceptual de los principales postulados de la teoría clásica y posteriormente de los algoritmos básicos de optimización.

En el Apéndice de fin de capítulo se presentan algunas definiciones que se utilizan en el desarrollo de las siguientes secciones.

XI.3.2.1 Teoría Clásica de la Programación No Lineal

A efectos de presentar los resultados teóricos, los NLP son clasificados en: NLP No condicionados (sin restricciones), NLP condicionados por igualdades y NLP condicionados por desigualdades. Los postulados teóricos son presentados a través del enunciado de teoremas sin demostración. Los resultados de los teoremas son aplicados a ejemplos académicos.

XI.3.2.1.1 Programas Matemáticos no Condicionados

Este tipo de NLP es representado a través de la siguiente estructura general: opt f(x) s.a: $\{x \in R^n\}$.

Teorema XI.1: *Condición Necesaria de Optimo Local No Condicionado*
Sea $f:R^n \rightarrow R^1$ continua y diferenciable en todo $x \in R^n$. Luego si x^* es un punto (finito) donde la función alcanza un óptimo local, se verifica $\nabla f(x^*)=0$. En particular, se dice que x^* es un óptimo local no condicionado.

Teorema XI.2: *Condición Suficiente de Optimo Local No Condicionado*
Sea $f:R^n \rightarrow R^1$ continua y diferenciable dos veces en todo $x \in R^n$. Si todas las primeras derivadas de f(x) se anulan en x^*, y si la matriz Hessiana evaluada en x^* es negativa (*positiva*) definida, luego x^* es un máximo (*mínimo*) local de f.
Extensión: Si x^* es un máximo (*mínimo*) local no condicionado de f(x), luego la matriz Hessiana evaluada en x^* es negativa (*positiva*) definida o semidefinida.

Teorema XI.3: *Condición Suficiente de Optimo Global No Condicionado*
Sea $f:R^n \rightarrow R^1$ continua y diferenciable dos veces y convexa (*cóncava*) $\forall x \in R^n$. Luego, todo mínimo (*máximo*) local de f(x) es global.

Ejemplo: Dado el NLP:

$$Opt\ f(x) = x_1^4 - 4x_1^3 + 4x_1^2 + x_2^4 - x_2^3 - 6x_2^2$$

encontrar máximos y mínimos.

Resolución: Por condición necesaria de óptimo local no condicionado (Teorema XI.1) se tiene:

ISBN: 950-42-0022-2 - ©1999

$$\nabla f(x^*) = \begin{bmatrix} \dfrac{\partial f(x^*)}{\partial x_1} \\ \dfrac{\partial f(x^*)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1^3 - 12x_1^2 + 8x_1 \\ 4x_2^3 - 3x_2^2 - 12x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Se obtiene el siguiente sistema de ecuaciones:

$$4x_1\left(x_1^2 - 3x_1 + 2\right) = 0 \qquad <1>$$
$$x_2\left(4x_2^2 - 3x_2^2 - 12\right) = 0 \qquad <2>$$

Resolviendo el sistema de ecuaciones resultante se obtienen las siguientes soluciones: x_1^*= (0; 0); x_2^*= (0; 2,147) ; x_3^*= (0; -1,397) ; x_4^*= (2; 0); x_5^*= (2; 2,147); x_6^*= (2; -1,397); x_7^*= (1; 0) ; x_8^*= (1; 2,147) ; x_9^*= (1; -1,397)
Luego, existen nueve puntos del dominio de la función que verifican las condiciones necesarias de óptimo local no condicionado.
A continuación se analiza si verifican la condición suficiente de óptimo local no condicionado (Teorema XI.2), la cual requiere evaluar, en cada uno de los puntos anteriores, la matriz Hessiana definida en el Apéndice de fin de capítulo:

$$Hf(x^*) = \begin{bmatrix} \dfrac{\partial^2 f(x^*)}{\partial x_1 x_1} & \dfrac{\partial^2 f(x^*)}{\partial x_1 x_2} \\ \dfrac{\partial^2 f(x^*)}{\partial x_2 x_1} & \dfrac{\partial^2 f(x^*)}{\partial x_2 x_2} \end{bmatrix} = \begin{bmatrix} 12x_1^2 - 24x_1 + 8 & 0 \\ 0 & 12x_2^2 - 6x_2 - 12 \end{bmatrix}$$

$$H\big|_{x_1^*} = \begin{bmatrix} 8 & 0 \\ 0 & -12 \end{bmatrix} \qquad h_{11} > 0 \quad |H| = -96 < 0 \rightarrow \textit{no verifica}$$

Luego, x_1^*= (0;0) no verifica la condición suficiente de óptimo local no condicionado.

$$H\big|_{x_2^*} = \begin{bmatrix} 8 & 0 \\ 0 & 30{,}43 \end{bmatrix} \qquad h_{11} > 0 \quad |H| = 243{,}4 > 0 \rightarrow H \textit{ es positiva definida}$$

ISBN: 950-42-0022-2 - ©1999

Luego, x_2^*=(0; 2,147) es un mínimo local.

$$H\big|_{x_3^*} = \begin{bmatrix} 8 & 0 \\ 0 & 19{,}80 \end{bmatrix} \quad h_{11} > 0 \quad |H| = 158{,}4 > 0 \rightarrow H \ es \ positiva \ definida$$

Luego, x_3^* = (0; -1,397) es un mínimo local.

$$H\big|_{x_4^*} = \begin{bmatrix} 8 & 0 \\ 0 & 0 \end{bmatrix} \quad h_{11} > 0 \quad |H| = 0 \rightarrow H \ es \ positiva \ semidefinida$$

Luego, x_4^*=(2; 0) no verifica condición suficiente. No obstante, por extensión del Teorema XI.2, es posible que x_4^* sea un óptimo local.

$$H\big|_{x_5^*} = \begin{bmatrix} 8 & 0 \\ 0 & 30{,}43 \end{bmatrix} \quad h_{11} > 0 \quad |H| = 243{,}4 > 0 \rightarrow H \ es \ positiva \ definida$$

Luego x_5^*= (2; 2,147) es un mínimo local

$$H\big|_{x_6^*} = \begin{bmatrix} 8 & 0 \\ 0 & 19{,}80 \end{bmatrix} \quad h_{11} > 0 \quad |H| = 158{,}0 > 0 \rightarrow H \ es \ positiva \ definida$$

Luego, x_6^*= (2; -1,397) es un mínimo local

$$H\big|_{x_7^*} = \begin{bmatrix} -4 & 0 \\ 0 & -12 \end{bmatrix} \quad h_{11} < 0 \quad |H| = 48 > 0 \rightarrow H \ es \ negativa \ definida$$

Luego, x_7^* = (1; 0) es un máximo local

$$H\big|_{x_8^*} = \begin{bmatrix} -4 & 0 \\ 0 & 30{,}43 \end{bmatrix} \quad h_{11} < 0 \quad |H| = -122 < 0 \rightarrow H \ no \ verifica \ definición$$

Luego x_8^* = (1; 2,147) no verifica condición suficiente

$$H\big|_{x_9^*} = \begin{bmatrix} -4 & 0 \\ 0 & 19{,}80 \end{bmatrix} \quad h_{11} < 0 \quad |H| = -79{,}2 < 0 \rightarrow H \ no \ verifica \ definición$$

ISBN: 950-42-0022-2 - ©1999

Luego $x_9^* = (1; -1,397)$ no verifica condición suficiente.

Como la función no es cóncava, ni convexa (Apéndice de final de capítulo) no verifica la condición suficiente de óptimo global no condicionado, Teorema XI.3.

XI.3.2.1.2 Programas Matemáticos Condicionados por Igualdades.

Este tipo de NLP es representado a través de la siguiente estructura general: opt f(x) s.a:$\{x \in R^n,\ g_i(x)=b_i \quad \forall\ i=1,..,m\}$ siendo $f:R^n \rightarrow R^1$ y $g_i:R^n \rightarrow R^1$ $\forall\ i=1,..,m$ donde m<n.

Sea $x^* \in R^n$ un óptimo local de este PNL, se define a la matriz (mxm) Jacobiana $J(x^*)$ como:

$$J(x^*) = \begin{bmatrix} \dfrac{\partial g_1(x)}{\partial x_1} & \cdots & \dfrac{\partial g_1(x^*)}{\partial x_m} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial g_m(x^*)}{\partial x_1} & \cdots & \dfrac{\partial g_m(x^*)}{\partial x_m} \end{bmatrix}$$

Teorema XI.4: *Condición necesaria de óptimo local restringido por igualdades.*

Sea $x^* \in R^n$ un óptimo local de f(x) sujeto a las restricciones de igualdad $g_i(x^*)=b_i$, $\forall$ i=1,...,m donde m<n. Si es posible elegir m variables tal que la matriz Jacobiana $J(x^*)$ sea no singular, $\exists$ un único conjunto de escalares $\lambda_1^*,...,\lambda_m^*$ satisfaciendo: $\nabla f(x^*) - \Sigma_{\forall i}\ \lambda_i^*\ \nabla g_i(x^*) = 0$.

La función Lagrangiana: Este sistema de ecuaciones puede obtenerse como condición necesaria de óptimo local del NLP no condicionado:

$$\text{opt}\quad L(x,\lambda) = f(x^*) - \Sigma_{\forall i}\ \lambda_i^*\,(gi(x^*)-b_i)$$

Sea (x^*, λ^*) óptimo local de $L(x,\lambda)$, luego debe verificar la condición necesaria de óptimo local no condicionado (Teorema XI.1) $\Lambda L(x^*,\lambda^*)=0$. Esto es:

$$\nabla_x L(x^*,\lambda^*) = \nabla f(x^*) - \Sigma_{\forall i}\ \lambda_i^*\ \nabla gi(x^*) = 0$$

$$\partial L(x^*,\lambda^*)/\partial\lambda_i = (gi(x^*) - b_i) = 0 \qquad \forall i=1,...,m$$

donde $\nabla_x L(x^*,\lambda^*)$ es un vector que contiene las derivadas parciales de $L(x,\lambda)$ respecto a x_j $\forall j=1,\Upsilon,n$.

Luego se puede decir que este problema irrestricto es equivalente al problema anterior en cuanto a que toda solución óptima de ambos problemas satisface el mismo conjunto de ecuaciones. $L(x,\lambda)$ se denomina función Lagrangiana y los λ_i $\forall$ i=1,...,m se denominan multiplicadores de Lagrange.

Teorema XI.5: *Condición suficiente de óptimo local restringido por igualdades.*

Sea $f(x):R^n \rightarrow R^1$ y $g_i(x):R^n \rightarrow R^1$ $\forall i=1,...,m$ continua y diferenciable dos veces en todo $x \in N_\delta(x^*)$ (entorno del punto x^* de radio δ). Sean $x^* \in R^n$ y $\lambda^* \in R^m / \nabla L(x^*,\lambda^*)=0$.

ISBN: 950-42-0022-2 - 81999

Si $\exists$ z $\in E^n$ no nulo cumpliendo $z^T \nabla g_i(x^*)=0$ $\wedge$ i=1,...,m, el cual satisface $z^T H_x L(x^*,\lambda^*) z > 0 \rightarrow x^*$ es un mínimo local estricto del problema: opt f(x) s.a:{$x \in R^n$, $g_i(x)=b_i$ $\forall$i=1,..,m}. Si $\exists$ $z^T H_x L(x^*,\lambda^*) z<0 \rightarrow x^*$ es un máximo local estricto del problema: opt f(x) s.a:{$x \in R^n$, $g_i(x)=b_i$ $\forall$i=1,..,m}.

Ejemplo: Dado el NLP:

$$
\begin{aligned}
&Opt\ f(x) = (5 - x_1)^2 + x_2^2 \\
&s.a; \qquad (1 - x_1)^2 + x_2^2 = 9
\end{aligned}
$$

se buscan máximos y mínimos.

Resolución: Por condición necesaria de óptimo local condicionado por igualdades (Teorema XI.4), derivada a través del problema equivalente, se tiene:

$$
\begin{aligned}
L(x,\lambda) &= (5 - x_1)^2 + x_2^2 - \lambda\left[(1 - x_1)^2 + x_2^2 - 9\right] \\
\nabla L(x,\lambda) = 0 &\Rightarrow -2(5 - x_1) + 2\lambda(1 - x_1) = 0 \\
&\qquad 2x_2 - 2\lambda x_2 = 0 \\
&\qquad (1 - x_1)^2 + x_2^2 - 9 = 0
\end{aligned}
$$

Resolviendo el sistema de ecuaciones resultante se obtienen las siguientes soluciones:
$x_1^* = (4; 0; -1/3)$; $x_2^* = (-2; 0; 7/3)$

Luego, existen dos puntos que verifican las condiciones necesarias de óptimo local restringido por igualdades.

A continuación se analiza si dichos puntos verifican la condición suficiente de óptimo local restringido por igualdades (Teorema XI.5):

Debe $\exists$ z $\in R^2$ no nulo / z^T $\nabla g(x^*)=0 \rightarrow (z_1; z_2)\begin{bmatrix} -2(1 - x_1) \\ 2x_2 \end{bmatrix} = 0$

$$
[z_1; z_2] H_x L(x^*) \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = [z_1; z_2] \begin{bmatrix} \dfrac{\partial^2 L(x^*)}{\partial x_1 x_1} & \dfrac{\partial^2 L(x^*)}{\partial x_1 x_2} \\ \dfrac{\partial^2 L(x^*)}{\partial x_2 x_1} & \dfrac{\partial^2 L(x^*)}{\partial x_2 x_2} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = [z_1; z_2] \begin{bmatrix} 2 - 2\lambda & 0 \\ 0 & 2 - 2\lambda \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 0
$$

Para $x_1^* = (4; 0; -1/3)$

ISBN: 950-42-0022-2 - ©1999

$$(z_1;z_2)\begin{bmatrix}6\\0\end{bmatrix}=0 \Rightarrow \textit{se verifica}\,\forall\, z\, \textit{no nulo} / z_1=0 \ y \ z_2 \neq 0$$

$$[z_1;z_2]H_xL(x^*)\begin{bmatrix}z_1\\z_2\end{bmatrix}=[z_1;z_2]\begin{bmatrix}8/3 & 0\\0 & 8/3\end{bmatrix}\begin{bmatrix}z_1\\z_2\end{bmatrix}=8/3\,z_2^2>0 \ \forall\ z_2\neq 0$$

Por lo tanto, $x_1^* = (4; 0; -1/3)$ es un mínimo local estricto del problema.

Para $x_2^* = (-2; 0; 7/3)$

$$(z_1;z_2)\begin{bmatrix}-6\\0\end{bmatrix}=0 \Rightarrow \textit{se verifica}\,\forall\, z\, \textit{no nulo} / \ z_1=0 \ y \ z_2 \neq 0$$

Luego:

$$[z_1;z_2]H_xL(x^*)\begin{bmatrix}z_1\\z_2\end{bmatrix}=[z_1;z_2]\begin{bmatrix}-8/3 & 0\\0 & -8/3\end{bmatrix}\begin{bmatrix}z_1\\z_2\end{bmatrix}=-8/3\,z_2^2<0 \ \forall\ z_2\neq 0$$

Por lo tanto, $x_2^* = (-2; 0; 7/3)$ es un máximo local estricto del problema.

XI.3.2.1.3 Problemas Condicionados por Desigualdades

Este tipo de NLP puede ser representado a través de la siguiente estructura general: max f(x) s.a:$\{x \in R^n,\ g_i(x) \leq b_i\ \ \forall i=1,..,m\}$ siendo $f:R^n \rightarrow R^1$ y $g_i:R^n \rightarrow R^1$ $\forall$ i=1,..,m. Todo NLP puede ser llevado a esta estructura, por lo tanto los teoremas siguientes son de aplicación general.

Teorema XI.6: *Condición Necesaria de Optimo Local de Kuhn Tucker.*
Sea x^* un máximo local restringido de f(x) sobre la región factible definida por $g_i(x) \leq b_i$ $\forall$ i=1,...,m, donde f(x) y $g_i(x)$ $\forall$ i=1,...,m son continuas y diferenciables. Si los $\nabla g_i(x^*)$ $\forall$i=1,...,m son linealmente independientes, luego $\exists\ \lambda^*=(\lambda_1^*,..,\lambda_m^*)$ tal que:

$$\nabla f(x^*) - \Sigma_{\forall i}\ \lambda_i^*\ \nabla gi(x^*) = 0$$
$$\lambda_i^*\ (gi(x^*) - b_i) = 0 \qquad \forall i=1,...,m$$
$$\lambda i^* \geq 0 \qquad \forall i=1,...,m$$

ISBN: 950-42-0022-2 - ©1999

Teorema XI.7: *Condición Suficiente de óptimo global*
Sea f(x):$R^n \rightarrow R^1$ una función cóncava a ser maximizada sobre una región factible convexa (ver Apéndice al final del capítulo), luego todo $x^* \in R^n$ que satisfaga las condiciones necesarias de óptimo de Kuhn-Tucker es un máximo global de PNL.
Sea f(x):$R^n \rightarrow R^1$ una función convexa a ser minimizada sobre una región factible convexa, luego todo $x^* \in E^n$ que satisfaga las condiciones necesarias de óptimo de Kuhn-Tucker es un mínimo global de NLP.

Ejemplo: Dado el NLP:

$$\begin{aligned} &Min\, f(x) = x_1^2 + x_2 \\ &s.a; \quad x_1^2 + x_2^2 \leq 9 \\ &\qquad\quad x_1 + x_2 \leq 1 \end{aligned}$$

Se resuelve el mismo usando condición necesaria y suficiente de óptimo.

Resolución: Se aplica condición necesaria de óptimo local (Teorema XI.6), para lo cual es necesario primero llevar el NLP a la estructura general en base a la cual se planteó el teorema XI.6, esto requiere en este caso: max -f(x). Luego se obtiene:

$$\nabla f(x) - \lambda_1 \nabla g_1(x) - \lambda_2 \nabla g_2(x) = 0$$

$$\lambda_1 [g_1(x) - 9] = 0$$

$$\lambda_2 [g_2(x) - 1] = 0$$

$$\lambda_1 \geq 0$$

$$\lambda_2 \geq 0$$

$$x_1^2 + x_2^2 \leq 9$$

$$x_1 + x_2 \leq 1$$

Se genera así el siguiente sistema de ecuaciones e inecuaciones:
1) $-2\, x_1 - \lambda_1\, 2\, x_1 - \lambda_2 = 0$
2) $-1 - 2\, \lambda_1\, x_2 - \lambda_2 = 0$
3) $\lambda_1\, (x_1^2 + x_2^2 - 9) = 0$
4) $\lambda_2\, (x_1 + x_2 - 1) = 0$
5) $\lambda_1 \geq 0$
6) $\lambda_2 \geq 0$
7) $x_1^2 + x_2^2 \leq 9$
8) $x_1 + x_2 \leq 1$

Resolviendo el sistema de ecuaciones e inecuaciones se obtienen el punto $(x^*, \lambda^*) = (0; -3; 1/6; 0)$ que verifica las condiciones necesarias de óptimo local.

ISBN: 950-42-0022-2 - ©1999

Se analiza a continuación la condición suficiente de óptimo, Teorema XI.7, obteniendo:

Función objetivo: $f(x) = x_1^2 + x_2$ es una función convexa.

Región factible: las funciones $g_1 = x_1^2 + x_2^2$ y $g_2 = x_1 + x_2$ son convexas y el signo de la restricción correspondiente es de $\leq$, → la región factible es convexa (ver Apéndice de final de capítulo).

Por lo tanto, todo x^* que satisface la condición necesaria de óptimo local Teorema XI.7, es mínimo global.

XI.3.2.2 Algoritmos para Resolver NLP

Son muchos los algoritmos desarrollados para la resolución de NLP. Cada uno de ellos adopta un tipo de NLP (convexo, cuadrático, con restricciones lineales, sin restricciones, etc.) y en función de las características particulares del mismo desarrolla una estrategia de búsqueda del óptimo.

El objetivo de esta sección es dar al lector una idea básica de la estrategia y la estructura de los algoritmos de optimización más convencionales. Siguiendo este propósito se describen a continuación algunos de los algoritmos básicos más difundidos en la bibliografía.

XI.3.2.2.1 Algoritmos para NLP Univariables sin Restricciones

Los problemas de optimización univariables sin restricciones han sido muy estudiados desde la perspectiva de lograr un método de búsqueda del óptimo que fuera lo más eficiente y robusto posible. El resultado ha sido la generación de casi una decena de algoritmos muy eficientes. La importancia asignada a la resolución de problemas de optimización tan pequeños radica en que la mayoría de los algoritmos de resolución de problemas de optimización multivariables realizan en cada iteración optimizaciones univariables, para lo cual requieren de estos algoritmos. Por ejemplo para resolver un problema de 20 variables un algoritmo puede requerir 20 optimizaciones univariables en cada iteración. Si el número de iteraciones requeridas para alcanzar la convergencia fuese 10, el algoritmo requerirá realizar 200 convergencias de un algoritmo de optimización univariable. Resulta entonces evidente la importancia de estos algoritmos.

Métodos de Reducción de Intervalo

Consideremos el problema: min $f(x)$ / $x \in [a,b]$. $[a,b]$ se denomina intervalo de incertidumbre, puesto que el óptimo $x^* \in [a,b]$ pero se desconoce su valor. Es necesario destacar que a y b son valores elegidos arbitrariamente, no son restricciones del problema, de manera tal que si el óptimo cae en uno de los extremos del intervalo, es necesario desplazar el valor del intervalo y resolver nuevamente el problema.

Los algoritmos de búsqueda del óptimo por reducción de intervalo, esencialmente se basan en reducir el intervalo de incertidumbre eliminando porciones del mismo en las cuales se tiene certeza que no se encuentra el óptimo. Los

ISBN: 950-42-0022-2 - ©1999

algoritmos más conocidos de este grupo son el método de la sección dorada y el método de Fibonacci. A continuación se describe uno de ellos.

Teorema XI.8: Sea $f:R^1 \rightarrow R^1$ estrictamente convexa sobre [a,b]. Sean $\lambda,\mu \in [a,b]$ / $\lambda<\mu$.
Luego:

(i) Si $f(\lambda)>f(\mu) \rightarrow f(x_3)>f(\mu) \quad \forall x_3 \in [a,\lambda]$.

(ii) Si $f(\lambda)\leq f(\mu) \rightarrow f(x_3)\geq f(\lambda) \quad \forall x_3 \in [\mu,b]$.

Método de la Sección Dorada: Dado el problema min f(x) / $x \in [a,b]$ con f(x) estrictamente convexa. Sea $[a_k,b_k]$ el intervalo de incertidumbre asociado a la iteración k, y sean λ_k y $\mu_k \in [a_k,b_k]$ / $\lambda_k<\mu_k$. Del Teorema XI.8 se tiene:

(i) Si $f(\lambda_k) > f(\mu_k) \rightarrow (b_{k+1}-a_{k+1}) = (b_k-\lambda_k)$.

(ii) Si $f(\lambda_k)$ # $f(\mu_k) \rightarrow (b_{k+1}-a_{k+1}) = (\mu_k -a_k)$.

Los puntos λ_k y μ_k se definen como:

(A) $\lambda_k=a_k+(1-\alpha)(b_k-a_k)$

(B) $\mu_k=a_k+ \alpha(b_k-a_k)$

Calculando λ_k y μ_k con las ecuaciones (A) y (B), es fácil deducir que los intervalos de incertidumbre de dos iteraciones sucesivas se relacionan a través de la siguiente ecuación: $(b_{k+1}-a_{k+1}) = \alpha(b_k-a_k)$.

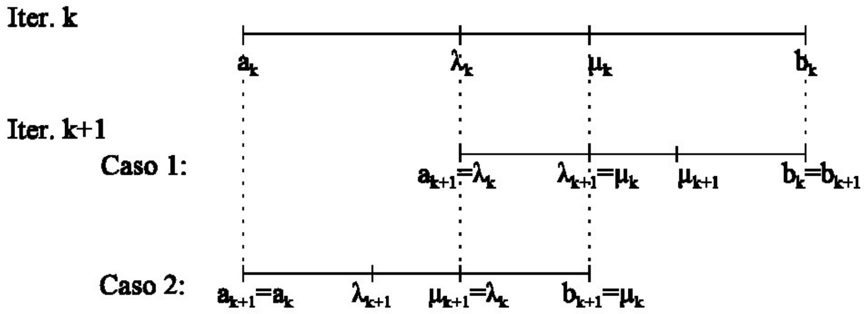
Es fácil mostrar que un valor de α=0,618 permite que la longitud del nuevo intervalo de incertidumbre no dependa del resultado de la iteración k (independiente de si resulta (i) o (ii)), y el punto que permanece en el intervalo sirve como uno de los puntos para la nueva iteración. Esto implica que en la primera iteración es necesario calcular dos puntos y en las restantes iteraciones sólo es necesario calcular un punto. Se presentarán dos casos:

***Caso 1**:* Si $f(\lambda_k)>f(\mu_k)$ por el Teorema XI.8 $a_{k+1}=\lambda_k$ y $b_{k+1}=b_k$ y $\lambda_{k+1}=\mu_k$ calculando μ_{k+1} con la ecuación (B).

***Caso 2**:* Si $f(\lambda_k)\leq f(\mu_k)$ por el Teorema XI.8 $a_{k+1}=a_k$ y $b_{k+1}=\mu_k$ y $\mu_{k+1}=\lambda_k$ calculando λ_{k+1} con la ecuación (A).

Representación gráfica:

ISBN: 950-42-0022-2 - 81999



ALGORITMO

Etapa Inicial

Elegir la longitud final del intervalo de incertidumbre $l>0$

Sea $[a_1,b_1]$ el intervalo inicial, calcular:

$\lambda_1=a_1+(1-\alpha)(b_1-a_1)$

$\mu_1=a_1+\alpha(b_1-a_1)$ $\alpha = 0{,}618$

evaluar $f(\lambda_k) \wedge f(\mu_k)$, hacer k=1 y continuar

Etapa principal

Mientras $(b_k-a_k) \geq l$

Si $f(\lambda_k)>f(\mu_k)$

Hacer $a_{k+1}=\lambda_k$ y $b_{k+1}=b_k$ y $\lambda_{k+1}=\mu_k$ y $\mu_{k+1}=a_{k+1}+\alpha(b_{k+1}-a_{k+1})$

Evaluar $f(\mu_{k+1})$

Si No

Hacer $a_{k+1}=a_k$ y $b_{k+1}=\mu_k$ y $\mu_{k+1}=\lambda_k$ y $\lambda_{k+1}=a_{k+1}+(1-\alpha)(b_{k+1}-a_{k+1})$

Evaluar $f(\lambda_{k+1})$

Fin Si

Hacer k=k+1

Fin Mientras

Ejemplo: Resolución del NLP min $f(x) = 2x^2+16/x$ en el intervalo 1#x#5 usando el algoritmo de la sección dorada, hasta una longitud final del intervalo de incertidumbre l=0,03.

Los resultados de las sucesivas iteraciones se muestran en la siguiente tabla:

k	a_k	b_k	λ_k	u_k	$f(\lambda_k)$	$f(u_k)$	b_k-a_k
1	1	5	2.528	3.472	19.1107	28.7179	4
2	1	3.472	1.9443	2.528	15.7898	19.1107	2.472

k	a_k	b_k	λ_k	u_k	$f(\lambda_k)$	$f(u_k)$	b_k-a_k
3	1	2.528	1.5831	1.9443	15.1191	15.7898	1.528
4	1	1.9443	1.3607	1.5837	15.4616	15.1191	0.9443
5	1.3607	1.9443	1.5837	1.7214	15.1191	15.2212	0.5836
6	1.3607	1.7214	1.4985	1.5837	15.1684	15.1191	0.3607
7	1.4985	1.7214	1.5837	1.6362	15.1191	15.1331	0.2288
8	1.4985	1.6362	1.5511	1.5837	15.1271	15.1191	0.1377
9	1.5511	1.6362	1.5837	1.6037	15.1191	15.1206	0.0851
10	1.5511	1.6037	1.5712	1.5837	15.1206	15.1191	0.0526
11	1.5712	1.6037	1.5837	1.5913	15.1191	15.1191	0.0325
12	1.5712	1.5913	1.5789	1.5837	15.1195	15.1191	0.02

De la tabla se obtiene que el óptimo pertenece al intervalo [1,5789; 1,5913]. Luego, puede adoptarse como óptimo a cualquier punto de dicho intervalo. En general se elige el punto medio.

Métodos que Utilizan Derivadas

Estos métodos se diferencian de los del grupo anterior porque utilizan la información de la derivada de la función para dirigir la búsqueda.

Generalmente convergen en un menor número de iteraciones. Requieren que la función sea diferenciable, y además es necesario calcular la derivada de la función ya sea en forma analítica o numérica.

En este grupo se pueden incluir el método de Newton, de la bisección, de la secante y búsqueda cúbica como los más convencionales. A continuación se describe el método de Newton.

Método de Newton: Se basa en aproximar a la función f(x) en el punto λ_k, a una función cuadrática q(x), y luego buscar el mínimo de q(x). La aproximación cuadrática se obtiene mediante una expansión de Taylor que incluye el término de segundo orden $q(x)=f(\lambda_k)+f'(\lambda_k)(x-\lambda_k)+\frac{1}{2} f''(\lambda_k)(x-\lambda_k)^2$. Si λ_k es un punto suficientemente próximo a un óptimo, todo punto x que verifica la condición necesaria de óptimo local no condicionado $q'(x)=0$, verifica también la condición suficiente de óptimo local no condicionado. Luego, λ_{k+1} es elegido como el punto que minimiza la función q(x). Es decir, que verifica $q'(x)=f'(\lambda_k)+f''(\lambda_k)(x-\lambda_k)=0$. De aquí se obtiene (haciendo $\lambda_{k+1}=x$): $\lambda_{k+1}=\lambda_k- f'(\lambda_k)/f''(\lambda_k)$.

El proceso termina cuando $|\lambda_{k+1}- \lambda_k| < \varepsilon$ o $|f'(\lambda_k)| < \varepsilon$ donde ε es un escalar

ISBN: 950-42-0022-2 - ©1999

prefijado.

Nota: El procedimiento puede ser aplicado a funciones diferenciables dos veces, y además es necesario que f''(λ_k)≠0. Puede demostrarse que este método converge si el punto inicial es suficientemente próximo a un óptimo.

ALGORITMO

Etapa Inicial

Elegir λ_1 y ε. hacer k=1

Calcular f'(λ_k) y f''(λ_k) y $\lambda_{k+1}= \lambda_k$ - f '(λ_k)/f'' (λ_k) y continuar

Etapa principal

Mientras (λ_{k+1}-λ_k) ∃ε

Hacer k=k+1

Calcular f'(λ_k), f''(λ_k) y $\lambda_{k+1}= \lambda_k - f'(\lambda_k)/f''(\lambda_k)$

Fin Mientras

Ejemplo: Se resuelve el NLP min f(x) = $2x^2+16/x$ usando el algoritmo de Newton, con ε=0,03.

Los resultados de las sucesivas iteraciones se muestran en la tabla siguiente.

$$f'(x) = 4x - \frac{16}{x^2}; \quad f''(x) = 4 + \frac{32}{x^3}$$

k	x_k	fN(x_k)	fO(x_k)	f(x_k)	$x_{(k+1)}$	$x_{(k+1)}$-x_k
1	3.0000	10.2222	5.1852	23.3333	1.0286	1.9714
2	1.0286	-11.0092	33.4067	17.6715	1.3581	0.3295
3	1.3581	-3.242	16.7742	15.47	1.5514	0.1933
4	1.5514	-0.4422	12.5701	15.127	1.5866	0.0352
5	1.5866	-0.0099	12.0125	15.1191	1.5874	0,0008<0,03

De la tabla se adopta el punto x^*=1,5874 como un mínimo local. El valor de la función objetivo en este punto es f(x^*)=15,1191.

XI.3.2.2.2 Algoritmos para NLP sin Restricciones Multivariables

Los problemas de optimización multivariables sin restricciones son muy comunes en todas las áreas del conocimiento, de allí el gran esfuerzo que se ha invertido en desarrollar algoritmos eficientes para resolver los mismos. Se puede rescatar de la bibliografía cerca de una veintena de algoritmos extensamente usados, no obstante es mucho mayor la lista de algoritmos que por determinadas circunstancias no fueron muy difundidos.

En su mayoría, estos algoritmos fueron desarrollados a partir de razonamiento heurístico y posteriormente justificados matemáticamente, cuando era posible. Una característica de estos algoritmos es que sólo convergen a óptimos locales.

En los últimos años se han desarrollado algunas estrategias que intentan encontrar óptimos globales. Esta es un área aún en desarrollo.

En lo que sigue, se describen algunos de los algoritmos clásicos, que constituyeron la base de la mayoría de los algoritmos desarrollados posteriormente.

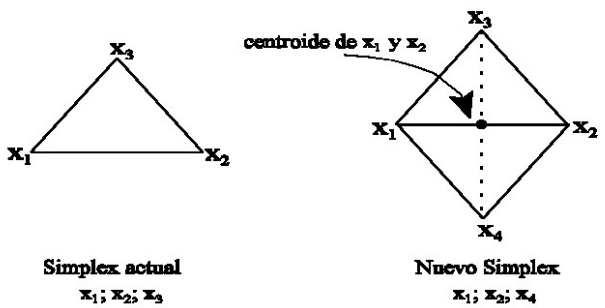
Método de Búsqueda Simplex: Este método, es una de las más antiguas y elementales estrategias de búsqueda del óptimo de una función multivariable sin restricciones. Se lo conoce con el mismo nombre del método para resolver LP descripto en secciones anteriores, pero su estrategia no guarda relación alguna.

Dado el problema min f(x) / $x \in R^n$, el método consiste en:

Seleccionar n+1 puntos equidistantes, los cuales corresponderían a los vértices de un poliedro regular que se denomina simplex. Por ejemplo, si n=2 el simplex es un triángulo equilátero, si n=3 el simplex es un tetraedro, etc.

Generar un nuevo Simplex seleccionando el vértice con el peor valor de la función objetivo y proyectándolo una distancia determinada a través del punto equidistante de todos los restantes vértices denominado ce*ntroide*.

En el gráfico de la figura siguiente se representa en forma esquemática el procedimiento para el caso de un problema de dos variables:



El procedimiento en su evolución utiliza las siguientes reglas:
Regla 1: Si el peor vértice fue generado en la iteración anterior, elija el vértice con el siguiente peor valor de la función objetivo.
Regla 2: Si un dado vértice se mantiene inalterado durante M iteraciones, reducir el tamaño del simplex por algún factor. Iniciar un nuevo simplex manteniendo el mejor punto.
Regla 3: Criterio de terminación. La búsqueda es concluida cuando el simplex es suficientemente pequeño, o cuando la desviación estándar del valor de la función objetivo en cada vértice se torna suficientemente pequeña.

Métodos de Búsqueda Directa

Estos algoritmos se basan en la siguiente **estrategia**:

- *dado un punto x,*
- *determinar una dirección d de búsqueda,*
- *búsqueda lineal: luego, f(x) es minimizada desde x en la dirección d. Para ello se requiere resolver el siguiente problema de optimización univariable:*

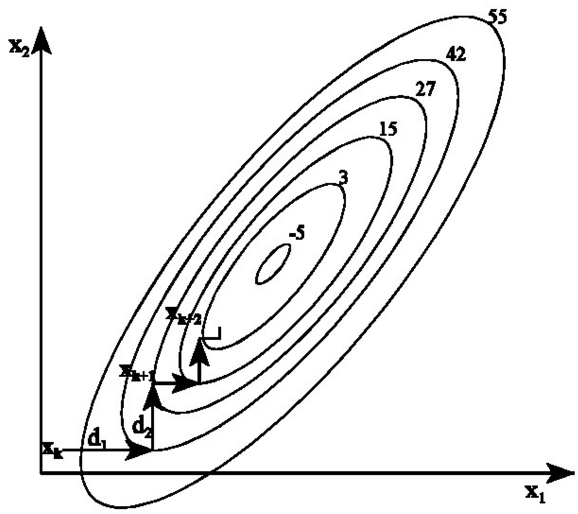
$$\min_{\lambda} f(x+\lambda d) \text{ s.a:}\{\lambda 0\ [a,b],\ \lambda \in R^1\}$$

Este subproblema se resuelve mediante alguno de los algoritmos para problemas de optimización univariable descriptos en la sección anterior. Esta estrategia es la base de muchos algoritmos. La diferencia entre ellos sólo radica en el método que usan para determinar la dirección de búsqueda. A continuación se describen dos de los algoritmos más simples a efectos de mostrar la arquitectura de los mismos.

Método Coordenado Cíclico: Utiliza los ejes coordenados como direcciones de búsqueda. Es decir, utiliza las direcciones $d_1,d_2,...,d_n$ donde d_j tiene un 1 en la posición j y ceros en todo el resto. Por lo tanto, cuando se optimiza desde el punto x^0 en la dirección d_j, el nuevo punto será $x^1=x^0+\lambda_j d_j = (x_1^{\,1}=x_1^{\,0};Y;\ x_j^{\,1} = x_j^{\,0}+\lambda_j;...\ ;x_n^{\,1} = x_n^{\,0})$.

Representación Esquemática

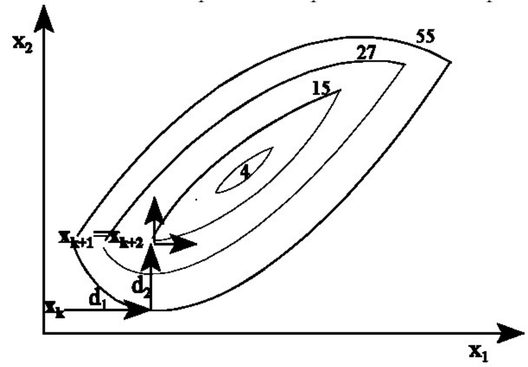
ISBN: 950-42-0022-2 - 81999



Una iteración se inicia en x_k y termina en x_{k+1} luego de haber optimizado (búsqueda lineal) en todas las direcciones. Es decir $x_{k+1}=x_k+\sum_{\forall i}\lambda_j d_j$ donde λ_j es el óptimo de $\min_\lambda f(x_j+\lambda d_j)$.

El *criterio de terminación* es $\|x_{k+1}-x_k\| < \varepsilon$, donde $\|.\|$ es la norma Euclídea.

Etapa de aceleración: Este método converge a un punto con gradiente cero (óptimo local) cuando se aplica a una función diferenciable. En ausencia de diferenciabilidad, el método puede detenerse en un punto no óptimo. Esto se representa en la figura:



En el punto x_{k+1} la función alcanza el valor mínimo según ambas direcciones $d_1=(1,0)$ y $d_2=(0,1)$ por lo tanto allí se detiene sin ser este punto el óptimo. Para salvar este problema, se propone una optimización univariable desde x_k en la dirección $(x_{k+1}-x_k)$.

ISBN: 950-42-0022-2 - ©1999

La búsqueda a lo largo de la dirección (x_{k+1}-x_k) suele usarse frecuentemente en el método coordenado cíclico, aún cuando f(x) es diferenciable, para acelerar la convergencia del método. De allí que se la denomina *etapa de aceleración*. Usualmente cada *p* iteraciones se aplica una etapa de aceleración. El parámetro *p* es determinado en base a criterios prácticos.

ALGORITMO

Etapa Inicial

Elegir $\varepsilon>0$ y $p\geq 1$ (entero). Hacer $dj=e_j$ $\forall j=1,..,n$. Elegir x_1. Hacer $y_1=x_1$ y k=j=i=1

```
Mientras j≤n
    Calcular λj solución óptima de minλ f(xj+λdj)  / λ∈R¹
    Hacer yj+1=yj+λjdj
    Hacer j=j+1
Fin Mientras
    Hacer xk+1= yn+1
```

Etapa principal

```
Mientras ‖xk+1- xk‖≥ε
    Mientras  i≤p
        Hacer j=1
        Hacer y1 = xk+1
        Hacer k=k+1
        Mientras j≤n
            Calcular λj solución óptima de minλ f(yj+λdj)  / λ∈R¹
            Hacer yj+1=yj+λjdj
            Hacer j=j+1
        Fin Mientras
        Hacer xk+1= yn+1
        Hacer i= i+1
    Fin Mientras
    Hacer  d=xk+1-xk
    Calcular λ* solución óptima de minλ f(xk+1+λd)  / λ∈R¹
    Hacer k=k+1
    Hacer  xk+1= xk+λ*d
    Hacer i=1
Fin Mientras
```

Ejemplo: Resolución del NLP min $f(x) = 4x_1^2 + 3x_2^2 - 4x_1x_2 + x_1$ utilizando el método coordenado cíclico desde el punto inicial x_0=(0;-1). Adoptar ε = 0,03.

Previo a iniciar el proceso iterativo es necesario realizar algunas consideraciones:

Las direcciones de búsqueda para este problema definido en R^2 son: d_1= (1; 0) y d_2= (0; 1).

Sea λ^* el valor que minimiza f(x+λd). Dicho valor se obtiene resolviendo el problema univariable $\min_\lambda f(x+\lambda d)$ s.a: {$\lambda \in [a,b]$, $\lambda \in R^1$ } utilizando para esto algún algoritmo de búsqueda univariable tal como el método de la sección dorada visto anteriormente.

No obstante, desde un punto de vista académico práctico, es más fácil resolver el problema $\min_\lambda f(x+\lambda d)$ s.a: {$\lambda \in [a,b]$, $\lambda \in R^1$ } en forma analítica utilizando elementos teóricos. Debe quedar claro que éste es sólo un recurso académico.

Luego, para calcular el valor de λ^* en cada una de las sucesivas búsquedas lineales se puede formular una ecuación en forma genérica asociada a la función a optimizar:

Sea el punto inicial y = (a; b) y la dirección de búsqueda d = (c; g) . Luego:

$$\min f(y+\lambda d) = f[a+\lambda c; b+\lambda g] = 4(a+\lambda c)^2 + 3(b+\lambda g)^2 - 4(a+\lambda c)(b+\lambda g) + (a+\lambda c)$$

Se obtiene así una función en λ a ser minimizada. Luego, si λ^* minimiza f(y +λd) debe verificar la condición necesaria y suficiente de óptimo local no condicionado:

$$\frac{df}{d\lambda} = 0 = 8c(a+\lambda^* c) + 6g(b+\lambda g) - 4c(b+\lambda^* g) - 4g(a+\lambda^* c) + c$$

$$\lambda^* = \frac{4c(b-2a) + 2g(2a-3b) - c}{8c^2 + 6g^2 - 8cg}$$

La ecuación anterior es aplicable a cualquier dirección de búsqueda. Para el caso particular de la dirección de búsqueda d_1= (1; 0), es c = 0 y g = 1, luego λ_1^* vendrá dado por:

$$\lambda_1^* = \frac{4(b-2a)-1}{8}$$

Mientras que para la dirección de búsqueda d_2=(0; 1), es c = 1 y g = 0, luego λ_2^*

$$\lambda_2^* = \frac{4a-6b}{6}$$

Con las ecuaciones generadas se calculará el valor de λ_1^* y λ_2^* de cada búsqueda lineal. Se inicia entonces la aplicación del algoritmo:

Iteración 1:

$$y_1^{(1)} = x_0 = [0;-1] \quad y_2^{(1)} = y_1^{(1)} + \lambda_1^* d_1 = [0+\lambda_1^*;-1]$$

$$\lambda_1^* = \frac{4(-1-0)-1}{8} = -0.625 \quad \Rightarrow y_2^{(1)} = [-0.625;-1]$$

$$y_3^{(1)} = [-0.625;-1+\lambda_2^*] \quad \lambda_2^* = \frac{(-4)0.625+6}{6} = 0.5833$$

$$\Rightarrow y_3^{(1)} = x_1 = [-0.625;-0.4167]$$

$$\|x_0 - x_1\| = \sqrt{(0.625)^2 + (-1+0.4167)^2} = 0.8549 > 0.03$$

Iteración 2:

$$y_1^{(2)} = x_1 = [-0.625;-0.4167]$$

$$y_2^{(2)} = [-0.625+\lambda_1^*;-0.4167] = [-0.33335;-0.4167] \quad (\lambda_1^* = 0.29165)$$

$$y_3^{(2)} = [-0.33335;-0.4167+\lambda_2^*] = [-0.33335;-0.2222] = x_2 \quad (\lambda_2^* = 0.19446)$$

$$\|x_2 - x_1\| = 0.3506 > 0.03$$

Los resultados de las 4 primeras iteraciones se presentan en la tabla siguiente:

Iteración i	$x_{i=1}=y_1^i$	λ_1^*	y_2^i	λ_2^*	$y_3^i=x_i$	$f(x_i)$	$\|x_{i-1}-x_i\|$
1	$\begin{pmatrix}0,0000\\-1,0000\end{pmatrix}$	-0.625	$\begin{pmatrix}-0,6250\\-1,0000\end{pmatrix}$	0.5833	$\begin{pmatrix}-0,6250\\-0,4167\end{pmatrix}$	0.4167	0.8549
2	$\begin{pmatrix}-0,6250\\-0,4167\end{pmatrix}$	0.2916	$\begin{pmatrix}-0,3334\\-0,4167\end{pmatrix}$	0.1946	$\begin{pmatrix}-0,3334\\-0,2222\end{pmatrix}$	-0.037	0.3506
3	$\begin{pmatrix}-0,3334\\-0,2222\end{pmatrix}$	0.0972	$\begin{pmatrix}-0,2361\\-0,2222\end{pmatrix}$	0.0648	$\begin{pmatrix}-0,2361\\-0,1574\end{pmatrix}$	-0.0874	0.1168
4	$\begin{pmatrix}-0,2361\\-0,1574\end{pmatrix}$	0.0324	$\begin{pmatrix}-0,2037\\-0,1574\end{pmatrix}$	0.0216	$\begin{pmatrix}-0,2037\\-0,1358\end{pmatrix}$	-0.0931	0.039

Luego de cuatro iteraciones (p=4) se decide aplicar una etapa de aceleración.
La dirección de búsqueda para la aceleración es:

$$d = x^4 - x^3 = [0.0324;\ 0.0216]$$

$$x^5 = x^4 + \lambda^* d = [-0.2037 + \lambda^*\ 0.0324; -0.1358 + \lambda^*\ 0.0216] \quad (\lambda^* = 0.5)$$

$$x^5 = [-0.1875;-0.1250] \quad \|x^4 - x^3\| = 0.019 < 0.03$$

Solución:

$$x^* = [-0.1875;-0.1250] \qquad f(x^*) = -0.09375$$

Método de Hooke & Jeeves *(Modificado)*. Este método se basa en realizar la

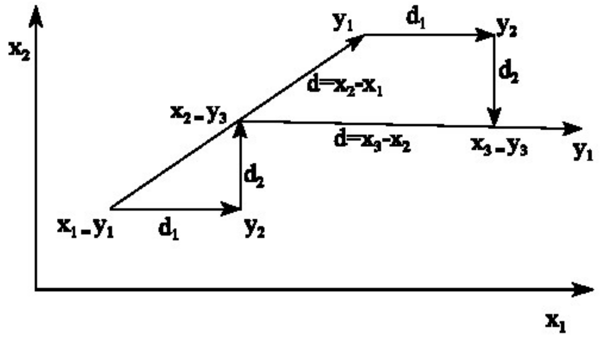
ISBN: 950-42-0022-2 - ©1999

búsqueda del óptimo según dos tipos de estrategias:

Búsqueda exploratoria: en la iteración k, partiendo del punto y_k genera el punto x_{k+1} mediante una búsqueda siguiendo las direcciones de los ejes coordenados, luego pasa a la etapa de búsqueda pautada.

Búsqueda pautada: generado el punto x_{k+1} mediante la etapa exploratoria, realiza una búsqueda a lo largo de la dirección $d=(x_{k+1}- x_k)$ generando el punto y_{k+1}, luego vuelve a la etapa de búsqueda exploratoria.

Gráfico:



Nota: El método de Hooke and Jeeves original, propone realizar las búsquedas a lo largo de los ejes coordenados y de la dirección $d=(x_{k+1}- x_k)$ a través de saltos discretos y evaluaciones de la función. La magnitud del salto se va ajustando a medida que avanza el cálculo. Es decir, en su versión original el método no propone realizar optimizaciones univariables según las distintas direcciones.

Ejemplo: Resolución del NLP min $f(x) = 4x_1^2 + 3x_2^2- 4x_1x_2 + x_1$ utilizando el método de Hooke & Jeeves desde el punto inicial $x_1=(0;-1)$.Adoptar $\varepsilon = 0{,}03$.
Se define $y_1= x_1= (0; -1)$.

Iteración 1:

Búsqueda exploratoria: en las dirección $d_1=(1; 0)$ y $d_2=(0; 1)$ desde $x_1=(0;-1)$.
Esta búsqueda es la misma que se realizó en la iteración 1 del método anterior, por lo tanto se utilizan aquí dichos resultados: $x_2= (-0{,}625; -0{,}4167)$

Búsqueda pautada: en la dirección $(x_2 - x_1) = (-0{,}625; 0{,}583)$

ISBN: 950-42-0022-2 - ©1999

$$\min f[x_2+\lambda(x_2-x_1)]=\min f[(-0.625;-0.417)+\lambda(-0.625;0.583)]=$$
$$\min f(-0.625-0.625\lambda;-0.417+0.583\lambda)$$

Luego, para realizar la búsqueda univariable y obtener el valor de λ^* puede utilizarse la ecuación desarrollada en el ejemplo anterior, o bien plantear y resolver el problema univariable correspondiente. Aquí se usa la segunda posibilidad simplemente a modo de muestra. Luego, se reemplazan los valores de las componentes del nuevo punto en la función objetivo original obteniéndose una función en λ:

$$\min F(\lambda)=4(-0{,}625-0{,}625\lambda)^2+3(-0{,}417+0{,}583\lambda)^2-4(-0{,}625-0{,}625\lambda)$$
$$(-0{,}417+0{,}583\lambda)-0{,}625-0{,}625\lambda$$

$$\min F(\lambda)=4(-0.625-0.625\lambda)^2+3(-0.417+0.583\lambda)^2-4(-0.625-0.625\lambda)$$
$$(-0.417+0.583\lambda)-0.625-0.625\lambda$$

Si λ^* es un mínimo local no condicionado de $F(\lambda)$ debe verificar la condición necesaria de óptimo:

$$\frac{dF(\lambda)}{d\lambda}=3.125+3.124\lambda-1.458+2.038\lambda-1.042+1.457+2.914\lambda-0.625=0$$
$$1.457+8.076\lambda=0 \quad \Rightarrow \lambda_3^*=0.180$$

Luego, el nuevo punto será:

$$y_1=x_2+\lambda_3^*(x_2-x_1)=[-0.625;-0.417]+(-0.180)[-0.625;0.583]$$
$$y_1=[-0.512;-0.521]$$

Iteración 2:

Búsqueda exploratoria, en las direcciones d_1=(1; 0) y d_2= (0; 1) desde el punto generado en la iteración anterior y_1=(-0,512; -0,521). Esto es:
Búsqueda univariable según d_1:

$$\min f(y_1+\lambda d_1)=\min f[(-0.512;-0.521)+\lambda(1;0)]=\min f(-0.512+\lambda;-0.521)$$

Reemplazando en la función objetivo y aplicando condición necesaria y suficiente de óptimo local no condicionado se obtiene λ_1^*:

$$F(\lambda)=1.048-4.096\lambda+4\lambda^2+0.814-1.067+2.084\lambda-0.512+\lambda$$
$$\frac{dF(\lambda)}{d\lambda}=8\lambda-1.012=0 \qquad \Rightarrow \lambda_1^*=0.126$$
$$y_2=y_1+\lambda_1^*d_1=(-0.512;-0.521)+0.126(1;0)=(-0.386;-0.521)$$

ISBN: 950-42-0022-2 - ©1999

Búsqueda univariable según d_2:

$$Min\, f(y_2 + \lambda\, d_2) = Min\, f[(-0.386; -0.521) + \lambda(_2 0;1)] = (-0.386; -0.521 + \lambda)$$

Reemplazando en la función objetivo y aplicando condición necesaria y suficiente de óptimo local no condicionado se obtiene λ_2^*:

$$F(\lambda) = 4(-0.386)^2 + 3(-0.521 + \lambda)^2 - 4(-0.386)(-0.521 + \lambda) - 0.386$$

$$\frac{dF(\lambda)}{d\lambda} = 6\lambda - 1.582 = 0 \quad \Rightarrow \lambda_2^* = 0.263$$

$$y_3 = x_3 = (y_2 + \lambda_2^*\, d_2) = (-0.386;\ -0.521) + 0.263(0;1) = (-0.386; -0.258)$$

Luego se verifica convergencia:

$$\|x_3 - x_2\| = 0.4642 > \varepsilon$$

Dado que no verifica convergencia se pasa a:

Búsqueda pautada en la dirección (x_3 - x_2)

$$\min f(x_2 + \lambda(x_3 - \bar{x}_2)) = \min f(-0.386 + 0.239\,\lambda;\ -0.258 + 0.159\ \lambda)$$

Reemplazando en la función objetivo y aplicando condición necesaria y suficiente de óptimo local no condicionado se obtiene λ_3^*:

$$F(\lambda) = 4(-0.386 + 0.239\,\lambda)^2 + 3(-0.258 + \lambda)^2$$
$$-4(-0.386 + 0.239\,\lambda)(-0.258 + 0.159\,\lambda) - 0.386 + 0.239\,\lambda$$

$$\frac{dF(\lambda)}{d\lambda} = 0.302\,\lambda - 0.207 = 0 \quad \Rightarrow \lambda_3^* = 0.685$$

$$y_1 = [x_3 + \lambda_2^*(x_3 + x_2)] = (-0.386;\ -0.258) + 0.685(0.239;\ 0.159)$$

$$y_1 = (-0.222;\ -0.149)$$

Iteración 3: a partir del punto generado en la iteración 2 y_1= (-0,222; -0,149) se

ISBN: 950-42-0022-2 - ©1999

vuelve a aplicar el algoritmo hasta verificar la condición de convergencia.

Métodos que Usan Derivadas

Método de la Dirección de Descenso más Empinada o de Cauchy: Requiere que la función sea diferenciable. Utiliza como dirección de búsqueda la dirección del gradiente de la función. Para un problema de minimización será: $d = -\nabla f(x)$ por lo tanto $x_{k+1} = x_k - \lambda \nabla f(x_k)$.

Inconvenientes del método (Zigzagging): Este método generalmente opera bien durante las primeras iteraciones, produciendo rápidas aproximaciones al óptimo, pero presenta dificultades en las proximidades del óptimo. Este fenómeno se puede corregir produciendo una desviación del gradiente. Es decir, en lugar de utilizar la dirección de desplazamiento $d = -\nabla f(x)$, se puede utilizar la dirección $d = -D \nabla f(x)$ o bien la dirección $d = -\nabla f(x) + g$ donde D y g son una matriz y un vector apropiados. Este procedimiento de corrección se analiza en detalle en los algoritmos posteriores.

Método de Newton *(multivariables)*: Requiere que la función sea contínua y diferenciable dos veces. Se basa en aproximar a la función f(x) en el punto x_k a una función cuadrática q(x) y luego buscar el mínimo de esta nueva función. La aproximación cuadrática se obtiene mediante una expansión de Taylor de segundo orden $q(x)=f(x_k)+ \nabla^T f(x_k)(x- x_k)+2(x-x_k)^T H(x_k)(x-x_k)$ donde $H(x_k)$ es la matriz Hessiana de f(x) en x_k. Si x_k es un punto suficientemente próximo a un óptimo, todo punto x que verifica la condición necesaria de óptimo local no condicionado $\nabla q(x)=0$, también verifica la condición suficiente de óptimo local no condicionado. Luego, x_{k+1} es elegido como el punto que minimiza la función q(x). Es decir $\nabla q(x)= \nabla f(x_k)+H(\lambda_k)(x-x_k)=0$. Asumiendo que $H(x_k)^{-1}$ ∃, (premultiplicando ambos miembros de la ecuación anterior por $H(x_k)^{-1}$ y reordenando) se tiene: $x_{k+1} = x_k - H(x_k)^{-1} \nabla f(x_k)$.

En esta ecuación recursiva, se ha definido al punto x que optimiza la función cuadrática q(x) como el punto x_{k+1}. De esta manera la ecuación permite calcular en forma recursiva, los puntos generados por el método de Newton, a partir de un punto inicial dado. De la ecuación de generación del nuevo punto se deriva como dirección de búsqueda $d = -H(x_k)^{-1} \nabla f(x_k)$. El procedimiento termina cuando $\| x_{k+1}- x_k \| < \varepsilon$ o $\| \nabla f(x_k) \| < \varepsilon$, donde ε es un escalar especificado previamente.

Análisis del método: Si x^* es un mínimo local estricto de f(x) ($\nabla f(x^*)=0$). Luego, asumiendo que f(x)) es continua y diferenciable dos veces, $H(x_k)$ será positiva definida para x_k suficientemente próximo a x^*, lo cual implica que $H(x_k)^{-1}$ ∃ y por lo tanto el punto x_{k+1} podrá ser calculado mediante la ecuación recursiva definida.

Convergencia: Los puntos generados por el método de Newton pueden no converger. La razón de que esto ocurra es que $H(x_k)$ puede ser singular, lo cual implica que

ISBN: 950-42-0022-2 - ©1999

$H(x_k)^{-1}$ no existe, o bien, aún si $H(x_k)^{-1}$ □, $f(x_{k+1})$ puede no ser necesariamente mejor que $f(x_k)$.
Puede demostrarse que el método converge si x_k es suficientemente próximo a un óptimo.

Conclusión respecto a los métodos que usan derivadas. El método de Cauchy tiene buena performance en las primeras iteraciones (lejos del óptimo) pero se torna ineficiente a medida que se aproxima al óptimo. En contraposición a éste, el método de Newton asegura convergencia en las proximidades del óptimo. Como consecuencia de las ventajas y desventajas de estas estrategias, se han desarrollado algoritmos que intentan *aprovechar* las ventajas de ambas. Algunos de ellos se describen a continuación.

Métodos que Usan Direcciones Conjugadas

Algunos de estos métodos utilizan derivadas, y otros evaluaciones de la función. Son, al igual que los anteriores, métodos de búsqueda directa cuya diferencia radica en que usan direcciones conjugadas.

Los distintos algoritmos de optimización que utilizan direcciones conjugadas, se diferencian entre si en la estrategia de generación de las direcciones conjugadas. Los más conocidos son el método de Davidon Fletcher and Powel, el método de Zangwill, y el método de Fletcher and Reeves entre otros. A modo de ejemplo, a continuación se describe el primero de estos algoritmos, luego de algunas definiciones necesarias.

*Vectores Conjugados***:** Sea H una matriz (nxn) simétrica. Los vectores $d_1,...,d_k$ son llamados H-conjugados o conjugados si ellos son l.i. y si $d_i^T H d_j=0 \quad \forall i \neq j$.

Teorema XI.9: Sea $f(x)=c^T x+2x^T Hx$, donde H es una matriz (nxn) simétrica. Sean $d_1,...,d_n$ direcciones H-conjugadas y sea x_1 un punto inicial arbitrario. Para k=1,...,n sea λ_k una solución óptima para el problema que implica: min $f(x_k+\lambda d_k)$, $\lambda \in E'$, $\wedge$ $x_{k+1}= x_k+ \lambda_k d_k$. Luego, para k=1,...,n se debe tener que x_{k+1} es una solución óptima del problema: min f(x) sa:$\{(x-x_1) \in L(d_1,...,d_k)\}$ donde $L(d_1,..,d_k) =\Sigma_{j=1,...k}\, \mu_j d_j \,/\, \mu_j \in R^1$ $\forall j$ es el subespacio lineal formado por $d_1,..,d_k$.

En particular, x_{n+1} es un mínimo del problema: min f(x) $x \in R^n$.

Este teorema asegura que optimizando una función cuadrática según un conjunto de direcciones conjugadas es posible alcanzar el óptimo en una sola iteración.

Método de Davidon, Fletcher & Powell (D.F.P): Requiere que la función sea diferenciable.
Dirección de búsqueda: $d_j = -D_j \nabla f(y_j)$ donde D_j es una matriz (nxn) positiva definida

y simétrica. Para la siguiente etapa, la matriz D_{j+1} se obtiene de:

$$D_{j+1}=D_j+p_jq_j^T/p_j^Tq_j-D_jp_jq_j^TD_j/p_j^TD_jq_j$$

$$\text{donde:} \quad p_j=\lambda_jd_j \qquad q_j=\nabla f(y_{j+1})-\nabla f(y_j) \qquad \wedge \qquad y_{j+1}=y_j+\lambda_jd_j.$$

Puede verse que este método utiliza como dirección de búsqueda la dirección del gradiente *modificado* por la matriz D_j.

Teorema XI.10: Considere el problema de minimizar $f(x)=c^Tx+2x^THx$ sujeto a $x\in R^n$, donde H es una matriz (nxn) simétrica y positiva definida. Suponga que el problema es resuelto utilizando el método (D.F.P.) comenzando con un punto inicial y_1 $\wedge$ una matriz simétrica (+)definida D_1. En particular, $\forall j=1,...,n$ sea λ_j la solución óptima del problema: min $f(y_j+\lambda d_j)$ sujeto a $\lambda\geq 0$ $\wedge$ sea $y_{j+1}= y_j+\lambda_jd_j$ donde $d_j= -D_j\nabla f(y_j)$ $\wedge$ $D_{j+1}=D_j+p_jq_j^T/p_j^Tq_j-D_jp_jq_j^TD_j/p_j^TD_jq_j$ donde $p_j=\lambda_jd_j$ $q_j=\nabla f(y_{j+1})-\nabla f(y_j)$. Si $\nabla f(y_j)\neq 0$ $\forall j=1,...,n$ luego las direcciones $d_1,...,d_n$ son H-conjugadas y $D_{n+1}=H^{-1}$. Además y_{n+1} es una solución óptima del problema.

Conclusiones del teorema XI.10 respecto del Método de (D.F.P.): Si la matriz inicialmente propuesta es $D_1=I$, luego $d_1= -\nabla f(y_1)$ con lo cual la primer búsqueda es una *búsqueda de Cauchy*, con lo cual el método estaría aprovechando en las primeras búsquedas las ventajas del método de Cauchy, esto es: rápida aproximación al óptimo.

Por otra parte, la matriz $D_{n+1}=H^{-1}$, lo cual indica que la etapa final es equivalente a la del método de Newton, con lo cual estaría aprovechando las propiedades de convergencia de éste en las proximidades del óptimo. Por esta razón se lo denomina *casi newtoniano*.

Si la f(x) es cuadrática, el método de (D.F.P.) genera direcciones conjugadas, por lo tanto, del Teorema XI.9 se tiene que el método convergerá en una iteración.

ALGORITMO

Etapa Inicial

Elegir $\varepsilon>0$, x_1 y una matriz inicial D_1 positiva definida y simétrica.

Hacer $y_1=x_1$ y $k=j=1$

Etapa principal

Mientras $\|\nabla f(y_j)\|\geq\varepsilon$ y $j\leq n$

Hacer $d_j= -D_j\nabla f(y_j)$

Calcular λ_j solución óptima de $\min_\lambda f(x_j+\lambda d_j)$ / $\lambda\in E^1$

Hacer $y_{j+1}=y_j+\lambda_jd_j$

Hacer $p_j=\lambda_jd_j$ $\qquad q_j=\nabla f(y_{j+1})-\nabla f(y_j)$

Hacer $D_{j+1}=D_j + p_jq_j^T/ p_j^Tq_j - D_j\, p_jq_j^TD_j / p_j^T D_j\, q_j$

Hacer $j=j+1$

Fin Mientras

Ejemplo: Resolución aplicando el método DFP desde el punto x=(-1; -1) del siguiente NLP no condicionado:

$$\min F(x) \quad 1/2\, x^T A\, x \qquad A = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \qquad \Rightarrow \quad \min F(x_1;\, x_2) = \frac{3}{2} x_1^2 + \frac{x_2^2}{2} - x_2\, x_1$$

Iteración 1:

$$k = 1 \qquad y_1 = x_1 = (-1;-1) \qquad\qquad D_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$d_1 = -\nabla f(y_1) = -[3\,x_1 - x_2\, ;\, x_2 - x_1]^T_{(-1;-1)} \quad \Rightarrow \; d_1 = (2;0)^T$$

Búsqueda univariable:

$$\min f(y_1 + \lambda\, d_1) = \min f\,[(-1;-1) + \lambda\,(2;0)] = \min\; f[-1 + 2\,\lambda; -1]$$

Reemplazando en f(x) se tiene:

$$\min F(\lambda) = \frac{3}{2}(-1 + 2\,\lambda)^2 + \frac{(-1)^2}{2} + (-1 + 2\,\lambda).(1)$$

Aplicando condición necesaria y suficiente de óptimo local no condicionado:

$$\frac{dF(\lambda)}{d\lambda} = 6(-1 + 2\,\lambda) + 2 = -6 + 12\,\lambda + 2 = 12\,\lambda - 4 = 0 \qquad \Rightarrow \lambda_1^* = 1/3$$

$$y_2 = y_1 + \lambda_1^*\, d_1 = (-1;-1) + 1/3\,(2;0) = (-0.333;-1)$$

Se actualiza la matriz D:

$$D_2 = D_1 + \frac{P_1\, P_1^T}{P_1^T\, q_1} - \frac{D_1\, q_1 q_1^T}{q_1^T\, D_1\, q_1} \quad P_1 = \lambda_1\, d_1 \qquad q_1 = \nabla f(y_2) - \nabla f(y_1)$$

$$D_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\begin{pmatrix} 2/3 \\ 0 \end{pmatrix}(2/3;0)}{(2/3;0)\begin{pmatrix} 2.001 \\ -0.667 \end{pmatrix}} - \frac{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{pmatrix} 2.001 \\ -0.667 \end{pmatrix}(2.001;-0.667)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}{(2.001;-0.667)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{pmatrix} 2.001 \\ -0.667 \end{pmatrix}}$$

$$D_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.333 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0.900 & -0.299 \\ -0.299 & 0.0099 \end{bmatrix} = \begin{bmatrix} 0.4331 & 0.299 \\ 0.299 & 0.901 \end{bmatrix}$$

ISBN: 950-42-0022-2 - ©1999

Luego, la segunda dirección de búsqueda viene dada por:

$$d_2 = -D_2 \nabla f(y_2) = \begin{bmatrix} 0.4331 & 0.299 \\ 0.299 & 0.901 \end{bmatrix} \begin{pmatrix} 1.10^{-3} \\ -0.667 \end{pmatrix} = \begin{pmatrix} 0.199 \\ 0.600 \end{pmatrix}$$

Búsqueda univariable en la dirección d_2:

$$\min f(y_2 + \lambda d_2) = \min f\left[\begin{pmatrix} -0.333 \\ -1 \end{pmatrix} + \lambda \begin{pmatrix} 0.199 \\ 0.600 \end{pmatrix}\right] = \min f\begin{bmatrix} -0.333 + 0.20\lambda \\ -1 + 0.60\lambda \end{bmatrix}$$

Reemplazando en f(x) se tiene:

$$F(\lambda) = 3/2(-0.333 + 0.20\lambda)^2 + 1/2(-1 + 0.60\lambda)^2 - [(-0.333 + 0.20\lambda)(-1 + 0.60\lambda)]$$

Aplicando condición necesaria y suficiente de óptimo local no condicionado:

$$\frac{dF(\lambda)}{d\lambda} = -0.2 + 0.12\lambda - 0.6 + 0.36\lambda - 0.333 + 0.40\lambda - 0.12\lambda^2 = 0 \qquad \Rightarrow \lambda^* = 1.6658$$

Luego, el nuevo punto será:

$$y_3 = y_2 + \lambda_2 d_2 = (-0.333;1) + 1.6658(0.2;0.6) = (0;0)$$

Se evalúa la condición necesaria de óptimo local no condicionado en y_3:

$$\nabla f(y_3) = (0;\ 0)^T \quad \Rightarrow \quad x_2 = y_3 = (0;\ 0) \quad \textit{es mínimo de } f(x)$$

Dado que f(x) es convexa, y_3 verifica también la condición suficiente de óptimo local no condicionado. Por lo tanto, el óptimo fue alcanzado en una sola iteración.

Las propiedades de este algoritmo son derivadas bajo la hipótesis de función cuadrática. Cuando la función no es cuadrática, puede aproximarse la misma a una función cuadrática y ser resuelta con este algoritmo, o bien, puede optimizarse la función original, en cuyo caso el número de iteraciones requerido será superior al indicado para funciones cuadráticas. En este caso el algoritmo no tiene ventajas comparativas con los anteriores.

XI.3.2.2.3 Algoritmos para NLP con Restricciones

En esta sección se describen algunos de los algoritmos desarrollados para

ISBN: 950-42-0022-2 - ©1999

resolver programas matemáticos no lineales con restricciones. Se han utilizado muchas estrategias como base de métodos para buscar el óptimo de este tipo de problemas de optimización. Los más conocidos son los métodos de penalización y barreras, los métodos de planos de corte y dual y los métodos de direcciones factibles. De estas tres estrategias, los métodos de direcciones factibles han dado origen a algoritmos muy difundidos y utilizados en la resolución de este tipo de problemas. La mayoría de los softwares comerciales incorporan estos algoritmos. En lo que sigue se presenta la estrategia de direcciones factibles y dos de los algoritmos más representativos.

Métodos de Direcciones Factibles y de Mejora

Dado el NLP min f(x) sa: x0S. Estos algoritmos resuelven este tipo de NLP desplazándose desde un punto factible (que pertenece a la región factible) a otro punto factible *no inferior* a través de la siguiente estrategia:

Estrategia:

Dado un punto x_k *factible*,

(i) Determinar una dirección d_k tal que para $\delta>0$ y suficientemente pequeño, las siguientes dos propiedades se cumplan:

$x_k+\lambda d_k$ es factible □ $\lambda 0(0, \delta)$

$f(x_k+\lambda d_k) < f(x_k)$ □ $\lambda 0(0, \delta)$ (para problemas de minimización)

(ii) Resolver el problema: min $f(x_k+\lambda d_k)$ s.a:$\lambda 0(0,\delta)$ / $x_{K+1}=(x_k+\lambda_k d_k)0S$

Dirección de Mejora, Factible: Sea f:$R^n 6 R^1$ y sea S un conjunto no vacío en R^n, considere el problema siguiente: min f(x) sa: x0S. Un vector d□0 define una *dirección factible* en x0S, si □ un $\delta>0$ / $(x+\lambda d)0S$ □ $\lambda 0(0,\delta)$. Además, d es una *dirección factible y de mejora* en S si □ $\delta>0$ / $f(x+\lambda d)<f(x)$ ϖ $(x+\lambda d)0S$ □ $\lambda 0(0,\delta)$.

Teorema XI.11: (*Caso de Restricciones Lineales*:) Sea $x0R^n$ una solución factible del problema: min f(x) s.a:{Ax#b; Ex=e; donde A(mxn); E(lxn)} / $A_1x=b_1$ ϖ $A_2x<b_2$ siendo $A^T=(A_1{}^T,A_2{}^T)$ ϖ $b^T=(b_1{}^T,b_2{}^T)$. Luego, un vector d□0 es *una dirección factible* en x si y solo si $A_1d\#0$ ϖ Ed=0. Por otra parte, $d0R^n$ es *una dirección de mejora* en x si $\Lambda f(x)d<0$. (para problema de minimización). La intersección del conjunto de direcciones factibles con el conjunto de direcciones de mejora, define el conjunto de *direcciones factibles y de mejora*.

Método del Gradiente Reducido de Wolfe (G.R.): Se basa en reducir la dimensión del dominio del problema representando todas las variables en término de un subconjunto de variables independientes. Sea el NLP min f(x) s.a:{Ax=b; x∃0;

$x0R^n\}$ donde A(mxn) es una matriz de rango *m* y f(x) es una función continua y diferenciable sobre R^n. Se considera además, que cualesquiera m columnas de A son linealmente independientes y todo punto extremo de la región factible tiene m variables > 0 (no degenerado). Con esta consideración, toda solución factible tiene como mínimo m variables > 0 y como máximo (n-m) variables iguales a cero. Siguiendo la estrategia general formulada arriba, en primer lugar el método debe proponer una estrategia para calcular una dirección factible y de mejora:

(i) *Cálculo de una dirección factible y de mejora*:

Con las consideraciones anteriores, A puede descomponerse en: A=(B,N) $\wedge$ $x=(x_B^T; x_N^T)^T$, donde B(mxm) es una matriz no singular, $x_B>0$ ϖ $x_N \geq 0$. Sea $\nabla f(x)=(\nabla_B f(x); \nabla \Lambda_N f(x))$ donde: $\nabla_B f(x)$ es el gradiente de f(x) asociado a x_B $\wedge$ $\nabla_N f(x)$ es el gradiente de f(x) asociado a x_N .

Por el Teorema XI.11 $d0R^n$ es una dirección:

factible en x si $Ad=0 \;\wedge\; d_j \geq 0 \;\;\forall\; x_j=0$, y

de mejora en x si $\nabla^T f(x)d<0$.

Análisis de la condición de factibilidad:

Siendo $d=(d_B^T; d_N^T)^T$ resulta:

$$Ad = (B,N)(d_B^T,d_N^T)^T= Bd_B+Nd_N=0.$$

Si B es no singular, su inversa $\exists$, luego, se obtiene: $d_B=-B^{-1}Nd_N$.

Luego, la condición de factibilidad se reduce a:

$$d_B= -B^{-1}Nd_N \;\wedge\; d_j \geq 0 \;\;\forall x_j=0,$$

Análisis de la condición de mejora:

$$\nabla f(x)d = (\nabla_B f(x),\; \nabla_N f(x))\, (d_B^T,d_N^T)^T = \nabla_B f(x)d_B+ \nabla_N f(x)\, d_N$$

Reemplazando el valor obtenido de d_B que verifica la condición de factibilidad se tiene:

$$\nabla f(x)d = \nabla_N f(x)d_N - \nabla_B f(x)B^{-1}Nd_N<0.$$

Luego, definiendo a $r_N = \nabla_N f(x)- \nabla_B f(x)B^{-1}N$ como el *Gradiente Reducido*, la condición de mejora queda formulada como:

$$\nabla f(x)^T d = r_N^T d_N<0$$

Por lo tanto, para que $d0R^n$ sea una dirección factible y de mejora en x, d_N debe elegirse tal que se verifique:

$$r_N^T d_N<0, \quad d_{Nj} \geq 0 \quad \forall\; x_{Nj} =0$$

Una regla práctica que permite obtener un d_N verificando estas condiciones es:

Si $r_{Nj} \leq 0$ asignar $d_{Nj} = - r_{Nj}$.

Si No, asignar $d_{Nj} = - x_{Nj}\, r_{Nj}$.

Obtenido d_N, d_B se calcula de manera tal que verifique la restante condición de factibilidad. Esto es:

$$d_B = -B^{-1}Nd_N.$$

El $d=(d_B^T; d_N^T)^T$ así obtenido será una *dirección factible y de mejora en x* para el problema considerado.

Una vez calculada la dirección factible y de mejora, siguiendo la estrategia general es necesario realizar la búsqueda univariable:

(ii) *Búsqueda Univariable*:

Sea $x_k \in E^n$ un punto factible asociado a la iteración k y $d_k \in R^n$ una dirección factible y de mejora, el siguiente punto está dado por $x_{k+1}= x_k+\lambda_k d_k$ donde λ_k se obtiene de resolver el siguiente problema univariable:

$$\min f(x_k+\lambda d_k) \quad \text{s.a:}\{ \text{(a) } A(x_k+\lambda d_k)=b; \quad \text{(b) } x_k+\lambda d_k\geq 0 \ ; \lambda>0\}$$

Suponiendo que en x_k el conjunto de restricciones toma los siguientes valores: $Ax_k=b$; $x_{Bk}>0$ y $x_{Nk} \geq 0$. Luego, por el Teorema XI.11 se tiene: $Ad_k=0$ y $x_{Nk}d_{Nk}\geq 0$ por ser d_k una dirección factible en x_k. Con esto, las restricciones resultarán:

(a) $Ax_k+\lambda_k Ad_k=b \rightarrow$ restricciones redundantes, y

(b) $x_k+\lambda_k d_k\geq 0 \qquad \rightarrow \lambda_k d_k\geq -x_k$.

De este modo, el problema univariable se reduce a:

$$\min f(x_k+\lambda d_k) \quad \text{s.a:}\{0<\lambda\leq\lambda_{MAX}\}$$

donde: $$\lambda_{MAX} = \text{Sup}\{\lambda \ / \ \lambda_k d_k\geq -x_k \}$$

Resulta así un problema de optimización univariable con un intervalo de incertidumbre definido. El mismo puede ser resuelto utilizando algún algoritmo de optimización univariable tal como el método de la Sección Dorada.

ALGORITMO

Dado el NLP $\min f(x)$ $x \in R^n$ s.a.:$\{Ax=b, x\geq 0 \}$

Etapa Inicial

Encontrar un punto x_1 factible. Hacer k=1, definir $\varepsilon> 0$, M suficientemente grande

Hacer I_k={índices de las m variables más positivas de x_k}

Hacer $B=(a_j \ / \ j \in I_k)$ (matriz asociada a las variables x_B)

Hacer $N=(a_j \ / \ j \notin I_k)$ (matriz asociada a las variables x_N)

Hacer $\Gamma_N = \nabla_N f(x_k)- \nabla_B f(x_k)B^{-1}N$

Etapa principal

Mientras $\|\Gamma_N\| >\varepsilon$

Hacer I_k={índices de las m variables más positivas de x_k}

Hacer $B=(a_j \ / \ j \in I_k)$ (matriz asociada a las variables x_B)

Hacer $N=(a_j \ / \ j \notin I_k)$ (matriz asociada a las variables x_N)

Hacer $\Gamma_N = \nabla_N f(x_k)- \nabla_B f(x_k)B^{-1}N$

Para todo $j \in I_k$

ISBN: 950-42-0022-2 - ©1999

Si $r_{Nj} \leq 0$
asignar $d_{Nj} = - r_{Nj}$.
Si No
asignar $d_{Nj} = - x_{Nj}\, r_{Nj}$.
Fin Si
Fin Para
Hacer $d_B = -B^{-1} N d_N$.
Hacer $d_k = (d_B^T, d_N^T)^T$
Hacer $\lambda_{MAX} = M$, $i=1$
Mientras $i \leq n$
Si $d_{ki} > 0$
asignar $\lambda_{MAX} = \text{mínimo}\{ \lambda_{MAX}, - x_{kj} / d_{kj} \}$
Fin Si
Hacer $i=i+1$
Fin Mientras
Calcular λ_k solución óptima de $\min_\lambda f(x_k+\lambda d_k)$ s.a.: $0 \leq \lambda \leq \lambda_{MAX}$
Hacer $x_{k+1}=x_k+\lambda_k d_k$
Hacer $k=k+1$
Fin Mientras

Ejemplo: Resolución del siguiente NLP usando el algoritmo de Wolfe.

$$\min f(x) = 2 x_1^2 +2 x_2^2 - 2 x_1 x_2 - 4 x_1 - 6 x_2$$

sujeto a:

$$x_1 + x_2 \leq 2$$
$$x_1 + 5x_2 \leq 5$$
$$x_1;\ x_2 \geq 0$$

Se transforma el problema al formato base del algoritmo:

$$\min f(x) = 2 x_1^2 + 2 x_2^2 - 2 x_1 x_2 - 4 x_1 - 6 x_2 + 0 x_3 + 0 x_4$$

s.a.:

$$x_1 + x_2 \leq 2$$
$$x_1 + 5x_2 \leq 5$$
$$x_1;\ x_2 \geq 0$$

Punto Inicial $x_1 = (0; 0; 2; 5)^T$. m = 2 (2 restricciones de =) → 2 variables básicas.

Iteración 1:

Se eligen las dos variables más positivas como x_B → $I_1 = \{3, 4\}$

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 5 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad N = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}$$

Cálculo del gradiente reducido $r_N = \nabla_N f(x_k) - \nabla_B f(x_k) B^{-1} N$:

$$\nabla_N^T f(x_1) = [(4x_1 - 2x_2 - 4);(4x_2 - 2x_1 - 6)]_{(0,0,2,5)} = (-4,-6)$$

$$r_N^T = (-4,-6) - (0,0).\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}$$

Luego, el gradiente reducido es: $r_N^T = (-4, -6)$

Aplicando la Regla se determina la componente no básica de la dirección como:
$d_N = (d_1; d_2) = (4; 6)$

La componente básica de la dirección se obtiene como: $d_B = -B^{-1}Nd_N$

$$d_B = (d_3; d_4) = .\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}\begin{pmatrix} 4 \\ 6 \end{pmatrix} = (-10; -34)$$

Luego, la dirección factible y de mejora obtenida es $d = (d_1;d_2;d_3;d_4) = (4;6;-10;-34)$

Ahora se resuelve el problema de búsqueda lineal:

$$\min f(x_1 + \lambda d_1) \qquad \text{sa. } 0 \le \lambda \le \lambda_{\text{máx}}$$

Cálculo de $\lambda_{\text{máx}}$:

$$\lambda_{\max} = \min[-2/-10; -5/-34] = 5/34$$

Reemplazando x_1 y d_1 se tiene:

$$\min f[(0;0;0;5) + \lambda(4;6;-10;-34)] = f(4\lambda; 6\lambda; 2-10\lambda; 5-34\lambda)$$

s.a.:

$$0 \le \lambda \le 5/34$$

Reemplazando en f(x) se tiene:

$$\text{mín } F(\lambda) = (56\lambda^2 - 52\lambda)$$

ISBN: 950-42-0022-2 - ©1999

s.a.: $0 \leq \lambda \leq 5/34$

Para un resultado académico, se resuelve aplicando condición necesaria de óptimo local no condicionado a $F(\lambda)$, esto es:

$$\frac{dF}{d\lambda} = 112\lambda - 52 = 0 \qquad \rightarrow \qquad \lambda = 58/112$$

El mínimo del problema irrestricto es λ= 52/112 el cual cae fuera del intervalo factible, por lo tanto se adopta como mínimo del problema el límite superior del intervalo: λ_1^*=5/34.

El nuevo punto será:

$$x_2 = x_1 + \lambda_1^* d_1 = (0;0;2;5) + \frac{5}{34}(4;6;-10;-4) = \left(\frac{10}{17};\frac{15}{17};\frac{9}{17};0\right)$$

Iteración 2:

Se eligen las dos variables más positivas como x_B → $I_2 = \{1, 2\}$

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix} \quad N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Se calcula el gradiente de la función en x_2:

$$\nabla f(x_2) = \left(-\frac{58}{17};-\frac{62}{17};0;0\right)$$

Cálculo del gradiente reducido $r_N = \nabla_N f(x_k) - \nabla_B f(x_k) B^{-1} N$:

$$r_N = (0;0) - \left(-\frac{58}{17};-\frac{62}{17}\right)\begin{bmatrix} 5/4 & -1/4 \\ -1/4 & 1/4 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \left(\frac{57}{17};\frac{1}{17}\right)$$

Aplicando la Regla se determina la componente no básica de la dirección como:

$$d_N = [d_3;d_4] = [-x_3\, r_{N3};-x_4\, r_{N4}] = [-513/289;0]$$

La componente básica de la dirección se obtiene como: $d_B = -B^{-1} N d_N$

ISBN: 950-42-0022-2 - ©1999

$$d_B^T = [d_1; d_2] = -\begin{bmatrix} 5/4 & -1/4 \\ -1/4 & 1/4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} -513/289 \\ 0 \end{pmatrix} = (2.565/1.156; -513/1.156)$$

Luego, la dirección factible y de mejora obtenida es:

$$d_2 = (d_1; d_2; d_3; d_4) = \left(\frac{2.565}{1.156}; -\frac{513}{1.156}; -\frac{513}{289}; 0 \right)$$

Ahora se resuelve el problema de búsqueda lineal:

$$\min f(x_2 + \lambda\, d_2)$$
$$\text{sa. } 0 \leq \lambda \leq \lambda_{\text{máx}}$$

Cálculo de $\lambda_{\text{máx}}$:

$$\lambda_{máx} = \min[-x_2/-d_2; -x_3/d_3] = \min[-151.156/-17513; -9289/-17513] = 17/57$$

Reemplazando x_2 y d_2 se tiene:

$$\text{mín } F(\lambda) = (12{,}21\, \lambda^2 - 5{,}95\, \lambda - 6{,}436)$$
$$\text{sa.: } 0 \leq \lambda \leq 17/57$$

Nuevamente, se resuelve aplicando condición necesaria de óptimo local no condicionado a F(λ), esto es:

$$\frac{dF}{d\lambda} = 24.42\,\lambda - 5.95 = 0 \qquad \rightarrow \qquad \lambda_2^* = 5.95/24.42$$

El mínimo del problema irrestricto es λ= 5,95/24,42 el cual cae dentro del intervalo

$$x_3 = x_2 + \lambda_2^*\, d_2 = \left[\frac{35}{31}; \frac{24}{31}; \frac{3}{31}; 0 \right]$$

factible, por lo tanto se adopta como mínimo del problema. El nuevo punto será entonces:

$$x_3 = x_2 + \lambda_2^*\, d_2 = \left[\frac{35}{31}; \frac{24}{31}; \frac{3}{31}; 0 \right]$$

Iteración 3:

Se eligen las dos variables más positivas como x_B → $I_3 = \{1, 2\}$

$$B = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix} \quad N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Se calcula el gradiente de la función en x_3:

ISBN: 950-42-0022-2 - ©1999

$$\nabla f(x_3) = \left(-\frac{32}{31}; -\frac{160}{31}; 0; 0 \right)$$

Cálculo del gradiente reducido $r_N = \nabla_N f(x_k) - \nabla_B f(x_k) B^{-1} N$:

$$r_N = (0;0) - \left(-\frac{32}{31}; -\frac{160}{31} \right) \begin{bmatrix} 5/4 & -1/4 \\ -1/4 & 1/4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \left(0; \frac{32}{31} \right)$$

Aplicando la Regla se determina la componente no básica de la dirección como:

$$d_N = [d_3; d_4] = [-x_3 r_{N3}; -x_4 r_{N4}] = [0;0]$$

La componente básica de la dirección se obtiene como: $d_B = -B^{-1} N d_N$
Dado que $d_N = (0; 0)$ resulta $d_B = (0; 0)$. Por lo tanto se obtiene que
$d_3 = (0; 0; 0; 0)$ lo cual implica que en x_3 no existe una dirección factible y de mejora. Esto nos indica que x_3 es un punto de Kuhn Tucker (Teorema XI.6).

Método del Gradiente Reducido Generalizado (GRG): La estrategia del *gradiente reducido* del algoritmo de Wolfe puede ser extendida para resolver NLP en general (con restricciones no lineales). En esta sección se describe una estrategia que constituye la base del método GRG. Para ello se considera el NLP:

$$\min f(x) \quad \text{s.a:} \{h_k(x)=0 \ \square\ k=1,...,m\}$$

Dado que las restricciones son no lineales, para representar las m variables dependientes x_B en función de las (n-m) variables independientes x_N se realiza una linealización de cada una de las restricciones en el punto factible actual mediante una expansión de Taylor de primer orden. Sea $x_1 \in R^n$ un punto factible del problema, la aproximación lineal en x_1 vendrá dada por $H_k(x,x_1)$. $\nabla^T h_k(x_1)(x-x_1) \in k=1,...,m$. Luego se utiliza la aproximación lineal para generar un nuevo punto x / $\nabla^T h_k(x_1)(x-x_1)=0 \in k=1,...,m$.

Siendo $x^T=(x_B^T, x_N^T)$, definiendo:

$$J^T(x_1) = (\nabla_B h_1(x_1),\ \nabla_B h_k(x_1),\ \nabla_B h_m(x_1))$$
$$C^T(x_1) = (\nabla_N h_1(x_1),\ \nabla_N h_k(x_1),\ \nabla_N h_m(x_1))$$

Suponiendo que $J(x_1)^{-1}\ \exists$, el nuevo punto verificará: $(x_B - x_{B1}) = -J(x_1)^{-1} C(x_1)(x_N - x_{N1})$.
Definiendo: $d_B = (x_B - x_{B1}) \ \wedge\ d_N = (x_N - x_{N1})$ la expresión anterior resulta en:

$$d_B = -J(x_1)^{-1} C(x_1) d_N.$$

Reemplazando en la función objetivo x_B por su relación con x_N utilizando la

ISBN: 950-42-0022-2 - 81999

aproximación lineal resulta:

$$f((x_{B,} x_N) = f((x_{B1}\text{-}J(x_1)^{-1}C(x_1)(x_N\text{-}x_{N1})), x_N)=F(x_N)$$

De esta manera se ha transformado el problema restringido en un problema irrestricto aproximado: min $F(x_N)$ $x_N \in R^{n-m}$, definido en el dominio reducido R^{n-m}. Usando la regla de la cadena para derivación de funciones implícitas se obtiene el gradiente de la función en el espacio reducido como:

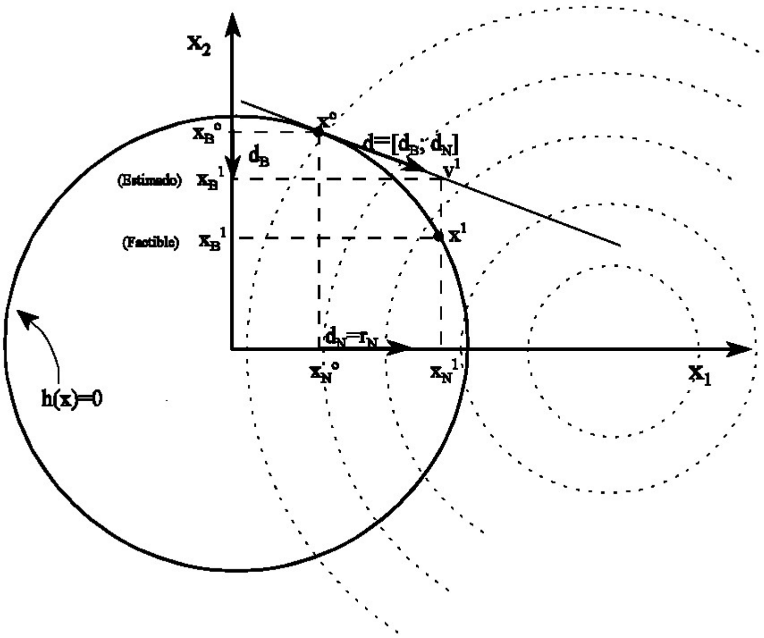
$$\nabla_N F(x) = \nabla_N f(x) - \nabla_B f(x) J(x_1)^{-1} C(x_1).$$

El vector $\nabla_N F(x)$ se denomina *Gradiente Reducido Generalizado*.

Asignando arbitrariamente, en el punto x_k actual, $d_N = - \nabla_N F(x_k)$ y luego calculando el valor de $d_B= -J(x_k)^{-1}C(x_k)d_N$, la dirección de búsqueda $d=(d_B,d_N) \in R^n$ será una *dirección descendente* y en general *no factible* para toda restricción $h_k(x)= 0$ no lineal.

Dado x^o factible, la linealización de $h(x)$ en x^o genera una recta tangente a la curva en x^o.

Se representan estos conceptos en el siguiente gráfico:



El punto $v^1= x^o + \lambda d$ alcanzado mediante un desplazamiento λ en la dirección d desde x^o es *no factible*; $h(v^1)\neq 0$. v_1 resulta no factible porque para una variación elegida en el valor de las variables independientes (x_N-x_N^o), el cambio en el valor de las variables dependientes (x_B-x_B^o) calculado usando la aproximación lineal no es suficientemente exacto como para dar un valor de x_B que satisfaga la restricción $h_k(x_N,x_B) = 0 \;\forall\;$ k=1,...,m. Esto se puede solucionar mediante la siguiente estrategia:

Buscar el mínimo de f(x) a lo largo de la *curva implícitamente definida* por los valores de λ y x_B que satisfacen las ecuaciones: $h_k(x_N+\lambda d_N,\ x_B+\lambda d_B) = 0 \;\forall$ k=1,...,m.

Luego, para cada valor de λ (en la búsqueda del óptimo), será necesario resolver el sistema de ecuaciones anterior para obtener el valor de x_B factible correspondiente. Este sistema se resuelve mediante el método de Newton:

$x_B^{i+1} = x_B^i - J^{-1}(x^i)h(x^i)$, donde $h^T(x^i)=(h_1(x^i),..\ h_k(x^i),..\ h_m(x^i))$ y $x^i = (x_N+\lambda d_N,\ x_B^i+\lambda d_B)$.

De este modo se puede obtener un nuevo punto x^1 factible que mejora el valor de la función objetivo.

ALGORITMO

Dado el NLP min f(x) $x \in R^n$ s.a.:{$h_l(x)=0,\ \forall\ l=1,\ldots m$ siendo $m<n$}

Etapa Inicial

Encontrar un punto x_1 factible. Hacer k=1, definir $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\varepsilon_3 > 0$, $\lambda=\lambda_0$, $0<\gamma<1$, MaxIter (número admitido de iteraciones sin converger)

Elegir una partición de x en x_B , x_N tal que $J(x_k)$ sea no singular

Calcular $r_N = \nabla_N f(x_k) - \nabla_B f(x_k) J^{-1}(x_k) C(x_k)$

Etapa principal

Mientras $\|r_N\| \geq \varepsilon_1$
 asignar $d_N = - r_N$
 Hacer $d_B = - J^{-1}(x_k)C(x_k)d_N$.
 Hacer $d_k = (d_B^T, d_N^T)^T$
 Hacer $\lambda=\lambda_0$, i=1, $v^{(i)}= x_k+\lambda d_k$
 Mientras $f(x_k)\leq f(v^{(i)})$ y $\exists\ \|h_l(v^{(i)})\| > \varepsilon_2,\ \forall\ l=1,\Upsilon m$
  Mientras $\exists\ \|h_l(v^{(i)})\| > \varepsilon_2,\ \forall l=1,\Upsilon m$
   Hacer: $v_B^{(i+1)}=v_B^{(i)}-J^{-1}(v^{(i)})h(v^{(i)})$ $\{h^T(v^{(i)})=(h_l(v^{(i)}))\}$
   Hacer $v_N^{(i+1)} = v_N^{(i)}$
   Mientras $\|v_B^{(i+1)}-v_B^{(i)}\| \geq \varepsilon_3$ o i $\leq$MaxIter
    Hacer i=i+1
    Hacer $v_B^{(i+1)} = v_B^{(i)}-J^{-1}(v^{(i)})h(v^{(i)})$ $\{h^T(v^{(i)})=(h_l(v^{(i)}))\}$
    Hacer $v_N^{(i+1)} = v_N^{(i)}$
   Fin Mientras
   Si $\exists\ \|h_l(v^{(i)})\| \geq \varepsilon_2$, $\forall l=1,\Upsilon m$
    Hacer $\lambda=\lambda\gamma$
    Hacer i=1, $v^{(i)}= x_k+\lambda d_k$
   Fin Si
  Fin Mientras
  Si $f(x_k)\leq f(v^{(i+1)})$

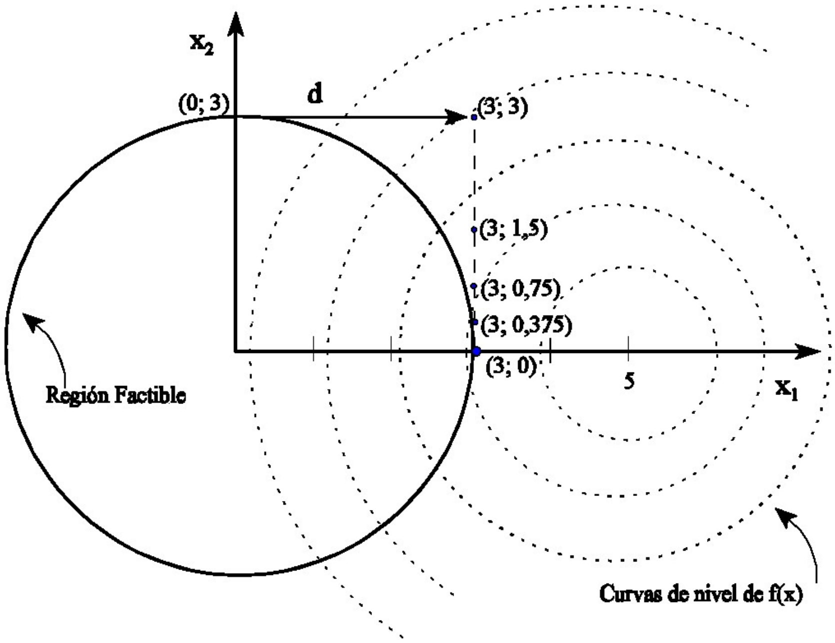
ISBN: 950-42-0022-2 - ©1999

Hacer $\lambda=\lambda\gamma$
Hacer $i=1$, $v^{(i)}= x_k+\lambda d_k$
Fin Si
Fin Mientras
Hacer $x_{k+1}=x_k+\lambda_k d_k$
Hacer $k=k+1$
Calcular $r_N = \nabla_N f(x_k) - \nabla_B f(x_k) J^{-1}(x_k) C(x_k)$
Fin Mientras

Ejemplo: Se resuelve utilizando el algoritmo GRG, el siguiente NLP:

$$\min f(x) = (x_1 - 5)^2 + x_2^2$$
$$\text{sa.:} \quad x_1^2 + x_2^2 = 9$$

Representación gráfica:



Iteración 1:

ISBN: 950-42-0022-2 - ©1999

Se elige arbitrariamente la partición $x_B = x_2$ y $x_N = x_1$

$$J = \left[\frac{\partial h}{\partial x_2}\right] = 2x_2 \quad J^{-1} = \frac{1}{2x_2} \quad J^{-1}_{(x^0)} = \frac{1}{6} \quad C = \left[\frac{\partial h}{\partial x_1}\right] = 2x_1 \quad C_{(x^0)} = 0$$

Cálculo del gradiente reducido:

$$r_N(x^0) = \nabla_{x_N} f(x^0) - \nabla_{x_B} f(x^0).J^{-1}(x^0).C(x^0) = \left[2(x_1 - 5) - 2x_2.\frac{1}{6}.0\right]_{x^0} = -10$$

Luego, la componente no básica de la dirección de búsqueda se elige como:
$d_N = -r_N = 10$
La componente básica de la dirección de búsqueda se calcula como: $d_B = -J(x^0)^{-1}C(x^0)d_N$
$d_B = - 1/6 . 0 . 10 = 0$
Luego, la dirección generada será: $d = (d_1; d_2) = (10; 0)$
Un desplazamiento λ desde x^0 en la dirección d genera un nuevo punto v^1

$$v^1 = x^0 + \lambda\, d = (0;3) + \lambda\,(10;0)$$

Se propone arbitrariamente $\lambda=3/10$, con lo cual el nuevo punto será $v^1=(3;3)$.
Se verifica si el punto $v^1=(3;3)$ es factible:

$$\| h(v^1) - b \| = \| 3^2 + 3^2 - 9 \| = 9 > \varepsilon_2$$

Luego, $v^1=(3;3)$ es no factible. Dado que v_N^1 es una variable independiente su valor pudo ser fijado arbitrariamente, por lo tanto se considera que la no factibilidad es causada por la variable dependiente v_B^1 la cual fue estimada usando la aproximación lineal de la restricción. Por esta razón se intentará modificar el valor de v_B^1 hasta un valor tal que haga cero (verifique) la restricción. Para ello se utiliza el método de Newton iterando sobre v_B^1.

i=1

$$v_B^2 = v_B^1 - j^{-1}_{(v^1)}.h_{(v^1)} \quad \Rightarrow \quad v_B^2 = 3 - (1/6)9 = 1.5$$
$$\|v_B^2 - v_B^1\| = \|3 - 15\| = 1.5 > \varepsilon_3$$

i=2

$$v_B^3 = 1.5 - 1/3\left[3^2 + (3/2)^2 - 9\right] = 0.75$$
$$\|v_B^3 - v_B^2\| = \|3/4 - 3/2\| = 0.75 > \varepsilon_3$$

i=3

ISBN: 950-42-0022-2 - ©1999

$$v_B^4 = 0.75 - 2/3\left[3^2 + 0.75^2 - 9\right] = 0.375$$

Se puede ver que v_B^{i+1} 6 0, luego se obtiene que $v^i = (3;0)$.
Se evalúa la función objetivo:

$$f\left(x^0\right) = 34 \quad y \quad f\left(v^i\right) = 4 \quad \Rightarrow \quad f\left(x^0\right) > f\left(x^i\right) \;\Rightarrow\; \textit{mejoró el valor de la función}$$

Luego, $x^1 = v^i = (3;0)$

Iteración 2:
Se elige arbitrariamente la partición $x_B = x_1$ y $x_N = x_2$

$$J = \left[\frac{\partial h}{\partial x_1}\right] = 2x_1 \quad J^{-1} = \frac{1}{2x_1} \quad J^{-1}_{\left(x^0\right)} = \frac{1}{6} \quad C = \left[\frac{\partial h}{\partial x_2}\right] = 2x_2 \quad C_{\left(x^1\right)} = 0$$

Cálculo del gradiente reducido:

$$r_N\left(x^1\right) = \nabla_{x_N} f\left(x^1\right) - \nabla_{x_B} f\left(x^1\right).J^{-1}\left(x^1\right).C\left(x^1\right) = \left[2\,x_2 - 2\left(x_1 - 5\right)\left(1/6\right).0\right]_{x^1} = 0$$

$r_N = 0 < \varepsilon_1$ → $x^1 = (3;0)$ verifica la condición necesaria de mínimo de Kuhn - Tucker

Finalmente, es necesario destacar que el objetivo de este capítulo ha sido realizar una descripción conceptual de la teoría y los algoritmos de optimización. Por esta razón su alcance se ha limitado a presentar los principales enunciados teóricos y los algoritmos básicos, apuntando así a dar una visión global con cierto detalle del concepto de optimización.

El lector que desee estudiar en profundidad la teoría y los algoritmos de optimización deberá recurrir a textos especializados. Algunos de los más convencionales se listan como referencia del presente capítulo.

EJERCICIOS PROPUESTOS

EJ.1: Una empresa industrial elabora 5 productos diferentes, designados genéricamente como Pj j=1,...5. Los productos son elaborados a partir de 4 materias primas diferentes designadas como MPi i=1,...4. En la tabla siguiente se especifican los precios de venta de cada uno de los productos, los costos de las materias primas, y las unidades de cada tipo de materia prima requerida para producir una unidad del producto.

Precio de venta del producto: ($ / unidad de producto)				
P1	P2	P3	P4	P5
3	5	2	1.7	4.5

ISBN: 950-42-0022-2 - 81999

Costo de la materia prima: ($ / unidad de materia prima)			
MP1	MP2	MP3	MP4
0.7	1.5	1.2	2

ISBN: 950-42-0022-2 - ©1999

Requerimiento de materia prima: (unidades de materia prima / unidades de producto)					
	P1	P2	P3	P4	P5
MP1	1	2	1.2	0	0.5
MP2	1.3	0	1.4	0.5	0
MP3	0	0.8	0.8	0.8	2
MP4	0.2	1.2	0	0	1.8

Disponibilidad de materia prima (unidades)			
MP1	MP2	MP3	MP4
250	180	300	12

Todos los costos son asignados a la materia prima, de modo tal que llegan al producto a través del requerimiento de materia prima por parte del producto.
Se requiere determinar:
a) La política de producción que maximice el ingreso bruto por ventas.
b) La política de producción que maximice el ingreso neto por ventas.
c) La política de producción que minimice el costo de las materias primas
d) Considerando ahora la información de la estimación de demanda dada en la siguiente tabla, determinar la política óptima de producción.

Estimación de demanda de producto (unidades de producto)				
P1	P2	P3	P4	P5
50	100	70	200	40

En cada uno de los casos modelar el problema y luego resolverlo usando algún optimizador disponible. (Por ejemplo pude utilizar el *solver* de la planilla Excel). Analizar y comparar los resultados de los distintos modelos.

EJ.2: Dado el LP: max $f(x) = 4\ x_1 + x_2$; s.a.: $\{x_1 + 3x_2 \leq 9$; $2\ x_1 - x_2 \leq 4$; $x_1 , x_2 \geq 0\}$
a) Resolver gráficamente
b) Resolver siguiendo el algoritmo SIMPLEX para LP.
c) Resolver usando un optimizador disponible.

ISBN: 950-42-0022-2 - ©1999

EJ.3: Dado el NLP min $f(x)= 2x_1 +x_1 x_2 -3x_1 x_2^2$
Resolver usando condición necesaria y suficiente de óptimo.

EJ.4: Dado el NLP opt $f(x) = (3-x_1)^2 +x_2^2$ s.a.:$\{x_1^2 +(2-x_2)^2 = 9 \}$
Resolver usando condición necesaria y suficiente de óptimo.

EJ.5: Dado el NLP min $f(x) = (3-x_1)^2 +x_2^2$ s.a.:$\{x_1^2 +(2-x_2)^2$ #9}
Resolver usando condición necesaria y suficiente de óptimo.
Nota: dado que la condición necesaria de óptimo fue desarrollada para un problema de maximización, es necesario tener en cuenta que min f(x) = - max - f(x)

EJ.6: Dado el NLP min $f(x) = 3x - (5- x)^2$
a) Resolver usando el método de la Sección Dorada
b) Resolver usando el método de Newton.

EJ.7: Dado el NLP min $f(x)= 2x_1 +x_1 x_2 -3x_1 x_2^2$
a) Resolver usando el método Coordenado Cíclico
b) Resolver usando el método de Hooke and Jeeves
c) Resolver usando el método de direcciones conjugadas D.F.P.

EJ.8: Dado el NLP min $f(x) = x_1^2 +(8+x_2)^2$ s.a.:$\{x_1 +3x_2 \leq 9$; $2 x_1 -x_2 \leq 4$; $x_1 , x_2 \geq 0\}$
a) Resolver usando condición necesaria y suficiente de óptimo
b) Resolver usando el algoritmo de Wolfe (GR)
c) Resolver utilizando un optimizador disponible (por ejemplo el solver de Excel)

EJ.9: Dado el NLP opt $f(x) = (2-x_1)^2 +x_2^2 - 4x_3$
s.a.:$\{x_1^2 +(2-x_2)^2 +x_3^2 = 9$; $(2-x_1)^2 +x_2^2 = 9\}$
a) Resolver usando condición necesaria y suficiente de óptimo.
b) Resolver usando el algoritmo GRG.
c) Resolver utilizando un optimizador disponible (por ejemplo el solver de Excel).

ISBN: 950-42-0022-2 - 81999

BIBLIOGRAFÍA

- Bazaraa,M.S., J.J.Jarvis & H.D. Sherali, *Linear Programinng and Network Flows.John* Wiley & Sons, New York, Second Edition (1990).
- Bazaraa,M.S., H.D. Sherali & C.M.Shetty, *Nonlinear Programinng. Theoty and Algorithms*, John Wiley & Sons, New York, Second Edition (1993).
- Budnick,F.S., R. Mojena & T.E.Vollmann, Principles of operations Reseach for Management, R.D. Irwin, Inc. (1977).
- Cohon, J.L., Multiobjective Programming and Planning, Academic press, N.Y., (1978).
- Fletcher, R., *Practical Methods of Optimization*, Vol I y II, John Wiley 7 Sons, New York (1981) .
- Halemane and Grossmann,I.E., 1983, Optimal process design under Uncertainty, AIChEJ, 29(3):425-433.
- Lasdon, L.S., optimization Theory for large Systems, The MacMillan Company, London, (1970).
- Luemberger,D.G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, (1973).
- Reklaitis G.V., & Ravindran A., *Engineering Optimization. Methods and Applications.* John Wiley & Sons, New York (1983).
- Siddall,J.N., Optimal Engineering Design, M. Dekker Inc. New York and Basel (1982).
- Simmons, D.M., *Nonlinear Programming for Operations Research*, Prentice-Hall Inc., New Jersey (1975).
- Taha,H.A., Investigación de Operaciones, 5/ED, ALFAOMEGA, (1994).
- Winston, Investigación de Operaciones, IIBEROAMERICA, (1994).

ISBN: 950-42-0022-2 - ©1999

APÉNDICE

Definiciones

El vector Gradiente: Sea f:$R^n \rightarrow R^1$ continua y diferenciable en $x \in R^n$, el vector gradiente de f en x $\exists$ y se define como:

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right]$$

e indica la dirección en la cual la función crece más rápidamente. Además, $\nabla f(x)$ es normal a toda dirección en la cual la velocidad (puntual) de cambio de la función es cero; esto es, $\nabla f(x)$ es perpendicular a toda superficie de la forma f(x)=k donde k es una constante.

La matriz Hessiana: Sea f:$R^n \rightarrow R^1$ continua y diferenciable dos veces en $x \in E^n$. Luego la matriz Hessiana de f en x $\exists$ y se define como:
Esta matriz es simétrica y única.

$$Hf(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{bmatrix}$$

Forma Cuadrática: Una forma cuadrática es cualquier f:$R^n \rightarrow R^1$ que tiene la siguiente forma:

$$q(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}\, x_i\, x_j$$

donde a_{ij} es un número real. Una forma cuadrática a diferencia de una función cuadrática, no posee términos lineales.

q(x) puede ser representada como q(x)= $x^T Ax$ donde a_{ij} de A(nxn) son los coeficientes de q(x).

Toda forma cuadrática puede representarse también como q(x)= $x^T Dx$ donde D es una matriz simétrica cuyos elementos son : $d_{ij} = d_{ji} = (a_{ij}+a_{ji})/2 \quad \forall \; i \neq j \;\wedge\; d_{ii} = a_{ii}$.

Propiedad: D es única.

ISBN: 950-42-0022-2 - ©1999

La forma cuadrática $q(x)= x^T Dx$ es *positiva (negativa) definida* si $q(x)>0$ $(q(x)<0)$ $\forall$ $x \neq 0$ en E^n.

La forma cuadrática $q(x)= x^T Dx$ es *positiva (negativa) semidefinida* si $q(x)\geq 0$ $(q(x)\leq 0)$ $\forall$ $x \in E^n$ pero q(x) no es positiva (negativa) definida.

Matriz positiva definida: La forma cuadrática $q(x)= x^T Dx$ y su matriz D asociada son *positivas definida*s si y solo si las variables pueden ser ordenadas de manera tal que $d_{11}>0$ y además se verifica:

$$\begin{vmatrix} d_{12} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} > 0 \quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix} > 0 \quad \ldots \quad |D| > 0$$

Matriz positiva semidefinida: La forma cuadrática $q(x)= x^T Dx$ y su matriz D asociada son *positivas semidefinida*s si y solo si las variables pueden ser ordenadas de manera tal que $d_{11}>0$ y además se verifica:

$$\begin{vmatrix} d_{12} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \geq 0 \quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix} \geq 0 \quad \ldots \quad |D| \geq 0$$

Matriz negativa definida: La forma cuadrática $q(x)= x^T Dx$ y su matriz D asociada son *negativas definida*s si y solo si las variables pueden ser ordenadas de manera tal que $d_{11}<0$ y además se verifica:

$$\begin{vmatrix} d_{12} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} > 0 \quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix} < 0 \quad \ldots \quad (-1)^n |D| > 0$$

Matriz negativa semidefinida: La forma cuadrática $q(x)= x^T Dx$ y su matriz D

ISBN: 950-42-0022-2 - ©1999

asociada son *negativas semidefinidas* si y solo si las variables pueden ser ordenadas de manera tal que $d_{11}<0$ y además se verifica:

$$\begin{vmatrix} d_{12} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \geq 0 \quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix} \leq 0 \quad \ldots \quad (-1)^n |D| \geq 0$$

Concavidad y Convexidad: Se dice que la función $f:S \in R^n \rightarrow R^1$ es convexa en el dominio S si para dos vectores cualesquiera $x_1 \wedge x_2 \in S$ se cumple: $f(\theta x_1+(1-\theta)x_2) \leq \theta f(x_1)+(1-\theta)f(x_2) \quad \forall \ \theta \in (0,1)$.
Se dice que la función $f:S \in E^n \rightarrow E^1$ es cóncava en el dominio S si para dos vectores cualesquiera $x_1 \wedge x_2 \in S$ se cumple: $f(\theta x_1+(1-\theta)x_2) \geq \theta f(x_1)+(1-\theta)f(x_2) \quad \forall \ \theta \in (0,1)$.

Propiedad 1: si f(x) es convexa, -f(x) es cóncava y viceversa.
Propiedad 2: Toda función lineal es cóncava y convexa a la vez.
Propiedad 3: la suma de dos o más funciones convexas (cóncavas) es otra función convexa (cóncava).
Propiedad 4: toda función convexa o cóncava es continua.

Otra caracterización de funciones convexas y cóncavas
Teorema: sea $f:S \in R^n \rightarrow R^1$ continua y diferenciable dos veces en todo punto $x \in S$, siendo S un conjunto convexo. Luego, f(x) es convexa (cóncava) en S si y solo si su matriz Hessiana es positiva (negativa) definida o semidefinida $\forall x \in S$.

Conjunto Convexo: El conjunto de punto $S \subset R^n$ se define como convexo si $\forall \ x_1 \wedge x_2 \in S$, el segmento de recta que los une está enteramente incluido en S.

Propiedad 1: si $f:R^n \rightarrow R^1$ es convexa sobre R^n, luego el conjunto de puntos $S=\{x/f(x) \leq b\}$ donde b es un escalar cualquiera, es un conjunto convexo.
Propiedad 2: si $f:R^n 6 R^1$ es cóncava sobre R^n, luego el conjunto de puntos $S=\{x/f(x) \geq b\}$ donde b es un escalar cualquiera, es un conjunto convexo.
Propiedad 3: sea $f:R^n \rightarrow R^1$, luego el conjunto de puntos $S=\{x/f(x)= b\}$ donde b es un escalar cualquiera, es un conjunto convexo si y solo si f(x) es lineal.
Propiedad 4: La intersección de dos o más conjuntos convexos es otro conjunto convexo.

ISBN: 950-42-0022-2 - 81999

Región Factible Convexa: La región factible de un programa matemático es convexa (es un conjunto convexo) si se cumplen las siguientes condiciones:
(1) Para todas las restricciones $g_i(x) \leq b_i$, $g_i(x)$ es convexa.
(2) Para todas las restricciones $g_i(x) \geq b_i$, $g_i(x)$ es cóncava.
(3) Para todas las restricciones $h_i(x)=b_i$, $h_i(x)$ es lineal.
Estas son condiciones suficientes.

Conceptos de máximo y mínimo local y global: Dada la $f:S \in R^n \rightarrow R^1$ siendo S un conjunto cerrado, se define:

1. todo punto $x^* \in S$ y cumple la condición $f(x^*) \geq f(x)$ $\forall x \in S$ define el máximo global de la función en S.
2. todo punto $x^* \in S$ y cumple la condición $f(x^*) \leq f(x)$ $\forall x \in S$ define el mínimo global de la función en S.
3. todo punto $x^* \in S$ y cumple la condición $f(x^*) \geq f(x^*+h)$ $\forall$ $h=(h1,...hn)$ / $|hj|$ es suficientemente pequeño $\forall$ $j=1,...,n$ y $(x^*+h) \in S$, define un máximo local de la función.
4. todo punto $x^* \in S$ y cumple la condición $f(x^*) \leq f(x^*+h)$ $\forall$ $h=(h1,...hn)$ / $|hj|$ es suficientemente pequeño $\forall$ $j=1,...,n$ y (x*+h)0S, define un mínimo local de la función.

ISBN: 950-42-0022-2 - 81999

ISBN: 950-42-0022-2 - ©1999