

*H
25*



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

**ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
ACATLAN**

**ANALISIS DE ALGORITMOS DE PROBLEMAS DE
PROGRAMACION NO LINEAL**

TESIS MANCOMUNADA

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A N

EDUARDO LUISILLO HERNANDEZ

MARTHA GUADALUPE PICHARDO ROJAS

MEXICO, D. F.



1992

**TESIS CON
FALLA DE ORIGEN**



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

INTRODUCCION		
CAPITULO I	OPTIMIZACION CLASICA	... 12
	I.1 Condiciones necesarias y suficientes de primer orden	... 14
	I.2 Condiciones necesarias y suficientes de segundo orden	... 21
	I.3 Funciones Convexas	... 26
	I.4 Ejemplos y Ejercicios	... 37
CAPITULO II	OPTIMIZACION SIN RESTRICCIONES METODOS DE BUSQUEDA	... 47
	II.1 Métodos de optimización de funciones unimodales de una sola variable en problemas no restringidos.	
	1a) Fibonacci.	... 48
	1b) Sección Dorada.	... 57
	II.2 Métodos de optimización de funciones multimodales de una sola variable en problemas no restringidos.	
	2a) Interpolación Cuadrática.	... 61
	2b) Interpolación cúbica.	... 66
	II.3 Métodos de optimización que utilizan derivadas para funciones de varias variables en problemas no restringidos.	
	3a) Método de Newton con aproximación cuadrática	... 71
	3b) Descenso acelerado caso cuadrático	... 78
	3c) Direcciones conjugadas	... 84
	3d) Método de Gradiente conjugado	... 90
	3e) Método de Quasi-Newton; Método de Davidon-Fletcher-Powell	... 96
	3f) Método de Hooke & Jeeves	... 104

CAPITULO III	APLICACIONES DE OPTIMIZACION SIN RESTRICCIONES	...	106
	III.1 Desarrollo del programa computacional para el Tema 1 del Capítulo II.	...	108
	III.2 Desarrollo del programa computacional para el Tema 2 del Capítulo II.	...	114
	III.3 Desarrollo del programa computacional para el Tema 3 del Capítulo II.	...	124
CAPITULO IV	METODO PARA OPTIMOS CON RESTRICCIONES	...	154
	IV.1 Método de Zoutendijk.	...	160
	IV.2 Método de Rosen de la Proyección del Gradiente.	...	162
	IV.3 Método del Gradiente Reducido.	...	167
	CONCLUSIONES		
	APENDICES TECNICOS		
	BIBLIOGRAFIA		

INTRODUCCION

Una de las materias que forma parte de la carrera de Actuaría y que es de carácter obligatorio es Seminario de Tesis I, en esta materia se estableció que para poder finalizar el curso se tendría que tener perfectamente definido el esquema de tesis, generando una primera reacción de asombro y desconcierto pues solamente se contaba con la teoría y ninguna experiencia profesional.

La idea que se desarrolló en el Seminario de Tesis llegó a concluirse originando este trabajo.

Simultáneamente al curso, se tuvo la fortuna de elegir la materia de programación matemática que forma parte de la especialización del Área de investigación de operaciones en donde es posible sensibilizarse de la necesidad de crear un documento fuente que sirviera de apoyo en el estudio de estos temas siendo esta la programación no lineal.

Lo anterior fue el principio del presente trabajo de tesis.

Posteriormente se tuvo la oportunidad de formar parte del personal docente para impartir la materia de Optimización I a la carrera de Matemáticas Aplicadas y Computación en cuyo temario se incluye la programación no lineal, sin duda alguna independientemente de que el destino de la vida profesional de quienes elaboran este trabajo los colocó en el campo de los Seguros, el interés persistió en seguir teniendo contacto con la programación no lineal del otro lado del escritorio, puesto que al impartir esta clase, se obtuvo una globalización de tal idea ya que ahora se cuenta tanto con el punto de vista de alumno como el de profesor.

El obstáculo encontrado en el estudio de esta materia fue la falta de bibliografía en español y que la poca bibliografía que existe en inglés se encuentra muy dispersa y en algunos casos muy especializada. Lo anterior da como resultado que en los primeros contactos con los temas de programación no lineal se producen confusiones o falta de entendimiento originando un desinterés en proseguir los estudios en estas preespecialidades.

Por lo tanto lo que se pretende con la Tesis es generar un documento básico que apoye en los estudios iniciales de esta área, para esto se decidió estructurar este trabajo en cuatro capítulos. El primero de ellos habla de Optimización Básica, el siguiente se refiere a Optimización sin restricciones, el tercer capítulo se presentan los programas computacionales para los algoritmos desarrollados en el capítulo II y por último se presentan algunos métodos para óptimos con restricciones.

Así mismo se pretende que el objetivo fundamental que es desarrollar los algoritmos de problemas de programación no lineal, esté completamente realizado.

CAPITULO I

CAPITULO I. OPTIMIZACION CLASICA

Lo que se pretende en este capítulo es establecer y analizar el problema de optimización no lineal para funciones vectoriales continuas $f(x)$ siendo x perteneciente a Ω , donde Ω es subconjunto de E^n y $f(x)$ posea un mínimo (o máximo), así mismo plantear las condiciones necesarias y suficientes de primer y de segundo orden, así como las demostraciones de proposiciones y teoremas que sustentan tales condiciones, también se resolverán ejemplos numéricos para asegurar la interpretación del concepto básico teórico.

Se Tratarán algunas propiedades de las funciones convexas (o cóncavas), las cuales pueden garantizar la existencia de puntos mínimos (o máximos), siendo estas propiedades el inicio de una búsqueda intuitiva hacia el punto mínimo (o máximo).

Por último se citarán unos ejemplos donde se aplica la teoría anterior, dando una motivación al planteamiento de algoritmos para la resolución de los mismos.

I.1 CONDICIONES NECESARIAS Y SUFICIENTES DE PRIMER ORDEN

Para efectos de simplificación, en este capítulo se consideran únicamente problemas de optimización de la forma:

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{s.a.} & x \in \Omega \end{array} \quad \dots \quad (II.1)$$

donde f es una función real y $\Omega \subset E^n$.

Primeramente, dentro del estudio de dichos problemas habrá que analizar la posibilidad de que exista una solución, para lo cual se puede utilizar el teorema de Weierstrass que establece que si f es continua y Ω cerrado, existe una solución.

Ahora bien, este análisis se enfoca en los puntos solución característicos y en idear métodos efectivos para encontrarlos. Se distinguen dos tipos de puntos solución: puntos mínimos locales y puntos mínimos globales.

Definición: Un punto $x^* \in \Omega$ es un mínimo relativo o mínimo local de f bajo Ω si existe un $\epsilon > 0$ tal que $f(x) \geq f(x^*)$ para todo $x \in \Omega$ en una distancia ϵ de x^* (esto es, $x \in \Omega$ y $\|x - x^*\| < \epsilon$). Si $f(x) > f(x^*)$ para todo $x \in \Omega$, en una distancia ϵ de x^* , entonces x^* es un mínimo relativo estricto de f bajo Ω .

Definición: Un punto x^* $\in \Omega$ es un mínimo global de f bajo Ω si $f(x) \geq f(x^*)$ para todo $x \in \Omega$. Si $f(x) > f(x^*)$ para todo $x \in \Omega$, $x \neq x^*$, entonces x^* es un mínimo global estricto de f bajo Ω .

Al formular y abordar el problema (II.1) se requiere explícitamente, y por definición, un punto mínimo global de f bajo el conjunto Ω . Sin embargo, la práctica dicta, tanto desde el punto de vista computacional como teórico, que en muchas ocasiones se debe conformar con obtener un mínimo relativo. Por ejemplo, al derivar condiciones necesarias basadas en el cálculo diferencial o cuando se busca un punto mínimo mediante un método iterativo convergente, lo único que se está haciendo son comparaciones entre puntos vecinos y fijando así la atención en puntos mínimos relativos.

Como regla general, solamente será posible encontrar una solución global si el problema posee ciertas propiedades de convexidad que garanticen que cualquier mínimo relativo es un mínimo global.

En conclusión se puede afirmar que lo más práctico al formular y abordar el problema (II.1) es considerar, implícitamente, que lo que se busca es un punto mínimo relativo.

Por otra parte, al aplicar condiciones necesarias satisfechas por

trayectorias lejos del punto en algunas direcciones dadas. A través de cualquiera de dichas direcciones, la función objetivo puede ser considerada como una función de una sola variable, el parámetro define la trayectoria en esta dirección, y de aquí será aplicable el cálculo ordinario de una sola variable.

Así dado $x \in \Omega$ podemos decir que un vector d es una dirección factible en x si existe un $\bar{\alpha} > 0$ tal que $x + \alpha d \in \Omega$ para todo α , $0 \leq \alpha \leq \bar{\alpha}$. Con este concepto tan simple es posible establecer algunas condiciones que deberán satisfacer los puntos mínimos relativos.

Proposición II.1.—(Condiciones necesarias de Primer orden).

Sea Ω un subconjunto de E^n y f una vez diferenciable, una función en Ω . Si x^* es un punto mínimo relativo de f bajo Ω , entonces para cualquier dirección factible $d \in E^n$ en x^* , tenemos $\nabla f(x^*)d \geq 0$.

Demostración:

Por hipótesis d es una dirección factible en x^* por lo tanto existe $\bar{\alpha} > 0$ tal que $x^* + \alpha d \in \Omega \forall \alpha$, $0 \leq \alpha \leq \bar{\alpha}$.

Aplicando la serie de Taylor:

$$f(x^* + \alpha d) = f(x^*) + (\nabla f(x^*)d)\alpha + R_2(x^*)$$

Restando $f(x^*)$ en la igualdad:

$$f(x^* + \alpha d) - f(x^*) = (\nabla f(x^*)d)\alpha + R_1(x^*) \quad \text{donde } R_1(x^*) \cong 0$$

$$\therefore f(x^* + \alpha d) - f(x^*) \cong (\nabla f(x^*)d)\alpha$$

Si $\nabla f(x^*)d < 0$ y $\alpha \geq 0 \Rightarrow f(x^* + \alpha d) - f(x^*) < 0$ contradice la hipótesis de que x^* es un mínimo relativo

$$\therefore \nabla f(x^*)d \geq 0 \quad \text{q.e.d}$$

A continuación se analiza el caso especial en el que x^* se encuentre dentro de Ω (como podría ocurrir si $\Omega = E^n$). En dicho caso habrán direcciones factibles emanadas en todas direcciones de x^* , y por tanto $\nabla f(x^*)d \geq 0$ para toda $d \in E^n$. Esto implica que $\nabla f(x^*) = 0$

Corolario II.1.- (Caso irrestricto)

Sea $\Omega \in E^n$, y f una vez diferenciable, una función en Ω . Si x^* es un mínimo relativo de f bajo Ω y si x^* es un punto interior de Ω , entonces $\nabla f(x^*) = 0$

Las condiciones necesarias en el caso irrestricto puro llevan a n ecuaciones (una para cada componente de ∇f) en n variables (las componentes de x^*), las cuales en muchos casos pueden ser resueltas para determinar la solución. En la práctica, sin embargo, un problema de optimización se resuelve directamente sin intentar explícitamente de resolver las ecuaciones que provienen de las condiciones necesarias, aún cuando estas condiciones

constituyan un fundamento para la teoría.

Por ejemplo, considérese el Problema

$$\text{minimizar } f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2$$

No existen restricciones por lo tanto $\Omega = E^2$. Obteniendo las derivadas parciales de f iguales a cero se tiene:

$$2x_1 - x_2 = 0 \quad \dots (1)$$

$$-x_1 + 2x_2 = 3 \quad \dots (2)$$

Resolviendo el sistema:

de (1) $x_2 = 2x_1$ sustituyendo en (2) $-x_1 + 2(2x_1) = 3$

$\therefore 3x_1 = 3 \Rightarrow x_1 = 1$, y $x_2 = 2$ siendo esta única solución, por lo tanto se trata de un mínimo global.

Sea el problema:

$$\begin{aligned} \text{minimizar } f(x_1, x_2) &= x_1^2 - x_1 + x_2 + x_1 x_2 \\ \text{s.a.} \quad x_1 &\geq 0, x_2 \geq 0 \end{aligned}$$

Este problema tiene un mínimo global en:

$$x_1 = \frac{1}{2}, x_2 = 0$$

Aplicando las derivadas parciales:

$$\frac{df}{dx_1} = 2x_1 - 1 + x_2 \quad 2x_1 - 1 + x_2 = 0 \Rightarrow x_2 = 3$$

$$\frac{df}{dx_2} = 1 + x_1 \qquad 1 + x_1 = 0 \quad \Rightarrow \quad x_1 = -1$$

Como puede verse la solución anterior ($x_1 = -1$, $x_2 = 3$) se sale del dominio establecido $x_1 \geq 0$ y $x_2 \geq 0$, por lo que se puede concluir que las derivadas parciales igualadas a cero no necesariamente conducen a una solución. Aplicando la Proposición II.1 tenemos $\nabla f(x^*)d \geq 0$ para toda $d \in E^2$ tal que d es una dirección factible en el punto x^* .

$$\begin{aligned} \nabla f(x^*)d &= \begin{bmatrix} x_2 + 2x_1 - 1 \\ x_1 + 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \\ &= d_1 x_2 + 2d_1 x_1 - d_1 + d_2 x_1 + d_2 \geq 0 \end{aligned}$$

Analizando la función objetivo $f(x_1, x_2) = x_1^2 - x_1 + x_2 + x_1 x_2$ como se está minimizando es claro que para alcanzar el mínimo es necesario que $x_2 = 0$, puesto que el dominio es $x_2 \geq 0$, entonces es factible este valor para x_2 .

$$-d_1(0) + 2d_1 x_1 - d_1 + d_2 x_1 + d_2 \geq 0$$

Ahora para poder garantizar la proposición II.1 para toda d dirección factible sería necesario que x_1 tuviera al menos el valor de $\frac{1}{2}$:

$$2d_1 \left(\frac{1}{2}\right) - d_1 + d_2 \left(\frac{1}{2}\right) + d_2 \geq 0$$

$$\frac{1}{2}d_2 + d_2 \geq 0 \Rightarrow \frac{3}{2}d_2 \geq 0 \Rightarrow d_2 \geq 0$$

Lo anterior demuestra que el mínimo se encuentra en $x_2=0$ y $x_1 \geq \frac{1}{2}$, retomando la función objetivo se tiene:

$$f(x_1, 0) = x_1^2 - x_1$$

Esta función se convierte en problema de una sola variable, el punto mínimo de la función aplicando la primera derivada:

$$f'(x_1) = 2x_1 - 1 = 0 \Rightarrow x_1 = \frac{1}{2}$$

$$f''(x_1) = 2 \geq 0 \quad \therefore x_1 = \frac{1}{2} \text{ es un punto mínimo}$$

\therefore El mínimo está en el punto $(\frac{1}{2}, 0)$ con imagen de $f(\frac{1}{2}, 0) = -0.25$ y el valor del gradiente estaría dado por

$$\frac{df}{dx_1} = 2x_1 - 1 + x_2 = 0$$

$$\frac{df}{dx_2} = 1 + x_1 = \frac{3}{2}$$

I.2 CONDICIONES NECESARIAS Y SUFICIENTES DE SEGUNDO ORDEN

La demostración de la proposición I1.1 está basada en hacer una aproximación de primer orden para la función f en la vecindad del punto mínimo relativo a través de la serie de Taylor. Mientras que las condiciones de segundo orden son definidas en términos de la matriz Hessiana $\nabla^2 f$ (1) de las segundas derivadas de f , como en seguida se verá:

Proposición I2.1.- (Condiciones necesarias de segundo orden)

Sea Ω un subconjunto de E^n y sea f dos veces diferenciable una función en Ω . Si x^* es un punto mínimo relativo de f sobre Ω , entonces para cualquier $d \in E^n$ esto es una dirección factible en x^* se tiene:

- i) $\nabla f(x^*) \cdot d \geq 0$
- ii) Si $\nabla f(x^*) \cdot d = 0 \Rightarrow d^T \nabla^2 f(x^*) d \geq 0$ (1)

La matriz hessiana se simbolizará indistintamente como $\nabla^2 f(x)$, $F(x)$ o $H(x)$.

Demostración:

- i) Proposición I1.1
- ii) Por hipótesis d es una dirección factible en x^* por lo tanto existe $\bar{\alpha} > 0$ tal que $x^* + \alpha d \in \Omega \quad \forall \alpha, 0 \leq \alpha \leq \bar{\alpha}$
Aplicando la serie de Taylor:

$$f(x^0 + \alpha d) = f(x^0) + (\nabla f(x^0)d)\alpha + \frac{1}{2}(d^T \nabla^2 f(x^0)d)\alpha^2 + R_2(x^0)$$

por hipótesis: $\nabla f(x^0)d = 0$

$$\Rightarrow f(x^0 + \alpha d) = f(x^0) + \frac{1}{2}(d^T \nabla^2 f(x^0)d)\alpha^2 + R_2(x^0)$$

$$\text{despejando } f(x^0 + \alpha d) - f(x^0) = \frac{1}{2}(d^T \nabla^2 f(x^0)d)\alpha^2 + R_2(x^0)$$

$$\text{como } R_2(x^0) \geq 0 \Rightarrow f(x^0 + \alpha d) - f(x^0) \geq \frac{1}{2}(d^T \nabla^2 f(x^0)d)\alpha^2$$

$$\text{si } d^T \nabla^2 f(x^0)d < 0 \text{ y } \frac{1}{2}\alpha^2 \geq 0 \Rightarrow f(x^0 + \alpha d) - f(x^0) < 0$$

contradictorio la hipótesis de que x^0 es un mínimo relativo

$$\therefore d^T \nabla^2 f(x^0)d > 0 \quad \text{q.e.d.}$$

Corolario I2.1 (Condiciones necesarias de segundo orden-caso sin restricciones)

Sea x^0 un punto interior del conjunto Ω y supóngase que x^0 es un punto mínimo relativo sobre Ω de la función f dos veces diferenciable, entonces

$$i) \nabla f(x^0) = 0$$

$$ii) \forall d, d^T \nabla^2 f(x^0)d \geq 0$$

Para una notación simplificada frecuentemente se denotará $\nabla^2 f(x)$ como la matriz de $n \times n$ de la segunda derivada parcial de f , el Hessiano de $f = F(x)$. La condición ii) es equivalente a establecer que la matriz $F(x^0)$ es positiva semidefinida. La matriz $F(x^0)$, la cual aparece aquí muy natural en una discusión de condiciones necesarias, juega un papel fundamental en el análisis de métodos iterativos para solucionar problemas de

optimización sin restricciones. La estructura de esta matriz es el determinante principal de la razón de convergencia de algoritmos a minimizar funciones f .

Retomando la información del problema anterior:

$$x^* = \left(\frac{1}{2}, 0\right), \quad \nabla f(x^*) = \left(0, \frac{2}{3}\right)$$

Aplicando la proposición I2.1

i) $\nabla f(x^*)d = \left(0, \frac{2}{3}\right) \cdot (d_1, d_2) = \frac{2}{3}d_2 \geq 0$ por definición de la región factible. Por lo tanto se cumple i)

ii) Si $\nabla f(x^*)d = 0 \Rightarrow d_2 = 0$

$$\nabla^2 f(x^*) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{aligned} d^T \nabla^2 f(x^*) d &= (d_1, d_2) \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 2d_1 & d_1 \\ d_1 & 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \\ &= 2d_1^2 + d_1 d_2 + d_1 d_2 = 2d_1^2 + 2d_1 d_2 \end{aligned}$$

pero $d_2 = 0$

$$= 2d_1^2 \geq 0 \Rightarrow d^T \nabla^2 f(x^*) d \geq 0$$

también se satisface la condición ii)

Existe un interés especial en los puntos mínimos que son puntos interiores del dominio, como son los problemas irrestrictos.

Considerar el problema

$$\text{minimizar } f(x_1, x_2) = x_1^2 - x_1^2 x_2 + 2x_2^2$$

s.a.

$$x_1 \geq 0, \quad x_2 \geq 0$$

Se obtiene el gradiente :

$$3x_1^2 - 2x_1x_2 = 0$$

$$-x_1^2 + 4x_2 = 0$$

Resolviendo el sistema de ecuaciones se tienen dos puntos solución

$$(x_1, x_2) = (0,0), \text{ que es un punto frontera, y}$$

$$(x_1, x_2) = (6,9)$$

Nótese que para x_1 fijo en 6, la función objetivo obtiene un mínimo relativo con respecto a x_2 en $x_2 = 9$.

Contrariamente, con x_2 fijo en 9, la función objetivo alcanza su mínimo con respecto a x_1 en $x_1 = 6$. Sin embargo la matriz Hessiana no cumple con

$$F = \begin{bmatrix} 6x_1 - 2x_2 & -2x_1 \\ -x_1^2 & 4 \end{bmatrix} = \begin{bmatrix} 18 & -12 \\ -12 & 4 \end{bmatrix}$$

en virtud de que no se cumple el inciso ii del Corolario I2.2 ya que F no es positiva definida, el punto (6,9) no es un mínimo relativo.

Para problemas sin restricciones o donde se puede garantizar que el punto mínimo es un punto interior de la región factible, se introduce una proposición que viene a reafirmar las condiciones anteriores:

Proposición I2.2: (Condiciones suficientes de segundo orden - Caso sin restricciones).

Sea f dos veces diferenciable definida, en una región en la cual el punto x^* es un punto interior. Supóngase en adición que:

- i) $\nabla f(x^*) = 0$
 - ii) $F(x^*)$ es definida positiva
- entonces x^* es punto mínimo relativo estricto de f .

Demostración:

Aplicando la serie de Taylor:

$$f(x^* + d) = f(x^*) + \nabla f(x^*)d + \frac{1}{2} d^T \nabla^2 f(x^*)d + R_2(x^*)$$

donde $R_2(x^*) = R(|d|^2)$

por hipótesis $\nabla f(x^*) = 0 \Rightarrow$

$$f(x^* + d) - f(x^*) = \frac{1}{2} d^T \nabla^2 f(x^*)d + R_2(x^*) \text{ por hipótesis}$$

$F(x^*)$ es definida positiva: entonces existe un escalar

$a > 0$ tal que $\forall d, d^T F(x^*)d \geq a|d|^2$, entonces

$$f(x^* + d) - f(x^*) \geq (a/2)|d|^2 + R(|d|^2)$$

q.e.d.

I.3 FUNCIONES CONVEXAS Y CONCAVAS

Para desarrollar una Teoría dirigida a caracterizar puntos mínimos globales más que locales, es necesario introducir alguna clase de supuestos de convexidad. Esto da como resultado no sólo una teoría más potente, aunque más restrictiva, sino también proporciona una interesante interpretación geométrica del resultado de las condiciones suficientes de segundo orden expuesto anteriormente.

Definición: Una función f definida en un conjunto convexo Ω se dice que es convexa si, para toda $x_1, x_2 \in \Omega$ y toda α tal que $0 \leq \alpha \leq 1$, se tiene:

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

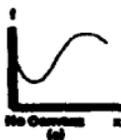
si para toda α , $0 < \alpha < 1$ y $x_1 \neq x_2$ se tiene

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha)f(x_2).$$

entonces se dice que f es estrictamente convexa.

En la Fig. 1 se muestran algunos ejemplos de funciones convexas y no convexas. Geométricamente, una función es convexa si la línea que une dos puntos en su gráfica no cae en ninguna parte debajo de la gráfica, como se muestra en la Fig. 1 o al pensar en una función de dos dimensiones, ésta es convexa si su gráfica forma una cavidad.

Definición: Una función g definida en un conjunto convexo, se dice que es cóncava si la función $f = -g$ es convexa. La función g es estrictamente cóncava si $-g$ es estrictamente convexa.



Combinaciones de funciones convexas.

Se muestra que las funciones convexas se pueden combinar para dar nuevas funciones convexas y que cuando son utilizadas como restricciones forman conjuntos convexas restringidos.

Proposición 13.1: Sean f_1 y f_2 funciones convexas en el conjunto convexo Ω . Entonces la función $f_1 + f_2$ es convexa en Ω .

Demostración:

Por hipótesis se tiene que f_1 y f_2 son funciones convexas, sea

$x_1, x_2 \in \Omega$ y $0 < \alpha < 1$:

$$f_1(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f_1(x_1) + (1-\alpha) f_1(x_2)$$

$$f_2(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f_2(x_1) + (1-\alpha) f_2(x_2)$$

sumando ambas desigualdades:

$$f_1(\alpha x_1 + (1-\alpha)x_2) + f_2(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f_1(x_1) + (1-\alpha) f_1(x_2) + \alpha f_2(x_1) + (1-\alpha) f_2(x_2)$$

$$f_1(\alpha x_1 + (1-\alpha)x_2) + f_2(\alpha x_1 + (1-\alpha)x_2) \leq \alpha (f_1(x_1) + f_2(x_1)) + (1-\alpha) (f_1(x_2) + f_2(x_2))$$

$\therefore f_1 + f_2$, es una función convexa.

Proposición 13.2: Sea f una función convexa sobre el conjunto convexo Ω . Entonces la función af es convexa para cualquier $a \geq 0$.

Demostración:

$$\text{Por hipótesis: } f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

multiplicando la desigualdad por $a \geq 0$:

$$af(\alpha x_1 + (1-\alpha)x_2) \leq \alpha(af(x_1)) + (1-\alpha)(af(x_2))$$

$\therefore af$ es convexa

Nótese que mediante la aplicación repetida de las dos proposiciones anteriores se encuentra que una combinación positiva $a_1 f_1 + a_2 f_2 + \dots + a_m f_m$ de funciones convexas es también convexa. Finalmente, considérense conjuntos definidos por restricciones convexas desiguales.

Proposición 13.3: Sea f una función convexa en un conjunto convexo Ω . El conjunto $\Gamma_c = \{x: x \in \Omega, f(x) \leq c\}$ es convexo para todo número real c .

Demostración:

Sea $x_1, x_2 \in \Gamma_c$, entonces $f(x_1) \leq c$ y $f(x_2) \leq c$

Además $0 \leq \alpha \leq 1$, multiplicando la primera desigualdad por α y la segunda por $(1-\alpha)$ se tiene:

$$\alpha f(x_1) \leq \alpha c, \quad (1-\alpha)f(x_2) \leq (1-\alpha)c$$

Sumando ambas desigualdades: $\alpha f(x_1) + (1-\alpha)f(x_2) \leq \alpha c + (1-\alpha)c = c$

por hipótesis $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$, $x_1, x_2 \in \Omega$

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) \leq c$$

$\Rightarrow \alpha x_1 + (1-\alpha)x_2 \in \Gamma_c \therefore \Gamma_c$ es convexo $\forall c$ real q.e.d.

Nótese que, puesto que la intersección de conjuntos convexas es también convexa, el conjunto de puntos que satisface simultáneamente $f_1(x) \leq c_1, f_2(x) \leq c_2, \dots, f_m(x) \leq c_m$, donde cada f_i es una función convexa, define un conjunto convexo. Esto es importante en Programación Matemática, ya que el conjunto

restringido es frecuentemente definido de esta forma.

Propiedades de funciones diferenciables convexas.

Si una función f es diferenciable, entonces existen caracterizaciones alternativas de convexidad.

Proposición 13.4: Sea f una vez diferenciable entonces f es convexa sobre un conjunto convexo si y sólo si $f(y) \geq f(x) + \nabla f(x)(y-x)$ para toda $x, y \in \Omega$

Demostración:

Primero supóngase que f es convexa. Entonces para todo α ,

$0 \leq \alpha \leq 1$, $f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha)f(x)$, reordenando:

$f(x + \alpha(y-x)) \leq \alpha f(y) + f(x) - \alpha f(x)$ Así para $0 < \alpha \leq 1$,

restando $f(x)$ y dividiendo entre α :

$$\frac{[f(x + \alpha(y-x)) - f(x)]}{\alpha} \leq f(y) - f(x)$$

$$\frac{f(x + \alpha(y-x)) - f(x)}{\alpha} (y-x) \leq f(y) - f(x)$$

$$\alpha (y-x)$$

Si $\alpha \rightarrow 0$ se tiene:

$$\nabla f(x) (y-x) \leq f(y) - f(x)$$

Esto prueba la parte "sólo si". \Rightarrow q.e.d.

Ahora supóngase:

$f(y) \geq f(x) + \nabla f(x) (y-x)$, para toda $x_1, x_2 \in \Omega$

$x_1, x_2 \in \Omega$ y $\alpha, 0 \leq \alpha \leq 1$.

Estableciendo $x = \alpha x_1 + (1-\alpha)x_2$ y alternativamente

$y = x_1$ ó $y = x_2$ se tiene

$$f(x_1) \geq f(x) + \nabla f(x)(x_1 - x) \dots\dots\dots (*)$$

$$f(x_2) \geq f(x) + \nabla f(x)(x_2 - x) \dots\dots\dots (**)$$

Multiplicando (*) por α y (**) por $(1-\alpha)$ y sumándolas, se obtiene:

$$\alpha f(x_1) + (1-\alpha)f(x_2) \geq f(x) + \nabla f(x) [\alpha x_1 + (1-\alpha)x_2 - x]$$

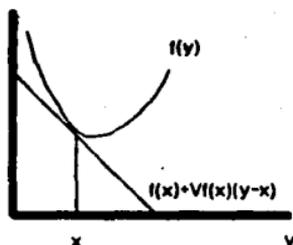
Sustituyendo $x = \alpha x_1 + (1-\alpha)x_2$, se obtiene

$$\alpha f(x_1) + (1-\alpha)f(x_2) \geq f(\alpha x_1 + (1-\alpha)x_2) \quad \leftarrow$$

q.e.d.

La declaración de la proposición anterior se ilustra en la Fig. 2. Esta se puede considerar como un tipo de caracterización dual de la definición original ilustrada en la Fig. 1. La definición original establece esencialmente que una interpolación lineal entre 2 puntos sobreestima la función, mientras que la proposición anterior establece que una aproximación lineal basada en la derivada local subestima la función.

Fig. 2



Para dos funciones continuamente diferenciables, existe otra caracterización de convexidad.

Proposición 13.5: Sea f dos veces diferenciable. Entonces f es convexa sobre un conjunto convexo Ω conteniendo un punto interior si y sólo si la matriz Hessiana F de f es positiva semidefinida sobre Ω

Demostración: Por el teorema de Taylor

$$f(y) = f(x) + \nabla f(x)(y-x) + \frac{1}{2}(y-x)^T F[x+\alpha(y-x)] (y-x)$$

para algún α , $0 \leq \alpha \leq 1$

Si F es positiva semidefinida $\frac{1}{2} (y-x)^T F(x + \alpha(y-x)) (y-x) \geq 0$
 $f(y) \geq f(x) + \nabla f(x)(y-x) \Rightarrow$ por la proposición 13.4 f es convexa

Suponiendo que el Hessiano no es positivo semidefinido en algún punto $x \in \Omega$. Por continuidad del Hessiano, se puede asumir sin perder generalidad, que x es un punto interior de Ω . Existe $y \in \Omega$ tal que $(y-x)^T F(x) (y-x) < 0$. Por continuidad del Hessiano, puede ser seleccionada de tal forma que para toda α , $0 \leq \alpha \leq 1$.

$$(y-x)^T F(x + \alpha(y-x)) (y-x) < 0$$

Nuevamente por el teorema de Taylor

$$f(y) - f(x) = \nabla f(x) (y-x) + \frac{1}{2} (y-x)^T F(x + \alpha(y-x)) (y-x)$$

contradice

$$f(y) \geq f(x) + \nabla f(x)(y-x)$$

lo que implica que f no es convexa.

q.e.d.

La matriz Hessiana es la generalización en E^n del concepto de curvatura de una función, asimismo, la definición (positiva, negativa, etc) del Hessiano es la generalización de la curvatura positiva. Las funciones convexas tienen curvatura positiva (o al menos no negativa) en cualquier dirección. Debido a estas observaciones, algunas veces se refiere a funciones como localmente convexas si su Hessiano es positivo definido en una pequeña región, y localmente estricta convexa si el Hessiano es positivo definido en la región. En estos términos se ve que las

condiciones suficientes de segundo orden requieren que la función sea estrictamente convexa en el punto x^* . Así, aún la teoría local, derivada solamente en términos de cálculo elemental, está relacionada estrechamente con la convexidad al menos localmente. Por esta razón se puede ver dos teorías, local y global, no como desarrollos separados sino complementarios e interactivos. Estos resultados están basados en la aplicación de convexidad aún en problemas no convexos en una región cerca de la solución, e inversamente, resultados locales se aplican a mínimos globales.

Minimización y maximización de funciones convexas.

Teorema I3.1: Sea f una función convexa definida en un conjunto convexo Ω . Entonces el conjunto donde f alcanza su mínimo es convexo, y cualquier mínimo relativo de f es un mínimo global.

Demostración:

Si f tiene un mínimo no relativo, el teorema es válido por default. Asígnese que C_0 es el mínimo de f .

Entonces $\Gamma = \{x: f(x) \leq C_0, x \in \Omega\}$ y ésta es convexa por la proposición I3.3.

Supóngase que $x^* \in \Omega$ es un mínimo relativo de f , pero que existe otro punto $y \in \Omega$ con $f(y) < f(x^*)$. Sobre la línea $\alpha y + (1-\alpha)x^*$, $0 < \alpha < 1$ se tiene que $f(\alpha y + (1-\alpha)x^*) \leq \alpha f(y) + (1-\alpha)f(x^*) < f(x^*)$, contradiciendo el hecho de que x^* es un mínimo relativo.

q.e.d.

Se podría decir por el teorema anterior que para funciones convexas, todos los puntos mínimos se localizan juntos (en un conjunto convexo) y todos los mínimos relativos son mínimos globales. El siguiente teorema dice que si f es continuamente diferenciable y convexo, entonces el satisfacer las condiciones necesarias de primer orden son necesarias y suficientes para un punto que sea un punto global mínimo.

Teorema 13.2: Sea f una vez diferenciable y convexa en el conjunto convexo Ω . Si existe un punto x^* en Ω tal que, para todo $y \in \Omega$, $\nabla f(x^*)(y-x) \geq 0$, entonces x^* es un punto mínimo global de f sobre Ω .

Demostración: Nótese que ya que $y-x^*$ es una dirección factible en x^* , la condición dada es equivalente a la condición necesaria de primer orden. La prueba de la proposición es inmediata, ya que por la proposición 13.4 de la última sección

$$f(y) \geq f(x^*) + \nabla f(x^*)(y-x^*) \geq f(x^*)$$

q.e.d.

Ahora volviendo a la cuestión de maximizar funciones convexas sobre un conjunto convexo. De hecho existe un teorema, sin embargo no es análogo al teorema 13.1; la tendencia es para la ocurrencia de numerosos máximos relativos no globales, pero es posible probar un resultado importante.

Sea f una función convexa definida en una vecindad, cerrada de un conjunto convexo Ω . Si f tiene un máximo sobre Ω este se logra en un punto extremo de Ω .

I.4 EJEMPLOS Y EJERCICIOS

La finalidad de este tema es dar una idea del panorama amplio de este concepto en programación no lineal de problemas irrestrictos, que va desde aplicaciones en procesos químicos, cotización de proyectos, problemas de diseño estructural, ajuste de curvas, asignación de recursos, problemas de producción; y así sucesivamente se podrían citar un sin fin de aplicaciones, aunque en la realidad la mayor parte de los problemas se reducen a una forma simple. También es importante mencionar que gran parte de la variedad de las aplicaciones de la programación no-lineal irrestricta se debe a que frecuentemente muchos problemas con restricciones se convierten a problemas irrestrictos, utilizando las restricciones para establecer relaciones entre variables, reduciendo así el número efectivo de variables.

A continuación se abordarán unos ejemplos donde se puede ver el alcance de la aplicación de esta teoría.

Ejemplo 1: (Producción).

Un problema común en la teoría económica consiste en determinar la mejor forma de combinar los gastos para obtener cierto bienestar, cierta comodidad.

Existe una función $f(x_1, x_2, \dots, x_n)$ que da la cantidad de bienestar producida como una función de los montos x_i de los gastos,

$i = 1, 2, \dots, n$. El precio unitario del bienestar producido es q y los precios unitarios de los gastos son p_1, p_2, \dots, p_n . Para maximizar el bienestar, se debe resolver el siguiente problema:

$$\text{Max } qf(x_1, x_2, \dots, x_n) - p_1x_1 - p_2x_2 - \dots - p_nx_n$$

Recurriendo a las condiciones necesarias de primer orden las derivadas parciales con respecto a cada x_i se igualan a cero. Esto genera las ecuaciones:

$$q \frac{df}{dx_i}(x_1, x_2, \dots, x_n) = p_i \quad i = 1, 2, \dots, n$$

$$\frac{df}{dx_i}(x_1, x_2, \dots, x_n) = \frac{p_i}{q} \quad i = 1, 2, \dots, n$$

Estas ecuaciones pueden interpretarse como que, en la solución, el valor marginal debido a un pequeño aumento en el i -ésimo gasto debe ser igual al precio p_i .

Ejemplo 2: (Aproximación)

Un uso común de la optimización es para el efecto de la función de aproximación. Supóngase, por ejemplo, que a través de un experimento el valor de una función g es observado en puntos, x_1, x_2, \dots, x_m . Así los valores $g(x_1), g(x_2), \dots, g(x_m)$ son conocidos. Se desea aproximar la función por un polinomio:

$$h(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$$

de grado n (o menos), donde $n < m$. Correspondiendo a cualquier elección del polinomio aproximado, habrá un conjunto de errores

$$E_k = g(x_k) - h(x_k).$$

Se define la mejor aproximación como el polinomio que minimiza la suma de los cuadrados de estos errores; esto es, minimiza:

$$\sum_{k=1}^m (E_k)^2$$

Esto por turno significa que se minimiza:

$$f(a) = \sum_{k=1}^m \left[g(x_k) - \left(a_n x_k^n + a_{n-1} x_k^{n-1} + \dots + a_0 \right) \right]^2$$

con respecto a $a = (a_0, a_1, \dots, a_n)$ para encontrar los mejores coeficientes. Esta es una expresión cuadrática en los coeficientes a . Para encontrar una representación compacta para este objetivo se define:

$$q_{ij} = \sum_{k=1}^m (x_k)^{i+j}, \quad b_j = \sum_{k=1}^m g(x_k) (x_k)^j, \quad y$$

$$c = \sum_{k=1}^m g(x_k)^2$$

Haciendo un poco de álgebra:

$$\begin{aligned} f(a) &= \sum_{k=1}^m \left[g(x_k) - h(x_k) \right]^2 = \sum_{k=1}^m \left[g(x_k)^2 - 2g(x_k)h(x_k) + h(x_k)^2 \right] \\ &= \sum_{k=1}^m g(x_k)^2 - 2 \sum_{k=1}^m g(x_k)h(x_k) + \sum_{k=1}^m h(x_k)^2 \end{aligned}$$

$$\sum_{k=1}^m g(x_k)^2 = c$$

$$\sum_{k=1}^m g(x_k)h(x_k) = \sum_{k=1}^m g(x_k) [a_n x_k^n + a_{n-1} x_k^{n-1} + \dots + a_0]$$

$$= g(x_1) [a_n x_1^n + a_{n-1} x_1^{n-1} + \dots + a_0]$$

$$\vdots$$

$$+ g(x_{m-1}) [a_n x_{m-1}^n + a_{n-1} x_{m-1}^{n-1} + \dots + a_0]$$

$$+ g(x_m) [a_n x_m^n + a_{n-1} x_m^{n-1} + \dots + a_0]$$

$$= a_n [g(x_1)(x_1)^n] + a_{n-1} [g(x_1)(x_1)^{n-1}] + \dots + a_0 g(x_1)$$

$$+ a_n [g(x_{m-1})(x_{m-1})^n] + a_{n-1} [g(x_{m-1})(x_{m-1})^{n-1}] + \dots + a_0 g(x_{m-1})$$

$$+ a_n [g(x_m)(x_m)^n] + a_{n-1} [g(x_m)(x_m)^{n-1}] + \dots + a_0 g(x_m)$$

$$= a_n \sum_{k=1}^m g(x_k)(x_k)^n + a_{n-1} \sum_{k=1}^m g(x_k)(x_k)^{n-1} + \dots + a_0 \sum_{k=1}^m g(x_k)(x_k)^0$$

$$= a_n b_{nn} + a_{n-1} b_{n-1, n-1} + \dots + a_0 b_{00} = b^T a$$

$$-2 \sum_{k=1}^m g(x_k) h(x_k) = -2b^T a$$

$$\begin{aligned} \sum_{k=1}^m h(x)^2 &= (a_n x_1^n + a_{n-1} x_1^{n-1} + \dots + a_0) (a_n x_1^n + a_{n-1} x_1^{n-1} + \dots + a_0) \\ &\quad \vdots \\ &\quad \vdots \\ &+ (a_n x_{m-1}^n + a_{n-1} x_{m-1}^{n-1} + \dots + a_0) (a_n x_{m-1}^n + a_{n-1} x_{m-1}^{n-1} + \dots + a_0) \\ &+ (a_n x_m^n + a_{n-1} x_m^{n-1} + \dots + a_0) (a_n x_m^n + a_{n-1} x_m^{n-1} + \dots + a_0) \end{aligned}$$

$$\begin{aligned} &= a_n x_1^{n+n} a_n + a_n x_1^{n+(n-1)} a_{n-1} + \dots + a_n x_1^n a_0 \\ &\quad \vdots \\ &\quad \vdots \\ &\quad \dots \dots \dots + a_0 x_1^n a_n + a_0 x_1^{n-1} a_{n-1} + \dots + a_0 x_1^0 a_0 \\ &+ a_n x_{m-1}^{n+n} a_n + a_n x_{m-1}^{n+(n-1)} a_{n-1} + \dots + a_n x_{m-1}^n a_0 \\ &\quad \vdots \\ &\quad \vdots \\ &\quad \dots \dots \dots + a_0 x_{m-1}^n a_n + a_0 x_{m-1}^{n-1} a_{n-1} + \dots + a_0 x_{m-1}^0 a_0 \\ &+ a_n x_m^{n+n} a_n + a_n x_m^{n+(n-1)} a_{n-1} + \dots + a_n x_m^n a_0 \\ &\quad \vdots \\ &\quad \vdots \\ &\quad \dots \dots \dots + a_0 x_m^n a_n + a_0 x_m^{n-1} a_{n-1} + \dots + a_0 x_m^0 a_0 \end{aligned}$$

$$= a_n \sum_{k=1}^m x_k^{n+n} a_n + a_n \sum_{k=1}^m x_k^{n+(n-1)} a_{n-1} + \dots + a_n \sum_{k=1}^m x_k^n a_0$$

$$+ a_{n-1} \sum_{k=1}^m x_k^{(n-1)+n} a_n + a_{n-1} \sum_{k=1}^m x_k^{(n-1)+(n-1)} a_{n-1} + \dots + a_{n-1} \sum_{k=1}^m x_k^{n-1} a_0$$

$$\dots \dots + a_0 \sum_{k=1}^m x_k^n a_n + a_0 \sum_{k=1}^m x_k^{n-1} a_{n-1} + \dots + a_0 \sum_{k=1}^m x_k^n a_0$$

$$= a_n q_{n,n} a_n + a_n q_{n,n-1} a_{n-1} + \dots + a_n q_{n,0} a_0$$

$$+ a_{n-1} q_{n-1,n} a_n + a_{n-1} q_{n-1,n-1} a_{n-1} + \dots + a_{n-1} q_{n-1,0} a_0$$

$$\dots \dots + a_0 q_{0,n} a_n + a_0 q_{0,n-1} a_{n-1} + \dots + a_0 q_{0,0} a_0$$

$$= (a_0, \dots, a_{n-1}, a_n) \begin{bmatrix} q_{0,0} & \dots & q_{0,n-1} & q_{0,n} \\ \vdots & & \vdots & \vdots \\ q_{n-1,0} & \dots & q_{n-1,n-1} & q_{n-1,n} \\ q_{n,0} & \dots & q_{n,n-1} & q_{n,n} \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \\ a_n \end{bmatrix}$$

$$= a^T Q a, \text{ donde } Q = [q_{ij}]$$

q.e.d.

por lo tanto:

$$f(a) = a^T Q a - 2b^T a + c$$

Las condiciones necesarias de primer orden establecen que el gradiente de f debe ser igual a cero. Esto lleva directamente al sistema de $n+1$ ecuaciones:

$$Qa = b$$

Las cuales pueden ser resueltas para determinar a .

Ejemplo 3: (Problemas de selección)

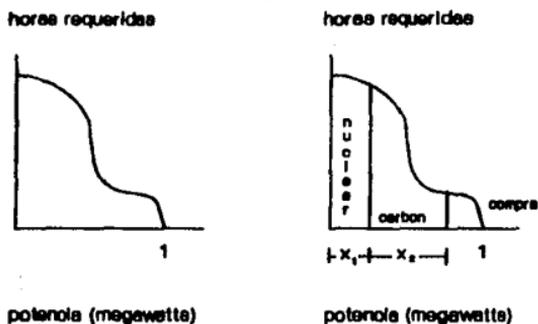
Frecuentemente es necesario seleccionar una clase de factores para conocer un conjunto dado de necesidades. Un ejemplo es el problema que afronta una planta eléctrica al seleccionar sus facilidades de generación de potencia.

El nivel de potencia que la compañía debe suministrar varía según la hora del día, el día de la semana, y según la estación del año. Sus necesidades de generación de potencia están resumidas en una curva, $h(x)$, como se muestra en la Fig. 3, la cual muestra el total de horas en un año que un nivel de potencia de al menos x_1 es requerido para cada x_i . Por conveniencia la curva es normalizada de manera que el límite superior sea unitario.

La compañía puede conocer estos requerimientos instalando equipo generador, tal como (1) nuclear o (2) carbón, o mediante la adquisición de potencia de una parrilla de energía central. Asociado con el tipo i ($i=1,2$) de equipo generador existe un costo de capital unitario anual b_i y un costo de operación unitario c_i . El precio unitario de la potencia comprada de la parrilla es g . Las plantas nucleares tienen un costo de capital más alto y un costo de operación menor, por lo que son utilizadas para suministrar una carga base.

Las plantas de carbón son usadas para niveles intermedios, y la potencia es comprada directamente sólo por períodos pico de demanda. Los requerimientos se satisfacen como se muestra en la siguiente figura:

Fig. 3



donde x_1 y x_2 denotan las capacidades de la planta nuclear y la planta de carbón respectivamente.

Por ejemplo, la potencia de la planta nuclear puede ser vista como compuesta de x_1/Δ pequeños generadores de capacidad Δ , donde Δ es pequeña. El primero de estos generadores se activa por $h(\Delta)$ horas, suministrando $\Delta h(\Delta)$ unidades de energía; el siguiente suministra $\Delta h(2\Delta)$ unidades, y así sucesivamente. La energía total suministrada por la planta nuclear se muestra en la gráfica.

El costo total es

$$f(x_1, x_2) = b_1 x_1 + b_2 x_2 + c_1 \int_0^{x_1} h(x) dx$$

$$+ c_2 \int_{x_1}^{x_1+x_2} h(x) dx + c_3 \int_{x_1+x_2}^1 h(x) dx$$

y la compañía desea minimizar esta función sobre el conjunto definido por $x_1 \geq 0$, $x_2 \geq 0$, $x_1 + x_2 \leq 1$.

Asumiendo que la solución se encuentra dentro de las restricciones, al igualar las derivadas parciales con cero, se obtienen dos ecuaciones

$$b_1 + (c_1 - c_2)h(x_1) + (c_2 - c_3)h(x_1 + x_2) = 0$$

$$b_2 + (c_2 - c_3)h(x_1 + x_2) = 0$$

que representan las condiciones necesarias.

Si $x_1 = 0$, entonces el teorema general de las condiciones necesarias muestra que la primera igualdad podría ser ≥ 0 . Así mismo, si $x_2 = 0$ entonces la segunda igualdad podría ser ≥ 0 . El caso $x_1 + x_2 = 1$ requiere un poco más de análisis (Ver ejemplo 2).

Ejemplo 4: (Control)

Problemas dinámicos, donde las variables corresponden a acciones tomadas en una secuencia de instantes de tiempo, frecuentemente pueden ser formulados como problemas de optimización irrestricta.

Como ejemplo, supóngase que la posición de un objeto largo es controlado por series de fuerzas de control correctivas. El error de posición (la distancia desde la posición deseada) está dada por la ecuación:

$$x_{k+1} = x_k + u_k$$

donde x_k es el error del instante de tiempo k , y u_k es la fuerza

efectiva aplicada en el tiempo u_k (después de ser normalizada para dar razón de la masa del objeto y la duración de la fuerza). El valor de x_0 es dado. La secuencia u_0, u_1, \dots, u_n debería elegirse de tal forma que minimice el objetivo:

$$J = \sum_{k=0}^n (x_k^2 + u_k^2)$$

esto representa un compromiso entre tener una deseada x_k igual a cero y el reconocimiento de que el control u_k es costosa.

El problema puede ser convertido en un problema irrestricto mediante la eliminación de variables $x_k, k=2, \dots, n$, de la función objetivo. Se puede ver fácilmente que.

$$x_k = x_0 + u_0 + u_1 + \dots + u_{k-1}$$

La función objetivo puede ser escrita como:

$$= \sum_{k=0}^n [(x_0 + u_0 + u_1 + \dots + u_{k-1})^2 + u_k^2]$$

Esta es una función cuadrática en las desconocidas u_k , tiene la misma estructura general como el ejemplo (2) y puede ser tratada de forma similar.

CAPITULO II

CAPITULO II OPTIMIZACION SIN RESTRICCIONES : METODOS DE BUSQUEDA.

II.1 METODOS DE OPTIMIZACION DE FUNCIONES UNIMODALES DE UNA SOLA VARIABLE EN PROBLEMAS NO RESTRINGIDOS.

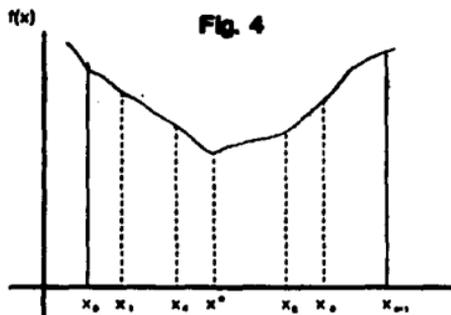
1a) METODO DE BUSQUEDA DE FIBONACCI

Los métodos analíticos son aplicables sólo a funciones diferenciables donde la evaluación de la función objetivo sobre el punto óptimo se hace hasta el final del proceso, a diferencia de los métodos numéricos de optimización cuyo proceso, es inverso al de los analíticos, en el sentido que primero se obtienen los valores de la función objetivo en varios puntos del dominio y después, mediante comparaciones de estos valores se hacen conclusiones sobre la optimalidad del problema.

Uno de los métodos numéricos de optimización muy popular es el método de búsqueda de Fibonacci, mismo que aparte de su popularidad tiene un cierto grado de elegancia teórica.

Para que este método determine un valor mínimo de cierta función f sobre un intervalo cerrado $[x_0, x_{n+1}]$, se basa en el supuesto de que f puede ser definida sobre un dominio extenso pero con la condicionante de que f sea una función unimodal, es decir, se tiene un mínimo relativo único en el intervalo de búsqueda, como

se muestra en la siguiente figura:



Si se tiene que f cumple con la anterior condicionante (ser función unimodal) y se desea obtener el punto mínimo de la función con cierto grado de aproximación, con sólo tener n puntos con sus respectivas imágenes se puede cumplir el objetivo de minimizar f , la interrogante que surge es cómo seleccionar los n puntos, siendo esto parte del principio del Método de Fibonacci planteado de la siguiente manera:

De acuerdo a la Fig. 4:

$$\text{Sea } x_2 - x_1 = L_1 \quad \text{y} \quad x_3 - x_2 = L_0 - L_1$$

$$\text{Donde } L_1 > L_0 - L_1 \quad \text{y} \quad x_3 - x_2 < x_4 - x_3$$

Planteando un análisis comparativo entre las imágenes $f(x_2)$ y $f(x_4)$ se generan los siguientes tres casos:

1) $f(x_4) < f(x_2)$

Es importante recordar que la hipótesis planteada para f es que tenga un solo punto mínimo en el intervalo de incertidumbre, es decir que f sea una función unimodal. En este caso el intervalo de incertidumbre se reduce a $[x_1, x_2]$; ya que no es posible que en el intervalo $[x_2, x_3]$ exista un mínimo, debido a que la función de x_2 a x_4 es decreciente ($f(x_4) < f(x_2)$) y para que existiera un mínimo en $[x_2, x_3]$ sería necesario que la función f de x_2 a x_3 decreciera, lo anterior sólo sería posible si en el intervalo (x_3, x_4) existiera un máximo, lo cual contradice la hipótesis planteada.

Concluyendo, sobre lo anterior, bajo las bases de la Fig.4 y que f es unimodal y además cuando $f(x_4) < f(x_2)$ el intervalo de incertidumbre de $[x_1, x_3]$ se reduce a $[x_1, x_2]$.

2) $f(x_4) > f(x_2)$

En este caso la reducción del intervalo de incertidumbre de $[x_1, x_3]$ es a $(x_4, x_3]$, donde el razonamiento para fundamentar esta reducción es análogo al caso anterior.

3) $f(x_4) = f(x_2)$

Este caso se deja como ejercicio para el lector.

Es evidente que el caso más conveniente del análisis anterior para poder lograr una mayor reducción en el intervalo de incertidumbre, sería el caso 2), puesto que el intervalo (x_4, x_5) resulta ser menor que el intervalo (x_1, x_2) . De lo anterior surge la interrogante: ¿Depende de la posición de los puntos interiores entre sí para lograr una rapidez en la reducción del intervalo de incertidumbre? La respuesta afirmativa a esta interrogante es indiscutible. Donde dicha situación resulta ser totalmente inconveniente, para lograr una convergencia uniforme hacia el mínimo.

Una forma de resolver el problema anterior es colocar los puntos interiores simétricamente dentro del intervalo, es decir $x_3 - x_2 = x_4 - x_1$, lo cual es indiferente si se trata del caso 1) o del 2) debido a que la reducción de incertidumbre es la misma. Por lo tanto para la siguiente iteración el punto x_6 se coloca simétricamente a x_2 o x_4 dependiendo del caso, para seguir con el mismo principio.

Con los anteriores razonamientos se intuye cierto control en la posición de los puntos a evaluar, para formalizar lo anterior, supóngase que se sigue el procedimiento de reducción de incertidumbre hasta el último intervalo como se muestra en la siguiente figura:

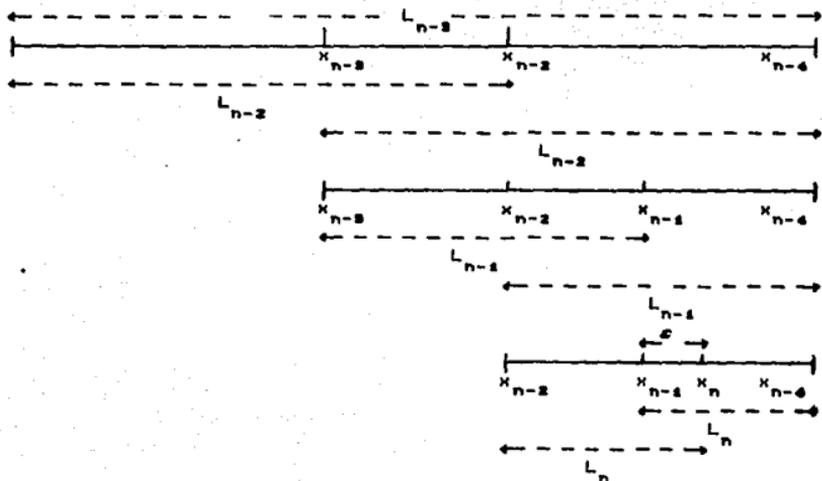


Fig. 5

Deduciendo gráficamente se tiene que:

$$L_{n-1} = 2 L_n - \varepsilon \quad \dots \quad \text{IIIa.1}$$

Demostración:

$$L_n = x_{n-4} - x_{n-1} \quad \circ \quad L_n = x_n - x_{n-2}$$

$$\varepsilon = x_n - x_{n-1}$$

Recordando la secuencia de los números Fibonacci:

$F_0 = 1, F_1 = 1$ y $F_k = F_{k-1} + F_{k-2}$ para $k = 2, 3,$

es decir: $F_0 = 1, F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, F_5 = 8, F_6 = 13,$
 $F_7 = 21, \dots$

Por lo tanto las anteriores igualdades se pueden generalizar bajo la siguiente fórmula:

$$L_{n-j} = F_{j+1} L_n - F_{j-1} \epsilon; \quad j=1, 2, \dots, n-1$$

Analícese para el caso $j=n-1$:

$$L_1 = F_n L_n - F_{n-2} \epsilon$$

despejando L_n :

$$L_n = \frac{L_1}{F_n} + \epsilon \frac{F_{n-2}}{F_n} \quad \dots \dots \dots \text{IIIIa.3}$$

Si $\epsilon = 0$ implica que al efectuar las "n" evaluaciones en el intervalo original de incertidumbre se reduce a una fracción $1/F_n$ de su valor.

Para iniciar el procedimiento es necesario obtener el valor de L_2 y de esa forma colocar los dos puntos a L_2 unidades de los

extremos respectivamente.

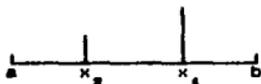
$$L_2 = F_{n-1} L_n - e F_{n-2}$$

Por lo tanto:
$$L_2 = \frac{F_{n-1}}{F_n} L_1 + \frac{(-1)^{n-2}}{F_n}$$

Demostración (Ver Apéndice A)

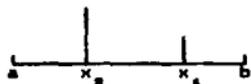
Una vez analizada la regla simétrica, lo que sólo queda por plantear es en qué forma van quedando los extremos de los intervalos de incertidumbre, los cuales están en función de la posición de los puntos interiores y sus respectivas imágenes. Siendo sólo cuatro casos posibles:

- 1) Si $x_1 > x_2$ y $f(x_1) > f(x_2)$



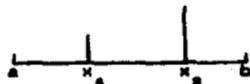
Nuevo intervalo de incertidumbre (a, x_1)

- 2) Si $x_1 > x_2$ y $f(x_1) < f(x_2)$



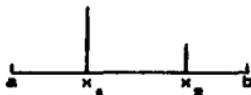
Nuevo intervalo de incertidumbre (x_2, b)

- 3) Si $x_1 < x_2$ y $f(x_1) < f(x_2)$



Nuevo intervalo de incertidumbre (a, x_2)

4) Si $x_1 < x_2$ y $f(x_1) > f(x_2)$



Nuevo intervalo de incertidumbre (x_1, b)

1b) METODO DE SECCION DORADA:

Para el planteamiento de este método es necesario tener las mismas hipótesis del método de Búsqueda de Fibonacci es decir la función a minimizar deberá ser unimodal, además utiliza el mismo análisis de reducción de los intervalos de incertidumbre, pero existe una gran diferencia entre estos dos métodos; el método de Sección Dorada no necesita fijar el número de iteraciones n para obtener la primera reducción del intervalo de incertidumbre, mientras que en el método de Búsqueda de Fibonacci es fundamental determinar n para obtener tal reducción.

Retomando las siguientes ecuaciones vistas en el tema anterior:

$$L_{j-1} = L_j + L_{j+1} \quad \dots \text{IIIb.1}$$

$$L_{n-1} = 2L_n - \epsilon \quad \dots \text{IIIb.2}$$

suponiendo que

$$\frac{L_1}{L_2} = \dots = \frac{L_{j-1}}{L_j} = \frac{L_j}{L_{j+1}} = \frac{L_{j+1}}{L_{j+2}} = \dots = \frac{L_{n-1}}{L_n} = k \quad \dots \text{IIIb.3}$$

dividiendo IIIb.1 entre L_j :

$$\frac{L_{j-1}}{L_j} = 1 + \frac{L_{j+1}}{L_j} \quad \dots \text{IIIb.4}$$

sustituyendo IIIb.3 en IIIb.4 :

$$k = 1 + \frac{1}{k} \quad \dots \text{IIIb.5}$$

donde la solución factible de la ecuación cuadrática de IIIb.5 es:

$$k = \frac{1 + \sqrt{5}}{2} \approx 1.618033$$

de la ecuación IIIb.3:

$$\begin{aligned} \frac{L_1}{L_2} &= k \\ \frac{L_1}{L_2} * \frac{L_2}{L_3} &= \frac{L_1}{L_3} = k^2 \\ \vdots & \\ \frac{L_1}{L_2} * \frac{L_2}{L_3} * \dots * \frac{L_{j+1}}{L_{j+2}} &= \frac{L_1}{L_{j+2}} = k^{j+1} \\ \vdots & \\ \frac{L_1}{L_2} * \frac{L_2}{L_3} * \dots * \frac{L_{j+1}}{L_{j+2}} * \dots * \frac{L_{n-1}}{L_n} &= \frac{L_1}{L_n} = k^{n-1} \end{aligned}$$

por lo tanto la reducción lograda con respecto al intervalo inicial es $\frac{1}{k^{n-1}}$:

$$L_n = \frac{L_1}{k^{n-1}} \quad \dots \text{IIIb.6}$$

recurriendo a la siguiente igualdad vista en el método de Búsqueda de Fibonacci y aplicando el límite cuando $n \rightarrow \infty$ se tiene:

$$\lim_{n \rightarrow \infty} \left[L_n = \frac{F_{n-1}}{F_n} L_1 + \frac{(-1)^n e}{F_n} \right]$$

$$\lim_{n \rightarrow \infty} \frac{L_2}{L_1} = \lim_{n \rightarrow \infty} \frac{F_{n-1}}{F_n} + \lim_{n \rightarrow \infty} \frac{(-1)^n \epsilon}{F_n} \frac{1}{L_1}$$

$$\lim_{n \rightarrow \infty} \frac{F_{n-1}}{F_n} = \lim_{n \rightarrow \infty} \frac{L_2}{L_1} = \frac{1}{k} \quad \dots \text{Ilib.7}$$

despejando: $L_2 = \frac{L_1}{k}$

Se puede deducir de lo anterior que bajo la suposición de que la reducción de los intervalos es proporcional, dentro de los casos del Método de Fibonacci, la mejor aproximación al mínimo sería cuando $n \rightarrow \infty$ y aplicando el límite, el resultado óptimo para Fibonacci es la proporción de la reducción de la constante "k".

Se puede afirmar que la hipótesis fundamental de este método es la proporción que guardan los cocientes:

$$\frac{L_1}{L_2} = \dots = \frac{L_{j-1}}{L_j} = \frac{L_j}{L_{j+1}} = \frac{L_{j+1}}{L_{j+2}} = \dots = \frac{L_{n-1}}{L_n} = k, \text{ complementando con}$$

la igualdad $L_2 = \frac{F_{n-1}}{F_n} L_1 + \frac{(-1)^n \epsilon}{F_n}$ vista en el método de Búsqueda

de Fibonacci y haciendo tender n a ∞ se concluye la relación Ilib.7, que marca el principio básico del método de Sección Dorada.

Se resume que el método de Sección Dorada aplica la misma secuencia iterativa del método de Búsqueda de Fibonacci con la gran diferencia de que la Sección Dorada utiliza la fórmula de reducción de incertidumbre $L_n = \frac{L_1}{k}$ mientras que Fibonacci

establece:

$$L_n = \frac{F_{n-1}}{F_n} L_1 + \frac{(-1)^n \epsilon}{F_n}$$

II.2: METODOS DE OPTIMIZACION DE FUNCIONES MULTIMODALES DE UNA SOLA VARIABLE EN PROBLEMAS NO RESTRINGIDOS.

2a) INTERPOLACION CUADRATICA

Este método está fundamentado a partir de tres puntos de la función a minimizar, donde dichos puntos pueden generar un polinomio, y la solución de tal polinomio es una aproximación al punto óptimo de la función, de ahí su nombre de Interpolación Cuadrática, también conocido como Método de la Parábola.

Una vez obtenida la solución del polinomio se procede a aplicar un procedimiento para descartar uno de los cuatro puntos, con los puntos seleccionados se generará otro polinomio y así sucesivamente hasta satisfacer la precisión deseada.

Para formalizar lo anterior se inicia con el desarrollo de la teoría que sustenta este método:

Recordando la fórmula general de un polinomio de segundo grado $Ax^2 + Bx + C = f(x)$, supóngase que existen tres puntos x_1, x_2, x_3 donde $x_1 \neq x_2, x_1 \neq x_3, x_2 \neq x_3$ y sus respectivas imágenes $f(x_1), f(x_2)$ y $f(x_3)$, sustituyendo estos valores en la fórmula general del polinomio cuadrático se obtiene el siguiente sistema de ecuaciones.

$$\begin{aligned} Ax_1^2 + Bx_1 + C &= f(x_1) \\ Ax_2^2 + Bx_2 + C &= f(x_2) \end{aligned} \quad \dots \quad \text{II2a.1}$$

$$Ax^2 + Bx + C = f(x)$$

donde

$$A = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{bmatrix} \quad D = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix} \quad X = \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

implica que: $X = A^{-1}D$

el polinomio resultante es:

$$\frac{A}{[(x_1^2 - x_2^2)f(x_3) + (x_1^2 - x_3^2)f(x_2) + (x_2^2 - x_3^2)f(x_1)]} x^2$$

$$\frac{(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)}{(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)}$$

$$\frac{B}{[(x_1^2 - x_2^2)f(x_1) + (x_2^2 - x_3^2)f(x_2) + (x_3^2 - x_1^2)f(x_3)]} x$$

$$\frac{(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)}{(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)}$$

$$\frac{C}{(x_1 x_2)(x_2 - x_3)f(x_1) + (x_1 x_3)(x_3 - x_2)f(x_2) + (x_2 x_3)(x_3 - x_1)f(x_3)} = D$$

$$\frac{(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)}{(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)}$$

por lo tanto la solución está dada por

$$x_4 = \frac{1}{2} \frac{(x_1^2 - x_2^2)f(x_3) + (x_2^2 - x_3^2)f(x_2) + (x_3^2 - x_1^2)f(x_1)}{(x_1 - x_2)f(x_1) + (x_2 - x_3)f(x_2) + (x_3 - x_1)f(x_3)} \quad \dots \quad II2a.2$$

$$x_4 \text{ es un m\u00ednimo si } 1 - \frac{(x_2 - x_3)f(x_1) + (x_3 - x_1)f(x_2) + (x_1 - x_2)f(x_3)}{(x_1 - x_2)(x_2 - x_3)(x_3 - x_1)} > 0$$

El m\u00e9todo puede ser utilizado en particular para funciones de una variable. Lo anterior puede ser de utilidad para llevar a cabo la b\u00fasqueda lineal requerida en los m\u00e9todos de gradientes.

En esos casos se desea encontrar el m\u00ednimo $f(x)$ en los puntos de la l\u00ednea $x_0 + \lambda d$, donde x_0 es un punto dado y d una direcci\u00f3n dada. Los valores de $f(x_0 + \lambda d)$ en esa l\u00ednea son funciones de variable λ es decir $g(\lambda) = f(x_0 + \lambda d)$.

Las ideas y resultados anteriores se describen en un proceso iterativo tal y como sigue. Se asume que se tiene una funci\u00f3n unimodal $f(x)$ de una sola variable, una aproximaci\u00f3n inicial hacia la posici\u00f3n donde se encuentra el punto \u00f3ptimo, una medida de magnitud D , donde dicha distancia es la que separa el punto x inicial del punto \u00f3ptimo x^* .

A continuaci\u00f3n se describen los pasos del algoritmo:

- 1) Calcular $f(x_1)$, y $f(x_1 + D)$ donde $x_2 = x_1 + D$
- 2) Si $f(x_1) < f(x_1 + D)$ entonces $x_3 = x_1 - D$, en caso contrario $x_3 = x_1 + 2D$ y encontrar $f(x_3)$
- 3) Utilizar los tres puntos para obtener x_4 con la f\u00f3rmula II2a.2

- 4) Si la diferencia entre las posiciones del valor de la función más pequeño y el siguiente valor de la función más pequeño, es menor que la aproximación requerida termina el proceso.
- 5) Si el proceso no ha terminado en la etapa 4), se procede a descartar un punto. La lógica indica descartar el punto de mayor imagen, pero si al retener el punto con mayor imagen se situa un intervalo definitivo en la localización del óptimo, entonces realmente se debe retener ese valor y regresar a la etapa 3). Lo anterior se puede analizar en la figura 6, en la cual se quiere retener x_1, x_2 , y x_4 en vez de x_1, x_2 , y x_3 .

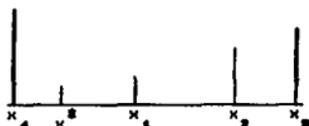


Fig. 6

La práctica ha establecido que cuando la aproximación E es demasiado pequeña x_1, x_2 , y x_3 con $f(x_1), f(x_2)$ y $f(x_3)$ estarían muy cerca entre sí y x_4 bajo la ecuación II2a.2, podría dejar de ser confiable. Para poder solucionar este problema se reestablece la ecuación II2a.2 de la siguiente forma:

$$x_4 = \frac{1}{2} (x_1 + x_2) + \frac{1/2 (f(x_1) - f(x_2))(x_2 - x_3)(x_3 - x_1)}{(x_3 - x_1)f(x_1) + (x_2 - x_1)f(x_2) + (x_2 - x_3)f(x_3)} \dots \text{II2a.3}$$

Para la segunda iteración en las interpolaciones subsiguientes.

Nota: Ver demostración de la ecuación II2a.3 en el Apéndice

Nota: Ver demostración de la ecuación II2a.3 en el Apéndice Técnico B.

2b) INTERPOLACION CUBICA

La idea básica de este método es tomar dos puntos p y q y de acuerdo a la función y su gradiente, se establece una aproximación a un polinomio cúbico, a través del cual se busca al punto mínimo.

Considerando el problema de minimizar $f(x)$ en puntos de línea $x_0 + \lambda d$, es decir:

$$\min \phi(\lambda) = f(x_0 + \lambda d)$$

donde x_0 es el punto de partida y d una dirección factible

$$f(x_0 + \lambda d) = f(x_{01} + \lambda d_1, x_{02} + \lambda d_2, \dots, x_{0n} + \lambda d_n) \quad y$$

$$\frac{d\phi}{d\lambda} = \nabla f(x_0 + \lambda d)^T d = g(x_0 + \lambda d)^T d$$

Ahora supóngase que se conocen los siguientes valores:

$$\phi(p) = \phi_p$$

$$\phi(q) = \phi_q$$

$$\frac{d\phi(p)}{d\lambda} = g_p$$

$$\frac{d\phi(q)}{d\lambda} = g_q$$

donde p y q serán los extremos del intervalo a analizar.

Para poder simplificar la teoría se hace $p=0$, es decir la búsqueda se hará en el intervalo $(0, q)$

$\phi(\lambda)$ se aproximará al siguiente polinomio

$$\phi(\lambda) \cong a + b\lambda + c\lambda^2 + d\lambda^3 \quad \dots \quad \text{II2b.1}$$

$$\phi'(\lambda) = b + 2c\lambda + 3d\lambda^2 \quad \dots \quad \text{II2b.2}$$

Valuando en p y q y haciendo p=0 se tiene

$$\phi(p) = a = \phi_p$$

$$\phi'(p) = b = G_p$$

$$\phi(q) = a + bq + cq^2 + dq^3 = \phi_q$$

$$\phi'(q) = b + 2cq + 3dq^2 = G_q$$

que es un sistema de cuatro ecuaciones con cuatro incógnitas, con el resultado siguiente:

$$a = \phi_p; \quad b = G_p; \quad c = -\frac{(G_p + z)}{q}; \quad d = \frac{G_p + G_q + 2z}{3q^2}$$

$$\text{donde } z = \frac{3(\phi_p - \phi_q)}{q} + G_p + G_q$$

De acuerdo a las condiciones de primer orden, la primera derivada debe ser igual a cero. Sustituyendo queda como sigue:

$$G_p - 2(G_p + z) \frac{\lambda}{q} + (G_p + G_q + 2z) \left[\frac{\lambda}{q} \right]^2 = 0$$

Esta última ecuación de forma cuadrática tiene la siguiente solución r:

$$\frac{r}{q} = \frac{(G_p + z) \pm \left[(G_p + z)^2 - G_p (G_p + G_q + 2z) \right]^{1/2}}{G_p + G_q + 2z}$$

$$\text{si } w = (z^2 - G_p G_q)^{1/2} \quad \text{entonces}$$

$$r = \left[\frac{G_p + z + w}{G_p + G_q + 2z} \right] q$$

$$r = \left[\frac{z + w - G_p}{G_q - G_p + 2w} \right] q \quad \dots \quad \text{II2b.3}$$

La demostración de la equivalencia entre estas dos últimas ecuaciones se encuentra en el Apéndice Técnico C.

Ahora, por las condiciones de segundo orden, la segunda derivada es (de II2b.2)

$$\phi''(\lambda) = 2c + 6d\lambda$$

Tomando w con signo positivo y sustituyendo II2b.3 en esta última expresión se tiene:

$$-2 \frac{(G_p + z)}{q} + \frac{2(G_p + G_q + 2z)}{q^2} \frac{q(G_p + z + w)}{G_p + G_q + 2z} =$$

$$\frac{1}{q} \left[-2G_p - 2z + 2G_p + 2z + 2w \right] = \frac{2w}{q}$$

como $\frac{2w}{q} > 0$, se trata de un mínimo.

Si $G_p < 0$ se toma q positiva, es decir hacia el decrecimiento de la función. El tamaño de q debe ser tal que el intervalo de $(0, q)$ contenga al punto mínimo. Esto es posible sí:

- a) $\phi_q > \phi_p$ ó
 b) $G_q > 0$

Considérense las siguientes figuras:

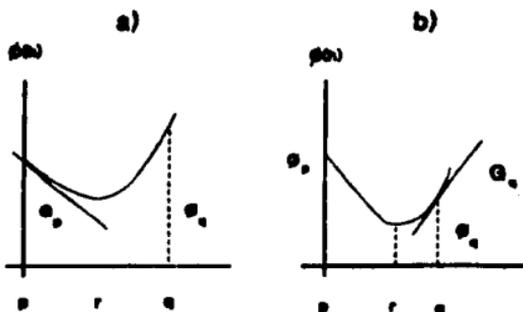


Fig. 7

Cuando no se cumple ninguna de estas dos condiciones, se duplica el valor de q , tantas veces como sea necesario hasta que el intervalo incluya al mínimo.

La elección de q es a discreción, sin embargo Davidon, Fletcher y Powell sugieren

$$q = \min \left\{ \eta, -2(\phi_p - \phi_m) / G_p \right\} \quad \dots \quad \text{II2b.4}$$

donde ϕ_m es un estimado, de preferencia abajo del valor mínimo de $\phi(\lambda)$, y η es una constante, generalmente 2 o 1.

Algoritmo

1) Dado x_0 y d

$$\text{Encuentre } \phi_p = f(x_0) \text{ y } G_p = [g(x_0)]^T d$$

2) Verifique que $G_p < 0$, si no busque a través de $-d$.

Escoja q mediante la ecuación II2b.4

3) Evalúe $\phi_q = f(x_0 + qd)$ y $G_q = [g(x_0 + qd)]^T d$

4) Si $G_q > 0$ ó $\phi_q > \phi_p$ se tiene un mínimo en el intervalo. Si no es así sustituya q por $2q$ y regrese al paso 3.

5) Utilice II2b.3 para aproximar el mínimo dentro de (p, q)

6) Si $\left| \frac{d\phi}{d\lambda} \right| = |G_r| < \epsilon$

donde ϵ es el grado de precisión, deténgase. Si no es así, regrese al paso 5 utilizando el intervalo (p, r) , si $G_r > 0$ o el intervalo (r, q) si $G_r \leq 0$.

11.3 METODOS DE OPTIMIZACION QUE UTILIZAN DERIVADAS PARA FUNCIONES DE VARIAS VARIABLES EN PROBLEMAS NO RESTRINGIDOS.

3a) METODO DE NEWTON CON APROXIMACION CUADRATICA

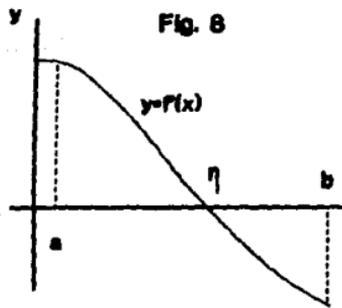
Para analizar la teoría de este método, se estudiará primeramente funciones de una sola variable, ya que esto facilitará el entendimiento de este método.

Considérese la función $f(x)$, la cual se optimizará. Por las condiciones de primer orden se debe cumplir que

$$f'(x) = 0 \dots \quad \text{II3a.1}$$

El problema de resolver esta ecuación puede a veces no ser tan sencillo, en este caso el Método de Newton puede ser de gran utilidad.

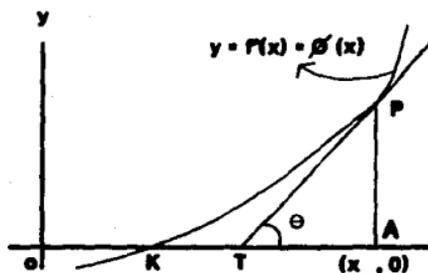
Se sabe que si se tienen dos valores a y b tales que $f'(a)$ y $f'(b)$ tienen signos opuestos y que además se cumplen ciertas hipótesis de continuidad, entonces existe una raíz η de la ecuación II3a.1 con $a < \eta < b$.



Lo que el método de Newton proporciona es una aproximación a la raíz de la función $f'(x)$.

Veáse desde un punto de vista geométrico, el procedimiento que establece el método.

Fig. 9



Sea PT la tangente en P a la curva $\phi(x)$. T es el punto donde la tangente cruza el eje de las x y k es la raíz de $\phi(x)$. Como puede observarse \overline{OT} es una aproximación a k . Encontrando el valor de esta aproximación:

$$\overline{OT} = \overline{OA} - \overline{TA} = x_0 - \overline{TA}$$

$$\tan \theta = \frac{\overline{PA}}{\overline{TA}} = \phi'(x_0)$$

despejando

$$\overline{TA} = \frac{\overline{PA}}{\phi'(x_0)}$$

como $\overline{PA} = \phi(x_0)$

$$\overline{TA} = \frac{\phi(x_0)}{\phi'(x_0)}$$

$$\overline{OT} = x_0 - \frac{\phi(x_0)}{\phi'(x_0)}$$

$$x_1 = x_0 - \frac{\phi(x_0)}{\phi'(x_0)}$$

en general

$$x_{p+1} = x_p - \frac{\phi(x_p)}{\phi'(x_p)} \quad \dots \quad \text{II3a.2}$$

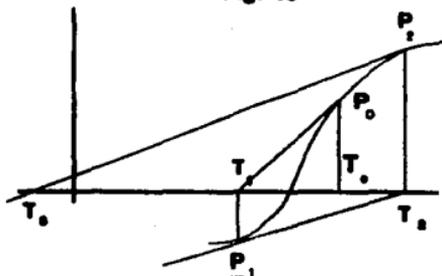
La ecuación II3a.2 muestra la regla que se utiliza para encontrar un punto mínimo, ya que como se puede observar en la Fig. 9 cuando la derivada (en este caso $\phi'(x)$), es negativa antes de llegar al punto óptimo, y después de él se vuelve positiva, se trata de un mínimo.

El mismo razonamiento puede utilizarse para encontrar un máximo, esto conduciría a la fórmula siguiente:

$$x_{p+1} = x_p + \frac{\phi(x_p)}{\phi'(x_p)}$$

Este método puede fallar si la primera aproximación a la raíz es tal que el valor de $\frac{\phi(x_0)}{\phi'(x_0)}$ no es lo suficientemente pequeño.

Fig. 10



Una forma de evitar esto es utilizar el método de Von Mises en el cual se sustituye $\phi'(x_p)$ por $\phi'(x_0)$. Lo que significa que se calcula la primera tangente y posteriormente se obtienen las aproximaciones por medio de paralelas a ésta. Este método se vuelve mas lento, pero es una buena solución.

Regresando al método de Newton, supóngase que $\phi(x)$ tiene derivadas en todos los órdenes, entonces se puede escribir $\phi(x)$ utilizando la serie de Taylor.

$$\phi(x_{p+1}) = \phi(x_p) + \frac{(x_{p+1} - x_p)}{1!} \phi'(x_p) + \frac{(x_{p+1} - x_p)^2}{2!} \phi''(x_p) + \dots \quad \text{II3a.3}$$

como x_{p+1} es el punto óptimo la primera derivada debe ser igual a cero, i.e.

$$\phi(x_{p+1}) \approx 0$$

$$\phi(x_p) + \frac{(x_{p+1} - x_p)}{1!} \phi'(x_p) + R \approx 0$$

despreciando R:

$$\phi(x_r) + (x_{r+1} - x_r) \phi'(x_r) = 0$$

despejando x_{r+1} :

$$x_{r+1} = x_r - \frac{\phi(x_r)}{\phi'(x_r)}$$

De esta forma se ha llegado a la misma fórmula II3a.2.

Utilizando la fórmula de Taylor para funciones de varias variables se tiene:

$$\begin{aligned} \phi(x_{r+1}) &= \phi(x_r) + \nabla\phi(x_r)(x_{r+1} - x_r) + \frac{1}{2}(x_{r+1} - x_r)^T F(x_r)(x_{r+1} - x_r) + R \\ \therefore x_{r+1} &= x_r - \frac{\phi(x_r)}{\nabla\phi(x_r)} \end{aligned}$$

Considérese la serie de Taylor, pero utilizando los tres primeros términos:

$$\begin{aligned} \phi(x_r) + \frac{(x_{r+1} - x_r)}{1!} \phi'(x_r) + \frac{(x_{r+1} - x_r)^2}{2!} \phi''(x_r) &\cong 0 \\ \phi(x_r) + (x_{r+1} - x_r) \left[\phi'(x_r) + \frac{(x_{r+1} - x_r)}{2} \phi''(x_r) \right] &\cong 0 \quad \dots \text{II3a.4} \end{aligned}$$

De II3a.2 se tiene que:

$$x_{r+1} - x_r = - \frac{\phi(x_r)}{\phi'(x_r)}$$

sustituyendo en II3a.4 el término $(x_{r+1} - x_r)$ que está dentro del paréntesis

$$\begin{aligned} \phi(x_r) + (x_{r+1} - x_r) \left[\phi'(x_r) - \frac{1}{2} \frac{\phi(x_r) \phi''(x_r)}{\phi'(x_r)} \right] &\cong 0 \\ x_{r+1} &= x_r + G \end{aligned}$$

donde

$$G = \frac{-\phi(x_p)}{\phi'(x_p) - \frac{1}{2} \frac{\phi(x_p) \phi''(x_p)}{\phi'(x_p)^2}}$$

por simplificación el inverso de G

$$\frac{1}{G} = -\frac{\phi'(x_p)}{\phi(x_p)} + \frac{1}{2} \frac{\phi''(x_p)}{\phi'(x_p)}$$

Este método es el llamado Método de Newton con aproximación cuadrática, o método de Newton de segundo orden.

3b) DESCENSO ACELERADO: CASO CUADRÁTICO

Este es un método muy importante desde el punto de vista teórico, puesto que parte de un análisis simple y llega a un resultado satisfactorio, de ahí que muchos algoritmos más avanzados son frecuentemente motivados por un intento a modificar la técnica de búsqueda basada en el descenso acelerado en tal forma que el nuevo algoritmo tenga propiedades de convergencia superior.

El método de descenso acelerado también es conocido con el nombre de método del gradiente.

Se puede afirmar que el principio de este método se basa en direcciones de curso ajustado es decir se modifica la dirección de búsqueda del método a fin de que cada paso de búsqueda por el mínimo se lleve a cabo a lo largo de la mejor dirección. De ahí, seguir la dirección opuesta a la dirección del gradiente, ya que ésta tiene un cierto recurso intuitivo en la búsqueda del mínimo. La dirección del gradiente es la dirección de ascenso acelerado, la dirección opuesta es así la dirección de descenso acelerado.

Para formalizar la propiedad anterior, se iniciará el desarrollo de la teoría que sustenta este método:

Sea $x \in E^n$ y un punto cercano $x+hd \in E^n$, donde d es alguna

**ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA**

dirección y cierta longitud. Si experimentamos un desplazamiento de (x_1, x_2, \dots, x_n) a $(x_1 + \delta x_1, x_2 + \delta x_2, \dots, x_n + \delta x_n)$ por lo tanto:

$$\delta x_l = h d_l, \quad l=1, \dots, n \quad \dots\dots \text{I13b.1}$$

donde d_l son los cosenos directores de d tal que:

$$\sum_{l=1}^n d_l^2 = 1 \quad \dots\dots \text{I13b.2}$$

El cambio en el valor de la función está dado por:

$$df = f(x_1 + \delta x_1, x_2 + \delta x_2, \dots, x_n + \delta x_n) - f(x_1, x_2, \dots, x_n)$$

Aplicando el cálculo vectorial (Serie de Taylor):

$$\begin{aligned} f(x_0 + k) - f(x_0) &= \sum_{l=1}^n k_l \frac{df(x_1, \dots, x_n)}{dx_l} + \frac{1}{2!} \sum_{l=1}^n \sum_{j=1}^n k_l k_j \frac{d^2 f(x_1, \dots, x_n)}{dx_l dx_j} + \dots \\ &= k^T \nabla f(x_0) + \frac{1}{2} k^T F(x_0) k + \dots \end{aligned}$$

Por lo tanto:

$$df = \frac{df}{dx_1} \delta x_1 + \frac{df}{dx_2} \delta x_2 + \dots + \frac{df}{dx_n} \delta x_n \quad \dots\dots \text{I13b.3}$$

para primer orden en los δx_l , donde las derivadas parciales son evaluadas en x .

Ahora la interrogante que surge es ¿Cómo elegir los d_l sujetos a la ecuación I13b.2 tal que se obtenga el valor más largo posible para df ? Lo anterior involucra problemas de maximización restringida, para solucionarlo, recórrase al método de

multiplicadores de Lagrange definidos:

$$\phi(d_1, d_2, \dots, d_n) = df + \lambda \left(\sum_{l=1}^n d_l^2 - 1 \right)$$

Aplicando la ecuación II3b.1 y II3b.3

$$\phi(d_1, d_2, \dots, d_n) = h \left(\frac{df}{dx_1} d_1 + \frac{df}{dx_2} d_2 + \dots + \frac{df}{dx_n} d_n \right) + \lambda (d_1^2 + d_2^2 + \dots + d_n^2 - 1)$$

El máximo de df sujeto a la restricción (ecuación II3b.2) surge cuando $\phi(d_1, d_2, \dots, d_n)$ es maximizado:

$$\frac{d\phi}{dd_l} = h \frac{df}{dx_l} + 2\lambda d_l \quad l=1, \dots, n \quad \dots \quad \text{II3b.4}$$

$$\text{cuando} \quad \frac{d\phi}{dd_l} = 0 \quad \Rightarrow \quad d_l = - \frac{h}{2\lambda} \frac{df}{dx_l} \quad \dots \quad \text{II3b.5}$$

$$\therefore \frac{d_1}{\frac{df}{dx_1}} = \frac{d_2}{\frac{df}{dx_2}} = \dots = \frac{d_n}{\frac{df}{dx_n}} \quad \dots \quad \text{II3b.6}$$

donde d_l es directamente proporcional a $\frac{df}{dx_l}$ y la dirección d es paralela a $\nabla f(x)$ en x .

Así el incremento local más grande en la función para un paso pequeño h dado, ocurre cuando d está en dirección del $\nabla f(x)$. De esta manera la dirección del descenso acelerado está en dirección de $-\nabla f(x)$.

Reescribiendo la ecuación II3b.3 de la siguiente forma:

$$df = |\nabla f(x)| |dx| \cos\theta$$

donde θ es el ángulo entre $\nabla f(x)$ y dx , si se selecciona $\theta = 180^\circ$ dx estará en dirección $-\nabla f(x)$, para una magnitud dx que minimiza df .

La dirección del gradiente es ortogonal al contorno de la función en cualquier punto, para que en un contorno el valor de la función no cambie, es necesario que (d_1, d_2, \dots, d_n) sea una pequeña medida a lo largo del contorno (Ver Fig. 11):

$$f(x_1 + d_1, x_2 + d_2, \dots, x_n + d_n) \cong f(x_1, x_2, \dots, x_n)$$

$$\therefore df = \sum_{l=1}^n \frac{df}{dx_l} d_l = \nabla f(x)^T d = 0 \quad \dots \text{II3b.7}$$

El método de descenso acelerado busca explotar esta propiedad de la dirección del gradiente. Así, al estar en el punto x_l en cualquier paso en el proceso se busca el mínimo de la función a lo largo de la dirección $-\nabla f(x_l)$.

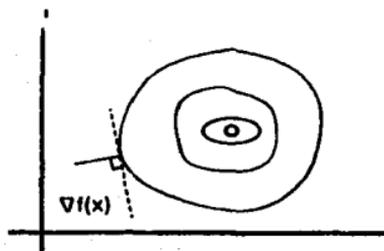


Fig. 11

En la iteración i se tiene una aproximación x_i para el punto mínimo, la siguiente aproximación sería:

$$x_{i+1} = x_i - \lambda \nabla f(x_i)$$

donde λ_i es el valor λ que minimiza:

$$\phi(\lambda) = f(x_i - \lambda \nabla f(x_i)) \quad \dots \quad 113b.8$$

La ecuación 113b.8 puede ser resuelta por los métodos definidos en el Tema II.1 o Tema II.2, eligiendo el Método de Interpolación Cuadrada para obtener los λ_i .

Para poder facilitar el entendimiento del proceso iterativo de este método se presenta el siguiente algoritmo:

- 1) Dar el valor inicial de x_0 .
- 2) Obtener la dirección $d_i = -\nabla f(x_i)$.

- 3) Encontrar λ_i que minimiza a $f(x_i + \lambda_i d_i)$.
- 4) Generar el nuevo punto $x_{i+1} = x_i + \lambda_i d_i$
- 5) Preguntar si x_{i+1} es el mínimo
si es falso: regresar al paso 2)
en caso contrario: El punto óptimo $x^* = x_{i+1}$.

3c). DIRECCIONES CONJUGADAS

Los métodos de direcciones conjugadas buscan acelerar la convergencia del método de descenso acelerado y a su vez reducir la información que se necesita para el método de Newton, esto es evitar la evaluación, suma e inversión de matrices.

Estos métodos se utilizan únicamente para el problema cuadrático siguiente:

$$\text{minimizar } \frac{1}{2} x^T Q x - b^T x$$

donde Q es una matriz de $n \times n$ definida positiva.

ORTOGONALIDAD

Definición.- Dada una matriz simétrica Q , dos vectores d_1 y d_2 se dice que son Q -ortogonales, o conjugados respecto a Q si $d_1^T Q d_2 = 0$.

Un conjunto finito de vectores d_1, d_2, \dots, d_n se dice que es un conjunto Q -ortogonal si $d_i^T Q d_j = 0 \quad \forall i, j$

Proposición II3c.1.- Si Q es positiva definida y el conjunto de vectores no nulos d_1, d_2, \dots, d_n son Q -ortogonales, entonces estos vectores son linealmente independiente. (Ver Apéndice Técnico D para la demostración de esta

proposición).

A continuación se verá el por qué la definición de ortogonalidad es necesaria.

Como se dijo anteriormente, se analizarán funciones del tipo

$$\text{minimizar } \frac{1}{2} x^T Q x - b^T x \quad \dots \text{II3c.1}$$

donde Q es positiva definida.

Es sabido que la solución a este problema es también la solución a (Ejemplo 2 del Capítulo I)

$$Q x = b \quad \dots \text{II3c.2}$$

Sea Q de $n \times n$ definida positiva, y d_1, d_2, \dots, d_n n vectores no nulos Q -ortogonales, ya que estos son mutuamente independientes, la solución x^* de II3c.1 y II3c.2 puede expresarse como una combinación lineal de éstos:

$$x^* = \alpha_0 d_0 + \dots + \alpha_{n-1} d_{n-1} = \sum_{l=0}^{n-1} \alpha_l d_l \quad \dots \text{II3c.3}$$

para algún conjunto de α_l .

premultiplicando por Q :

$$Q x^* = \alpha_0 Q d_0 + \dots + \alpha_{n-1} Q d_{n-1}$$

y por d_l $0 \leq l \leq n-1$

$$d_l^T Q x^* = \alpha_0 d_l^T Q d_0 + \dots + \alpha_l d_l^T Q d_l + \dots + \alpha_{n-1} d_l^T Q d_{n-1}$$

ya que los vectores son Q -ortogonales

$$d_l^T Q x^* = \alpha_l d_l^T Q d_l \Rightarrow \alpha_l = \frac{d_l^T Q x^*}{d_l^T Q d_l} \quad \dots \text{II3c.4}$$

Sustituyendo II3c.4 en II3c.3 y como $Q x^* = b$

$$x^* = \sum_{l=0}^{n-1} \frac{d_l^T b}{d_l^T Q d_l} d_l \quad \dots \text{II3c.5}$$

Teorema II3c.1. (Dirección Conjugada)

Sea $\{d_l\}_{l=0}^{n-1}$ un conjunto de vectores no nulos y además ortogonales a Q . Para cualquier $x_0 \in E^n$ la secuencia $\{x_k\}$ generada de acuerdo a

$$x_{k+1} = x_k + \alpha_k d_k \quad , k \geq 0 \quad \dots \text{II3c.6}$$

donde

$$\alpha_k = - \frac{g_k^T d_k}{d_k^T Q d_k} \quad \dots \text{II3c.7}$$

$$g_k = Qx_k - b$$

converge a la solución única x^* de $Qx = b$ después de n pasos, esto es $x_n = x^*$.

Demostración: Debido a que los vectores d_l son linealmente independientes se puede escribir $x^* - x_0$ como una combinación lineal de éstos

$$x^* - x_0 = \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{n-1} d_{n-1}$$

para algún conjunto α_k . Siguiendo el mismo procedimiento que se utilizó para llegar a II3c.4 se obtiene:

$$\alpha_k = \frac{d_k^T Q(x^* - x_0)}{d_k^T Q d_k} \quad \dots \text{II3c.8}$$

de II3c.6

$$x_1 = x_0 + \alpha_0 d_0$$

$$x_2 = x_1 + \alpha_1 d_1 = x_0 + \alpha_0 d_0 + \alpha_1 d_1$$

$$x_k = x_{k-1} + \alpha_{k-1} d_{k-1} = x_0 + \alpha_0 d_0 + \dots + \alpha_{k-1} d_{k-1}$$

$$x_k - x_0 = \alpha_0 d_0 + \dots + \alpha_{k-1} d_{k-1}$$

multiplicando por $d_k^T Q$

$$d_k^T Q (x_k - x_0) = \alpha_0 d_k^T Q d_0 + \dots + \alpha_{k-1} d_k^T Q d_{k-1}$$

por ortogonalidad

$$d_k^T Q (x_k - x_0) = 0$$

de ahí que $d_k^T Q x_k = d_k^T Q x_0$

sustituyendo esta última ecuación en II3c.8

$$\alpha_k = \frac{d_k^T Q (x^* - x_k)}{d_k^T Q d_k}$$

$$= \frac{d_k^T Q x^* - d_k^T Q x_k}{d_k^T Q d_k}$$

$$= \frac{(b - Q x_k) d_k}{d_k^T Q d_k}$$

$$= \frac{-g_k^T d_k}{d_k^T Q d_k}$$

esta igualdad es idéntica a II3c.7

Ahora véase que x_k minimiza II3c.1.

Se define β_k como el subespacio de E^n generado por $\{d_0, d_1, \dots, d_{k-1}\}$

Teorema II3c.2.- (Expansión de un subespacio)

Sea $\{d_l\}_{l=0}^{n-1}$ una secuencia de vectores no nulos

ortogonales a Q en E^n . Entonces para cualquier

$x_0 \in E^n$ la secuencia $\{x_k\}$ generada de acuerdo a

$$x_{k+1} = x_k + \alpha_k d_k$$

$$\alpha_k = - \frac{g_k^T d_k}{d_k^T Q d_k}$$

x_k minimiza $f(x) = \frac{1}{2} x^T Q x - b^T x$ sobre la línea

$$x = x_{k-1} + \alpha d_{k-1}, -\infty < \alpha < \infty \text{ tanto como } x_0 + \beta_k$$

Demostración: Como $x_0 + \beta_k$ contiene la línea $x = x_{k-1} + \alpha d_{k-1}$ es suficiente demostrar que x_k minimiza a f sobre $x_0 + \beta_k$.

Debido a que f es una función estrictamente convexa, la conclusión se obtendrá al demostrar que g_k^* es ortogonal a β_k (es decir el gradiente de f en x_k es ortogonal al subespacio β_k).

Prueba por inducción.- Se desea demostrar que $g_k \perp \beta_k$.

1) $k = 0$

β_0 está vacía, entonces es cierto para $k = 0$

2) Se supone que es cierto para k , es decir $g_k \perp \beta_k$

$$g_{k+1} = g_k + \alpha_k Q d_k$$

por lo tanto

$$d_k^T g_{k+1} = d_k^T g_k + \alpha_k d_k^T Q d_k = 0$$

por definición de α_k . También para $l < k$

$$d_l^T g_{k+1} = d_l^T g_k + \alpha_k d_l^T Q d_k = 0$$

El primer término del lado derecho de esta ecuación desaparece por la hipótesis de inducción, mientras que el segundo desaparece por la Q -ortogonalidad de d_l 's.

Así $g_{k+1} \perp \beta_{k+1}$

■ q.e.d.

Corolario: En el método de direcciones conjugadas el gradiente

$$g_k \quad k=0,1,\dots,n \text{ satisface}$$
$$g_k^T d_l = 0 \quad \text{para } i < k$$

Algoritmo

- 1) $i = 1$
- 2) Obtener el Hessiano de f
- 3) Si $i > 1$ ir al paso 4 en caso contrario ir al paso 5.
- 4) Calcular d_i donde

$$d_i^T H d_{i-1} = 0$$

$$d_i^T d_i = 1$$

- 5) Calcular $x_{i+1} = x_i + \alpha_i d_i$ donde

$$g_i = Hx_i - b$$

$$\alpha_i = - \frac{g_i^T d_i}{d_i^T H d_i}$$

- 6) Si $i=n$ el procedimiento ha terminado en caso contrario hacer $i = i + 1$, y en seguida ir al paso 4.

3d) METODO DE GRADIENTE CONJUGADO

El método más popular de los Métodos de Direcciones Conjugadas es el Método de Gradiente Conjugado, el cual fue propuesto por Hestenes and Stiefel (1952) como un método para resolver ecuaciones lineales de la forma $b+Bx=0$, siendo esto equivalente a minimizar la función $f(x) = a + b^T x + \frac{1}{2} x^T Bx$.

La versión más común de este método es la que propusieron Fletcher and Reeves (1964) la cual es presentada aquí.

Se puede afirmar que este método no es tan eficiente como el Método de Davidon-Fletcher-Powell pero es más eficiente que los Métodos de Descenso Acelerado y el de Newton.

Hay tres ventajas primarias para el Método de Gradiente Conjugado:

- 1) El gradiente es siempre no-nulo y linealmente independiente de todos los vectores directores previos
- 2) La ventaja más importante del método es la fórmula que utiliza para determinar el vector director nuevo.
- 3) Porque las direcciones son basadas en el gradiente, lo que genera un progreso uniforme hacia la solución en todo paso, esto es en contraste a la situación del Método de Direcciones Conjugadas en el cual la selección de las direcciones se efectúan en forma arbitraria.

Retomando la función anteriormente mencionada:

$$f(x) = a + b^T x + \frac{a}{2} x^T G x$$

La búsqueda será a lo largo de las direcciones que son conjuntamente ortogonales o conjuntamente conjugadas a G.

Supóngase x_1 , como el punto inicial, la primera dirección de búsqueda es hacer en la dirección de descenso acelerado, es decir:

$$d_1 = -g_1 \quad \dots \quad \text{II3d.1}$$

donde $g_1 = \nabla f(x_1)$

obteniendo el valor λ_1 , el cual minimiza:

$$f(x_1 + \lambda d_1)$$

Formulando a x_2 se tiene:

$$x_2 = x_1 + \lambda_1 d_1 \quad \dots \quad \text{II3d.2}$$

y se elige la dirección d_2 de tal forma que sea una combinación lineal de d_1 y $-g_2$ para encontrar x_2 :

$$x_2 = x_1 + \lambda_2 d_2$$

donde λ_2 es el mínimo de $f(x_2 + \lambda_2 d_2)$

La dirección de búsqueda d_2 desde x_2 es conjugada a d_1 y d_2 y g_2 .

En la iteración $(k+1)$ se elige a d_{k+1} como una combinación lineal de $-g_{k+1}, d_1, d_2, \dots, d_k$ donde es conjugado para todo d_1, d_2, \dots, d_k .

$$\therefore d_{k+1} = -g_{k+1} + \sum_{r=1}^k \alpha_r d_r \quad , k=2, \dots$$

Donde todos los α_r son cero excepto para α_k por lo que:

$$d_{k+1} = -g_{k+1} + \alpha_k d_k \quad \dots \quad \text{II3d.3}$$

y

$$\alpha_k = \frac{q_{k+1}^2}{q_k^2}$$

... II3d.4

se demostrará la ecuación II3d.4 por inducción matemática:

1) Para $k=1$

$$\nabla f(x) = b + Gx \quad \therefore g_1 = b + Gx_1 \quad \text{y} \quad g_2 = b + Gx_2$$

Si d_1 y d_2 son conjugadas $\therefore d_2 = -g_2 + \alpha_1 d_1$

$$\text{además } d_2^T G d_1 = 0$$

Sustituyendo el valor de d_2 en la anterior igualdad

$$-g_2^T G d_1 + \alpha_1 d_1^T G d_1 = 0$$

Por otro lado se tiene que,

$x_2 = x_1 + \lambda_1 d_1$ donde λ_1 es el mínimo de $f(x_1 + \lambda d_1)$

$$\therefore d_1 = \frac{x_2 - x_1}{\lambda_1}$$

Sustituyendo el anterior resultado y aplicando la ecuación II3d.1:

$$\frac{(-g_2^T - \alpha_1 g_1^T) G (x_2 - x_1)}{\lambda_1} = 0$$

$$\text{pero } Gx_2 - Gx_1 = (b + Gx_2) - (b + Gx_1)$$

$$\therefore (-g_2^T - \alpha_1 g_1^T) (g_2 - g_1) = 0$$

$$\Rightarrow -g_2^2 + \alpha_1 g_1^2 = 0$$

Despejando α_1

$$\alpha_1 = \frac{g_2^2}{g_1^2}$$

q.e.d. para $k=1$

2) Se supone verdadero para k y por demostrar para $k+1$.

Por hipótesis d_1, d_2, \dots, d_k son mutuamente conjugados.

Y sea $x_{k+1} = x_k + \lambda_k d_k$.

Ahora minimizando $f(x_k + \lambda_k d_k)$ a lo largo de la línea

$$x_k + \lambda d_k.$$

$$\text{Así } g_{k+1}^T d_k = 0 \quad \dots \text{II3d.5}$$

$$\begin{aligned} \text{Se tiene: } x_{k+1} &= x_k + \lambda_k d_k \\ &= x_{k-1} + \lambda_{k-1} d_{k-1} + \lambda_k d_k \end{aligned}$$

$$\vdots \quad \vdots$$

$$\therefore x_{k+1} = x_j + \sum_{l=j}^k \lambda_l d_l \quad \text{donde } j=1, \dots, k \quad \dots \text{II3d.6}$$

Pre-multiplicando por G :

$$Gx_{k+1} = Gx_j + \sum_{l=j}^k \lambda_l G d_l$$

sumando el vector b en ambos miembros de la igualdad:

$$b + Gx_{k+1} = b + Gx_j + \sum_{l=j}^k \lambda_l G d_l$$

$$g_{k+1} = g_j + \sum_{l=j}^k \lambda_l G d_l$$

$$g_{k+1}^T = g_j^T + \sum_{l=j}^k \lambda_l d_l^T G$$

post-multiplicando por d_m la igualdad

$$g_{k+1}^T d_m = g_j^T d_m + \sum_{l=j}^k \lambda_l d_l^T G d_m \quad \text{donde } m=j, m=1, \dots, k$$

Ahora $g_j^T d_m = 0$ y d_1, \dots, d_k son mutuamente conjugados, por hipótesis por lo que $d_l^T G d_m = 0$ para $l \neq m$

$$\therefore g_{k+1}^T d_m = 0, \quad m=1, 2, \dots, k \quad \dots \text{II3d.7}$$

aplicando la ecuación II3d.5 se tiene:

$$g_{k+1}^T d_m = 0, \quad m=1, 2, \dots, k \quad \dots \text{II3d.8}$$

Se ha demostrado que g_{k+1} es ortogonal a cada una de las direcciones d_1, d_2, \dots, d_k .

También se demostrará que g_{k+1} es ortogonal a cada g_1, g_2, \dots, g_k .

Por la hipótesis de inducción:

$$d_j = -g_j + \alpha_{j-1} d_{j-1}$$

Premultiplicando g_{k+1}^T la anterior igualdad:

$$g_{k+1}^T d_j = -g_{k+1}^T g_j + \alpha_{j-1} g_{k+1}^T d_{j-1}$$

Utilizando la ecuación II3d.8 haciendo $j=m$:

$$\therefore -g_{k+1}^T g_j = 0$$

Por lo que $g_{k+1}^T g_j = 0$ para $j=1, 2, \dots, k$...II3d.9

Para completar la demostración de la inducción, es necesario demostrar que d_{k+1} definido por II3d.3 es conjugado a

$$d_1, d_2, \dots, d_k$$

$$d_{k+1}^T = -g_{k+1}^T + \alpha_k d_k^T$$

Postmultiplicando $G d_j$

$$d_{k+1}^T G d_j = -g_{k+1}^T G d_j + \alpha_k d_k^T G d_j$$

Como d_1, \dots, d_{k-1} son conjugados por hipótesis:

$$\therefore d_{k+1}^T G d_j = -g_{k+1}^T G d_j \quad \text{para } j=1, \dots, k-1$$

Ahora

$$x_{j+1} = x_j + \lambda_j d_j$$

$$-d_{k+1}^T G d_j = -g_{k+1}^T G \frac{(x_{j+1} - x_j)}{\lambda_j}$$

$$= -g_{k+1}^T \frac{(g_{j+1} - g_j)}{\lambda_j} = 0$$

por la ecuación II3d.9.

Así $d_{k+1}^T G d_j = 0$ para $j=1, 2, \dots, k-1$ para cualquier valor de α_k .

Por último se determinará α_k tal que

$$d_{k+1}^T G d_k = 0$$

$$d_{k+1}^T G d_k = -g_{k+1}^T G d_k + \alpha_k d_k^T G d_k$$

$$= -g_{k+1}^T \frac{(g_{k+1} - g_k)}{\lambda_k} + \alpha_k (-g_k^T + \alpha_{k-1} d_{k-1}^T) \frac{(g_{k+1} - g_k)}{\lambda_k}$$

Pero $g_{k+1}^T d_j = 0$ (II3d.7) y $g_{k+1}^T d_j = 0$ (II3d.8) para $j = s, s+1, \dots, k$

$$\therefore d_{k+1}^T g_k = \frac{-g_{k+1}^2 + \alpha_k g_k^2}{\lambda_k}$$

Así para que d_{k+1} sea conjugado a d_k , $\alpha_k = \frac{g_{k+1}^2}{g_k^2}$

q.e.d.

Para simplificar el algoritmo iterativo del anterior Método se presentan los pasos que lo componen:

- 1) $t=1$, proporcionar x_t
- 2) $d_t = -g_t$
- 3) Encontrar λ_t tal que minimice $f(x_t + \lambda_t d_t)$
- 4) Si x_{t+1} es el mínimo el proceso ha terminado, si no es así ir a 5.
- 5) Si $t=n$ ir a 9 si no ir a 6.
- 6) $d_{t+1} = -g_{t+1} + \frac{g_{t+1}^2}{g_t^2} d_t$
- 7) $t=t+1$
- 8) ir a 3
- 9) $t=1$
- 10) ir a 2.

3e) EL METODO DE DAVIDON-FLETCHER-POWELL

El método de Davidon-Fletcher-Powell (D.F.P.) está basado en las ecuaciones

$$x_m = x_0 - G^{-1}(x_0)g(x_0)$$
$$x_{i+1} = x_i - \lambda_i G^{-1}(x_i)g(x_i)$$

aunque esto evita el cálculo de la inversa de la Hessiana $G^{-1}(x_i)$ en cada paso al establecer la dirección de búsqueda en la iteración i como $-H_i g(x_i)$, donde H_i es una matriz simétrica definida positiva que es actualizada en cada iteración en una forma que será explicada posteriormente. Finalmente H viene siendo la inversa de la Hessiana.

Se inicia con un punto inicial x_0 y una matriz inicial H_0 , generalmente la matriz unitaria, aunque cualquier matriz simétrica positiva puede utilizarse. El procedimiento iterativo se justificará después de lo siguiente [Nota: Es conveniente escribir $g(x_i)$ como g_i].

Algoritmo.

- 1) En la iteración i se tiene un punto x_i y una matriz definida positiva H_i .
- 2) Establecer la dirección de búsqueda
 $d_i = -H_i g_i$...II3e.1
- 3) Llevar a cabo una búsqueda lineal sobre la línea $x_i + \lambda d_i$ para encontrar el valor λ_i que minimiza $f(x_i + \lambda_i d_i)$.
- 4) Calcular $v_i = \lambda_i d_i$...II3e.2

- 5) Calcular $x_{l+1} = x_l + v_l$...II3e.3
 6) Encontrar $f(x_{l+1})$ y g_{l+1} . Terminar el procedimiento si $|g_{l+1}|$ o $|v_l|$ son suficientemente pequeños. En caso contrario

$$g_{l+1}^T v_l = 0 \quad \dots \text{II3e.4}$$

- 7) Calcular $u_l = g_{l+1} - g_l$...II3e.5

- 8) Actualizar la matriz mediante $H_{l+1} = H_l + A_l + B_l$...II3e.6

donde

$$A_l = v_l v_l^T / (v_l^T u_l) \quad \dots \text{II3e.7}$$

$$B_l = -H_l u_l u_l^T H_l / (u_l^T H_l u_l) \quad \dots \text{II3e.8}$$

- 9) Incrementar l ($l+1$) y regresar al paso 2.

Justificación del procedimiento de Fletcher y Powell.

- a) El proceso es estable, v_l está en declive y λ_l es positiva. Ya que g_l es la dirección de ascenso acelerado, v_l irá en descenso si y sólo si

$$-v_l^T g_l = -g_l^T v_l \quad \dots \text{II3e.9}$$

es positiva.

Esto será si H_l es simétrica definida positiva para toda l . H_0 tiene estas propiedades porque son supuestas. La actualización en las ecuaciones II3e.6, II3e.7 y II3e.8 mantiene la simetría.

Se prueba que H_l permanece positiva definida por inducción. Así se asume que H_l es simétrica positiva

definida. Entonces tiene una raíz cuadrada C_l tal que

$$C_l^T C_l = C_l C_l^T = H_l \quad \dots \text{II3e.10}$$

Hacer

$$p = C_l w \quad \text{y} \quad q = C_l u_l \quad \dots \text{II3e.11}$$

donde w es cualquier vector.

Entonces

$$\begin{aligned} w^T H_{l+1} w &= w^T H_l w + \frac{w^T v_l v_l^T w}{v_l^T u_l} - \frac{w^T H_l u_l u_l^T H_l w}{u_l^T H_l u_l} \\ &= p^T w - \frac{(p^T q)^2}{q^2} + \frac{(w^T v_l)^2}{v_l^T u_l} \\ &= \frac{p^T q^2 - (p^T q)^2}{q^2} + \frac{(w^T v_l)^2}{v_l^T u_l} \\ &\geq \frac{(w^T v_l)^2}{v_l^T u_l} \quad \dots \text{II3e.12} \end{aligned}$$

porque $p^T q^2 \geq (p^T q)^2$ por la desigualdad de Schwarz.

Ahora el denominador de la ecuación II3e.12 es positivo porque

$$\begin{aligned} v_l^T u_l &= v_l^T [g_{l+1} - g_l] \\ &= -v_l^T g_l \text{ ya que } v_l^T g_{l+1} = 0 \text{ de la ecuación II3e.4} \\ &= \lambda_l g_l^T H_l g_l \\ &> 0 \text{ porque } \lambda_l > 0 \text{ y } H_l \text{ es positiva definida} \quad \dots \text{II3e.13} \end{aligned}$$

Así $\det H_{l+1} > 0$ lo que prueba que H_{l+1} es positiva definida.

Así la prueba de inducción es completa.

b) Se Mostrará que si el método de D.F.P. se aplica a la función cuadrática

$$F(x) = a + x^T b + \frac{1}{2} x^T G x$$

con G definida simétrica positiva, entonces $H_n = G^{-1}$ y el proceso será terminado después de n iteraciones. Se hará esto al demostrar que v_0, v_1, \dots, v_k son eigenvectores linealmente independientes de $H_{k+1} G$ con eigenvalor 1. Así H_n deberá ser matriz unitaria.

Nótese de la ecuación II3e.5 que

$$\begin{aligned} u_l &= g_{l+1} - g_l \\ &= Gx_{l+1} - Gx_l \\ &= Gv_l \end{aligned} \quad \dots \text{II3e.14}$$

Además

$$H_{l+1} G v_l = H_{l+1} u_l \quad (\text{de la ecuación II3e.14})$$

$$= H_l u_l + A_l u_l + B_l u_l$$

$$= H_l u_l + v_l - H_l u_l$$

usando las ecuaciones II3e.7, II3e.8 y II3e.9 y notando que $v_l^T u_l$ y $u_l^T H_l u_l$ son escalares y que pueden ser cancelados.

$$\text{Así} \quad H_{l+1} G v_l = v_l \quad \dots \text{II3e.15}$$

Se demostrará que para $k = 2, 3, \dots, n$

$$v_l^T G v_j = 0 \quad 0 \leq l < j < k \quad \dots \text{II3e.16}$$

$$H_k G v_l = v_l \quad 0 \leq l < k \quad \dots \text{II3e.17}$$

Se hará por inducción sobre k . Si se hace $l=0$ en la ecuación II3e.15, se obtiene $H_1 G v_0 = v_0$ que es la ecuación

II3e.17 cuando $k=1$. Las ecuaciones II3e.17 con $k=2$ son

$$H_2 G v_0 = v_0 \quad \text{y} \quad H_2 G v_1 = v_1$$

La segunda de ellas viene de la ecuación II3e.15 con $l=1$.

Para el primero

$$H_2 G v_0 = H_1 G v_0 + v_1 \frac{v_1^T G v_0}{v_1^T u_1} - \frac{H_1 u_1 u_1^T H_1 G v_0}{u_1^T H_1 u_1}$$

Los últimos dos términos de la derecha son ambos ceros ya que

$$\begin{aligned} v_1^T G v_0 &= v_0^T G v_1 = v_0^T G (-\lambda_1 H_1 g_1) \\ &= -\lambda_1 g_1^T H_1 G v_0 \\ &= -\lambda_1 g_1^T v_0 \quad (\text{de la ecuación II3e.15 con } l=0) \end{aligned}$$

También $u_1 = G v_1$, tal que $u_1^T H_1 G v_0 = v_1^T G v_0$. Así las ecuaciones II3e.17 son verdaderas para $k=2$ y el resultado justo establecido, cuando la transpuesta de

$$v_0^T G v_1 = 0$$

que es la ecuación II3e.16 cuando $k=2$.

Ahora se procederá con la inducción y se demostrará que si las ecuaciones II3e.18 y II3e.17 son verdaderas para los valores de k son también verdaderas para valores $k+1$.

Se tiene

$$\begin{aligned} g_k &= b + G x_k \\ &= b + G(x_{k-1} + v_{k-1}) \end{aligned}$$

$$\begin{aligned}
 &= b + G(x_{k-2} + v_{k-2} + v_{k-1}) \text{ etc.} \\
 &= g_{l+1} + G(v_{l+1} + v_{l+2} + \dots + v_{k-1}) \quad \text{II3e.18}
 \end{aligned}$$

Así la ecuación II3e.16, cuando $l < k-1$

$$v_l^T g_k = v_l^T g_{l+1} = 0 \text{ mediante la ecuación II3e.4}$$

y de la ecuación II3e.4 directamente

$$v_{k-1}^T g_k = 0$$

Así

$$v_l^T g_k = 0 \quad 0 \leq l < k \quad \dots \text{II3e.19}$$

tal que de la ecuación II3e.17

$$-v_l^T B H_k g_k = 0$$

tal que

$$-v_l^T B v_k = 0 \quad \text{ya que } v_k = \lambda_k d_k$$

Así se ha establecido que

$$v_l^T B v_k = 0 \quad \text{para } 0 \leq l < k \quad \dots \text{II3e.20}$$

$$\text{i.e. } v_l^T B v_j = 0 \quad \text{para } 0 \leq l < j < k \dots \text{II3e.21}$$

Se tiene de las ecuaciones II3e.14 y II3e.17

$$\begin{aligned}
 (u_k)^T H_k B v_l &= u_k^T v_l \\
 &= v_k^T B v_l = 0 \quad \text{para } 0 \leq l < k \dots \text{II3e.22}
 \end{aligned}$$

Por lo tanto de las ecuaciones II3e.6, II3e.7, II3e.8 y II3e.17

$$H_{k+1} B v_l = H_k B v_l + \frac{v_k v_k^T B v_l}{v_k^T u_k} - \frac{H_k u_k u_k^T H_k B v_l}{u_k^T H_k u_k}$$

$$= H_k G v_l, \quad v_k^T G v_l = 0$$

y

$$\begin{aligned} u_k^T H_k G v_l &= u_k^T v_l \\ &= v_k^T G v_l = 0 \quad \text{para } 0 \leq l < k \end{aligned}$$

Así se ha demostrado que

$$H_{k+1} G v_i = H_k G v_i = v_i \quad \text{para } 0 \leq i < k \quad \text{II3e.23}$$

También para $l=k$

$$\begin{aligned} H_{k+1} G v_k &= H_k G v_k + \frac{v_k v_k^T G v_k}{v_k^T u_k} - \frac{H_k u_k u_k^T H_k G v_k}{u_k^T H_k u_k} \\ &= H_k u_k + v_k - H_k u_k \end{aligned}$$

ya que

$$v_k^T G v_k^T = v_k^T u_k \quad \text{y} \quad u_k^T H_k G v_k = u_k^T H_k u_k$$

Así $H_{k+1} G v_k = v_k^T u_k$ y ésta combinada con la ecuación II3e.23 da

$$H_{k+1} G v_l = v_l \quad 0 \leq l < k+1 \quad \dots \text{II3e.24}$$

Las ecuaciones II3e.21 y II3e.24 extienden las ecuaciones II3e.16 y II3e.17 para el siguiente valor de k y completan la prueba de inducción para el último.

La ecuación II3e.16 demuestra que v_0, v_1, \dots, v_{n-1} son linealmente independientes y mutuamente conjugadas con respecto a G . De las ecuaciones II3e.17 v_0, v_1, \dots, v_{n-1} son eigenvectores de $H_n G$ con eigenvalor 1. Así $H_n G$ debe ser una matriz unitaria por lo tanto:

$$H_n = G^{-1} \quad \dots \text{II3e.25}$$

El hecho de que el mínimo es encontrado en n iteraciones

se concluye de la ecuación II3e.19 g_n debe ser ortogonal a cada uno de los n vectores independientes v_0, v_1, \dots, v_{n-1} . Así

$$g_n = 0 \quad \dots \text{II3e.26}$$

c) La actualización de la matriz H se concluye de II3e.6

$$H_n = H_0 + \sum_{l=0}^{n-1} A_l + \sum_{l=0}^{n-1} B_l$$

3f) METODO DE HOOKE AND JEEVES

Este método es un método de búsqueda directo, esto es, únicamente utiliza los valores de la función para encontrar el óptimo de la función. Aunque este método no utiliza derivadas, se introdujo dentro de este tema por ser un algoritmo que se utiliza para funciones de varias variables.

La idea general es evaluar la función en un punto inicial, analizar el entorno de este punto para saber como se comporta la función, hasta que se encuentra un punto tal que el valor de la función es menor, se continua iterando hasta encontrar el mínimo.

A continuación se describe el algoritmo.

1. Se escoge un punto inicial x_1 , y longitudes h_j para cada x_j , $j=1, \dots, n$.
2. Se procede a explorar alrededor de x_1 .
 - a) Evaluar $f(x_1)$
 - b) Evaluar $f(x_1 + h_1 e_1)$; donde e_1 es un vector unitario en dirección del eje x_1 . Si la función se reduce, reemplazar x_1 por $x_1 + h_1 e_1$. Si no se decrementa, evaluar $f(x_1 - h_1 e_1)$ y reemplazar x_1 por $x_1 - h_1 e_1$. Si ninguno de los dos proporciona un valor menor de la función, muévase a través del eje x_2 , es decir calcule $f(x_1 - h_2 e_2)$. Cuando se hayan considerado las n variables se tendrá el nuevo punto x_2 .
 - c) Si $x_1 = x_2$, se comienza una nueva exploración, esta vez reduciendo el tamaño h (se aconseja reducirlo una décima

parte de su valor anterior).

- d) Si $x_2 \neq x_1$ se ha encontrado un modelo a seguir en cuanto a la dirección que se debe utilizar.

- 3) a) Evaluar la función en el siguiente punto

$$P_1 = x_1 + 2(x_2 - x_1)$$

En general

$$P_i = x_i + 2(x_{i+1} - x_i)$$

- b) Realizar una exploración alrededor de $P_1 (P_i)$

- c) Si el valor más bajo obtenido en 3a) es menor que el valor en el punto $x_2 (x_{i+1})$ entonces se ha encontrado un nuevo punto. En este caso repetir 3a), de otra forma continuar la exploración en x_2 sin utilizar 3a).

- 4) El proceso termina cuando la longitud de los pasos ha sido reducida a un predeterminado valor pequeño.

CAPITULO III

CAPITULO III. APLICACIONES DE OPTIMIZACION SIN RESTRICCIONES

En esta sección se desarrollan los programas de computación de cada un de los algoritmos cuya base teórica fue explicada en el capítulo anterior, por lo que se irán presentando con la misma secuencia de dicho capítulo.

Estos programas están desarrollados en BASIC. Son varias las razones por las cuales fue utilizado este lenguaje, las principales causas son dos, una de ellas es que el BASIC es un lenguaje sencillo y de fácil comprensión para las personas relacionadas con la computación y la otra es que el BASIC forma parte del sistema operativo MS-DOS que manejan las computadoras personales IBM y compatibles (P.C.) y como es sabido el uso de estos equipos se ha ido generalizando, de esta forma será más sencillo para los lectores aplicar estos programas.

III.1 DESARROLLO DEL PROGRAMA COMPUTACIONAL PARA EL TEMA 1 DEL
CAPITULO II.

1a) FIBONACCI

El programa presentado minimiza a la función $x^4 - 14x^3 + 60x^2 - 70x$. la cual está desarrollada en la línea 2000.

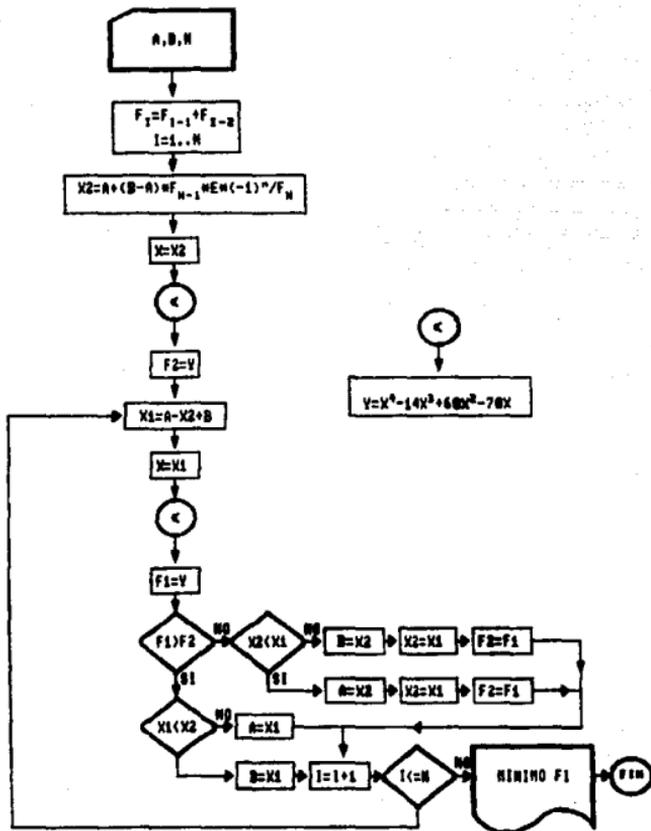
Variables de entrada: A, B, E, N.

A Extremo izquierdo del intervalo
B Extremo derecho del intervalo
E grado de precisión
N Número de iteraciones

Variabes internas: F_0, F_1, F_2
 F_0, F_1 Primeros números de la serie de Fibonacci.
 F_2, F_3 Números de Fibonacci $2^a, \dots, n$

Variables de salida: X1, X2 puntos interiores
F1 valor de la función en el punto óptimo

DIAGRAMA DE FLUJO (FIBONACCI)



```

10 REN PROGRAMA: FIBO.BAS
15 REN PROCESO: ALGORITMO DEL METODO DE FIBONACCI.
20 REN
25 REN ESTE ALGORITMO MINIMIZA LA FUNCION  $X^4 - 14(X)^3 + 60(X)^2 - 70X$ 
26 CLR
29 DEFDBL A-B, F, X-Y
30 DIM F(50)
35 PRINT "CUANTAS EVALUACIONES DEBEAS REALIZAR : "; INPUT N
40 PRINT "PROPORCIONAR EL VALOR DE EPSILON": INPUT E
45 PRINT "PROPORCIONAR LOS EXTREMOS DEL INTERVALO DE INCERTIDUMBRE (A,B): "; INPUT A, B
70 PRINT "      ALGORITMO DEL METODO DE FIBONACCI": PRINT
71 PRINT "Funcion a minimizar  $F(X) = X^4 - 14X^3 + 60X^2 - 70X$ "
72 PRINT "en el intervalo ("; A; ", "; B; ") con "; B; " iteraciones y E="; E; ",."
77 F(0) = 1; F(1) = 1
78 FOR I = 2 TO N
79 F(I) = F(I - 1) + F(I - 2)
80 NEXT I
83 X2 = A + ((B - A) * F(N - 1) + E * (-1) ^ N) / F(N)
85 X = X2: GOSUB 2000: F2 = Y
90 I = 0
100 X1 = A - X2 + B
105 X = X1: GOSUB 2000: F1 = Y
106 PRINT : PRINT
107 PRINT "Iteracion "; I; " : "; PRINT
108 PRINT "A="; A; " B="; B; PRINT "X1="; X1; " X2="; X2: PRINT "F(X1)="; F1; " F(X2)="; F2
110 IF F1 > F2 THEN GOTO 140
115 IF X2 < X1 THEN GOTO 130
120 B = X2: X2 = X1: F2 = F1
125 GOTO 160
130 A = X2: X2 = X1: F2 = F1
135 GOTO 160
140 IF X2 < X1 THEN GOTO 155
145 A = X1
150 GOTO 160
155 B = X1
160 I = I + 1
165 IF I <= N THEN GOTO 100
167 PRINT : PRINT : PRINT
200 PRINT "Valor minimo de F(X): ", F1
220 END
2000 Y = X * X * X * X - 14 * X * X * X + 60 * X * X - 70 * X
2010 RETURN
9999 SAVE "FIBO", A

```

1b) SECCION DORADA

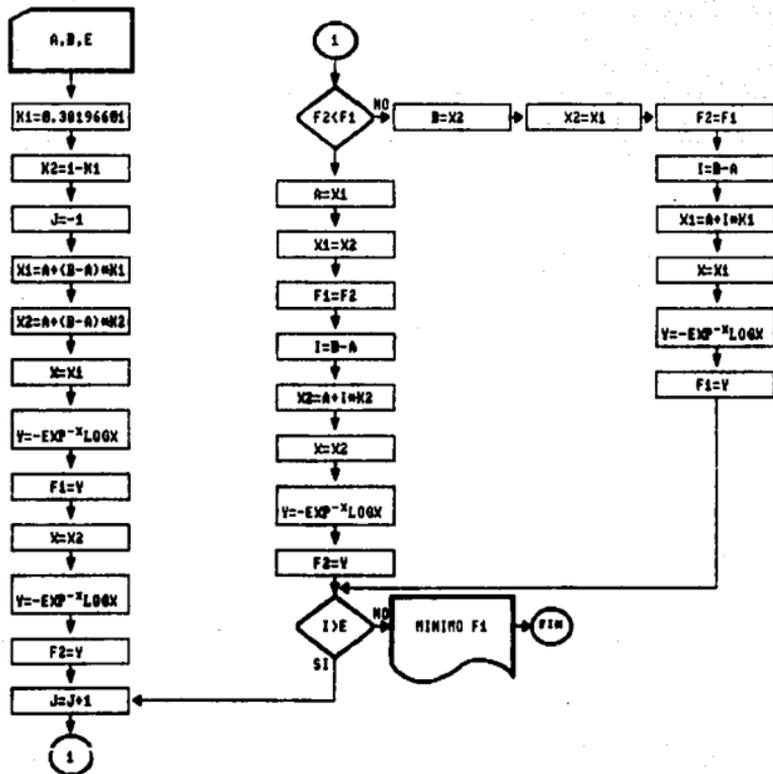
Función a optimizar: $Y = -e^{-X} \log x$ (línea 2000)

Variables de entrada: A extremo izquierdo del intervalo
B extremo derecho del intervalo
E grado de precisión

Variables internas: K1: constante de reducción

Variables de salida: X_1, X_2 puntos intermedios
F1 valor de la función en el punto óptimo

DIAGRAMA DE FLUJO (SECCION ROMANA)



```

10 REN PROGRAMA : DORA.BAS
15 REN PROCESO : ALGORITMO DEL METODO SECCION BORADA
20 REN
25 REN Este algoritmo minimiza la funcion F(X)= (-e-X)* (ln(X))
26 CLS
35 DEFDBL A-B, F, K, X
40 PRINT "ESPECIFICAR LA PRECISION DESEADA : "; INPUT E
45 PRINT "PROPORCIONAR LOS EXTREMOS DEL INTERVALO DE INCERTIDUMBRE (A,B): "; INPUT A, B
50 PRINT " ALGORITMO DEL METODO SECCION BORADA"; PRINT
52 PRINT "Funcion a minimizar F(X) = (-e-X)* (ln(X)) en el =
53 PRINT " intervalo ("; A; ", "; B; ") con precision de "; E; "."
55 K1 = .38196601138; K2 = 1 - K1; J = -1
60 X1 = A + (B - A) * K1; X2 = A + (B - A) * K2
62 X = X1: GOSUB 2000: F1 = Y
63 X = X2: GOSUB 2000: F2 = Y
65 PRINT : PRINT : J = J + 1
70 PRINT "Iteracion "; J; " : "; PRINT
75 PRINT "A="; A; " B="; B; PRINT "X1="; X1; " X2="; X2; PRINT "F(X1)="; F1; " F(X2)="; F2
80 IF F2 < F1 THEN 95
85 B = X2; X2 = X1; F2 = F1; I = B - A; X1 = A + I * K1
90 X = X1: GOSUB 2000: F1 = Y: GOTO 105
95 A = X1; X1 = X2; F1 = F2; I = B - A; X2 = A + I * K2
100 X = X2: GOSUB 2000: F2 = Y
105 IF I > E THEN 65
110 PRINT : PRINT : PRINT
115 PRINT " Valor minimo de F(X): ", F1
120 END
2000 Y = -EXP(-X) * LOG(X)
2010 RETURN
9999 SAVE "DORA", A

```

III.2 DESARROLLO DEL PROGRAMA COMPUTACIONAL PARA EL TEMA 2 DEL
CAPITULO II

2a) INTERPOLACION CUADRATICA.

Función a optimizar: $f(x)=2x^2-e^x$ (línea 800)

Variables de entradas: A,H,E

A: punto inicial

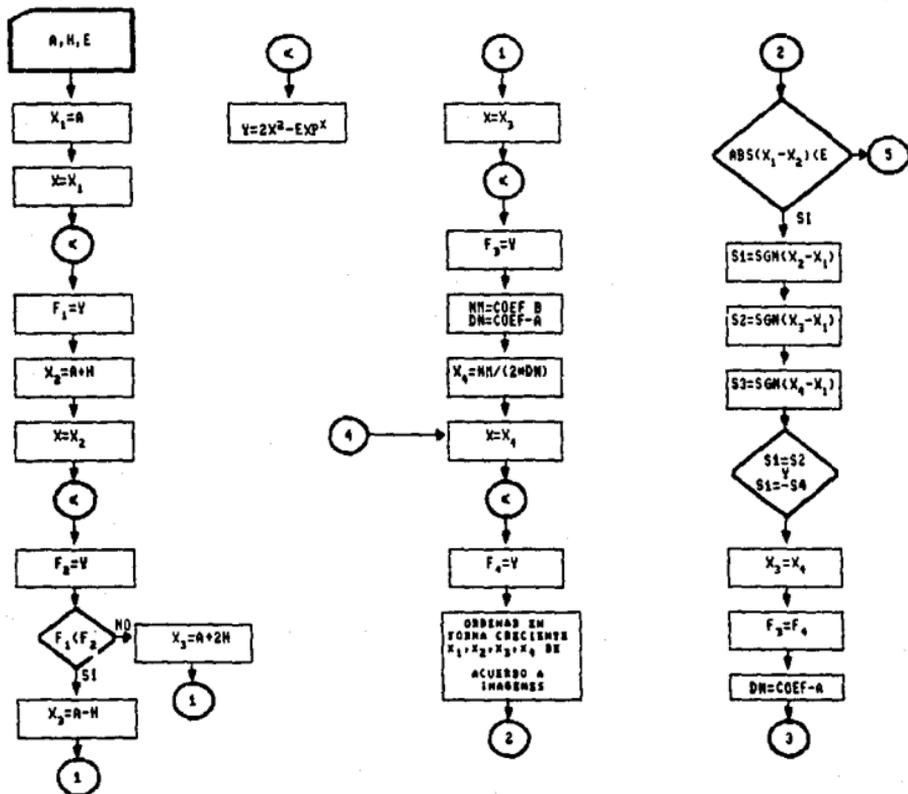
H: distancia estimada entre el punto inicial y el punto óptimo

E: grado de precisión

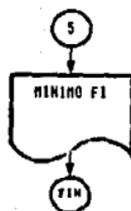
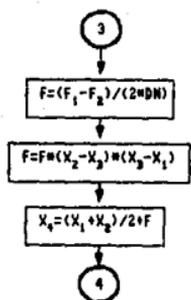
Variable internas: NM: coeficiente B del polinomio cuadrado
DN: Coeficiente -A del polinomio cuadrado
S1: Función signo de x_2-x_1
S2: Función signo de x_3-x_1
S3: Función signo de x_4-x_1

Variables de salida: X_1, X_2, X_3, X_4 : puntos a evaluar
 F_1 : valor de la función en el punto óptimo

DIAGRAMA DE FLUJO (INTERPOLACION CUBICA)



... (INTERPOLACION CUADRÁTICA)



```

10 REM PROGRAMA:INTCLM.BAS
15 REM PROCESO:ALGORITMO DE INTERPOLACION CUADRATICA
20 REM
25 REM ESTE ALGORITMO MINIMIZA LA FUNCION 2X**2-6**X
30 CLS
37 DEFDBL F, X
40 PRINT "PROPORCIONAR EL PUNTO INICIAL:": INPUT A
45 PRINT "PROPORCIONAR EL VALOR DE H:": INPUT H
50 PRINT "PROPORCIONAR LA APROXIMACION:": INPUT E
51 CLS
55 X(1) = A: X = X(1): GOBUB BOO: F(1) = Y
56 X(2) = A + H: X = X(2): GOBUB BOO: F(2) = Y
57 IF F(1) < F(2) THEN X(3) = A - H: X = X(3): GOBUB BOO: F(3) = Y: GOTO 60
58 X(3) = A + 2 * H: X = X(3): GOBUB BOO: F(3) = Y
60 PRINT "ALGORITMO DEL METODO DE INTERPOLACION CUADRATICA": PRINT
61 PRINT "FUNCION A MINIMIZAR F(X)=2X**2-6**X CON UNA APROXIMACION"
62 PRINT "DE": E; ", CON PUNTO INICIAL DE": A; "Y VALOR DE H": H; ".": PRINT
63 PRINT "      X(1)      F(1)"
65 DH = (X(2) - X(3)) * F(1)
70 DH = DH + (X(3) - X(1)) * F(2) + (X(1) - X(2)) * F(3)
75 HH = (X(2) * X(2) - X(3) * X(3)) * F(1)
80 HH = HH + (X(3) * X(3) - X(1) * X(1)) * F(2)
85 HH = HH + (X(1) * X(1) - X(2) * X(2)) * F(3)
90 X(4) = HH / (2 * DH): X = X(4): GOBUB BOO: F(4) = Y
100 FOR J = 1 TO 3
105 FOR K = J + 1 TO 4
110 IF F(J) <= F(K) THEN GOTO 125
115 X = X(J): X(J) = X(K): X(K) = X
120 F = F(J): F(J) = F(K): F(K) = F
125 NEXT K: NEXT J
130 FOR I = 1 TO 4: PRINT X(I), F(I): NEXT I
135 PRINT : PRINT
140 IF ABS(X(1) - X(2)) < E THEN GOTO 250
145 S1 = SIGN(X(2) - X(1)): S2 = SIGN(X(3) - X(1)): S3 = SIGN(X(4) - X(1))
150 IF S1 = S2 AND S1 = -S4 THEN X(3) = X(4): F(3) = F(4)
160 DH = (X(2) - X(3)) * F(1) + (X(3) - X(1)) * F(2) + (X(1) - X(2)) * F(3)
170 F = (F(1) - F(2)) / (2 * DH)
180 F = F + (X(2) - X(3)) * (X(3) - X(1))
190 X(4) = (X(1) + X(2)) / 2 + F
200 X = X(4): GOBUB BOO: F(4) = Y
210 GOTO 100
250 PRINT "SOLUCION FINAL"
260 PRINT "X=": X(1), "Y=": F(1)
300 END
800 Y = 2 * X * X - EXP(X)
810 RETURN
9999 SAVE "INTCLM", A

```

2b) INTERPOLACION CUBICA

Función a minimizar: $z=100(x_2-x_1^2)^2+(1-x_1)^2$ (línea 5000)

VARIABLES DE ENTRADA: N: número de variables

X_1 : vector punto inicial

D_x : vector dirección factible

E: precisión

FM: mínimo supuesto

VARIABLES INTERMEDIAS: P_x : punto inicial

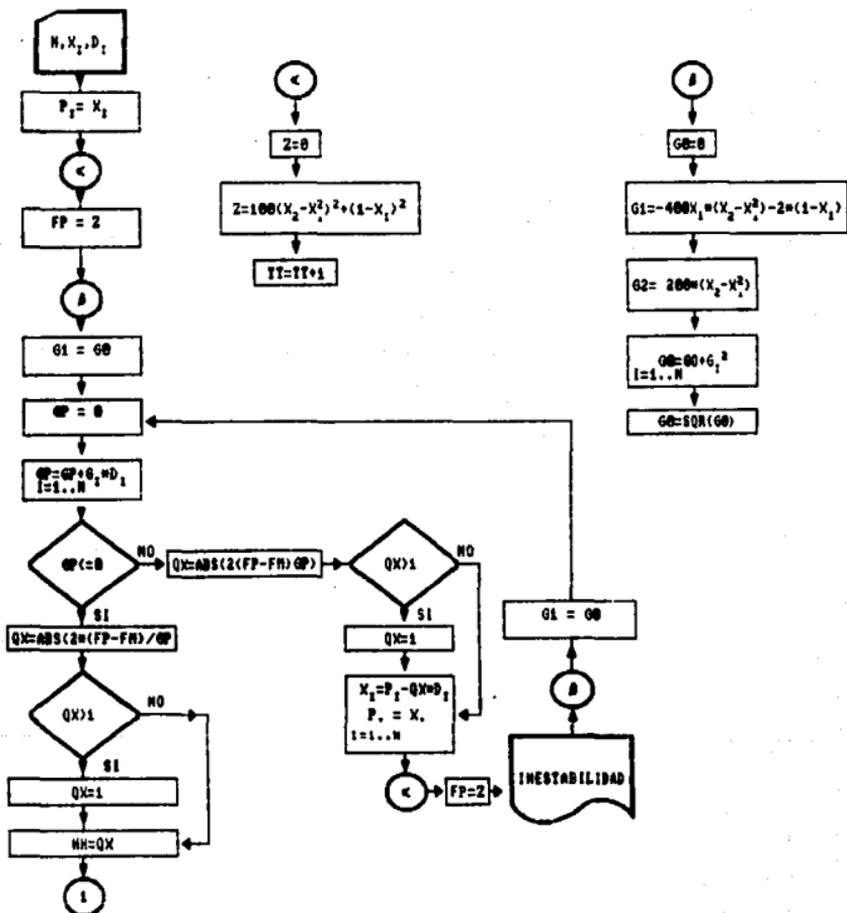
G1, G2: gradientes de la función

Q: Matriz semidefinida positiva.

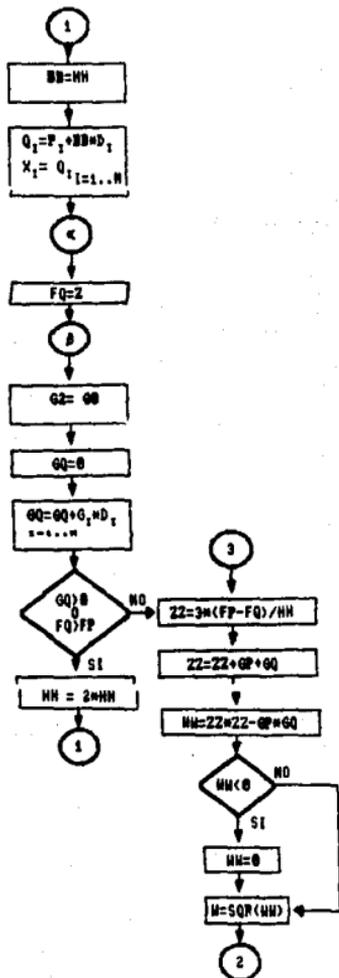
VARIABLES DE SALIDA: CC: número de iteraciones

Z: valor de la función en el punto óptimo

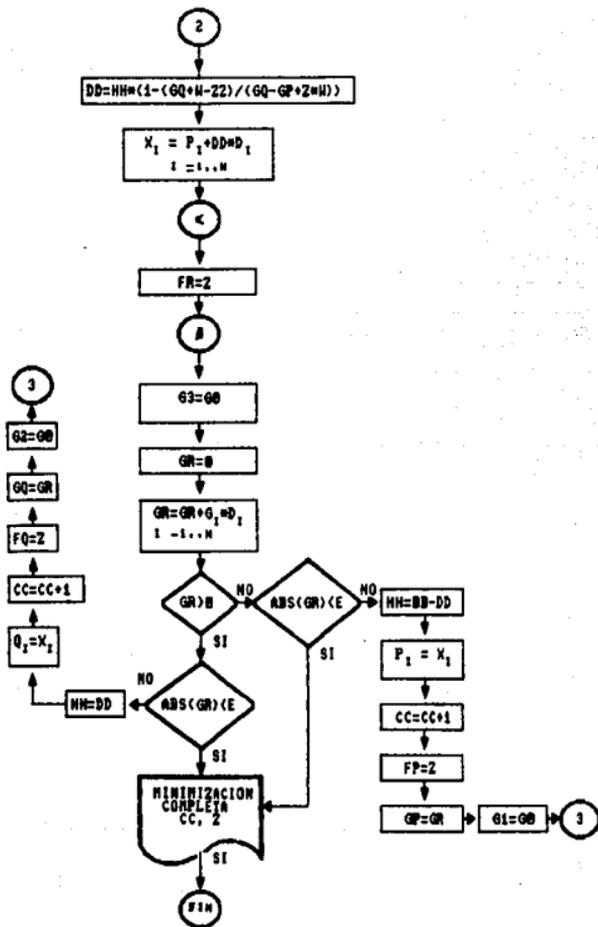
DIAGRAMA DE FLUJO (INTERPOLACION CUBICA)



...(INTERPOLACION CUBICA)



...(INTERPOLACION CUBICA)



```

10 REM PROGRAMA: INTCLUB.BAS
20 REM PROCESO: ESTE PROGRAMA REALIZA EL ALGORITMO DE
30 REM INTERPOLACION CUBICA. ENCUENTRA EL MINIMO DE
40 REM  $F(X + \text{LAM} * D)$  SOBRE LA LINEA  $X + \text{LAM} * D$ .
50 REM  $Z = F(X1, X2, \dots, XN)$  SE CALCULA EN SUB 5000 Y SU GRADIENTE
60 REM  $G(1), G(2), \dots, G(N)$  EN SUB 6000
65 DEFDBL D, F, P, Q, X
70 CLR
80 PRINT "INTERPOLACION CUBICA"
90 INPUT "INTRODUZCA EL NUMERO DE VARIABLES "; N
100 DIM X(N), P(N), G(N), D(N)
110 PRINT "INTRODUZCA EL PUNTO INICIAL "
120 FOR I = 1 TO N: INPUT X(I): NEXT I
130 PRINT "INTRODUZCA LA DIRECCION FACTIBLE"
140 FOR I = 1 TO N: INPUT D(I): NEXT I
150 INPUT "INTRODUZCA LA PRECISION DESEADA "; E
160 INPUT "INTRODUZCA EL MINIMO SUPUESTO F(X1, ..., XN) "; FN
170 REM PUNTO P INICIAL
180 FOR I = 1 TO N: P(I) = X(I): PRINT "X": I, " = ", X(I): NEXT I
190 GOBUB 5000
200 PRINT "ITERACION": CC; "VALOR DE F": Z
210 FP = Z: GOBUB 6000: G1 = G0
220 GP = 0
230 FOR I = 1 TO N: GP = GP + G(I) * D(I): NEXT I
240 IF GP <= 0 THEN GOTO 300
241 REM BUSCA EN SENTIDO CONTRARIO A LA DIRECCION DADA
242 GK = ABS(2 * (FP - FN) * GP): IF GK > 1 THEN GK = 1
243 FOR I = 1 TO N
244 X(I) = P(I) - GK * D(I): P(I) = X(I): NEXT I
245 GOBUB 5000: FP = Z: PRINT "INESTABILIDAD?"
246 GOBUB 6000: G1 = G0: GOTO 220
250 REM BUSCA EL PUNTO INICIAL SI EL GRADIENTE > 0
300 GK = ABS(2 * (FP - FN) / GP): IF GK > 1 THEN GK = 1
310 NN = GK
320 REM ENCUENTRA EL SIGUIENTE Q
330 QQ = NN
340 FOR I = 1 TO N: Q(I) = P(I) + QQ * D(I): X(I) = Q(I): NEXT I
350 GOBUB 5000: FQ = Z
360 GOBUB 6000: G2 = G0
370 QQ = 0
380 FOR I = 1 TO N: QQ = QQ + G(I) * D(I): NEXT I
390 IF QQ > 0 OR FQ > FP THEN GOTO 420
400 REM SI NO ENCUENTRA Q ADECUADO, LO DÚPLICA
410 NN = 2 * NN: GOTO 320
420 REM EJECUTA LA INTERPOLACION CUBICA
430 ZZ = 3 * (FP - FQ) / NN: ZZ = ZZ + GP * G0
440 WW = ZZ * ZZ - GP * G0: IF WW < 0 THEN WW = 0
450 W = SQR(WW)
460 DO = NN * (1 - (G0 + W - ZZ) / (G0 - GP + 2 * W))
470 FOR I = 1 TO N: X(I) = P(I) + DO * D(I): NEXT I
480 GOBUB 5000: FR = Z

```

```

490 GOSUB 6000: G3 = G0
500 REM CALCULA EL GRADIENTE EN EL NUEVO PUNTO
510 GR = 0
520 FOR I = 1 TO N: GR = GR + G(I) * D(I): NEXT I
530 IF GR > 0 GOTO 600
540 REM ENCUENTRA EL NUEVO INTERVALO Y VERIFICA PARA TERMINAR
550 IF ABS(GR) < E THEN GOTO 670
560 HH = BB - DD
570 FOR I = 1 TO N: P(I) = X(I): PRINT "X"; I, "=", X(I): NEXT I
580 CC = CC + 1: PRINT "ITERACION"; CC, "VALOR=", Z
590 PP = Z: GP = GR: Q1 = G0: GOTO 430
600 IF ABS(GR) < E THEN GOTO 670
610 REM SE REPITE LA INTERPOLACION EN EL NUEVO INTERVALO
620 REM BB-DD,540, O DD, 630
630 HH = DD
640 FOR I = 1 TO N: Q(I) = X(I): PRINT "X"; I, "=", X(I): NEXT I
650 CC = CC + 1: PRINT "ITERACION"; CC, "VALOR =", Z
660 PP = Z: QQ = GR: Q2 = G0: GOTO 430
670 PRINT "MINIMIZACION COMPLETA"
680 PRINT "ITERACIONES="; CC, "VALOR="; Z
690 FOR I = 1 TO N
700 PRINT "X"; I, X(I)
710 NEXT I
720 END
5000 Z = 0
5010 Z = 100 * (X(2) - X(1) ^ 2) ^ 2 + (1 - X(1)) ^ 2
5020 TT = TT + 1
5030 RETURN
6000 G0 = 0
6010 G(1) = -400 * X(1) * (X(2) - X(1) ^ 2) - 2 * (1 - X(1))
6020 G(2) = 200 * (X(2) - X(1) ^ 2)
6030 FOR I = 1 TO N: G0 = G0 + G(I) ^ 2: NEXT I
6040 G0 = SQR(G0)
6050 RETURN
9999 SAVE "INTCLUB", A

```

III.3 DESARROLLO DEL PROGRAMA COMPUTACIONAL PARA EL TEMA 2 DEL CAPITULO II.

3a) METODO DE NEWTON CON APROXIMACION CUADRATICA

El programa optimiza a la función $f = \frac{x^2}{2} - \text{sen}x$ en la línea 2000.

Variables de entrada: E, Z

E: Valor de la precisión

Z: Valor inicial

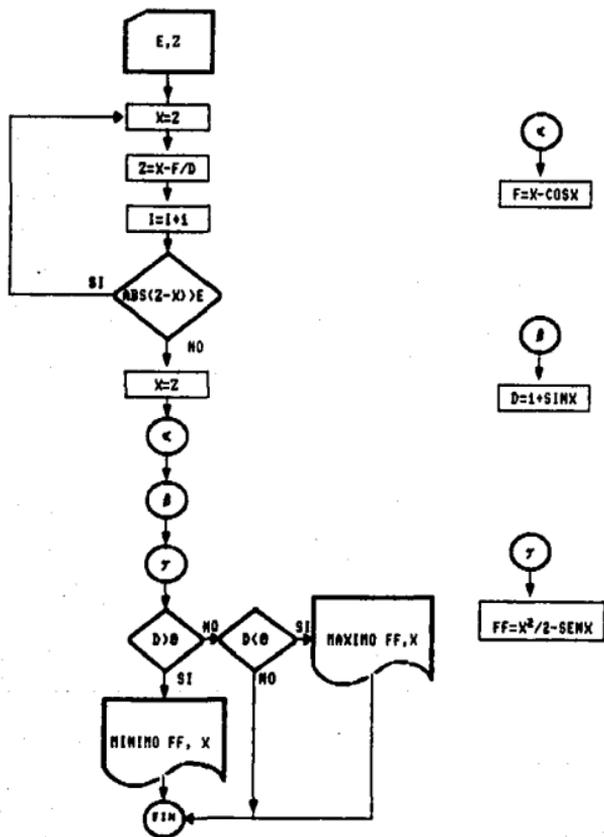
Variables internas: F: Primera derivada

D: Segunda derivada

Variables de salida: X: punto óptimo

FF: Valor de la función en el punto
óptimo

DIAGRAMA DE FLUJO (METODO DE NEWTON CON APROXIMACION CUADRATICA)



```

10 REM PROGRAMA: NEWTON.BAS
15 PRINT "PROGRAMA PARA ENCONTRAR LA RAIZ DE LOS PUNTOS DE F(X)"
20 REM F(X) SE EVALUA EN 3000
30 REM F'(X) SE EVALUA EN 1000
40 REM F''(X) SE EVALUA EN 2000
50 PRINT "PRECISION REQUERIDA: "; INPUT E
60 PRINT "VALOR INICIAL: "; INPUT Z
70 PRINT " "; PRINT "APROXIMACIONES SUCEESIVAS "
80 X = Z
90 GOSUB 1000; GOSUB 2000
100 Z = X - F / D
105 I = I + 1
110 PRINT " ITERACION "; I; " X = "; X; " SIGUIENTE PUNTO = "; Z
120 IF ABS(Z - X) > E THEN GOTO 80
130 PRINT "SOLUCION FINAL "
140 X = Z; GOSUB 1000; GOSUB 2000; GOSUB 3000
150 IF D > 0 THEN PRINT "EL MINIMO ES "; FF; " EN "; X; GOTO 200
160 IF D < 0 THEN PRINT "EL MAXIMO ES "; FF; " EN "; X; GOTO 200
200 END
1000 F = X - COS(X)
1010 RETURN
2000 D = 1 + SIN(X)
2010 RETURN
3000 FF = X * X / 2 - SIN(X)
3010 RETURN
9999 SAVE "NEWTON", A

```

3b) DESCENSO ACELERADO CASO CUADRATICO.

El programa minimiza a la función $(x_1-1)^2+(x_2-3)^2+4(x_3+5)^2$. la cual está desarrollada en la línea 2000.

Variables de entrada: N, X(I), L

N: Número de variables

X(I): Punto inicial

L: Paso inicial

Variables internas: G(I), D(I), L(I)

G(I): Derivadas parciales

D(I): Dirección de $-\nabla f(x_i)$

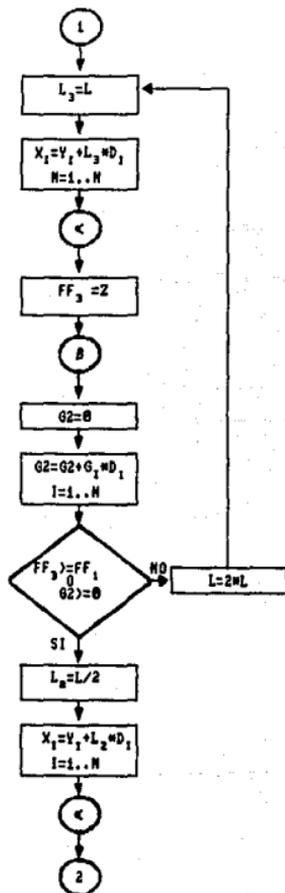
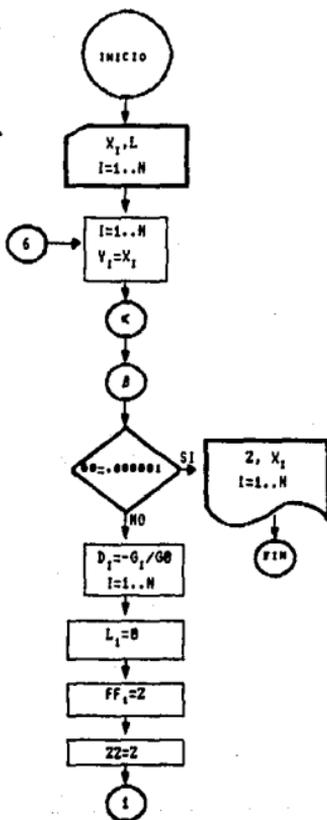
L(I): Valor de las λ_i

Variables de salida: X(I), Z

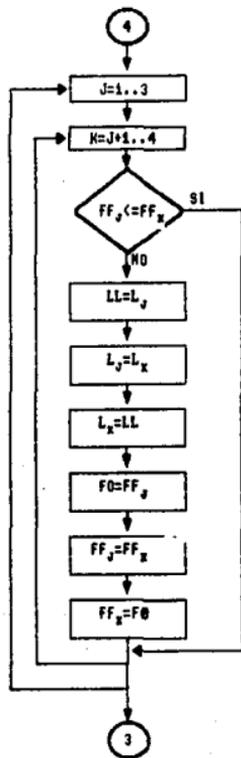
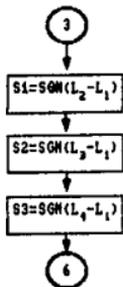
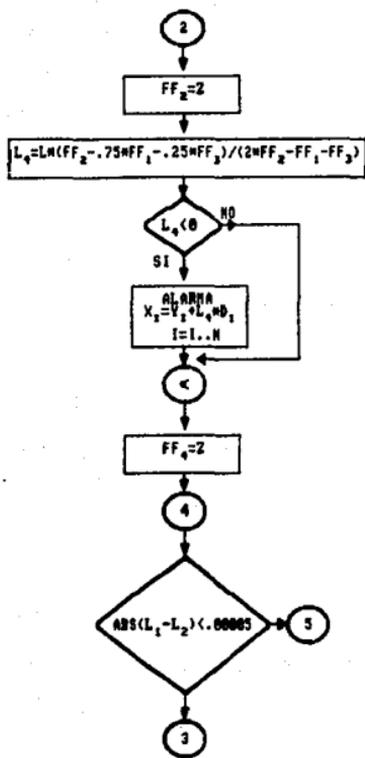
X(I): Punto a optimizar

Z: Mínimo de la función

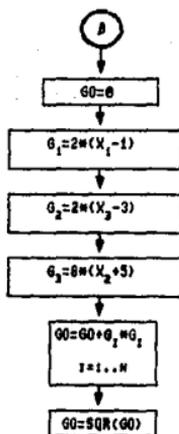
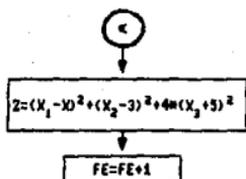
DIAGRAMA DE FLUJO (METODO DE DESCENSO ACCELERADO CASO CUADRATICO)



...(DESCENSO ACCELERADO CASO CUMMATIVO)



...(METODO DE DESCENSO ACELERADO)



```

20 PRINT "          DESCENSO ACCELERADO"
40 PRINT "  Z = F(X1,X2,...,XN) EN 2000"
60 PRINT "  DERIVADAS PARCIALES  G(1),G(2),...,G(N) EN 3000 "
80 PRINT "  BUBBLEDA LINEAL POR INTERPOLACION CUADRATICA"
100 PRINT "NUMERO DE VARIABLES ": INPUT N
110 DEFDBL D, P, G, L, X-Z
120 DIM X(N), Y(N), G(N), D(N), L(4), FF(4)
140 REM L(4) & FF(4) POR INT. CUAD.
160 PRINT "PUNTO INICIAL: "
180 FOR I = 1 TO N: INPUT X(I): NEXT I
200 PRINT "PASE INICIAL (L): ": INPUT L
250 PRINT "  VALOR EN LA ITERACION"
260 REM
300 FOR I = 1 TO N: Y(I) = X(I): NEXT I
320 GOSUB 2000: GOSUB 3000: IF G0 < .000001 THEN GOTO 1200
340 FOR I = 1 TO N: D(I) = -G(I) / G0: NEXT I
360 REM ENVIO DE DIRECCION DEL DESCENSO ACCELERADO EN Y(I)
370 REM [=X(I)] EN PRINCIPIO
380 L(1) = 0: FF(1) = Z: ZZ = Z
400 L(3) = L
410 FOR I = 1 TO N: X(I) = Y(I) + L(3) * D(I): NEXT I
430 GOSUB 2000: FF(3) = Z
440 GOSUB 3000
450 G2 = 0
460 FOR I = 1 TO N: G2 = G2 + G(I) * D(I): NEXT I
470 IF FF(3) >= FF(1) OR G2 >= 0 THEN GOTO 500
480 L = 2 * L: GOTO 400
500 REM SI OBEDECENOS 480 DOBLAMOS EL PASE DE LONGITUD
510 REM PARA CONTENER EL MINIMO
520 L(2) = L / 2
540 FOR I = 1 TO N: X(I) = Y(I) + L(2) * D(I): NEXT I
560 GOSUB 2000: FF(2) = Z
580 REM AHORA HACEMOS LA PRIMERA INTERPOLACION CUADRATICA
600 L(4) = L * (FF(2) - .75 * FF(1) - .25 * FF(3)) / (2 * FF(2) - FF(1) - FF(3))
620 IF L(4) < 0 THEN PRINT "ALARMA"
640 FOR I = 1 TO N: X(I) = Y(I) + L(4) * D(I): NEXT I
660 GOSUB 2000: FF(4) = Z
680 REM AHORA TENEMOS 4 VALORES DE LAMBIDAS Y 4 VALORES DE LA FUNCION
690 REM LOS COLOCAMOS EN ORDEN ASCENDENTE
700 FOR J = 1 TO 3
710 FOR K = J + 1 TO 4
720 IF FF(J) <= FF(K) THEN GOTO 760
730 LL = L(J): L(J) = L(K): L(K) = LL
740 FO = FF(J): FF(J) = FF(K): FF(K) = FO
750 REM CAMBIAMOS FF(J) & FF(K) SI ORDENAMOS INCORRECTO
760 NEXT K: NEXT J
790 REM FINALIZA BUBBLEDA EN ESTA DIRECCION SI LA APROXIMACION SE CUMPLE EN 800
800 IF ABS(L(1) - L(2)) < .0005 THEN GOTO 1000
810 REM RETENER LOS TRES PUNTOS MEJORES
820 B1 = SGN(L(2) - L(1)): B2 = SGN(L(3) - L(1))
830 B3 = SGN(L(4) - L(1))

```

```

040 IF S1 = S2 AND S1 = -S4 THEN L(3) = L(4): FF(3) = FF(4)
050 REN SEGUNDA Y DESPUES INTERPOLANDO
060 S0 = (L(2) - L(3)) * FF(1) + (L(3) - L(1)) * FF(2) + (L(1) - L(2)) * FF(3)
070 F = (FF(1) - FF(2)) / (2 * S0)
080 F = F * (L(2) - L(3)) + (L(3) - L(1))
090 L(4) = (L(1) + L(2)) / 2 + F
100 FOR I = 1 TO N: X(I) = V(I) + L(4) * D(I): NEXT I
110 GOTO 2000: FF(4) = Z
120 REN REPETIR SEGUNDA INTERPOLACION
130 GOTO 700
140 FOR I = 1 TO N: X(I) = V(I) + L(1) * D(I): V(I) = X(I): PRINT "X"; I; "Y"; X(I)
150 NEXT I
160 PRINT ""
170 GOTO 2000: GOTO 3000
180 PRINT "Y"; Z: PRINT ""
190 L = L / 2
200 IF S0 > .00001 THEN GOTO 340
210 REN ESTO SERA EL INICIO DE OTRA BARRERA DESDE LA POSICION ACTUAL
220 PRINT " ": PRINT
230 PRINT " SOLUCION FINAL="
240 FOR I = 1 TO N: PRINT "X"; I; "Y"; X(I): NEXT I
250 PRINT "MUNDO DE F(X1,X2,...,XN)"; Z
260 END
2700 Z = (X(1) - 1) ^ 2 + (X(2) - 3) ^ 2 + 4 * (X(3) + 5) ^ 2
2800 PE = PE + 1
2900 RETURN
3000 S0 = 0
3100 G(1) = 2 * (X(1) - 1)
3200 G(2) = 2 * (X(2) - 3)
3300 G(3) = 8 * (X(3) + 5)
3400 FOR I = 1 TO N: S0 = S0 + G(I) * D(I): NEXT I
3500 S0 = SQR(S0)
4000 RETURN
9999 SAVE "BERACE", A

```

3c) DIRECCIONES CONJUGADAS

Función a minimizar: $z=(x_1^2+x_2^2-4)$ en la línea 2000.

VARIABLES DE ENTRADA: N, X(I), D(I)

N: Número de variables

X(I): Punto inicial

D(i): Dirección inicial

VARIABLES INTERNAS: H(I,J): Matriz Hessiana

G(K): g_k

$$A = \alpha_l = \frac{g_l^T d_l}{d_l^T H d_l}$$

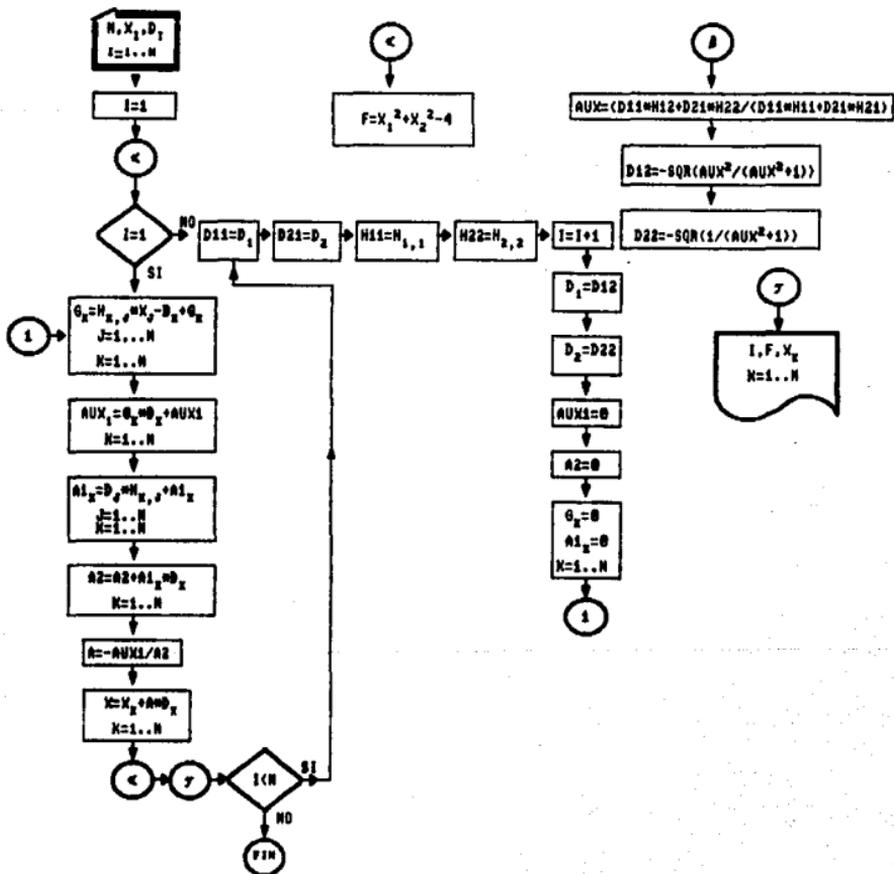
VARIABLES DE SALIDA: I, F, X(K)

I: Número de iteraciones

F: Valor de la función en el punto óptimo

X(K): Punto mínimo

DIAGRAMA DE FLUJO (DIRECCIONES CONJUGADAS)



```

10 REN DIRECCIONES CONJUGADAS
20 REN EN 1000 SE CALCULA EL VALOR DE F
30 REN
35 DEFDBL A, D, G, X
40 ON ERROR GOTO 9000
50 PRINT "NUMERO DE VARIABLES ": INPUT N
60 PRINT " PUNTO INICIAL "
65 FOR I = 1 TO N
70 PRINT " X (": I; ": ) ": INPUT X(I)
75 NEXT I
80 PRINT " DIRECCION INICIAL "
85 FOR I = 1 TO N
90 PRINT " D (": I; ": ) ": INPUT D(I)
95 NEXT I
100 I = 1: GOSUB 1000
110 IF I = 1 GOTO 130
115 D11 = D(1): D21 = D(2): N11 = N(1, 1): N22 = N(2, 2): I = I + 1
120 GOSUB 2000: D(1) = D12: D(2) = D22
125 REN CALCULA EL SIGUIENTE PUNTO
129 ALX1 = 0: A2 = 0: FOR K = 1 TO N: G(K) = 0: A1(K) = 0: NEXT K
130 FOR K = 1 TO N
140 FOR J = 1 TO N: G(K) = N(K, J) * X(J) - B(K) + G(K)
150 NEXT J: NEXT K
160 FOR K = 1 TO N: ALX1 = G(K) * D(K) + ALX1: NEXT K
170 FOR K = 1 TO N
180 FOR J = 1 TO N
190 A1(K) = D(J) * N(K, J) + A1(K)
195 NEXT J
196 NEXT K
200 FOR K = 1 TO N: A2 = A2 + A1(K) * D(K): NEXT K
220 A = -ALX1 / A2
222 FOR K = 1 TO N
225 X(K) = X(K) + A * D(K)
227 NEXT K
230 GOSUB 1000: GOSUB 900: IF I < N GOTO 115
240 END
900 PRINT "ITERACION ": I; " VALOR DE LA FUNCION ": F
910 FOR K = 1 TO N
920 PRINT "X("": K; ") = ": X(K)
930 NEXT K
940 RETURN
1000 F = X(1) * X(1) + X(2) * X(2) - 4: N(1, 1) = 2: N(2, 2) = 2: RETURN
2000 ALX = (D11 * N12 + D21 * N22) / (D11 * N11 + D21 * N21)
2010 D12 = -DBR(ALX * 2 / (ALX * 2 + 1))
2020 D22 = DBR(1 / (ALX * 2 + 1))
2030 RETURN
9000 PRINT ERR, ERR: STOP
9999 SAVE "DIRCON", A

```

3d) METODO DE GRADIENTE CONJUGADO.

El programa presentado minimiza a la función $100(x_2 - x_1)^2 + (1 - x_1)^2$ la cual está desarrollada en la línea 2000.

Variables de entrada: N, X(I)

N: Número de variables

X(I): Punto inicial

Variables internas: G(I), D(I), AK, K

G(I): Derivadas parciales

D(I): Dirección de $-g_i$ ($g_i = \nabla f(x_i)$)

AK: Es g_{i+1}^T / g_i^T

K: Auxiliar λ

Variables de salida: DV, X(I), Z, SV

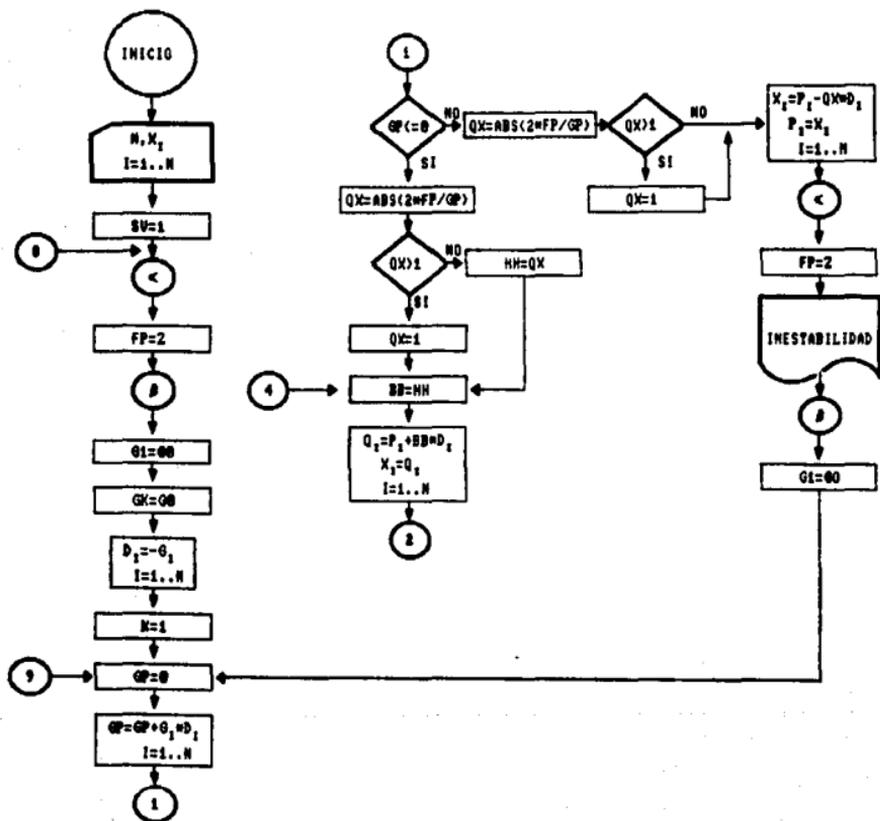
DV: Número de dirección

X(I): Punto de la iteración

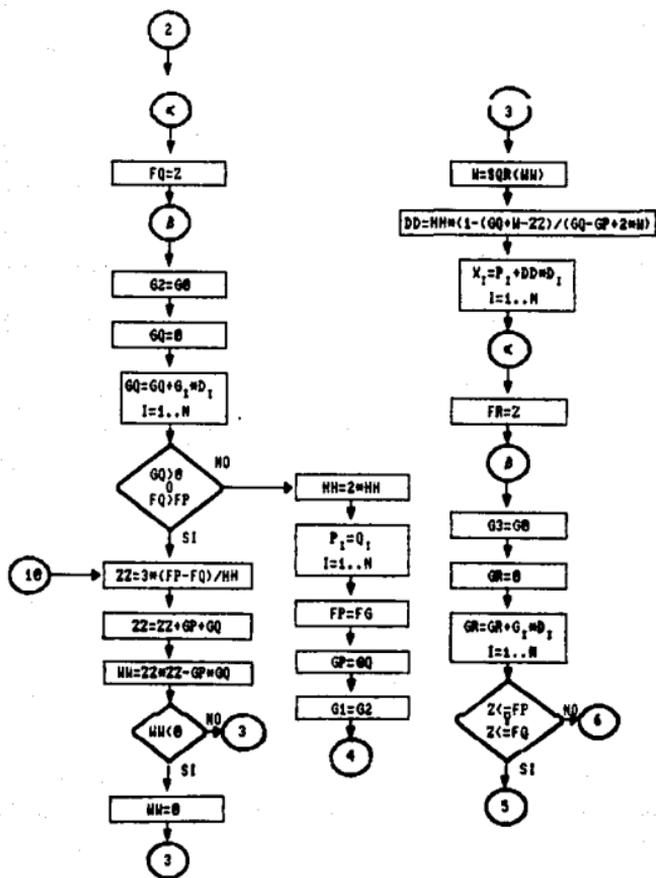
Z: Mínimo de la iteración

SV: Número de la iteración

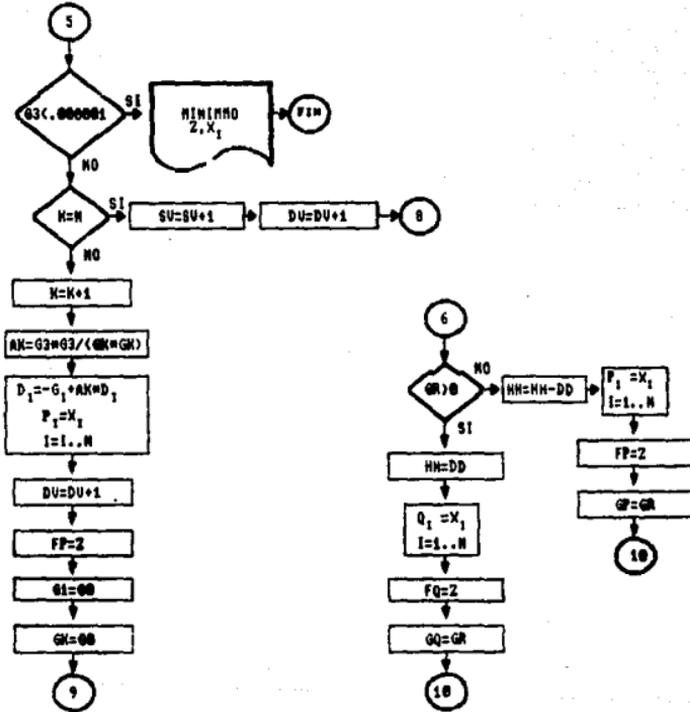
DIAGRAMA DE FLUJO (MÉTODO DEL GRADIENTE CONJUGADO)



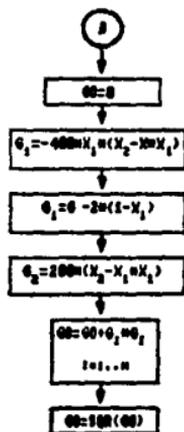
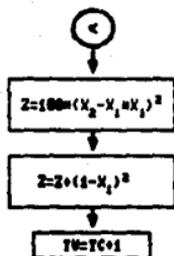
... (METODO DEL GRADIENTE CONJUGADO)



...(METODO DEL GRADIENTE CONJUGADO)



... (METODO DEL GRADIENTE CONJUGADO)



```

10 REN PROGRAMA: GRACOM.BAS
90 CLS
100 PRINT "      MINIMIZACION POR GRADIENTE CONJUGADO (FLETCHER-REEVES)"
120 REN BUSQUEDA LINEAL POR INTERPOLACION CUBICA"
150 REN F(X1,X2,...,XN) EN 5000 ; G(1),G(2),...,G(N) EN 6000
200 PRINT "NUMERO DE VARIABLES :": INPUT N
210 DEFDBL A, D, F-N, P-Q, X-Z
220 DIM X, Y, P, Q, D, G(N)
300 PRINT "VALORES INICIALES:"
320 FOR I = 1 TO N: PRINT "X"; I, : INPUT X(I): NEXT I: PRINT " = "
340 REN ESTABLECER EL CALCULO DE LA BUSQUEDA
350 SV = 1
360 REN INTERMEDIATE OUTPUT
380 PRINT "      VALOR EN LA ITERACION"
550 FOR I = 1 TO N: P(I) = X(I): PRINT "X"; I, " = "; X(I): NEXT I
560 GOBUB 5000: FP = Z: PRINT "Z="; Z
570 GOBUB 6000: G1 = G0: G2 = G0
575 REN ESTABLECCE LA PRIMERA BUSQUEDA EN DIRECCION DESCENSO ACCELERADO
580 FOR I = 1 TO N: D(I) = -G(I): NEXT I
585 REN K CONTADOR DE LAS ITERACIONES
590 K = 1
600 GP = 0
610 FOR I = 1 TO N: GP = GP + G(I) * D(I): NEXT I
620 IF GP <= 0 THEN GOTO 680
625 REN ENCONTRAR EL PASO INICIAL Y SI ES NECESARIO REGRESAR
630 GX = ABS(Z * FP / GP): IF GX > 1 THEN GX = 1
640 FOR I = 1 TO N
650 X(I) = P(I) - GX * D(I): P(I) = X(I): NEXT I
660 GOBUB 5000: FP = Z: PRINT "INESTABILIDAD"
670 GOBUB 6000: G1 = G0: GOTO 600
680 GX = ABS(Z * FP / GP): IF GX > 1 THEN GX = 1
690 NN = GX
700 REN ENCONTRAR EL PROXIMO PUNTO
710 Q0 = NN
720 FOR I = 1 TO N
730 Q(I) = P(I) + Q0 * D(I): X(I) = Q(I)
740 NEXT I
750 GOBUB 5000: FQ = Z
760 GOBUB 6000: G2 = G0
770 Q0 = 0
780 FOR I = 1 TO N
790 Q0 = Q0 + Q(I) * D(I)
800 NEXT I
810 IF Q0 > 0 OR FQ > FP THEN GOTO 860
815 REN HACER PASO DE DOBLE INTERPOLACION Y CAMBIAR A Q
820 NN = 2 * NN
830 FOR I = 1 TO N: P(I) = Q(I): NEXT I
840 FP = FQ: GP = G0: G1 = G2
850 GOTO 710
860 ZZ = 3 * (FP - FQ) / NN: Z2 = Z2 + GP + Q0
870 W0 = Z2 * Z2 - GP * Q0: IF W0 < 0 THEN W0 = 0

```

```

080 M = 90R(LM)
090 DD = HH * (1 - (GR + M - Z) / (GO - GP + 2 * M))
900 FOR I = 1 TO N: X(I) = P(I) + DD * D(I): NEXT I
910 GOBUS 5000: FR = Z
920 GOBUS 6000: G3 = GO
925 REM ENCONTRAR GRADIENTE EN EL NUEVO PUNTO
930 GR = 0
940 FOR I = 1 TO N: GR = GR + G(I) * D(I): NEXT I
950 IF Z <= FP AND Z <= FQ THEN GOTO 1100
960 IF GR > 0 THEN GOTO 1020
990 HH = NN - DD
1000 FOR I = 1 TO N: P(I) = X(I): NEXT I
1010 FP = Z: GP = GR: GOTO 860
1020 NN = DD
1030 FOR I = 1 TO N: G(I) = X(I): NEXT I
1040 FQ = Z: GO = GR: GOTO 860
1100 REM PRUEBA TERMINADA
1110 IF G3 < .000001 THEN GOTO 1300
1120 IF K = N THEN GOTO 1250
1130 REM NUMERO DE ITERACIONES
1140 K = K + 1
1150 REM ENCONTRAR DIRECCION CONJUGADA
1160 AK = G3 * G3 / (GK * GK)
1170 FOR I = 1 TO N: D(I) = -G(I) + AK * D(I): P(I) = X(I): NEXT I
1200 PRINT "NUEVA DERECCION"; : DV = DV + 1: PRINT " BUSQUEDA "; DV
1210 FP = Z: G1 = GO: GK = GO
1220 FOR I = 1 TO N: PRINT "X"; I; " = "; X(I): NEXT I: PRINT "Z="; Z
1230 GOTO 800
1250 PRINT "REINICIAR"; : SV = SV + 1: DV = DV + 1
1260 PRINT " ITERACION"; SV; "BUSQUEDA"; DV
1270 PRINT ""
1280 GOTO 550
1300 PRINT "MINIMO ENCONTRADO"
1320 FOR I = 1 TO N: PRINT "X"; I; " = "; X(I): NEXT I
1340 PRINT "MINIMO DE LA FUNCION="; Z
1350 PRINT "EVALUACIONES DE LA FUNCION="; TV
1400 END
5000 Z = 100 * (X(2) - X(1) * X(1)) ^ 2
5010 Z = Z + (1 - X(1)) ^ 2
5020 TV = TV + 1
5100 RETURN
6000 BB = 0
6100 G(1) = -400 * X(1) * (X(2) - X(1) * X(1))
6110 G(1) = G(1) - 2 * (1 - X(1))
6200 G(2) = 200 * (X(2) - X(1) * X(1))
7000 FOR I = 1 TO N: GO = GO + G(I) * G(I): NEXT I
7100 BB = DSR(GO)
7500 RETURN
9999 SAVE "GRACON", A

```

3e) METODO DE QUASI-NEWTON; METODO DE DAVIDON FLETCHER Y POWELL.

La función a optimizar en este programa es:

$$z=(x_1+10x_2)^2+5(x_3-x_4)^2+(x_1-2x_3)^4+10(x_1-x_4)^4$$

Variables de entrada: N, X(I)

N: Número de variables

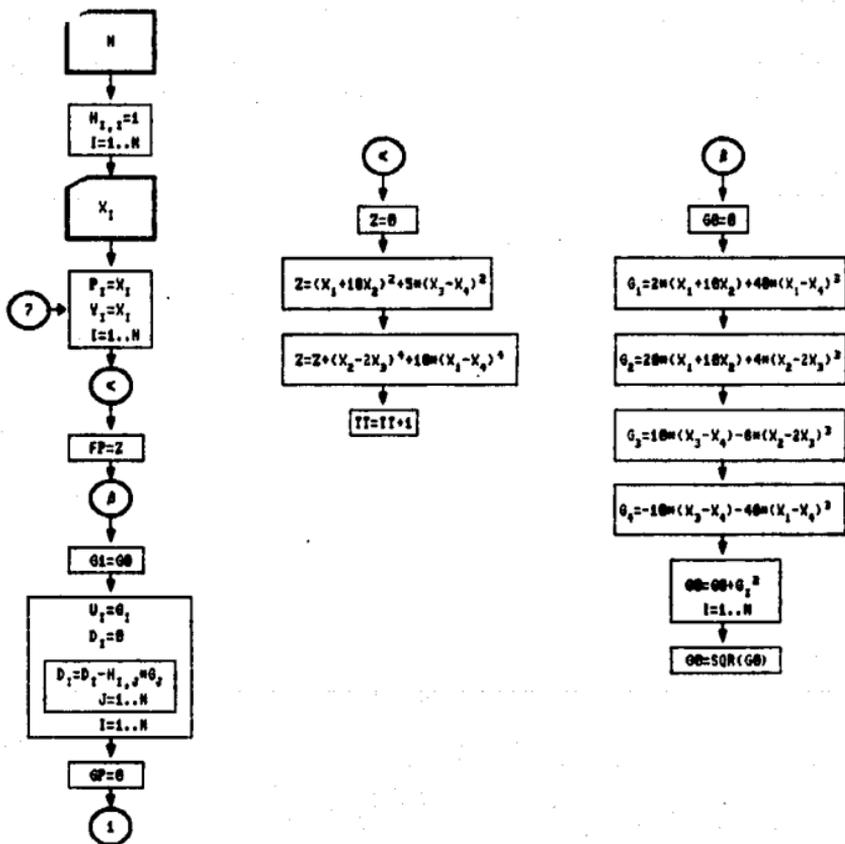
X(I): Punto inicial

Variables internas: H, G(I)

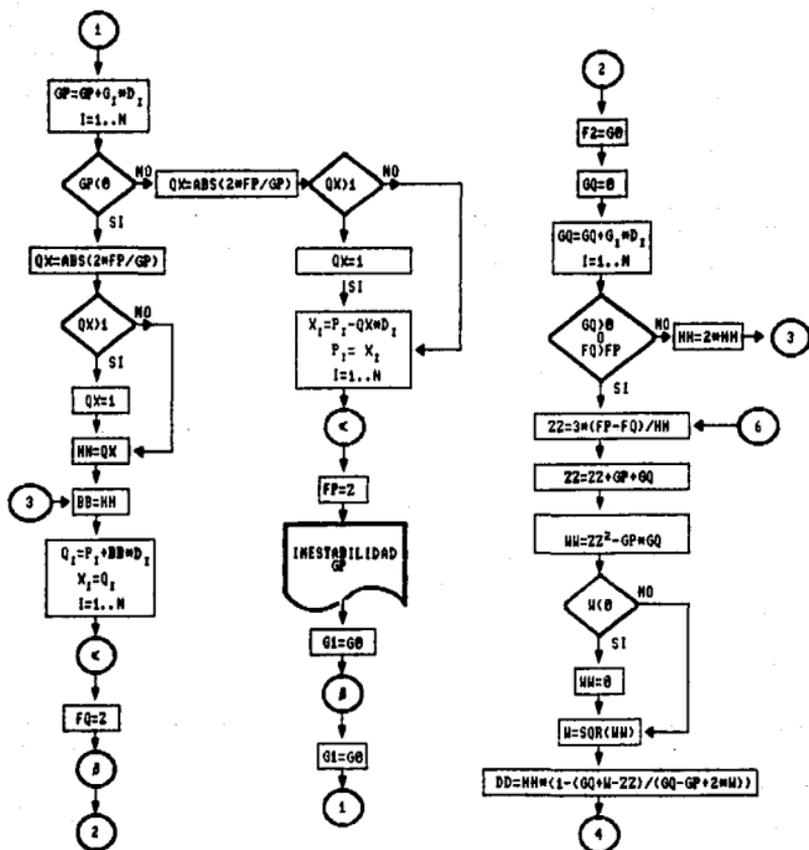
H: Matriz hessiana inicialmente la
matriz identidad

Z: Valor de la función en el punto
mínimo

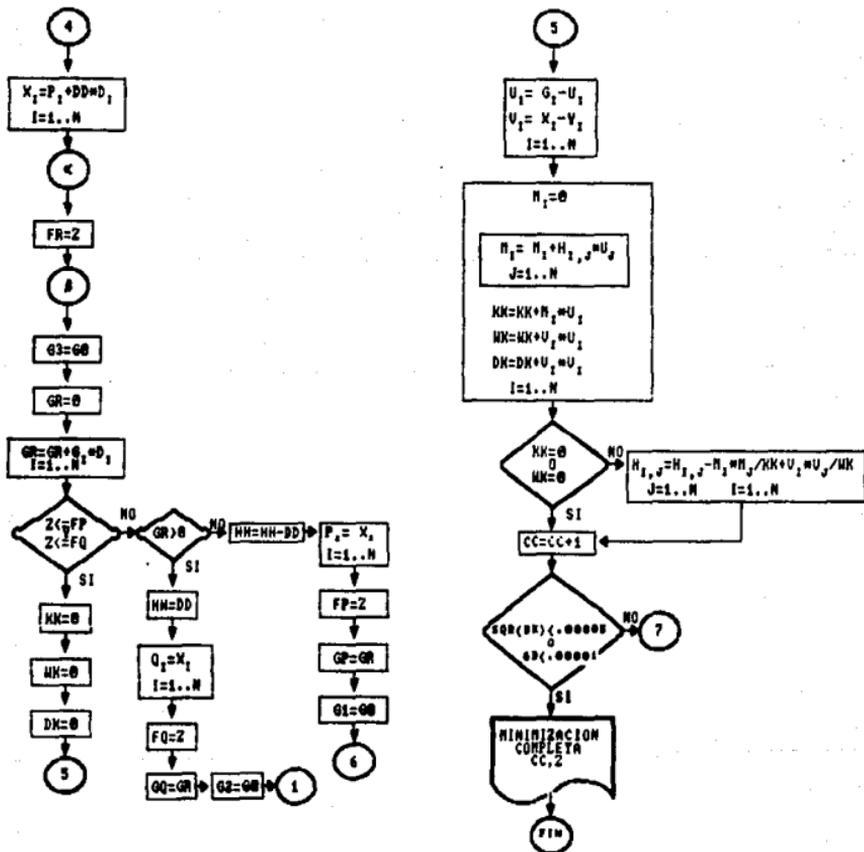
DIAGRAMA DE FLUJO (METODO DE QUASI NEWTON)



... (MÉTODO DE QUINZI-NEWTON)



...(METODO DE QUASI NEWTON)



```

100 PRINT "MINIMIZACION POR D-F-P"
120 REM SUBRUEDAS LINEALES POR INTERPOLACION CUBICA
150 REM P(X1,X2,...,XN) EN 5000; G(1),G(2),...,G(N) EN 6000
200 PRINT "NUMERO DE VARIABLES " : INPUT N
210 DEFDBL D, G, H, M, P, R, U, V, W, Z
220 DIM X, P, G, H, D, G, U, V, W, H(N)
240 DIM N(N, N)
245 ON ERROR GOTO 9000
300 REM ESTABLECER N INICIALMENTE MATRIZ IDENTIDAD
320 FOR I = 1 TO N: N(I, I) = 1: NEXT I
330 PRINT "VALORES INICIALES "
340 FOR I = 1 TO N: PRINT "X " ; I ; : INPUT X(I): NEXT I: PRINT " = "
360 REM SALIDA INTERMEDIA
380 PRINT " VALORES"
400 FOR I = 1 TO N: P(I) = X(I): V(I) = X(I): PRINT "X " ; I, X(I): NEXT I
410 GOSUB 5000
420 PRINT "ITERACION " ; CC ; : VALOR " ; Z
430 FP = Z: GOSUB 6000: G1 = G0
440 REM AGREGAR EL GRADIENTE EN U Y DIRECCION DE SUBRUEDA D
450 FOR I = 1 TO N
460 U(I) = G(I): D(I) = 0
470 FOR J = 1 TO N
480 D(I) = D(I) - N(I, J) * G(J)
490 NEXT J
500 NEXT I
600 GP = 0
610 FOR I = 1 TO N: GP = GP + G(I) * D(I): NEXT I
620 IF GP < 0 THEN GOTO 680: REM *****
625 REM ENCUENTRA EL PUNTO INICIAL Y SI ES NECESARIO REGRESAR
630 GX = ABS(2 * FP / GP): IF GX > 1 THEN GX = 1
640 FOR I = 1 TO N
650 X(I) = P(I) - GX * D(I): P(I) = X(I): NEXT I
660 GOSUB 5000: FP = Z: PRINT "INESTABILIDAD? " ; GP: REM *****
670 GOSUB 6000: G1 = G0: GOTO 600
680 GX = ABS(2 * FP / GP): IF GX > 1 THEN GX = 1
690 HX = GX
700 REM ENCUENTRA EL PUNTO SIGUIENTE Q
710 QB = HX
720 FOR I = 1 TO N
730 Q(I) = P(I) + QB * D(I): X(I) = Q(I)
740 NEXT I
750 GOSUB 5000: FQ = Z
760 GOSUB 6000: G2 = G0
770 QB = 0
780 FOR I = 1 TO N
790 QR = QR + G(I) * D(I)
800 NEXT I
810 IF G0 > 0 OR FQ > FP THEN GOTO 830
815 REM REALIZA LA INTERPOLACION CUBICA O DUPLICA EL INTERVALO DEL MINIMO
820 HX = 2 * HX: GOTO 700
830 ZZ = 3 * (FP - FQ) / HX: ZZ = ZZ + GP + QR
840 HX = ZZ * ZZ - GP * G0: IF HX < 0 THEN HX = 0
850 M = SQR(HX)
860 DD = HX * (1 - (G0 + M - ZZ) / (G0 - GP + 2 * M))
870 FOR I = 1 TO N: X(I) = P(I) + DD * D(I): NEXT I
880 GOSUB 5000: FR = Z
890 GOSUB 6000: G3 = G0
895 REM ENCUENTRA EL GRADIENTE EN EL NUEVO PUNTO
900 GR = 0
910 FOR I = 1 TO N: GR = GR + G(I) * D(I): NEXT I
920 IF Z <= FP AND Z <= FQ THEN GOTO 1100
930 IF GR > 0 THEN GOTO 990
960 HX = HX - DD
970 FOR I = 1 TO N: P(I) = X(I): NEXT I
980 FP = Z: GP = GR: G1 = G0: GOTO 830
990 HX = G0
1000 FOR I = 1 TO N: G(I) = X(I): NEXT I
1010 P0 = Z: QR = GR: G2 = G0: GOTO 830
1050 REM ACTUALIZAR LA MATRIZ
1100 KX = G: KX = 0: KX = 0
1110 FOR I = 1 TO N
1120 KX(I) = G(I) - KX(I): V(I) = X(I) - V(I)
1130 NEXT I
1140 FOR I = 1 TO N: KX(I) = 0
1150 FOR J = 1 TO N
1160 KX(I) = KX(I) + N(I, J) * U(J)
1170 NEXT J

```

```

1180 KK = KK + H(I) * U(I); MK = MK + V(I) * U(I)
1190 DK = DK + V(I) * V(I)
1200 NEXT I
1205 IF KK = 0 OR MK = 0 THEN GOTO 1260
1210 FOR I = 1 TO N
1220 FOR J = 1 TO N
1230 H(I, J) = H(I, J) - H(I) * H(J) / KK + V(I) * V(J) / MK
1240 NEXT J
1250 NEXT I
1260 CC = CC + 1
1265 REM test para firmlizar
1270 IF ABS(DK) < .0005 OR CC < .00001 THEN GOTO 1300
1275 REM COMENZAR NUEVA SUBLEDA
1280 GOTO 400
1300 PRINT "MINIMIZACION COMPLETA"
1310 PRINT "ITERACION "; CC; " VALOR "; Z
1320 FOR I = 1 TO N
1330 PRINT " X "; I, X(I)
1340 NEXT I
1380 END
5000 Z = 0
5010 Z = (X(1) + 10 * X(2)) ^ 2 + 5 * (X(3) - X(4)) ^ 2
5020 Z = Z + (X(2) - 2 * X(3)) ^ 4 + 10 * (X(1) - X(4)) ^ 4
5100 TT = TT + 1
5200 RETURN
6000 GO = 0
6100 G(1) = 2 * (X(1) + 10 * X(2)) + 40 * (X(1) - X(4)) ^ 3
6200 G(2) = 20 * (X(1) + 10 * X(2)) + 4 * (X(2) - 2 * X(3)) ^ 3
6300 G(3) = 10 * (X(3) - X(4)) - 5 * (X(2) - 2 * X(3)) ^ 3
6400 G(4) = -10 * (X(3) - X(4)) - 40 * (X(1) - X(4)) ^ 3
7000 FOR I = 1 TO N: GO = GO + G(I) * G(I): NEXT I
7010 GO = ABS(GO)
7500 RETURN
9000 PRINT ERR, ERR; STOP
9999 REM SAVE"PCMELL",A

```

3f) METODO DE HOOKE & JEEVES

La función a optimizar en este programa es

$$z=(x_1-2)^2+(x_2-5)^2+(x_3+2)^4.$$

Variables de entrada: N, X(I), H

N: Número de variables

X(I): Punto inicial

H: Longitud de paso

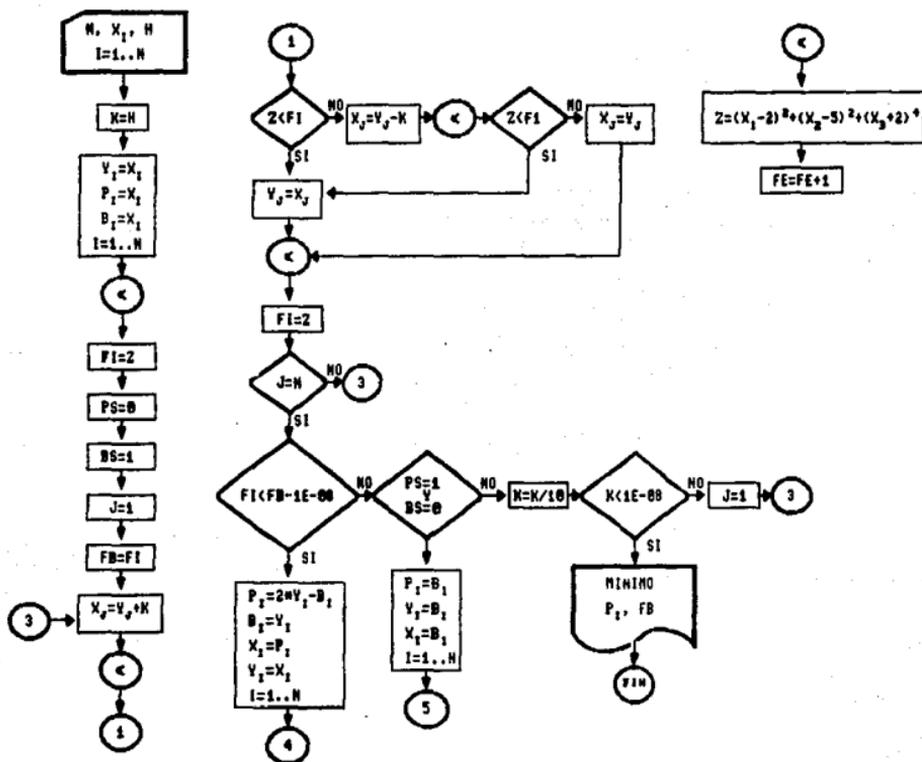
Variables de salida: P(I), FB, FE

P(I): Punto mínimo

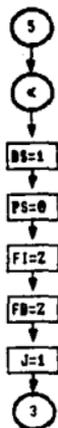
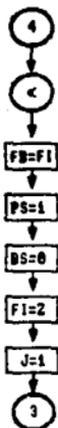
FB: Valor de la función en el punto
mínimo

FE: Número de valuación de la función.

DIAGRAMA DE FLUJO (METODO DE HOOKE & JEEVES)



...(METODO DE HOOKE & JEEVES)



```

10 PRINT "      HOOKE & JEEVES"
20 REM CALCULA Z=F(X1,X2,...,XN) EN 2000
30 PRINT "NUMERO DE VARIABLES ": INPUT N
40 DIM X(N), B(N), Y(N), P(N)
50 PRINT "PUNTO INICIAL X1, X2,..., XN"
60 FOR I = 1 TO N: INPUT X(I): NEXT I
70 PRINT "LONGITUD DE PASO": INPUT H
80 K = H
90 FOR I = 1 TO M
100 Y(I) = X(I): P(I) = X(I): B(I) = X(I): NEXT I
110 GOSUB 2000: F1 = Z
120 PRINT "VALOR INICIAL ": Z
130 FOR I = 1 TO N: PRINT X(I); : NEXT I: PRINT " "
140 PS = 0: BS = 1
150 REM EXPLORACION ALREDEDOR DEL PUNTO BASE
180 J = 1: FB = F1
200 X(J) = Y(J) + K
210 GOSUB 2000
220 IF Z < F1 THEN GOTO 280
230 X(J) = Y(J) - K
240 GOSUB 2000
250 IF Z < F1 THEN GOTO 280
260 X(J) = Y(J)
270 GOTO 290
280 Y(J) = X(J)
290 GOSUB 2000
300 F1 = Z
310 PRINT "PASO DE EXPLORACION ": Z
320 FOR I = 1 TO N: PRINT X(I); : NEXT I: PRINT " "
330 IF J = N THEN GOTO 360
340 J = J + 1
350 GOTO 200
360 IF F1 < FB - 1E-08 THEN GOTO 540
370 REM DESPUES DE 360 REALIZA UN PATRON DE MOVIMIENTO SI LA FUNCION HA SIDO REDUCIDA
380 IF PS = 1 AND BS = 0 THEN GOTO 420
390 REM PERO SI LA EXPLORACION FUE ALREDEDOR DEL PATRON PY
400 REM Y NO HUBO REDUCCION SE CAMBIA LA BASE EN 420
410 GOTO 490
420 FOR I = 1 TO N: P(I) = B(I): Y(I) = B(I): X(I) = B(I): NEXT I
430 GOSUB 2000: BS = 1: PS = 0
440 F1 = Z: FB = Z
450 PRINT "CAMBIO DE BASE ": Z
460 FOR I = 1 TO N: PRINT X(I); : NEXT I: PRINT " "
470 REM (SIGUE DE 395) Y EXPLORA ALREDEDOR DEL NUEVO PUNTO BASE
480 J = 1: GOTO 200
490 K = K / 10
500 PRINT "SE CONTRAE LA LONGITUD DEL PASO"
510 IF K < 1E-08 THEN GOTO 700
520 REM SI NO MENOS TERMINADO SE HACE UNA NUEVA EXPLORACION
525 REM ALREDEDOR DEL ULTIMO PUNTO BASE
530 J = 1: GOTO 200

```

```

535 REN MODELO DE MOVIMIENTO
540 FOR I = 1 TO N: P(I) = 2 * Y(I) - B(I)
550 B(I) = Y(I): X(I) = P(I): Y(I) = X(I)
560 NEXT I
570 GOBUN 2000: FB = FI: PS = 1: BS = 0: FI = 2
580 PRINT "MODELO DE MOVIMIENTO ": Z
590 FOR I = 1 TO N: PRINT X(I); : NEXT I: PRINT " =
600 REN ENTONCES SE EXPLORA ALREDEDOR DEL ULTIMO MODELO DE MOVIMIENTO
610 J = 1: GOTO 200
700 PRINT "          MINIMO ENCONTRADO"
710 FOR I = 1 TO N: PRINT "X "; I; " = "; P(I): NEXT I: PRINT " =
750 PRINT "MINIMO DE LA FUNCION = "; FB
760 PRINT "NO. DE VALUACIONES DE LA FUNCION = "; FE
790 END
2000 Z = (X(1) - 2) ^ 2 + (X(2) - 5) ^ 2 + (X(3) + 2) ^ 4
2010 FE = FE + 1
2020 REN CUENTA LAS VALUACIONES DE LA FUNCION
2030 RETURN
9999 SAVE "HOOKE", A

```

CAPITULO IV

ayudarán a establecer las condiciones necesarias y suficientes para la existencia de puntos óptimos.

Definición.- Un punto x^* que satisface la restricción $h(x^*)=0$ se dice que es un punto regular de la restricción si los vectores gradiente $\nabla h_1(x^*), \nabla h_2(x^*), \dots, \nabla h_m(x^*)$, son linealmente independientes.

Teorema IV0.1.- En un punto regular x^* de la superficie S definida por $h(x) = 0$ el plano tangente es igual a:

$$M = \{ y; \nabla h(x^*) y = 0 \}$$

CONDICIONES NECESARIAS DE PRIMER ORDEN (Restricciones de igualdad)

Lemma IV0.1.- Sea x^* un punto regular de las restricciones $h(x)=0$ y un punto local extremo (máximo o mínimo) de f sujeto a esas restricciones, entonces $\forall y \in E^n$ que satisface

$$\nabla h(x^*) y = 0$$

también se cumple que

$$\nabla f(x^*) y = 0$$

Teorema IV0.2.- Sea x^* un punto extremo local de f sujeto a $h(x)=0$. Si además se supone que x^* es un punto regular de esas restricciones, entonces existe una $\lambda \in E^m$ tal que

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

Definase el lagrangiano asociado al problema restringido

$$\phi(x, \lambda) = f(x) + \lambda^T h(x)$$

Las condiciones necesarias podrían ser expresadas por

$$\nabla_x \phi(x, \lambda) = 0$$

$$\nabla_x \phi(x, \lambda) = 0$$

CONDICIONES NECESARIAS DE SEGUNDO ORDEN

Supóngase que x^* es un mínimo local de f sujeto a $h(x)=0$ y que x^* es un punto regular de estas restricciones.

Entonces existe una $\lambda \in E^m$ tal que

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

Si se denota M (plano tangente) como

$$M = \{y; \nabla h(x^*)^T y = 0\}$$

entonces la matriz

$$L(x^*) = F(x^*) + \lambda^T H(x^*)$$

es positiva semidefinida sobre M , esto es, $y^T L(x^*) y \geq 0 \forall y \in M$.

CONDICIONES SUFICIENTES DE SEGUNDO ORDEN

Supóngase que existe un punto x^* que satisface $h(x^*) = 0$ y una $\lambda \in E^m$ tal que

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0$$

Supóngase también que la matriz $L(x^*) = F(x^*) + \lambda^T H(x^*)$ es positiva definida sobre $M = \{y; \nabla h(x^*)^T y = 0\}$, es decir, para $y \in M$, $y \neq 0$ se cumple que $y^T L(x^*) y > 0$. Entonces x^* es un mínimo local estricto de f sujeto a $h(x) = 0$.

RESTRICCIONES DE DESIGUALDAD

Considérese el problema

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & h(x) = 0 \quad \dots \text{IVO.3} \\ & g(x) \leq 0 \end{aligned}$$

CONDICIONES DE PRIMER ORDEN

Definición.- Sea x^* un punto que satisface las siguientes restricciones:

$$h(x^*) = 0, \quad g(x^*) \leq 0, \quad \dots \text{IVO.4}$$

y sea J un conjunto de índices j para cada $g_j(x^*) = 0$. Entonces x^* que es un punto regular de las restricciones IVO.4, si los vectores gradientes $\nabla h_i(x^*)$, $\nabla g_j(x^*)$, $1 \leq i \leq m$ e J son linealmente independientes.

Condiciones de Kuhn-Tucker.- Sea x^* un punto mínimo relativo para el problema

$$\begin{aligned} \text{minimizar} \quad & f(x) \\ \text{s.a.} \quad & h(x) = 0, \quad g(x) \leq 0 \end{aligned}$$

y supóngase que x^* es un punto regular para las restricciones. Entonces existe un vector $\lambda \in E^m$ y un vector $\mu \in E^n$ con $\mu \geq 0$ tal que

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0 \quad \dots \text{IVO.5}$$

$$\mu^T \nabla g(x^*) = 0 \quad \dots \text{IVO.6}$$

Condiciones necesarias de segundo orden.- Supóngase que las

funciones $f, g, h \in C^2$ y que x^* es un punto regular de las restricciones IVO.4. Si x^* es un mínimo relativo de IVO.3, entonces existe una $\lambda \in E^m$, $\mu \in E^p$, $\mu \geq 0$ tal que (IVO.5) y (IVO.6) se cumplen y tal que

$$L(x^*) = F(x^*) + \lambda^T H(x^*) + \mu^T G(x^*)$$

es positiva semidefinida sobre el subespacio tangente de las restricciones activas en x^* .

Condiciones suficientes de segundo orden.- Sea $f, g, h \in C^2$. Una condición suficiente para que un punto x^* que satisface IVO.4 sea un mínimo relativo del problema (IVO.3) es que existe $\lambda \in E^m$, $\mu \in E^p$, tal que

$$\mu \geq 0$$

$$\mu^T q(x^*) = 0$$

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g(x^*) = 0$$

y la matriz Hessiana

$$L(x^*) = F(x^*) + \lambda^T H(x^*) + \mu^T G(x^*)$$

es positiva definida sobre el subespacio

$$M' = \{y : \nabla h(x^*)y = 0, \nabla g_j(x^*)y = 0 \forall j \in J\}$$

donde

$$J = \{j : g_j(x^*) = 0, \mu_j > 0\}$$

IV.1. METODO DE ZOUTENDIJK

Para este algoritmo considérese el problema IVO.2

La forma general de éste y los siguientes métodos es la siguiente:

$$x_{k+1} = x_k + \alpha_k d_k \quad \dots IV1.1$$

donde d_k es una dirección factible y de descenso y α_k se escoge como

$$\begin{aligned} \min. f(x_k + \alpha d_k) & \quad \dots IV1.2 \\ 0 \leq \alpha \leq \bar{\alpha} \end{aligned}$$

Recordando que d es una dirección factible en x si existe $\bar{\alpha} > 0$ tal que $x + \alpha d$ es factible para $0 \leq \alpha \leq \bar{\alpha}$.

A continuación se verán dos casos:

1) Problema con restricciones lineales

min $f(x)$

$$\text{s.a. } a_l^T x \leq b_l, \quad l = 1, \dots, m$$

Dado el punto x_0 se obtendrán los siguientes puntos de esta forma:

dado x_k se resolverá el siguiente problema acotado de programación no lineal:

$$\begin{aligned} \min \nabla f(x_k)^T d \\ \text{s.a. } a_l^T x_k - b_l + a_l^T d \leq 0 \\ |d_j| \leq 1 \quad j = 1, \dots, n \end{aligned}$$

Se llama d_k a la solución de este problema, como $d=0$ es un punto factible, $\nabla f(x_k)^T d_k \leq 0$; luego d_k es una dirección de

descenso. Además:

$$a_l^T(x_k + \alpha d_k) = a_l^T x_k + \alpha a_l^T d_k \leq b_l \text{ para } \alpha \geq 0$$

luego d_k es una dirección factible. Ahora se elige:

$$x_{k+1} = x_k + \alpha_k d_k$$

donde α_k es el mínimo para $0 \leq \alpha \leq 1$ de $f(x_k + \alpha d_k)$. La condición $|d_j| \leq 1$ tiene como objetivo que el problema IV1.3 sea acotado y, por lo tanto tenga solución.

2) Problema con restricciones no lineales

$$\min f(x)$$

$$\text{s.a. } g_l(x) \leq 0 \quad (l = 1, \dots, m)$$

En este caso, dado x_k factible considérese el problema

$$\min \nabla f(x_k)^T y$$

$$\text{s.a. } g_l(x_k) + \nabla g_l(x_k)^T y \leq 0 \quad l = 1, \dots, m \quad \dots \text{IV1.4}$$

$$|y_j| \leq 1 \quad j = 1, \dots, n$$

Sea y_k el óptimo de IV1.4. Supóngase que existe $\bar{\alpha}$, $0 \leq \bar{\alpha} \leq 1$ tal que $x_k + \alpha y_k$ es factible para $0 < \alpha < \bar{\alpha}$. Sea ahora:

$$x_{k+1} = x_k + \alpha_k d_k$$

donde α_k es el mínimo de $f(x_k + \alpha y_k)$ para

$$0 < \alpha < \bar{\alpha}.$$

El problema esencial de este método es la determinación de α , ya que en algunos casos podría no existir.

IV.2 METODO DE ROSEN DE LA PROYECCION DEL GRADIENTE

Considérese el siguiente problema de optimización:

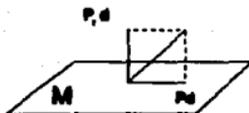
$$\begin{aligned}
 & \min f(x) \\
 & \text{s.a. } (a_i)^T x \leq b_i \quad i \in I_1 \\
 & \quad (a_i)^T x = b_i \quad i \in I_2 \quad \dots \quad \text{IV2.1}
 \end{aligned}$$

Sea x un punto factible, $I(x) = \{ i : (a_i)^T x = b_i \}$ o sea $I(x)$ corresponde al conjunto de índices de las restricciones activas en x . Esto quiere decir que I_2 está incluido en $I(x)$.

Sea M el subespacio definido por:

$$M = \{ y : (a_i)^T y = 0, i \in I(x) \}$$

Dado un vector $d \in \mathbb{R}^n$, $d = P_d d + P_\perp d$ donde $P_d d \in M$ y $P_\perp d$ es ortogonal a M como se muestra en la siguiente figura:



Nótese que $a_i, i \in I(x)$ genera el subespacio ortogonal a M . luego

$$P_d^\perp = \sum u_i a_i \quad i \in I(x).$$

P_d es la proyección ortogonal de d en M .

Ahora bien el negativo de la proyección del gradiente de una función $f(x)$ es una dirección de descenso, ya que $\nabla f(x)^T (-P \nabla f(x)) \leq 0$ como se puede observar:

Haciendo $d = -\nabla f(x)$ se tiene

$$\begin{aligned} \nabla f(x)^T (-P\nabla f(x)) &= -d^T P d \\ &= -(P_d d + P d)^T P d \\ &= -(P d)^T P d \\ &= -\|P d\|^2 \leq 0 \end{aligned}$$

Además, $-P\nabla f(x)$ es una dirección factible porque se cumplen:

$$\begin{aligned} \text{i)} \quad a_l^T (x - \alpha P\nabla f(x)) &= b_l, & l \in I(x) \\ \text{ii)} \quad a_l^T (x - \bar{\alpha} P\nabla f(x)) &\leq b_l, & l \in I(x) \end{aligned}$$

para un $\bar{\alpha}$ que se definirá posteriormente.

En el caso de $l \in I(x)$ se tiene

$$a_l^T (x - \alpha P\nabla f(x)) = a_l^T x - \alpha a_l^T P\nabla f(x) = b_l$$

porque por definición $P\nabla f(x) \in M$

En el caso $l \notin I(x)$ se tiene:

$$a_l^T x < b_l$$

Definiendo un α_l positivo para el cual se cumple

$$a_l^T (x - \alpha_l d) = b_l$$

$$\text{Esto es } \alpha_l = \frac{b_l - a_l^T x}{a_l^T d}$$

tomando $\bar{\alpha} = \min \{ \alpha_l; l \in I(x), \alpha_l > 0 \}$ se cumple la condición (ii).

Se comienza con un x_0 factible y dado x_k se determina:

$$d_k = -P\nabla f(x_k)$$

donde P es la proyección sobre el subespacio determinado por las restricciones activas en x_k . Luego se determina α como se indicó

antes y definimos:

$$x_{k+1} = x_k + \alpha d_k$$

donde α_k es el $\min f(x_k + \alpha d_k)$
 $0 \leq \alpha \leq \bar{\alpha}$

Sólo falta determinar la proyección P . Obsérvese que $P_1 d$ es un vector en el espacio generado por los gradientes de las restricciones activas, esto es, existen escalares u_i , tal que:

$$\begin{aligned} P_1 d &= \sum_{i \in I(x)} u_i a_i \\ &= A^T u \end{aligned} \quad \dots \text{IV2.2}$$

donde cada columna de la matriz A^T corresponde a alguna a_i , con $i \in I(x)$.

Sea cualquier $d \in \mathbb{R}^n$, entonces:

$$\begin{aligned} P d &= d - P_1 d \\ &= d - A^T u \end{aligned} \quad \dots \text{IV2.3}$$

Como $A P d = 0$, por definición, multiplicando IV2.3 por A y se obtiene:

$$0 = A d - A A^T u$$

de donde

$$u = (A A^T)^{-1} A d \quad \dots \text{IV2.4}$$

Sustituyendo IV2.4 en IV2.3 se tiene

$$P d = d - A^T (A A^T)^{-1} A d$$

¿Qué sucede si $P \nabla f(x) = 0$? Entonces $-\nabla f(x) + A^T u = 0$

Si $u \geq 0$, x es el mínimo del problema. Si existe $u_i < 0$, es

posible obtener una dirección factible y de descenso, sacando de A la fila a_i^T que representa el u_i más negativo.

Algoritmo

- 1) Escoger x_0 factible $k = 0$
- 2) Para x_k , hallar la matriz A de restricciones activas.
- 3) Calcular $P = I - A^T(AA^T)^{-1}A$ y $d_k = -P\nabla f(x_k)$
- 4) Si $d_k \neq 0$, hallar $\bar{\alpha}$ tal que
 $\bar{\alpha} = \max \{ \alpha > 0, x_k + \alpha d_k \text{ es factible} \}$
 α_k el mín $f(x_k + \alpha d_k)$
 $0 \leq \alpha \leq \bar{\alpha}$
 $x_{k+1} = x_k + \alpha_k d_k$
 $k = k+1$ y regresar a 2
- 5) Si $d_k = 0$, calcular $u = (AA^T)^{-1}A\nabla f(x_k)^T$
 - a) $u > 0$, detenerse $x^* = x_k$
 - b) Existe $u_i < 0$. Sacar de A la fila correspondiente al u_i más negativo y regresar a 3

IV.3 METODO DEL GRADIENTE REDUCIDO

En este tema se comenzará con el caso de restricciones lineales:

$$\begin{aligned} \min f(x) \\ \text{s.a. } Ax = b \\ x \geq 0 \end{aligned} \quad \dots \text{IV3.1}$$

donde $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, A es de $m \times n$ y $m \leq n$ supónse la condición de regularidad, todo conjunto de m columnas de A es linealmente independiente, y toda solución básica (i.e.: hay $n-m$ variables iguales a cero) tiene exactamente m valores mayores que cero. Dado x factible, se escribe a $x = (y, z)$ donde $y \in \mathbb{R}^m$, $z \in \mathbb{R}^{n-m}$ son coordenadas de x y $y > 0$.

El problema IV2.2 se convierte en

$$\begin{aligned} \min f(y, z) \\ \text{s.a. } By + Cz = b \\ y \geq 0, z \geq 0 \end{aligned}$$

donde B es de $m \times m$ y C de $m \times (n - m)$ y se obtiene partiendo las columnas de A en la misma forma que se partió x en (y, z) . Se dice que z son las variables independientes y que y las dependientes ya que

$$y = B^{-1}b - B^{-1}Cz$$

y ahora el problema IV2.2 se reduce a

$$\min f(B^{-1}b - B^{-1}Cz, z) = g(z) \quad \dots \text{IV3.2}$$

salvo que, al resolver este problema, deberá resultar $B^{-1}b - B^{-1}Cz \geq 0$

Al calcular el gradiente reducido de f se tiene

$$r = \nabla_x g(x) = \nabla_y f(y, z) - \nabla_y f(y, z) B^{-1} C$$

considerando

$$(\nabla z)_i = \begin{cases} r_i & \text{si } r_i > 0 \text{ o } z_i > 0 \\ 0 & \text{de otra forma} \end{cases}$$

Obsérvese que si $\Delta_n = 0$, z es solución de IV3.1 si $y \geq 0$. Entonces (y, z) es solución de IV3.1.

Algoritmo del gradiente reducido para IV2.1

- 1) Escoger x_0 factible y poner $s = 0$.
- 2) Partir x_k en dos vectores z y y de variables independientes y dependientes. Partir en forma similar A en B y C.
- 3) Calcular $r = \nabla_x f(y, z) = \nabla_y f(y, z) B^{-1} C$

4) Hacer ∇_z el vector

$$(\nabla z)_i = \begin{cases} r_i & \text{si } r_i > 0 \text{ o } z_i > 0 \\ 0 & \text{de otra forma} \end{cases}$$

- 5) Si $\Delta z = 0$ el vector $x = (y, z)$ es solución del problema IV2.1
- 6) Si $\nabla z \neq 0$, hacer $\nabla y = B^{-1} C \nabla_{zy} d_k = (\nabla_y, \nabla_n)$
- 7) Calcular

$$\beta_1 = \max (\alpha: y + \alpha \nabla_y \geq 0)$$

$$\beta_2 = \max (\alpha: z + \alpha \nabla_z \geq 0)$$

$$\beta_0 = \min (\beta_1, \beta_2)$$

$$\alpha_k \text{ el } \min f(x_k + \alpha d_k) \\ 0 \leq \alpha \leq \beta_3$$

- 8) Si $\beta_3 < \beta_1$, hacer $x_{k+1} = x_k + \alpha_k d_k$, $k = k+1$ y regresar a 2.
- 9) Si $\beta_3 = \beta_1$, hay alguna variable independiente igual a cero, cambiarla por alguna dependiente. Recalcular (y, z) y (B, C) y regresar a 3.

CONCLUSIONES

CONCLUSIONES

La mejor forma de establecer las conclusiones del presente trabajo, es reflejarlas tal y como se estructura este mismo, por lo tanto se darán conclusiones siguiendo el orden de los capítulos y temas.

Definitivamente acerca del Capítulo I se puede afirmar que es el capítulo más difícil en decisión, puesto que éste es el origen donde nacen las directrices a seguir de todo el trabajo, por tal motivo la importancia de establecer perfectamente los objetivos y alcances del mismo, para no caer con un desarrollo largo y tedioso que no estableciera los pilares de la teoría para el entendimiento de los algoritmos a tratar. Sin duda alguna hablar de optimización clásica por sí sola podría ser un tópico único a desarrollar como tema de tesis, de ahí la importancia de sólo seleccionar los temas que nutrieran a quienes elaboran este trabajo de los fundamentos y conocimientos para desarrollar los algoritmos que se implementaran en el capítulo siguiente, siendo estos temas las condiciones necesarias y suficientes de primer y segundo orden, así un análisis de la funciones convexas que establecen nuestros los conocimientos basados en el cálculo, geometría, álgebra matricial y topología. Los corolarios fueron reafirmados y posteriormente para lograr una comprensión de tales conceptos. citamos un tema de aplicaciones.

Resumiendo, el objetivo de establecer las bases teoricas de la programación no lineal en este capítulo fue cubierto.

Referente al Capítulo II este es considerado la columna vertebral de todo el trabajo, puesto que se desarrollan todos los algoritmos en este capítulo. La forma de presentar el análisis se establece desde el planteamiento del origen del problema, la generación de teoremas y simultáneamente su demostración, la presentación del algoritmo y la secuencia de pasos para converger al punto óptimo. Se pretende abordar la solución a tres tipos de funciones: funciones unimodales de una sola variable, funciones multimodales de una sola variable y funciones de varias variables, todos los problemas no restringidos, a su vez en cada tema se citan los algoritmos mas populares por su elegancia teórica, por su principio intuitivo a la convergencia del óptimo, por sus propiedades de convergencia, y por su técnica de búsqueda, por lo que se consideró que los algoritmos citados en este trabajo son los representativos en cada método, se presentan tanto sus ventajas como sus desventajas, se hizo un análisis de su técnica de convergencia hacia el óptimo, todo desde el punto de vista teórico, y también los objetivos fueron satisfechos.

El capítulo III es una sección cien por ciento de aplicación puesto que en este se presentan los programas de los algoritmos analizados en el capítulo II, el objetivo en este capítulo es generar programas en lenguaje de fácil y común comprensión y definitivamente el BASIC cumple con todos estos requerimientos ya que es un lenguaje muy simple de fácil acceso, puesto que cualquier computadora personal compatible con IBM lo tiene, y tiene un buen manejo tanto de variables vectoriales y matriciales, y lo anterior complementado con una definición de las variables principales y el diagrama de flujo de cada uno de los algoritmos para una comprensión total del mismo.

Todo lo anterior se enfocó a problemas no restringidos de alguna forma pudo finalizarse este trabajo, pero para que fuera completo se presentó el análisis de tres algoritmos de problemas no restringidos, que si no va más a fondo sobre éstos, puesto que cada uno de ellos por sí solo pudieran ser un tema para tesis.

Se concluye en términos generales que todavía hay mucho por hacer en la programación no lineal, debido a la complejidad de la misma, a la falta de interés de su aplicación y a la poca información que existe, pero se espera que con este trabajo se aporte un granito de arena para despertar inquietudes a las futuras generaciones de crear un interés sobre estos temas y de esa forma motivar a realizar estudios más profundos o investigaciones de nuevos algoritmos y éstos sean aplicados para resolver problemas de la vida diaria.

APENDICE TECNICO A

$$L_2 = \frac{F_{n-1}}{F_n} L_1 + \frac{(-1)^n \epsilon}{F_n}$$

Demostración:

Aplicando la definición generalizada de los intervalos vista en el Capítulo II (a), siendo ésta la siguiente:

$$L_{n-j} = F_{j+1} L_n - F_{j+1} \epsilon, \quad j = 1, 2, \dots, n-1$$

Obtenemos los valores L_2 y L_1 :

$$j=n-1, \quad L_{n-(n-1)} = F_{(n-1)+1} L_n - F_{(n-1)+1} \epsilon$$

$$L_1 = F_n L_n - F_{n-1} \epsilon$$

$$j=n-2, \quad L_{n-(n-2)} = F_{(n-2)+1} L_n - F_{(n-2)+1} \epsilon$$

$$L_2 = F_{n-1} L_n - F_{n-2} \epsilon$$

Sustituyendo los valores de L_1 y L_2 en la igualdad a demostrar:

$$(F_{n-1} L_n - F_{n-2} \epsilon) = \frac{F_{n-1}}{F_n} (F_n L_n - F_{n-1} \epsilon) + \frac{(-1)^n \epsilon}{F_n}$$

$$F_{n-1} L_n - F_{n-2} \epsilon = F_{n-1} L_n - \frac{F_{n-1} F_{n-1} \epsilon}{F_n} + \frac{(-1)^n \epsilon}{F_n}$$

$$F_{n-2} \epsilon - \frac{F_{n-1} F_{n-1} \epsilon}{F_n} + \frac{(-1)^n \epsilon}{F_n} = 0$$

Sacando término común a ϵ :

$$e \left(F_{n-3} - \frac{F_{n-1} F_{n-2}}{F_n} + \frac{(-1)^n}{F_n} \right) = 0$$

$e \cong 0$ pero $e \neq 0$ por lo tanto

$$F_{n-3} - \frac{F_{n-1} F_{n-2}}{F_n} + \frac{(-1)^n}{F_n} = 0$$

Multiplicando la igualdad por F_n :

$$F_n F_{n-3} - F_{n-1} F_{n-2} + (-1)^n = 0 \dots\dots \text{Al}$$

Por lo tanto nuestra demostración se reduce a demostrar Al; la cual la vamos a demostrarla por inducción matemática:

1) Por demostrar el primer caso factible:
 $n = 3$

$$F_{(3)} F_{(3)-3} - F_{(3)-1} F_{(3)-2} + (-1)^{(3)} = 0$$

$$F_3 F_0 - F_2 F_1 + (-1)^3 = 0$$

$$(3)(1) - (2)(1) - 1 = 3 - 2 - 1 = 0 \quad \text{q.e.d.}$$

2) Por hipótesis se considera verdadero el caso $n = k$:

$$F_k F_{k-3} - F_{k-1} F_{k-2} + (-1)^k = 0$$

3) Por demostrar el caso $n = k + 1$:

$$F_{(k+1)} F_{(k+1)-3} - F_{(k+1)-1} F_{(k+1)-2} + (-1)^{(k+1)} = 0$$

$$F_{k+1} F_{k-2} - F_k F_{k-1} + (-1)^{k+1} = (F_k + F_{k-1}) F_{k-2} - (F_{k-2} + F_{k-3}) F_k + (-1)^{k+1}$$

$$F_k F_{k-2} + F_{k-1} F_{k-2} - F_{k-2} F_k - F_{k-3} F_k + (-1)^{k+1} = F_{k+1} F_{k-2} - F_{k-3} F_k + (-1)^{k+1}$$

Dividiendo entre -1 :

$$F_k F_{k-2} - F_{k+1} F_{k-1} + (-1)^k = 0, \quad \text{por hipótesis es verdadero}$$

Por lo tanto :

$$F_{(k+1)} F_{(k+1)-2} - F_{(k+1)-1} F_{(k+1)-1} + (-1)^{(k+1)} = 0 \quad \text{q.e.d.}$$

Concluyendo que nuestra igualdad :

$$L_2 = \frac{F_{n-1}}{F_n} L_1 + \frac{(-1)^n}{F_n} \quad \text{q. e. d.}$$

APENDICE TECNICO B

$$\frac{G_p + z + w}{G_p + G_q + 2z} = \frac{z + w - G_p}{G_q - G_p + 2w}$$

$$(G_p + z + w)(G_q - G_p + 2w) = (G_p + G_q + 2z)(z + w - G_p)$$

$$G_p G_q - G_p^2 + 2w G_p + z G_q - z G_p + 2zw + w G_q - w G_p + 2w^2 =$$

$$z G_p + w G_p - G_p^2 + z G_q + w G_q - G_p G_q - 2z^2 + 2zw - 2z G_p$$

Reduciendo términos:

$$2G_p G_q + 2w^2 - 2z^2 = 0$$

$$\text{como } w = (z^2 - G_p G_q)^{\frac{1}{2}}$$

$$2G_p G_q + 2(z^2 - G_p G_q) - 2z^2 = 0$$

$$2G_p G_q + 2z^2 - 2G_p G_q - 2z^2 = 0$$

a.e.d.

APENDICE TECNICO C

$$\frac{1}{2} \frac{(x_2^2 - x_3^2)f(x_1) + (x_3^2 - x_1^2)f(x_2) + (x_1^2 - x_2^2)f(x_3)}{(x_2 - x_3)f(x_1) + (x_3 - x_1)f(x_2) + (x_1 - x_2)f(x_3)}$$

$$\frac{1}{2} \frac{(x_1 + x_2) \left[f(x_1) - f(x_2) \right] (x_2 - x_3)(x_3 - x_1)}{(x_2 - x_3)f(x_1) + (x_3 - x_1)f(x_2) + (x_1 - x_2)f(x_3)}$$

La igualdad anterior es equivalente a

$$A = (x_2^2 - x_3^2)f(x_1) + (x_3^2 - x_1^2)f(x_2) + (x_1^2 - x_2^2)f(x_3)$$

$$B = (x_1 + x_2) \left[(x_2 - x_3)f(x_1) + (x_3 - x_1)f(x_2) + (x_1 - x_2)f(x_3) \right] \left[f(x_1) - f(x_2) \right] (x_2 - x_3)(x_3 - x_1)$$

Desarrollando B llegamos a la expresion:

$$x_2^2 f(x_1) - x_3^2 f(x_2) + x_1^2 f(x_3) - x_2^2 f(x_3) - x_3^2 f(x_1) + x_1^2 f(x_2)$$

que es la misma que A.

APENDICE TECNICO D

Demostración:

(Por reducción al absurdo)

Sean d_1, d_2, \dots, d_n son vectores linealmente dependientes.

Por lo tanto existe al menos un vector que puede generarse como una combinación de los vectores restantes y supongamos que es el vector d_r , entonces:

$$d_r = c_1 d_1 + c_2 d_2 + \dots + c_{r-1} d_{r-1} + c_{r+1} d_{r+1} + \dots + c_n d_n \quad \text{ó}$$

$$d_r = \sum_{k=1}^n c_k d_k \quad \text{con } k \neq r \text{ y al menos algún } c_k \text{ diferente de cero}$$

Por hipótesis, el conjunto de vectores d_1, d_2, \dots, d_n son Q -ortogonales por lo tanto:

$$d_i Q d_j = 0 \quad \forall i, j$$

Se elije la pareja entre el vector d_r y un d_i con $c_i \neq 0$

$$d_i Q d_r = 0 \Rightarrow d_i Q \left(\sum_{k=1}^n c_k d_k, k \neq r \right) = 0$$

$$d_i Q (c_1 d_1 + c_2 d_2 + \dots + c_{r-1} d_{r-1} + c_{r+1} d_{r+1} + \dots + c_n d_n) = 0$$

Aplicando la propiedad distributiva en la suma de matrices:

$$d_i Q (c_1 d_1) + d_i Q (c_2 d_2) + \dots + d_i Q (c_i d_i) + \dots + d_i Q (c_n d_n) = 0$$

Propiedad asociativa con los múltiplos de matrices:

$$c_1 d_i Q d_1 + c_2 d_i Q d_2 + \dots + c_i d_i Q d_i + \dots + c_n d_i Q d_n = 0$$

Todos los términos de la izquierda son cero por ser Q -ortogonales exceptuando el término

$$c_i d_i Q d_i \neq 0$$

$c_i \neq 0$ por ser uno de los coeficientes de los d_i 's que generan a d_r , y al menos existe alguno que no es cero.

$d_i Q d_i \neq 0$ por ser Q definida positiva y d_i no nulo.

$\Rightarrow d_i Q d_r \neq 0$ lo que contradice la hipótesis de que los vectores son Q -ortogonales.

Por reducción al absurdo

$\therefore d_1, d_2, \dots, d_r, \dots, d_n$ deben ser linealmente independientes

q.e.d.

BIBLIOGRAFIA

- Programación BASIC para microcomputadoras Joyanes Aguilar Luis, McGraw Hill, México (1986).
- MS GW-BASIC intérprete bajo MS-DOS Guía del Usuario, Olivetti Personal Computer Hispano Olivetti, Barcelona España (1987).
- Nonlinear Optimization Jackobs, Muwickz, Pizzo Pretince Hall
- Mathematical Programming Jeremy F. Shapiro John Wiley, New York (1979).
- Optimality in Nonlinear Programming: a feasible directions approach. Ben-Israel, Ben-Tal, S. Zibec John Wiley, New York (1981).
- Numerical Methods for Constrained Optimization P.E Gill and W. Murray Academic Press, London (1974).
- Basic Optimization Methods Brian D. Bunday Edward Arnold School of Mathematical Sciences, University of Bradford (1984).
- Computer methods for Mathematical Computations George E. Forsythe, Michael A. Malcom, Cleve B. Moler Prentice Hall
- Nonlinear Optimization Dixon L.C.W. English University Press, London(1972)
- Nonlinear Programming Bazaraa, M.S., and Shetty, C. M (1979) John Wiley, New York
- Constrained Optimization Using a Nondifferentiable Penalty Function John A. R. SIAM J. (1973)