

I do hate sums. There is no greater mistake than to call arithmetic an exact science. There are ... hidden laws of number which it requires a mind like mine to perceive. For instance, if you add a sum from the bottom up, and then again from the top down, the result is always different.

MRS. LA TOUCHE (siglo XIX)



LAS TECNICAS DE PROGRAMACION MATEMATICA
EN OPTIMIZACION ESTRUCTURAL

CAPITULO III
LAS TECNICAS DE PROGRAMACION MATEMATICA
EN OPTIMIZACION ESTRUCTURAL

III.1 INTRODUCCION

Como se ha puesto de manifiesto en el capítulo precedente, la concepción moderna de las técnicas de optimización estructural se desarrolla a partir de los planteamientos de Klein [1955a], quien considera que la mayor parte de los problemas de optimización estructural pueden ser formulados matemáticamente en la forma estándar:

$$\begin{aligned} \text{minimizar: } & f(\bar{x}) & ; \bar{x} = \{x_i\} & ; i = 1, \dots, n \\ \text{verificando: } & g_j(\bar{x}) \leq 0 & ; j = 1, \dots, m \end{aligned} \quad (3.1)$$

A lo largo de la exposición siguiente observaremos cómo técnicas de optimización estructural aparentemente alejadas de este enfoque, tales como el diseño a máxima tensión o los criterios de optimalidad, pueden ser consideradas como particularizaciones del planteamiento general anterior.

Las técnicas que permiten resolver el problema (3.1) se enmarcan en el área de conocimiento conocida como PROGRAMACION MATEMATICA. Si bien la definición de este término es un tanto confusa, y existen discrepancias entre diferentes autores acerca

de su utilización, en lo sucesivo denominaremos programación matemática al conjunto de métodos que permiten analizar y resolver problemas de minimización de funciones reales de varias variables.

La programación matemática abarca un gran número de técnicas para la solución de diversas familias de problemas, algunas de los cuales exceden claramente las pretensiones de este estudio, y cuya clasificación es compleja. Por este motivo nos limitaremos al análisis de los métodos de solución de aquellos problemas que pueden formularse en la forma estándar (3.1), o en la forma general equivalente (3.2), más adelante expuesta, en los cuales el rango de definición de las variables es continuo y las funciones que definen el problema son diferenciables, ya que consideramos que esta formulación cubre el porcentaje más significativo de los problemas de optimización estructural.

En esta exposición pretendemos sintetizar de forma racional y estructurada el conjunto de tales técnicas, y para un estudio más detallado nos remitimos a tratados monográficos de optimización y optimización estructural [Cea 1971, Himmelblau 1972, Martos 1975, Fletcher 1981, Morris 1982, Bertsekas 1982, Scales 1985]. Igualmente omitiremos las referencias en las observaciones no demostradas que se expongan, ya que pueden encontrarse en los textos anteriores.

Este enfoque, por tanto, no contempla formulaciones propias de la PROGRAMACION ENTERA, en las cuales el rango de definición de las variables es discreto, ni de la PROGRAMACION NO DIFERENCIABLE, cuyo ámbito de aplicación a la optimización

estructural es más restrictivo, aunque de indudable interés (problemas de selección de materiales, etc.). Otros planteamientos de interés residen en los ámbitos de la OPTIMIZACION MULTIOBJETIVO, la PROGRAMACION GEOMETRICA, la PROGRAMACION DINAMICA y el CONTROL OPTIMO, para cuyo estudio nos remitimos a tratados especializados [Fletcher 1981, Sawaragi, Nakayama y Tanino 1985]

Dividiremos la exposición en los cuatro apartados siguientes:

- Definiciones.
- Minimización no restringida.
- Minimización con restricciones de igualdad.
- Minimización con restricciones de tipo general.

III.2 DEFINICIONES

Denominaremos problema general de programación matemática al problema de minimización de una función real de varias variables con restricciones de tipo general, expresadas como un sistema de ecuaciones e inecuaciones que condicionan la solución. El problema general, por tanto, puede escribirse en la forma:

$$\begin{aligned}
 \text{minimizar} & : f(\bar{x}) & ; \bar{x} = \{x_i\} & ; i=1, \dots, n \\
 \text{dados} & : \bar{g}(\bar{x}) = \{g_j(\bar{x})\} & & ; j=1, \dots, m \\
 & \bar{h}(\bar{x}) = \{h_l(\bar{x})\} & & ; l=1, \dots, p \\
 \text{verificando:} & g_j(\bar{x}) \leq 0 & & ; j=1, \dots, m \\
 & h_l(\bar{x}) = 0 & & ; l=1, \dots, p \\
 & a_i \leq x_i \leq b_i & & ; i=1, \dots, n
 \end{aligned} \tag{3.2}$$

donde admitiremos que las funciones "f", "g_j", y "h_l" son dos veces diferenciables, y las "h_l" funcionalmente independientes.

Denominaremos VARIABLES DE DISEÑO al conjunto " \bar{x} " de "n" variables reales del problema, aunque esta denominación solo tenga sentido en el contexto de la optimización del diseño, y FUNCION OBJETIVO a la función real "f(\bar{x})" a minimizar. Se denomina ESPACIO DE DISEÑO al espacio n-dimensional formado por las variables de diseño.

Las inecuaciones "g" reciben el nombre de RESTRICCIONES EN DESIGUALDAD, y las ecuaciones "h" se denominan RESTRICCIONES DE IGUALDAD. Las desigualdades que conforman el último conjunto de

condiciones reciben el nombre de RESTRICCIONES LATERALES.

Un punto " \bar{x} " del espacio de diseño se dice FACTIBLE si en él se verifican todas las restricciones impuestas, y no factible en caso contrario. Se denomina REGION FACTIBLE al subconjunto de puntos factibles del espacio de diseño.

Una restricción se denomina VIOLADA si no se cumple en un determinado punto no-factible. Una restricción en desigualdad no violada en un punto dado del espacio de diseño se dice ACTIVA si se cumple estrictamente la relación de igualdad, e INACTIVA en caso contrario.

Si existen restricciones laterales para una determinada variable, ésta se dice LIBRE si su valor se encuentra dentro del rango permitido de variación, pero no coincide con ninguno de sus extremos. Por contra, la variable se dirá FIJA si su valor coincide estrictamente con el máximo o el mínimo de su rango de variación. Numerosos autores utilizan los términos variable activa y variable pasiva en lugar de variable libre y variable fija. Esta terminología goza de amplia difusión. No obstante puede dar lugar a confusiones, ya que de alguna forma entra en contradicción con los términos introducidos previamente en relación con las restricciones en desigualdad. Por ello no será empleada en esta exposición

Existen diversas transformaciones elementales que permiten modificar la formulación del problema general, dando lugar a diversos problemas equivalentes. Entre las más significativas se cuentan las siguientes:

- Un problema de minimización puede transformarse en un problema de maximización, y viceversa, puesto que minimizar una función equivale a maximizar su opuesta; esto es:

$$\text{minimizar: } f(\bar{x}) \quad \Leftrightarrow \quad \text{maximizar: } -f(\bar{x}) \quad (3.3)$$

- Una restricción de igualdad puede sustituirse por dos restricciones en desigualdad en la forma:

$$h(\bar{x}) = 0 \quad \Leftrightarrow \quad \begin{cases} h(\bar{x}) \leq 0 \\ -h(\bar{x}) \leq 0 \end{cases} \quad (3.4)$$

- Una restricción en desigualdad puede transformarse en una restricción de igualdad, si bien para ello es preciso introducir una nueva variable en el problema, conocida como variable de separación o variable muda. La transformación puede escribirse en la forma:

$$\begin{aligned} g(\bar{x}) \leq 0 \quad \Leftrightarrow \quad g(\bar{x}) + P(\beta) = 0 \\ \text{siendo: } P(\beta) \geq 0 \quad \forall \beta \in \mathbb{R} \end{aligned} \quad (3.5)$$

siendo elecciones habituales para la función no negativa "P(β)" la función cuadrado y la exponencial.

Empleando adecuadamente las transformaciones anteriores, puede obtenerse la formulación estándar (3.1) a partir de la formulación general (3.2), y viceversa. Ambas formulaciones, por tanto, son equivalentes, y pueden derivarse con sencillez otras formas alternativas.

Desde esta perspectiva general, el problema de minimización con restricciones de igualdad y el problema de minimización no restringida no son sino particularizaciones de la formulación (3.2).

Un punto factible se dice MINIMO LOCAL o MINIMO RELATIVO del problema general (3.2) si, de entre todos los puntos factibles en un cierto entorno, la función objetivo adquiere su menor valor en dicho punto. Un minimo local se dice MINIMO GLOBAL, MINIMO ABSOLUTO, u OPTIMO del problema general (3.2), si en él la función objetivo toma un valor menor que en cualquier otro punto de la región factible.

No es sencillo, en general, probar la existencia y unicidad de solución para un determinado problema de minimización del tipo (3.2). La obtención de la solución, si existe, no suele ser trivial, y es preciso recurrir a algoritmos numéricos de tipo iterativo para abordar su cálculo. Es importante consignar que los algoritmos desarrollados para resolver el problema (3.2), tratan de obtener -en general- mínimos locales, y por tanto siempre existirá una cierta incertidumbre en cuanto a la globalidad de la solución obtenida mediante su aplicación.

Se han desarrollado numerosos algoritmos de diversa complejidad y eficiencia para abordar el problema, basándose la mayor parte de los mismos en ciertas hipótesis que son válidas exclusivamente para una determinada variedad de problemas que denominaremos PROBLEMAS CONVEXOS. Un problema del tipo (3.2) se dice convexo, si su región factible es un CONJUNTO CONVEXO, y su función objetivo es una FUNCION CONVEXA, conceptos éstos que

definimos a continuación y de los cuales se ofrece una interpretación gráfica (Fig. 3.1):

Un conjunto $A \subset \mathbb{R}^n$ se dice CONVEXO, si se verifica:

$$\forall \bar{a}, \bar{b} \in A \quad \forall t \in [0,1] \text{ se cumple,} \quad (3.6)$$

$$X = \{ \bar{x} \mid \bar{x} = (1-t)\bar{a} + t\bar{b} \} \subset A$$

Una función $f(\bar{x})$ se dice CONVEXA en un conjunto convexo A , si se verifica:

$$\forall \bar{a}, \bar{b} \in A \quad \forall t \in [0,1] \text{ se cumple,} \quad (3.7)$$

$$f(\bar{x}) \leq (1-t)f(\bar{a}) + tf(\bar{b})$$

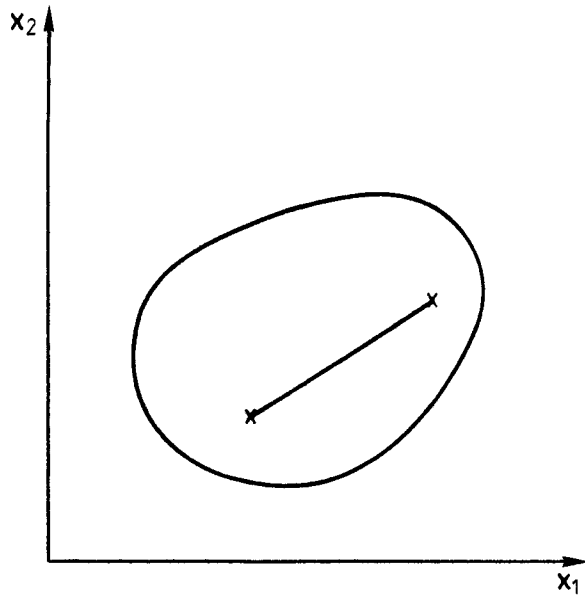
con : $\bar{x} = (1-t)\bar{a} + t\bar{b}$

Una función se dice cóncava, si su opuesta es convexa. Obviamente, el espacio " \mathbb{R}^n " es un conjunto convexo. En lo sucesivo hablaremos de funciones convexas -sin especificar el conjunto en que se verifica esta condición- para referirnos a funciones convexas en su campo de definición, siendo éste el espacio " \mathbb{R}^n " o un subconjunto convexo del mismo.

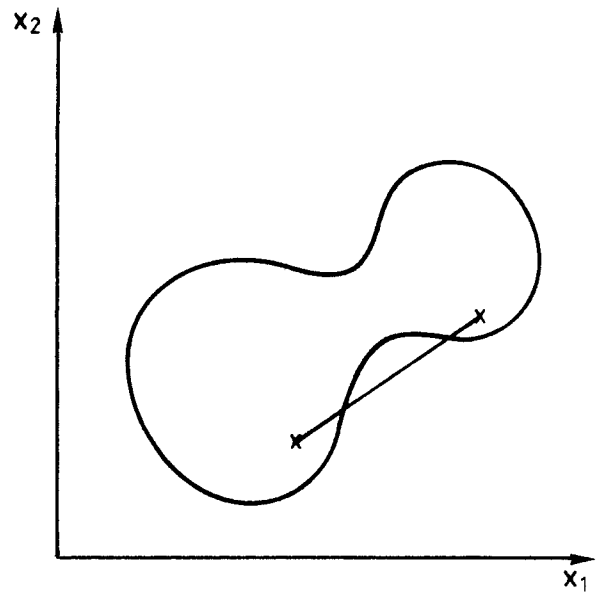
Si la función " $f(\bar{x})$ " es diferenciable, es condición necesaria y suficiente de convexidad en un conjunto " A " convexo la siguiente:

$$\forall \bar{x}, \bar{y} \in A, \quad f(\bar{y}) \geq f(\bar{x}) + \nabla f(\bar{x}) (\bar{y} - \bar{x}) \quad (3.8)$$

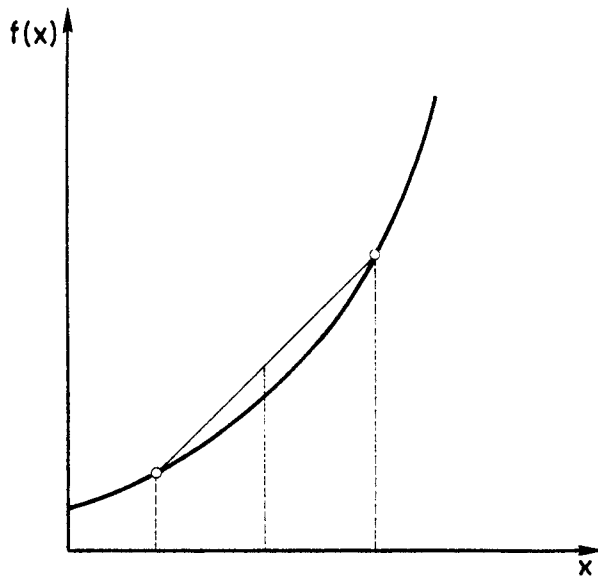
Si la función " $f(\bar{x})$ " es dos veces diferenciable, es condición necesaria y suficiente de convexidad en un conjunto " A " convexo, el que su hessiano sea semidefinido positivo en todo punto del



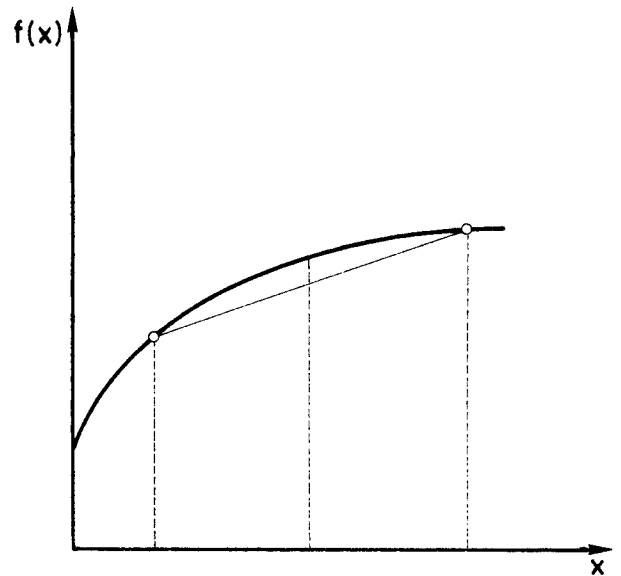
a)



b)



c)



d)

Figura 3.1.- Interpretación gráfica de diversos conceptos relativos a la convexidad.

- a) Conjunto convexo.
- b) Conjunto cóncavo.
- c) Función convexa.
- d) Función cóncava.

conjunto dado.

Si las funciones " g_j " que definen las restricciones en desigualdad del problema (3.2) son convexas, y las restricciones de igualdad son lineales, la región factible es necesariamente un conjunto convexo. Si la función objetivo es igualmente convexa, el problema se denomina problema general de programación convexa.

El concepto de convexidad adquiere una importancia fundamental en el ámbito de la programación matemática, puesto que en el estado del conocimiento actual, las demostraciones de existencia y unicidad de solución para los problemas de minimización, así como los algoritmos de solución del problema, son válidos en esencia exclusivamente para problemas convexas.

III.3 MINIMIZACIÓN NO RESTRINGIDA

El problema general de minimización no restringida puede escribirse en la forma:

$$\text{minimizar: } f(\bar{x}) \quad ; \bar{x} = \{x_i\} \quad ; i= 1, \dots, n \quad (3.9)$$

III.3.1 Condición de mínimo

La condición necesaria de extremo para el problema de minimización no restringida se expresa como:

$$\frac{\partial}{\partial \bar{x}} f(\bar{x}) = \bar{0} \quad (3.10)$$

es decir, que se anule el gradiente de la función objetivo.

Evidentemente, la condición anterior no es suficiente para asegurar la existencia de un mínimo. Para obtener condiciones suficientes de mínimo es necesario emplear derivadas de orden superior y en la práctica, con frecuencia, estas derivadas no son conocidas. Una condición suficiente de mínimo local, adicional a la condición (3.10) para este problema, puede escribirse en la forma:

$$\frac{\partial^2}{\partial \bar{x}^2} H(\bar{x}) \quad \bar{v} \geq 0 \quad ; \quad \forall \bar{v} \in R^n \quad (3.11)$$

es decir, que el hessiano " H " de la función objetivo sea semidefinido positivo. Si la función objetivo es convexa, la condición (3.11) se cumple necesariamente, y la condición (3.10) es entonces necesaria y suficiente de mínimo global del problema.

III.3.2 Métodos de solución del problema

La solución del problema general de minimización no restringida (3.9) puede abordarse desde dos perspectivas; analítica o numérica.

La solución analítica del problema se plantea resolviendo el sistema de "n" ecuaciones con "n" incógnitas (3.10). Este sistema será en general no lineal, y será preciso obtener todas sus soluciones. Posteriormente es necesario discriminar los mínimos relativos del problema entre todas las soluciones obtenidas, habitualmente comprobando si se verifican condiciones del tipo (3.11). Si la función objetivo está acotada inferiormente, entre los mínimos relativos se encuentra el mínimo absoluto, que puede hallarse con sencillez comparando los valores de la función objetivo.

Evidentemente este procedimiento no es aplicable a muchos problemas habituales, puesto que no siempre es posible conocer una expresión del gradiente de la función objetivo, y aún en el caso de que éste fuese conocido, no existe ningún procedimiento general que permita obtener analíticamente todas las soluciones de un sistema no lineal de ecuaciones. Además, aunque en muchas ocasiones es posible averiguar si la función está acotada inferiormente, para verificar si se cumplen condiciones de mínimo del tipo (3.11) es necesario conocer derivadas de orden superior de la función, de las cuales tampoco se dispone en general, y por tanto es difícil discriminar los mínimos relativos del resto de las soluciones.

Numéricamente, la solución del problema se plantea aplicando métodos iterativos para la solución de sistemas de ecuaciones no lineales (Newton-Raphson, Newton, etc.) al sistema (3.10), o mediante métodos iterativos que suelen denominarse métodos de descenso. Ambos enfoques, si bien conceptualmente distintos, son formalmente idénticos, y es importante resaltar la íntima relación existente entre las técnicas de minimización no restringida y las técnicas de solución de sistemas de ecuaciones no lineales, con su consiguiente similitud. Sin embargo los métodos de descenso implican conceptos fructíferos y clarificadores, característicos de los problemas de minimización, motivo por el cual realizaremos la exposición según este punto de vista.

El método general de descenso puede escribirse en la forma:

$$\begin{array}{l}
 \text{dado:} \quad \bar{x}^k \\
 \text{obtener:} \quad \bar{x}^{k+1} = \bar{x}^k + \theta \bar{s}^k \\
 \text{que verifique: } f(\bar{x}^{k+1}) < f(\bar{x}^k)
 \end{array} \tag{3.12}$$

El método de descenso consta de dos partes claramente diferenciadas:

- Obtener la dirección de descenso " \bar{s} ".
- Obtener el factor de avance " θ ".

que analizaremos separadamente. La obtención del factor de avance para una dirección de descenso conocida, es evidentemente un problema de minimización no restringida de funciones de una sola variable, que denominaremos minimización unidimensional o unidireccional, y que en la literatura inglesa suele conocerse

con el nombre de "line-search".

III.3.2.1 Métodos de obtención del factor de avance

Una vez conocida la dirección de descenso " \bar{s} " en cada iteración, tal y como se define en el algoritmo general de descenso (3.12), la obtención del factor de avance se plantea como un problema de minimización unidimensional, esto es:

$$\text{minimizar: } \varphi(\theta) = f(\bar{x} + \theta \bar{s}) \quad (3.13)$$

La condición necesaria de mínimo para el problema anterior se expresa como la anulación de la derivada de la función " $\varphi(\theta)$ ". Desarrollando esta condición se obtiene:

$$d\varphi(\theta)/d\theta = \bar{\nabla}f(\bar{x} + \theta \bar{s}) \cdot \bar{s} = 0 \quad (3.14)$$

condición que implica la solución de una ecuación de una variable (Fig. 3.2-a) que puede abordarse mediante numerosos métodos ampliamente difundidos. Entre los más comúnmente empleados se encuentran los bien conocidos métodos de Newton -si es posible calcular la función " $\varphi(\theta)$ " y sus derivadas primeras y segundas punto a punto y la raíz de la ecuación (3.14) es simple-, y los métodos de Whittaker y Regula-Falsi en caso de que las derivadas segundas sean desconocidas o la raíz sea de orden superior. Otros métodos aplicables son los de bisección, interpolación cuadrática, e interpolación cúbica.

En ciertos problemas de minimización pueden existir restricciones en el rango de variación del factor de avance " θ ",

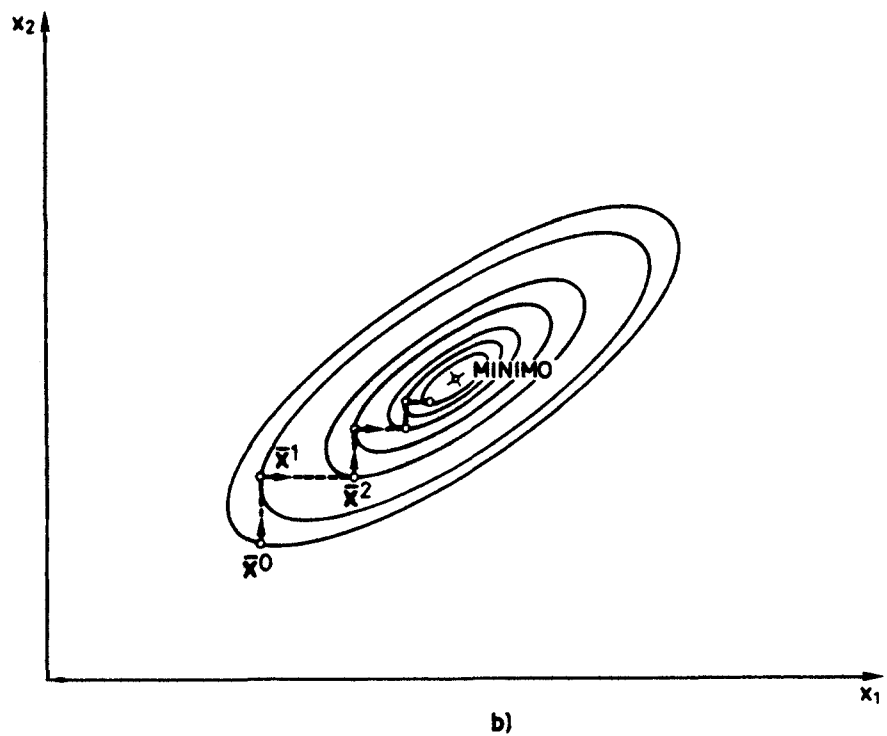
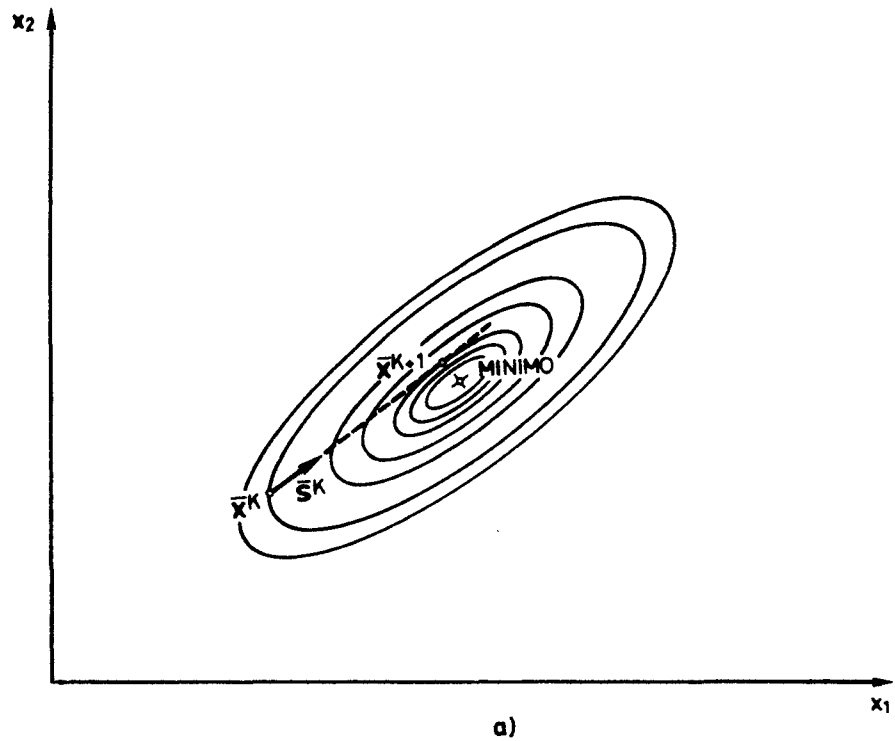


Figura 3.2.- Interpretación gráfica de la condición de obtención del factor de avance en problemas de minimización no condicionada y del método de máximo descenso.

a) El factor de avance se calcula de forma que la dirección de avance sea tangente a la línea de nivel (en general la hiper superficie de nivel) de la función objetivo.

b) En el método de máximo descenso, cada dirección de avance es ortogonal a la precedente, produciéndose frecuentemente fenómenos de zig-zag y una baja velocidad de convergencia.

que en su forma más frecuente, acotan inferior o superiormente su rango de variación. Evidentemente la dificultad introducida por estas restricciones es mínima, y bastará considerar que en este caso la función " $\varphi(\theta)$ " puede alcanzar su mínimo en el extremo inferior o superior de su rango de variación; en este supuesto, no es condición necesaria la anulación de su derivada en el mínimo.

III.3.2.2 Métodos de obtención de la dirección de descenso

Para la obtención de la dirección de descenso existen varias familias de métodos que podemos clasificar en dos grandes grupos:

- Métodos de Newton.
- Métodos derivados de los métodos de Newton.

III.3.2.2.1 Métodos de Newton

Los métodos de Newton parten del desarrollo de Taylor de la función objetivo hasta un cierto término. Si bien para problemas de minimización unidimensional pueden derivarse métodos de cualquier orden con relativa simplicidad, ello no es factible en general para problemas generales n-dimensionales dada la complejidad de los desarrollos. Por lo tanto contemplaremos únicamente dos tipos de métodos:

- Método de primer orden, o aproximación lineal.
- Método de segundo orden, o aproximación cuadrática.

Método de primer orden.

Desarrollando la función objetivo en serie de Taylor hasta el primer término obtenemos:

$$f(\bar{x} + \theta \bar{s}) \approx f(\bar{x}) + \theta \nabla f(\bar{x}) \bar{s} \quad (3.15)$$

El desarrollo anterior, indica que la disminución en el valor de la función objetivo es más acusada, en primera aproximación, cuanto menor es el producto escalar de su gradiente con el vector que define la dirección de descenso " \bar{s} ". Siguiendo este razonamiento, es obvio que el vector " \bar{s} " más favorable es paralelo al gradiente de la función objetivo y de sentido contrario. Sin embargo la aproximación no está acotada inferiormente, y por tanto no proporciona información alguna sobre el módulo del vector " \bar{s} ", que puede ser arbitrario dado que posteriormente se obtendrá un factor de avance mediante minimización unidireccional.

Utilizando esta aproximación deducimos, por tanto, que la dirección de descenso más favorable es:

$$\bar{s} = - \nabla f(\bar{x}) \quad (3.16)$$

recibiendo este algoritmo la denominación de método de máximo descenso, o "steepest descent method", por motivos obvios.

Este método tiene un orden de convergencia lineal. Por ello, aunque en sucesivas iteraciones se obtienen valores descendentes de la función objetivo, el número de iteraciones que es necesario realizar hasta obtener una solución suficientemente

aproximada a la óptima es habitualmente muy elevado, y en muchos tipos de problemas la naturaleza de la función objetivo y las imprecisiones de cálculo dan lugar a diversas anomalías (fenómenos de zig-zag, Fig. 3.2-b) que reducen fuertemente el progreso obtenido en cada iteración. Ello indica que la dirección de máximo descenso no es normalmente la más eficiente. Efectivamente, los métodos de segundo orden proporcionan en general un orden de convergencia mayor, si bien a costa de condiciones de convergencia más restrictivas.

Método de segundo orden.

Desarrollando la función objetivo en serie de Taylor hasta el segundo término obtenemos:

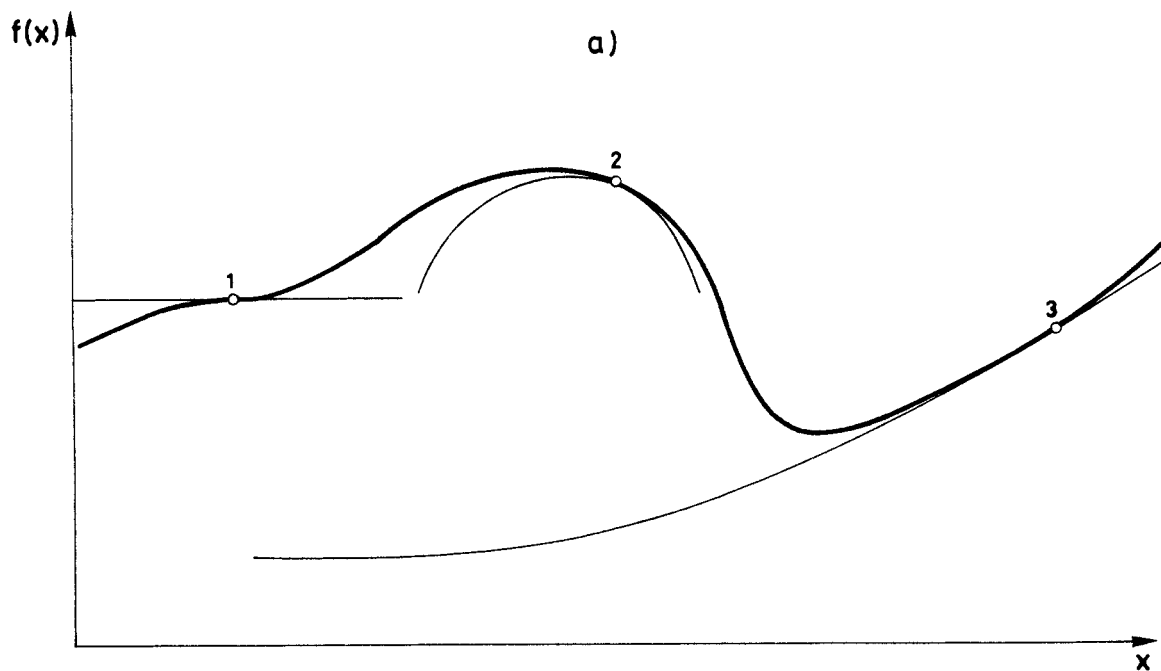
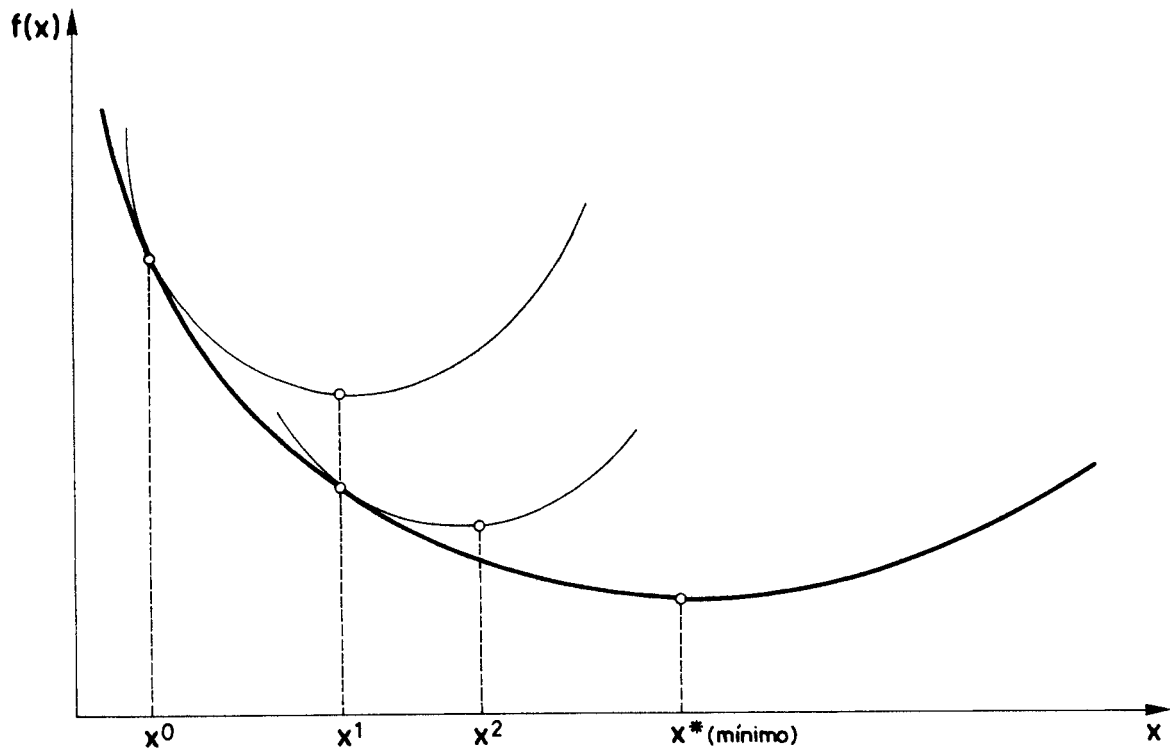
siendo: $\bar{r} = \theta \bar{s}$

$$\phi(\bar{r}) = f(\bar{x} + \bar{r}) \approx f(\bar{x}) + \nabla f(\bar{x}) \cdot \bar{r} + 1/2 \bar{r}^t \underline{H}(\bar{x}) \cdot \bar{r} \quad (3.17)$$

Si el hessiano " \underline{H} " de la función objetivo es definido positivo, la aproximación cuadrática (3.17) tiene un mínimo, y éste ha de cumplir la condición de extremo (3.10), es decir, el gradiente ha de ser nulo. Planteando esta condición, obtenemos:

$$\nabla \phi(\bar{r}) \approx \nabla f(\bar{x}) + \underline{H}(\bar{x}) \cdot \bar{r} = 0 \quad (3.18)$$

y por tanto, el factor de avance y la dirección de descenso más favorables para esta aproximación se obtienen simultáneamente (Fig. 3.3-a), siendo:



b)

Figura 3.3.- Interpretación gráfica del método de Newton de segundo orden y modos de fallo más característicos del algoritmo.

a) Las variables obtenidas en cada iteración minimizan la aproximación cuadrática a la función objetivo en la iteración anterior.

b) Fallo del algoritmo en los casos en que el hessiano de la función objetivo es singular (1), no definido positivo (2), o el factor de avance es muy elevado (3).

$$\theta \quad \underline{s} = \underline{r} = - \underline{H}^{-1}(\underline{x}) \cdot \underline{\nabla} f(\underline{x}) \quad (3.19)$$

Este método tiene orden de convergencia cuadrático, y por tanto, es en principio preferible al método de máximo descenso. No obstante, no asegura que el valor de la función objetivo en el nuevo punto sea menor que el anterior. El fallo del algoritmo puede deberse a varias causas (Fig. 3.3-b), a saber:

- que el hessiano sea singular
- que el hessiano sea regular pero no definido positivo
- que el hessiano sea definido positivo, pero el factor de avance sea muy elevado

Por estas causas, en una determinada iteración pueden presentarse varias anomalías, y entre ellas:

- no es posible calcular una dirección.
- la dirección es ortogonal al gradiente.
- la dirección es de ascenso (el producto escalar del vector " \underline{r}^k " con el gradiente de la función objetivo es positivo).
- la dirección es de descenso (el producto escalar del vector " \underline{r}^k " con el gradiente de la función objetivo es negativo), pero para el factor de avance dado la función objetivo resultante es mayor.

Estas anomalías pueden resolverse de diversas maneras. Si en una cierta iteración se presenta una de las dos primeras anomalías, puede tomarse como dirección la de máximo descenso. Si se presenta la tercera, puede invertirse la dirección, y si se presenta la cuarta, puede buscarse un factor de avance distinto, mediante minimización unidireccional conservando la dirección. Existen otras técnicas que intentan resolver las anomalías anteriores sumando al hessiano una matriz diagonal de elementos

positivos, de forma que el hessiano modificado sea definido positivo, y que pueden entenderse como una generalización del método de Levenberg-Marquardt (3.43) que se expondrá posteriormente.

En cualquier caso es indicado realizar una minimización unidimensional en la dirección obtenida, corrigiendo el módulo del vector de avance " \bar{r} ", ya que el proporcionado por el algoritmo solo es estrictamente válido para funciones cuadráticas. El algoritmo resultante suele denominarse método de Newton de segundo orden generalizado.

Si no se realiza esta minimización unidireccional, es buena práctica imponer una acotación superior al módulo del vector " \bar{r} " y reducir su valor si se excede el límite impuesto. De esta forma, no se permite al algoritmo modificar excesivamente las variables del problema en cada iteración, y se evita que el error de truncamiento cometido al aproximar la función a minimizar por su desarrollo cuadrático (3.17) sea excesivo, y el valor de la función objetivo en el nuevo punto sea imprevisible.

III.3.2.2.2 Métodos derivados del método de Newton

El método de Newton de segundo orden requiere el cálculo del hessiano de la función y su inversión para obtener las sucesivas direcciones de descenso. Para evitar una u otra operación se han desarrollado diversas técnicas, dando lugar a varias familias de métodos, que podemos clasificar de la siguiente forma:

- Métodos de direcciones conjugadas.

- Métodos Quasi-Newton.
- Métodos para problemas específicos.

Métodos de direcciones conjugadas

Los métodos de direcciones conjugadas se desarrollan inicialmente para minimización de funciones cuadráticas, aunque es posible utilizarlos para minimizar en general funciones no lineales, aplicándolos de forma iterativa a los desarrollos de Taylor de segundo orden de las funciones a minimizar. En este sentido, son derivados del método de Newton de aproximación cuadrática, resolviendo el sistema de ecuaciones (3.18) mediante un procedimiento alternativo. Previamente a su exposición, es necesario introducir el concepto de vectores conjugados respecto a una matriz simétrica.

Dada una matriz simétrica de dimensión n : A

los vectores $\{s_i\}; i=1, \dots, n$

se dicen conjugados respecto a A , si y solo si: (3.20)

$$s_i^t A s_j = \begin{cases} = 0, & \text{para } i \neq j \\ \neq 0, & \text{para } i=j \end{cases}$$

Además, si la matriz " A " es definida positiva, es sencillo demostrar que todo conjunto de vectores conjugados es linealmente independiente.

Por tanto, la solución de un sistema de ecuaciones lineales con matriz de coeficientes simétrica y definida positiva, puede escribirse como combinación lineal de un conjunto cualquiera de

vectores conjugados. Haciendo uso de esta propiedad, los coeficientes de la combinación lineal pueden obtenerse de forma sencilla:

Sea A una matriz simétrica y definida positiva
 Sean $\{s_i\}; i=1, \dots, n$; conjugados respecto a A
 Sea \bar{x} , la solución del sistema $A\bar{x} + \bar{b} = \bar{0}$
 Sea \bar{x}^0 una aproximación inicial a \bar{x}^*

(3.21)

Entonces,
$$\bar{x}^* = \bar{x}^0 + \sum_{i=1}^n \alpha_i \bar{s}_i$$

con
$$\alpha_i = - \frac{\bar{s}_i^t (\bar{b} + A\bar{x}^0)}{\bar{s}_i^t A \bar{s}_i}$$

Como hemos visto anteriormente (3.18), la minimización de una función cuadrática con hessiano definido positivo, implica la solución de un sistema lineal de ecuaciones, esto es:

si \bar{x}^* minimiza $f(\bar{x}) = c + \bar{b}^t \bar{x} + 1/2 \bar{x}^t A \bar{x}$

con A simétrica y definida positiva, (3.22)

entonces: $A\bar{x} + \bar{b} = \bar{0}$

Haciendo uso del desarrollo (3.21), podemos escribir el algoritmo general de direcciones conjugadas para funciones cuadráticas en la forma iterativa siguiente:

Sea \bar{x}^* que minimiza: $f(\bar{x}) = c + \bar{b}^t \bar{x} + 1/2 \bar{x}^t A \bar{x}$

con A simétrica y definida positiva,

Sea \bar{x}^0 una aproximación inicial a \bar{x}^*

Sea \bar{s}^1 una dirección inicial arbitraria

Entonces: siendo, $\bar{x}^k = \bar{x}^{k-1} + \alpha_k \bar{s}^k$; $k=1, \dots, n$

$$\text{con, } \begin{cases} \bar{s}^k \cdot A \bar{s}^i = 0 & ; i < k \\ \bar{s}^k \cdot A \bar{s}^k \neq 0 & ; i = k \end{cases} \quad (3.23)$$

$$\text{con, } \alpha_k = - \frac{\nabla f(\bar{x}^{k-1}) \cdot \bar{s}^k}{\bar{s}^k \cdot A \bar{s}^k}$$

se cumple que: $\bar{x}^n = \bar{x}^*$

Además:

los α_k minimizan: $\varphi(\alpha) = f(\bar{x}^{k-1} + \alpha \bar{s}^k)$

$$\nabla f(\bar{x}^k) \cdot \bar{s}^i = 0 ; i=1, \dots, k ; k=1, \dots, n$$

La obtención de los coeficientes puede por tanto realizarse mediante una minimización unidireccional.

Este algoritmo converge al mínimo de una función cuadrática con hessiano definido positivo en un número finito de pasos, igual al número de variables, independientemente de la aproximación inicial a la solución y la elección de las direcciones conjugadas.

Una particular elección de éstas como combinación lineal de los gradientes de la función objetivo en las iteraciones

anteriores, conduce al método de los gradientes conjugados. Formulando de esta manera las sucesivas direcciones conjugadas, y haciendo uso de las propiedades anteriores, se obtienen inmediatamente las

Fórmulas de Hestenes-Stiefel:

$$\begin{aligned} \bar{s}^1 &= -\bar{\nabla} f(\bar{x}^0) \\ \bar{s}^k &= -\bar{\nabla} f(\bar{x}^{k-1}) + \beta_k \bar{s}^{k-1} \quad ; k=2, \dots, n \end{aligned} \quad (3.24)$$

con:

$$\beta_k = \frac{[\bar{\nabla} f(\bar{x}^{k-1}) - \bar{\nabla} f(\bar{x}^{k-2})] \bar{\nabla} f(\bar{x}^{k-1})}{[\bar{\nabla} f(\bar{x}^{k-1}) - \bar{\nabla} f(\bar{x}^{k-2})] \bar{s}^{k-1}}$$

Las fórmulas de Hestenes-Stiefel pueden simplificarse empleando diversas propiedades de los gradientes conjugados -válidas para problemas cuadráticos y para otros tipos de problemas bajo ciertas condiciones- dando lugar a las conocidas expresiones siguientes para la obtención de β_k :

Fórmulas de Fletcher-Reeves:

$$\beta_k = \frac{\bar{\nabla} f(\bar{x}^{k-1}) \bar{\nabla} f(\bar{x}^{k-1})}{\bar{\nabla} f(\bar{x}^{k-2}) \bar{\nabla} f(\bar{x}^{k-2})} \quad ; k=2, \dots, n \quad (3.25)$$

Fórmulas de Polak-Ribiere:

$$\beta_k = \frac{[\nabla f(\bar{x}^{k-1}) - \nabla f(\bar{x}^{k-2})] \nabla^t f(\bar{x}^{k-1})}{\nabla f(\bar{x}^{k-2}) \nabla^t f(\bar{x}^{k-2})} ; k=2, \dots, n \quad (3.26)$$

Fórmulas de Myers:

$$\beta_k = \frac{\nabla f(\bar{x}^{k-1}) \nabla^t f(\bar{x}^{k-1})}{[\nabla f(\bar{x}^{k-1}) - \nabla f(\bar{x}^{k-2})] \nabla^t f(\bar{x}^{k-1})} ; k=2, \dots, n \quad (3.27)$$

Quando se aplica el algoritmo de los gradientes conjugados a la minimización de funciones no lineales, la convergencia no está asegurada en "n" iteraciones. Es práctica habitual reinicializar el algoritmo cada "n" iteraciones, aunque algunos autores han encontrado empíricamente que la eficiencia es mayor si se permite al algoritmo proseguir durante algunas iteraciones más antes de reinicializar los cálculos.

Métodos Quasi-Newton

El método de Newton de aproximación cuadrática, conduce a un sistema lineal de ecuaciones (3.19) que es preciso resolver en cada iteración, y cuya matriz de coeficientes es el hessiano de la función objetivo particularizado en el último punto obtenido. Los métodos Quasi-Newton permiten aproximar el hessiano, o su inversa, directamente, a partir de los valores previamente obtenidos de la función objetivo y su gradiente.

Sean:

$$\underline{R}^k = \underline{H}(\underline{x}^k) \quad (3.28)$$

$$\underline{S}^k = \underline{H}^{-1}(\underline{x}^k)$$

aproximaciones al hessiano y su inversa en la iteración "k",
obtenidas de forma que se verifiquen las ecuaciones de
Quasi-Newton:

$$\begin{aligned} \underline{y}^k &= \underline{R}^{k+1} \underline{r}^k \\ \underline{r}^k &= \underline{S}^{k+1} \underline{y}^k \end{aligned} \quad (3.29)$$

siendo:

$$\begin{aligned} \underline{y}^k &= \underline{\nabla} f(\underline{x}^{k+1}) - \underline{\nabla} f(\underline{x}^k) \\ \underline{r}^k &= \underline{x}^{k+1} - \underline{x}^k \end{aligned}$$

Sustituyendo estas aproximaciones en la ecuación de Newton
(3.19) se obtiene el algoritmo Quasi-Newton que puede escribirse
en las formas alternativas:

$$\begin{aligned} \underline{x}^k &= \underline{x}^{k-1} - [\underline{R}^{k-1}]^{-1} \underline{\nabla} f(\underline{x}^{k-1}) \\ & \quad ; k=1, \dots, n \end{aligned} \quad (3.30)$$

$$\underline{x}^k = \underline{x}^{k-1} - \underline{S}^{k-1} \underline{\nabla} f(\underline{x}^{k-1})$$

y para cuya aplicación se parte de una aproximación inicial a la
solución " \underline{x}^0 " y una matriz simétrica definida positiva " \underline{S}^0 ". Los
diversos métodos Quasi-Newton difieren en la elección de las
sucesivas matrices " \underline{S} " y " \underline{R} " que verifiquen las ecuaciones

(3.29).

La construcción de estas aproximaciones se plantea actualizando en cada iteración el valor de la aproximación en la iteración precedente, en la forma:

$$\underset{\sim}{S}^{k+1} = \underset{\sim}{S}^k + \underset{\sim}{C}^k \quad (3.31)$$

donde la matriz " $\underset{\sim}{C}$ " se elige adecuadamente como más adelante se expone. Evidentemente, para que en cada iteración el problema de minimización esté bien definido, la aproximación al hessiano debe ser una matriz simétrica y definida positiva, además de verificar la ecuación de Quasi-Newton (3.29).

Existen fundamentalmente dos grupos de métodos Quasi-Newton, que se diferencian por las propiedades exigidas a la matriz de actualización " $\underset{\sim}{C}$ ", y que denominaremos:

- métodos de rango uno
- métodos de rango dos

Dentro de cada grupo existen a su vez diversas familias que se diferencian en la forma de realizar la elección de la matriz citada, y numerosas variantes dentro de cada familia. Por ello extractaremos únicamente los algoritmos más representativos.

Metodos Quasi-Newton de rango uno.

Los métodos de rango uno son aquellos en que cada aproximación a la inversa del hessiano se obtiene sumando a la anterior una matriz de actualización " $\underset{\sim}{C}$ " simétrica, semidefinida

positiva y de rango uno, que puede escribirse en la forma:

$$\bar{C}^k = \gamma \frac{\bar{z}^k}{z^k} \frac{\bar{z}^k}{z^k} t \quad (3.32)$$

Por tanto, si la aproximación inicial es simétrica, se garantiza la simetría, aunque no la definición positiva, de las aproximaciones siguientes.

Es sencillo demostrar que la elección del escalar " γ_k " y el vector " \bar{z}^k " en cada iteración de forma que se satisfaga la ecuación de Quasi-Newton conduce a la fórmula de rango uno:

$$\bar{S}^{k+1} = \bar{S}^k + \frac{\frac{\bar{z}^k}{z^k} \frac{\bar{z}^k}{z^k} t}{\frac{\bar{z}^k}{z^k} \frac{\bar{z}^k}{z^k} t} \frac{[\bar{r} - \bar{S}^k \bar{y}^k] [\bar{r} - \bar{S}^k \bar{y}^k]}{[\bar{r} - \bar{S}^k \bar{y}^k]} \quad (3.33)$$

El algoritmo (3.30), aplicado ahora a un problema cuadrático y partiendo de una aproximación inicial simétrica y definida positiva a la inversa del hessiano, no requiere el uso de una minimización unidireccional en cada iteración, converge en "n" iteraciones, y la aproximación n-ésima de la inversa del hessiano es exacta, independientemente de la aproximación inicial a la solución.

Si se aplica a funciones no lineales es conveniente, en general, emplear una minimización unidireccional, si bien en este caso no está asegurada la convergencia, y la principal forma de fallo del algoritmo se produce por pérdida de la definición positiva de la aproximación.

Metodos Quasi-Newton de rango dos.

Los métodos de rango dos son aquellos en que se construyen las sucesivas aproximaciones a la inversa del hessiano mediante matrices de actualización simétricas y semidefinidas positivas de rango dos, de forma que partiendo de una aproximación inicial simétrica y definida positiva, se asegure la simetría y la definición positiva de las aproximaciones siguientes.

Una posible elección de la matriz " \tilde{C} " es de la forma:

$$\tilde{C}^k = \gamma_k \frac{\underline{k} \underline{k}^t}{\underline{r} \underline{r}} + \rho_k \frac{[\underline{S} \underline{y}] [\underline{S} \underline{y}]}{\underline{y} \underline{S} \underline{y}} \quad (3.34)$$

lo que garantiza la simetría de la aproximación.

Una elección de los escalares " γ_k " y " ρ_k " en cada iteración, conduce a la fórmula de Davidon- Fletcher- Powel, DFP, que garantiza la conservación de la aproximación a la inversa del hessiano como matriz definida positiva siempre que la aproximación inicial lo sea, y se utilice una minimización unidireccional en cada iteración:

$$\tilde{S}^{k+1} = \tilde{S}^k + \frac{\frac{\underline{k} \underline{k}^t}{\underline{r} \underline{r}}}{\frac{\underline{k}^t \underline{k}}{\underline{r} \underline{y}}} - \frac{\frac{[\underline{S} \underline{y}] [\underline{S} \underline{y}]}{\underline{y} \underline{S} \underline{y}}}{\frac{\underline{k}^t \underline{k}}{\underline{y} \underline{S} \underline{y}}} \quad (3.35)$$

Cuando se aplica el algoritmo anterior a problemas cuadráticos se generan siempre direcciones conjugadas. Además, si la aproximación inicial a la inversa del hessiano es la matriz identidad, coincide exactamente con el algoritmo de los

gradientes conjugados. Converge por tanto en "n" iteraciones, y la n-ésima aproximación de la inversa del hessiano es exacta, independientemente de la aproximación inicial a la solución.

La fórmula anterior puede invertirse obteniendo directamente una aproximación del hessiano:

$$\begin{aligned} \bar{R}^{k+1} = & \left(\begin{array}{c} \bar{I} - \frac{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}}{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}} \\ \bar{I} - \frac{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}}{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}} \end{array} \right) \bar{R}^k \left(\begin{array}{c} \bar{I} - \frac{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}}{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}} \\ \bar{I} - \frac{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}}{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}} \end{array} \right) + \\ & + \frac{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}}{\frac{\bar{y}^k}{\bar{y}} \frac{\bar{r}^k}{\bar{r}}} \end{aligned} \quad (3.36)$$

La existencia de las dos fórmulas anteriores para aproximar el hessiano y su inversa, sugiere la posibilidad de invertir los términos, y emplear para aproximar el hessiano y su inversa fórmulas similares a (3.35) y (3.36) respectivamente, en las cuales las posiciones de las matrices "S" y "R" y los vectores "r" e "y" se intercambian entre si, obteniendo dos fórmulas complementarias de las anteriores, denominadas fórmulas de Broyden- Fletcher- Goldfarb- Shanno, o BFGS:

$$\begin{aligned} \bar{S}^{k+1} = & \left(\begin{array}{c} \bar{I} - \frac{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}}{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}} \\ \bar{I} - \frac{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}}{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}} \end{array} \right) \bar{S}^k \left(\begin{array}{c} \bar{I} - \frac{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}}{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}} \\ \bar{I} - \frac{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}}{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}} \end{array} \right) + \\ & + \frac{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}}{\frac{\bar{r}^k}{\bar{r}} \frac{\bar{y}^k}{\bar{y}}} \end{aligned} \quad (3.37)$$

$$\bar{R}^{k+1} = \bar{R}^k + \frac{\bar{y}^k \bar{y}^k t}{\bar{y}^k \bar{r}^k} - \frac{[\bar{R}^k \bar{r}^k] [\bar{R}^k \bar{r}^k t]}{\bar{r}^k \bar{r}^k} \quad (3.38)$$

Las propiedades del algoritmo anterior son similares a las ya citadas para el algoritmo DFP (3.35), aunque se estima en general que su eficiencia es más elevada. Tal vez pueda ser considerado en la actualidad como uno de los métodos más potentes y firmemente establecidos de la programación matemática, ya que ha permitido en la práctica abordar eficazmente un grupo considerablemente numeroso de problemas de minimización no restringida.

Métodos para problemas específicos

En la aplicación de los métodos citados a problemas específicos puede hacerse uso de peculiaridades de los mismos para obtener algoritmos de mayor eficiencia, aunque de menor generalidad. Un caso interesante por la frecuencia con que aparece en la práctica es el de mínimos cuadrados, cuya expresión general puede reducirse a la forma:

$$\begin{aligned} \text{minimizar: } F(\bar{x}) &= \sum_i^t \bar{f}_i(\bar{x}) \bar{f}_i(\bar{x}) \quad ; \bar{x} = \{x_i\} ; i=1, \dots, n \\ & ; \bar{f}_j(\bar{x}) = \{f_j(\bar{x})\} ; j=1, \dots, m \end{aligned} \quad (3.39)$$

Evidentemente, la forma particular de la función objetivo anterior da lugar a expresiones particulares para su gradiente y hessiano, que podemos escribir en la forma:

Siendo:

$$\begin{aligned}
 \underline{J}(\bar{x}) &= \underline{\nabla} f(\bar{x}) \\
 \underline{T}_{\sim j}(\bar{x}) &= \underline{\nabla}^2 f(\bar{x}) \\
 \underline{Q}(\bar{x}) &= \sum_{j=1}^m \underline{T}_{\sim j}(\bar{x}) \underline{f}_{\sim j}(\bar{x}) \quad (3.40)
 \end{aligned}$$

Entonces:

$$\begin{aligned}
 \underline{\nabla} F(\bar{x}) &= 2 \underline{f}^t(\bar{x}) \underline{J}(\bar{x}) \\
 \underline{\nabla}^2 F(\bar{x}) &= 2 \underline{J}^t(\bar{x}) \underline{J}(\bar{x}) + 2 \underline{Q}(\bar{x})
 \end{aligned}$$

Aplicando al problema (3.39) el método de Newton de segundo orden, y utilizando las expresiones anteriores, obtenemos:

$$\begin{aligned}
 \underline{x}_{k+1} &= \underline{x}_k + \underline{r}_k \\
 [\underline{J}^t(\underline{x}_k) \underline{J}(\underline{x}_k) + \underline{Q}(\underline{x}_k)] \underline{r}_k &= - \underline{J}^t(\underline{x}_k) \underline{f}(\underline{x}_k) \quad (3.41)
 \end{aligned}$$

Evidentemente, el mayor coste operativo, en lo que respecta al cálculo de los términos que intervienen en la fórmula anterior, reside en la obtención de la matriz "Q", cuyo cálculo puede además ser inviable o cuanto menos muy costoso. Para obviar su cálculo existen dos tipos de técnicas:

- despreciar la matriz
- aproximar la matriz

Despreciando la matriz "Q" se obtiene el algoritmo de Gauss-Newton:

$$\begin{aligned} \underline{x}^{k+1} &= \underline{x}^k + \underline{r}^k \\ [\underline{J}^t(\underline{x}^k) \underline{J}(\underline{x}^k)] \underline{r}^k &= - \underline{J}^t(\underline{x}^k) \underline{f}(\underline{x}^k) \end{aligned} \quad (3.42)$$

Aproximando la matriz " \underline{Q} " de diferentes formas se obtiene toda una familia de métodos, la más representativa de las cuales es la de Levenberg-Marquardt:

$$\begin{aligned} \underline{x}^{k+1} &= \underline{x}^k + \underline{r}^k \\ [\underline{J}^t(\underline{x}^k) \underline{J}(\underline{x}^k) + \gamma_k \underline{D}(\underline{x}^k)] \underline{r}^k &= - \underline{J}^t(\underline{x}^k) \underline{f}(\underline{x}^k) \end{aligned} \quad (3.43)$$

donde la matriz " \underline{D} " es diagonal y definida positiva, y el coeficiente " γ_k " puede obtenerse de formas diversas, debidas fundamentalmente a Levenberg, Marquardt y Fletcher.

III.3.2.2.3 Enfoque global

Todos los métodos de minimización no restringida anteriormente expuestos pueden considerarse como particularizaciones de un solo enfoque globalizador, que podemos expresar como:

$$\text{Siendo: } \bar{x}^{k+1} = \bar{x}^k + \bar{r}^k$$

$$\text{Hallar: } \bar{r}^k \quad ; k \geq 1$$

de forma que:

$$\bar{r}^k, \text{ minimiza } a(\bar{r}) = \begin{bmatrix} \bar{\nabla} f(\bar{x}^k) & \bar{r}^k \end{bmatrix} \quad (3.44)$$

$$\text{verificando: } \begin{bmatrix} \bar{r}^k & \bar{M} \bar{r}^k \end{bmatrix} = 1$$

donde " \bar{M} " es una matriz simétrica definida positiva. El planteamiento anterior puede interpretarse como un proceso iterativo en el que se obtienen sucesivamente vectores en cuyas direcciones se maximiza el descenso de la función objetivo, imponiendo una restricción en la medida de los módulos de tales vectores. La matriz " \bar{M} " define la métrica empleada. Desde este punto de vista, según adoptemos como matriz " \bar{M} " la matriz identidad, el hessiano de la función objetivo, o una aproximación al mismo, obtendremos respectivamente los métodos de máximo descenso, los métodos de Newton de segundo orden o de direcciones conjugadas, y los métodos Quasi-Newton.

III.4 MINIMIZACIÓN CON RESTRICCIONES DE IGUALDAD

El problema general de minimización con restricciones de igualdad puede escribirse en la forma:

$$\begin{aligned}
 \text{minimizar} & : f(\bar{x}) & ; \bar{x} = \{x_i\} & ; i=1, \dots, n \\
 \text{verificando:} & \bar{h}(\bar{x}) = \bar{0} & ; \bar{h}(\bar{x}) = \{h_l(\bar{x})\} & ; l=1, \dots, p \\
 & & & p \leq n
 \end{aligned}
 \tag{3.45}$$

donde supondremos que las funciones "h_l" son funcionalmente independientes.

III.4.1 Condición de mínimo

La condición necesaria de extremo para el problema de minimización con restricciones de igualdad, se expresa habitualmente en función de los denominados multiplicadores de Lagrange " λ ", en la forma siguiente:

Si: \bar{x}^* es solución del problema (3.45), entonces

existe : $\bar{\lambda}^* = \{ \lambda_l^* \} ; l=1, \dots, p$

verificándose:

$$\begin{aligned}
 [\bar{\nabla} f(\bar{x}^*) - \bar{\lambda}^{*t} \bar{\nabla} h(\bar{x}^*)] & = \bar{0} & (3.46) \\
 \bar{h}(\bar{x}^*) & = \bar{0}
 \end{aligned}$$

Definiendo una función, denominada lagrangiano del problema (3.45) en la forma:

$$L(\bar{x}, \bar{\lambda}) = f(\bar{x}) - \bar{\lambda}^t \bar{h}(\bar{x})
 \tag{3.47}$$

La condición necesaria (3.46) puede escribirse:

Si: \bar{x}^* es solución del problema (3.45), entonces

existe : $\bar{\lambda} = \{ \lambda_l \} ; l=1, \dots, p$

verificándose:

$$\frac{\partial}{\partial x} L(\bar{x}^*, \bar{\lambda}^*) = 0 \quad (3.48)$$

$$\frac{\partial}{\partial \lambda} L(\bar{x}^*, \bar{\lambda}^*) = 0$$

condición que implica la anulación del gradiente del lagrangiano.

Posteriormente analizaremos este aspecto con mayor detalle.

Es posible obtener condiciones suficientes de mínimo para el problema general de minimización con restricciones de igualdad, si bien su utilidad es escasa en la práctica.

III.4.2 Métodos de solución del problema

Desde un punto de vista analítico es posible, al menos teóricamente, resolver el sistema no lineal de "n+p" ecuaciones con "n+p" incógnitas que representa la anulación del gradiente del lagrangiano, y posteriormente discriminar entre las soluciones de este sistema las que son mínimos locales, y el mínimo global (si existe). Desde un punto de vista numérico puede plantearse la resolución de este sistema mediante un algoritmo iterativo general para solución de sistemas de ecuaciones no lineales (Newton, Newton-Raphson, etc). Dada la inherente dificultad que representa en general la solución de este tipo de sistemas, este procedimiento no siempre es adecuado,

y por ello se emplean frecuentemente particularizaciones de algoritmos numéricos basados en otras consideraciones, diseñados para resolver el problema de minimización con restricciones de tipo general. No obstante, es importante reseñar que, como veremos posteriormente, las ideas fundamentales desarrolladas para analizar el problema general se obtienen a partir de las condiciones (3.48), y que diversos algoritmos para problemas de minimización con restricciones en desigualdad plantean una reducción del problema a una forma con restricciones de igualdad del tipo (3.45). Este planteamiento, por tanto, es sumamente importante, y la solución directa del sistema de ecuaciones no lineales mencionado no debe descartarse de antemano. Ambos enfoques, de hecho, y al igual que se expuso en la descripción de los métodos de descenso, si bien conceptualmente distintos, son formalmente idénticos, y una vez más se pone de manifiesto la íntima relación existente entre los métodos de minimización y los métodos de solución de sistemas de ecuaciones no lineales.

III.5 MINIMIZACIÓN CON RESTRICCIONES DE TIPO GENERAL

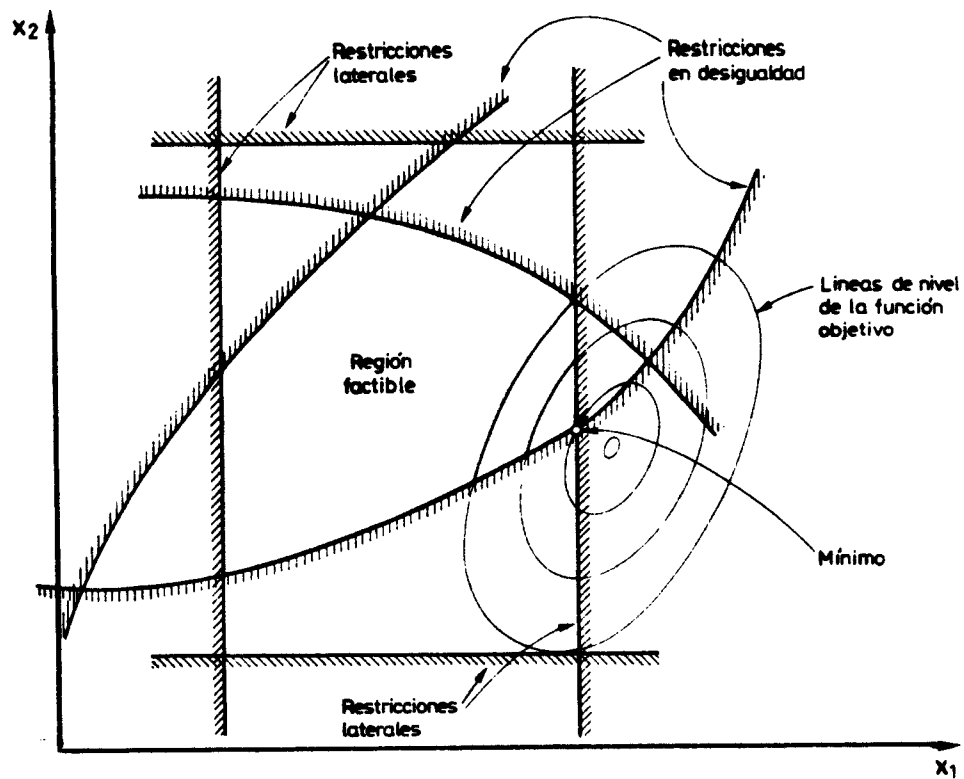
En la exposición siguiente se analiza el problema general de minimización con restricciones (3.2), o problema general de programación matemática, cuya diferencia fundamental respecto a los problemas analizados previamente reside en la existencia de restricciones en desigualdad (Fig. 3.4-a).

Como se ha expuesto con anterioridad, las restricciones laterales pueden considerarse englobadas en el grupo de restricciones generales en desigualdad, si bien la mayor parte de los algoritmos de programación matemática permiten tratar las restricciones laterales separadamente con mayor sencillez y eficiencia. Además, se ha descrito previamente la posibilidad de sustituir cada restricción de igualdad por dos restricciones en desigualdad, y por tanto, el problema general de programación matemática (3.2) y el problema estándar (3.1) son formalmente equivalentes.

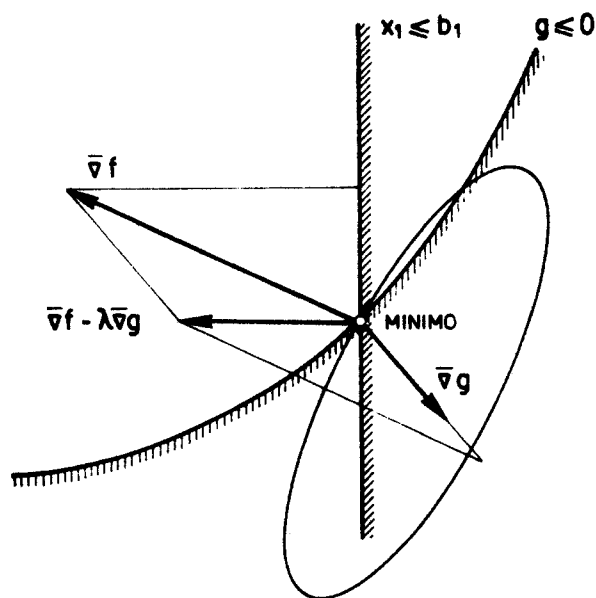
Por estos motivos, en lo sucesivo emplearemos una u otra formulación y trataremos las restricciones laterales separada o conjuntamente con las restantes restricciones en desigualdad, según se estime más conveniente para la claridad de las exposiciones.

III.5.1 Condiciones de mínimo. Condiciones de Kuhn-Tucker

Las restricciones de desigualdad del problema general (3.2) pueden tratarse como restricciones de igualdad, introduciendo variables adicionales en el problema, en la forma siguiente:



a)



b)

Figura 3.4.- Representación gráfica de un problema general de minimización con restricciones, e interpretación geométrica de las condiciones de Kuhn-Tucker.

a) Representación de un problema general de programación matemática bidimensional.

b) Interpretación geométrica de las condiciones de Kuhn-Tucker en la solución del problema anterior.

$$\begin{aligned}
g_j(\bar{x}) + \gamma_j^2 &= 0 && ;j=1, \dots, m \\
(a_i - x_i) + \alpha_i^2 &= 0 && ;i=1, \dots, n \quad (3.49) \\
(x_i - b_i) + \beta_i^2 &= 0 && ;i=1, \dots, n
\end{aligned}$$

La introducción de estas variables permite reducir el problema general de minimización con restricciones en desigualdad a un problema de minimización con restricciones de igualdad. No obstante, el problema reducido está por su propia naturaleza mal condicionado numéricamente en general, y su resolución directa no es viable mediante técnicas habituales. Sin embargo, a partir de la condición necesaria de mínimo en función de los multiplicadores de Lagrange, pueden obtenerse las condiciones de mínimo para el problema general.

En efecto, el problema reducido puede escribirse como:

$$\begin{aligned}
\text{minimizar} &: f(\bar{x}) && ;\bar{x} = \{x_i\} && ;i=1, \dots, n \\
\text{verificando:} & g_j(\bar{x}) + \gamma_j^2 &= 0 && ;j=1, \dots, m \\
& h_l(\bar{x}) &= 0 && ;l=1, \dots, p \quad (3.50) \\
& (a_i - x_i) + \alpha_i^2 &= 0 && ;i=1, \dots, n \\
& (x_i - b_i) + \beta_i^2 &= 0 && ;i=1, \dots, n
\end{aligned}$$

El lagrangiano del problema (3.50), tal como ha sido definido en (3.47), puede expresarse como:

$$\begin{aligned}
L(\bar{x}, \bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\lambda}^h, \bar{\lambda}^g, \bar{\lambda}^a, \bar{\lambda}^b) = & f(\bar{x}) - \sum_{l=1}^p \lambda_l^h h_l(\bar{x}) - \\
& - \sum_{j=1}^m \lambda_j^g (g_j(\bar{x}) + \gamma_j^2) - \\
& - \sum_{i=1}^n \lambda_i^a (a_i - x_i + \alpha_i^2) - \quad (3.51) \\
& - \sum_{i=1}^n \lambda_i^b (x_i - b_i + \beta_i^2)
\end{aligned}$$

En virtud de (3.48), la anulaci3n del gradiente del lagrangiano respecto a variables y multiplicadores de Lagrange es condici3n necesaria de extremo. Planteando esta condici3n para el lagrangiano anterior, es inmediato comprobar que puede obtenerse un conjunto de condiciones generales de extremo que incluyen a las obtenidas para minimizaci3n con restricciones de igualdad (3.46) como caso particular; en estas condiciones no intervienen las variables adicionales introducidas en el problema, ni los multiplicadores correspondientes a las restricciones laterales.

Adem1s, si en el m3nimo de "f(x)" los gradientes de las restricciones de igualdad y los gradientes de las restricciones activas de desigualdad son linealmente independientes, es sencillo demostrar que es condici3n necesaria de m3nimo el que los multiplicadores asociados a todas las restricciones en desigualdad sean no positivos. Esta condici3n es necesaria siempre que el m3nimo verifique unos requisitos m1s amplios y menos restrictivos que la mencionada independencia lineal de los gradientes, requisitos que se conocen con el nombre gen3rico de

regularidad, y que no expondremos aquí por exceder claramente de nuestro contexto. La condición de no positividad de los multiplicadores correspondientes a restricciones en desigualdad reviste una importancia excepcional, como se observará a lo largo del desarrollo siguiente.

Las condiciones necesarias de extremo, junto con la anterior condición necesaria de mínimo, configuran las condiciones necesarias de mínimo de Kuhn-Tucker que pueden escribirse en la forma:

Si: \bar{x}^* es un mínimo del problema (3.2), entonces

$$\text{siendo : } L(\bar{x}, \bar{\lambda}, \bar{\lambda}) = f(\bar{x}) - \sum_j \bar{\lambda}_j g_j(\bar{x}) - \sum_l \bar{\lambda}_l h_l(\bar{x})$$

el lagrangiano del problema general,

$$\text{existen : } \bar{\lambda}_j^*, \bar{\lambda}_l^*$$

que verifican:

$$\begin{aligned} g_j(\bar{x}^*) \leq 0 \quad ; \quad \bar{\lambda}_j^* \leq 0 \quad ; \quad \bar{\lambda}_j^* g_j(\bar{x}^*) = 0 \quad ; j=1, \dots, m \\ h_l(\bar{x}^*) = 0 \quad ; l=1, \dots, p \end{aligned} \tag{3.52}$$

$$\frac{\partial}{\partial x_i} L(\bar{x}, \bar{\lambda}, \bar{\lambda}) \Big|_{\bar{x}^*, \bar{\lambda}^*, \bar{\lambda}^*} \Rightarrow \begin{cases} \geq 0 \quad ; \text{si } x_i = a_i \\ = 0 \quad ; \text{si } a_i < x_i < b_i \\ \leq 0 \quad ; \text{si } x_i = b_i \end{cases} ; i=1, \dots, n$$

Si las funciones "f" y "g" son convexas, y las funciones "h" lineales, las condiciones de Kuhn-Tucker son necesarias y

suficientes de mínimo global. Si las mencionadas convexidad y linealidad se verifican únicamente en un entorno del punto dado, las condiciones de Kuhn-Tucker son entonces necesarias y suficientes de mínimo local (Fig. 3.4-b).

La forma general de las condiciones de Kuhn-Tucker admite diversas particularizaciones según existan o no restricciones de igualdad, restricciones laterales en varias variables, etc.

Cuando se aplican las condiciones de Kuhn-Tucker a problemas convexos permiten definir el denominado problema dual, a partir del cual se derivan diversos métodos de minimización con restricciones.

Los valores de los multiplicadores " λ^g " y " λ^h " de las condiciones de Kuhn-Tucker en un mínimo, poseen además un significado sutil en lo que respecta a la sensibilidad de la solución del problema de minimización respecto a las restricciones impuestas cuyo conocimiento es de gran importancia y aplicación en el desarrollo de algoritmos basados en el uso de multiplicadores.

En efecto, consideremos el problema perturbado

$$\begin{aligned}
\text{minimizar} & : f(\bar{x}) & ; \bar{x} = \{x_i\} & ; i=1, \dots, n \\
\text{dados} & : g(\bar{x}) = \{g_j(\bar{x})\} & & ; j=1, \dots, m \\
& & h(\bar{x}) = \{h_l(\bar{x})\} & ; l=1, \dots, p \\
\text{verificando:} & g_j(\bar{x}) \leq \epsilon_j & & ; j=1, \dots, m \\
& & h_l(\bar{x}) = \delta_l & ; l=1, \dots, p \\
& & a_i \leq x_i \leq b_i & ; i=1, \dots, n
\end{aligned} \tag{3.53}$$

correspondiente al problema general (3.2).

Podemos escribir inmediatamente el lagrangiano del problema anterior en la forma:

$$L(\bar{x}, \bar{\lambda}, \bar{\lambda}) = f(\bar{x}) - \sum_j \bar{\lambda}_j (g_j(\bar{x}) - \epsilon_j) - \sum_l \bar{\lambda}_l (h_l(\bar{x}) - \delta_l) \tag{3.54}$$

De donde se deduce que:

Denominando: $\bar{x}(\bar{\epsilon}, \bar{\delta})$ a la solución del problema perturbado

$$y: \bar{\lambda}_j(\bar{\epsilon}, \bar{\delta}), \bar{\lambda}_l(\bar{\epsilon}, \bar{\delta})$$

a los multiplicadores que verifican las condiciones de Kuhn-Tucker (3.55)

$$\text{entonces} : f(\bar{x}(\bar{\epsilon}, \bar{\delta})) = L(\bar{x}(\bar{\epsilon}, \bar{\delta}), \bar{\lambda}_j(\bar{\epsilon}, \bar{\delta}), \bar{\lambda}_l(\bar{\epsilon}, \bar{\delta}))$$

esto es, el valor de la función objetivo y el del lagrangiano coinciden en la solución del problema.

Derivando esta igualdad respecto a las perturbaciones " $\bar{\epsilon}$ ", y " $\bar{\delta}$ ", y haciendo uso de las condiciones de Kuhn-Tucker para eliminar los términos restantes, se obtiene con sencillez:

$$\begin{aligned} \frac{\partial}{\partial \bar{\epsilon}} f(\bar{x}(\bar{\epsilon}, \bar{\delta})) &= \bar{\lambda}^g \\ \frac{\partial}{\partial \bar{\delta}} f(\bar{x}(\bar{\epsilon}, \bar{\delta})) &= \bar{\lambda}^h \end{aligned} \tag{3.56}$$

ecuaciones que demuestran que el valor de los multiplicadores que verifican las condiciones de Kuhn-Tucker es una medida de la sensibilidad del valor de la función objetivo en la solución respecto a una perturbación en las restricciones.

Las ecuaciones (3.56) permiten ofrecer, además, una explicación intuitiva de parte de las condiciones de Kuhn-Tucker. En efecto, si una restricción en desigualdad está inactiva en la solución, una pequeña perturbación en la misma no debe alterar el valor de la función objetivo en la solución, y por tanto su multiplicador asociado debe anularse. Por el contrario, si una restricción en desigualdad está activa en la solución, una pequeña perturbación en la misma no debe dar lugar a un incremento en el valor de la función objetivo, y por tanto su multiplicador asociado debe ser no positivo.

III.5.2 Dualidad

Si el problema (3.2) es convexo, se puede demostrar fácilmente a partir de las condiciones de Kuhn-Tucker que en su solución el lagrangiano del problema, tal como se define en (3.52), tiene un punto de ensilladura (Fig. 3.5). Esto es:

Si: $\bar{x}^*, \bar{g}^*, \bar{h}^*$ verifican las condiciones (3.52)

siendo : $L(\bar{x}, \bar{g}, \bar{h}) = f(\bar{x}) - \bar{\lambda}^g \text{ t } \bar{g}(\bar{x}) - \bar{\lambda}^h \text{ t } \bar{h}(\bar{x})$

$$\bar{x} = \{ \bar{x} \mid a_i \leq x_i \leq b_i \quad ; i=1, \dots, n \} \quad (3.57)$$

$$\bar{\Lambda} = \{ \bar{\lambda}^g \mid \lambda_j \leq 0 \quad ; j=1, \dots, m \}$$

se verifica:

$$L(\bar{x}^*, \bar{g}, \bar{h}) \leq L(\bar{x}^*, \bar{\lambda}^g, \bar{h}^*) \leq L(\bar{x}^*, \bar{g}^*, \bar{h}^*)$$

$$\forall \bar{\lambda}^g \in \bar{\Lambda}, \forall \bar{\lambda}^h \quad \quad \quad \forall \bar{x} \in \bar{x}$$

La condición de punto de ensilladura puede expresarse de una forma alternativa como:

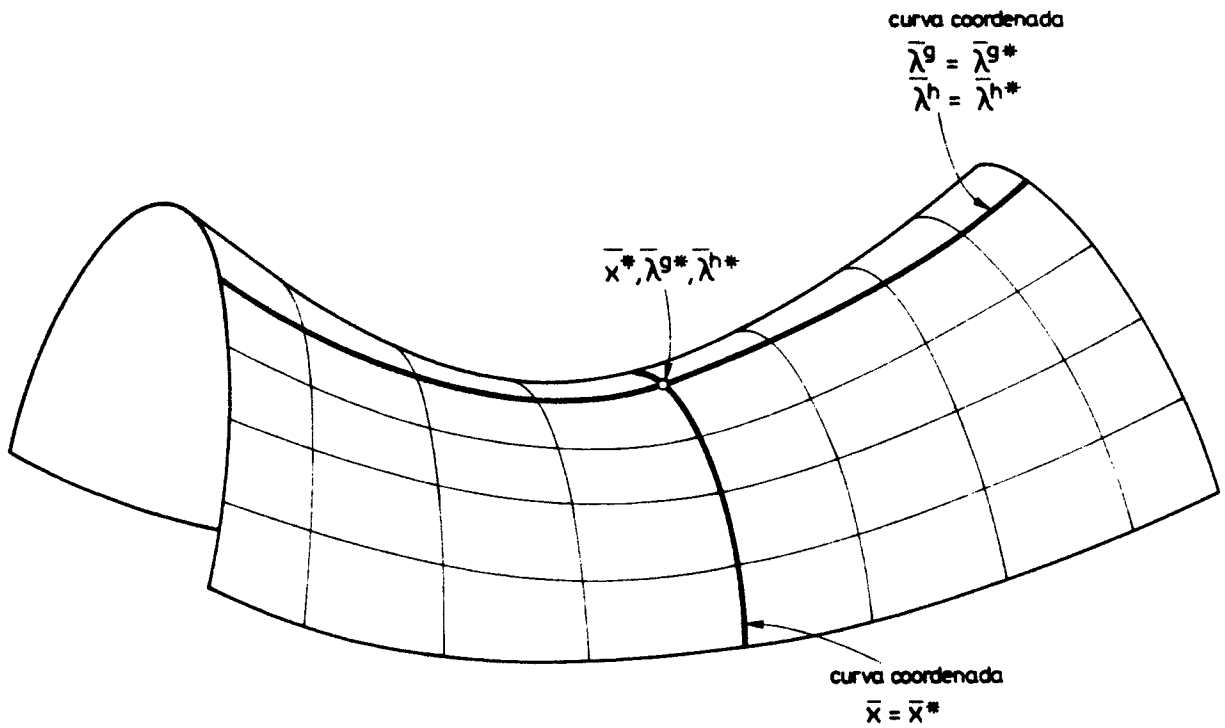


Figura 3.5.- Idealización de la función lagrangiano de un problema de programación convexa y condición de punto de ensilladura del óptimo.

Si: $\bar{x}^*, \bar{g}^*, \bar{h}^*$ verifican las condiciones (3.52)

dados: $\bar{g} \in \Lambda, \bar{h}$

sea: $\bar{x} = \bar{x}(\bar{g}, \bar{h})$

la solución del problema:

$$\begin{cases} \text{minimizar} & : \mathcal{J}(\bar{x}) = L(\bar{x}, \bar{g}, \bar{h}) \\ \text{verificando} & : \bar{x} \in X \end{cases}$$

y dado: $\bar{x} \in X$

sea: $\bar{\lambda} = \bar{\lambda}(\bar{x}), \bar{\lambda} = \bar{\lambda}(\bar{x})$ (3.58)

la solución del problema:

$$\begin{cases} \text{maximizar} & : \mathcal{J}(\bar{\lambda}, \bar{\lambda}) = L(\bar{x}, \bar{g}, \bar{h}) \\ \text{verificando} & : \bar{\lambda} \in \Lambda \end{cases}$$

entonces:

$\forall \bar{x} \in X, \forall \bar{\lambda} \in \Lambda, \forall \bar{h}$, se cumple:

$$L(\bar{x}(\bar{\lambda}, \bar{g}), \bar{\lambda}, \bar{h}) \leq L(\bar{x}^*, \bar{g}^*, \bar{h}^*) \leq L(\bar{x}, \bar{g}(\bar{x}), \bar{h}(\bar{x}))$$

A partir del resultado anterior, podemos formular dos problemas equivalentes al problema general (3.2) convexo:

Problema primal.

Hallar: $\bar{x} = \bar{x}^*$

que minimiza: $L_p(\bar{x}) = L(\bar{x}, \bar{\lambda}^g(\bar{x}), \bar{\lambda}^h(\bar{x}))$

verificando : $\bar{x} \in X$

donde: $\bar{\lambda}^g = \bar{\lambda}^g(\bar{x}), \bar{\lambda}^h = \bar{\lambda}^h(\bar{x})$ (3.59)

maximizan : $\vartheta(\bar{\lambda}^g, \bar{\lambda}^h) = L(\bar{x}, \bar{\lambda}^g, \bar{\lambda}^h)$

verificando : $\bar{\lambda} \in \Lambda$

Problema Dual.

Hallar: $\bar{\lambda}^g = \bar{\lambda}^{g*}, \bar{\lambda}^h = \bar{\lambda}^{h*}$

que maximizan: $L_d(\bar{\lambda}^g, \bar{\lambda}^h) = L(\bar{x}(\bar{\lambda}^g, \bar{\lambda}^h), \bar{\lambda}^g, \bar{\lambda}^h)$

verificando : $\bar{\lambda} \in \Lambda$

donde: $\bar{x} = \bar{x}(\bar{\lambda}^g, \bar{\lambda}^h)$ (3.60)

minimizan : $\vartheta(\bar{x}) = L(\bar{x}, \bar{\lambda}^g, \bar{\lambda}^h)$

verificando : $\bar{x} \in X$

Los problemas primal y dual son problemas de minimización con restricciones laterales en los que existen variables independientes, y variables dependientes que son, a su vez, solución de otros problemas de minimización con restricciones laterales en función de las primeras.

El valor del lagrangiano en la solución del problema coincide evidentemente con el valor de la función objetivo. Por ello, a la vista del resultado (3.58), la solución de ambos problemas mediante algún procedimiento de tipo iterativo

proporciona dos sucesiones de valores del lagrangiano que convergen superior e inferiormente al mínimo de la función objetivo, acotando por tanto su valor.

Para problemas de minimización con restricciones exclusivamente laterales las condiciones de Kuhn-Tucker se simplifican notablemente, y pueden obtenerse algoritmos iterativos de relativa simplicidad. Sin embargo, pese a que la solución de ambos problemas, tal como los hemos definido, es la solución del problema general convexo del que deriva, su tratamiento es complicado numéricamente, puesto que al evaluar sus correspondiente lagrangianos " L_p " y " L_d " para un cierto valor de sus respectivas variables independientes es preciso calcular, en principio, sus correspondientes variables dependientes. El problema interno de minimización que permite obtener las variables dependientes en función de las variables independientes, pese a la sencillez que implica su condición de problema de minimización con restricciones exclusivamente laterales, (y a su linealidad en el caso del problema primal), puede manifestar diversos aspectos de difícil tratamiento. De hecho puede tener soluciones de valor infinito, y las funciones " $\bar{x}(\bar{g}, \bar{h})$ ", " $\bar{\lambda}^g(x)$ " y " $\bar{\lambda}^h(x)$ " pueden tener discontinuidades y no estar definidas en parte de sus respectivos intervalos. Ello dificulta extraordinariamente el tratamiento numérico de estos problemas. No obstante, a partir de estos principios pueden derivarse algoritmos que permiten actualizar iterativamente los valores de las variables " \bar{x} ", y los multiplicadores " $\bar{\lambda}$ ", que denominaremos en lo sucesivo variables primales y variables duales, respectivamente.

Puesto que de hecho son equivalentes, denominaremos indistintamente problema primal al problema definido anteriormente (3.59) y al problema inicial convexo (3.2).

En lo que respecta al planteamiento dual (3.60), es sencillo demostrar que el gradiente de la función lagrangiano dual "Ld", respecto a sus variables independientes (esto es, las variables duales), tiene una expresión sencilla que coincide directamente con las funciones cambiadas de signo que definen las restricciones correspondientes, esto es:

$$\begin{aligned} \frac{\partial}{\partial \lambda^g} L_d(\bar{\lambda}^g, \bar{\lambda}^h) &= -g(x(\bar{\lambda}^g, \bar{\lambda}^h)) \\ \frac{\partial}{\partial \lambda^h} L_d(\bar{\lambda}^g, \bar{\lambda}^h) &= -h(x(\bar{\lambda}^g, \bar{\lambda}^h)) \end{aligned} \tag{3.61}$$

De forma similar pueden obtenerse expresiones de su hessiano, y de ambas puede hacerse uso extensivo en el desarrollo de algoritmos de solución del problema. Además, para evitar referencias a la función " $x(\bar{\lambda}^g, \bar{\lambda}^h)$ ", que puede no existir, es frecuente expresar el problema dual sustituyendo el problema de minimización interior por sus condiciones de Kuhn-Tucker. De esta forma, el problema dual puede expresarse como:

Problema DUAL

Hallar: $\bar{\lambda} = \bar{\lambda}^*$, $\bar{\lambda} = \bar{\lambda}^*$

que maximizan: $Ld(\bar{\lambda}, \bar{\lambda}) = L(\bar{x}, \bar{\lambda}, \bar{\lambda})$

verificando :

(3.62)

$$\bar{\lambda} \in \Lambda; \bar{x} \in X$$

$$\frac{\partial}{\partial x_i} L(\bar{x}, \bar{\lambda}, \bar{\lambda}) \Rightarrow \begin{cases} \geq 0 & ; \text{si } x_i = a_i \\ = 0 & ; \text{si } a_i < x_i < b_i \\ \leq 0 & ; \text{si } x_i = b_i \end{cases}$$

; i=1, ..., n

III.5.3 Métodos De Solución Del Problema

Es evidente la dificultad que conlleva el calcular directamente el conjunto de soluciones de las condiciones de Kuhn-Tucker y discriminar entre ellas el mínimo del problema. Este planteamiento analítico solo es aplicable a problemas de gran sencillez. Desde un punto de vista numérico no existe ningún método conocido en la actualidad que permita resolver eficientemente, y con total generalidad, el problema de minimización con restricciones cualesquiera (3.2); de todo lo que disponemos es de un conjunto de diversas técnicas cuya eficiencia es, en gran medida, dependiente del problema específico a que se apliquen. Además, algunas de las herramientas matemáticas disponibles para abordarlo, son únicamente aplicables a ciertas clases de problemas (tal es el caso de la dualidad, aplicable a problemas convexos).

Como ya hemos visto al exponer las condiciones de Kuhn-Tucker, la introducción de variables adicionales permite reducir un problema con restricciones cualesquiera a un problema con restricciones de igualdad. No obstante, este planteamiento conduce a problemas mal condicionados numéricamente. Por otro lado, los métodos de solución para problemas con restricciones de igualdad no son sino particularizaciones de métodos más generales para restricciones cualesquiera. Por ambos motivos, este tipo de tratamiento no suele emplearse en la práctica.

Como seguidamente expondremos, es posible desarrollar algoritmos de relativa simplicidad y eficiencia para ciertos problemas particulares de minimización con restricciones, tales como problemas con restricciones exclusivamente laterales, programación lineal y programación cuadrática, que configuran una primera familia de métodos de carácter específico.

Una segunda familia de métodos se basa en la generalización de los métodos de descenso para minimización sin restricciones de forma que permitan el tratamiento de restricciones de tipo general, y reciben la denominación de métodos de descenso generalizados o métodos de gradiente proyectado.

Una tercera familia de métodos se basa directamente en las condiciones de Kuhn-Tucker, ya sea a través de formulaciones duales, ya sea mediante la identificación de las restricciones activas y la reducción subsiguiente a problemas de minimización con restricciones de igualdad. Entre estos últimos tienen particular interés los denominados métodos de criterios de optimalidad, donde realizando hipótesis sobre las restricciones

que se encuentran activas en el óptimo, se sustituye el problema inicial de minimización por un problema diferente y más simple.

Por último, existe una cuarta familia de métodos que permiten reducir los problemas de minimización con restricciones de tipo general, a problemas de minimización no restringida, y una quinta familia que permite reducir los problemas de minimización con restricciones de tipo general a problemas particulares del mismo tipo para cuya solución se conocen algoritmos eficientes.

Por tanto, estructuraremos los métodos de solución del problema general con restricciones cualesquiera de la siguiente forma (Fig. 3.6):

- Métodos para problemas específicos de minimización con restricciones de desigualdad.
- Métodos de descenso generalizados.
- Métodos basados en las condiciones de Kunh-Tucker.
- Métodos de reducción a problemas de minimización no restringida.
- Métodos de reducción a problemas específicos de minimización con restricciones cualesquiera.

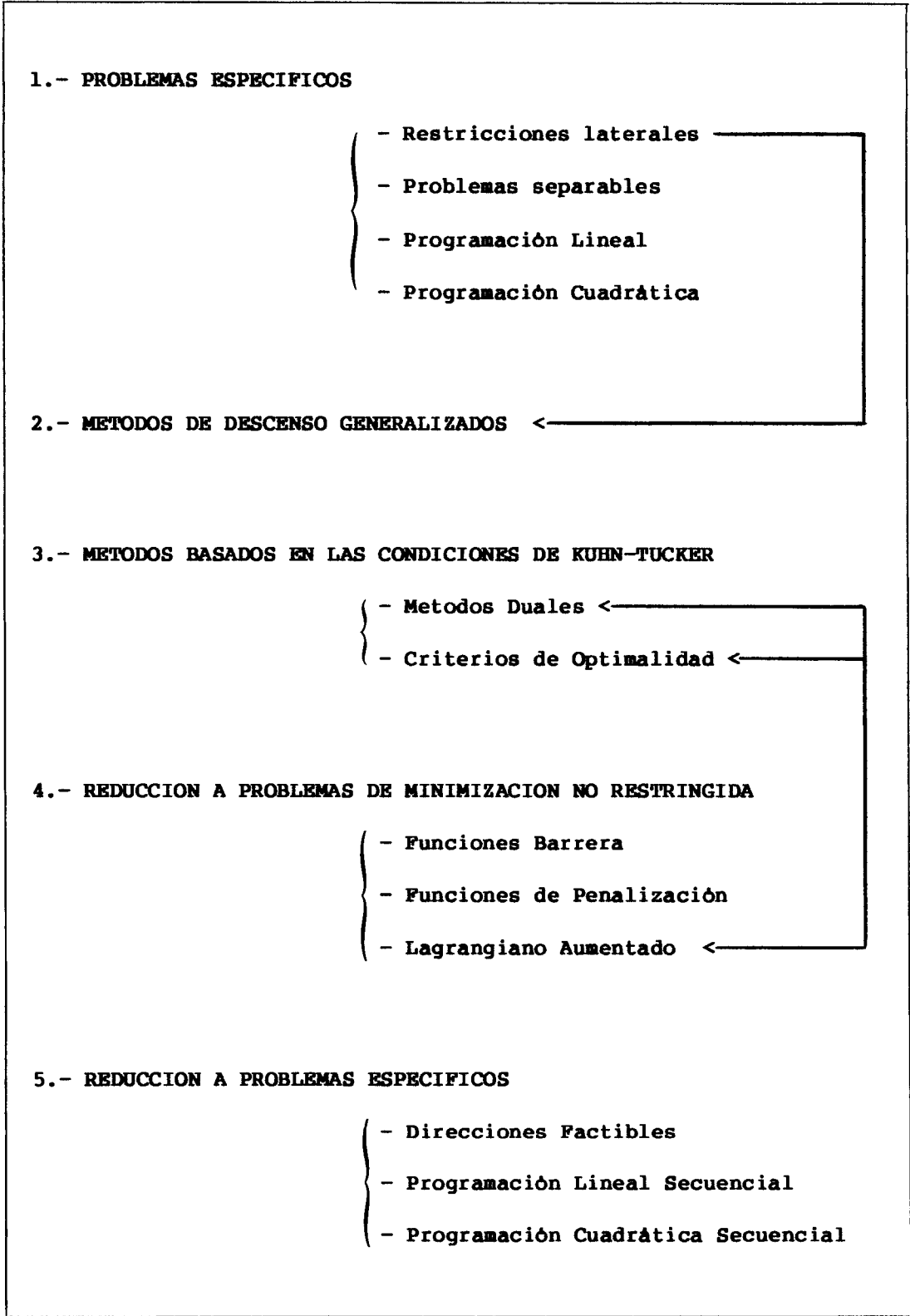


Figura 3.6.- Clasificación de los métodos de solución para problemas de minimización con restricciones, e interrelaciones entre los mismos.

III.5.3.1 Métodos para problemas específicos

Para ciertos problemas de minimización con restricciones en desigualdad, pueden desarrollarse algoritmos específicos de gran eficiencia. Nos referiremos en particular a los siguientes problemas:

- Problemas con restricciones exclusivamente laterales.
- Problemas separables.
- Problemas de Programación Lineal.
- Problemas de Programación Cuadrática.

Problemas con restricciones exclusivamente laterales.

Sea el problema de minimización con restricciones laterales:

$$\begin{aligned} \text{minimizar} & : f(\bar{x}) & \bar{x} & = \{x_i\} & ; i=1, \dots, n \\ & & & & \\ \text{verificando:} & a_i \leq x_i \leq b_i & & & ; i=1, \dots, n \end{aligned} \quad (3.63)$$

A partir de las condiciones de Kuhn-Tucker (3.52) particularizadas para este problema, pueden plantearse numerosos algoritmos iterativos, el más sencillo de los cuales puede escribirse en la forma:

$$\bar{x}^{k+1} = \bar{x}^k + \Theta \bar{s}^k$$

$$\bar{s}_i^k \Rightarrow \left\{ \begin{array}{l} = 0 \quad \text{si} \quad \bar{x}_i^k = a_i \quad \text{y} \quad \frac{\partial}{\partial x_i} f(\bar{x}^k) \geq 0 \\ = 0 \quad \text{si} \quad \bar{x}_i^k = b_i \quad \text{y} \quad \frac{\partial}{\partial x_i} f(\bar{x}^k) \leq 0 \\ = - \frac{\partial}{\partial x_i} f(\bar{x}^k) \quad \text{en otro caso} \end{array} \right.$$

$$\Theta \Rightarrow \left\{ \begin{array}{l} \text{minimiza } f(\bar{x}^k + \Theta \bar{s}^k) \\ \text{verificando } \Theta \leq \min(\Theta^-, \Theta^+) \end{array} \right. \quad (3.64)$$

con:

$$\Theta^- = \min_{\substack{k \\ s_i < 0}} \left\{ \frac{a_i - \bar{x}_i^k}{s_i} \right\}$$

$$\Theta^+ = \min_{\substack{k \\ s_i > 0}} \left\{ \frac{b_i - \bar{x}_i^k}{s_i} \right\}$$

donde la aproximación inicial a la solución debe ser factible.

El algoritmo anterior es similar al método de descenso de Newton de primer orden, pero incorpora además el control de las restricciones laterales (Fig. 3.7). De forma similar pueden derivarse algoritmos de orden superior.

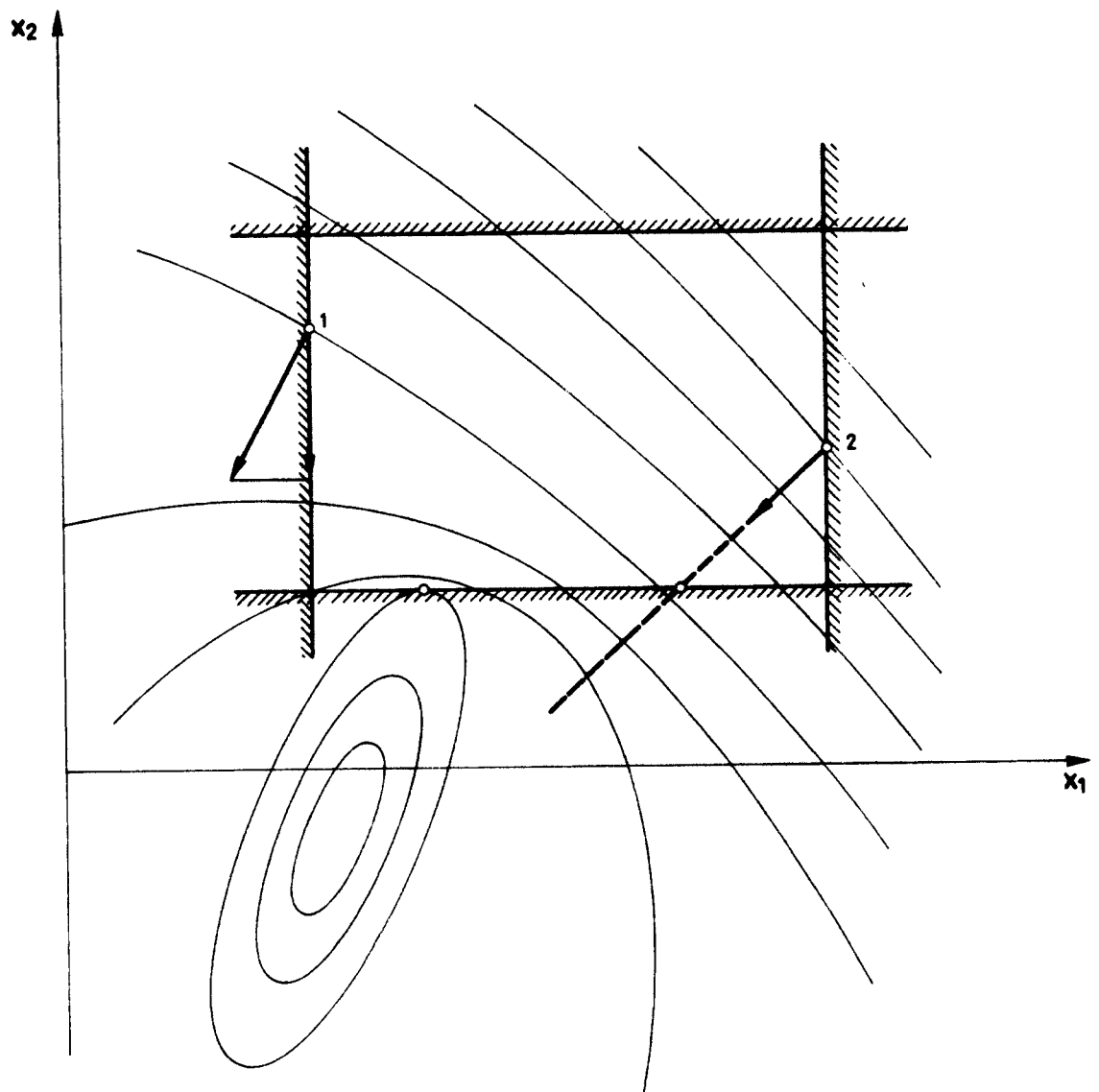


Figura 3.7.- Método de primer orden para problemas de minimización con restricciones exclusivamente laterales.

Se evita la violación de las restricciones proyectando la dirección de avance sobre las restricciones activas (1) y limitando el factor de avance (2).

Problemas Separables

El problema general de minimización con restricciones cualesquiera (3.2) se dice separable si se cumple:

$$\begin{aligned} f(\bar{x}) &= \sum_{i=1}^n f_i(x_i) && ; i=1, \dots, n \\ g_j(\bar{x}) &= \sum_{i=1}^n g_{ji}(x_i) && ; j=1, \dots, m \\ h_l(\bar{x}) &= \sum_{i=1}^n h_{li}(x_i) && ; l=1, \dots, p \end{aligned} \quad (3.65)$$

Un problema separable goza de dos propiedades importantes tanto desde el punto de vista analítico, como numérico: los hessianos de la función objetivo y las restricciones son diagonales, y si el problema es además convexo, en la formulación dual asociada el problema de minimización interno puede desacoplarse en "n" problemas de minimización unidireccionales independientes, uno por cada variable primal.

Habitualmente, un problema separable puede resolverse de forma más eficiente que un problema general, mediante métodos específicos o simplificaciones de métodos generales basadas en sus propiedades y peculiaridades, y que no detallaremos aquí debido a su gran diversidad y dependencia de la forma del problema.

Un caso particular de problema separable convexo, es el problema de programación lineal que expondremos a continuación.

Programación Lineal

El problema general de programación lineal puede escribirse en la forma:

$$\begin{aligned} \text{minimizar} \quad & f(\bar{x}) = \sum_{i=1}^n f_i x_i \quad ; i=1, \dots, n \\ \text{verificando:} \quad & g_j(\bar{x}) = \sum_{i=1}^n g_{ji} x_i - g_{j0} \leq 0 ; j=1, \dots, m \quad (3.66) \\ & h_l(\bar{x}) = \sum_{i=1}^n h_{li} x_i = 0 \quad ; l=1, \dots, p \end{aligned}$$

donde las restricciones laterales, si existen, se consideran englobadas en el conjunto de restricciones en desigualdad. El problema es obviamente convexo, y por tanto, todo mínimo local es también global. No obstante, para que el problema tenga solución única, es necesario que existan al menos tantas restricciones independientes como variables primales, y que la región factible no sea vacía.

En el problema anterior, si hay únicamente restricciones en desigualdad, cada una limita la posición de la solución en el hiperespacio n-dimensional a un semiespacio limitado por un hiperplano. La región factible consiste, por tanto, en el conjunto formado por los puntos situados en el interior y la frontera del hiperpoliedro convexo limitado por tales hiperplanos. Si existen únicamente restricciones de igualdad, la región factible está formada por la intersección de todos los hiperplanos correspondientes a tales restricciones. Si existen ambos tipos de restricciones simultáneamente, la región factible es la intersección de las regiones factibles correspondientes a

cada uno de los dos tipos de restricción.

Dado que la función objetivo es lineal, sus hipersuperficies de nivel son hiperplanos ortogonales al vector gradiente. En consecuencia, la solución del problema es el punto de la región factible cuyo hiperplano de nivel de la función objetivo corresponde al menor valor de ésta. Excepto en casos singulares la solución es única, y por tanto se trata de un vértice de la región factible. Además, el hiperplano de nivel de la función objetivo al que pertenece, es tangente al hiperpoliedro que configura la frontera de la región factible.

Ello sugiere que la solución del problema podría obtenerse por inspección simple, evaluando la función objetivo en todos los vértices del hiperpoliedro que limita la región factible. No obstante, el crecimiento del número de vértices con el número de variables primales y de restricciones no permite en general realizar esta operación excepto en el caso de problemas con muy pocas variables y restricciones.

El algoritmo denominado Método del Simplex permite resolver el problema general de programación lineal mediante una selección orientada de sucesivos vértices. El método converge en un número finito de pasos, y experimentalmente se ha demostrado que éste es notablemente inferior al número total de vértices. Su eficiencia es muy elevada, y tanto los tiempos de cálculo requeridos como la necesidad de almacenamiento en memoria son razonables. El Método del Simplex constituye una herramienta de extremada complejidad teórica, aunque su aplicación es sumamente sencilla, sobre la que existen tratados monográficos, y admite numerosas variantes y

diversas formas de evitar problemas de tipo operativo. Podemos considerarlo como una herramienta indispensable para resolver problemas generales de programación no lineal, ya que muchas de las técnicas de minimización con restricciones generales se basan en aproximaciones lineales. El algoritmo del Simplex fue desarrollado para resolver fundamentalmente problemas económicos. Por ello, está implícito en su formulación el que las variables primales sean positivas o nulas. No obstante, ello no constituye ningún problema en su aplicación a problemas cuyas variables no sean necesariamente positivas o nulas, ya que basta con una traslación de ejes adecuada para imponer la no negatividad de las mismas.

Programación Cuadrática

El problema general de programación cuadrática puede expresarse en la forma:

$$\text{minimizar : } f(\bar{x}) = \bar{b}^t \bar{x} + 1/2 \bar{x}^t \underline{A} \bar{x}$$

$$\text{con: } \bar{x} = \begin{Bmatrix} x \\ \vdots \\ x \end{Bmatrix} \quad ; i=1, \dots, n$$

\underline{A} , semidefinida positiva

$$\text{verificando: } g_j(\bar{x}) = \sum_{i=1}^n g_{ji} x_i - g_{j0} \leq 0 ; j=1, \dots, m \quad (3.67)$$

$$h_l(\bar{x}) = \sum_{i=1}^n h_{li} x_i = 0 \quad ; l=1, \dots, p$$

Para este problema existen diversos algoritmos iterativos de gran eficiencia, entre los más conocidos de los cuales se cuentan el método de Wolfe y sus derivados. El problema es obviamente convexo, y por tanto al igual que en el caso anterior,

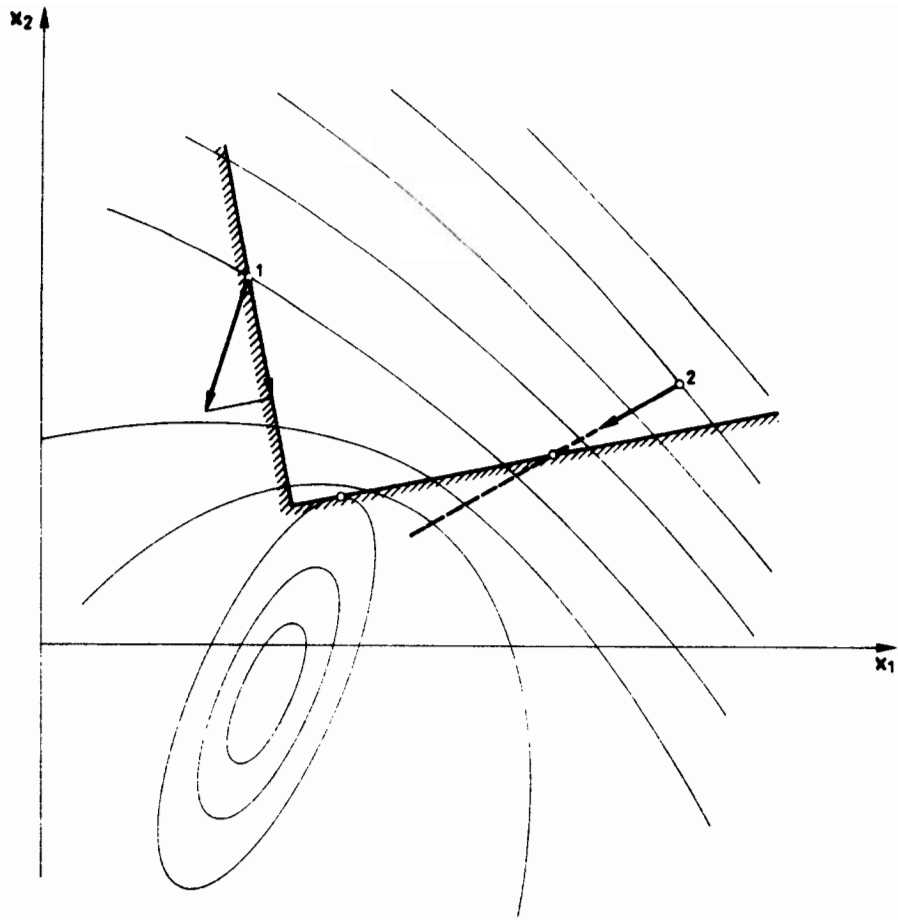
todo mínimo local es también global. A diferencia del problema de programación lineal, el problema de programación cuadrática tiene siempre solución, independientemente del número de restricciones, si la matriz "A" es semidefinida positiva, siempre y cuando la región factible no sea vacía.

III.5.3.2 Métodos de descenso generalizados

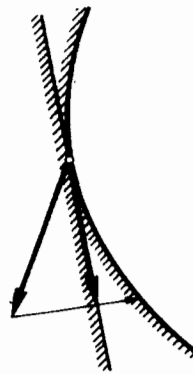
Los métodos de descenso para problemas de minimización no restringida han sido ya generalizados de alguna forma al desarrollar el algoritmo (3.64) para problemas con restricciones laterales. Si bien este algoritmo fue desarrollado a partir de las condiciones de Kuhn-Tucker, puede interpretarse su forma de trabajo como un método de descenso, en el cual si la dirección de descenso conduce al exterior de la región factible violando una determinada restricción lateral, se modifica la dirección proyectándola sobre la superficie de tal restricción, que en este caso es un hiperplano coordenado.

Esta idea, aplicada a restricciones en desigualdad, da lugar a los métodos de descenso generalizados o métodos de gradiente proyectados, desarrollados inicialmente para restricciones lineales (Fig. 3.8-a), pero que pueden modificarse adecuadamente para permitir el tratamiento de restricciones no lineales.

Sea el problema de minimización con restricciones lineales en desigualdad:



a)



b)

Figura 3.8.- Métodos de descenso generalizados para problemas de minimización con restricciones lineales en desigualdad, y fallo del algoritmo en su aplicación a problemas no lineales.

a) Se evita la violación de las restricciones proyectando la dirección de avance sobre las restricciones activas (1) y limitando el factor de avance (2).

b) Fallo del algoritmo en su aplicación a problemas con restricciones no lineales.

$$\begin{aligned}
\text{minimizar} & : f(\bar{x}) & ; \bar{x} = \{x_i\} ; i=1, \dots, n \\
\text{verificando} & : g_j(\bar{x}) = \sum_{i=1}^n g_{ji} x_i - g_{j0} \leq 0 & ; j=1, \dots, m
\end{aligned} \tag{3.68}$$

donde si existen restricciones laterales, pueden tratarse separadamente con sencillez y eficiencia, y no se considerarán aquí por simplicidad.

Planteemos un algoritmo de descenso del tipo (3.12), y definamos:

$$\begin{aligned}
\text{para } \bar{x} & = \bar{x}^k \\
Q & = \text{conjunto de restricciones activas} & (3.69) \\
N & = \{g_{ji}\} ; i=1, \dots, n ; j \in Q
\end{aligned}$$

Derivaremos el método para obtener la dirección de avance, desde un punto de vista global, similar al planteamiento (3.44) para minimización no restringida. Formularemos el problema en la forma:

$$\begin{aligned}
\text{Hallar: } \bar{s}^k & ; k \geq 1 \\
\bar{x}^{k+1} & = \bar{x}^k + \theta \bar{s}^k \\
\text{de forma que:} & & (3.70)
\end{aligned}$$

$$\bar{s}^k, \text{ minimiza } a(\bar{s}) = [\nabla f(\bar{x}^k) \quad \bar{s}^k]$$

$$\text{verificando: } [\bar{s}^t \quad \bar{M} \bar{s}^t] = 1$$

$$\sum_{i=1}^n g_{ji} \bar{s}_i \leq 0 \quad \forall j \in Q$$

siendo " \underline{M} " una matriz simétrica definida positiva.

El planteamiento anterior se interpreta inmediatamente como un intento de obtener el vector en cuya dirección el descenso de la función objetivo es más acusado, imponiendo una restricción en su norma (definida mediante la matriz " \underline{M} "), y de forma que no se violen las restricciones que se encuentran activas.

Formando el lagrangiano, y planteando las condiciones de Kuhn-Tucker para el problema anterior, se obtiene:

$$L(\underline{s}, \underline{\lambda}, \underline{\lambda}) = \nabla f(\underline{x}) \underline{s} - \lambda (1 - \underline{s}^t \underline{M} \underline{s}) - \underline{\lambda}^t \underline{N} \underline{s}$$

para: $\underline{s} = \underline{s}^k$, $\lambda = \lambda^m$, $\underline{\lambda} = \underline{\lambda}^q$

se verifica:

$$\begin{aligned} (\nabla f(\underline{x}) - \underline{\lambda}^t \underline{N}) + 2 \lambda \underline{M} \underline{s} &= \underline{0} & (3.71) \\ \underline{s}^t \underline{M} \underline{s} &= 1 ; \sum_{i=1}^n g_{ji} \underline{s}_i \leq 0 \quad \forall j \in Q \\ \underline{\lambda}^t \underline{N} \underline{s} &= 0 ; \underline{\lambda}_j \leq 0 \quad \forall j \in Q \end{aligned}$$

Combinando adecuadamente las relaciones anteriores, se obtiene:

$$\begin{aligned} \bar{s}^k &= - \frac{1}{2\lambda} \bar{M}^{-1} \bar{e}^k \\ \bar{\lambda}^q &= [\begin{matrix} (N & M & N) \\ \bar{q} & \bar{q} & \bar{q} \end{matrix} \quad \bar{N} \quad \bar{M} \quad \bar{N}] \quad \bar{\nabla} f(\bar{x}^k) \end{aligned} \tag{3.72}$$

$$\bar{\lambda}^m = 1/2 [\bar{e}^k \quad \bar{t} \quad \bar{M}^{-1} \quad \bar{e}^k \quad -1/2]$$

$$\text{donde, } \bar{e}^k = (\bar{\nabla} f(\bar{x}^k) - \bar{\lambda}^q \bar{t} \quad \bar{N})$$

Se observa que el multiplicador " λ^m " correspondiente a la norma del vector de descenso, afecta tan solo al módulo del vector de descenso " s^k ", pero no a su dirección. Puesto que posteriormente se realizará una minimización unidireccional, el módulo del vector es irrelevante.

Operando con las fórmulas anteriores, y adoptando un módulo arbitrario para el vector de descenso, se obtiene:

$$\begin{aligned} \bar{s}^k &= - P \bar{M}^{-1} \bar{\nabla} f(\bar{x}^k) \\ \text{con:} & \quad \bar{P} = I - \bar{M} \bar{N}^{-1} (\bar{N} \bar{M} \bar{N}) \bar{N}^{-1} \end{aligned} \tag{3.73}$$

siendo " P_q " un operador de proyección. La dirección de avance es la proyección del opuesto del vector producto de la inversa de la matriz " M " por el gradiente de la función objetivo.

Si alguno de los multiplicadores " λ_j^q " es positivo, el gradiente de la restricción correspondiente es negativo en la

dirección adoptada, y por tanto la restricción deja de ser activa. Este factor permite controlar en cada iteración las restricciones que dejan de ser activas al avanzar en una determinada dirección.

Se comprueba fácilmente que si el vector de avance obtenido es nulo, y todos los multiplicadores " λ_j^q " son no positivos, coinciden con los multiplicadores de Lagrange no nulos del problema original (3.68) que verifican sus condiciones de Kuhn-Tucker, siendo nulos todos los restantes por pertenecer a restricciones inactivas, y por tanto se habrá alcanzado el mínimo del problema.

En cambio, si el vector de avance es nulo, pero existe algún multiplicador positivo, debe eliminarse la restricción correspondiente del conjunto de restricciones activas, y calcular una nueva dirección.

Por otro lado, durante el proceso de minimización unidireccional es necesario verificar que las restricciones se cumplan, lo cual es sencillo dada su naturaleza lineal. Evidentemente, al finalizar el proceso de minimización unidireccional en cada iteración es preciso comprobar si se han activado nuevas restricciones, hecho que deberá ser tenido en cuenta en las operaciones posteriores.

Cuando la matriz " \tilde{M} " anterior es la matriz identidad " \tilde{I} ", el método obtenido es un método de máximo descenso generalizado, en el cual el operador " \tilde{P}_q " es un operador ortogonal que proyecta el gradiente de la función objetivo ortogonalmente a los gradientes de las restricciones activas sobre la intersección de

sus hiperplanos correspondientes. Cuando la matriz " \tilde{M} " es el hessiano de la función objetivo, se obtiene un método de descenso de Newton de segundo orden generalizado, y el operador " \tilde{P}_q " es un operador de proyección oblicuo. De forma similar pueden derivarse métodos Quasi-Newton generalizados.

Un aspecto importante de estos métodos es que las inversiones de los productos matriciales que es necesario realizar para construir los operadores de proyección pueden efectuarse directamente actualizando los operadores obtenidos en iteraciones previas, siendo necesario realizar la inversión completa únicamente en la primera iteración. El uso de procedimientos de actualización reduce fuertemente el coste operativo de estos métodos.

El algoritmo iterativo para resolver el problema escrito en la forma estándar (3.68), con restricciones lineales en desigualdad, puede escribirse finalmente en la forma general:

$$\underline{x}^{k+1} = \underline{x}^k + \theta \underline{s}^k$$

DIRECCION DE DESCENSO:

$Q =$ conjunto de restricciones activas en \underline{x}^k

$$N = \{g_{ji}\} \quad ; i=1, \dots, n \quad ; j \in Q$$

$$\underline{s}^k = -P_{-q}^{-1} M_{-q}^{-1} \nabla f(\underline{x}^k)$$

$$\underline{\lambda} = \left[\begin{array}{cc} (N_{-q} \ M_{-q} \ N_{-q}) & N_{-q} \end{array} \right]^{-1} \nabla f(\underline{x}^k)$$

$$P_{-q} = I_{-q} - M_{-q}^{-1} N_{-q} (N_{-q} \ M_{-q} \ N_{-q})^{-1} N_{-q}$$

si: $\underline{s}^k = \bar{0} \rightarrow$ si: $\left| \begin{array}{l} \lambda_j \leq 0 \forall j \in Q \implies \underline{x}^k = \text{OPTIMO} \\ \lambda_j > 0 \rightarrow \text{eliminar } j \text{ de } Q, \\ \text{y calcular nueva } \underline{s}^k \end{array} \right.$

(3.74)

si: $\underline{s}^k \neq \bar{0} \rightarrow$ minimización unidireccional

MINIMIZACION UNIDIRECCIONAL:

$$\theta \text{ minimiza } : f(\underline{x}^k + \theta \underline{s}^k)$$

$$\text{verificando: } 0 \leq \theta \leq \theta_{\max}$$

$$\text{con: } \theta = \min_{\max} \left\{ \frac{g_{j0} - \sum_{i=1}^n g_{ji} x_i^k}{d_j} \mid d_j > 0; j \notin Q \right\}$$

$$d_j = \sum_{i=1}^n g_{ji} s_i^k$$

en función de la matriz "M" simétrica y definida positiva.

Evidentemente, en la aplicación de este algoritmo, es preciso partir de una aproximación inicial factible. El algoritmo asegura que las sucesivas aproximaciones a la solución que se obtienen son factibles, y los correspondientes valores de la función objetivo son decrecientes.

El algoritmo anterior puede aplicarse a problemas con restricciones en desigualdad no lineales adoptando ciertas precauciones. En este caso, al proyectar el gradiente sobre los hiperplanos tangentes a las restricciones activas para obtener la dirección de descenso, no hay garantía alguna de que tal dirección esté parcialmente contenida en la región factible. Por tanto, es probable que para cualquier valor del factor de avance distinto de cero se viole alguna restricción. Además, si se realiza una minimización unidireccional, teniendo en cuenta las aproximaciones lineales de las restricciones, y no sus verdaderas expresiones, la nueva aproximación a la solución puede ser no factible (Fig. 3.8-b). Estos inconvenientes pueden evitarse sumando a la aproximación obtenida una combinación lineal de los gradientes de las restricciones violadas, de forma que la aproximación así modificada sea factible, si bien no existe ningún procedimiento general que permita obtener una tal combinación, y ésta debe hallarse mediante métodos intuitivos. No obstante, en problemas fuertemente no lineales, con muchas restricciones activas en el mínimo, el problema puede revestir una gran complejidad, y de hecho, llegar a ser virtualmente intratable mediante este tipo de algoritmos.

III.5.3.3 Métodos basados en las condiciones de Kuhn-Tucker.

Criterios de optimalidad

A partir de las condiciones de Kuhn-Tucker, pueden derivarse numerosos algoritmos. Podemos considerar como métodos derivados de las condiciones de Kuhn-Tucker los siguientes:

- Métodos Duales para problemas convexos.
- Métodos basados en la identificación de restricciones activas.

La resolución del problema dual ha sido tratada previamente desde un punto de vista analítico. Desde un punto de vista numérico existen numerosos algoritmos de solución del problema dual que suelen denominarse métodos de multiplicadores. Los métodos de multiplicadores permiten reducir el problema de minimización con restricciones cualesquiera a problemas de minimización no restringida, o secuencias de problemas de este tipo. Por ello introduciremos estos métodos, aunque desde una perspectiva distinta, al tratar los métodos de reducción a problemas de minimización sin restricciones.

La identificación del conjunto de restricciones activas en el óptimo, permite reducir el problema de minimización con restricciones cualesquiera a un problema de minimización con restricciones de igualdad. De lo ya expuesto, se deduce que la ventaja de este tratamiento no es sustancial a menos que el número de restricciones activas sea muy inferior al número total de restricciones del problema. Existen diversas técnicas basadas en la estimación del valor de los multiplicadores que permiten discriminar las restricciones activas de las inactivas, en intima

relación con los métodos duales. Por ello consideraremos a estas técnicas englobadas dentro de los métodos de multiplicadores, a los que haremos referencia posteriormente.

Sin embargo, si se formulan ciertas hipótesis sobre la identificación de las restricciones activas en conexión con la propia naturaleza del problema, es posible obtener ciertas relaciones entre las variables que deben verificarse en el óptimo, si las hipótesis de partida son ciertas. De hecho, se prescinde entonces del problema de minimización, y se sustituye por un problema distinto, que en general se materializará en un sistema de ecuaciones no lineales. Denominaremos a estos métodos, Criterios de Optimalidad, puesto que de hecho, a partir de ciertas hipótesis sobre las restricciones que condicionan el óptimo, se obtienen criterios que permiten obtener el óptimo si las hipótesis son ciertas. La denominación de criterios de optimalidad es particularmente indicada para el caso de optimización en diseño estructural, puesto que los criterios de optimalidad se plasman entonces en criterios de diseño.

Criterios de Optimalidad

Para facilitar la comprensión del concepto de Criterio de Optimalidad formularemos un caso de gran sencillez.

Sea un problema de minimización en la forma estándar (3.1).

Formulemos la hipótesis de que en el óptimo solo es activa la primera restricción. Si la hipótesis anterior es correcta, las restricciones restantes adquieren un valor negativo en el óptimo y por tanto, de las condiciones de Kuhn-Tucker (3.52) se

infiere la anulación de sus correspondientes multiplicadores, y las condiciones se reducen a:

en $\bar{x} = \bar{x}^*$ se verifica:

$$\frac{\partial f(\bar{x})}{\partial x_i} - \lambda_1 \frac{\partial g_1(\bar{x})}{\partial x_i} = 0 \quad ; i=1, \dots, n \quad (3.75)$$

$$g_1(\bar{x}) = 0 \quad ; \lambda_1 < 0$$

siendo necesario para que la hipótesis sea correcta que se cumpla:

$$g_j(\bar{x}^*) < 0 \quad ; j=2, \dots, m \quad (3.76)$$

Si los componentes del gradiente de la función objetivo no se anulan en el óptimo, las condiciones (3.75) pueden escribirse en la forma:

en $\bar{x} = \bar{x}^*$ se verifica:

$$\frac{\frac{\partial g_1(\bar{x})}{\partial x_i}}{\frac{\partial f(\bar{x})}{\partial x_i}} = \lambda_1^{-1} < 0 \quad ; i=1, \dots, n \quad (3.77)$$

$$g_1(\bar{x}) = 0$$

Es importante considerar que para ciertos problemas específicos, las relaciones anteriores pueden resolverse explícitamente, o cuanto menos simplificarse notablemente; si la función objetivo es lineal, y la restricción es cuadrática, se

comprueba inmediatamente que el problema se reduce a resolver una ecuación de segundo grado en el inverso del multiplicador; si el problema inicial es separable, se obtiene un conjunto de "n" ecuaciones no lineales desacopladas, una por variable primal, que con la condición de anulación de la restricción, permiten resolver el problema.

Las condiciones anteriores indican que en problemas de minimización con restricciones en desigualdad en los cuales una restricción es predominante y determina el óptimo, en éste los gradientes de la función objetivo y de la restricción citada son paralelos y de sentido contrario. Para ciertos problemas de optimización estructural, en que la función objetivo es el peso estructural, lineal con las variables de diseño, y la restricción predominante es igualmente lineal, las relaciones anteriores implican que la distribución de la densidad de energía es uniforme en el óptimo, siendo éste el criterio de optimalidad. El problema inicial de minimización del peso estructural puede sustituirse por otro problema, consistente en obtener un diseño para el cual la densidad de energía está uniformemente distribuida.

Para el caso en que existan simultáneamente varias restricciones activas en el óptimo, se obtiene a partir de las condiciones de Kuhn-Tucker un sistema de ecuaciones no lineales. Para ciertos problemas específicos, es posible obtener algoritmos iterativos sencillos que conduzcan a la solución de tales sistemas.

Evidentemente, la desventaja principal de la aplicación de

criterios de optimalidad consiste en la dificultad de estimar a priori el conjunto de restricciones activas. Además, para ofrecer ventajas sustanciales, la naturaleza del problema ha de ser tal que permita obtener expresiones iterativas sencillas que conduzcan a la solución del sistema no lineal de ecuaciones de forma eficiente, o interpretar de forma adecuada el significado real de tal sistema de forma que el criterio de optimalidad indique la estrategia a seguir.

El diseño a máxima tensión como Criterio de Optimalidad

Puede considerarse el método de diseño a máxima tensión como un criterio de optimalidad, en el cual se formula la hipótesis de que son activas en el óptimo las restricciones en tensión máxima. Para aclarar este concepto, formulemos el problema de optimización de estructuras de barras articuladas siguiente:

$$\begin{aligned}
 \text{minimizar} & : W(\bar{x}) = \sum_{i=1}^n L_i x_i \\
 \text{verificando} & : s_i(\bar{x}) \leq T_i \\
 & \quad - s_i(\bar{x}) \leq c_i
 \end{aligned}
 \tag{3.78}$$

siendo : i = número de barra ; $i=1, \dots, n$

- x_i = sección. L_i = longitud
- s_i = tensión.
- T_i = tensión máxima a tracción
- c_i = tensión máxima a compresión
- $W(\bar{x})$ = peso estructural

donde no consideramos restricciones laterales por simplicidad, y suponemos que la estructura se encuentra sometida a un solo estado de carga.

Las condiciones de Kuhn-Tucker para el problema anterior, se reducen a:

$$\text{para: } \bar{x} = \bar{x}^* \text{, existen: } \bar{\lambda}_\tau = \bar{\lambda}_\tau^* \text{, } \bar{\lambda}_c = \bar{\lambda}_c^*$$

verificando:

$$L - \sum_{j=1}^n (\lambda_{\tau j} - \lambda_{c j}) \frac{\partial s_j(\bar{x})}{\partial x_i} = 0 \quad ; i=1, \dots, n \quad (3.79)$$

$$\lambda_{\tau j} (s_j(\bar{x}) - \tau_j) = 0 \quad ; \lambda_{\tau j} \leq 0 \quad ; s_j(\bar{x}) \leq \tau_j \quad ; j=1, \dots, n$$

$$\lambda_{c j} (s_j(\bar{x}) + c_j) = 0 \quad ; \lambda_{c j} \leq 0 \quad ; s_j(\bar{x}) \geq -c_j \quad ; j=1, \dots, n$$

siendo " $\bar{\lambda}_\tau$ " y " $\bar{\lambda}_c$ " los multiplicadores de las restricciones en tracción y compresión respectivamente.

Formulemos la siguiente hipótesis:

$$\begin{aligned} \lambda_{\tau j} &= 0 \quad \forall j \in C \\ \lambda_{c j} &= 0 \quad \forall j \in T \end{aligned} \quad (3.80)$$

$$C \cup T = \{1, \dots, n\}$$

$$C \cap T = \emptyset$$

siendo "T" y "C" dos conjuntos complementarios, cuya unión es el conjunto de todas las barras. Evidentemente, la hipótesis anterior equivale a presuponer que todas las barras están traccionadas o comprimidas a sus tensiones máximas respectivas, y que se conoce el estado de cada barra.

Introduciendo esta hipótesis en parte de las condiciones de Kuhn-Tucker obtenemos:

$$\text{para: } \bar{x} = \bar{x}^*$$

$$s_j(\bar{x}) = \tau_j \quad ; j \in T \quad (3.81)$$

$$s_j(\bar{x}) = -c_j \quad ; j \in C$$

sistema de "n" ecuaciones con "n" incógnitas, no lineal en general, para cuya resolución ha de emplearse un algoritmo iterativo, y que no es sino la aplicación del método de diseño a máxima tensión. Resolviendo el sistema anterior mediante el método de Newton para sistemas de ecuaciones no lineales, obtenemos el algoritmo iterativo siguiente:

$$x_j^{k+1} = x_j^k + \frac{\tau_j - s_j(\bar{x}^k)}{\partial s_j(\bar{x}^k) / \partial x_j} \quad j \in T \quad (3.82)$$

$$x_j^{k+1} = x_j^k - \frac{c_j + s_j(\bar{x}^k)}{\partial s_j(\bar{x}^k) / \partial x_j} \quad j \in C$$

No obstante, una solución del sistema (3.81) no verifica

forzosamente las restantes condiciones de Kuhn-Tucker, esto es:

$$\text{para: } \bar{x} = \bar{x}^{\bullet}, \text{ existen: } \bar{\lambda}_{\tau} = \bar{\lambda}_{\tau}^{\bullet}, \bar{\lambda}_{c} = \bar{\lambda}_{c}^{\bullet}$$

verificando:

$$L - \sum_i \lambda_{\tau j} \frac{\partial}{\partial x_j} s_i(\bar{x}) + \sum_{j \in C} \lambda_{c j} \frac{\partial}{\partial x_j} s_i(\bar{x}) = 0 \quad (3.83)$$

$; i=1, \dots, n$

$$\lambda_{\tau j} \leq 0 \quad ; j \in T$$

$$\lambda_{c j} \leq 0 \quad ; j \in C$$

Tan sólo es posible probar que las condiciones (3.83) se satisfacen inmediatamente para el diseño a máxima tensión (3.81) para el caso de estructuras isostáticas. En efecto, en este caso los esfuerzos en las barras son independientes de las propiedades mecánicas y resistentes de las barras. Así pues, puede calcularse la estructura y averiguarse cuales están traccionadas y cuales comprimidas, independientemente de las secciones de las mismas. Además, una vez calculados los esfuerzos, la tensión en cada barra solo depende del esfuerzo axil en la misma y de su sección, esto es:

$$s_j(\bar{x}) = N_j / x_j \quad j=1, \dots, n$$

$$N_j = \text{esfuerzo axil } \begin{cases} > 0 \text{ si } j \in T \\ < 0 \text{ si } j \in C \end{cases} \quad (3.84)$$

Evidentemente, en este caso las ecuaciones (3.81) se resuelven de forma trivial:

$$\begin{aligned}
 x_j &= N_j / \tau_j \quad \text{si } N_j \geq 0 \\
 x_j &= -N_j / c_j \quad \text{si } N_j < 0
 \end{aligned}
 \tag{3.85}$$

y las condiciones (3.83) se verifican de forma automática, puesto que los multiplicadores no nulos pueden obtenerse explícitamente en la forma:

$$\begin{aligned}
 \lambda_{\tau_j} &= \frac{L_j}{\partial s_j / \partial x_j} \quad j \in T \\
 \lambda_{c_j} &= - \frac{L_j}{\partial s_j / \partial x_j} \quad j \in C
 \end{aligned}
 \tag{3.86}$$

y en virtud de (3.84) son forzosamente negativos o nulos.

De lo anterior, se deduce que para estructuras articuladas isostáticas sometidas a un solo estado de carga, con restricciones en tensión y utilizando como variables de diseño las secciones transversales de las barras, el diseño a máxima tensión verifica las condiciones necesarias de mínimo de Kuhn-Tucker. Se trata efectivamente del diseño de peso mínimo que verifica las restricciones. Este análisis no aporta ningún elemento nuevo, puesto que para un caso de tal sencillez el resultado es evidente sin necesidad de aplicar las condiciones de Kuhn-Tucker, pero indica con claridad por qué el método de diseño a máxima tensión no proporciona necesariamente diseños óptimos para otros tipos de estructuras, hiperestáticas o continuas, con múltiples estados de carga, u otro tipo de restricciones. Es además muy ilustrativo acerca de la forma de desarrollar

criterios de optimalidad.

III.5.3.4 Métodos de reducción a problemas de minimización no restringida

La reducción a problemas de minimización sin restricciones de un problema general de programación matemática puede realizarse a través de los métodos siguientes:

- Método de funciones Barrera.
- Método de funciones de Penalización.
- Métodos de lagrangiano aumentado.

Método de funciones Barrera

Sea un problema general de programación matemática expresado en la forma estándar (3.1).

Una función " $B(\bar{g})$ " se denomina función Barrera para el problema anterior si se verifica:

$$\begin{aligned}
 B(\bar{g}) &\geq 0 && \forall \bar{g} \mid g_j \leq 0 ; j=1, \dots, m \\
 \frac{\partial B(\bar{g})}{\partial g_k} &\geq 0 && \forall \bar{g} \mid g_j \leq 0 ; j=1, \dots, m \\
 &&& ; k=1, \dots, m \\
 B(\bar{g}) &\rightarrow +\infty && \text{cuando } g_j \rightarrow 0 \text{ para algún } j
 \end{aligned}
 \tag{3.87}$$

Una función barrera, por tanto, está definida solamente en la región factible, y alcanza un valor infinito en la frontera de la misma.

En particular, son frecuentes las dos siguientes elecciones de la función barrera:

$$B(\bar{g}) = \sum_{j=1}^m b_j(g_j)$$

con: $b_j(g_j) = -\ln(-g_j)$;(f. b. logarítmica)

$b_j(g_j) = -1/g_j$;(f. b. inversa)

(3.88)

Definamos el problema de minimización no restringida, dependiente del parámetro "w", siguiente:

$$\text{minimizar: } \phi(\bar{x}, w) = f(\bar{x}) + w B(\bar{g}(\bar{x})) \quad ; w > 0 \quad (3.89)$$

Bajo ciertas condiciones favorables, dependientes del problema inicial, se cumple:

$$\text{si } \lim_{k \rightarrow \infty} \{w^k\} = 0, \quad \text{con } w^{k+1} < w^k$$

y \bar{x}^k es solución de (3.89) para $w = w^k$

(3.90)

$$\text{entonces: } \lim_{k \rightarrow \infty} \{\bar{x}^k\} = \bar{x}^*$$

donde \bar{x}^* es solución del problema inicial (3.1)

es decir, que si se construye una sucesión de problemas de minimización sin restricciones del tipo (3.89) con parámetros "w" decrecientes, se espera que la sucesión de soluciones de tales problemas converja a la solución del problema inicial de minimización con restricciones.

La denominación de función barrera alude claramente a la

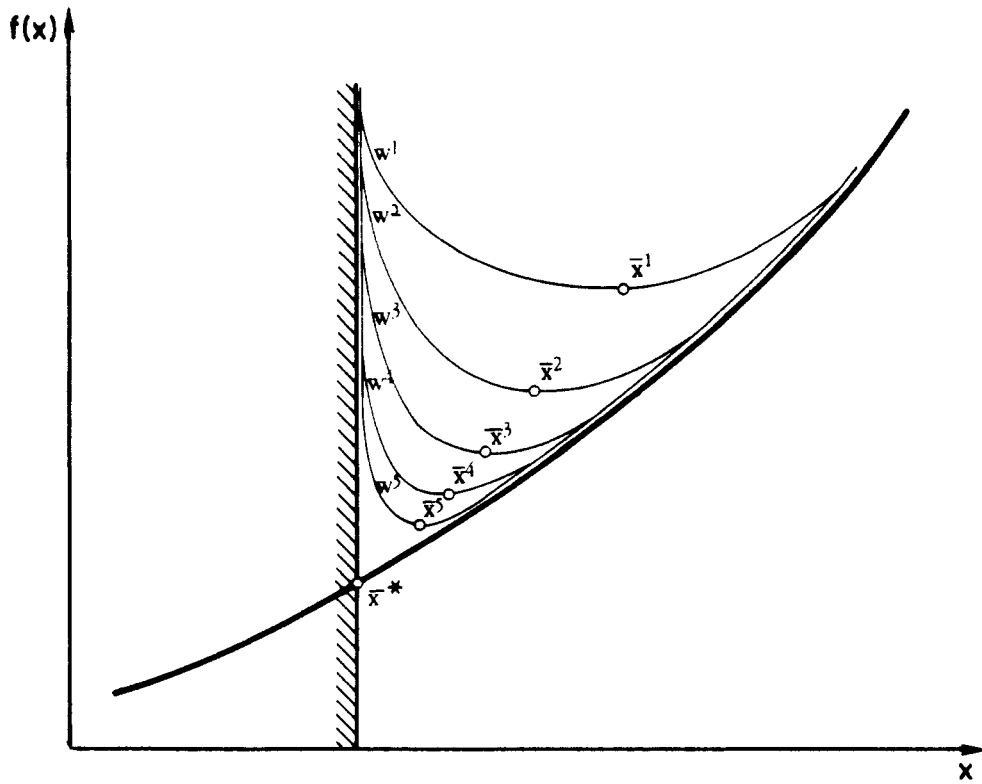
modificación que introduce en la función objetivo de forma que se satisfagan las restricciones (Fig. 3.9-a).

Este método tiene la gran ventaja, de que todas las soluciones de la sucesión " $\{\bar{x}^{-k}\}$ " son factibles, dada la naturaleza del problema (3.89).

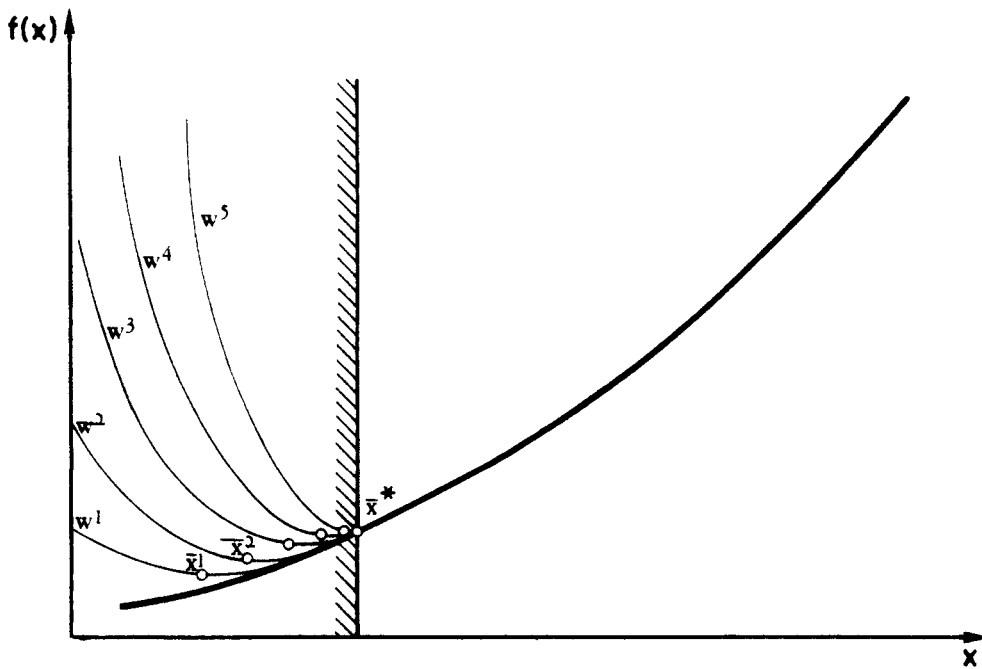
Por contra, ofrece las siguientes desventajas:

- El método no es válido si hay restricciones de igualdad, aunque estén encubiertas en la formulación estándar a través de dos restricciones en desigualdad cada una, puesto que en la frontera de la región factible, la función " $\phi(\bar{x}, w)$ " siempre alcanza un valor infinito.
- Para resolver cada problema de minimización sin restricciones del tipo (3.89) mediante un algoritmo iterativo, es necesario partir de una aproximación inicial factible, puesto que fuera de la región factible el comportamiento de la función " $\phi(\bar{x}, w)$ " es incierto, y en la frontera de ésta, la función alcanza un valor infinito.
- Al adoptar parámetros " w " cada vez más pequeños, se mal condiciona el hessiano de la función " $\phi(\bar{x}, w)$ " a minimizar.

Parte de estas desventajas pueden paliarse empleando funciones barrera extendidas, de forma que su rango de definición se amplie al exterior de la región factible, renunciando a la condición de que el valor de la función barrera en la frontera sea infinito, esto es:



a)



b)

Figura 3.9.- Interpretación gráfica de los métodos de funciones Barrera y los métodos de funciones de Penalización.

a) Método de funciones Barrera. Secuencia de soluciones aproximadas factibles.

b) Método de funciones de Penalización. Secuencia de soluciones aproximadas no factibles.

$$B(\bar{g}) = \sum_{j=1}^m b_j^e(g_j)$$

con: $b_j^e(g_j) = b_j(g_j) \quad ; \text{ si } g_j \leq -\epsilon$

$b_j^e(g_j) = e_j(g_j, \epsilon) \quad ; \text{ si } g_j > -\epsilon$

(3.91)

; j=1, ..., m

verificando: $e_j(g_j, \epsilon) = b_j(g_j) \quad ; \text{ si } g_j = \epsilon$

$e_j(g_j, \epsilon)$ creciente en g_j si $g_j > -\epsilon$

$\epsilon > 0$

; j=1, ..., m

siendo las funciones " b_j ", funciones barrera cualesquiera para cada restricción.

Método de funciones de Penalización

Sea un problema general de programación matemática expresado en la forma estándar (3.1).

Una función " $P(\bar{g})$ " se denomina función de Penalización para el problema anterior si se verifica:

$$P(\bar{g}) = 0 \quad \forall \bar{g} \mid g_j \leq 0 \quad ; j=1, \dots, m$$

$$P(\bar{g}) > 0 \quad \forall \bar{g} \mid g_j > 0 \text{ para algún } j \quad (3.92)$$

$$\frac{\partial}{\partial g_k} P(\bar{g}) > 0 \quad \forall \bar{g} \mid g_j > 0 \text{ para algún } j$$

; k=1, ..., m

Una función de penalización, por tanto, está definida en todo el espacio, anulándose en el interior y la frontera de la

región factible, y creciendo a medida que se incumplen las restricciones.

En particular, son frecuentes las dos elecciones siguientes de la función de penalización:

$$P(\bar{g}) = \sum_{j=1}^m p_j(g_j)$$

con: $p_j(g_j) = (\max\{0, g_j\})^2$; (f. de p. cuadrática) (3.93)

$p_j(g_j) = |\max\{0, g_j\}|$; (f. de p. de Zangwill)

Definamos el problema de minimización no restringida, dependiente del parámetro "w", siguiente:

$$\text{minimizar: } \zeta(\bar{x}, w) = f(\bar{x}) + 1/w P(\bar{g}(\bar{x})) \quad ; w > 0 \quad (3.94)$$

Bajo ciertas condiciones favorables, dependientes del problema inicial, se cumple:

$$\text{si } \lim_{k \rightarrow \infty} \{w^k\} = 0, \quad \text{con } w^{k+1} < w^k$$

y \bar{x}^k es solución de (3.94) para $w = w^k$ (3.95)

$$\text{entonces: } \lim_{k \rightarrow \infty} \{\bar{x}^k\} = \bar{x}^*$$

donde \bar{x}^* es solución del problema inicial (3.1)

es decir, que si se construye una sucesión de problemas de minimización sin restricciones del tipo (3.94) con parámetros "w" decrecientes, se espera que la sucesión de soluciones de tales problemas converja a la solución del problema inicial de

minimización con restricciones.

La denominación de función de penalización alude claramente a la modificación que introduce en la función objetivo de forma que las restricciones se incumplan lo menos posible (Fig. 3.9-b).

Este método tiene dos ventajas fundamentalmente frente al de funciones barrera:

- en la resolución de cada problema de minimización sin restricciones no es necesario partir de una aproximación inicial factible.
- El método es válido si hay restricciones de igualdad, y de hecho fue inicialmente desarrollado para este tipo de problemas. En este caso, sustituyendo cada restricción de igualdad por dos restricciones en desigualdad, se comprueba que los términos de la función de penalización de ambas pueden agruparse en uno solo.

Por contra, ofrece las siguientes desventajas:

- Las soluciones de la sucesión " $\{\bar{x}^k\}$ " son generalmente no factibles, dada la naturaleza del problema (3.94).
- Al adoptar parámetros " w " cada vez más pequeños, se mal condiciona el hessiano de la función " $\zeta(\bar{x}, w)$ " a minimizar.

Métodos de lagrangiano aumentado.

Los métodos de lagrangiano aumentado, o métodos de multiplicadores, fueron derivados inicialmente a partir de los métodos de penalización, que a su vez fueron concebidos en principio para problemas con restricciones de igualdad. No obstante, puede extenderse su aplicación a problemas de tipo general con restricciones cualesquiera, como expondremos seguidamente.

Sea un problema general de minimización con restricciones

exclusivamente de igualdad, del tipo (3.45).

Obviamente, puede formularse un problema equivalente al anterior en la forma:

$$\begin{aligned}
 \text{minimizar} & : f(\bar{x}) + 1/w A(\bar{h}(\bar{x})) ; \bar{x} = \{x_i\}; i=1, \dots, n \\
 & ; w > 0 \\
 \text{verificando: } & \bar{h}(\bar{x}) = \bar{0} \quad ; \bar{h}(\bar{x}) = \{h_l(\bar{x})\} \quad ; l=1, \dots, p \\
 & ; p \leq n
 \end{aligned}
 \tag{3.96}$$

donde la función $A(\bar{h})$ verifica:

$$\bar{h} = \bar{0} \Rightarrow \begin{cases} A(\bar{h}) = 0 \\ \nabla A(\bar{h}) = \bar{0} \end{cases}$$

$$\frac{\partial^2}{\partial \bar{v}} A(\bar{h}) \quad \bar{v} \geq 0 \quad ; \forall \bar{h}, \bar{v}$$

siendo habitual la siguiente elección para la función "A(h)":

$$A(\bar{h}) = \bar{h}^t \bar{h}
 \tag{3.97}$$

De hecho, si se prescinde de la condición de anulación de las restricciones de igualdad, y se resuelve el problema (3.96) para una secuencia de valores decrecientes del parámetro "w", lo que obtenemos es la aplicación del método de penalización al problema general de minimización con restricciones de igualdad (3.45). Sin embargo, el método de multiplicadores resuelve el problema de forma diferente, evitando el mal condicionamiento del hessiano para valores pequeños del parámetro "w".

Denominaremos lagrangiano aumentado del problema de minimización con restricciones de igualdad, al lagrangiano del problema equivalente (3.96). Esto es:

$$\chi(\bar{x}, \bar{\lambda}, w) = L(\bar{x}, \bar{\lambda}) + 1/w \quad A(\bar{h}(\bar{x})) \quad ; w > 0 \quad (3.98)$$

siendo: $L(\bar{x}, \bar{\lambda}) = f(\bar{x}) - \bar{\lambda}^t \bar{h}(\bar{x})$,

el lagrangiano (3.47)

y evidentemente, por la equivalencia de ambos problemas, la aplicación de las condiciones de extremo (3.48) al lagrangiano aumentado " $\chi(\bar{x}, \bar{\lambda}, w)$ ", da lugar a un sistema equivalente al obtenido para el lagrangiano " $L(\bar{x}, \bar{\lambda})$ ". Igualmente se comprueba, sin más que aplicar la formulación del problema primal (3.59), que la minimización sin restricciones del lagrangiano aumentado, particularizado para el valor de los multiplicadores que verifican tales condiciones, " $\bar{\lambda}^*$ ", respecto a las variables primales " \bar{x} ", conduce a la solución del problema original independientemente del valor del parámetro " w ". Ello sugiere que al resolver una sucesión de tales problemas con valores de los multiplicadores que converjan a los " $\bar{\lambda}^*$ " obteniendo la secuencia de soluciones correspondiente, es razonable esperar su convergencia a la solución del problema original. Esto es:

$$\text{si } \lim_{k \rightarrow \infty} \{ \bar{\lambda}^k \} = \bar{\lambda}^*$$

y \bar{x}^k es la solución del problema:

$$\left\{ \begin{array}{l} \text{minimizar: } \chi(\bar{x}, \bar{\lambda}, w) \\ \text{con } \bar{\lambda} = \bar{\lambda}^k ; w > 0 \end{array} \right. \quad (3.99)$$

$$\text{entonces: } \lim_{k \rightarrow \infty} \{ \bar{x}^k \} = \bar{x}^* ; \forall w > 0$$

donde \bar{x}^* es solución del problema original (3.45)

y dado que no es necesario adoptar valores excesivamente pequeños

para el parámetro "w", se evita el mencionado mal condicionamiento del problema.

El problema original puede resolverse mediante el tratamiento (3.99) si se dispone de una sucesión de valores de los multiplicadores que verifiquen las condiciones requeridas. Su obtención es trivial, dado que comparando la condición de anulación del gradiente del lagrangiano aumentado (3.98), con su equivalente (3.48) para el lagrangiano, se obtiene de forma inmediata una fórmula de actualización de los multiplicadores que podemos escribir:

$$\bar{\lambda}^{k+1} = \bar{\lambda}^k - \frac{2}{w^k} \bar{h}(\bar{x}^k) \quad (3.100)$$

donde el parámetro " w^k " puede modificarse en cada iteración de forma conveniente.

Además, la fórmula (3.100) puede escribirse en función del gradiente de la función lagrangiano dual (3.61) para el problema (3.45), en la forma:

$$\bar{\lambda}^{k+1} = \bar{\lambda}^k + \frac{2}{w^k} \frac{\partial L_d}{\partial \lambda}(\bar{\lambda}^k) \quad (3.101)$$

Se deduce que el método (3.99), con la actualización (3.101) de los multiplicadores, es en realidad un método de solución del problema dual (3.60) asociado al problema inicial, en el que se aplica para la maximización del lagrangiano dual respecto a los multiplicadores una fórmula de máximo ascenso. Siguiendo este planteamiento, pueden obtenerse fórmulas de

actualización de los multiplicadores de mayor eficiencia, basadas en métodos de Newton de segundo orden, en función del hessiano del lagrangiano dual.

Los métodos de multiplicadores, por tanto, pueden contemplarse como métodos de solución de problemas de minimización generales, basados en formulaciones duales, donde los términos que se añaden al lagrangiano, tienen por objeto, de alguna forma, asegurar la convexidad del problema.

Es posible generalizar el método para problemas con restricciones en desigualdad, mediante una adecuada elección de la función lagrangiano aumentado. En particular para problemas en la forma estándar (3.1) es habitual la adopción de la función siguiente:

$$\chi(\bar{x}, \bar{\lambda}, w) = f(\bar{x}) + 1/w \sum_{j=1}^m \xi_j(\bar{x}, \bar{\lambda}, w) \quad ; w > 0$$

$$\xi_j(\bar{x}, \bar{\lambda}, w) \rightarrow \begin{cases} = -(1/2 w \lambda_j)^2 & ; \text{si } g_j(\bar{x}) < 1/2 w \lambda_j \\ = g_j(\bar{x}) - w \lambda_j g_j(\bar{x}) & ; \text{si } g_j(\bar{x}) \geq 1/2 w \lambda_j \end{cases} \quad (3.102)$$

siendo sencillo comprobar que las condiciones de Kuhn-Tucker de mínimo para el problema de partida coinciden con la condición de extremo para la función anterior. No es difícil comprobar que al minimizar la función anterior respecto a las variables primales " \bar{x} ", si los multiplicadores " $\bar{\lambda}$ " toman los valores que cumplen las condiciones de Kuhn-Tucker del problema inicial, la solución obtenida es solución del problema original. De igual forma que en el caso anterior, pueden derivarse fórmulas de actualización

para los multiplicadores de forma que converjan a los valores que verifican las condiciones de Kuhn-Tucker, obteniendo:

$$\lambda_j^{k+1} = \lambda_j^k - \frac{2}{w} \max \left\{ \frac{1}{2} w \lambda_j^k, g_j(\bar{x}^k) \right\} \quad (3.103)$$

Se observa que la fórmula de actualización (3.103) para problemas de minimización con restricciones en desigualdad, se obtiene directamente de la fórmula (3.101) para problemas con restricciones de igualdad, introduciendo una corrección que evite que los multiplicadores correspondientes a restricciones en desigualdad adopten valores positivos. De forma similar al caso anterior, podemos interpretar la fórmula (3.103) como una particularización del algoritmo (3.64), para minimización con restricciones exclusivamente laterales. Se trata por tanto de un método de máximo ascenso para la maximización del lagrangiano dual, sometido a restricciones laterales (de no positividad) de los multiplicadores. Desde este punto de vista, la fórmula de actualización tiene un significado inmediato, y se comprenden de forma intuitiva las razones que conducen a adoptar para el problema de minimización con restricciones en desigualdad un lagrangiano aumentado del tipo (3.102).

Unificando ambos lagrangianos aumentados, y desde una perspectiva de dualidad, puede obtenerse un método de multiplicadores, para el problema general (3.2) de minimización con restricciones cualesquiera, tratando las restricciones laterales separadamente. Esto no es más que una particularización del problema dual (3.60), sustituyendo el lagrangiano del problema por un lagrangiano aumentado que combina

los términos de los lagrangianos (3.98) y (3.102) para cada restricción, según se trate de restricciones de igualdad o desigualdad respectivamente, y resolviendo el problema de maximización con restricciones laterales mediante un método de máximo ascenso.

III.5.3.5 Métodos de reducción a problemas específicos de minimización con restricciones cualesquiera

En secciones previas se ha puesto de manifiesto la existencia de algoritmos de gran eficiencia para resolver dos problemas específicos de minimización con restricciones en desigualdad, a saber:

- Problemas de programación lineal.
- Problemas de programación cuadrática.

Para problemas de minimización de funciones no lineales con restricciones en desigualdad se han desarrollado varios métodos con numerosas variantes basados en la reducción del problema original a una secuencia de problemas de programación lineal o cuadrática, que expondremos a continuación, y que podemos clasificar en la forma:

- Métodos de direcciones factibles.
- Métodos de programación lineal secuencial.
- Métodos de programación cuadrática secuencial.

En lo sucesivo, consideraremos el problema general de programación matemática escrito en la forma estándar (3.1), donde supondremos por simplicidad, que si hay restricciones laterales,

se encuentran incluidas en el conjunto de restricciones en desigualdad, aunque pueden tratarse separadamente en la mayoría de los casos.

Métodos de Direcciones Factibles

Dado un punto factible " \bar{x}^k " del problema (3.1), una dirección " \bar{s} " se dice factible (o más propiamente, factible en primera aproximación), si se cumple:

$$\bar{\nabla}g_j(\bar{x}^k) \bar{s} \leq 0 \quad ; \forall j \in Q \quad (3.104)$$

siendo "Q" el conjunto de las "q" restricciones que se encuentran activas en el punto dado (Fig. 3.10).

Los métodos de direcciones factibles no son sino métodos de descenso del tipo (3.12) en los que se impone además la condición de que las sucesivas direcciones de descenso sean factibles, es decir, que verifiquen la condición anterior. Existen diversos algoritmos que se diferencian en la forma de estimar una dirección factible de avance, entre los cuales el más ampliamente utilizado es el de Zoutendijk. En este algoritmo, la dirección de descenso se obtiene como solución de un problema de programación lineal, formulado como:

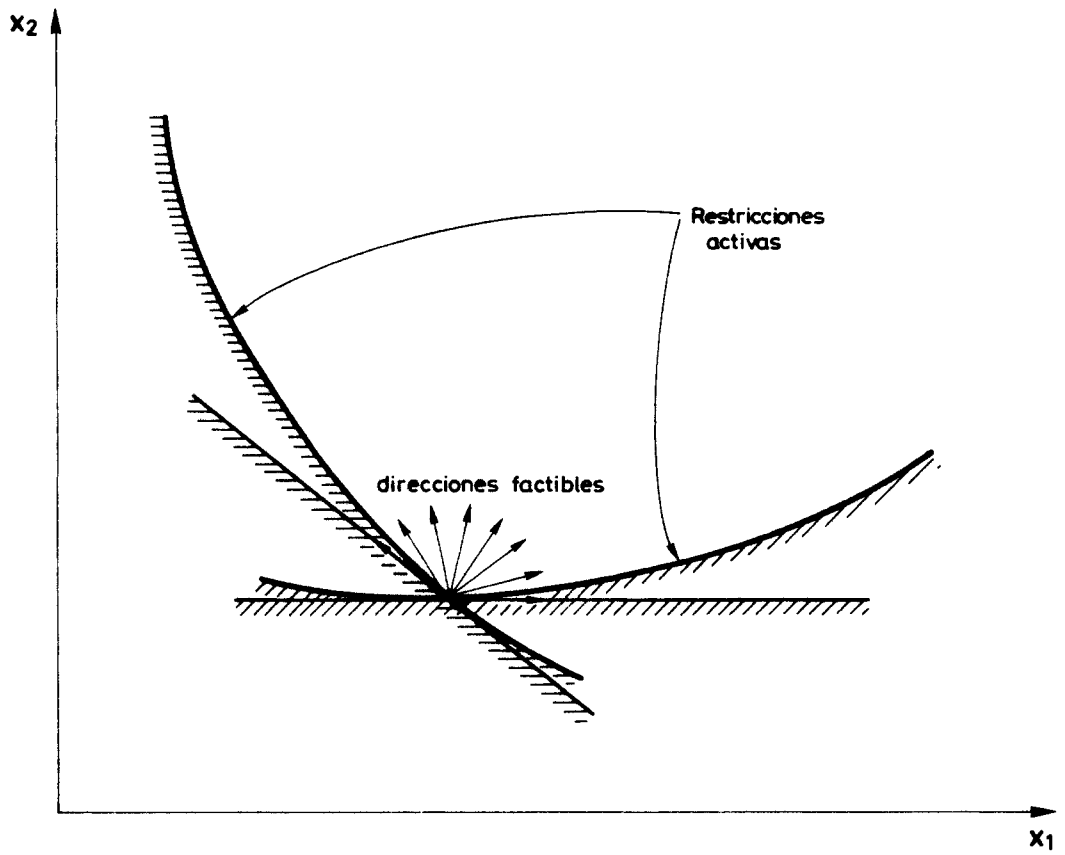


Figura 3.10.- Interpretación geométrica del concepto de direcciones factibles.

maximizar : β

verificando : $\beta \geq 0$

$$\bar{\nabla} f(\bar{x}) \cdot \bar{s} + \beta \leq 0 \quad (3.105)$$

$$\bar{\nabla} g_j(\bar{x}) \cdot \bar{s} + \varphi_j \beta \leq 0 \quad ; \forall j \in Q$$

con: $\varphi_j \geq 0 \quad ; \forall j \in Q$

y se impone además alguna restricción que limite el módulo del vector de descenso, para que exista una solución de valor finito.

Los coeficientes " φ_j " condicionan la dirección obtenida. Si un coeficiente tiende a anularse, la dirección de avance tiende a ser ortogonal al gradiente de la restricción correspondiente, y por tanto a encontrarse en el hiperplano tangente a tal restricción, predominando en la selección de la dirección el que el descenso de la función objetivo sea lo más acusado posible. Por el contrario, si el coeficiente tiende a infinito, el descenso de la función objetivo tiende a anularse a lo largo de tal dirección, y ésta tiende a ser ortogonal al gradiente de la misma, y por tanto la dirección obtenida tiende a encontrarse en las hipersuperficies de nivel de la función objetivo. Por tanto, para valores pequeños de los coeficientes se obtienen direcciones de fuerte descenso pero que conducen rápidamente al exterior de la región factible, mientras que para valores grandes, se obtienen direcciones que conducen al interior de la región factible, pero de leve descenso de la función objetivo.

Si la solución del problema (3.105) es $\beta = 0$, el punto factible es un mínimo local del problema inicial. En caso contrario, la dirección es factible y puede realizarse una minimización unidireccional para obtener una nueva aproximación a la solución del problema. En el proceso pueden añadirse o eliminarse elementos del conjunto de restricciones activas. Este factor es de gran interés, pues ofrece la posibilidad de controlar con sencillez la convergencia del algoritmo.

Es conveniente imponer la limitación en el módulo del vector de avance mediante restricciones lineales, puesto que de lo contrario el problema asociado (3.105) no podría resolverse mediante programación lineal, perdiéndose su ventaja fundamental. Es importante considerar que para puntos interiores de la región factible no existen restricciones activas, y por tanto la dirección de descenso dependerá exclusivamente de la forma de imponer tal limitación. En este caso, si la limitación se impusiese mediante una restricción de igualdad sobre la norma del vector, similar a la empleada en (3.44), el método coincidiría exactamente con un método de Newton para minimización no restringida.

La principal desventaja del algoritmo estriba en la necesidad de adoptar los coeficientes " φ_j " adecuados. Su estimación no es sencilla, y unos coeficientes incorrectos motivarán en general la presentación de efectos de zig-zag y un consiguiente bajo orden de convergencia. Sin embargo, el método permite proseguir el proceso de minimización en puntos del contorno con varias restricciones activas, en casos en que otros algoritmos no evolucionan favorablemente o sencillamente se

estancan.

Métodos de Programación Lineal Secuencial

Los métodos de programación lineal secuencial o recursiva, también conocidos como métodos de linealización, son variantes de un solo método que consiste en sustituir un problema de minimización con restricciones en desigualdad, por una secuencia de problemas de programación lineal que se obtienen aproximando la función objetivo y las restricciones mediante sus desarrollos de Taylor truncados por los términos de primer orden (Fig. 3.11-a). El algoritmo de programación lineal secuencial para el problema en la forma estándar (3.1) puede escribirse por tanto en la forma siguiente:

Dado: \bar{x}^k , Obtener: $\bar{x}^{k+1} = \bar{x}^k + \bar{r}^k$, donde:

$$\bar{r}^k \text{ minimiza: } F(\bar{r}) = f(\bar{x}^k) + \nabla f(\bar{x}^k) \cdot \bar{r} \quad (3.106)$$

$$\text{verificando: } G_j(\bar{r}) = g_j(\bar{x}^k) + \nabla g_j(\bar{x}^k) \cdot \bar{r} \leq 0$$

; j=1, ..., m

Para resolver cada problema linealizado es preciso realizar una traslación de ejes en el espacio de diseño, dado que la aplicación del Método del Simplex implica la no negatividad de las variables primales. Por ello, el algoritmo anterior se escribe habitualmente en la forma:

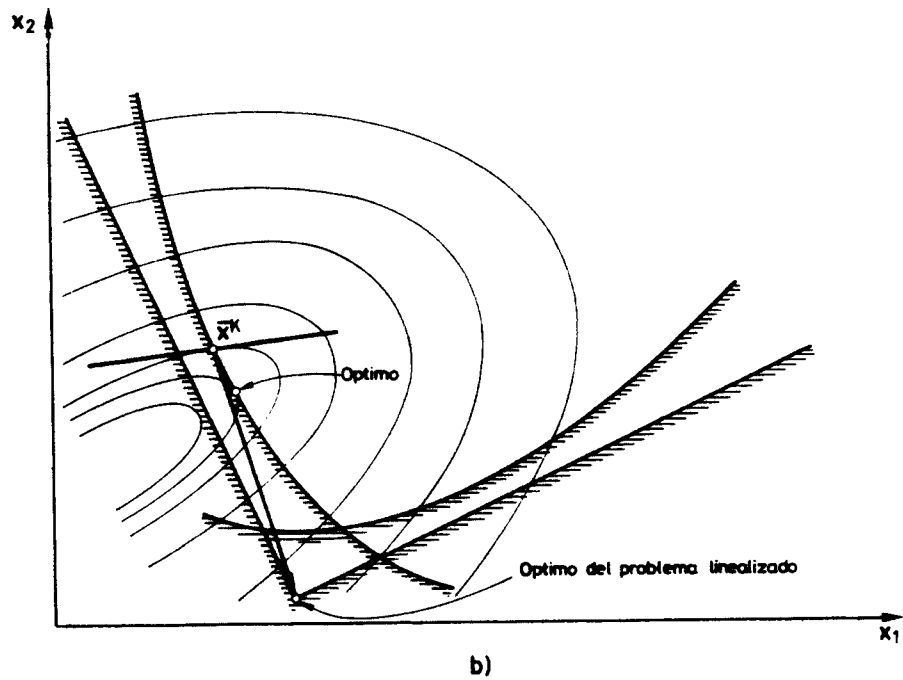
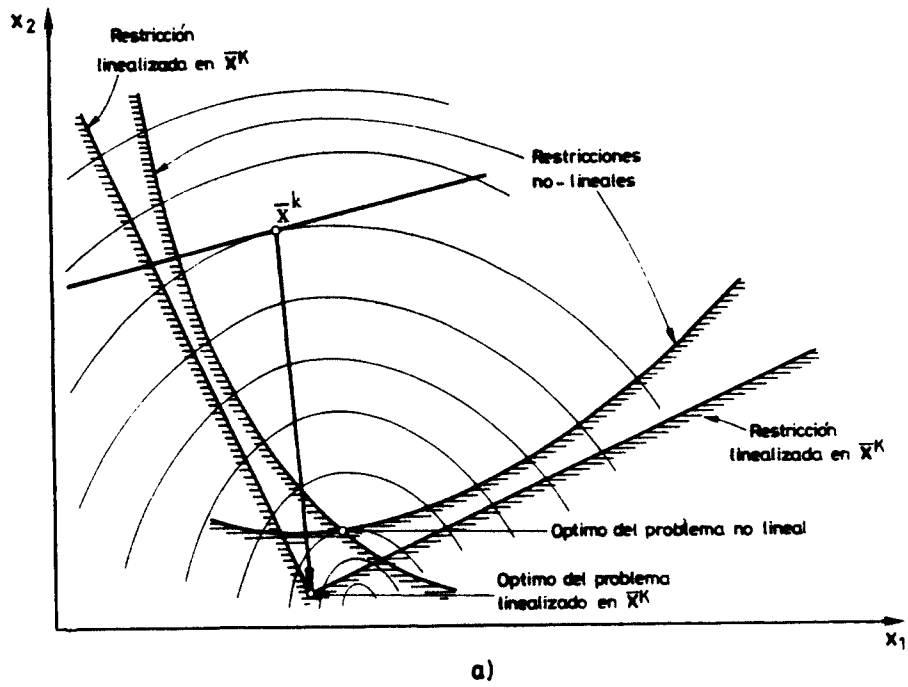


Figura 3.11.- Interpretación geométrica del algoritmo de Programación Lineal Secuencial y de las oscilaciones cerca del óptimo.

a) Comparación entre el problema de programación no lineal y el problema linealizado en torno a un punto.

b) Ejemplo de fuertes oscilaciones en torno al óptimo en un problema cuya solución no está en un vértice del contorno de la región factible.

Dado: \bar{x}^k , Obtener: $\bar{x}^{k+1} = \bar{x}^k + \bar{r}^k = \bar{t} + \bar{u}^k$, donde:

$$\bar{u}^k \text{ minimiza: } F(\bar{u}) = F(\bar{0}) + \bar{\nabla} f(\bar{x}^k) \cdot \bar{u}$$

$$\begin{aligned} \text{verificando: } G_j(\bar{u}) &= G_j(\bar{0}) + \bar{\nabla} g_j(\bar{x}^k) \cdot \bar{u} \leq 0 && ; j=1, \dots, m \\ u_i &\geq 0 && ; i=1, \dots, n \end{aligned} \quad (3.107)$$

$$\text{con: } F(\bar{0}) = f(\bar{x}^k) + \bar{\nabla} f(\bar{x}^k) \cdot (\bar{t} - \bar{x}^k)$$

$$\begin{aligned} G_j(\bar{0}) &= g_j(\bar{x}^k) + \bar{\nabla} g_j(\bar{x}^k) \cdot (\bar{t} - \bar{x}^k) \\ &&& ; j=1, \dots, m \end{aligned}$$

siendo " \bar{t} " un vector que acota inferiormente los valores de las variables de diseño. Si no existe tal acotación en el problema inicial, basta adoptar un vector de cotas suficientemente pequeño, cuya única misión es permitir el tratamiento mediante el Método del Simplex de la secuencia de problemas de programación lineal. No obstante, es frecuente en problemas de optimización estructural la imposición de restricciones laterales en los valores de las variables de diseño, y como veremos a continuación, ciertos defectos del método de programación lineal secuencial pueden soslayarse en mayor o menor grado introduciendo restricciones laterales adicionales en cada iteración. Por ello la traslación de ejes no supone habitualmente un encarecimiento o una pérdida de eficiencia del algoritmo.

La resolución de la secuencia de problemas linealizados que proporciona el algoritmo anterior es sencilla, y sus requerimientos de almacenamiento en memoria y tiempos de cálculo

son razonables. Por ello, el algoritmo de programación lineal secuencial goza de gran aceptación en problemas de optimización estructural. Sin embargo, es preciso reconocer la existencia de graves inconvenientes que exponemos a continuación:

- Para que la secuencia de problemas linealizados tenga solución, es necesario en general que existan al menos tantas restricciones como variables de diseño. Este defecto no se manifiesta en la versión (3.107) del algoritmo, dado que la traslación de ejes realizada, introduce automáticamente "n" restricciones adicionales.
- Aunque se cumpla la condición anterior y el problema de programación no lineal esté bien definido, en su aproximación lineal la región factible puede ser ilimitada en una dirección de descenso, y el problema de programación lineal tendría solución en el infinito.
- Puesto que la solución de un problema de programación lineal siempre se encuentra en un vértice del hiperpoliedro que configura su región factible, el algoritmo anterior no converge si la solución del problema inicial de programación no lineal es un mínimo no restringido o un mínimo condicionado por un número de restricciones inferior al número de variables de diseño (Fig. 3.11-b). En estos casos se producen oscilaciones entre vértices de los problemas linealizados.

Para evitar los defectos anteriores se han desarrollado diversas modificaciones del algoritmo original de programación lineal secuencial, entre las que citaremos los siguientes métodos:

- Método de Kelley, o del plano secante.
- Método de la hiperesfera.
- Método de las hiperesferas inscritas.
- Método de límites móviles.

El método del plano secante se basa en el principio de que en problemas convexos, la región factible de cada problema linealizado del algoritmo de programación lineal secuencial

(3.106) contiene integramente a la región factible del problema inicial (3.1). Por tanto, las sucesivas linealizaciones de las restricciones en la secuencia de problemas lineales a resolver, configuran una envolvente convexa de la región factible del problema inicial. El método consiste simplemente en retener en cada iteración las restricciones linealizadas de las iteraciones anteriores. Para problemas convexos puede probarse la convergencia del algoritmo al mínimo global, aunque su orden de convergencia es bajo. Sin embargo, cuando se aplica a problemas no convexos, la retención de linealizaciones sucesivas de las restricciones puede excluir de la región factible linealizada el mínimo global. En cualquier caso, la principal desventaja del algoritmo estriba en el crecimiento del número de restricciones en cada iteración.

El método de la hiperesfera consiste en la aplicación del método de programación lineal secuencial con una corrección adicional a posteriori en cada iteración, de forma que cada aproximación a la solución se encuentre en el interior de una hiperesfera de radio prefijado centrada en la aproximación precedente, respetando la dirección de modificación de las variables de diseño " \bar{r}^{-k} " proporcionada por la solución del problema linealizado. Podemos escribir el algoritmo de corrección aplicable al algoritmo (3.107) en la forma:

$$\begin{aligned}
 \text{Si:} \quad & \|\bar{r}^{-k}\| > R \\
 \text{entonces:} \quad & \bar{x}^{k+1} = \bar{x}^k + \frac{R}{\|\bar{r}^{-k}\|} \bar{r}^{-k}
 \end{aligned} \tag{3.108}$$

donde "R" es el radio de la hiperesfera arbitrariamente adoptado. La importancia de la corrección anterior decrece a medida que el proceso converge al óptimo. De hecho, su único efecto es impedir que la modificación de las variables de diseño en cada iteración sea excesiva, de forma que la aproximación lineal mantenga hasta cierto punto su validez. Sin embargo, el método no converge, al igual que el algoritmo general, si la solución del problema no lineal no se encuentra condicionada por un número suficientemente elevado de restricciones.

El método de las hiperesferas inscritas es una reformulación del método de programación lineal secuencial, construida con el objetivo de intentar evitar la convergencia a óptimos locales, y forzar la convergencia al óptimo global del problema. Sin embargo adolece de los mismos defectos del algoritmo anterior, en cuanto a que la convergencia no es posible si la solución es un mínimo no condicionado, o se encuentra condicionada por un número insuficiente de restricciones. El algoritmo se basa en la obtención del centro de la hiperesfera de mayor radio inscrita en la región factible linealizada, y de este hecho recibe su nombre.

El método de los límites móviles, consiste en la imposición de restricciones laterales adicionales en cada iteración, de forma que la posición de la nueva aproximación se encuentre restringida al interior de un paralelepípedo n-dimensional centrado en la aproximación precedente. El algoritmo (3.106), se completa por tanto con el conjunto de "2n" restricciones lineales:

\bar{r}^k verifica:

$$-c_i \leq r_i^k \leq c_i \quad ; i=1, \dots, n \quad (3.109)$$

con: $c_i > 0 \quad ; i=1, \dots, n$

siendo " c_i " un conjunto de límites móviles arbitrario, que puede mantenerse fijo a lo largo de todo el proceso, o actualizarse en cada iteración con objeto de mejorar la convergencia. Evidentemente es posible, y en ocasiones recomendable, adoptar un valor distinto para el límite inferior y superior de cada variable, aunque en este desarrollo emplearemos límites idénticos a ambos lados por simplicidad.

El algoritmo de programación lineal secuencial (3.106) con los límites móviles (3.109), formulado de forma que pueda ser resuelto mediante el Método del Simplex, puede escribirse en una forma similar a la (3.107) en la que no es necesario introducir un vector " \bar{t} " de cotas inferiores de las variables de diseño, dado que los límites móviles (3.109) sugieren inmediatamente una traslación de ejes que asegure la no negatividad de las variables primales de los problemas linealizados. El algoritmo resultante puede escribirse en la forma:

Dado: \bar{x}^k , Obtener: $\bar{x}^{k+1} = \bar{x}^k + r = \bar{x}^k - c + \bar{u}$, donde:

$$\bar{u}^k \text{ minimiza: } F(\bar{u}) = F(\bar{0}) + \bar{\nabla} f(\bar{x}^k) \bar{u}$$

$$\text{verificando: } G_j(\bar{u}) = G_j(\bar{0}) + \bar{\nabla} g_j(\bar{x}^k) \bar{u} \leq 0 \quad ; j=1, \dots, m$$

$$0 \leq u_i \leq 2c_i \quad ; i=1, \dots, n \quad (3.110)$$

$$\text{con: } F(\bar{0}) = f(\bar{x}^k) - \bar{\nabla} f(\bar{x}^k) \bar{c}$$

$$G_j(\bar{0}) = g_j(\bar{x}^k) - \bar{\nabla} g_j(\bar{x}^k) \bar{c} \quad ; j=1, \dots, m$$

La imposición de límites móviles proporciona ventajas de importancia. Los límites móviles permiten controlar la modificación de las variables de diseño, de forma que la validez de las aproximaciones lineales se mantenga dentro de unos márgenes razonables. Introducen de forma natural "2n" restricciones adicionales sencillas que obvian los dos primeros defectos del método de programación lineal secuencial anteriormente expuestos. Además, el método puede aproximarse a soluciones no necesariamente condicionadas por un número suficientemente elevado de restricciones no lineales, y si se presentan oscilaciones en la convergencia, su amplitud está restringida por el valor de los límites impuestos. Por tanto, una adecuada actualización de los límites móviles en cada iteración, reduciéndolos en mayor o menor grado si se presentan oscilaciones cerca del óptimo, permite obtener buenas aproximaciones a la solución del problema no lineal. Si además

hay restricciones laterales en el problema, bastará imponer en cada iteración, y para cada variable, las restricciones predominantes, ya sean las laterales iniciales o las impuestas por los límites móviles. En general, puede afirmarse que el algoritmo (3.110) es un método de considerable potencia para abordar la solución del problema (3.1) de programación no lineal, ampliamente utilizado en el ámbito de la optimización estructural, y capaz de proporcionar aproximaciones razonablemente buenas a la solución, si bien en las proximidades del óptimo es necesario recurrir a tratamientos más sofisticados para obtener resultados de mayor exactitud y evitar oscilaciones.

Métodos de Programación Cuadrática Secuencial

Los métodos de programación cuadrática secuencial o recursiva consisten en la sustitución de un problema general de minimización con restricciones (3.2) por una secuencia de problemas aproximados de programación cuadrática (3.67), de forma que la secuencia de soluciones obtenida converja a la solución del problema general inicial.

Una forma razonable de obtener una secuencia de problemas de programación cuadrática aproximados, consiste en desarrollar la función objetivo y las restricciones mediante sus series de Taylor, y truncarlas por sus términos de segundo y primer orden respectivamente. Consecuentemente, podemos definir un algoritmo de programación cuadrática secuencial para el problema general escrito en la forma estándar (3.1) como:

Dado: \bar{x}^k , Obtener: $\bar{x}^{k+1} = \bar{x}^k + \bar{r}^k$, donde:

$$\bar{r}^k \text{ minimiza: } F(\bar{r}) = f(\bar{x}^k) + \nabla f(\bar{x}^k) \bar{r} + \frac{1}{2} \bar{r}^t \underline{H}(\bar{x}^k) \bar{r} \quad (3.111)$$

$$\text{verificando: } G_j(\bar{r}) = g_j(\bar{x}^k) + \nabla g_j(\bar{x}^k) \bar{r} \leq 0 \quad ; j=1, \dots, m$$

donde " \underline{H} " es el hessiano de la función objetivo, y que, como es bien sabido, debe ser una matriz semidefinida positiva para que el problema pueda considerarse de programación cuadrática. El cálculo del hessiano puede obviarse mediante la aplicación de un algoritmo de Quasi-Newton (DFP, BFGS, etc.).

Este algoritmo de programación cuadrática recursiva proporciona un mayor orden de convergencia que los métodos de programación lineal secuencial, y la convergencia es segura para problemas convexos. En problemas no convexos, es posible que el hessiano de la función objetivo sea no definido positivo, y el algoritmo puede no funcionar adecuadamente.

No obstante, en su aplicación a problemas generales de optimización estructural, este algoritmo de programación cuadrática secuencial no aporta grandes ventajas, dado que en general, en este tipo de problemas, las funciones altamente no lineales son las restricciones, y no la función objetivo. Su principal ventaja frente a los métodos de programación lineal secuencial se deben a que la aproximación cuadrática de la función objetivo es mejor que la aproximación lineal, y las soluciones de cada problema cuadrático no tienen por qué hallarse

en un vértice de la región factible, y por tanto es de esperar una convergencia más rápida, así como la no aparición de oscilaciones y otras mal funciones características de los métodos de linealización. Su principal desventaja reside en sus mayores requerimientos de almacenamiento de memoria y tiempo de cálculo por iteración.

Es posible desarrollar otros algoritmos de programación cuadrática recursiva a partir de consideraciones distintas. En particular, si se emplea una estrategia basada en la identificación de las restricciones activas para resolver el problema (3.1), en cada iteración se dispone de un problema de minimización con restricciones de igualdad. Para esta secuencia de problemas, las condiciones (3.46) son condiciones de extremo. Consecuentemente, si se plantea la solución de las ecuaciones (3.48), que implican la anulación del gradiente del lagrangiano mediante el método iterativo de Newton-Raphson, se obtiene un algoritmo equivalente a un método de programación cuadrática recursiva que podemos escribir en la forma:

$$\begin{aligned} \text{Dado: } \bar{x}^k, \text{ Obtener: } \bar{x}^{k+1} &= \bar{x}^k + \bar{r}^k, \text{ donde:} \\ \bar{r}^k \text{ minimiza: } F(\bar{r}) &= f(\bar{x}^k) + \bar{\nabla} f(\bar{x}^k) \bar{r}^k + \frac{1}{2} \bar{r}^t \bar{W}(\bar{x}^k) \bar{r}^k \quad (3.112) \\ \text{verificando: } G_j(\bar{r}) &= g_j(\bar{x}^k) + \bar{\nabla} g_j(\bar{x}^k) \bar{r}^k \leq 0 \\ & \quad ; j=1, \dots, m \end{aligned}$$

donde "W" es el hessiano del Lagrangiano del problema respecto a las variables primales. La diferencia esencial entre los algoritmos (3.111) y (3.112) se basa en que en el primero el

hessiano de la función cuadrática a minimizar coincide con el de la función objetivo, mientras que en el segundo coincide con el lagrangiano del problema. La desventaja de la formulación (3.112) estriba en que es obligado el cálculo de los hessianos de las funciones que definen las restricciones, lo cual no siempre es posible o admisible en términos de tiempo de cálculo. Su ventaja fundamental es que introduce de alguna forma en cada uno de los problemas cuadráticos la no linealidad de las restricciones.

Parte de las desventajas de ambos algoritmos pueden paliarse mediante la aplicación de métodos Quasi-Newton para aproximar los hessianos.