

CAPÍTULO 3. OPTIMIZACIÓN NO LINEAL.

3.1. Introducción.

En el presente capítulo se describe el método de optimización no lineal con restricciones que se ha desarrollado. Un proceso de optimización consta de varias partes a definir:

- Función objetivo: Es la variable o combinación de variables que se quieren minimizar o maximizar.
- Variables de diseño: Son las variables independientes que definen el diseño del problema mecánico. Normalmente son parámetros geométricos.
- Restricciones: Son inecuaciones que se deben cumplir para que el óptimo alcanzado sea válido. Éstas podrán incluir tanto variables dependientes como independientes.

Si por ejemplo se quiere optimizar la cantidad de material utilizado en una viga de sección cuadrada, y se quiere que esta sección mantenga una cierta relación de aspecto, entonces se tendrá:

- Función objetivo: Volumen del material utilizado en la viga
- Variables de diseño: El ancho y el canto de la sección de la viga
- Restricciones: Que no se supere el límite elástico del material, que la flecha sea menor que una dada y que la relación entre ancho y canto se encuentre dentro de cierto intervalo.

Además las variables de diseño deberán estar dentro de cierto intervalo. La formulación de este problema se presenta a continuación:

Función objetivo.	Minimizar Volumen de material.
	$\sigma_{eqv} \leq \sigma_e$
Restricciones.	$u \leq u_{max}$
	$H/B \leq rel_1$
	$rel_2 \leq H/B$
	$H_1, B_1 < H, B < H_2, B_2$

donde H y B son el canto y ancho de la viga.

El primer paso para resolver este tipo de problemas es transformar el problema con restricciones en uno sin restricciones. Esto se consigue mediante la penalización de la función objetivo que se realiza mediante funciones de penalización del tipo interior extendida. De esta manera se pueden aplicar los algoritmos de optimización sin restricciones. Posteriormente se

estudia la optimización multidimensional sin restricciones, haciendo consideraciones teóricas de los métodos de primer y de segundo orden, en los que se explica el método del gradiente (o método de Cauchy) y el de Davidon-Fletcher-Powell [14] (algoritmo de métrica variable) respectivamente. Estos métodos se diferencian principalmente en la forma en la que se obtiene la dirección de búsqueda.

Ambos métodos descritos anteriormente necesitan de una búsqueda monodimensional en la dirección de búsqueda que se obtiene. Para ello se describe el método de búsqueda monodimensional que es el método más básico así como el primordial en cualquier proceso de optimización.

Todo el proceso de optimización se puede condensar en la siguiente fórmula:

$$\underline{x}^{k+1} = \underline{x}^k + s_k \cdot \underline{d}_k \quad (3.1.1)$$

donde s_k es el parámetro de búsqueda monodimensional en la iteración k -ésima, \underline{x}^k son las variables de diseño en la iteración k -ésima, y \underline{d}_k es la dirección de búsqueda en la iteración k -ésima.

En cada iteración se calcula una dirección de búsqueda \underline{d}_k . La forma de calcularla dependerá de si se utiliza el método del gradiente o el de Davidon-Fletcher-Powell. El siguiente paso consiste en buscar el mínimo de la función objetivo penalizada en dicha dirección mediante la búsqueda monodimensional. Esta búsqueda, que se realiza como se ha comentado con el método de la sección áurea, proporciona un valor de s_k que será el que se utilice para el cálculo del siguiente estado \underline{x}^{k+1} . Una vez que este proceso converge se actualiza el parámetro de penalización y se obtiene una nueva función objetivo penalizada que vuelve a ser optimizada. Este proceso se termina cuando se da la convergencia del proceso de optimización global.

Por último se explica la implementación tanto del algoritmo de primer orden como del de segundo orden. Esta implementación se realiza tanto en lenguaje de programación de ANSYS® como en lenguaje de programación de FORTRAN.

El programa de elementos finitos que se va a utilizar para la resolución del problema incluye un módulo de optimización. A pesar de esto se ha decidido realizar la implementación de los algoritmos de optimización debido a ciertas incongruencias que aparecen en el manual de dicho módulo. En primer lugar se presenta un método de orden 0. Este método no requiere la evaluación de las derivadas y utiliza una aproximación de la función objetivo penalizada mediante ajustes cuadráticos.

Estas incongruencias son las siguientes:

- En el método de orden 0:

- Las funciones de penalización, que según dice son del tipo interior extendida, quedan definidas en función de ciertos parámetros:

$$W(w_i) = \begin{cases} d_1 + d_2 / (\bar{w} - \hat{w}_i) & \text{if } \hat{w}_i < \bar{w}_i - \varepsilon(\bar{w}_i - \underline{w}_i) \\ d_3 + d_4(\hat{w}_i - \bar{w}) & \text{if } \hat{w}_i \geq \bar{w}_i - \varepsilon(\bar{w}_i - \underline{w}_i) \end{cases} \quad (i = 1, 2, 3, \dots, m_1) \quad (3.1.2)$$

donde d_1 , d_2 , d_3 y d_4 son parámetros “internamente calculados” y ε es un “número positivo muy pequeño”, lo que impide cualquier tipo de control sobre las funciones de penalización.

- Dice que el parámetro de penalización se incrementa cuando debe decrementarse como corresponde a la penalización interior extendida.
- El parámetro que utiliza para la actualización de las variables de diseño tampoco se indica como se obtiene.

$$\mathbf{x}^{(j+1)} = \mathbf{x}^{(b)} + C(\hat{\mathbf{x}}^{(j)} - \mathbf{x}^{(b)}) \quad (3.1.3)$$

donde $x^{(b)}$ es el estimado anterior y C es un parámetro que varía entre 0 y 1 “internamente elegido”.

- En el método de primer orden:

- Las funciones de penalización son dejadas en función de parámetros no conocidos.

$$P_g(g_i) = \left(\frac{g_i}{g_i + \alpha_i} \right)^{2\lambda} \quad (3.1.4)$$

donde λ es “un entero de valor muy elevado”.

- Indica que para las variables de diseño se aplica penalización exterior, pero no la muestra.
- No se define el parámetro de penalización.

Además, *ANSYS*[®] carece de método de segundo orden. Estos algoritmos proporcionan una convergencia más rápida en problemas donde la región factible sea muy amplia o problemas altamente no lineales.

Estos hechos justifican la implementación de dos algoritmos de optimización: uno de primer orden y otro de segundo orden, que conseguiría mejorar la convergencia en problemas en los que la función objetivo no se aproxime bien de forma lineal.

3.2. Optimización con restricciones.

Se trata del problema más general. Optimización de una función no lineal sujeto también a restricciones, que en el caso más general serán no lineales. Este problema se formula como sigue:

$$\text{Minimizar: } f(\underline{x}) \quad (3.2.1)$$

$$\text{Sujeto a: } g_j(\underline{x}) \leq 0 \quad j = 1, m \quad (3.2.2)$$

$$h_k(\underline{x}) = 0 \quad k = 1, l \quad (3.2.3)$$

$$\underline{x}_i \in [\underline{x}_i^l, \underline{x}_i^u]. \quad (3.2.4)$$

El procedimiento que se sigue es intentar formular el problema como un problema sin restricciones. Para ello se penaliza la violación de las restricciones. Si inicialmente se penalizaran completamente las restricciones se obtendrían discontinuidades. Estas discontinuidades en la frontera de la región factible daría lugar a problemas mal condicionados. Por ello se impone inicialmente la penalización de una forma moderada, aumentándose dicha penalización a medida que avanza la penalización. Esto implica que se tengan que resolver varios problemas de optimización sin restricciones para la resolución de uno con restricciones. De ahí el nombre de estos métodos: *sequential unconstrained minimization techniques (SUMT)*. Existe multitud de teoremas y aspectos teóricos sobre *SUMT* pero dado que el propósito que se tiene es el de implementar un algoritmo para un diseño de ingeniería se limitará la discusión a un aspecto más general y práctico de estos métodos.

La pseudo-función objetivo que se emplea en estos métodos son de la forma:

$$\phi(\underline{x}, \lambda_p) = f(\underline{x}) + \lambda_p \cdot P(\underline{x}) \quad (3.2.5)$$

donde $f(\underline{x})$ es la función objetivo original y $P(\underline{x})$ es la función de penalización.

La forma de la función de penalización es lo que diferencia los distintos métodos que existen para transformar un problema de optimización no lineal con restricciones en un problema de optimización no lineal sin restricciones. Los métodos que se discuten a continuación son tres. El primero y más fácil de implementar es el de la penalización exterior, llamado así porque sólo se penaliza la función objetivo cuando las restricciones no se cumplen y se puede partir de un estimado fuera de la región factible. El segundo se conoce como el método de la penalización interior, en el que se va penalizando conforme se acerca a la frontera de la región factible pero sin dejar nunca que se viole. En este segundo hay que partir de un estimado factible. Las ventajas de estos dos métodos se incorporan al tercero, conocido como el método de penalización interior extendida. Estudios más recientes indican que cuando se dan

problemas numéricos por mal condicionamiento del mismo estos se pueden reducir con el uso de los multiplicadores de Lagrange.

3.2.1. Método de la penalización exterior.

En este método la función penalización viene dada de la siguiente forma:

$$P(\underline{x}) = \sum_{j=1}^m \{\max[0, g_j(\underline{x})]\}^2 + \sum_{k=1}^l [h_k(\underline{x})]^2 \quad (3.2.6)$$

Como se observa, en caso de que no se cumplan las restricciones existe penalización de éstas, pero si se están cumpliendo la penalización es nula ($g_j(\underline{x}) \leq 0$, $h_k(\underline{x}) = 0$). La penalización es la propia restricción al cuadrado. Así se consigue continuidad de la derivada primera de la pseudo-función objetivo $\phi(\underline{x}, \lambda_p)$ en la frontera. Pero al no ser continua la derivada segunda, puede conllevar problemas numéricos en caso de usar un método de segundo orden.

Considerando ahora el factor λ_p , se observa en la figura 3.1 que un valor pequeño del mismo facilitarían la minimización pero implicaría una mayor violación de las restricciones. Un valor alto conllevaría un cumplimiento bastante satisfactorio de las restricciones pero implicaría un peor condicionamiento del problema desde el punto de vista numérico. De este modo generalmente se comienza con un valor bajo del factor λ_p , optimizando la pseudo-función objetivo $\phi(\underline{x}, \lambda_p)$ posteriormente. A continuación se aumenta el valor de λ_p para volver a optimizar $\phi(\underline{x}, \lambda_p)$ partiendo de la solución anterior, creándose así un proceso iterativo que termina cuando se obtienen unos resultados satisfactorios. La aplicación de este método al problema siguiente se muestra en la figura 3.2. (b), mientras que en la 3.2. (a) se muestra la función objetivo y las restricciones originales:

Minimizar:	$\frac{(x+2)^2}{20}$
	$\frac{1-x}{2} \leq 0$
Restricciones.	$\frac{x-2}{2} \leq 0$

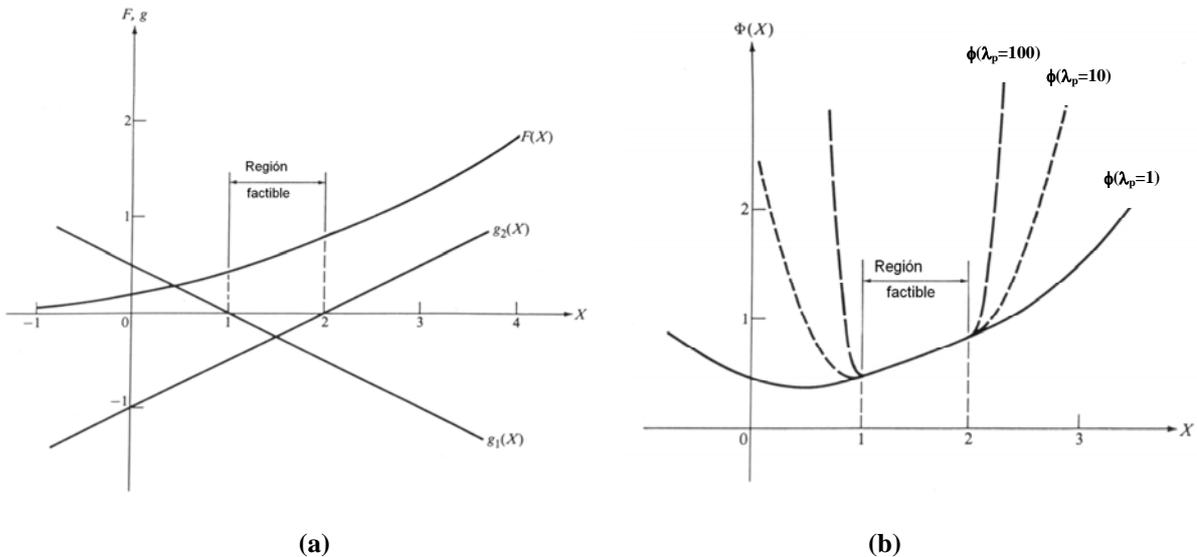


Figura 3.1. (a) Función objetivo $F(x)$ y funciones de penalización $g(x)$; (b) Pseudo función objetivo mediante el método de la penalización externa.

3.2.2. Método de la penalización interior

La ventaja de este método respecto al anterior reside en que la serie de soluciones que se van alcanzando a lo largo del proceso iterativo son todas factibles. Como desventaja se tiene que el problema de minimización resultante es algo más complejo y que es necesario que el estimado inicial sea factible.

La función de penalización más comúnmente utilizada en este método es, para las restricciones de desigualdad:

$$P(\underline{x}) = \sum_{j=1}^m \frac{-1}{g_j(\underline{x})} \quad (3.2.7)$$

mientras que para las de igualdad son las dadas por el método de la penalización exterior ya que se busca que $h_k(\underline{x})$ sea lo más cercana posible a cero. Quedando la pseudo-función objetivo como sigue:

$$\phi(\underline{x}, \lambda_p, \lambda_p') = f(\underline{x}) + \lambda_p' \cdot \sum_{j=1}^m \frac{-1}{g_j(\underline{x})} + \lambda_p \cdot \sum_{k=1}^l [h_k(\underline{x})]^2 \quad (3.2.8)$$

que es una pseudo-función objetivo bastante más compleja que la del método anterior. Cabe destacar que el factor λ_p' va decreciendo a lo largo del proceso, mientras que el λ_p crece, ya que se trata de penalización exterior. Este proceso se muestra en la siguiente figura, sobre la misma función objetivo y restricciones utilizadas en el método de penalización anterior.

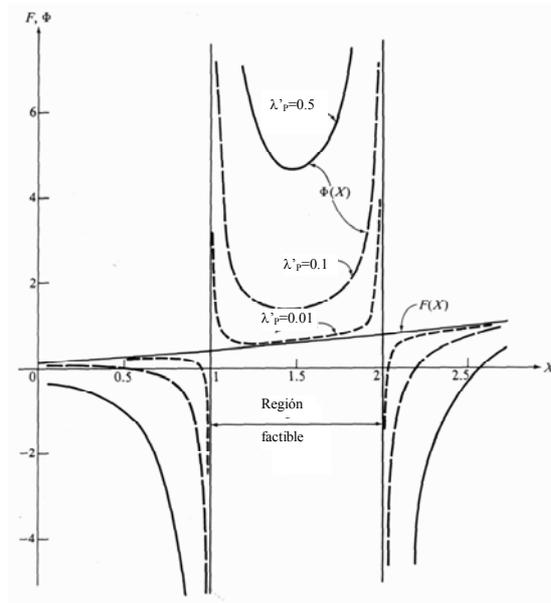


Figura 3.2. Pseudo función objetivo mediante el método de la penalización interna.

Para que la solución sea factible es necesario que $g_j(\underline{x})$ sea menor que cero, observándose que $-1/g_j(\underline{x})$ es mayor que cero, tendiendo además a infinito conforme se acerca a la restricción.

3.2.3. Método de la penalización interior extendida.

Este método intenta incorporar las ventajas de los anteriores para las restricciones de desigualdad. Para las restricciones de igualdad utiliza penalización exterior, por lo que se omitirán en la explicación por brevedad.

La función de penalización tiene la siguiente forma:

$$P(\underline{x}) = \sum_{j=1}^m \tilde{g}_j(\underline{x}) \quad (3.2.9)$$

donde:

$$\tilde{g}_j(\underline{x}) = \frac{-1}{g_j(\underline{x})} \quad \text{si} \quad g_j(\underline{x}) \leq \varepsilon \quad (3.2.10)$$

$$\tilde{g}_j(\underline{x}) = -\frac{2 \cdot \varepsilon - g_j(\underline{x})}{\varepsilon^2} \quad \text{si} \quad g_j(\underline{x}) > \varepsilon \quad (3.2.11)$$

El parámetro ε es un valor pequeño que indica el punto en el que se pasa de penalización interior de la ecuación (3.2.9) a la extendida de la ecuación (3.2.10). La elección de ε debe ser tal que la pendiente de la pseudo-función objetivo en la frontera de la región factible sea positiva. El valor recomendado por Haftka y Starnes [16] es:

$$\varepsilon = -C \cdot (\lambda'_p)^a \quad \text{con} \quad 1/3 \leq a \leq 1/2 \quad (3.2.12)$$

donde C es una constante. Al principio de la optimización estos autores eligen un $\varepsilon \in [-0.3, -0.1]$, y escogen λ_p' de tal forma que los dos términos de la derecha de la ecuación (3.2.5) sean iguales. De este modo se define el valor de C . Con esta elección de valores se asegura pendiente positiva en la frontera, lo que conlleva una secuencia de soluciones factibles.

Durante la optimización se va disminuyendo λ_p' de forma que la solución se va acercando al óptimo deseado como se muestra a continuación.

Este es el método que se ha utilizado en el presente proyecto debido a sus ventajas ya comentadas, es decir, que la serie de soluciones que se van alcanzando a lo largo del proceso de optimización son factibles y que el estimado inicial puede encontrarse fuera de la región factible.

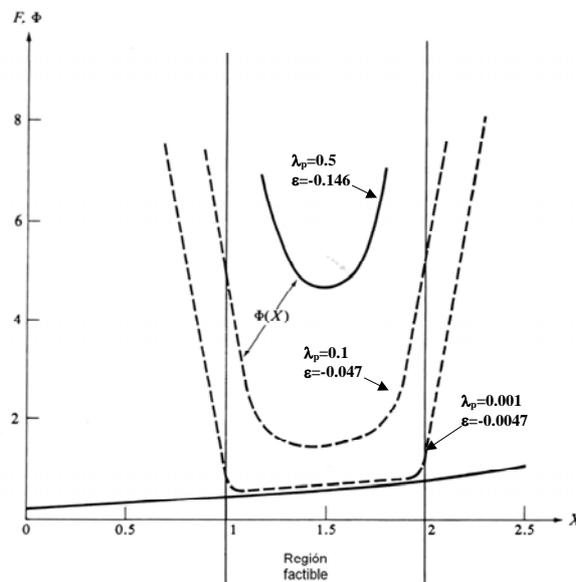


Figura 3.3. Pseudo función objetivo mediante el método de la penalización interna.

3.3. Optimización sin restricciones

En muchas ocasiones se presentan problemas de optimización sin restricciones o uno con restricciones que se ha transformado en otro sin restricciones. En estos problemas sólo se busca el mínimo de la función objetivo. Hay que aclarar que en caso de que lo que se desee sea maximizar, basta con minimizar la función cambiada de signo, por ello, en lo que sigue se habla de minimizar sin pérdida alguna de generalidad.

Los métodos de optimización se pueden clasificar en directos y en indirectos. Los indirectos son bastante complejos ya que hay que resolver unas ecuaciones que son fuertemente no lineales. Éstas son obtenidas tras imponer la condición de mínimo, que no es otra que anular el gradiente.

Los métodos directos de optimización son iterativos. Se parte de un estimado inicial x^0 y tras calcular la dirección de búsqueda d^l , se llega al punto x^l tras la búsqueda monodimensional en dicha dirección. Según el método que se utilice se calculará la dirección de búsqueda de una u otra manera. Generalizando para la iteración k-ésima se tiene:

$$\underline{x}^{k+1} = \underline{x}^k + s_k \cdot \underline{d}_k \quad (3.3.1)$$

donde s_k es el parámetro calculado en la búsqueda monodimensional. En ésta se minimiza la función objetivo mediante la variación del parámetro s_k dentro de su intervalo factible $[s_{k1}, s_{k2}]$. Este intervalo vendrá dado por la imposición de los límites de las variables de diseño.

$$\text{Min. } f(\underline{x}^k + s_k \cdot \underline{d}_k) \quad s_k \in [s_{k1}, s_{k2}] \quad (3.3.2)$$

Según cómo se obtenga la dirección de búsqueda los métodos se clasifican principalmente en primer y segundo orden como se explica a continuación.

3.3.1. Métodos de primer orden.

Reciben su nombre porque utilizan el gradiente para el cálculo de la dirección de búsqueda. Son métodos lentos en general pero que en muchos casos, en los que la función a minimizar se asemeje a una función lineal en el intervalo, da muy buenos resultados debido a su simplicidad. Así mismo dan malos resultados para funciones con derivadas primeras discontinuas en los que son más convenientes métodos de optimización sin derivadas.

La forma de evaluar el gradiente puede ser analítica o numérica por diferencias finitas. Si se toma como dirección de búsqueda la dada por el gradiente se tiene el método del gradiente, también llamado de Cauchy.

El método del gradiente es el más inmediato, siendo en el que se basan otros más complejos.

El algoritmo es el siguiente:

- i. Se parte de un punto inicial \underline{x}^k .
- ii. Si $\|\nabla f(\underline{x}^k)\| < \varepsilon$ se termina la optimización. \underline{x}^k es el óptimo buscado y ε es un valor lo suficientemente pequeño para suponer que el gradiente se anula (condición de mínimo).

En otro caso se ir a iii.

- iii. Se define la dirección de búsqueda: $\underline{d}_k = -\frac{\nabla f(\underline{x}^k)}{\|\nabla f(\underline{x}^k)\|}$

- iv. Búsqueda monodimensional en la dirección \underline{d}_k . Se minimiza $f(\underline{x}^k + s_k \cdot \underline{d}_k)$ con $s_k \in [s_{k1}, s_{k2}]$.

- v. Se define el nuevo estado $\underline{x}^{k+1} = \underline{x}^k + s_k \cdot \underline{d}_k$

vi. Ir a i.

3.3.2. Métodos de segundo orden.

Estos métodos utilizan las derivadas primeras y las derivadas segundas de la función a optimizar. El método de Newton es el que mejor representa a los métodos de esta familia. Se utiliza para la búsqueda de raíces de funciones de una variable o para resolver sistemas de ecuaciones. Además, como al optimizar se busca que las derivadas parciales primeras se anulen, el método de Newton también se puede aplicar a este tipo de problemas.

El método de Newton aproxima la función en el entorno del mínimo mediante una función cuadrática, que es la que se minimiza exactamente. Dicha aproximación cuadrática es la representación de Taylor hasta el segundo orden de la función. En esta aproximación se impone gradiente nulo $\nabla f(\underline{x}^k) = 0$ y se aprovecha que la matriz hessiana $\underline{H}(\underline{x}^k) = \nabla^2 f(\underline{x}^k)$ es simétrica para llegar a la ecuación recurrente del método de Newton:

$$\underline{x}^{k+1} = \underline{x}^k - \underline{H}^{-1}(\underline{x}^k) \cdot \nabla^T f(\underline{x}^k) \quad (3.3.3)$$

En la práctica el cálculo de la matriz hessiana puede ser muy costoso, por lo que se suele recurrir a otros métodos. También puede ocurrir que la matriz hessiana sea singular o no sea definida positiva, por lo que este método no podría usarse o tendría que usarse conjuntamente con otros que actuaran cuando el invertir la matriz hessiana no fuera posible. Debido a esto se han desarrollado los métodos de Newton Modificados. Uno de ellos consiste en modificar la matriz hessiana, sustituyéndola por otra próxima a ella pero que sí sea definida positiva. Aún así no se soluciona el problema del coste computacional que conlleva el cálculo de la matriz hessiana. Además de ser costoso el cálculo de la matriz hessiana, también lo es el cálculo de su inversa.

Para solventar estos problemas surgen los algoritmos de métrica variable. En éstos la dirección de búsqueda viene definida como $\underline{d}_k = -\underline{S}^k \cdot \nabla^T f(\underline{x}^k)$, donde \underline{S}^k es una matriz definida positiva. La matriz \underline{S}^k debe reflejar las propiedades de la hessiana de la función para conseguir una convergencia rápida del problema. Para el cálculo de \underline{S}^k no se necesitan las derivadas segundas, reduciéndose así considerablemente el coste del cálculo de la dirección de búsqueda. El método de Cauchy, visto anteriormente, no es más que un caso particular de este método, en el que $\underline{S}^k = \underline{I}$. La velocidad de convergencia de los métodos de métrica variable dependerá de lo cerca que se encuentre \underline{S}^k de $\underline{H}^{-1}(\underline{x}^k)$.

La idea fundamental de estos métodos es la construcción de una aproximación a la matriz inversa del hessiano. Para una mejor convergencia esta aproximación debe ser buena en el entorno del óptimo. La aproximación que se hace del hessiano es la siguiente:

$$\underline{H}(\underline{x}^{k+1}) \approx \frac{\nabla f(\underline{x}^{k+1}) - \nabla f(\underline{x}^k)}{\underline{x}^{k+1} - \underline{x}^k} \quad (3.3.4)$$

denominando $\underline{q}^k = \nabla f(\underline{x}^{k+1}) - \nabla f(\underline{x}^k)$ y $\underline{p}^k = \underline{x}^{k+1} - \underline{x}^k$ la expresión anterior resulta en:

$$\underline{q}^k = \underline{H}(\underline{x}^{k+1}) \cdot \underline{p}^k \quad (3.3.5)$$

Si se supone que el hessiano puede considerarse constante en n sucesivos pasos del algoritmo se podrá hacer:

$$\underline{q}^k = \underline{H} \cdot \underline{p}^k \quad \forall k = 1, 2, \dots, n \quad (3.3.6)$$

y queda determinado \underline{H} para que se cumplan el anterior sistema de ecuaciones.

Para un problema en vías de resolución, si el número de pasos resueltos es $k < n$, la matriz \underline{H}^1 no podrá ser construida totalmente y el sistema de ecuaciones

$$\underline{S}^{i+1} \cdot \underline{q}^i = \underline{p}^i \quad 1 \leq i < n \quad (3.3.8)$$

admite infinitas soluciones de la matriz \underline{S} . Los distintos métodos se distinguen en cómo se defina \underline{S} . Existen varios métodos, pero el que se ha usado es el método de Davidon-Fletcher-Powell. La matriz aproximada de la inversa del hessiano se actualiza en cada paso del algoritmo sumándole dos matrices simétricas de rango unidad. Por ello también se conoce como método de *corrección de rango dos*. Partiendo de (3.3.8) se supone la matriz \underline{S}^{k+1} de la forma:

$$\underline{S}^{k+1} = \underline{S}^k + \underline{A}^k \quad (3.3.9)$$

Sustituyendo la ecuación anterior en (3.3.8) se obtiene un sistema de ecuaciones:

$$\underline{A}^k \cdot \underline{q}^k = \underline{p}^k - \underline{S}^k \cdot \underline{q}^k \quad (3.3.10)$$

cuya solución es de la forma:

$$\underline{A}^k = \frac{\underline{p}^k \cdot \underline{y}^{kT}}{\underline{y}^{kT} \cdot \underline{q}^k} - \frac{\underline{S}^k \cdot \underline{q}^k \cdot \underline{z}^{kT}}{\underline{z}^{kT} \cdot \underline{q}^k} \quad (3.3.11)$$

donde \underline{y}^{kT} , \underline{z}^{kT} son vectores arbitrarios no ortogonales a \underline{q}^k . Éstos se elegirán de tal manera que \underline{A}^k sea simétrica ya que se sabe que \underline{H}^1 lo es. La forma más inmediata de imponer la simetría de \underline{A}^k es hacer $\underline{y}^k = \underline{p}^k$ y $\underline{z}^k = \underline{q}^k$. Sustituyendo en (3.3.11) se obtiene la fórmula de Davidon-Fletcher-Powell:

$$\underline{S}^{k+1} = \frac{\underline{p}^k \cdot \underline{p}^{kT}}{\underline{p}^{kT} \cdot \underline{q}^k} - \frac{\underline{S}^k \cdot \underline{q}^k \cdot \underline{q}^{kT} \cdot \underline{S}^k}{\underline{q}^{kT} \cdot \underline{S}^k \cdot \underline{q}^k} + \underline{S}^k \quad (3.3.12)$$

Se demuestra que si \underline{S}^k es definida positiva \underline{S}^{k+1} también lo será. Por esto basta escoger una matriz inicial \underline{S}^0 definida positiva, como puede ser una matriz diagonal con escalares positivos a lo largo de la diagonal, por ejemplo la matriz identidad.

El algoritmo basado en el presente método es el siguiente:

- i. Definición de una matriz inicial \underline{S}^0 definida positiva y de un punto inicial \underline{x}^0 .
- ii. Obtención de la dirección de búsqueda $\underline{d}_k = \underline{S}^k \cdot \underline{g}^k$, donde $\underline{g}^k = \nabla^T f(\underline{x}^k)$.
- iii. Minimización de $f(\underline{x}^k + s \cdot \underline{d}_k)$ para $s \geq 0$, obteniéndose el nuevo estimado \underline{x}^{k+1} y por tanto también $\underline{g}^{k+1} = \nabla^T f(\underline{x}^{k+1})$
- iv. Definir de $\underline{q}^k = \nabla f(\underline{x}^{k+1}) - \nabla f(\underline{x}^k) = \underline{g}^{k+1} - \underline{g}^k$, $\underline{p}^k = \underline{x}^{k+1} - \underline{x}^k$ y

$$\underline{S}^{k+1} = \frac{\underline{p}^k \cdot \underline{p}^{kT}}{\underline{p}^{kT} \cdot \underline{q}^k} - \frac{\underline{S}^k \cdot \underline{q}^k \cdot \underline{q}^{kT} \cdot \underline{S}^k}{\underline{q}^{kT} \cdot \underline{S}^k \cdot \underline{q}^k} + \underline{S}^k$$

3.4. Búsqueda monodimensional.

Para la búsqueda monodimensional es necesario tener definido un intervalo de la variable independiente en el que se va a realizar la búsqueda del mínimo. La función a optimizar debe ser unimodal en dicho intervalo, es decir, que sólo presente un óptimo local en el intervalo, garantizándose así la unicidad de la solución.

En este caso la búsqueda monodimensional se encuentra integrada en un proceso de búsqueda multivariable. Por ello para aplicar este método se crea el parámetro s_k que varía en la dirección de búsqueda. Éste será el que se utilice como variable independiente en dicho proceso.

$$\underline{x}^{k+1} = \underline{x}^k + s_k \cdot \underline{d}_k \quad (3.4.1)$$

El método que se ha empleado para la búsqueda monodimensional es el de la Sección Áurea. La principal ventaja de este método es que reduce considerablemente el número de evaluaciones de la función objetivo con el consiguiente ahorro de tiempo, especialmente en resoluciones de elevado coste computacional.

Para la aplicación de este método se requiere tener un intervalo $[a^n, b^n]$ de aplicación del mismo. Como en cualquier método de búsqueda se requiere encontrar el mínimo en dicho intervalo mediante la evaluación de la función en distintos puntos. La particularidad del método de la sección áurea reside en la relación existente entre dos intervalos consecutivos. De esta manera en un intervalo se utilizan valores del intervalo anterior, evitándose así nuevas evaluaciones.

En primer lugar se realiza la evaluación de la función objetivo en dos puntos x_1^n y x_2^n que cumplen la relación:

$$\frac{x_2^n - a^n}{b^n - a^n} = \frac{b^n - x_1^n}{b^n - a^n} = \lambda \quad (3.4.2)$$

Al evaluarse la función en dichos puntos pueden darse dos casos:

- $f(x_1^n) < f(x_2^n)$. El nuevo intervalo de incertidumbre será $[a^n, x_2^n]$ y los nuevos puntos en los que evaluar la función serían $x_2^{n+1} = x_1^n$ y x_1^{n+1} se obtiene de la expresión (3.4.2), quedando:

$$\frac{x_2^{n+1} - a^{n+1}}{b^{n+1} - a^{n+1}} = \frac{x_1^n - a^n}{x_2^n - a^n} = \lambda \quad (3.4.3)$$

- $f(x_1^n) > f(x_2^n)$. El nuevo intervalo de incertidumbre será $[x_1^n, b^n]$, luego la función se evaluará en los puntos $x_1^{n+1} = x_2^n$ y el nuevo x_2^{n+1} vendrá dado por la expresión:

$$\frac{b^{n+1} - x_1^{n+1}}{b^{n+1} - a^{n+1}} = \frac{b^n - x_2^n}{b^n - x_1^n} = \lambda \quad (3.4.4)$$

De las expresiones anteriores se deduce que $\lambda = \frac{-1 + \sqrt{5}}{2} \approx 0.618$, que es el número áureo, el cual da nombre al presente algoritmo.

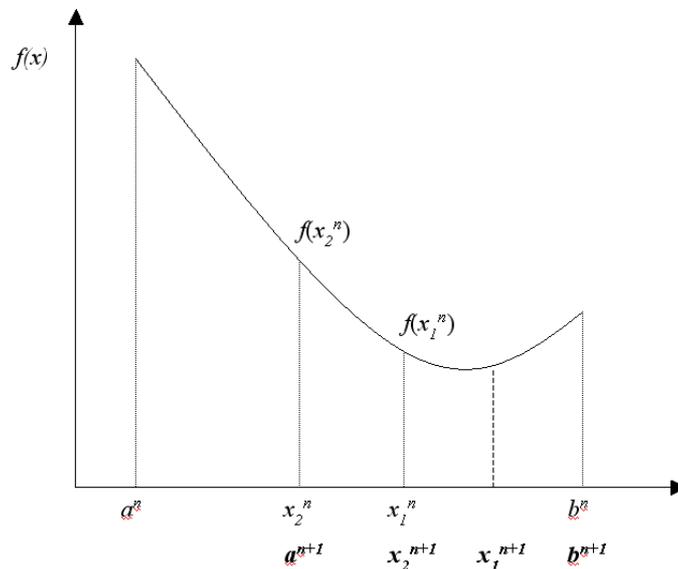


Figura 3.4. Ejemplo de dos intervalos consecutivos del algoritmo de la Sección Áurea

Iterando el proceso anterior se va cerrando el intervalo entorno al mínimo de la función. Para salir de la iteración se definen unas tolerancias para el tamaño del intervalo de incertidumbre y para la variación de la función entre dos intervalos consecutivos. Una vez que

se sale del proceso iterativo por cualquiera de estos dos motivos, se realiza un ajuste cuadrático de los puntos del último intervalo:

$$f(x) = A \cdot x^2 + B \cdot x + C \quad (3.4.5)$$

De esta manera se puede calcular el mínimo y su posición derivando la función anterior.

$$x = -\frac{B}{2 \cdot A} \quad (3.4.6)$$

La obtención de los parámetros se realiza a partir de los tres puntos conocidos del último intervalo mediante el siguiente sistema de ecuaciones de tres incógnitas. Invertiendo la matriz se llega a obtener A , B y C :

$$\begin{pmatrix} (a^{n+1})^2 & a^{n+1} & 1 \\ (b^{n+1})^2 & b^{n+1} & 1 \\ (x_2^{n+1})^2 & x_2^{n+1} & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} f(a^{n+1}) \\ f(b^{n+1}) \\ f(x_2^{n+1}) \end{pmatrix} \quad (3.4.7)$$

Por último basta la sustitución de la ecuación (3.4.6) en (3.4.5) para la obtención del valor de la función en el mínimo.

3.5. Implementación de los algoritmos de optimización.

En primer lugar se han implementado los algoritmos en *FORTRAN* para poder comprobar su funcionamiento con problemas matemáticos. A continuación se procedió a su implementación en *APDL*, lenguaje paramétrico de Ansys. Se han comparado los resultados dados por *ANSYS* y *FORTRAN* mediante su aplicación a problemas de solución analítica conocida. De esta manera se comprueba el correcto funcionamiento de los algoritmos implementados en *APDL*. Los archivos que componen los programas de optimización tanto en *APDL* como en *FORTRAN* se adjuntan en el *ANEXO A*.

3.5.1. Algoritmo de primer orden.

Se ha implementado el algoritmo del método del gradiente, método básico de optimización no lineal. Es un método bastante general aunque de convergencia lenta, característica que se atenúa frente a otros métodos cuando se tiene una región factible pequeña, en la que la función no se diferencia en gran medida de una función lineal.

La implementación, tanto en *FORTRAN* como en *APDL*, ha requerido la creación de varias funciones debido a la complejidad del algoritmo resultante, creándose así una estructura clara que ayuda a la comprensión del proceso de optimización. Hay que decir que los programas de optimización también son paramétricos, es decir, sirven para cualquier problema

con sólo cambiar algunos parámetros como el número de restricciones o de variables de diseño, así como sus valores iniciales e intervalos.

En primer lugar hay que tener bien definido el problema mecánico a estudiar mediante el MEF. De esta manera es necesario que el problema esté definido de forma paramétrica, para que pueda recibir en cada resolución las variables de diseño que el programa de optimización vaya requiriendo. Una vez resuelto el problema, éste debe devolver al programa de optimización el valor de las distintas variables que forman parte de la función objetivo o de las restricciones. A continuación se forma la función objetivo penalizada con las variables devueltas por el modelo. A esta función se le busca el óptimo mediante el método del gradiente y se comprueba la convergencia, actualizando a su vez el parámetro de penalización. Una vez actualizado éste, se tiene otra función penalizada que se vuelve a optimizar mediante el método del gradiente.

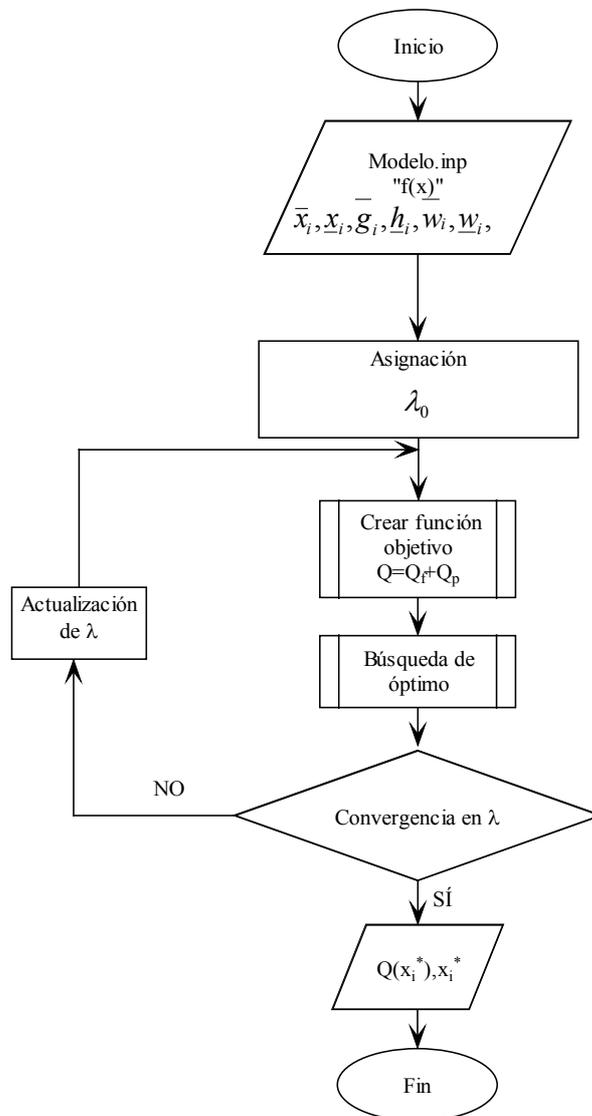


Figura 3.5. Diagrama de flujo del proceso de optimización.

La función objetivo penalizada resultante es del siguiente tipo:

$$Q(\lambda) = P_f + \lambda \cdot (P_g + P_h + P_w) \tag{3.5.1}$$

donde Q es la función objetivo penalizada, λ el parámetro de penalización, P_f la función objetivo adimensionada y P_g , P_h y P_w las funciones de penalización de las restricciones según el signo de la inecuación. La función objetivo original se adimensiona con su valor en el estimado inicial según la siguiente ecuación:

$$P_f = F/F_0 \tag{3.5.2}$$

donde F_0 es el valor de la función objetivo en el estimado inicial.

En los siguientes diagramas de flujo se muestra el proceso de penalización siguiendo el método de penalización interior extendida que se explica en el apartado 3.2.3. así como la generación de la función objetivo penalizada.

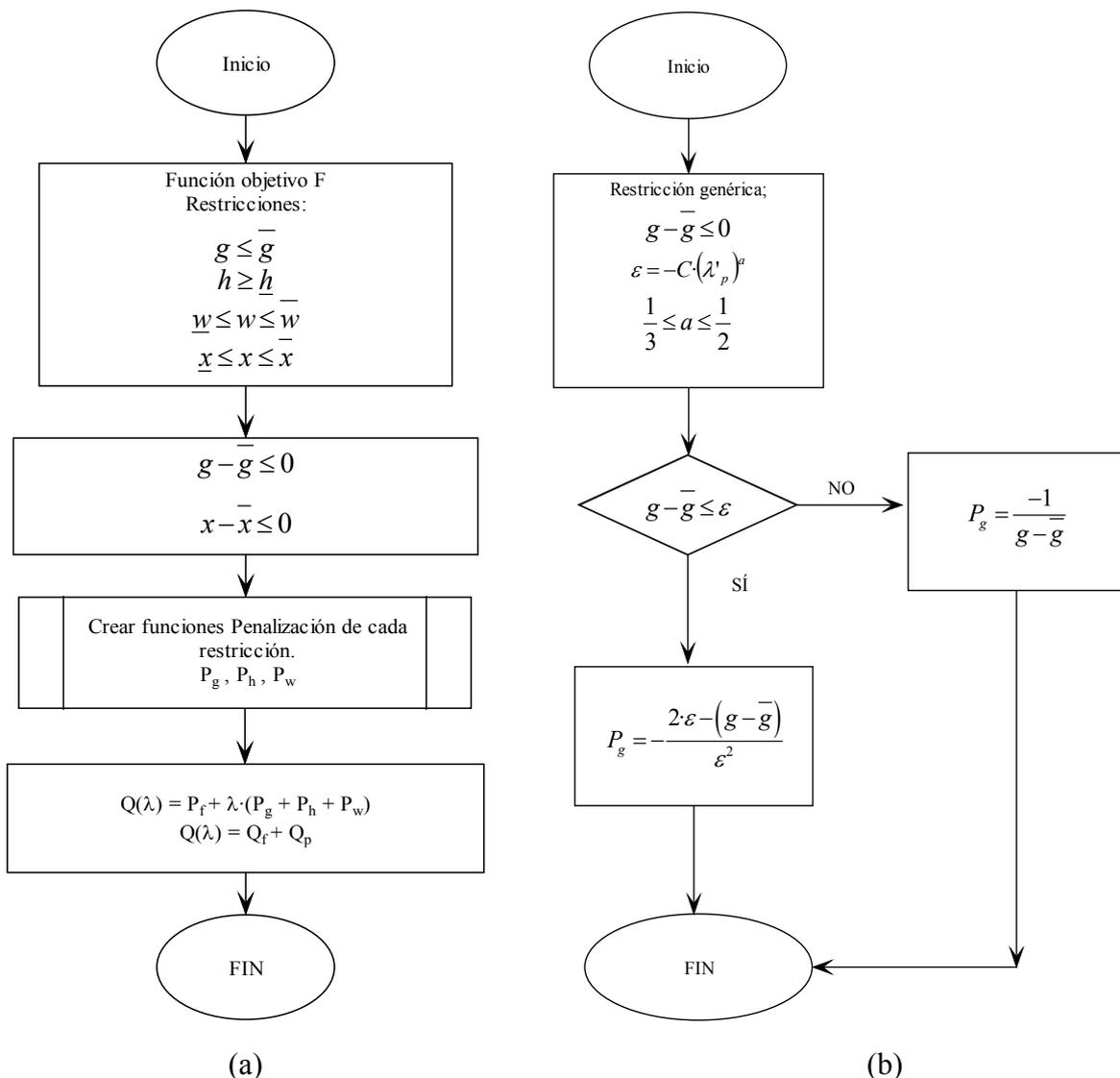


Figura 3.6. Creación de la función objetivo penalizada (a) y penalización de una restricción genérica.

Una vez que se calcula la función objetivo se procede al cálculo de la dirección de búsqueda, que en el método del gradiente viene dada por la de máxima disminución de la función objetivo. El cálculo del gradiente se realiza mediante diferencias finitas. Para ello, partiendo del valor inicial, se aplican pequeños incrementos a cada variable de diseño, calculándose el valor de la función objetivo penalizada para el incremento de cada variable por separado. Por cada variable se tendrá que formar una función objetivo penalizada, lo que implica la resolución de un modelo mecánico por cada variable de diseño para el cálculo del gradiente. Una vez se tiene la dirección se calcula el intervalo de búsqueda que será entre el estimado actual y el valor de las variables de diseño que haga que una llegue a uno de sus límites del rango de validez. De esta manera se imponen la ecuación (3.4.4). El siguiente paso es la minimización monodimensional para la que se utiliza el algoritmo de la sección áurea. Esta búsqueda se realiza según la ecuación (3.3.2). Este proceso de minimización mediante el método del gradiente se muestra en el siguiente diagrama de flujo.

En la figura 3.8 se muestra el proceso de obtención del valor límite de las variables de diseño en la dirección de búsqueda. Como se observa se utiliza un parámetro s que va a caracterizar todo el proceso monodimensional, en primer lugar calculando su valor máximo s_{max} para que las variables de diseño se encuentren en su intervalo de validez y posteriormente siendo el parámetro a variar durante la búsqueda monodimensional mediante el método de la sección áurea.

El proceso de búsqueda monodimensional, mostrado en la figura 3.9, mediante el método de la sección áurea se realiza evaluando la función objetivo en unos puntos determinados en dirección del gradiente. Se comparan sus valores y se va reduciendo el intervalo de minimización, aprovechando para cada iteración de búsqueda todos los valores calculados anteriormente menos uno. Es decir, para cada iteración en la búsqueda monodimensional se evalúa una vez la función objetivo. Llega un momento en que el intervalo o la variación de la función objetivo en dicho intervalo es tan pequeño que se pueden ajustar los puntos mediante una parábola con suficiente aproximación. Lo que se hace llegado a este punto es obtener el mínimo de dicha parábola.

Una vez finalizada cada iteración de optimización se actualiza el punto inicial de las variables de diseño con el óptimo de la iteración anterior. Cuando el óptimo obtenido en una iteración se encuentra lo suficientemente próximo al óptimo anterior se termina el proceso de optimización mediante el método del gradiente. Si el proceso global de optimización no ha terminado se actualiza el parámetro de penalización λ_p y se vuelve a realizar el proceso de

optimización de la función objetivo, penalizada con el nuevo λ_p , mediante el método del gradiente.

El proceso global de optimización finaliza si se dan alguna de las dos situaciones siguientes: que el óptimo obtenido se encuentre próximo al óptimo obtenido en la anterior iteración o que el gradiente sea nulo (condición de óptimo).

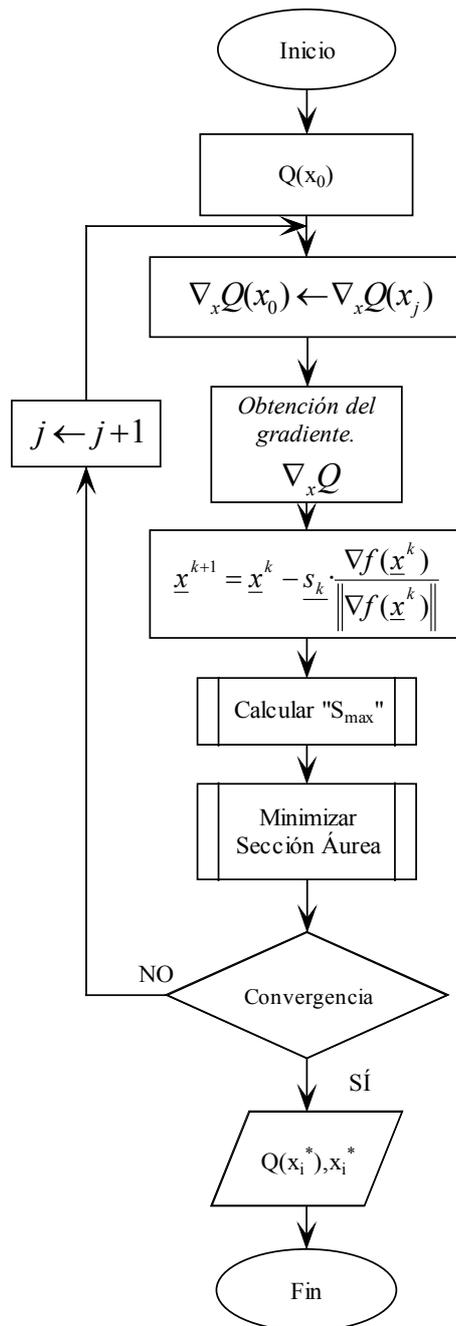


Figura 3.7. Proceso de minimización mediante el método del gradiente.

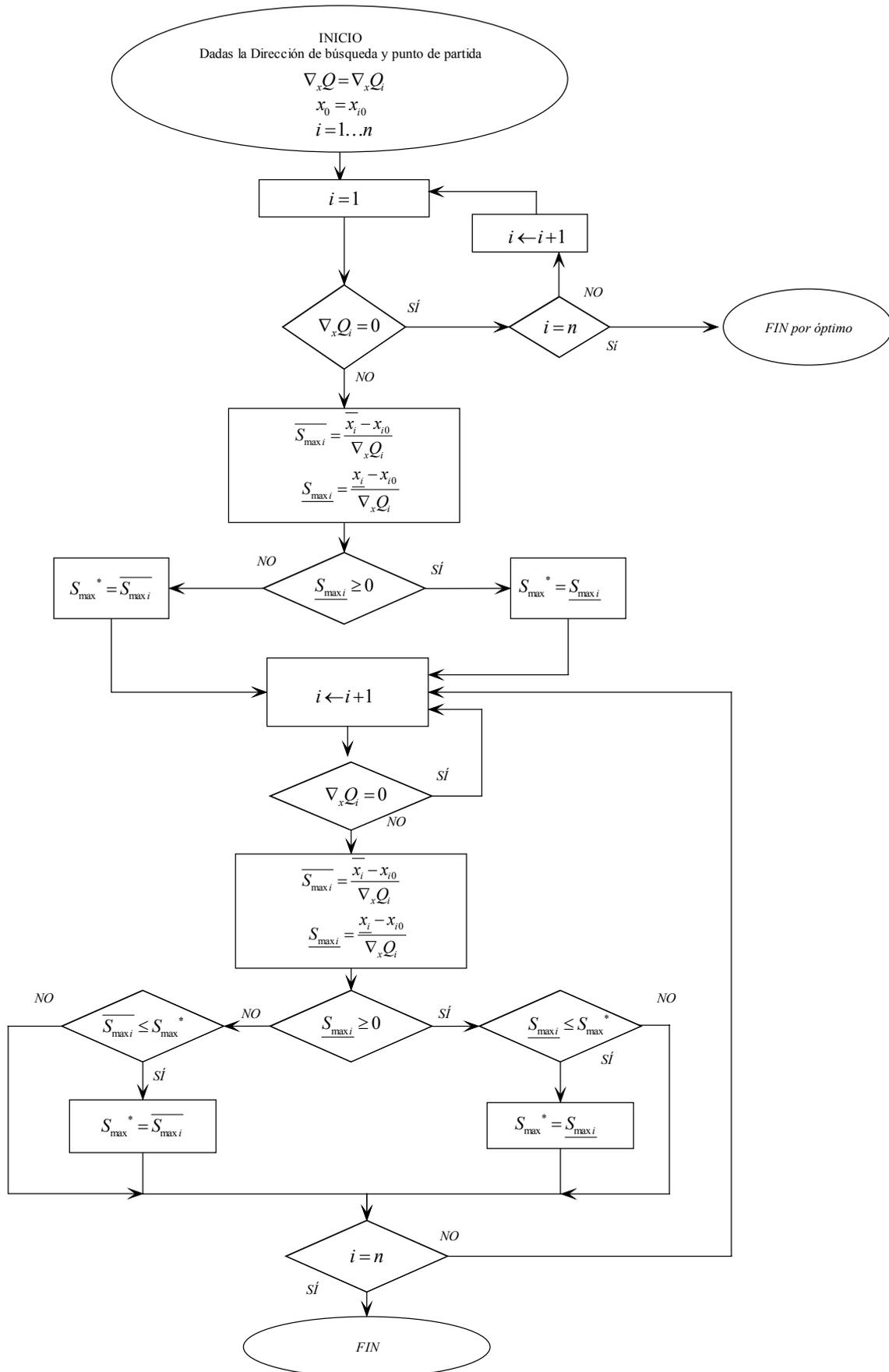


Figura 3.8. Obtención del valor límite de las variables de diseño en la dirección de búsqueda.

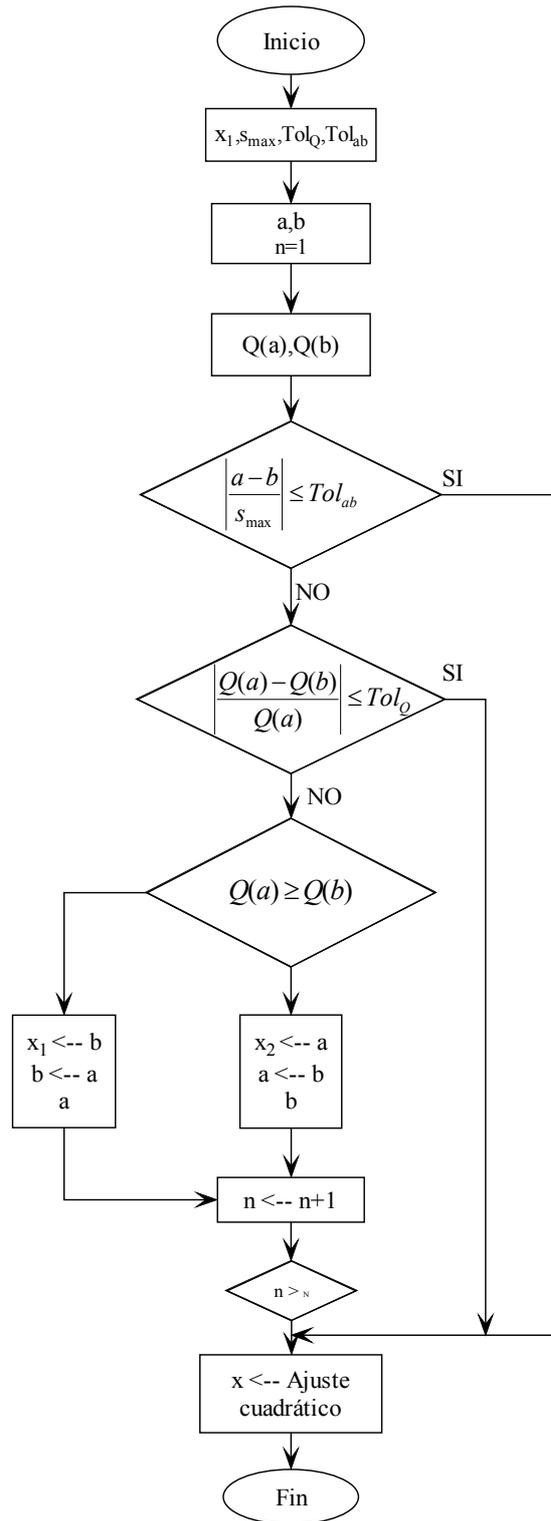


Figura 3.9. Proceso de búsqueda monodimensional mediante el algoritmo de la Sección Áurea.

3.5.2. Algoritmo de segundo orden.

El método del gradiente presenta una convergencia lenta, necesitándose gran número de iteraciones para llegar a la solución óptima. Cuando el tiempo de resolución del modelo mecánico es elevado, resulta interesante la aplicación de métodos de segundo orden. Estos métodos consiguen una convergencia más rápida, reduciéndose así, en general el tiempo de resolución. A pesar de requerir menos iteraciones necesita realizar más operaciones en cada iteración, pero son operaciones matriciales y de almacenamiento de datos que no suponen pérdida de tiempo en la resolución a no ser que se tuviera un número muy elevado de variables de diseño. Aún así, el tiempo requerido para estas operaciones adicionales al método del gradiente es varios órdenes de magnitud inferior, en general, al de resolución de un problema mecánico complejo mediante el MEF. Sólo en el caso de un problema de muy rápida resolución en MEF, y con un número muy elevado de variables de diseño se vería incrementado apreciablemente el tiempo de resolución del cálculo de la dirección de búsqueda mediante un método de segundo orden.

El algoritmo de segundo orden implementado es el método de Davidon-Fletcher-Powell. Lo que lo diferencia del método del gradiente es el cálculo de la dirección de búsqueda. Como se comentó en la explicación del método, se aproxima la inversa de la matriz hessiana por una matriz \underline{S}^k que se construye gracias a los gradientes ya calculados en iteraciones anteriores. Como en la primera iteración no se posee la información de gradientes en iteraciones anteriores, se elige arbitrariamente la matriz \underline{S}^0 con la única condición de que sea definida positiva. En la implementación se ha tomado la matriz identidad, por lo que la primera dirección de búsqueda es la dada por el método del gradiente. A partir de aquí el proceso es idéntico al del método del gradiente sólo que cuando se calcula la dirección de búsqueda ésta viene dada por:

$$\underline{d}_k = -\underline{S}^k \cdot \nabla f(\underline{x}^k) \quad (3.5.1)$$

El diagrama de flujo de este método se muestra en la figura 3.10.

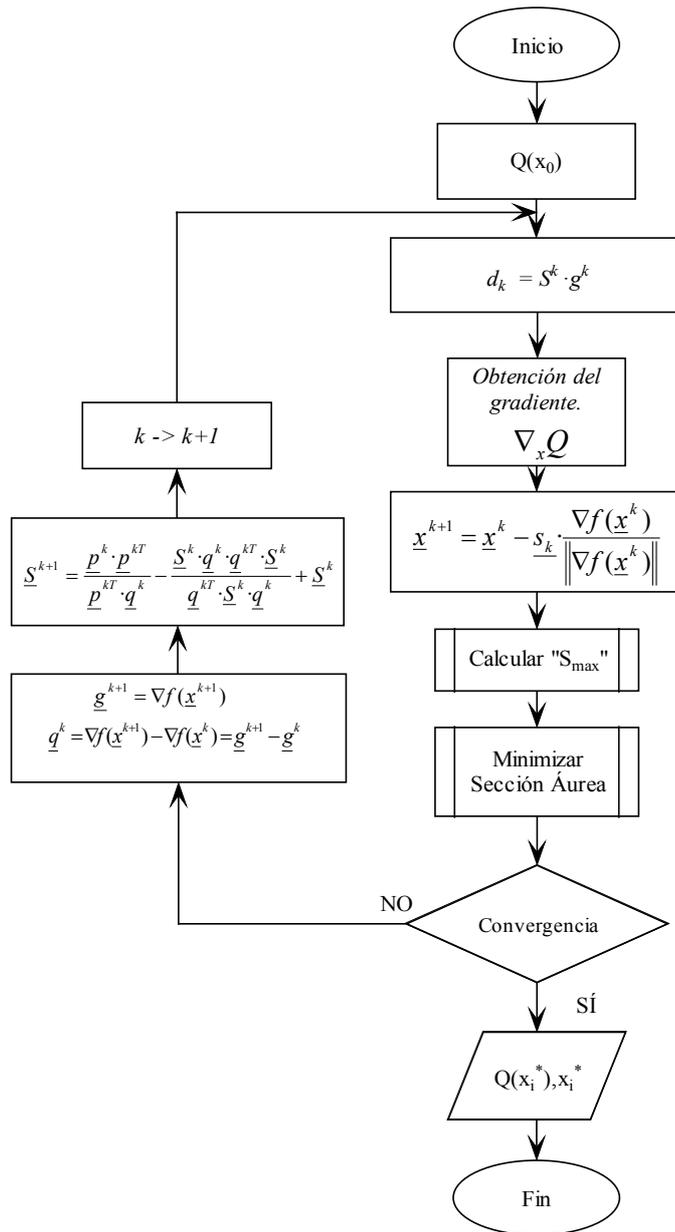


Figura 3.10. Proceso de optimización multidimensional mediante el algoritmo de Davidon-Fletcher-Powell.