

TEMA 10: MÉTODOS NUMÉRICOS PARA LA OPTIMIZACIÓN DE PROBLEMAS NO LINEALES CON RESTRICCIONES

1.- INTRODUCCIÓN

2.- METODOS DE PENALIZACIÓN, BARRERA Y LAGRANGIANA AUMENTADA

2.1.- Métodos de Penalización

2.2.- Método de Penalización Exacta.

2.3.- Lagrangiana Aumentada.

2.4.- Métodos de Barrera.

3.- PROGRAMACIÓN CUADRÁTICA SUCESIVA

4.- MÉTODO DEL GRADIENTE REDUCIDO

4.1.- Gradiente reducido con restricciones lineales.

4.2.- Extensión a sistemas con restricciones no lineales.

5.-ALGUNAS REGLAS PRACTICAS PARA LA FORMULACIÓN DE NLPs.

TEMA 10.

MÉTODOS NUMÉRICOS PARA LA OPTIMIZACIÓN DE PROBLEMAS NO LINEALES CON RESTRICCIONES

1.- INTRODUCCIÓN

En todos los casos la idea básica de los métodos de optimización con restricciones es transformar el problema en otro sin restricciones. Como se comentó anteriormente los métodos más usados son:

1. Métodos de penalización externa.
2. Métodos de penalización interna
3. Métodos de programación lineal sucesiva.
4. Métodos de gradiente reducido
5. Programación cuadrática sucesiva.

Veremos una pequeña introducción a los métodos de penalización. Sin entrar en demasiados detalles, y nos centraremos sobre todo en los métodos de gradiente reducido y de programación cuadrática sucesiva.

2.- METODOS DE PENALIZACIÓN BARRERA Y LAGRANGIANA AUMENTADA.

2.1.- MÉTODOS DE PENALIZACIÓN

La idea esencial de los métodos de penalización consiste en transformar un problema no lineal (NLP) en una secuencia de problemas sin restricciones que presentan la misma solución final.

$$\left. \begin{array}{l} \min : f(x) \\ m.s.a. \quad g(x) \leq 0 \\ h(x) = 0 \end{array} \right\} \longrightarrow \min : P(f, g, h, r)$$

donde P es una función de penalización y 'r' es un parámetro de penalización positivo.

Por ejemplo, considere el siguiente problema:

$$\begin{array}{l} \min : f(x) = (x_1 - 1)^2 + (x_2 - 2)^2 \\ s.a. \quad h(x) = x_1 + x_2 - 4 = 0 \end{array}$$

Se puede transformar en el problema equivalente:

$$\min : P(x, r) = (x_1 - 1)^2 + (x_2 - 2)^2 + r(x_1 + x_2 - 4)^2$$

El principal problema que plantea la función de penalización anterior estriba en que el valor del parámetro r debe aumentar a medida que nos acercamos a la solución. En el límite el producto rh^2 se aproxima a cero, y el valor de $P(x, r)$ se aproxima al de $f(x)$. En el problema anterior el punto óptimo es $x^* = (1.5, 2.5)^T$, y el mínimo del mismo problema

si eliminamos la restricción es $x^*=(1,2)^T$. La Tabla siguiente muestra el efecto del valor del parámetro r en el valor que resulta al resolver el problema P.

r	x_1	x_2	f
0	1.0000	2.0000	0.0000
0.1	1.0833	2.0833	0.0833
1	1.3333	2.3333	0.3333
10	1.4762	2.4762	0.4762
100	1.4975	2.4975	0.4975
1000	1.4998	2.5000	0.4998

El aumento en el valor de r acarrea un problema de “mal condicionamiento”, haciendo enorme el número de condición de la hessiana. $(\nabla^2 P(x, r))$.

La forma general de la función de penalización cuadrática clásica es:

$$P_2(x, r) = f(x) + r \left[\sum_j h_j^2(x) + \sum_j \max(0, g_j(x))^2 \right]$$

2.2.- MÉTODO DE PENALIZACIÓN EXACTA

La penalización exacta trata de evitar los problemas de condicionamiento que aparecen en los problemas que utilizan funciones de penalización cuadráticos clásicos. En ellos se define una función de penalización de la siguiente manera:

$$P_1(x, w_1, w_2) = f(x) + \left[\sum_j w_{1j} |h_j(x)| + \sum_j w_{2j} \max\{0, g_j(x)\} \right]$$

donde w_1, w_2 son parámetros de penalización positivos. De tal manera que cuando se viola una restricción aparece una contribución positiva.

En este caso se evita el problema de condicionamiento debido al **Teorema de la Penalización Exacta** que dice que:

Si x^ es un mínimo local del problema original, entonces si se cumple que:*

$$w_{1j} \geq \left| \lambda_j^* \right| \quad j = 1 \dots m$$

$$w_{2j} \geq \left| \mu_j^* \right| \quad j = 1 \dots r$$

x^ es también un mínimo del problema $P_1(x, w_1, w_2)$*

En este caso no existe problema de mal condicionamiento y no es tampoco necesario aumentar el valor de los parámetros de penalización hasta el infinito.

Intuitivamente se puede observar que el problema de la penalización cuadrática se hace mal condicionado porque la velocidad a la que tiende a cero el término h^2 tiene que ser compensada con el aumento del término de penalización, para llegar a la misma solución que el problema original.

A pesar de todo, los métodos de penalización exacta presentan un grave inconveniente, y es que tanto el término $|h(x)|$ como el término $\max\{0, g(x)\}$ no son diferenciables y por lo tanto no se puede utilizar, de forma fiable, con ellos ningún método de optimización que incluya el cálculo de derivadas.

2.3.- LAGRANGIANA AUMENTADA.

La Lagrangiana aumentada es una función de penalización derivable exacta. Solucionando tanto los problemas que presentan los métodos clásicos de penalización tanto cuadrática como exacta. Por simplicidad consideraremos sólo restricciones de igualdad:

$$AL(x, \lambda, r) = f(x) + \sum_i \lambda_i h_i(x) + r \sum_j h_j^2(x)$$

donde 'r' es un parámetro positivo.

Sea x^* un mínimo local del problema:

$$\begin{aligned} \min f(x) \\ \text{s.a. } h_j(x) = 0 \quad \forall j \end{aligned}$$

y sean λ^* , x^* los valores de los multiplicadores en la solución y el punto solución respectivamente.

El gradiente de AL vendrá dado por:

$$\nabla_x AL = \nabla f(x) + \sum_j \lambda_j \nabla h_j(x) + 2r \sum_j h_j(x) \nabla h_j(x)$$

dado que en el punto solución x^* se debe cumplir que $h(x^*) = 0$ entonces:

$$\nabla_x AL(x^*, \lambda^*, r) = \nabla f(x^*) + \sum_j \lambda_j^* \nabla h_j(x^*) = 0$$

Así pues x^* es un punto estacionario de $AL(x^*, \lambda^*, r)$ para cualquier r . Sin embargo no todos los puntos son mínimos. Para que lo sean es necesario que $\nabla_x^2 AL(x^*, \lambda^*, r)$ sea definida positiva. En general esto se cumple para valores de r por encima de un límite cierto valor límite. El principal inconveniente es que ni el valor límite que debe tomar el parámetro r ni el valor del multiplicador de Lagrange en la solución se conocen a priori.

Existen en la actualidad algunos códigos que trabajan con formulaciones de lagrangiana aumentada, sin embargo, aunque desde un punto de vista teórico se pueden utilizar sin problema, su eficacia ha sido superada por los métodos de gradiente reducido o de programación cuadrática sucesiva.

2.4.- MÉTODOS DE BARRERA

Los métodos de barrera convierten un problema con restricciones en una sucesión de problemas sin restricciones. La solución óptima de estos problemas cae en el interior de la región factible y converge a la solución del problema con restricciones a medida que el parámetro de barrera tiende a cero.

Por ejemplo:

$$\begin{aligned} \min : f(x) &= (x_1 - 1)^2 + (x_2 - 2)^2 \\ \text{s.a. } h(x) &= x_1 + x_2 - 4 = 0 \end{aligned}$$

Se puede transformar utilizando una de las funciones de barrera más conocidas, la función de barrera logarítmica, en el siguiente problema:

$$B(x, r) = f(x) - r \ln[g(x)] = (x_1 - 1)^2 + (x_2 - 2)^2 - r \ln[x_1 + x_2 - 4]$$

donde 'r' es un escalar positivo llamado parámetro de barrera.

A medida que x se aproxima a la restricción, $g(x)$ se aproxima a cero y el término $-r \ln[g(x)]$ se aproxima a infinito. La minimización fuerza a B a encontrar un mínimo sin restricciones en el interior de la región factible, y su localización dependerá del parámetro 'r'. A medida que r tiende a cero la solución tiende a ser del problema original.

Nuevamente la Hessiana se hace mal condicionada cuando r tiende a cero $\nabla_x^2 B(x(r), r)$ tiende a infinito.

3.- PROGRAMACIÓN CUADRÁTICA SUCESIVA (SQP)

La idea básica de la programación cuadrática sucesiva es resolver las condiciones necesarias de optimalidad de Karush – Kuhn –Tucker utilizando el método de Newton (o métodos cuasi Newton).

Dado un problema general de programación no lineal:

$$\begin{aligned} \text{Min } & f(x) \\ \text{s.a. } & h(x)=0 \\ & g(x)\leq 0 \end{aligned} \quad (1)$$

Las condiciones de KKT se pueden escribir como:

$$\begin{aligned} \nabla_x L(x^*, \lambda^*, \mu^*) &= \nabla f(x^*) + \lambda^T \nabla h(x^*) + \mu^T \nabla g_j(x^*) = 0 \quad j \in J \\ h(x^*) &= 0 \\ g_j(x^*) &= 0 \quad j \in J \end{aligned} \quad (2)$$

donde J representa el conjunto de restricciones activas.

Si suponemos un punto (x^i, λ^i, μ^i) y aplicamos el método de Newton, a las condiciones necesarias nos queda:

$$\left. \begin{aligned} \nabla_x L(x^i, \lambda^i, \mu^i) &\approx \nabla_x f(x^i) + \nabla h(x^i) \lambda^i + \nabla_x g_j(x^i) \mu^i + \nabla_{xx} f(x^i) (x - x^i) + \\ &(\lambda^i)^T \nabla_{xx} h(x^i) (x - x^i) + \nabla_x h(x^i) (\lambda - \lambda^i) + (\mu^i)^T \nabla_{xx} g_j(x^i) (x - x^i) + \nabla_x g_j(x^i) (\mu - \mu^i) = 0 \\ h(x^i) + \nabla h(x^i)^T (x - x^i) &= 0 \\ g_j(x^i) + \nabla g_j(x^i)^T (x - x^i) &= 0 \end{aligned} \right\} \quad (3)$$

Quizás en su forma matricial el problema aparezca más claro que lo único que se ha hecho hasta el momento es aplicar un paso del método de Newton.

$$\begin{bmatrix} \nabla_{xx}L & \nabla g & \nabla h \\ \nabla g & 0 & 0 \\ \nabla h & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \mu \\ \nabla \lambda \end{bmatrix} = - \begin{bmatrix} \nabla_x L(x^i, \lambda^i, \mu^i) \\ g(x^i) \\ h(x^i) \end{bmatrix}$$

El sistema anterior lo podemos re-escribir como:

$$\left. \begin{aligned} \nabla f(x^i) + \nabla_x h(x^i) \lambda + \nabla_x g_j(x^i) \mu + \nabla_{xx} L(x^i, \lambda^i, \mu^i) (x - x^i) &= 0 \\ h(x^i) + \nabla h(x^i)^T (x - x^i) &= 0 \\ g_j(x^i) + \nabla g_j(x^i)^T (x - x^i) &= 0 \end{aligned} \right\} \quad (4)$$

donde:

$$\nabla_{xx} L(x^i, \lambda^i, \mu^i) (x - x^i) = \nabla_{xx} f(x^i) (x - x^i) + (\lambda^i)^T \nabla_{xx} h(x^i) (x - x^i) + (\mu^i)^T \nabla_{xx} g_j(x^i) (x - x^i)$$

Si nos fijamos en la ecuación (4) nos podemos dar cuenta de que son las condiciones necesarias de optimalidad de otro problema que podemos escribir como:

$$\begin{aligned} \text{Min} \quad & \nabla f(x^i)^T d + \frac{1}{2} d^T \nabla_{xx} L(x^i, \lambda^i, \mu^i) d \\ \text{s. a.} \quad & h(x^i) + \nabla h(x^i)^T d = 0 \\ & g_j(x^i) + \nabla g_j(x^i)^T d = 0 \end{aligned} \quad (5)$$

donde:

$$d = (x - x^i)$$

El problema planteado por las ecuaciones (5) es lo que se conoce como programación cuadrática. Es decir hemos transformado nuestro problema original en una sucesión de problemas cuadráticos (señalar que existen métodos muy eficientes para resolver un problema cuadrático). Sin embargo rara vez se utiliza el algoritmo de forma directa tal como lo hemos planteado. Las razones son fundamentalmente dos, la primera es que puede no converger a una solución local del problema y la segunda es que resulta costoso desde un punto de vista numérico.

Comencemos por el costo numérico. La razón por la que el método es numéricamente costoso es que hay que calcular la matriz hessiana de la función objetivo y de las restricciones activas. Esto implica calcular las derivadas segundas con respecto a las variables. Una forma de reducir el costo numérico del algoritmo es utilizar aproximaciones de tipo cuasi Newton de la hessiana de la lagrangiana. Evitándonos así el cálculo de las derivadas. A pesar de que la velocidad de convergencia teórica es ligeramente inferior a la que se obtiene con el método de Newton tiene dos ventajas, la actualización permite mantener una hessiana simétrica definida positiva, característica que podría perderse debido a los errores de redondeo al calcular derivadas o que simplemente la matriz hessiana real no fuese definida positiva, lo que nos asegura que la solución del problema cuadrático (QP) es única. Y además en problemas grandes permite mantener la matriz en una factorización adecuada para su resolución.

Por otra parte los resultados de convergencia para programación cuadrática sucesiva se obtienen insistiendo en que $(x^{i+1}, \lambda^{i+1}, \mu^{i+1})$ son una estimación mejor que la solución en (x^i, λ^i, μ^i) . En un problema sin restricciones el progreso se mide en términos de la función objetivo. En el caso de optimización con restricciones es común utilizar una función auxiliar llamada “función de mérito” o en ocasiones también lagrangiana aumentada. Habitualmente una función de mérito es una suma de términos que incluye la función objetivo y la cantidad de infactibilidad de las restricciones. Si el nuevo punto reduce la función objetivo y reduce las infactibilidades, entonces la función de mérito se reduce. Sin embargo, en muchas ocasiones la reducción de la función de mérito es a costa de las infactibilidades o viceversa.

El uso de la función de mérito elimina además otro problema, y es que el método de Newton podría no converger desde puntos no lo suficientemente cercanos al óptimo. Así pues en lugar de realizar un paso completo de Newton hacemos una búsqueda unidireccional en la dirección d .

Nos queda el punto de como decidir qué restricciones van a ser activas y cuales no. Una posibilidad es introducirlas todas en el programa cuadrático y tomar como conjunto activo aquellas que son activas en el QP.

Algoritmo SQP Básico

0. Suponer x^0 , Hacer $H^0=I$ (matriz identidad) y evaluar $f(x^0)$, $g(x^0)$, y $h(x^0)$
1. En el punto x^i calcular $\nabla f(x^i)$, $\nabla g(x^i)$, $\nabla h(x^i)$
2. Si $i>0$ actualizar H^i utilizando la fórmula BFGS:

$$H^{i+1} = H^i - \frac{H^i s s^T H^i}{s^T H^i s} + \frac{y y^T}{s^T y} \quad \text{con}$$

$$s = x^{i+1} - x^i$$

$$y = \nabla L(x^{i+1}, \lambda^{i+1}, \mu^{i+1}) - \nabla L(x^i, \lambda^{i+1}, \mu^{i+1})$$

3. Resolver el siguiente QP (utilizando una estrategia activa como la que vimos en capítulos anteriores, un método basado en programación lineal o cualquiera de los disponibles para este tipo de problemas):

$$\begin{aligned} \underset{d}{\text{Min}} \quad & \nabla f(x^i)^T d + \frac{1}{2} d^T H^i d \\ \text{s.a.} \quad & h(x^i) + \nabla h(x^i)^T d = 0 \\ & g(x^i) + \nabla g(x^i)^T d \leq 0 \end{aligned}$$

4. Encontrar un α tal que una función de mérito cumpla:

$$\Psi(x^i + \alpha d) < \Psi(x^i) \quad 0 < \alpha \leq 1$$

Como funciones de mérito se pueden usar:

Función de penalización exacta:

$$\Psi_p(x) = f(x) + \rho \left[\sum_j \max(0, g_j(x)) + \sum_{i=1}^m |h_i(x)| \right] \quad j \in J$$

$$\rho > \max_{i,j} \{ |\lambda_i|, |\mu_j| \}$$

donde ρ es el parámetro de penalización

Función de Lagrange aumentada

$$\Psi_p(x) = f(x) + \mu_j^T g_j(x) + \lambda_i^T h(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^m (h_i(x))^2 + \sum_j (\max(0, g_j(x)))^2 \right\}$$

donde el parámetro de penalización ρ debe cumplir:

$$-\frac{1}{2\rho} [h^T(x)\lambda + g^T(x)\mu] - h^T(x)h(x) - g_j^T(x)g_j(x) \leq 0$$

5. Hacer $x^{i+1} = x^i + d$, y $i=i+1$. Volver al paso 1.

Si después del paso 3 el error de las condiciones de KKT es menor que una tolerancia pre especificada entonces parar.

$$|\nabla L^T d| \leq |\nabla f^T d| + \sum |\mu_i g_i| + \sum |\lambda_i h_i| \leq \varepsilon$$

o bien

$$\|\nabla L\| \text{ y } \|g_j\|, \|h\| \leq \varepsilon$$

La gran ventaja de los métodos de programación cuadrática sucesiva es que convergen de forma muy rápida con muy poco número de evaluación de las funciones. Sin embargo para problemas con un elevado número de grados de libertad quizás sea mejores los métodos de gradiente reducido, en tanto que reducen la dimensionalidad del problema a resolver.

4.- MÉTODOS DE GRADIENTE REDUCIDO

Dado un problema de optimización con m restricciones que vienen dadas en forma de igualdad (forma estándar), y n variables ($n \geq m$). Una forma directa de resolver el problema consiste en plantear el problema en función de los $n-m$ grados de libertad. En otras palabras, sería posible pensar que se pueden despejar m variables (tantas como restricciones) en función de las $n-m$ restantes y resolver el problema como un problema sin restricciones. Esta es la idea básica que está detrás de los métodos de gradiente reducido. Por supuesto, no siempre es posible hacer esta eliminación de variables de forma algebraica, al menos con ecuaciones no lineales. Sin embargo, en un problema donde las restricciones son lineales bien planteado (no tenemos problemas con el cálculo de las inversas asociadas a los diferentes conjuntos de variables dependientes) esta eliminación sí que es posible. Comencemos considerando el caso del problema con restricciones lineales.

4.1.- GRADIENTE REDUCIDO CON RESTRICCIONES LINEALES

Sea el problema:

$$\begin{array}{ll} \min : & f(x) \\ \text{s.a} & Ax=b \end{array}$$

Donde $f(x)$ es una función objetivo escalar, continua diferenciable al menos dos veces y no necesariamente lineal. Y el conjunto $Ax = b$ define un sistema de m restricciones lineales con n variables; donde $n \geq m$.

Es posible hacer una partición de variables

$$x = \begin{pmatrix} x_D \\ x_I \end{pmatrix} \begin{array}{l} \longrightarrow \text{Variables dependientes o básicas (m variables)} \\ \longrightarrow \text{Variables independientes o superbásicas (n-m variables)} \end{array}$$

Así mismo también podemos dividir la matriz A en dos partes:

$A=(B|N)$ B es una matriz $m \times m$ invertible que tiene asociados los coeficientes de los vectores de las variables dependientes
N es una matriz $m \times (n-m)$ formada por los coeficientes de los vectores de las variables independientes

Así pues :

$$Ax=b \Rightarrow Bx_D + Nx_I = b \quad \text{Despejando } x_D$$

$$x_D = B^{-1}b - B^{-1}Nx_I \quad \text{o sea:}$$

$$x = \begin{pmatrix} B^{-1}b - B^{-1}Nx_I \\ x_I \end{pmatrix}$$

Sustituyendo en la función objetivo se obtiene que:

$$f(x) = f(x_D, x_I) = f(B^{-1}b - B^{-1}Nx_I)$$

Se llama **gradiente reducido** al gradiente respecto a las $n-m$ variables independientes, así pues:

$$g_R^T = \frac{\partial f}{\partial x_I} + \frac{\partial f}{\partial x_D} \frac{\partial x_D}{\partial x_I} = \nabla_{x_I} f(x)^T - \nabla_{x_D} f(x)^T B^{-1}N$$

Matriz de transformación

Considere, por otra parte, un Δx :

$$\Delta x = \begin{pmatrix} x_D^{k+1} - x_D^k \\ x_I^{k+1} - x_I^k \end{pmatrix} = \begin{pmatrix} B^{-1}b - B^{-1}Nx_I^{k+1} - B^{-1}b + B^{-1}Nx_I^k \\ x_I^{k+1} - x_I^k \end{pmatrix} = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} \Delta x_I = Z \Delta x_I$$

A la matriz $Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}$ se la conoce como **matriz de transformación**.

El gradiente reducido y la hessiana reducida en función de la matriz de transformación vendrán dados por:

$$g_R^T = \nabla f(x)^T Z$$

$$H_R = Z^T \nabla_{xx} f(x) Z$$

Se ha transformado un problema con restricciones en otro sin restricciones en el espacio de variables reducidas. Se puede, por lo tanto, aplicar los mismos métodos que se utilizaban en el caso de problemas sin restricciones. Dado que los métodos más efectivos son los basados en el método de Newton veamos como podemos aplicar dicho método en el espacio de variables reducidas.

Método de Newton en el espacio de variables reducidas:

Igual que se procedía en la deducción del método de Newton en problemas sin restricciones comenzamos por desarrollar en serie alrededor del punto x hasta segundo orden, así :

$$f(x^{k+1}) = f(x^k) + \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T \nabla_{xx} f(x) \Delta x$$

Recordando que $\Delta x = Z \Delta x_I$ y sustituyendo, se tiene que

$$f(x^{k+1}) = f(x^k) + \nabla f(x)^T Z \Delta x_I + \frac{1}{2} \Delta x_I^T Z^T \nabla_{xx} f(x) Z \Delta x_I =$$

$$= f(x^k) + g_R^T \Delta x_I + \frac{1}{2} \Delta x_I^T H_R \Delta x_I$$

Recordando que la condición de optimalidad es:

$$\frac{\partial f(x)}{\partial x_I} = 0 \quad \Rightarrow \quad g_R + H_R \Delta x_I = 0$$

Así pues un paso del método de Newton es equivalente a resolver:

$$H_R \Delta x_I = -g_R \quad \text{o bien} \quad \Delta x_I = -H_R^{-1} g_R$$

NOTAS:

1. La Hessiana reducida se puede sustituir por una estimación utilizando un método cuasi-Newton, como, por ejemplo BFGS.
2. Es posible calcular g_R sin necesidad de hallar explícitamente B^{-1} ; para ello como:

$$g_R^T = \nabla f(x)^T Z = -\nabla_{x_D} f(x)^T B^{-1} N + \nabla_{x_I} f(x)^T$$

si se resuelve el sistema auxiliar : $B^T w = -\nabla_{x_D} f(x)$

Como $(B^T w) = w^T B$

resulta que : $w^T = -\nabla_{x_D} f(x)^T B^{-1}$

o sea : $g_R^T = w^T N + \nabla_{x_I} f(x)^T$

4.2.- EXTENSIÓN A SISTEMAS CON RESTRICCIONES NO LINEALES

El problema que vamos a considerar en este caso puede escribirse como:

$$\begin{aligned} \min & f(x) \\ \text{s.a.} & h_j(x) = 0 \quad j = 1, \dots, m \\ & L_i \leq x_i \leq U_i \quad i = 1, \dots, n \end{aligned}$$

donde L_i y U_i son los límites inferior y superior de x_i respectivamente. Estos límites se tratan como vectores separados, en lugar de ser clasificados como desigualdades (restricciones) porque se tratan de forma diferente cuando se determina la longitud de paso en la dirección de búsqueda.

El problema debe venir dado en forma estándar, así pues para aquellas restricciones en forma de desigualdad debemos añadir variables de holgura de tal forma que:

$$\begin{aligned} h_j(x) &= g_j(x) + \sigma = 0 \\ \sigma &\geq 0 \end{aligned}$$

Al igual que en apartado anterior podemos dividir las variables en dependientes e independientes, sin embargo, no es posible, en general, en un sistema de m restricciones no lineales m variables en función de las n-m restantes. No obstante, si linealizamos las restricciones –las aproximamos por su plano tangente en un punto dado- podríamos proceder como en el caso anterior, así pues, sustituimos momentáneamente el problema original por otro con las restricciones linealizadas:

$$\begin{aligned} \min : & f(x) \\ \text{s.a.} & h(x^i) + \nabla h(x^i) \Delta x = 0 \end{aligned}$$

Donde la matriz $\nabla h(x^i)$ que vamos a llamar $J(x^i)$, (Matriz Jacobiana de las restricciones). También la puedo dividir en dos partes, una asociada a las variables dependientes y otra asociada a las variables independientes:

$$J(x^i) = [J_D(x), J_I(x)]$$

Podemos aplicar, por lo tanto, el método visto en el apartado referente a problemas con restricciones lineales. En este caso tendremos que:

$$Z = \begin{pmatrix} -J_D^{-1} J_I \\ I \end{pmatrix}$$

Y por lo tanto:

$$g_R^T = \nabla f(x)^T Z = \nabla_{x_I} f(x)^T - \nabla_{x_D} f(x)^T J_D^{-1} J_I$$

Una vez que se ha efectuado un paso del algoritmo de optimización (por ejemplo del método de Newton) dado que se está haciendo una aproximación lineal del conjunto de restricciones, lo más probable es que el conjunto de restricciones original ($h(x)=0$) no se satisfaga. En otras palabras, el nuevo punto obtenido es un punto no factible.

Los algoritmos se dividen en dos tipos dependiendo del tipo de procedimiento seguido desde este punto: algoritmos sin restauración de la factibilidad y algoritmos con restauración de la factibilidad.

Algoritmo básico sin restauración de la factibilidad.

Quizás el algoritmo más conocido que trabaja sin restauración de la factibilidad está implementado en el solver MINOS. Para evitar minimizar el efecto de las no

factibilidades resuelve una sucesión de problemas utilizando el procedimiento anterior añadiendo a la función objetivo un término de penalización (lagrangiana aumentada):

$$\begin{aligned} \text{Pk:} \quad \min : & \quad f(x) + (\lambda^k)^T (h(x) - h(x^k)) \\ \text{s.a.} \quad & \quad h(x^k) + \nabla h(x^k) \Delta x = 0 \end{aligned}$$

Donde λ^k es el multiplicador de lagrange en la iteración k. El problema puede seguir un camino no-factible hasta el final.

Algoritmo básico con restauración de factibilidad (Gradiente Reducido Generalizado)

Resumiremos aquí los pasos básicos del algoritmo de gradiente reducido generalizado. Cualquier código puede presentar variaciones basadas en procedimientos para la manipulación de matrices, selección de elementos de dirección de búsqueda, u otras decisiones basadas en la experiencia. Se presenta aquí el algoritmo básico, su equivalente sin restricciones sería el método de máximo descenso (aunque es fácil convertirlo en un método tipo Newton o cuasi-newton). Este es el algoritmo básico que aparece en el conocido Solver de Excel.

Suponemos que $\left(\frac{\partial h}{\partial x_D} \right)$ no es singular. Dividiremos el algoritmo en 5 pasos.

1. Determinación de las componentes de búsqueda para las variables independientes.

Comenzamos el algoritmo desde un punto factible x^k (en la etapa k) y se calcula el gradiente reducido.

$$(g_R^k)^T = \left(\frac{\partial f^k}{\partial x_I^k} \right)^T - \left(\frac{\partial f^k}{\partial x_D^k} \right)^T \left(\frac{\partial h^k}{\partial x_D^k} \right)^{-1} \left(\frac{\partial h^k}{\partial x_I^k} \right)$$

Nótese que en la primera etapa ($x^k = x^0$) se deben designar las variables independientes. En principio la designación de dichas variables se puede hacer o bien seleccionando las cantidades controlables del proceso, o bien por consideraciones estructurales.

Las componentes de la dirección de búsqueda, en el espacio de las variables independientes se establece a partir de los elementos del gradiente reducido como sigue:

- a. Si x_i^k está en uno de los límites, la dirección del componente es $\Delta_i^k=0$ si el paso excede el límite. Esto es:

$$\begin{aligned} x_i^k &= U_i & g_{Ri} < 0 \\ x_i^k &= L_i & g_{Ri} > 0 \end{aligned}$$

- b. Si $L_i < x_i < U_i$ la dirección de búsqueda es la negativa del correspondiente elemento del gradiente reducido: $\Delta_i^k = -g_{Ri}$. En algunos algoritmos solamente la dirección original es la negativa al correspondiente gradiente reducido, las iteraciones siguientes usan la fórmula del gradiente conjugado.

2. Determinación de las componentes de búsqueda de las variables dependientes

Se calculan a partir de la relación:

$$\Delta_D^k = - \left(\frac{\partial h^k}{\partial x_D} \right)^{-1} \left(\frac{\partial h^k}{\partial x_I} \right) \Delta_I^k$$

La relación anterior viene de la generalización de la ecuación

$$\begin{aligned} dh &= \frac{\partial h}{\partial x_I} dx_I + \frac{\partial h}{\partial x_D} dx_D = 0 \\ \frac{dx_D}{dx_I} &= - \left(\frac{\partial h}{\partial x_D} \right)^{-1} \left(\frac{\partial h}{\partial x_I} \right) \quad \text{ya que } dx_D \text{ se deba calcular a partir de } dx_I \end{aligned}$$

3. Mejora de la función objetivo:

Se minimiza $f(x_I^k + \alpha \Delta_I^k, x_D^k + \alpha \Delta_D^k)$ con respecto a α ($\alpha > 0$) mediante una búsqueda unidimensional.. La fase 3 nos lleva a:

$$\begin{aligned}x_I^{k+1} &= x_I^k + \alpha^k \Delta_I^k \\ \tilde{x}_D^{k+1} &= x_D^k + \alpha^k \Delta_D^k\end{aligned}$$

Dado que el algoritmo que nos interesa es de camino factible el valor de \tilde{x}_D^{k+1} solamente un valor posible. En otras palabras, debido a que las restricciones pueden no ser lineales, la búsqueda unidimensional puede llevar a valores no factibles de las variables x_D .

4. Búsqueda de una solución factible de las variables dependientes

Dado que las componentes de \tilde{x}_D^{k+1} podrían ser puntos no factibles, es decir $h(x_I^{k+1}, \tilde{x}_D^{k+1}) \neq 0$ debemos modificar \tilde{x}_D^{k+1} para volver a la región factible, para ello utilizamos el método de Newton:

$$x_D^{k+1} = \tilde{x}_D^{k+1} - \left[\frac{\partial h(x_I^{k+1}, \tilde{x}_D^{k+1})}{\partial x_D^k} \right]^{-1} h(x_I^{k+1}, \tilde{x}_D^{k+1})$$

5. ¿Fallo del método de Newton?

- Si x^{k+1} es un punto factible y si $f(x_I^{k+1}, x_D^{k+1}) < f(x_I^k, x_D^k)$ se toma x^{k+1} y se pasa a la etapa 1.
- Si x^{k+1} es un punto factible, pero $f(x_I^{k+1}, x_D^{k+1}) > f(x_I^k, x_D^k)$ se reduce el valor de α (en 1/4 o 1/10) y se comienza otra vez la fase 3.
- Si el método de Newton falla en encontrar un punto factible (en digamos unas 20 iteraciones) se reduce α en factor de 1/2 o 1/10 y se vuelve a la fase 4.
- Si todos los pasos anteriores fallasen se podría recurrir a un cambio de base, intercambiando las variables dependientes y las independientes.

Acotaciones:

Si la variable x_i alcanza uno de sus límites (acotación) entonces es eliminada de la base tomando su valor límite y se cambia por otra variable x_D . La iteración de Newton se lleva a cabo para el nuevo conjunto de variables dependientes.

Terminación:

Se puede aplicar cualquier criterio de terminación, pero el mas simple es comprobar si $\left| \Delta_I^k \right| \leq \varepsilon$ después de la fase 3.

Gradiente reducido para n variables y m restricciones.

x_I Variables independientes. Vector de (n - m) variables
 x_D Variables dependientes. Vector de m variables

$$h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix} \quad \text{vector (m x 1)} \quad \text{con} \quad x^T = [x_I, x_D]$$

$$\frac{\partial h}{\partial x_D} = \begin{bmatrix} \frac{\partial h_1(x)}{\partial x_1} & \dots & \frac{\partial h_1(x)}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial h_m(x)}{\partial x_1} & \dots & \frac{\partial h_m(x)}{\partial x_m} \end{bmatrix} \quad \text{matriz (m x m) matriz base.}$$

$$\left(\frac{\partial f(x)}{\partial x_I} \right)^T = \nabla_{x_I}^T f(x) = \left[\frac{\partial f(x)}{\partial x_{m+1}}, \dots, \frac{\partial f(x)}{\partial x_n} \right] \quad \text{vector} \quad 1 \times (n-m)$$

$$\left(\frac{\partial f(x)}{\partial x_D} \right)^T = \nabla_{x_D}^T f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_m} \right] \quad \text{vector} \quad (1 \times m)$$

$$\frac{dx_D}{dx_I} = \begin{bmatrix} \frac{dx_1}{dx_{m+1}} & \dots & \frac{dx_1}{dx_n} \\ \vdots & & \vdots \\ \frac{dx_m}{dx_{m+1}} & \dots & \frac{dx_m}{dx_n} \end{bmatrix} \quad \text{matriz} \quad m \times (n-m)$$

$$\frac{dh(x)}{dx_I} = \begin{bmatrix} \frac{dh_1}{dx_{m+1}} & \dots & \frac{dh_1}{dx_n} \\ \vdots & & \vdots \\ \frac{dh_m}{dx_{m+1}} & \dots & \frac{dh_m}{dx_n} \end{bmatrix} \quad \text{matriz} \quad m \times (n-m)$$

$$g_R^T = \left(\frac{\partial f(x)}{\partial x_I} \right)^T - \left(\frac{\partial f(x)}{\partial x_D} \right)^T \left(\frac{\partial h(x)}{\partial x_D} \right)^{-1} \left(\frac{\partial h(x)}{\partial x_I} \right) \quad \text{o bien}$$

$$g_R = \left(\frac{\partial f(x)}{\partial x_I} \right) - \left(\frac{\partial h(x)}{\partial x_I} \right)^T \left[\left(\frac{\partial h(x)}{\partial x_D} \right)^{-1} \right]^T \left(\frac{\partial f(x)}{\partial x_D} \right)$$

5.- ALGUNAS REGLAS PRÁCTICAS PARA LA FORMULACIÓN DE NLPs.

1. Tratar de especificar los límites superior e inferior de todas las variables. Esto no significa introducir nuevas restricciones al problema, ya que estos límites se tratan de forma independiente en todos los algoritmos de programación matemática y ayudan a definir el problema.
2. Tratar de formular el problema utilizando restricciones lineales:

$$\frac{x_1}{x_2} \leq 2 \quad \longrightarrow \quad x_1 = 2 x_2$$

3. Eliminar variables sólo si no se introducen nuevas no linealidades o se pierden las características de convexidad del problema:

Ej₁:

$$\left. \begin{array}{l} \text{Min } x_1 + x_2 \\ \text{s.a. } x_1 x_2 - 2 = 0 \\ 0 \leq x_1 \leq 2 \\ 0 \leq x_2 \leq 2 \end{array} \right\} \xrightarrow{x_2 = \frac{2}{x_1}} \left. \begin{array}{l} \text{Min } x_1 + \frac{2}{x_1} \\ \text{s.a. } 1 \leq x_1 \leq 2 \end{array} \right\}$$

no convexo

convexo

Ej2:

$$\left. \begin{array}{l} \text{Min } x_1 + 2x_2 + 3x_3 \\ \text{s.a. } e^{x_1} (x_2 + x_3)^2 = 10 \\ 1 \leq x_1 \leq 2 \\ 0.5 \leq x_2 \leq 2 \\ 0.5 \leq x_3 \leq 2 \end{array} \right\} \xrightarrow{x_1 = \text{Ln}\left(\frac{10}{(x_2 + x_3)^2}\right)} \begin{array}{l} \text{Min } \text{Ln}\left(\frac{10}{(x_2 + x_3)^2}\right) + 2x_2 + 3x_3 \\ \text{s.a. } \text{Ln}\left(\frac{10}{(x_2 + x_3)^2}\right) \leq 2 \\ -\text{Ln}\left(\frac{10}{(x_2 + x_3)^2}\right) + 1 \leq 0 \\ 0.5 \leq x_2 \leq 2 \\ 0.5 \leq x_3 \leq 2 \end{array}$$

Aunque se reduce la dimensionalidad se añade una nueva restricción no lineal.

4. Tratar de evitar overflows, divisiones por cero, etc:

Ej1.

$$x + y - \ln(z) = 0 \longrightarrow \begin{array}{l} x + y - u = 0 \\ \exp(u) - z = 0 \end{array}$$

Ej2

$$v(xy - z^2)^{1/2} = 3 \longrightarrow \begin{array}{l} vu = 3 \\ u^2 - (xy - z^2) = 0 \\ u \geq 0 \end{array}$$

5. Explotar las propiedades globales cuando estas existan: Programa lineal, cuadrático, convexo, convexificar si es posible, etc.

6. Explotar la estructura del problema.

Ej.

$$\min z = Tx - 3y$$

$$s.a. \quad xT + y - yT^2 = 5$$

$$4x - 5yT + xT = 7$$

$$0 \leq T \leq 1$$

Se fija T y se resuelve un problema lineal. T se optimiza en un bucle exterior

7. Tratar de escalar todas las variables y las restricciones, esto facilita la convergencia y evita problemas de condicionamiento.
8. Tratar de proporcionar buenos valores iniciales a las variables.