13. Métodos iterativos de minimización con restricciones

El propósito de este capítulo es describir y usar en ejemplos, algunos métodos para resolver el problema general de programación no lineal siguiente:

$$(P) \quad \begin{array}{ll} \text{mín} & f(\vec{x}), \\ \text{sujeto a:} & g(\vec{x}) \leq 0, \\ & h(\vec{x}) = 0, \\ & \vec{x} \in X \subseteq R^n. \end{array}$$

Suponemos que todas las funciones tienen derivadas parciales continuas de orden 3.

Trabajaremos con MÉTODOS PRIMALES, que son aquellos que: El método de búsqueda de la solución óptimal del problema original lo hace directamente en la región factible.

Estos métodos presentan las siguientes ventajas:

a) cada punto generado por el procedimiento de búsqueda es factible

> Esto da la idea de que cuando finalicemos el procedimiento de búsqueda, el último punto hallado es factible y probablemente estará cerca de la solución óptima de (P).

b) si se genera una sucesión convergente, el punto límite de esta sucesión debe ser al menos un mínimo restringido local.

c) la mayoría de los métodos primales no dependen de la estructura especial del problema (convexidad), de ahí su aplicabilidad.

13.1. Método de la dirección factible

Consideramos el problema:

$$(P1) \quad \begin{array}{ll} \text{mín} & f(\vec{x}), \\ \text{sujeto a:} & \vec{a}_i^T \cdot \vec{x} \leq b_i \quad i = 1.....m \end{array}$$

en matrices

$$\begin{array}{ll} \text{mín} & f(\vec{x}), \\ \text{sujeto a:} & A\vec{x} \leq b. \end{array}$$

Inicialmente, para facilitar la comprensión del método, describiremos la búsqueda del mínimo a lo largo de una línea, posteriormente, la búsqueda del mínimo a lo largo de un segmento dentro del conjunto factible y finalmente el método de la dirección factible.

13.2. Mínimo a lo largo de una línea

El procedimiento para el problema (P1) es como sigue:

Paso1. Seleccionamos un punto base arbitrario $\vec{x}_o$.

Paso 2. Elegimos una dirección $\vec{d}$ y nos desplazamos por la línea que pasa por $\vec{x}_o$ y tiene la dirección $\vec{d}$, es decir, $\vec{x}_o + \lambda \vec{d}$ donde λ es un número real.

Paso 3. Resolvemos el programa

$$\min_{\lambda} f(\vec{x}_o + \lambda\vec{d}),$$

y encontramos el valor λ_1, de tal manera que el mínimo es el punto es

$$\vec{x}_1 = \vec{x}_o + \lambda_1\vec{d}$$

.

13.3. Mínimo a lo largo de un segmento dentro del conjunto factible

El procedimiento para el problema (P1) es como sigue:

Paso1. Seleccionamos un punto base arbitrario $\vec{x}_o$, pero, dentro del conjunto factible.

Paso 2. Elegimos una dirección $\vec{d}$ y nos desplazamos por la línea que pasa por $\vec{x}_o$ y tiene la dirección $\vec{d}$, es decir, $\vec{x}_o + \lambda \vec{d}$ donde λ es un número real positivo.

Paso 3 Buscamos el valor $\lambda_{\text{máx}}$ tal que, el segmento $\left[\vec{x}_o, \vec{x}_o + \lambda_{\text{máx}}\vec{d}\right]$ está contenido en el conjunto de las soluciones factibles. Para ello elegiremos el valor de $\lambda_{\text{máx}}$ de la siguiente forma

$$\lambda_{\text{máx}} = \begin{cases} \text{mín} \left\{ \dfrac{\hat{b}_i}{\hat{d}_i} \mid \hat{d}_i > 0 \right\} & \text{si } \hat{d} \not\leq 0, \\ \infty & \text{si } \hat{d} \leq 0, \end{cases} \quad \text{donde} \quad \begin{array}{c} \hat{b} = b - A\vec{x}_0, \\ \hat{d} = A\vec{d}. \end{array} \tag{1}$$

Paso 4. Calculamos

$$\begin{array}{cc} \text{mín} & f(\vec{x}_o + \lambda\vec{d}), \\ \text{sujeta a:} & 0 \leq \lambda \leq \lambda_{\text{máx}}. \end{array}$$

y encontramos el valor λ_1, de tal manera que el mínimo es el punto es $\vec{x}_1 = \vec{x}_o + \lambda_1 \vec{d}$.

□

Para poder explicar (1) nos basaremos en la forma matricial del programa (P1).

Suponemos ahora la matriz A descompuesta en dos matrices A_1 y A_2 y $\vec{b}$ descompuesto en $\vec{b}_1$ y $\vec{b}_2$ tales que $A_1\vec{x}_0 = \vec{b}_1$ y $A_2\vec{x}_0 < \vec{b}_2$. La información que nos proporciona la igualdad es redundante

Entonces $A_2\left(\vec{x}_o + \lambda\vec{d}\right) < \vec{b}_2$ para todo valor de $\lambda \geq 0$, pero, nosotros sólo tenemos que restringir el valor de λ .

Operando $A_2\vec{x}_o + \lambda A_2\vec{d} < \vec{b}_2$, de aquí $\lambda A_2\vec{d} < \vec{b}_2 - A_2\vec{x}_o$ resultando la expresión (1), teniendo en cuenta que en el problema la matriz A es la matriz A_2 de la explicación.

Observamos que, el tema importante es la elección de la dirección adecuada, por consiguiente, se impone una modificación del método con la idea siguiente: Nos moveremos por la región factible del problema (P1) de la forma $\vec{x}_{k+1=}\vec{x}_k + \lambda_k\ \vec{d}_k$ donde λ_k es un escalar no negativo que se elige para minimizar la función objetivo, con la condición de que el punto $\vec{x}_{k+1}$ y el segmento $[\vec{x}_k, \vec{x}_{k+1}]$ sean factibles.

13.4. Búsqueda de la dirección factible (Caso lineal)

Ahora nuestro objetivo es buscar una dirección $\vec{d}_k$ adecuada. Se obtiene resolviendo el problema lineal:

$$\begin{array}{ll} \text{mín} & \nabla f(\vec{x}_k)^T \cdot \vec{d}, \\ \text{sujeto a:} & \vec{a}_i^T \cdot \vec{d} \leq 0, \quad i \in I, \\ & -1 \leq\ d_i \leq 1 \quad \forall i = 1, \cdots, n. \end{array}$$

donde:

a) la función objetivo nos indica que el vector $\vec{d}_k$ se tiene que alinear lo más próximo posible al vector negativo del gradiente (recordar que es el descenso al mínimo).

b) la primera restricción asegura que los vectores $\vec{d}_k$ serán factibles para $\lambda_k > 0$ suficientemente pequeña, por ello se eligen aquellas restricciones que estén saturadas y el índice I es el conjunto de índices de las mismas.

c) la última restricción es una ecuación de normalización que nos asegura una solución acotada.

d) la justificación de la primera restricción es:

$$\vec{a}_i^T \cdot (\ \vec{x}_k +\ \lambda_k\ \vec{d}_k) = \vec{a}_i^T \cdot \vec{x}_k + \alpha \cdot \vec{a}_i^T \cdot\ \vec{d}_k \leq b_i \text{ para } \alpha \geq 0$$

si sólo tomamos las restricciones saturadas, se cumple $\vec{a}_i^T \cdot \vec{x}_k = b_i$ por consiguiente, de la expresión anterior obtendremos

$$\vec{a}_i^T \cdot\ \vec{d}_k \leq 0$$

El resultado de este problema es la **mejor dirección local posible** en la que continuar.

13.5. Algoritmo del método de la dirección factible

Paso 1. Búsqueda de la dirección adecuada mediante el problema lineal siguiente:

$$\begin{array}{ll} \text{mín} & \nabla f(\vec{x}_k)^T \cdot \vec{d}, \\ \text{sujeto a:} & \vec{a}_i^T \cdot \vec{d} \leq 0, \quad i \in I \text{ (restricciones activas)}, \\ & -1 \leq \ d_i \leq 1 \quad \forall i. \end{array}$$

Paso 2. Búsqueda del mínimo a lo largo de la dirección encontrada en el Paso 1 mediante el programa siguiente:

$$\begin{array}{cc} \text{mín} & f(\vec{x}_o + \ \lambda \vec{d}), \\ \text{sujeta a:} & 0 \leq \lambda \leq \lambda_{\text{máx}}. \end{array} \quad \text{donde}$$

$$\lambda_{\text{máx}} \ = \begin{cases} \text{mín} \left\{ \dfrac{\hat{b}_i}{\hat{d}_i} \mid \hat{d}_i > 0 \right\} & \text{si } \hat{d} \nleq 0, \\ \infty & \text{si } \hat{d} \leq 0, \end{cases} \quad \text{donde} \quad \begin{array}{c} \hat{b} = b - A\vec{x}_0, \\ \hat{d} = A\vec{d}. \end{array}$$

Paso 3. Hacemos $\vec{x}_1 \ = \ \vec{x}_0 + \ \ \lambda_0 \vec{d}_0$. Si el resultado es el adecuado, el procedimiento se para, en caso contrario encontramos el nuevo conjunto de índices correspondientes a las restricciones que satura la nueva solución $\vec{x}_1$ y volvemos al Paso 1 con el nuevo punto.

13.6. Inconvenientes

Entre los posibles inconvenientes que se encuentran al método podemos citar:

En los casos generales pueden no existir direcciones factibles, como es el caso de restricciones de igualdad no lineales. En estos casos sería necesario flexibilizar el requerimiento de factibilidad permitiendo, por ejemplo, que los puntos se desvíen ligeramente de la superficie de restricción, o bien, moviéndose por curvas.

Pueden no ser convergentes, es decir, sujetos a zigzagueo donde la sucesión de puntos converge a un punto que ni siquiera es un mínimo local restringido.

13.7. Ejemplo

Usando el procedimiento anterior resuelve el programa matemático siguiente:

$$\begin{array}{ll} \text{mín} & 2x_1^2 - 4x_1 + 2x_2^2 - 6x_2 - 2x_1 x_2, \\ \text{sujeto a:} & x_1 + x_2 \leq 2, \\ & x_1 + 5x_2 \leq 5, \\ & -x_1 \leq 0, \\ & -x_2 \leq 0. \end{array}$$

Solución:

Elegimos un punto inicial factible $\vec{x}^0 = [0, 0]$

Calculamos el gradiente de la función objetivo

$$\nabla f\left(\vec{x}\right) = [4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6]$$

Valoramos el gradiente en $\vec{x}^0$ resultando $[-4, -6]$

Para la búsqueda de la dirección factible $\vec{d} = [d_1, d_2]$ resolveremos el siguiente programa matemático:

$$\left.\begin{array}{lr}
\text{mín} & [-4,-6]\cdot\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -4d_1 - 6d_2 \\
\text{sujeta a:} & -d_1 \le 0 \\
& -d_2 \le 0 \\
& -1 \le d_1 \le 1 \\
& -1 \le d_2 \le 1
\end{array}\right\} \Rightarrow \text{la solución es}$$

$$\vec{d} = [d_1, d_2] = [1, 1]$$

por consiguiente la recta a considerar será

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}$$

sustituido en la función

$$f(\lambda, 3\lambda) = 2\lambda^2 - 10\lambda$$

resolviendo

$$\begin{array}{lc}
\text{mín} & 2\lambda^2 - 10\lambda \\
\text{sujeta a;} & 0 \le \lambda \le \lambda_{\text{máx}}
\end{array}$$

donde $\lambda_{\text{máx}} = \text{mín}\left\{\frac{2}{2}, \frac{5}{6}\right\}$ encontramos que $\lambda = \frac{5}{6}$ que nos proporciona el punto $\vec{x}^1 = \left[\frac{5}{6}, \frac{5}{6}\right]$ que pertenece al conjunto factible y nos da la primera aproximación a la solución del problema.

Decidimos si esta solución es la que nos conviene o no, en caso de decidir que no, continuamos con la iteración.

Tomamos $\vec{x}^1 = \left[\frac{5}{6}, \frac{5}{6}\right]$ como punto inicial de nuestro problema

Valoramos el gradiente en $\vec{x}^1$ resultando $\nabla f\left(\frac{5}{6}, \frac{5}{6}\right) = \left[-\frac{7}{3}, -\frac{13}{3}\right]$

Calculamos la dirección factible:

$$\left.\begin{array}{lr}
\text{mín} & \left[-\frac{7}{3}, -\frac{13}{3}\right]\cdot\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = -\frac{7}{3}d_1 - \frac{13}{3}d_2 \\
\text{sujeta a:} & d_1 + 5d_2 \le 0 \\
& -1 \le d_1 \le 1 \\
& -1 \le d_2 \le 1
\end{array}\right\} \Rightarrow \text{la solución es}$$

$$\vec{d} = [d_1, d_2] = \left[1, -\frac{1}{5}\right]$$

por consiguiente la recta a considerar será

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{5}{6} \\ \frac{5}{6} \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -\frac{1}{5} \end{bmatrix} = \begin{bmatrix} \frac{5}{6} + \lambda \\ \frac{5}{6} - \frac{1}{5}\lambda \end{bmatrix}$$

sustituido en la función

$$f\left(\frac{5}{6} + \lambda, \frac{5}{6} - \frac{1}{5}\lambda\right) = \frac{62}{25}\lambda^2 - \frac{22}{15}\lambda - \frac{125}{8}$$

resolviendo

$$\begin{array}{lc}
\text{mín} & \frac{62}{25}\lambda^2 - \frac{22}{15}\lambda - \frac{125}{8} \\
\text{sujeta a:} & 0 \le \lambda \le \lambda_{\text{máx}}
\end{array}$$

donde

$$\lambda_{\text{máx}} = \text{mín}\left\{\frac{1/3}{4/5}, \frac{5/6}{1/6}\right\}$$

encontramos que $\lambda = \frac{55}{186}$ que nos proporciona el punto $\vec{x}^2 = \left[\frac{35}{31}, \frac{24}{31}\right]$ que pertenece al conjunto factible y nos da la segunda aproximación a la solución del problema.

13.8. Caso de restricciones no lineales

Tenemos el problema

$$\begin{array}{ll} \text{mín} & f(\vec{x}) \\ \text{sujeto a:} & g_i(\vec{x}) \leq 0 \quad i = 1, ...m \end{array}$$

Consideramos $\vec{x}_k$ solución factible del problema, tenemos que buscar una dirección $\vec{y}$ tal que, al desplazarnos desde el punto $\vec{x}_k$ en la dirección $\vec{d}$ encontrada, un determinado paso, nos encontremos con otra solución factible y una disminución del valor de la función en ese nuevo punto con respecto al anterior.

Para el cálculo de la dirección $\vec{y}$ resolveremos el siguiente problema:

$$\begin{array}{ll} \text{mín} & \nabla f(\vec{x}_k)^T \vec{d}, \\ \text{sujeto a:} & g_i(\vec{x}_k) + \nabla g_i(\vec{x}_k)^T \vec{d} \leq 0 \quad i = 1, ..., m, \\ & -1 \leq d_j \leq 1 \quad j = 1, ..., n. \end{array}$$

si $\vec{d}$ es el óptimo de este programa. Buscaremos un valor α tal que $\vec{x}_k + \alpha\vec{d}$ sea factible, resolviendo para ello el siguiente programa

$$\alpha_k = \text{máx}\{\alpha \mid g_i(\vec{x}_k + \alpha\vec{d}) \leq 0 \ i = 1, ..., m\}$$

$\vec{x}_k + \alpha\vec{d}$ será factible para $0 \leq \alpha \leq \alpha_k$

$$\begin{array}{ll} \text{mín} & f(\vec{x}_k + \alpha\vec{d}), \\ \text{sujeto a:} & 0 < \alpha < \alpha_k, \end{array}$$

aquí obtendremos $\vec{x}_{k+1} = \vec{x}_k + \alpha\vec{d}$. En caso de ser una solución aceptable paramos el proceso, en caso contrario, volvemos a empezar el proceso con la solución hallada como valor origen.

13.9. Ejemplo

$$\begin{array}{ll} \text{mín} & x_1 + 2x_2 + x_3, \\ \text{sujeto a:} & 3x_1^2 + 2x_2^2 + \frac{1}{3}x_3^2 - 6x_1 - 4x_2 - 2x_3 + 7{,}25 \leq 0. \end{array}$$

Solución:

Elegimos $x_0 = (1, 1, 3)$ un punto factible que cumple la restricción.

Tenemos que $\nabla f(\vec{x}) = (1, 2, 1)$ y

$\nabla g_i(\vec{x}_k) = (6x-6, 4y-4, \frac{2z}{3}-2)_{(1,1,3)} = (0,0,0)$ $g_i(\vec{x}_k) + \nabla g_i(\vec{x}_k)^T \vec{d} =$
$-0{,}75 + (0,0,0) \cdot \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} \leq 0$

Planteamos el problema de encontrar la dirección

$$\left.\begin{array}{ll} \text{mín} & d_1 + 2d_2 + d_3 \\ \text{sujeto a:} & -0{,}75 \leq 0 \\ & -1 \leq d_1, d_2, d_3 \leq 1 \end{array}\right\} \rightarrow (d_1, d_2, d_3) = (-1, -1, -1).$$

Planteamos encontrar el valor de α máximo para que el recorrido sea factible

$$\alpha_k = \text{máx}\{\alpha \mid g_i(1-\alpha, 1-\alpha, 3-\alpha)\}, \frac{64\alpha^2 - 9}{12} \leq 0 \implies \alpha = 0{,}375.$$

Planteamos el problemas de encontrar el mínimo en el recorrido

$$\left.\begin{array}{ll} \text{mín} & f(1 + \alpha \cdot (-1), 1 + \alpha \cdot (-1), 3 + \alpha \cdot (-1) \\ \text{sujeto a:} & 0 < \alpha < 0{,}375 \end{array}\right\} \Rightarrow$$

$$\left.\begin{array}{ll} \text{mín} & 6 - 4x \\ \text{sujeto a:} & 0 < \alpha < 0{,}375 \end{array}\right\} \Rightarrow \alpha = 0{,}375,$$

el nuevo punto será: $(x_1, x_2, x_3) = (0{,}625, 0{,}625, 2{,}625)$, reiterando el proceso más de 10 veces el resultado es:

$$(x_1, x_2, x_3) = (0{,}875, 0{,}625, 1{,}875).$$

14. Método del conjunto activo

En el desarrollo del método anterior hemos observado que distinguíamos entre restricciones activas y restricciones inactivas, estas últimas son ignoradas. Basándonos en este hecho proponemos un método en el que sólo trataremos de restricciones activas. El problema es el siguiente:

$$\begin{array}{ll} \text{mín} & f(\vec{x}) \\ \text{sujeto a:} & g(\vec{x}) \leq 0 \end{array}$$

Las condiciones de K-T para este problema son:

$$\left.\begin{array}{l} \nabla f(\vec{x}) + \lambda^T \nabla g(\vec{x}) = 0 \\ g(\vec{x}) \leq 0 \\ \lambda^T g(\vec{x}) = 0 \\ \lambda \geq 0 \end{array}\right\} \begin{array}{l} \text{Expresadas} \\ \text{en términos} \\ \text{de restricciones} \\ \text{activas} \end{array} \left\{\begin{array}{l} \nabla f(\vec{x}) + \sum\limits_{i \in A} \lambda_i^T \nabla g_i(\vec{x}) = 0 \\ g_i(\vec{x}) = 0, \ i \in A \\ g_i(\vec{x}) \leq 0, \ i \notin A \\ \lambda i \geq 0, \ i \in A \\ \lambda_i = 0, \ i \notin A \end{array}\right.$$

A es el conjunto de índices de las restricciones activas. Es obvio que, de conocerse las restricciones activas, el problema se podría sustituir por otro más sencillo en el que todas las restricciones son de igualdad. Nuestra idea es definir en cada momento el conjunto que índices que se considera activo, si los multiplicadores de Lagrange correspondientes a ese problema son todos no negativos, la solución se considerará correcta. Si llamamos *conjunto de trabajo* a las restricciones activas en ese momento en el punto que nosotros estamos trabajando, el algoritmo se moverá por la superficie definida por el conjunto de trabajo hacia un punto mejorado. En este momento se podrá cambiar el conjunto de trabajo. Repetiremos este proceso hasta llegar a la solución que consideremos adecuada.

14.1. Algoritmo

Supongamos A el conjunto de índices del conjunto de trabajo. Nuestro problema queda:

$$\begin{array}{ll} \text{mín} & f(\vec{x}) \\ \text{sujeto a:} & g_i(\vec{x}) = 0,\ i \in A \end{array}$$

Resolvemos el problema produciendo un punto $\vec{x}_0$ que satisface $g_i(\vec{x}) < 0,\ i \notin A$. Este punto satisface las condiciones necesarias

$$\nabla f(\vec{x}_0) + \sum_{i \in A} \lambda_i^T \nabla g_i(\vec{x}_0) = 0$$

Si $\lambda_i \geq 0\ \forall i \in A$ el punto $\vec{x}_0$ es una solución local del problema original.

Si $\lambda_i < 0$ para algún $i \in A$ entonces este objetivo se puede eliminar del conjunto (esto está en relación con el análisis de sensibilidad de los multiplicadores de Lagrange, que nos indica que una pequeña variación en el valor de la restricción nos produce una variación en el valor de la función objetivo). El multiplicador de Lagrange nos sirve como indicador si las restricción de trabajo debe eliminarse o no.

Debemos tener en cuenta las otras restricciones, en el sentido de que, cuando calculamos la solución del problema nuevo no se incumplan, en caso de incumplimiento añadiremos dicha restricción al problema.

14.2. Ejemplo

15. Método de la proyección del gradiente

Podemos considerar este método una extensión para restricciones del método de máximo descenso en el problema de optimización sin restricciones. Uno de los objetivos que persigue es abreviar el método de programación lineal en su complicación de cálculo.

Consideramos dos casos:

a) Con restricciones lineales.

b) Con restricciones no lineales.

15.1. Con restricciones lineales

Consideramos el problema

$$\begin{array}{lll} \text{mín} & f(\vec{x}), & \\ \text{sujeto a:} & \vec{a}_i^T \vec{x} \leq b_i & i \in I_1, \\ & \vec{a}_i^T \vec{x} \leq b_i & i \in I_2, \end{array}$$

I_1 y I_2 son el conjunto de índices de las restricciones saturadas y no saturadas respectivamente. Elegimos un punto factible $\vec{x}_0$ como primera solución. Sabemos que, para este punto, existirán un cierto número de restricciones activas I , entre las que están incluidas las I_1. Nosotros elegiremos solamente las restricciones del conjunto de índices I. El problema queda reducido a:

$$\begin{array}{ll} \text{mín} & f(\vec{x}), \\ \text{s. a:} & \vec{a}_i^T \vec{x} = b_i \; i \in I, \end{array} \quad \text{en forma matricial} \quad \begin{array}{ll} \text{mín} & f(\vec{x}), \\ \text{s. a:} & A\vec{x} = \vec{b} \; i \in I, \end{array}$$

a partir del punto elegido buscamos un vector dirección factible $\vec{d}$, que en principio tiene que satisfacer la condición $\nabla f\,(\vec{x}_0) \cdot \vec{d} \leq 0$ de modo que el movimiento en la dirección del vector produzca una disminución de la función. Suponemos la regularidad de las restricciones. El requerimiento de factibilidad nos dice que el desplazamiento lo tendremos que hacer a través del subespacio tangente en $\vec{x}_0$ definido por las restricciones con las que trabajamos. Si definimos este subespacio como $T = \{\vec{y} \;\; | \;\; \vec{a}_i^T \vec{y} = 0, \;\; i \in I\}$ en nuestro caso la condición se escribe $A \cdot \vec{d} = 0$ el vector director que se utilizará es la proyección de $\vec{d}$ sobre el subespacio T. Definimos el subespacio O como el subespacio ortogonal a T que está definido por las filas de la matriz A. Al ser dos espacios complementarios, cualquier vector del espacio se puede poner como suma de un vector de T y otro vector de O . Concretando esto para nuestro método, $-\nabla f(\vec{x}_0)$ se puede expresar de la forma

$$\nabla f(\vec{x}_0) = \vec{d} + A^T \cdot \lambda,$$

despejamos λ exigiendo que $A \cdot \vec{d} = 0$

$$-A \cdot \nabla f(\vec{x}_0) = A \cdot \vec{d} + A \cdot A^T \cdot \lambda \;\; \lambda = -(A \cdot A^T)^{-1} \cdot A \cdot \nabla f(\vec{x}_0),$$

luego $\vec{d}$ lo podemos encontrar

$$\begin{aligned} \vec{d} &= -\nabla f(\vec{x}_0) - A^T \lambda \\ &= -\nabla f(\vec{x}_0) + A^T \cdot (A \cdot A^T)^{-1} \cdot A \cdot \nabla f(\vec{x}_0) \\ &= -[I - A^T \cdot (A \cdot A^T)^{-1} \cdot A] \cdot \nabla f(\vec{x}_0) \\ &= -P \cdot \nabla f(\vec{x}_0), \end{aligned}$$

donde P es la matriz de proyección correspondiente al subespacio T.

Podemos señalar dos casos:

Caso $\vec{d} \neq 0$ Comprobamos que es una dirección de descenso de la función: Sabemos que $\vec{d}$ es ortogonal a $\vec{d} + \nabla f(\vec{x}_0)$ luego

$$\vec{d}^T \cdot \left(\vec{d} + \nabla f(\vec{x}_0)\right) = \vec{d} \cdot \vec{d} + \nabla f(\vec{x}_0) \cdot \vec{d} = -\left|\vec{d}\right|^2 .$$

Luego cumple una de las condiciones impuestas. A continuación, pasamos a la selección de la medida del paso α ya que el nuevo punto $\vec{x}_1 = \vec{x}_0 + \alpha\vec{d}$ debe de ser factible y, lo hacemos, resolviendo los problemas

$$\begin{array}{l} \text{máx } \{\alpha \mid \vec{x}_0 + \alpha\vec{d} \text{ es factible } \} = \alpha_0. \\ \text{mín } f(\vec{x}_0 + \alpha\vec{d}) \text{ tal que } 0 \leq \alpha \leq \alpha_0. \end{array}$$

el resultado del último problema será α_1 que dará origen al punto x_1 y a partir de aquí volvemos a comenzar el proceso teniendo en cuenta que este punto puede dar alguna restricción nueva activa, en cuyo caso la añadiremos a las restricciones que teníamos al principio del problema.

Caso $\vec{d} = 0$ Este caso nos dice que la proyección del gradiente sobre el plano tangente es cero. En este caso tenemos que hallar

$$\lambda = -(A \cdot A^T)^{-1} \cdot A \cdot \nabla f(\vec{x}_0).$$

Con las casuísticas siguientes:
a) Si $\lambda_i \geq 0$ $i \in I$ correspondientes a las desigualdades activas, se para el proceso ya que x_0 satisface las condiciones de KT.
b) En caso contrario, se elimina de la matriz A la fila correspondiente al valor de λ_i negativo menor, es decir se elimina la restricción correspondiente al conjunto de restricciones activas y se vuelve al principio del proceso.

15.2. Ejemplo

Sea el problema

$$\begin{array}{ll} \text{mín} & x_1^2 + x_2^2 + 2x_3^2 x_4^2, \\ \text{Sujeto a:} & x_1 + x_3 - x_4 = 2, \\ & x_2 + x_3 + x_4 = 3, \\ & x_1,\ x_2,\ x_3.\ x_4\ \geq 0. \end{array}$$

Comenzamos en un punto factible que consideraremos solución aproximada $\vec{x}_0 = (2, 3, 0, 0)$ y observamos que nuestra solución satura las dos restricciones, por tanto

$$A = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad \nabla f(\vec{x}_0) = \begin{pmatrix} 2x_1 \\ 2x_2 \\ 4x_3 x_4^2 \\ 4x_3^2 x_4 \end{pmatrix}_{(2,3,0,0)} = \begin{pmatrix} 4 \\ 6 \\ 0 \\ 0 \end{pmatrix}.$$

Como sabemos que $\lambda = -(A \cdot A^T)^{-1} \cdot A \cdot \nabla f(\vec{x}_0)$ obtenemos

$$\lambda = -\left(\begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{pmatrix}\right)^{-1} \cdot \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 6 \\ 0 \\ 0 \end{pmatrix},$$

que resolviendo queda $\lambda = (4, 6, -10, -2)$. Calculamos $\vec{d}$,

$$\vec{d} = -\begin{pmatrix} 4 \\ 6 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 6 \\ -10 \\ -2 \end{pmatrix} = \begin{pmatrix} -8 \\ -12 \\ 0 \\ 0 \end{pmatrix}.$$

Consideramos ahora

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -8 \\ -12 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2-8\alpha \\ 3-12\alpha \\ 0 \\ 0 \end{pmatrix},$$

sustituyendo en la función objetivo resolvemos el programa

$$\begin{aligned} &\min\, f(\vec{x}) = (2-8\alpha)^2 + (3-12\alpha)^2, \\ &\text{sujeta a: } 0 \le \alpha \le \alpha_{máx}. \end{aligned}$$

El calculo de $\alpha_{máx}$ se hace cogiendo las restricciones no saturadas y sustituyendo los valores de $\vec{x}_o$, en nuestro caso $x_1 \ge 0 \Rightarrow 2 - 8\alpha \ge 0 \Rightarrow \alpha \le \frac{1}{4}$ $x_2 \ge 0 \Rightarrow 3 - 12\alpha \ge 0 \Rightarrow \alpha \le \frac{1}{4}$ $\alpha_{máx} = \text{mín}\left\{\frac{1}{4}, \frac{1}{4}\right\}$.

Derivando con respecto a α resulta $f'(\vec{x}) = -16(2-8\alpha) - 24(3-12\alpha)$, igualando a cero y resolviendo la ecuación $\alpha = \frac{13}{52} = \frac{1}{4} = 0{,}25$ por tanto el nuevo punto $\vec{x}_1$ es:

$$\vec{x}_1 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

y en este caso se vuelve a empezar dado que no hemos alcanzado un solución óptima.

15.3. Con restricciones no lineales

16. Métodos de penalización

Los métodos de penalización transforman un problema de optimización con restricciones en otro con una sola restricción o en una sucesión de problemas sin restricciones. Las restricciones son introducidas en la función objetivo vía un parámetro de penalización de una forma que penaliza cualquier violación de las restricciones.

Consideraremos dos tipos de problemas:

a)

$$\begin{aligned} &\text{mín} && f(\vec{x}), \\ &\text{s. a:} && h(\vec{x}) = 0, \end{aligned}$$

este problema es reemplazado por el problema sin restricciones siguiente:

$$\begin{aligned} &\text{mín} && f(\vec{x}) + \mu h^2(\vec{x}), \\ &\text{Sujeto a:} && \vec{x} \in R^n, \end{aligned}$$

intuitivamente observamos que en la solución óptima es preciso que $h^2(\vec{x})$ sea cero, porque en caso contrario incurriremos en una gran penalización.

b) El problema del tipo

$$\begin{array}{ll} \text{mín} & f(\vec{x}), \\ \text{Sujeto a:} & g(\vec{x}) \leq 0, \end{array}$$

es claro que la forma anterior no es la apropiada puesto que la penalización se incurre cuando $g(\vec{x}) < 0$ o $g(\vec{x}) > 0$. Necesitamos que la penalización sólo se cumpla en los puntos no factibles, esto es, cuando $g(\vec{x}) > 0$.Por consiguiente se propone el problema

$$\begin{array}{ll} \text{mín} & f(\vec{x}) + \mu \, \text{máx}\{0,\ g(\vec{x})\}, \\ \text{Sujeto a:} & \vec{x} \in R^n. \end{array}$$

En este caso, si $g(\vec{x}) \leq 0$ el máximo de $\{0,\ g(\vec{x})\}$ es 0, y no se incurre en penalización y si $g(\vec{x}) > 0$ el máximo de $\{0,\ g(\vec{x})\}$ es mayor que cero y el término de penalización $\mu \, g(\vec{x})$ se realiza.

En el caso de que la diferenciabilidad sea deseable podemos considerar el término del tipo $\mu \left[\text{máx}\{0,\ g(\vec{x})\}\right]^2$.

Ejemplo 16.1 *considera el siguiente problema*

$$\begin{array}{ll} \text{mín} & x, \\ \textit{Sujeto a:} & -x + 2 \leq 0, \end{array}$$

Sea $\alpha(x) = \left[\text{máx}\{0,\ g(\vec{x})\}\right]^2$, *entonces*

$$\alpha(x) = \begin{cases} 0 & \textit{si } x \geq 2, \\ (-x+2)^2 & \textit{si } x < 2. \end{cases}$$

El problema queda reducido a

$$\begin{array}{ll} \text{mín} & x + \mu\alpha(x), \\ \textit{Sujeto a:} & x \in R, \end{array}$$

derivando e igualando a cero la derivada obtenemos $1 - 2\mu(-x+2) = 0 \Rightarrow x = 2 - \dfrac{1}{2\mu}$, *que se aproxima al punto mínimo* $x = 2$ *cuando* $\lim\limits_{\mu \to \infty} \left(2 - \dfrac{1}{2\mu}\right) = 2$.

Ejemplo 16.2 *Sea el problema*

$$\begin{array}{ll} \text{mín} & x_1^2 + x_2^2, \\ \textit{Sujeto a:} & x_1 + x_2 - 1 = 0. \end{array}$$

El programa transformado es

$$\begin{array}{ll} \text{mín} & x_1^2 + x_2^2 + \mu(x_1 + x_2 - 1), \\ \textit{Sujeto a:} & (x_1, x_2) \in R^2. \end{array}$$

Derivando e igualando a cero las derivadas, obtenemos

$$\left.\begin{array}{l} 2x_1 + 2\mu\,(x_1 + x_2 - 1) = 0 \\ 2x_2 + 2\mu\,(x_1 + x_2 - 1) = 0 \end{array}\right\} \Rightarrow x_1 = x_2 = \frac{\mu}{2\,(\mu+1)}.$$

Tomando límites, $\lim_{\mu\to\infty} \frac{\mu}{2\,(\mu+1)} = \frac{1}{2}$, *por tanto la solución es:*

$$(x_1, x_2) = \left(\frac{1}{2}, \frac{1}{2}\right).$$

16.1. Generalización del método

Dado el programa

$$\begin{array}{lll} \text{mín} & f\,(\vec{x})\,, & \\ \text{Sujeto a:} & g_i\,(\vec{x}) \leq 0, & i = 1, \ldots, m, \\ & h_j\,(\vec{x}) = 0, & j = 1, \ldots, l. \end{array}$$

la función de penalización está definida por

$$\alpha\,(x) = \sum_{i=1}^{m} \Phi\,[g_i\,(\vec{x})] + \sum_{j=1}^{l} \Psi\,[h_j\,(\vec{x})]$$

donde Φ y Ψ son funciones continuas que satisfacen las siguientes condiciones

$$\Phi\,(y) = \begin{cases} 0 & \text{si } y \leq 0, \\ > 0 & \text{si } y > 0. \end{cases}$$
$$\Psi\,(y) = \begin{cases} 0 & \text{si } y \leq 0, \\ > 0 & \text{si } y > 0. \end{cases}$$

La forma típica de Φ y Ψ es

$\Phi\,(y) = [\text{máx}\,\{0, y\}]^p$ con p un entero positivo.

$\Psi\,(y) = |y|^p$con p un entero positivo.

Dando origen a la forma usual de la función de penalización

$$\alpha\,(x) = \sum_{i=1}^{m} [\text{máx}\,\{0, g_i\,(\vec{x})\}]^p + \sum_{j=1}^{l} |h_j\,(\vec{x})|^p$$

y el programa queda

$$\begin{array}{ll} \text{mín} & f\,(\vec{x}) + \mu \left(\sum_{i=1}^{m} [\text{máx}\,\{0, g_i\,(\vec{x})\}]^p + \sum_{j=1}^{l} |h_j\,(\vec{x})|^p \right), \\ \text{Sujeto a:} & \vec{x} \in R^n. \end{array}$$