

# THI, Optimización

C. Palencia (palencia.math@gmail.com)  
ETSI Telecomunicación, UVA

Dada una función objetivo

$$f : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$$

definida en un conjunto abierto  $\Omega \subset \mathbb{R}^d$ , tratamos de resolver el problema de minimización

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \Omega} f(\mathbf{x}).$$

Supondremos que  $f$  es regular (admite derivadas continuas, al menos hasta el segundo orden). Ya sabemos que los puntos extremos, globales o locales, son puntos críticos de  $f$ , es decir son solución del sistema de ecuaciones en  $\Omega$

$$\nabla f(\mathbf{x}) = 0.$$

Asímismo, son frecuentes funciones objetivo para las cuales suceden alguna de las siguientes situaciones: (i) carecen de extremos, (ii) tienen extremos locales pero no globales, (iii) hay puntos críticos que son extremos.

La teoría rigurosa se apoya en la hipótesis de que realmente  $f$  alcanza un mínimo, al menos local. Vamos a introducir diversos métodos iterativos, muchas veces sin saber a priori si van a converger. Cuando realmente converjan, podemos intentar justificar a posteriori que el punto alcanzado es está realmente cerca de un mínimo local. Para ello hemos de estudiar la matriz hessiana, pues si ésta resulta ser definida positiva y si el gradiente en el punto límite es despreciable, podemos asegurar que hay un mínimo local muy próximo.

Un primer enfoque consiste simplemente en resolver el sistema de ecuaciones

$$\nabla f(\mathbf{x}) = 0,$$

por ejemplo vía Newton, y después estudiar el hessiano. Notemos que la matriz jacobiana del gradiente

$$D\nabla f(\mathbf{x}) = D \begin{bmatrix} \partial f(\mathbf{x})/\partial x_1 \\ \partial f(\mathbf{x})/\partial x_2 \\ \vdots \\ \partial f(\mathbf{x})/\partial x_n \end{bmatrix}$$

no es otra cosa que la matriz hessiana de  $f$ , que en lo sucesivo pasamos a denotar por  $Hf(\mathbf{x})$ . Dado  $\mathbf{x}_0 \in \Omega$ , intentamos pues avanzar en la iteración

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{c}_n, \quad \text{con} \quad Hf(\mathbf{x}_n)\mathbf{c}_n = f(\mathbf{x}_n).$$

El inconveniente principal es la evaluación de  $Hf(\mathbf{x}_n)$  resulta muy costosa incluso para valores moderados de  $d$ . Se podría implementar, alternativamente, la versión con matriz jacobiana aproximada. Pero en todo caso es preciso, en la práctica, orientar el avance de los puntos  $\mathbf{x}_n$ , pues no tendremos convergencia hasta que estemos en las proximidades del punto límite y, en general, no hay criterio para seleccionar  $\mathbf{x}_0$  adecuadamente.

Los algoritmos que vamos a presentar se inspiran no obstante en el método de Newton, pero introduciendo modificaciones para garantizar que  $f(\mathbf{x}_n)$  sea decreciente, lo que permite orientar la evolución de los puntos en la forma deseada.

Fijada la función objetivo  $f : \Omega \subset \mathbb{R}^d$ , el campo de sus gradientes se denotará por  $\mathbf{g}$ :

$$\mathbf{g}(\mathbf{x}) = \nabla f(x), \quad \mathbf{x} \in \Omega.$$

Dado un punto  $\mathbf{x} \in \Omega$ , buscamos direcciones  $\mathbf{p} \in \mathbb{R}^n$  tales que la función objetivo  $f$  decrezca al movernos a lo largo de la recta  $\mathbf{x} + \lambda \mathbf{p}$ , al menos durante cierto intervalo inicial  $[0, \delta]$ , con  $\delta > 0$ . Expresando la derivada direccional  $D_{\mathbf{p}}f(\mathbf{x})$  en la forma

$$D_{\mathbf{p}}f(\mathbf{x}) = \frac{d}{d\lambda} f(\mathbf{x} + \lambda \mathbf{p})|_{\lambda=0} = \langle \mathbf{g}(\mathbf{x}), \mathbf{p} \rangle,$$

vemos que una condición suficiente para nuestro propósito es que

$$\langle \mathbf{g}(\mathbf{x}), \mathbf{p} \rangle < 0.$$

Llamaremos dirección de descenso (para la función objetivo  $f$  en un punto  $\mathbf{x} \in \Omega$ ) a todo vector  $\mathbf{p} \in \mathbb{R}^d$  para el que se cumpla la condición anterior. La longitud de una dirección de descenso  $\mathbf{p}$  no es realmente importante, si bien afectará a la longitud del intervalo  $[0, \delta]$  sobre el que  $f(\mathbf{x} + \lambda \mathbf{p})$  es decreciente. Si  $\mathbf{x} \in \Omega$  no es un punto crítico de  $f$ , entonces  $\mathbf{p} = -\mathbf{g}(\mathbf{x})$  es siempre una dirección de descenso, denominada la dirección de máximo descenso. En efecto: (a) Por una parte  $\langle \mathbf{g}(\mathbf{x}), -\mathbf{g}(\mathbf{x}) \rangle = -\|\mathbf{g}(\mathbf{x})\|^2 < 0$ , de forma que  $\mathbf{p} = -\mathbf{g}(\mathbf{x})$  es ciertamente una dirección de descenso. (b) Para cualquier otra dirección  $\mathbf{p} \in \mathbb{R}^d$  de la misma longitud que  $\mathbf{g}(\mathbf{x})$ , tenemos

$$\langle \mathbf{g}(\mathbf{x}), -\mathbf{g}(\mathbf{x}) \rangle = -\|\mathbf{g}(\mathbf{x})\|^2 = -\|\mathbf{g}(\mathbf{x})\| \|\mathbf{p}\| \leq \langle \mathbf{g}(\mathbf{x}), -\mathbf{p} \rangle,$$

y así la menor derivada posible de  $f(\mathbf{x} + \lambda \mathbf{p})$ , en  $\lambda = 0$ , se consigue precisamente con  $\mathbf{p} = -\mathbf{g}(\mathbf{x})$ .

Observemos que si  $A \in M_{d,d}$  es una matriz definida positiva y si  $\mathbf{x}$  no es un punto crítico, entonces  $\mathbf{p} = -A\mathbf{g}(\mathbf{x})$  es siempre una dirección de descenso, pues

$$\langle \mathbf{g}(\mathbf{x}), \mathbf{p} \rangle = -\langle \mathbf{g}(\mathbf{x}), A\mathbf{g}(\mathbf{x}) \rangle < 0,$$

salvo si  $\mathbf{g}(\mathbf{x}) = 0$ , situación excluida.

Los métodos de descenso, que son los que vamos a estudiar, adoptan siempre la misma estructura: se parte de un vector inicial  $\mathbf{x}_0 \in \Omega$ . Alcanzado un punto  $\mathbf{x}_n \in \Omega$ , si éste es un punto crítico de ( $\mathbf{g}(\mathbf{x}) = 0$ ) terminamos el algoritmo. Caso contrario, tomamos una dirección de descenso  $\mathbf{p}_n$  de  $f$  en  $\mathbf{x}_n$  de la forma

$$\mathbf{p}_n = -A_n \mathbf{g}(\mathbf{x}_n),$$

donde  $A_n \in M_{d,d}$  es una matriz definida positiva, y definimos

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \lambda_n \mathbf{p}_n, \quad \text{con } \lambda_n > 0.$$



Los distintos métodos han de explicitar la elección de  $A_n$  y de  $\lambda_n > 0$ . En todos ellos se pretende que cuando  $g(\mathbf{x}_n) \neq 0$  (si  $g(\mathbf{x}_n) = 0$  hemos terminado la búsqueda)

$$f(\mathbf{x}_{n+1}) < f(\mathbf{x}_n).$$

Ejemplos básicos:

- 1)  $A_n = I$ ,  $\lambda_n = 1$ , método del máximo descenso.
- 2)  $A_n = Hf(\mathbf{x}_n)^{-1}$ ,  $\lambda_n = 1$ , método de Newton (observemos que esta matriz es realmente definida positiva en las proximidades de un mínimo para el que se cumpla la condición suficiente).

Incluso con la misma elección de  $A_n$  podemos implementar distintas versiones, según tomemos  $\lambda_n$ . En algunos casos (por ejemplo para los funcionales cuadráticos) se puede llegar a tomar

$$\lambda_n = \arg \min_{\lambda > 0} f(\mathbf{x}_n + \lambda \mathbf{p}_n),$$

o una aproximación, y hablamos entonces de método de descenso con búsqueda exacta a lo largo de las direcciones.

Nos vamos a decantar por el algoritmo de Armijo, para la selección de  $\lambda_n > 0$ . Hay que fijar tres parámetros  $\theta \in (0, 1)$ ,  $\alpha \in (0, 1)$  y  $\beta > 0$ . Por ejemplo  $\theta = 1/2$ ,  $\alpha = 1/2$  y  $\beta = 1$ .

Partimos de  $\mathbf{x}_n$ , que no es un punto crítico, de  $\mathbf{p}_n$  y ya hemos calculado también  $\mathbf{g}_n = \mathbf{g}(\mathbf{x}_n)$ . Formamos la función

$$\varphi(\lambda) = f(\mathbf{x}_n + \lambda \mathbf{p}_n), \quad \lambda > 0.$$

Al ser

$$\varphi'(0) = \langle \mathbf{g}_n, \mathbf{p}_n \rangle < 0,$$

la función comienza siendo decreciente, pero al aumentar  $\lambda$  no tenemos información y podría hasta crecer. Queremos a garantizar que realmente se produce un descenso, por ello vamos a intentar tomar  $0 < \lambda_n \leq \beta$  lo mayor posible pero de forma que

$$\varphi(\lambda_n) \leq \varphi(0) + \theta \varphi'(0) \lambda_n.$$

(en el punto  $\lambda_n$  el valor de  $\varphi(\lambda_n)$  permanece por debajo del correspondiente valor sobre la recta que en 0 vale  $\varphi(0)$  pero decrece sólo con la pendiente  $\theta \varphi'(0)$ ). Como

$$\lim_{k \rightarrow \infty} \frac{\varphi(\alpha^k \beta) - \varphi(0)}{\alpha^k \beta} = \varphi'(0) < \theta \varphi'(0),$$

tiene sentido el siguiente algoritmo (Armijo).

Dados  $\mathbf{x}_n$ ,  $\mathbf{g}_n$   $\mathbf{p}_n$ :

$m = \mathbf{g}'_n * \mathbf{p}_n$ ;  $\lambda = \beta$ ;  $\text{phi} = f(\mathbf{x})$ ;

$\text{philambda} = f(\mathbf{x} + \lambda * \mathbf{p}_n)$ ;  $\text{dif} = \text{phi} + \lambda * m * \theta - \text{philambda}$ ;

if  $((\|\mathbf{g}_n\| == 0) \parallel (m \geq 0))$ ;  $\lambda = 0$ ;

else; while  $(\text{dif} \leq 0)$ ;  $\lambda = \alpha * \lambda$ ;

$\text{philambda} = f(\mathbf{x} + \lambda * \mathbf{p}_n)$ ;  $\text{dif} = \text{phi} + \lambda * m * \theta - \text{philambda}$ ;

end (while); end (if)

y tomaremos luego  $\lambda_n = \lambda$ .

Pasamos a comentar sobre la filosofía para construir las matrices  $A_n$ . Idealmente nos gustaría adoptar

$$A_n \approx Hf(\mathbf{x}_n)^{-1},$$

o alternatively  $A_n = B_n^{-1}$ , con  $B_n \approx Hf(\mathbf{x}_n)$  (resultarán dos versiones diferentes):

$$\text{Tipo 1 : } \quad \mathbf{p}_n = A_n \mathbf{g}(\mathbf{x}_n); \quad \mathbf{x}_{n+1} = \mathbf{x}_n + \lambda_n \mathbf{p}_n;$$

$$\text{Tipo 1 : } \quad B_n \mathbf{p}_n = \mathbf{g}(\mathbf{x}_n); \quad \mathbf{x}_{n+1} = \mathbf{x}_n + \lambda_n \mathbf{p}_n.$$

La diferencia es clara, en la segunda alternativa hay que resolver un sistema lineal de ecuaciones en cada etapa. Vamos a referirnos (terminología que no es de uso común, es únicamente para el contexto de esta presentación) como implementación de tipo 1, respectivamente de tipo 2.

La idea es renunciar a aproximar las matrices  $Hf$  o  $Hf^{-1}$  y conformarnos con reflejar un rasgo significativo de las mismas, la llamada condición casi-Newton o de la secante, que pasamos a describir.

Ya observamos que  $D\nabla f(\mathbf{x})$  es la matriz hessiana  $Hf(\mathbf{x})$ . Por lo tanto

$$g(\mathbf{x} + \mathbf{h}) - g(\mathbf{x}) = Hf(\mathbf{x})\mathbf{h} + O(\|\mathbf{h}\|^2) \approx Hf(\mathbf{x})\mathbf{h}.$$

Aplicado a  $\mathbf{x}_{n+1} - \mathbf{x}_n + \lambda_n \mathbf{p}_n$ , resultará

$$g(\mathbf{x}_{n+1}) - g(\mathbf{x}_n) \approx Hf(\mathbf{x}_n)(\mathbf{x}_{n+1} - \mathbf{x}_n),$$

al menos cuando los vectores estén próximos a un límite.

En principio nos podríamos plantear que las matrices  $B_n$  cumplieran

$$B_n(\mathbf{x}_{n+1} - \mathbf{x}_n) = \mathbf{g}(\mathbf{x}_{n+1}) - \mathbf{g}(\mathbf{x}_n), \quad n \geq 0,$$

pero únicamente será operativo el imponer

$$B_{n+1}(\mathbf{x}_{n+1} - \mathbf{x}_n) = \mathbf{g}(\mathbf{x}_{n+1}) - \mathbf{g}(\mathbf{x}_n), \quad n \geq 0,$$

y esta es la condición casi-Newton o de la secante.

Si trabajamos en el formato de las matrices  $A_n$ , la condición casi-Newton convierte en

$$A_{n+1}(\mathbf{g}(\mathbf{x}_{n+1}) - \mathbf{g}(\mathbf{x}_n)) = \mathbf{x}_{n+1} - \mathbf{x}_n, \quad n \geq 0.$$

Partiendo pues de cierta matriz inicial  $B_0$  (resp.  $A_0$ ) definida positiva (por ejemplo una aproximación de  $Hf(\mathbf{x}_0)$ , la matrix identidad, ...), las matrices de índice posterior se van a ir construyendo a partir de las del índice previo mediante un proceso de puesta al día (up-dating): para obtener  $B_{n+1}$  (resp.  $A_{n+1}$ ) se va a modificar  $B_n$  (resp.  $A_n$ ) de la manera más sencilla posible para que se cumple la condición casi-Newton. La modificación consistirá en sumar una matriz de rango uno o dos.



Nos planteamos ahora un problema puramente algebraico: Dada una matriz  $E \in M_{d,d}$ , definida positiva, y dados dos vectores  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ , encontrar una matriz  $R \in M_{d,d}$ , de rango uno o dos, tal que la suma  $F = E + R \in M_{d,d}$  sea de nuevo definida positiva y cumpla la condición

$$F\mathbf{u} = \mathbf{v}.$$

Este problema, de antemano, no siempre tiene solución. En primer lugar, si  $\mathbf{u} = 0$ , la única opción será que  $\mathbf{v} = 0$ , en cuyo caso bastaría tomar  $F = E$ . Supondremos pues en lo sucesivo que  $\mathbf{u} \neq 0$ . Notemos que si existe  $F$  entonces

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, F\mathbf{u} \rangle > 0$$

y surge así una condición necesaria a cumplir por  $\mathbf{u}$  y  $\mathbf{v}$ .

Vamos a comenzar por las modificaciones de rango uno. Si  $R \in M_{d,d}$  es de rango unidad es porque todas sus columnas son proporcionales a un vector dado  $\xi \in \mathbb{R}^d$ , por lo tanto, para cierto  $\eta = [\eta_1, \dots, \eta_d]^T \in \mathbb{R}^d$ , tendremos

$$R = [\eta_1 \xi, \eta_2 \xi, \dots, \eta_d \xi] = \xi \eta^T.$$

Observemos que

$$R\mathbf{w} = \xi \eta^T \mathbf{w} = \xi \langle \eta, \mathbf{w} \rangle = \langle \mathbf{w}, \eta \rangle \xi, \quad \mathbf{w} \in \mathbb{R}^d.$$

Una matriz de este tipo será simétrica si

$$R^T = \eta \xi^T = \xi \eta = R,$$

para lo cual es necesario y suficiente que  $\eta = \mu \xi$ , para cierto  $\mu \in \mathbb{R}$ . En esta situación

$$\langle R\mathbf{w}, \mathbf{w} \rangle = \mu \langle \mathbf{w}, \xi \rangle^2, \quad \mathbf{w} \in \mathbb{R}^d$$

y la matriz  $R$  es semidefinida positiva si, y sólo si,  $\mu \geq 0$ .

Como pretendemos que  $E + R$  sea de carácter definido positivo y no utilizamos otra información que el hecho de que  $E$  tiene ese carácter, hemos de restringirnos pues a modificaciones de la forma anterior,  $R = \mu \xi \xi^T$ , con  $\mu \geq 0$ .

Imponemos ahora la condición

$$F\mathbf{u} = E\mathbf{u} + R\mathbf{u} = \mathbf{v} \Rightarrow R\mathbf{u} = \delta := \mathbf{v} - E\mathbf{u}.$$

Recordando que  $R\mathbf{u} = \langle \mathbf{u}, \xi \rangle \xi$ , la condición se transforma en

$$\mu \langle \mathbf{u}, \xi \rangle \xi = \delta,$$

y la única opción que tenemos es que para algún  $\gamma \in \mathbf{R}$ ,

$$\xi = \gamma \delta,$$

luego o bien  $\delta = 0$ , y entonces  $R = 0$ , o bien

$$\mu \gamma^2 \langle \mathbf{u}, \delta \rangle = 1.$$

Como  $\mu$  ha de ser positivo, para que el problema tenga solución resulta necesario que

$$\langle \mathbf{u}, \delta \rangle = \langle \mathbf{u}, \mathbf{v} - E\mathbf{u} \rangle > 0,$$

en cuyo caso podemos adoptar  $\gamma = 1$ ,  $\xi = \delta$  y finalmente

$$R = \mu \delta \delta^T, \quad \text{con } \delta = \mathbf{v} - E\mathbf{u}, \quad \mu = 1/\langle \mathbf{u}, \delta \rangle.$$

Notemos que

$$\langle \mathbf{u}, \delta \rangle = \langle \mathbf{u}, \mathbf{v} - E\mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, E\mathbf{u} \rangle < \langle \mathbf{u}, \mathbf{v} \rangle,$$

pues  $E$  es definida positiva. Así, la condición encontrada  $\langle \mathbf{u}, \delta \rangle > 0$  para construir la modificación de rango 1 es más exigente que la condición necesaria  $\langle \mathbf{u}, \mathbf{v} \rangle > 0$ .

Pasamos a discutir el problema con modificaciones de rango 2. Observamos que toda matriz  $F \in M_{d,d}$  de la forma

$$F = (I - \gamma \mathbf{v} \mathbf{u}^T) E (I - \gamma \mathbf{u} \mathbf{v}^T) + \gamma \mathbf{v} \mathbf{v}^T,$$

con  $\gamma \geq 0$  es definida positiva. En efecto, si  $\gamma = 0$ , entonces  $F = E$  y ya tenemos la conclusión. En general, llamemos  $P = (I - \gamma \mathbf{u} \mathbf{v}^T)$  de suerte que

$$F = P^T E P + \gamma \mathbf{v} \mathbf{v}^T$$

es siempre simétrica. Además, para  $\mathbf{w} \in \mathbb{R}^d$ ,

$$\langle F \mathbf{w}, \mathbf{w} \rangle = \langle E P \mathbf{w}, P \mathbf{w} \rangle + \gamma \langle \mathbf{w}, \mathbf{v} \rangle^2 \geq 0,$$

y de momento  $F$  será semidefinida positiva. Si estudiamos la anulación de la expresión anterior, como consta de dos sumandos positivos, podemos afirmar que, cuando  $\gamma > 0$ ,

$$\langle F \mathbf{w}, \mathbf{w} \rangle = 0 \Rightarrow \langle E P \mathbf{w}, P \mathbf{w} \rangle = 0 \quad \text{y} \quad \langle \mathbf{w}, \mathbf{v} \rangle = 0.$$

Ahora bien, si  $\langle \mathbf{w}, \mathbf{v} \rangle = 0$ , encontramos que

$$P\mathbf{w} = \mathbf{w} - \gamma \langle \mathbf{w}, \mathbf{v} \rangle \mathbf{u} = \mathbf{w},$$

y, como  $E$  es definida positiva, entonces  $P\mathbf{w} = \mathbf{w} = 0$ . En conclusión  $F$  es también definida.

Por otra parte, si desarrollamos la expresión de  $F$

$$F = E - \gamma \mathbf{v} \mathbf{u}^T E - \gamma E \mathbf{u} \mathbf{v}^T + \gamma^2 \mathbf{v} \mathbf{u}^T \mathbf{u} \mathbf{v}^T + \gamma \mathbf{v} \mathbf{v}^T$$

y agrupamos los términos de la diferencia  $R = F - E$  en la forma

$$R = \mathbf{v}(-\gamma(E\mathbf{u}) + \gamma^2 \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v} + \gamma \mathbf{v})^T + (E\mathbf{u})\mathbf{v}^T,$$

vemos que  $R$  es la suma de dos matrices de rango unidad, luego  $R$  es de rango  $\leq 2$ .

Finalmente, si queremos que  $F\mathbf{u} = \mathbf{v}$ , recordando que

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, F\mathbf{u} \rangle > 0,$$

basta tomar

$$\gamma = \frac{1}{\langle \mathbf{u}, \mathbf{v} \rangle},$$

pues entonces

$$\gamma \mathbf{v} \mathbf{v}^T \mathbf{u} = \gamma \mathbf{v} \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}$$

y

$$P\mathbf{u} = \mathbf{u} - \gamma \mathbf{u} \mathbf{v}^T \mathbf{u} = \mathbf{u} - \gamma \mathbf{u} \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{0},$$

y así claramente

$$F\mathbf{u} = \mathbf{v}.$$

En resumen, la modificación de rango 2 es siempre posible cuando se cumple la condición necesaria del problema.

En los programas se incluye la función

$$F = \text{update}(E, \mathbf{u}, \mathbf{v}, \text{rango})$$

que lleva a cabo la modificacion cuando ésta es posible y, en otro caso, produce  $F = E$ .

Las modificaciones de rango 1, combinadas con la condición de la secante, dan origen a los métodos DFP (Davidon, Fletcher and Powell). Las de rango 2, a los llamados métodos BFGS (Broyden, Fletcher, Goldfarg and Shanno). Cada una de estas opciones se puede utilizar en el formato  $A's$  (tipo 1 en programa principal, que evita la resolución de sistemas lineales) o en el de las  $B's$  (tipo 2 en el programa principal).



Un método de descenso, bajo condiciones razonables (el punto de partida  $\mathbf{x}_0$  es tal que

$$\{ \mathbf{x} \in \Omega : f(\mathbf{x}) \leq f(\mathbf{x}_0) \}$$

es compacto) produce una sucesión que admite una subsucesión que converge hacia un punto crítico  $\mathbf{x}^*$ .

Si además  $Hf(\mathbf{x}^*)$  es definida positiva, es realmente sorprendente que los métodos BFGS son convergentes (de forma superlineal) y las matrices  $B_n \rightarrow Hf(x^*)$ ,  $A_n \rightarrow Hf(x^*)^{-1}$ . La convergencia superlineal significa que

$$\lim_{n \rightarrow +\infty} \frac{\|x_{n+1} - \mathbf{x}^*\|}{\|x_n - \mathbf{x}^*\|} = 0.$$