

|          |   |            |
|----------|---|------------|
| <b>5</b> | <b>Métodos iterativos para la resolución de ecuaciones algebraicas lineales</b> | <b>95</b>  |
| 5.1      | Método de Gauss-Jacobi . . . . .  | 95         |
| 5.2      | Método de Gauss-Seidel . . . . .  | 97         |
| 5.3      | Método de sobrerrelajación sucesiva . . . . .                                   | 100        |
| 5.4      | Convergencia de los métodos iterativos estacionarios . . . . .                  | 103        |
| 5.5      | Método del descenso más rápido . . . . .  | 104        |
| 5.6      | Método del gradiente conjugado y preconditionamiento . . . . .                  | 105        |
|          | <b>Bibliografía</b>   | <b>109</b> |

24 de octubre de 2001

© Francisco R. Villatoro, Carmen M. García, Juan I. Ramos. Estas notas están protegidas por derechos de copyright y pueden ser distribuidas libremente sólo con propósitos educativos sin ánimo de lucro. *These notes are copyright-protected but may be freely distributed for instructional nonprofit purposes.*



## CAPÍTULO 5

# MÉTODOS ITERATIVOS PARA LA RESOLUCIÓN DE ECUACIONES ALGEBRAICAS LINEALES

### 5.1 Método de Gauss-Jacobi

Los métodos iterativos para la resolución de un sistema lineal  $Ax = b$  parten de una aproximación inicial  $x^{(0)}$  a la solución del sistema lineal y obtienen una sucesión de soluciones  $x^{(k)}$  que, bajo determinadas hipótesis, converge a la solución del sistema lineal.

El método de Gauss-Jacobi o de Jacobi para la resolución del sistema lineal consiste en obtener una nueva aproximación a la solución  $x^{(k+1)}$  a partir de una previa  $x^{(k)}$  mediante el siguiente proceso

$$\begin{aligned}x_1^{(k+1)} &= \frac{1}{a_{11}} (b_1 - a_{12} x_2^{(k)} - \dots - a_{1n} x_n^{(k)}), \\x_2^{(k+1)} &= \frac{1}{a_{22}} (b_2 - a_{21} x_1^{(k)} - a_{23} x_3^{(k)} - \dots - a_{2n} x_n^{(k)}), \\&\dots \\x_n^{(k+1)} &= \frac{1}{a_{nn}} (b_n - a_{n1} x_1^{(k)} - \dots - a_{n(n-1)} x_{n-1}^{(k)}).\end{aligned}$$

Este proceso iterativo se puede representar matricialmente si descomponemos la matriz

$$A = L + D + U, \quad a_{ij} = l_{ij} + d_{ij} + u_{ij},$$

donde  $L$  y  $U$  son matrices triangulares estrictas inferior y superior, respectivamente, y  $D$  es una matriz diagonal tales que

$$D = (d_{ij}), \quad d_{ij} = \delta_{ij} a_{ii},$$

$$L = (l_{ij}), \quad l_{ij} = \begin{cases} a_{ij}, & j < i \\ 0, & i \leq j, \end{cases}$$

$$U = (u_{ij}), \quad u_{ij} = \begin{cases} a_{ij}, & j > i \\ 0, & i \geq j. \end{cases}$$

De esta forma podemos escribir el sistema lineal como  $(L + D + U)x = b$  o, lo que es lo mismo,

$$Dx = b - (L + U)x, \quad (5.1)$$

y vemos que la iteración de Gauss-Jacobi consiste en resolver en cada iteración los sistemas diagonales siguientes

$$Dx^{(k+1)} = b - (L + U)x^{(k)}. \quad (5.2)$$

Para estudiar la condición de convergencia del método de Gauss-Jacobi, definiremos el error

$$e^{(k)} = x - x^{(k)},$$

con lo que, restando las ecuaciones (5.1) y (5.2), obtenemos la siguiente ecuación para el error

$$De^{(k+1)} = -(L + U)e^{(k)},$$

cuya solución es muy fácil de obtener,

$$e^{(k+1)} = -D^{-1}(L + U)e^{(k)}.$$

Aplicando normas a esta expresión

$$\|e^{(k+1)}\| = \|D^{-1}(L + U)e^{(k)}\| \leq \|D^{-1}(L + U)\| \|e^{(k)}\|,$$

obtenemos que la condición de convergencia es

$$\|D^{-1}(L + U)\| < 1,$$

para cualquier norma. Como el radio espectral es el ínfimo de todas las normas, es condición necesaria y suficiente para la convergencia del método de Jacobi que

$$\rho(D^{-1}(L + U)) < 1.$$

Podemos obtener una condición suficiente para la convergencia del método de Jacobi si utilizamos la norma 1, con lo que

$$\begin{aligned} \|D^{-1}(L+U)\|_1 &= \max_{1 \leq j \leq n} \sum_i \left| \frac{1}{d_{ii}} (l_{ij} + u_{ij}) \right| \\ &= \max_{1 \leq j \leq n} \sum_{j \neq i=1}^n \left| \frac{a_{ij}}{a_{ii}} \right| < 1, \end{aligned}$$

y obtenemos

$$\sum_{i=1, i \neq j}^n |a_{ij}| < |a_{jj}|, \quad \forall 1 \leq j \leq n,$$

es decir, que la matriz  $A$  ha de ser diagonalmente dominante por filas. Si se toma norma infinito se obtiene que  $A$  ha de ser diagonalmente dominante por columnas. Estas condiciones son suficientes para la convergencia del método de Jacobi, pero no son necesarias. Por ejemplo, para la matriz

$$A = \begin{pmatrix} 1 & -1 & 1 & 0 \\ -1/2 & 1 & 0 & -1/2 \\ -1/2 & 0 & 1 & -1/2 \\ 0 & -1/2 & -1/2 & 1 \end{pmatrix}, \quad (5.3)$$

no es diagonalmente dominante y sin embargo

$$\rho(D^{-1}(L+U)) = \frac{1}{\text{sqrt}2} < 1,$$

el método de Jacobi converge.

## 5.2 Método de Gauss-Seidel

El método de Gauss-Seidel para la resolución del sistema lineal consiste en obtener una nueva aproximación a la solución  $x^{(k+1)}$  a partir de una previa  $x^{(k)}$  mediante el siguiente proceso

$$\begin{aligned} x_1^{(k+1)} &= \frac{1}{a_{11}} (b_1 - a_{12} x_2^{(k)} - \dots - a_{1n} x_n^{(k)}), \\ x_2^{(k+1)} &= \frac{1}{a_{22}} (b_2 - a_{21} x_1^{(k+1)} - a_{23} x_3^{(k)} - \dots - a_{2n} x_n^{(k)}), \\ &\dots \\ x_n^{(k+1)} &= \frac{1}{a_{nn}} (b_n - a_{n1} x_1^{(k+1)} - \dots - a_{2(n-1)} x_{n-1}^{(k+1)}). \end{aligned}$$

El sistema de ecuaciones se puede escribir como

$$(D + L)x = b - Ux, \quad (5.4)$$

y vemos que la iteración de Gauss-Seidel consiste en resolver los sistemas

$$(D + L)x^{(k+1)} = b - Ux^{(k)}. \quad (5.5)$$

La ecuación para el error se obtiene restando las ecuaciones (5.4) y (5.5),

$$(D + L)e^{(k+1)} = -Ue^{(k)}, \quad e^{(k)} = x - x^{(k)},$$

cuya solución es

$$e^{(k+1)} = -(D + L)^{-1} Ue^{(k)}. \quad (5.6)$$

Aplicando normas a esta expresión

$$\|e^{(k+1)}\| \leq \|(D + L)^{-1} U\| \|e^{(k)}\| \leq \|(D + L)^{-1}\| \|U\| \|e^{(k)}\|.$$

La condición necesaria y suficiente para la convergencia del método de Gauss es

$$\rho((D + L)^{-1} U) < 1.$$

Vamos a obtener una condición suficiente (pero no necesaria) para la convergencia que sea más manejable que la del radio espectral, y que evite su cálculo. Escribiendo la ecuación (5.6) con subíndices

$$e_i^{(k+1)} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} e_j^{(k+1)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} e_j^{(k)},$$

y buscando la norma infinito introduciendo valores absolutos, obtenemos

$$\begin{aligned} |e_i^{(k+1)}| &\leq \sum_{j=1}^{i-1} \left| \frac{a_{ij}}{a_{ii}} \right| |e_j^{(k+1)}| + \sum_{j=i+1}^n \left| \frac{a_{ij}}{a_{ii}} \right| |e_j^{(k)}| \\ &\leq \sum_{j=1}^{i-1} \left| \frac{a_{ij}}{a_{ii}} \right| \|e^{(k+1)}\|_\infty + \sum_{j=i+1}^n \left| \frac{a_{ij}}{a_{ii}} \right| \|e^{(k)}\|_\infty \\ &= \alpha_i \|e^{(k+1)}\|_\infty + \beta_i \|e^{(k)}\|_\infty. \end{aligned}$$

donde  $\alpha_i \geq 0$  y  $\beta_i \geq 0$  son números reales positivos. Sea  $m$  el índice de la mayor componente del vector  $e^{(k+1)}$ , de forma que

$$\|e^{(k+1)}\|_\infty = |e_m^{(k+1)}|,$$

y, por tanto,

$$\|e^{(k+1)}\|_\infty \leq \alpha_m \|e^{(k+1)}\|_\infty + \beta_m \|e^{(k)}\|_\infty,$$

de donde despejando

$$\|e^{(k+1)}\|_\infty \leq \frac{\beta_m}{1 - \alpha_m} \|e^{(k)}\|_\infty \leq \eta \|e^{(k)}\|_\infty,$$

con

$$\eta = \max_i \frac{\beta_i}{1 - \alpha_i}.$$

Si notamos

$$\mu = \max_i (\alpha_i + \beta_i) = \max_{1 \leq j \leq n} \sum_{j \neq i=1}^n \left| \frac{a_{ij}}{a_{ii}} \right|,$$

podemos escribir

$$\alpha_i + \beta_i - \frac{\beta_i}{1 - \alpha_i} = \frac{\alpha_i}{1 - \alpha_i} (1 - (\alpha_i + \beta_i)) \geq \frac{\alpha_i}{1 - \alpha_i} (1 - \mu) \geq 0,$$

con lo que observamos que si  $\mu < 1$ , es decir, la matriz es diagonalmente dominante por filas, entonces

$$\alpha_i + \beta_i \geq \frac{\beta_i}{1 - \alpha_i},$$

y por tanto  $1 > \mu \geq \eta$ , por lo que el método de Gauss-Seidel converge en dicho caso. La dominancia diagonal es una condición suficiente que no es necesaria, por ejemplo, para la matriz (5.3),

$$\rho((L + D)^{-1}(U)) = \frac{1}{2} < 1,$$

y el método de Seidel converge.

Podemos relacionar la convergencia de los métodos de Jacobi y Seidel. Como en el método de Gauss-Jacobi teníamos

$$|e_i^{(k+1)}| \leq \sum_{j=1, i \neq j}^n \left| \frac{a_{ij}}{a_{ii}} \right| |e_j^{(k)}| \leq \sum_{j=1, i \neq j}^n \left| \frac{a_{ij}}{a_{ii}} \right| \|e^{(k)}\|_\infty = \mu \|e^{(k)}\|_\infty,$$

entonces

$$\|e^{(k)}\|_\infty \leq \mu \|e^{(k)}\|_\infty,$$

con lo que tenemos que, si el método de Gauss-Jacobi converge entonces también converge el de Gauss-Seidel ( $1 > \mu \geq \eta$ ) y además converge linealmente con mayor rapidez que la del de Gauss-Jacobi.

En resumen, una condición suficiente para que el método de Seidel converja es que la matriz  $A$  sea diagonalmente dominante, y en ese caso ya sabemos que también convergerá el método de Jacobi. Además, si el método de Jacobi converge, aunque la matriz no sea diagonalmente

dominante, también lo hará el de Seidel, que lo hará más rápido, algo que hemos visto con la matriz (5.3) de los ejemplos anteriores.

Aunque los métodos iterativos requieren un número infinito de iteraciones para converger, en la práctica se realiza solamente un número finito de iteraciones hasta que se alcanza alguna tolerancia de error. Existen muchos procedimientos para aplicar dicha tolerancia de error, por ejemplo,

1. utilizando el error absoluto entre iterados (recuerde que se si  $x^{(k)}$  converge es una sucesión de Cauchy),

$$\|x^{(k+1)} - x^{(k)}\| \leq \text{TOL}_{\text{abs}},$$

2. utilizando el error relativo,

$$\frac{\|x^{(k+1)} - x^{(k)}\|}{\|x^{(k+1)}\|} \leq \text{TOL}_{\text{rel}},$$

3. acotando el residuo en cada iteración,

$$\|Ax^{(k+1)} - b\| = \|r^{(k+1)}\| \leq \text{TOL}_{\text{abs}},$$

4. una cota muy utilizada es acotar el residuo de forma relativa al vector no homogéneo,

$$\frac{\|r\|}{\|b\|} \leq \text{TOL}_{\text{rel}}.$$

La ventaja fundamental de los métodos iterativos es que minimizan el efecto de los errores de redondeo ya que su propagación se puede controlar mediante la determinación iterativa de la solución.

### 5.3 Método de sobrerrelajación sucesiva

El método de sobrerrelajación sucesiva se puede interpretar como un método predictor-corrector, cuyo predictor es un método iterativo como el de Gauss-Jacobi o Gauss-Seidel y como corrección realiza un promedio entre la solución en la iteración anterior y la obtenida por el predictor.

Para obtener un método de sobrerrelajación sucesiva a partir de la expresión iterativa del método de Gauss-Jacobi, escribimos este método de la siguiente forma

$$x_{Ji}^{(k+1)} = \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right)$$



$$\begin{aligned}
&= \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - a_{ii} x_i^{(k)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} + a_{ii} x_i^{(k)} \right) \\
&= x_i^{(k)} + \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^n a_{ij} x_j^{(k)} \right).
\end{aligned}$$

El método de sobrerelajación sucesiva introduce un coeficiente  $w$  de sobrerelajación

$$x_i^{(k+1)} = w x_{ji}^{(k+1)} + (1 - w) x_i^{(k)}.$$

De esta forma, se obtiene la siguiente iteración

$$x_i^{(k+1)} = (1 - w) x_i^{(k)} + \frac{w}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right).$$

Matricialmente se puede escribir como ( $A = L + D + U$ )

$$\begin{aligned}
D x^{(k+1)} &= (1 - w) D x^{(k)} + w (b - (L + U) x^{(k)}) \\
&= D x^{(k)} + w (b - (L + D + U) x^{(k)}) \\
&= D x^{(k)} + w (b - A x^{(k)}).
\end{aligned}$$

El parámetro de relajación nos permite controlar el método utilizado

$$\left\{ \begin{array}{l} w < 1 \quad \text{subrelajación} \\ w = 1 \quad \text{método predictor} \\ w > 1 \quad \text{sobrerelajación} \end{array} \right.$$

Se puede obtener otro método de sobrerelajación sucesiva a partir de la expresión iterativa del método de Gauss-Seidel,

$$\begin{aligned}
x_{Si}^{(k+1)} &= \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right) \\
&= \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} + a_{ii} x_i^{(k)} - a_{ii} x_i^{(k)} \right) \\
&= x_i^{(k)} + \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i}^n a_{ij} x_j^{(k)} \right),
\end{aligned}$$

para el que la expresión

$$x_i^{(k+1)} = w x_{S_i}^{(k+1)} + (1-w) x_i^{(k)}.$$

conduce al siguiente método

$$x_i^{(k+1)} = (1-w) x_i^{(k)} + \frac{w}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right).$$

Este método se escribe en forma matricial como

$$D x^{(k+1)} = (1-w) D x^{(k)} + w (b - L x^{(k+1)} - U x^{(k)}),$$

$$(D + w L) x^{(k+1)} = (1-w) D x^{(k)} + w (b - U x^{(k)})$$

$$= ((1-w) D - w U) x^{(k)} + w b.$$

La ecuación del error para este método, se escribe

$$(D + w L) e^{(k+1)} = ((1-w) D - w U) e^{(k)},$$

$$(I + w D^{-1} L) e^{(k+1)} = ((1-w) I - w D^{-1} U) e^{(k)},$$

$$e^{(k+1)} = (I + w D^{-1} L)^{-1} ((1-w) I - w D^{-1} U) e^{(k)},$$

y tomando normas,

$$\|e^{(k+1)}\| \leq \|(I + w D^{-1} L)^{-1}\| \|((1-w) I - w D^{-1} U)\| \|e^{(k)}\|.$$

Se debe elegir  $w$  de tal forma que minimice el error, es decir,  $w$  debe minimizar la norma de la matriz

$$(I + w D^{-1} L)^{-1} ((1-w) I - w D^{-1} U).$$

Los métodos iterativos son más útiles que los directos siempre y cuando su coste sea menor. Sea  $m$  el número de iteraciones necesarias para alcanzar una tolerancia dada de error, como el número de operaciones aritméticas por iteración es del orden de  $O(n^2)$ , el número total de operaciones es  $O(mn^2)$ , que deberá ser menor que el coste de los métodos directos, por ejemplo,  $O(2n^3/3)$  para la eliminación de Gauss. Para los métodos de sobrerrelajación que hemos presentado su convergencia es lineal, por tanto, tras  $m$  iteraciones el error está acotado por

$$\|x - x^{(k)}\| \leq c^m \|x - x^{(0)}\|,$$

y si se requiere una tolerancia de error absoluto  $\epsilon$ , entonces el número de iteraciones debe ser mayor que

$$m \geq -\frac{\ln \epsilon}{c}.$$

## 5.4 Convergencia de los métodos iterativos estacionarios

Los métodos de Gauss-Jacobi, Gauss-Seidel y de sobrerelajación son casos particulares de los métodos iterativos estacionarios, que tienen la forma

$$x^{(k+1)} = B x^{(k)} + c.$$

Se debe aclarar que se denominan métodos iterativos no estacionarios a los métodos en los que la matriz  $B$  y el vector  $c$  cambian en cada iteración, es decir, se sustituyen por  $B^{(k)}$  y  $c^{(k)}$ .

La condición de consistencia requiere que el punto fijo de esta iteración sea solución del sistema lineal a resolver, es decir,

$$x = B x + c \quad \Rightarrow \quad A x = b,$$

lo que fija el valor de  $c$  como función de  $B$ ,

$$(I - B) x = c, \quad x = A^{-1} b, \quad \Rightarrow \quad c = (I - B) A^{-1} b.$$

Si escribimos el método iterativo estacionario como

$$x^{(k+1)} = B x^{(k)} + E b,$$

la condición de consistencia es

$$E = (I - B) A^{-1} \quad \Rightarrow \quad I = B + E A.$$

El método iterativo estacionario más simple es el de Picard

$$(I + A) x = x + b \quad \Rightarrow \quad x^{(k+1)} = (I + A) x^{(k)} - b.$$

Para estudiar la convergencia de los métodos iterativos estacionarios, como el error cumple

$$e^{(k+1)} = x^{(k+1)} - x = B x^{(k)} + E b - B x - E b = B e^{(k)},$$

se dará la convergencia si

$$\lim_{k \rightarrow \infty} \|e^{(k+1)}\| = \lim_{k \rightarrow \infty} \|B e^{(k)}\| = 0,$$

y como  $\|e^{(k+1)}\| \leq \|B\| \|e^{(k)}\|$ , el criterio de Cauchy para la convergencia de una sucesión nos dice que

$$\frac{\|e^{(k+1)}\|}{\|e^{(k)}\|} \leq \|B\| < 1,$$

para alguna norma. Esta es la condición necesaria y suficiente para la convergencia, que se suele escribir en la forma equivalente como  $\rho(B) < 1$ .

## 5.5 Método del descenso más rápido

Para matrices simétricas definidas positivas  $A = A^\top \in \mathbb{R}^{n \times n}$ , y para matrices hermíticas  $A = A^* \in \mathbb{C}^{n \times n}$ , se puede interpretar la resolución del sistema lineal de ecuaciones  $Ax = b$  como la solución de un problema de optimización cuadrático. En concreto, consideraremos el caso real (matriz simétrica), para el que

$$Ax = b \quad \Leftrightarrow \quad x = \min f(y), \quad \forall y \in \mathbb{R}^n,$$

donde

$$f(x) = \frac{1}{2} x^\top A x - b^\top x = \frac{1}{2} \langle x, Ax \rangle - \langle b, x \rangle.$$

La demostración es sencilla si consideramos la función aplicada a un rayo  $x + ty$ , con  $x$  e  $y$  dos vectores cualesquiera y  $t$  un escalar real. Como

$$\begin{aligned} 2f(x + ty) &= \langle x + ty, A(x + ty) \rangle - 2\langle b, x + ty \rangle \\ &= \langle x, Ax \rangle + t\langle x, Ay \rangle + t\langle y, Ax \rangle + t^2\langle y, Ay \rangle - 2\langle b, x + ty \rangle \\ &= 2f(x) + 2t\langle y, Ax \rangle - 2t\langle y, b \rangle + t^2\langle y, Ay \rangle \\ &= 2f(x) + 2t\langle y, Ax - b \rangle + t^2\langle y, Ay \rangle, \end{aligned}$$

donde se ha aplicado que  $A = A^\top$ . Como  $\langle y, Ay \rangle$  es positivo por ser  $A$  definida positiva, la función cuadrática tiene un mínimo sobre el rayo y no un máximo, **cqtd**. El valor de  $t$  que nos da el mínimo sobre el rayo se obtiene de la condición de que la derivada sea nula,

$$\frac{d}{dt} f(x + ty) = \langle y, Ax - b \rangle + t\langle y, Ay \rangle = 0,$$

por lo que el mínimo se obtiene para  $\hat{t}$ ,

$$\hat{t} = -\frac{\langle y, Ax - b \rangle}{\langle y, Ay \rangle}.$$

Sustituyendo este valor mínimo obtenemos

$$\begin{aligned} 2f(x + \hat{t}y) &= 2f(x) + \hat{t}(2\langle y, Ax - b \rangle + \hat{t}\langle y, Ay \rangle) \\ &= 2f(x) + \hat{t}\langle y, Ax - b \rangle \\ &= 2f(x) - \frac{\langle y, Ax - b \rangle^2}{\langle y, Ay \rangle}. \end{aligned}$$

Este cálculo demuestra que  $f(x + \hat{t}y)$  es más pequeño que  $f(x)$  a menos que  $\langle y, Ax - b \rangle = 0$ . Si  $Ax \neq b$ , entonces existen vectores  $y$  que no son ortogonales al residuo  $\langle y, Ax - b \rangle \neq 0$ , por lo que  $\hat{t}$  no es el mínimo. Por tanto, debe ser  $Ax - b = 0$ , como se quería demostrar.

El método del descenso más rápido es un algoritmo iterativo que se basa en la demostración precedente. Se basa en la siguiente ecuación iterativa

$$x^{(k+1)} = x^{(k)} + t_k y^{(k)},$$

donde

$$t_k = \frac{\langle y^{(k)}, Ax^{(k)} - b \rangle}{\langle y^{(k)}, Ay^{(k)} \rangle}$$

y  $y^{(k)}$  es la dirección del gradiente negativo de  $f$  en  $x^{(k)}$ , que se puede calcular fácilmente como

$$\begin{aligned} \frac{\partial f}{\partial x_k} &= \frac{\partial}{\partial x_k} \left( \frac{1}{2} \sum_{i,j} x_i a_{ij} x_j - \sum_i x_i b_i \right) \\ &= \frac{1}{2} \sum_{i,j} a_{ij} x_j \delta_{ik} + \frac{1}{2} \sum_{i,j} x_i a_{ij} \delta_{jk} - \sum_i b_i \delta_{ik} \\ &= \sum_{i,k} a_{ik} x_k - b_k = Ax - b, \end{aligned}$$

ya que  $A$  es simétrica, por lo que

$$y^{(k)} = -(Ax^{(k)} - b).$$

El método del descenso más rápido es un método iterativo que se puede aplicar para minimizar una función no lineal  $f(x)$  cualquiera. Consiste en elegir un camino  $x^{(0)}, x^{(1)}, \dots$  de búsqueda del mínimo mirando en la dirección en la que  $f$  decrece más rápidamente (de ahí el nombre). Esta dirección es la del gradiente negativo de  $f$ , es decir,

$$n = -\frac{\nabla f}{|\nabla f|}.$$

El inconveniente fundamental del método del descenso más rápido es su lenta convergencia. En problemas no lineales en los que  $f$  tiene mínimos locales, este método puede converger a un mínimo local que no sea el mínimo global que se desea como solución.

## 5.6 Método del gradiente conjugado y preconditionamiento

Asumamos que  $x_0 = 0$ . Si no lo es, definimos

$$Az = b - Ax_0,$$

cuya solución  $z$  define  $x = x_0 + z$ , y por tanto, la estimación inicial  $x_0$  corresponde a  $z_0 = 0$ .

Dos vectores  $p_i$  y  $p_j$  se denominan vectores  $A$ -conjugados si  $\langle p_i, A p_j \rangle = p_i^\top A p_j = 0$ , para  $i \neq j$ .

Los vectores conjugados definen direcciones conjugadas. Sea la norma- $A$  de un vector  $x$ ,

$$\|x\|_A = \sqrt{\langle x, x \rangle_A} = \sqrt{x^\top A x}.$$

Sea  $A x_e = b$  y definamos el error de una aproximación  $x$

$$E(x) = \frac{1}{2} \|x_e - x\|_A^2.$$

Como  $n$  vectores conjugados definen una base de  $\mathbb{R}^n$ , podemos escribir la solución exacta como

$$x_e = \alpha_1 p_1 + \cdots + \alpha_n p_n$$

donde

$$\alpha_n = \frac{\langle p_k, x_e \rangle_A}{\langle p_k, p_k \rangle_A} = \frac{p_k^\top A x_e}{p_k^\top A p_k} = \frac{p_k^\top b}{p_k^\top A p_k}.$$

El método de la iteración en las direcciones conjugadas consiste en iterar a partir de  $x_0 = 0$ , en la forma

$$x_k = x_{k-1} + \alpha_k p_k,$$

$$x_k = \alpha_1 p_1 + \cdots + \alpha_k p_n,$$

y como

$$r_k = b - A x_k = -\nabla f(x_k), \quad r_0 = b,$$

$$r_k = b - A x_{k-1} - \alpha_k A p_k = r_{k-1} - \alpha_k A p_k.$$

De esta forma, en el  $k$ -ésimo paso  $k = n$  tendremos  $x_n = x_e$  y  $r_n = 0$ .

El método del gradiente conjugado parte de  $x_0 = 0$ , y opera como sigue

$$p_1 = -\nabla f(x_0) = r_0 = b,$$

$$p_{k+1} = r_k + \beta_{k+1} p_k,$$

que si imponemos que  $p_k$  y  $p_{k+1}$  son conjugados,

$$p_k^\top A p_{k+1} = 0 \quad \Rightarrow \quad \beta_{k+1} = -\frac{p_k^\top A r_k}{p_k^\top A p_k}.$$

De esta forma, definimos la iteración

$$x_k = x_{k-1} + \alpha_k p_k,$$

donde exigiendo

$$x_e = \alpha_1 p_1 + \cdots + \alpha_n p_n,$$

obtenemos, como antes,

$$\alpha_k = \frac{p_k^\top b}{p_k^\top A p_k}.$$

El método del gradiente conjugado tiene la ventaja de que determina la solución en  $m \leq n$  iteraciones.

La velocidad de convergencia de todos los métodos iterativos que hemos presentado y, en particular, el método del gradiente conjugado, dependen del número de condicionamiento de la matriz de coeficientes. Con objeto de reducir este número, se puede utilizar la técnica de preconditionamiento. Sea  $Q$  una matriz cuya inversa  $Q^{-1}$  exista y sea fácil de calcular, entonces

$$Ax = b \quad \Rightarrow \quad Q^{-1} A Q^{-1} Qx = Q^{-1} b,$$

y  $x$  se puede determinar resolviendo los siguientes problemas

$$Qx = y, \quad By = Q^{-1} b,$$

donde  $B = Q^{-1} A Q^{-1}$  y  $Q$  se debe elegir de tal forma que  $\text{cond}(B) \ll \text{cond}(A)$ .

## Apéndice A: Teorema de Householder-John

Sea una matriz  $A$  simétrica (real), no singular, y una descomposición  $A = M - N$ , donde  $M$  es no singular tal que

$$Q = M + M^\top - A = N + M^\top,$$

sea definida positiva. Entonces

$$\rho(M^{-1} N) < 1,$$

si y sólo si  $A$  es definida positiva.

**Demostración.** Llamemos  $B = M^{-1} N$ . Como

$$M^{-1} A = I - M^{-1} N = I - B,$$

$B = I - M^{-1}A$  y  $A = M - MB$ . Supongamos que  $A$  es definida positiva (en  $\mathbb{R}$ ), es decir, para todo vector  $x$ ,

$$\langle x, Ax \rangle = x^\top Ax > 0.$$

Cojamos un autovalor y un autovector de  $B$ , sean  $Bx = \lambda x$ , con lo que

$$Ax = (1 - \lambda)Mx.$$

Como  $A$  es definida positiva,  $\lambda \neq 1$ , sino sería  $x = 0$ . Aplicando el producto interior (real) tenemos

$$\langle \bar{x}, Ax \rangle = \bar{x}^\top Ax = (1 - \lambda) \langle \bar{x}, Mx \rangle,$$

y su complejo conjugado

$$\overline{\langle \bar{x}, Ax \rangle} = \langle \bar{x}, Ax \rangle = (1 - \bar{\lambda}) \langle \bar{x}, Mx \rangle,$$

ya que  $M$  es una matriz real. Sumando las dos expresiones anteriores obtenemos

$$\left( \frac{1}{1 - \lambda} + \frac{1}{1 - \bar{\lambda}} \right) \langle \bar{x}, Ax \rangle = \langle \bar{x}, (M + M^\top)x \rangle,$$

es decir,

$$2 \operatorname{Re} \left\{ \left( \frac{1}{1 - \lambda} \right) \right\} = \frac{\langle \bar{x}, (M + M^\top)x \rangle}{\langle \bar{x}, Ax \rangle} = 1 + \frac{\langle \bar{x}, Qx \rangle}{\langle \bar{x}, Ax \rangle},$$



## BIBLIOGRAFÍA

- [1] Granero Rodríguez, Francisco, “Álgebra y geometría analítica,” McGraw-Hill / Interamericana de España, 1985. [FTM-4-c/GRA/alg (5)]
- [2] Hernández Rodríguez, Eugenio, “Álgebra y geometría,” (2<sup>a</sup> ed.), Addison-Wesley Iberoamericana España, 1998. [FTM-4/HER (5)]
- [3] Burgos Román, Juan de, “Álgebra lineal,” McGraw-Hill / Interamericana de España, 1993. [FTM-4-c/B (9)]