MÉTODOS ITERATIVOS

EDUARDO MARTÍNEZ

Se denominan métodos iterativos o métodos de aproximaciones sucesivas a los que consisten en la obtención de una sucesión de valores $\{x^{(k)}\}$ que converja a la solución de nuestro problema. En general, se dá un procedimiento para hallar una aproximación $x^{(k+1)}$ a partir de la aproximación anterior $x^{(k)}$, de manera que la sucesión de aproximaciones queda determinada al dar un primer valor inicial $x^{(0)}$.

En el caso de la solución de sistemas lineales $Ax = b$, que es el objeto de este tema, la regla para hallar una aproximación a partir de la anterior es también lineal $x^{(k+1)} = Tx^{(k)} + c$, y habrá que obtener condiciones para que la sucesión de aproximaciones sucesivas converja a la solución del sistema dado.

Veremos en primer lugar dos de los métodos iterativos más conocidos: el método de Jacobi y el método de Gauss-Seidel. Dejando el estudio de la convergencia de estos y otros métodos para secciones posteriores.

1. Método de Jacobi

En el método de Jacobi se despeja la primera variable de la primera ecuación, la segunda variable de la segunda ecuación y así sucesivamente. Si tenemos la k-ésima aproximación $x^{(k)}$ entonces substituimos en los términos de la derecha y obtenemos el valor de $x^{(k+1)}$. Como valor inicial se suele tomar $x^{(0)} = 0$, aunque puede elegirse cualquier otro punto.

Ejemplo: Uno convergente 3x3. [[Burden pag.444]]

Ha de notarse que estamos usando la palabra 'aproximaciones' en un sentido muy amplio (es una declaración de intenciones, más que una propiedad), ya que de momento nada nos asegura que los sucesivos valores que vamos obteniendo vayan aproximándose a la solución.

Ejemplo: Uno divergente 3x3

Ejemplo: Uno oscilante

En general tenemos el sistema $Ax = b$ explícitamente,

$$\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1\\
&a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2\\
&\vdots\\
&a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n
\end{aligned}$$

y, siempre que cada a_{ii} sea no nulo, despejamos

$$\begin{aligned}
x_1 &= \big(b_1 - a_{12}x_2 - a_{13}x_3 - \cdots - a_{1n}x_n\big)/a_{11}\\
x_2 &= \big(b_2 - a_{21}x_1 - a_{23}x_3 - \cdots - a_{2n}x_n\big)/a_{22}\\
&\vdots\\
x_n &= \big(b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{nn-1}x_{n-1}\big)/a_{nn}
\end{aligned}$$

. Sustituyendo $x^{(k)}$ en la derecha obtenemos el valor de $x^{(k+a)}$ a la izquierda,

$$\begin{aligned}
x_1^{(k+1)} &= \big(b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots - a_{1n}x_n^{(k)}\big)/a_{11} \\
x_2^{(k+1)} &= \big(b_2 - a_{21}x_1^{(k)} - a_{23}x_3^{(k)} - \cdots - a_{2n}x_n^{(k)}\big)/a_{22} \\
&\vdots \\
x_n^{(k+1)} &= \big(b_n - a_{n1}x_1^{(k)} - a_{n2}x_2^{(k)} - \cdots - a_{nn-1}x_{n-1}^{(k)}\big)/a_{nn}
\end{aligned}$$

o de forma resumida,

$$x_i^{(k+1)} = \Big(b_i - \sum_{j=1 j\neq i}^{n} a_{ij}x_j^{(k)}\Big)\Big/a_{ii} \qquad\qquad i = 1, \ldots, n.$$

En términos matriciales, si llamamos M a la matriz diagonal $M = \mathrm{diag}(a_{11}, \ldots, a_{nn})$, y $N = M - A$, de forma que $A = M - N$,

$$M = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \qquad \text{y} \qquad N = \begin{bmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & 0 & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & 0 \end{bmatrix}$$

tenemos que el sistema $Ax = b$ se escribe $(M - N)x = b$ de donde despejando $Mx = Nx + b$ y finalmente $x = M^{-1}Nx + M^{-1}b$, que es de la forma $x = Tx + c$ con $T = M^{-1}N$ y $c = M^{-1}b$.

2. Método de Gauss-Seidel

En el método de Jacobi, empleamos siempre los valores antiguos de las variables aunque ya hayamos calculado los nuevos valores de algunas de ellas. Parece lógico pensar que, si los valores nuevos se aproximan más a la solución, entonces se obtendrán mejores resultados usando ya dichos valores. En esto consiste el método de Gauss-Seidel, se sustituyen en la derecha los valores actualizados que ya tenemos en cada momento.

Ejemplo: El mismo que Jacobi.

Ejemplo: El divergente o el oscilante de antes.

Si se piensa bien lo que se está haciendo es lo siguiente. Se escribe el sistema $Ax = b$ en la forma

$$\begin{aligned}
&a_{11}x_1 && = b_1 - a_{12}x_2 - a_{13}x_3 - \cdots a_{1n}x_n \\
&a_{21}x_1 + a_{22}x_2 && = b_2 - a_{23}x_3 - \cdots a_{2n}x_n \\
& && \vdots \\
&a_{21}x_1 + a_{22}x_2 + \cdots + a_{nn}x_n && = b_n
\end{aligned}$$

se substituye el valor $x^{(k)}$ a la derecha y se resuelve el sistema anterior (por sustitución progresiva) obteniendo $x^{(k+1)}$. Por tanto,

$$\begin{aligned}
&a_{11}x_1^{(k+1)} && = b_1 - a_{12}x_2^{(k)} - a_{13}x_3^{(k)} - \cdots a_{1n}x_n^{(k)} \\
&a_{21}x_1^{(k+1)} + a_{22}x_2^{(k+1)} && = b_2 - a_{23}x_3^{(k)} - \cdots a_{2n}x_n^{(k)} \\
& && \vdots \\
&a_{21}x_1^{(k+1)} + a_{22}x_2^{(k+1)} + \cdots + a_{nn}x_n^{(k+1)} && = b_n.
\end{aligned}$$

De forma abreviada, se tiene pues

$$x_i^{(k+1)} = \Big(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)}\Big) \Big/ a_{ii} \qquad i = 1, \ldots, n.$$

De nuevo, nótese que, para poder obtener $x^{(k+1)}$ los elementos de la diagonal de la matriz A deben ser todos no nulos.

Para expresarlo matricialmente, llamemos M y N a las matrices

$$M = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,1} & a_{n-1,2} & a_{n-1,3} & \cdots & 0 \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$

y

$$N = \begin{bmatrix} 0 & -a_{12} & -a_{13} & \cdots & -a_{1n} \\ 0 & 0 & -a_{23} & \cdots & -a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -a_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

de manera que $A = M - N$ y el sistema se escribe $(M - N)x = b$. Igual que en el método de Jacobi, tenemos que $Mx = Nx + b$ y por tanto $x = M^{-1}Nx + M^{-1}b$, que es de la forma $x = Tx + c$ con $T = M^{-1}N$ y $c = M^{-1}b$.

3. Convergencia de un método iterativo

Hasta este momento, sólo hemos indicado dos formas de obtener las aproximaciones sucesivas $\{x^{(k)}\}$ de manera consistente, es decir, de manera que el sistema $x = Tx + c$ sea equivalente a $Ax = b$. (En general, para que sean sistemas equivalentes, debe existir una matriz regular Q tal que $Q(Ax - b) = x - Tx - c$, para todo x, de donde se obtiene que $T = I - QA$ y $c = Qb$. Otra forma es como se ha hecho en los métdos anteriores, se escribe $A = M - N$ con M invertible; de esta forma $x = M^{-1}N + M^{-1}b$). Ahora nos preocuparemos de encontrar condiciones para que la sucesión $\{x^{(k)}\}$ tenga límite, ya que en este caso, el límite $x = \lim_{k\to\infty} x^{(k)}$ cumple

$$x = \lim_{k\to\infty} x^{(k+1)} = \lim_{k\to\infty} Tx^{(k)} + c = T \lim_{k\to\infty} x^{(k)} + c = Tx + c,$$

es decir, x es la solución del sistema $x = Tx + c$ que equivale a $Ax = b$.

Consideremos pues una sucesión $\{x^{(k)}\}$ definida por recurrencia

$$\begin{cases} x^{(k+1)} = Tx^{(k)} + c \qquad k = 1, 2, \ldots. \\ x^{(0)} \text{ dado.} \end{cases}$$

Definición: Decimos que el método anterior es convergente si la sucesión $\{x^{(k)}\}$ converge, cualquiera que sea el valor inicial $x^{(0)}$.

Denotaremos por $d^{(k)}$ la diferencia entre la k-ésima aproximación $x^{(k)}$ y la solución del sistema x, es decir

$$d^{(k)} = x^{(k)} - x.$$

Es obvio que $x^{(k)} \to x$ es equivalente a $d^{(k)} \to 0$, de forma que un método es convergente si y sólo si $d^{(k)} \to 0$, independientemente del valor inicial $d^{(0)}$. Nótese que cada término de la sucesión $\{d^{(k)}\}$ se obtiene multiplicando el anterior por T, o lo que es lo mismo $d^{(k)} = T^k d^{(0)}$.

Definición: Decimos que una matriz T cuadrada $n \times n$ es contractiva si $||Tv|| < ||v||$ para todo vector $v \in \mathbb{R}^n$.

Se deduce que T es contractiva si y sólo si existe alguna constante L con $0 < L < 1$ tal que $||Tv|| \leq L||v||$.

Proposición: Una matriz T es contractiva si y sólo si todos sus valores propios son de módulo menor que la unidad.

Dem. Si T tiene un valor propio λ tal que $|\lambda| \geq 1$ entonces tomando v un vector propio asociado a λ, tenemos que $||Tv|| = ||\lambda v|| = |\lambda| ||v|| \geq ||v||$, y por tanto, T no es contractiva.

La demostración del recíproco es complicada, y, aunque el resultado es válido para cualquier matriz, la demostración sólo la haremos para una matriz simétrica T. En este caso T es diagonalizable con valores propios λ_i reales y podemos encontrar una base ortonormal $\{v_i\}$ de vectores propios. Si $v = \sum a_i v_i$, entonces $Tv = T(\sum a_i v_i) = \sum \lambda_i a_i v_i$, de donde, si los valores propios son de módulo menor que 1,

$$||Tv||^2 = \sum \lambda_i^2 a_i^2 < \sum a_i^2 = ||v||^2,$$

de donde $||Tv|| < ||v||$, es decir, T es contractiva. □

En realidad, de la demostración anterior se deduce que $||Tv|| \leq |\lambda_M| ||v||$, donde λ_M es el mayor de todos los autovalores de T. Aunque la matriz T no sea simétrica, el resultado anterior sigue siendo cierto, como se indica a continuación.

Al mayor de los módulos de los valores propios de T se le llama radio espectral de T

$$\rho(T) = \max \{ \, |\lambda| \mid \lambda \text{ valor propio de } T \, \} .$$

Entonces, se puede demostrar que el radio espectral satisface

$$||Tx|| \leq \rho(T) ||x||,$$

y es el menor número real que satisface dicha desigualdad.

Teorema: El método iterativo $x^{(k+1)} = Tx^{(k)} + c$ es convergente si y sólo si la matriz T es contractiva.

Dem. Si T es contractiva, como la sucesión $\{d^{(k)}\}$ satisface $d^{(k+1)} = Td^{(k)}$, dicha sucesión es de norma exponencialmente decreciente $||d^{(k+1)}|| \leq L^k ||d^{(0)}||$. En consecuencia su límite es cero, y el método es convergente.

Recíprocamente, si T no es contractiva, podemos encontrar un vector propio $v \neq 0$, asociado a un valor propio λ con $|\lambda| \geq 1$. Tomando como condición inicial $x^{(0)} = x + v$, de forma que $d^{(0)} = v$, tenemos que $d^{(k)} = \lambda^k v$. Así, $||d^{(k)}|| \geq ||v|| \neq 0$, de donde se deduce que la sucesión $\{d^{(k)}\}$ no puede converger a cero, y el método no es convergente. □

De esta manera, ya sabemos si un determinado método iterativo converge hacia la solución. Todo lo que nos falta es algún criterio que nos diga cuando podemos conformarnos con la aproximación $x^{(k)}$ de x. Fijamos ϵ y queremos que el error cometido sea de longitud menor que ϵ, es decir, queremos saber un valor de k tal que $||x^{(k)} - x|| < \epsilon$. Para ello utilizamos el siguientes resultado.

Proposición: Sea T una matriz contractiva y $L < 1$ un número real tal que $||Tx|| \leq L||x||$. Entonces,

$$||x^{(k)} - x|| \leq \frac{L}{1-L} ||x^{(k)} - x^{(k-1)}||.$$

Dem. En efecto, $d^{(k-1)} = (x^{(k-1)} - x^{(k)}) + d^{(k)}$ y por la desigualdad triangular, $||d^{(k-1)}|| \leq ||x^{(k)} - x^{(k-1)}|| + ||d^{(k)}||$. Además, $||d^{(k)}|| \leq L||d^{(k-1)}||$, por lo que

$$||d^{(k-1)}|| \leq ||x^{(k)} - x^{(k-1)}|| + ||d^{(k)}|| \leq ||d^{(k-1)}|| \leq ||x^{(k)} - x^{(k-1)}|| + L||d^{(k-1)}||.$$

Despejando $||d^{(k-1)}||$ de esta inecuación,

$$(1-L)||d^{(k-1)}|| \leq ||x^{(k)} - x^{(k-1)}||.$$

Finalmente, como $||d^{(k)}|| \leq L||d^{(k-1)}||$, y $L \leq 1$, llegamos a

$$||d^{(k)}|| \leq L||d^{(k-1)}|| \leq L\frac{1}{1-L}||x^{(k)} - x^{(k-1)}||,$$

como queríamos demostrar. □

Una consecuencia del resultado anterior es que

$$||x^{(k)} - x|| \leq \frac{L^k}{1-L}||x^{(1)} - x^{(0)}||,$$

que se prueba sin más que tener en cuenta que $||x^{(k+1)} - x^{(k)}|| \leq L||x^{(k)} - x^{(k-1)}||$.

Ejemplo: Uno que no provenga de Gauss-Seidl ni de Jacobi.

De todo lo anterior se deduce que cuanto más pequeño sea el radio espectral más rapidamente converge el método.

4. Convergencia de los métodos de Jacobi y Gauss-Seidel

Como acabamos de mostrar, para ver si un método iterativo es convergente, todo lo que hay que hacer es calcular los valores propios y asegurar que son menores que 1 en módulo. Veamos cómo calcular los valores propios en los casos de los métodos de Jacobi y Gauss-Seidel. En general, hemos escrito $A = M - N$, con M invertible, de donde el polinomio característico es

$$\kappa(x) = \det(xI - T) = det(xI - M^{-1}N) = \det(M)^{-1}\det(xM - N).$$

Por tanto, los valores propios son las raices del polinomio

$$p(x) = \det(xM - N).$$

En el caso de Jacobi, tenemos

$$p_J(x) = \begin{vmatrix} xa_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & xa_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & xa_{2n} \end{vmatrix}$$

mientras que en el caso de Gauss-Seidel,

$$p(x)_G = \begin{vmatrix} xa_{11} & a_{12} & \cdots & a_{1n} \\ xa_{21} & xa_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ xa_{n1} & xa_{n2} & \cdots & xa_{2n} \end{vmatrix}$$

Ejemplo: La convergente.

Definición: Una matriz A se dice diagonal dominante si

$$|a_{ii}| \geq \sum_{j=1 j\neq i}^{n} |a_{ij}|, \qquad \text{para } i = 1, \dots, n.$$

Se dice que A es estrictamente diagonal dominante si

$$|a_{ii}| > \sum_{j=1 j\neq i}^{n} |a_{ij}|, \qquad \text{para } i = 1, \dots, n.$$

Teorema: Si A es una matriz diagonal dominante entonces los métodos de Jacobi y de Gauss-Seidel son ambos convergentes. Además los radios espectrales de las matrices de dichos métodos satisfacen

$$\rho(T_J) < c \qquad \text{y} \qquad \rho(T_G) < c$$

donde $c = \max_i \{\sum_{j=1 j\neq i}^{n} |a_{ij}|/|a_{ii}|\}$.

Teorema: Si A es una matriz simétrica tridiagonal, entonces los métodos de Jacobi y de Gauss-Seidel son ambos convergente o ambos divergentes. Los radios espectrales de los métodos satisfacen $\rho(T_J) = \rho(T_G)^2$, por lo que, en caso de converger, el método de Gauss-Seidel lo hace más rápidamente.

Teorema: Sea A es una matriz simétrica con elementos diagonales positivos. Entonces el método de Gauss-Seidel es convergente si y sólo si la matriz A es definida positiva.