UN METODO DE ITERACION PARA LA RESOLUCION DE SISTEMAS DE ECUACIONES LINEALES ALGEBRAICAS

por PEDRO E. ZADUNAISKY, La Plata (Argentina)

1.-*Introducción.* Un sistema de ecuaciones lineales algebraicas puede representarse, usando la notación de las matrices, por la expresión

$$A\,x = k \tag{1}$$

en la que A es la matriz de los coeficientes de las incógnitas, x es la matriz de una sola columna cuyos elementos son las incógnitas $x_1, x_2 \dots x_n$ y k otra matriz también de una sola columna cuyos elementos son los segundos miembros $k_1, k_2, \dots, k_n$.

Los procedimientos usuales de iteración para resolver tales sistemas pueden resumirse así [1] (1). Se busca de establecer una relación de la forma

$$x = M\,x + N \tag{2}$$

que sea equivalente a (1) y en la que M y N son matrices de orden n convenientemente elegidas. Si se conoce una primera aproximación $x^{(0)}$ de la solución, se obtendrá la segunda sustituyendo en el segundo miembro de (2).

Siguiendo el proceso se obtendrá, después de ν iteraciones,

$$x^{(\nu)} = M\,x^{(\nu-1)} + N. \tag{3}$$

(1) Los números entre corchetes se refieren a la bibliografía citada al final.

Resulta entonces

$$x^{(\nu)} - x = M\,(x^{(\nu-1)} - x) \tag{4}$$

o bien

$$x^{(\nu)} - x = M^{\nu}\,(x^{(0)} - x). \tag{5}$$

Se puede entonces demostrar que si todos los autovalores de la matriz M son menores que la unidad el proceso es convergente (2).

La tarea de determinar, o por lo menos de acotar, los autovalores de una matriz es tan larga como la resolución misma del sistema (1) y por otra parte la rapidez de convergencia del proceso depende del acierto con que se hayan elegido las matrices M y N. Por estas razones los métodos de iteración tienen importancia práctica y son de uso corriente sólo en algunos casos ya clásicos. Tal, por ejemplo, el caso en que los elementos de la diagonal principal de la matriz de los coeficientes son mayores que los restantes y es aplicable entonces el método llamado de Gauss-Seidel [2]. En éste método el proceso de iteración se puede sintetizar en la fórmula siguiente:

$$x_i^{(\nu)} = \frac{k_i}{A_{ii}} - \left\{ \sum_{j=1}^{n-i} \frac{a_{ij}}{A_{ii}}\, x_j^{(\nu)} + \sum_{j=n-i}^{n} \frac{a_{ij}}{A_{ii}}\, x_j^{(\nu-1)} \right\} \quad \begin{matrix} (i = 1, 2, \ldots, n) \\ (j \neq i) \end{matrix} \tag{6}$$

donde hemos querido indicar con las mayúsculas A_{ii} los coeficientes predominantes, que forman la diagonal principal de la matriz del sistema. Se puede demostrar que el proceso converge si se verifican las condiciones

$$s_i = \sum_{j=1}^{n} \left| \frac{a_{ij}}{A_{ii}} \right| < 1. \tag{7}$$

(2) Recordamos que los autovalores de una matriz son las raíces de la ecuación de grado enésimo en λ.

$$\begin{vmatrix} m_{11}-\lambda & m_{12} & \ldots & m_{1n} \\ m_{21} & m_{22}-\lambda & \ldots & m_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ m_{n1} & m_{n2} & \ldots & m_{nn}-\lambda \end{vmatrix} = 0 .$$

Otro caso en que resulta practicable un método de iteración es aquel en que no solo predominan los elementos de la diagonal principal sino también los de las diagonales advacentes [3].

Un caso más general que éstos se nos presentó en un trabajo astronómico [4], pues el sistema a resolver tenía la forma siguiente

$$
\left.\begin{array}{l}
A_{11}\,x_1 + A_{12}\,x_2 + \dots + A_{1r}\,x_r + a_{1,r+1}\,x_{r+1} + \\
\qquad + a_{1,r+2}\,x_{r+2} + \dots + a_{1n}\,x_n = k_1, \\
A_{21}\,x_1 + A_{22}\,x_2 + \dots + A_{2r}\,x_r + a_{2,r+1}\,x_{r+1} + \\
\qquad + a_{2,r+2}\,x_{r+2} + \dots + a_{2n}\,x_n = k_2, \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
A_{r1}\,x_1 + A_{r2}\,x_2 + \dots + A_{rr}\,x_r + a_{r,r+1}\,x_{r+1} + \\
\qquad + a_{r,r+2}\,x_{r+2} + \dots + a_{rn}\,x_n = k_r, \\
a_{r+1,1}\,x_1 + a_{r+1,2}\,x_2 + \dots + a_{r+1,r}\,x_r + A_{r+1,r+1}\,x_{r+1} + \\
\qquad + A_{r+1,r+2}\,x_{r+2} + \dots + A_{r+1,n}\,x_n = k_{r+1}, \\
a_{r+2,1}\,x_1 + a_{r+2,2}\,x_2 + \dots + a_{r+2,r}\,x_r + A_{r+2,r+1}\,x_{r+1} + \\
\qquad + A_{r+2,r+2}\,x_{r+2} + \dots + A_{r+2,n}\,x_n = k_{r+2}, \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
a_{n1}\,x_1 + a_{n2}\,x_2 + \dots + a_{nr}\,x_r + A_{n,r+1}\,x_{r+1} + \\
\qquad + A_{n,r+2}\,x_{r+2} + \dots + A_{n,n}\,x_n = k_n,
\end{array}\right\} \quad (8)
$$

donde los coeficientes escritos con mayúsculas son los predominantes.

Queremos describir aquí el método de iteración que hemos aplicado para resolverlo y que involucra como caso particular el de Gauss-Seidel. Daremos también un criterio de convergencia de aplicación relativamente simple y que confiere a nuestro método una generalidad muy amplia ([3]).

([3]) En R. v. MISES, H. POLLACZEK-GEIRINGER, *Praktische Verfahren der Gleichungsauflösung*. Z. angew. Math. Mech. 9 (1929) S. 76, se encuentra un método análogo que los autores denominan "iteración en grupos" aunque el procedimiento de cálculo difiere del nuestro y no se da un criterio general de convergencia como lo hacemos aquí.

2. — El método consiste, en primer lugar, en separar las incógnitas en dos grupos: $x_j = \{x_1, x_2, \ldots, x_r\}$ y $x_k = \{x_{r+1}, x_{r+2}, \ldots, x_n\}$. Por otra parte, con los símbolos $A_{ij}, A_{ik}, a_{ik}, a_{lj}$ indicaremos matrices cuadradas de orden n cuyos elementos son todos nulos excepto los que tengan $\{i \text{ o } j\} \leq r$ y $\{k \text{ o } l\} > r$. Los símbolos x_j, k_i y x_k, k_l también indicarán matrices cuadradas de orden n cuyos elementos son todos nulos excepto los r primeros o los $n-r$ últimos de la primera columna respectivamente.

Entonces el sistema (8) se puede descomponer en las dos ecuaciones siguientes:

$$A_{ij}\, x_j = k_i - a_{ik}\, x_k \tag{9}$$

$$A_{lk}\, x_k = k_l - a_{lj}\, x_j \tag{10}$$

de donde

$$x_j = A_{ij}^{-1} \{k_i - a_{ik}\, x_k\} \tag{11}$$

$$x_k = A_{lk}^{-1} \{k_l - a_{lj}\, x_j\} \tag{12}$$

siendo A_{ij}^{-1} y A_{lk}^{-1} las matrices recíprocas de A_{ij} y A_{lk} respectivamente. Ahora, si indicamos con $x^{(\nu)}$ la aproximación ν-ésima, el proceso de iteración se resume en las siguientes fórmulas, a aplicar en forma sucesiva:

$$x_j^{(\nu)} = A_{ij}^{-1} \{k_i - a_{ik}\, x_k^{(\nu-1)}\} \tag{13}$$

$$x_k^{(\nu)} = A_{lk}^{-1} \{k_l - a_{lj}\, x_j^{(\nu)}\}. \tag{14}$$

Si el proceso es convergente, el cálculo continuará hasta que la diferencia entre dos aproximaciones sucesivas sea menor que el error admisible.

Para que el método sea practicable el proceso debe pues converger y hacerlo rápidamente en modo que el número de iteraciones necesarias no sea excesivo. Deduciremos por tanto un criterio que permite establecer estas circunstancias.

Restando (11) de (13) se obtiene

$$x_j^{(\nu)} - x_j = A_{ij}^{-1} a_{ik} \left(x_k - x_k^{(\nu-1)}\right);$$

y poniendo

$$A_{ij}^{-1}\, a_{ik} = - X_{jk}, \tag{15}$$

cuyo significado aclararemos más adelante, resulta

$$x_j^{(\nu)} - x_j = X_{jk}\,(x_k - x_k^{(\nu-1)}).$$

Entonces para el primer grupo de incógnitas se obtiene explícitamente

$$x_j^{(\nu)} - x_j = \sum_{k=r+1}^{n} X_{jk}\,(x_k^{(\nu-1)} - x_k)\,;$$

y en forma análoga para el segundo grupo

$$x_k^{(\nu)} - x_k = \sum_{j=1}^{r} X_{kj}\,(x_j^{(\nu)} - x_j),$$

con

$$A_{lk}^{-1}\, a_{lj} = - X_{kj}. \tag{16}$$

Si ponemos ahora

$$s_k = \sum_{j=1}^{r} X_{jk}, \quad s_j = \sum_{k=r+1}^{n} X_{kj} \tag{17}$$

resulta

$$\sum_{j=1}^{r} (x_j^{(\nu)} - x_j) = \sum_{k=r+1}^{n} s_k\,(x_k^{(\nu-1)} - x_k)$$

$$\sum_{k=r+1}^{n} (x_k^{(\nu)} - x_k) = \sum_{j=1}^{r} s_j\,(x_j^{(\nu)} - x_j)$$

e indicando con S_j la mayor de las s_j y con S_k la mayor de las s_k resulta, en valor absoluto,

$$\sum_{j=1}^{r} |x_j^{(\nu)} - x_j| \leqq |S_k| \sum_{k=r+1}^{n} |x_k^{(\nu-1)} - x_k|$$

luego

$$\sum_{k=r+1}^{n} |x_k^{(\nu)} - x_k| \leqq |S_j| \sum_{j=1}^{r} |x_j^{(\nu)} - x_j| \leqq |S_j|\ |S_k| \sum_{k=r+1}^{n} |x_k^{(\nu-1)} - x_k|\,;$$

y en forma análoga podríamos deducir

$$\sum_{j=1}^{r} |x_j^{(\nu)} - x_j| \leqq |S_k|\ |S_j| \sum_{j=1}^{r} |x_j^{(\nu-1)} - x_j|\,.$$

Ahora, si $x^{(0)}$ es la primera aproximación y ponemos $\sigma = |S_j|\,.\,|S_k|$ resulta evidentemente, al cabo de ν iteraciones

$$\sum_{j=1}^{r} |x_j^{(\nu)} - x_j| \leqq \sigma^{\nu} \sum_{j=1}^{r} |x_j^{(\nu-1)} - x_j|\,.$$

o bien

$$\sum_{k=r+1}^{n} |x_k^{(\nu)} - x_k| \leqq \sigma^{\nu} \sum_{k=r+1}^{n} |x_k^{(\nu-1)} - x_k|\,.$$

Por tanto el proceso será convergente si $\sigma < 1$ y lo será tanto más rápido cuanto más pequeño sea σ.

La aplicación de este criterio de convergencia es menos complicada de lo que a primera vista parece si se atiende al significado de σ que se obtiene de las expresiones (15), (16) y (17). En efecto, s_k es la suma de los valores que tomarían las incógnitas x_j si el segundo miembro de (9) se redujera a la k-ésima columna de la matriz a_{ik} y s_j es la suma de los valores que tomarían las incógnitas x_k si el segundo miembro de (10) se redujera a la j-ésima columna de la matriz a_{ij}. Entonces σ es el producto del valor absoluto de la mayor de las s_j por la mayor de las s_k.

En nuestro estudio hemos descompuesto el sistema de ecuaciones en dos grupos cuyos respectivos determinantes $|A_{ij}|$ y $|A_{lk}|$ son menores principales del determinante total [4]. Si el caso lo requiere, se puede hacer la descomposición en un número cualquiera p de menores principales y análogamente puede demostrarse que una condición práctica, aunque no rigurosa de convergencia será

$$\sigma = |S_1| \, . \, |S_2| \, . \, |S_3| \ldots |S_p| < 1. \tag{18}$$

Nada obsta para que algunos de aquellos grupos, en que se ha subdividido el sistema total, consten de una sola ecuación; en el caso particular de que todos los grupos consten de una sola ecuación estaremos en el proceso de Gauss-Seidel descripto en el primer párrafo. En este caso las expresiones (17) toman la forma

$$S_i = \sum_{j=1}^{n} - \frac{a_{ij}}{A_{ii}} \, (i = 1, 2, \ldots, n; \, i \neq j)$$

y la condición de convergencia sería

$$\sigma = |S_1| \, |S_2| \ldots |S_n| < 1$$

menos restringida que la expresada por (7).

3. - *Resumen de fórmulas.*

El sistema (8) a resolver se puede escribir en la forma, equivalente a (9) y (10),

$$\sum_{j=1}^{r} A_{ij} \, x_j = K_i, \quad (i = 1, 2, \ldots, r)$$

$$\sum_{k=r+1}^{n} A_{lk} \, x_k = K_l, \quad (l = r+1, r+2, \ldots, n)$$

(4) Se llama principal a todo menor cuya diagonal principal forma parte de la diagonal principal del determinante total.

habiendo puesto

$$K_i = k_i - \sum_{k=r+1}^{n} a_{ik}\, x_k$$

$$K_l = k_l - \sum_{j=1}^{r} a_{lj}\, x_j.$$

Ahora indicamos con α_{ij} el adjunto del elemento A_{ij} en el determinante $|A_{ij}|$ dividido por dicho determinante y análogo significado atribuimos a α_{lk}. Entonces aplicando la regla de Cramer tendremos

$$x_j = \sum_{i=1}^{r} \alpha_{ij}\, K_i$$

$$x_k = \sum_{l=r+1}^{n} \alpha_{lk}\, K_l.$$

Luego las fórmulas a aplicar en orden sucesivo para el proceso de iteración son las siguientes:

$$K_i^{(\nu-1)} = k_i - \sum_{k=r+1}^{n} a_{ik}\, x_k^{(\nu-1)}$$

$$x_j^{(\nu-1)} = \sum_{i=1}^{r} \alpha_{ij}\, K_i^{(\nu-1)}$$

$$K_l^{(\nu)} = k_l = \sum_{j=1}^{r} a_{lj}\, x_j^{(\nu-1)}$$

$$x_k^{(\nu)} = \sum_{l=r+1}^{n} \alpha_{lk}\, K_l^{(\nu)}.$$

Y para la aplicación del criterio de convergencia serán

$$s_k = \sum_{j=1}^{r} X_{jk} = \sum_{j=1}^{r} \sum_{i=1}^{r} \alpha_{ij}\, a_{ik}$$

$$s_j = \sum_{k=r+1}^{n} X_{kj} = \sum_{k=r+1}^{n} \sum_{l=r+1}^{n} \alpha_{lk}\, a_{lj}.$$

Los adjuntos α_{ij} y α_{lk} se calculan de una vez por todas y sirven en primer lugar para aplicar el criterio de convergencia y luego para realizar las sucesivas iteraciones en las que sólo van variando las $K_i^{(v)}$ y $K_l^{(v)}$.

La aplicación de la regla de Cramer es especialmente ventajosa en sistemas cuyo determinante es simétrico y esta ventaja se conserva en nuestro método ya que los menores principales son también simétricos.

En general la eficacia del método dependerá del acierto con que se haya hecho la subdivisión de las incógnitas en grupos. Pero por otra parte su campo de aplicación es amplio ya que la condición de convergencia (18) queda asegurada con que una sola de las S sea suficientemente pequeña. En el caso estudiado, por ejemplo, basta con que una sola de las matrices a_{ik} o a_{lj} tenga sus elementos suficientemente pequeños aunque los de la otra sean grandes.

BIBLIOGRAFIA

[1] W. E. Milne, *Numerical Solution of Differential Equations* (Wiley, 1953); cap. 9.

[2] R. Zurmühl, *Matrizen* (Springer, 1950); cap. VI.

[3] Fr. A. Willers, *Practical Analysis* (Dover Publications, 1947); cap. 4.

[4] P. E. Zadunaisky, *A Determination of new Elements of the Orbit of Phoebe, ninth satellite of Saturn* (Astr. Journal, January, 1954).

OBSERVATORIO ASTRONÓMICO DE LA
UNIVERSIDAD NAC. DE LA PLATA.