ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIV, 1 (1974)

S. LEWANOWICZ (Wrocław)

ALGORITHM 29

SOLVING SYSTEMS OF LINEAR EQUATIONS BY SOKOLOV'S METHOD

1. Procedure declaration. Procedure *sleS2* solves by Sokolov's iterative method (method of averaged functional corrections) [1] the system of linear equations

$$Ax = b, \tag{1}$$

where A is a square matrix of order n, and $x,\ b \in \mathbf{R}^n$.

Data:

n — order of the system,
$a[1:n,\ 1:n]$ — array of coefficients of the matrix A,
$b[1:n]$ — vector b of free terms,
$x[1:n]$ — initial solution x_0 which may be arbitrary,
$fi1,\ fi2[1:n]$ — orthogonal vectors φ_1 and φ_2 which have to be chosen so that the necessary condition of the convergence of the method (see section 2) be satisfied,
eps — positive number ε characterizing the relative error of the solution; the calculations are finished if, for the approximations $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ and $x_{i-1} = (x_{i-1,1}, x_{i-1,2k}, \ldots, x_{i-1,n})$ of the solution of (1), there holds

$$\max_{1\leqslant k\leqslant n} |x_{ik} - x_{i-1,k}|/|x_{ik}| < \varepsilon,$$

maxit — maximum number of iterations to be performed during the calculations.

Results:

$x[1:n]$ — solution of (1),
maxit — number of performed iterations.

Other parameters:

ns — label (outside of the body of procedure *sleS2*) to which a jump is made when after *maxit* iterations the required accuracy has not been obtained.

Remark. The elements on the main diagonal of A must be different from zero.

```
procedure sleS2(n,a,b,x,fi1,fi2,eps,maxit,ns);
 value n,eps;
 integer n,maxit;
 real eps;
 array a,b,x,fi1,fi2;
 label ns;
 begin
  integer i,j,m;
  real aij,a1,a2,c1i,c2i,d,s11,s12,s21,s22,t1,t2,yi;
  Boolean conv;
  array c1,c2,y[1:n];
  s11:=s12:=s21:=s22:=.0;
  for i:=1 step 1 until n do
   begin
    c1i:=c2i:=.0;
    for j:=i+1 step 1 until n do
     begin
      aij:=a[i,j];
      c1i:=c1i-aij×fi1[j];
      c2i:=c2i-aij×fi2[j]
     end j;
    for j:=i-1 step -1 until 1 do
     begin
      aij:=a[i,j];
      c1i:=c1i-aij×c1[j];
      c2i:=c2i-aij×c2[j]
     end j;
    d:=1.0/a[i,i];
    c1i:=c1[i]:=c1i×d;
```

```
   c2i:=c2[i]:=c2i×d;
   yi:=fi1[i];
   s11:=s11+yi×(yi-c1i);
   s12:=s12-yi×c2i;
   yi:=fi2[i];
   s21:=s21-yi×c1i;
   s22:=s22+yi×(yi-c2i);
   y[i]:=x[i]
  end i;
 d:=1.0/(s11×s22-s12×s21);
 s11:=s11×d;
 s12:=s12×d;
 s21:=s21×d;
 s22:=s22×d;
 for m:=1 step 1 until maxit do
  begin
   t1:=t2:=.0;
   for i:=1 step 1 until n do
    begin
     yi:=b[i];
     for j:=i-1 step -1 until 1,i+1 step 1 until n do
      yi:=yi-a[i,j]×y[j];
     yi:=y[i]:=yi/a[i,i];
     yi:=yi-x[i];
     t1:=t1+yi×fi1[i];
     t2:=t2+yi×fi2[i]
    end i;
   a1:=s22×t1-s12×t2;
   a2:=s11×t2-s21×t1;
```

```
                conv:=true;
                for i:=1 step 1 until n do
                 begin
                  yi:=y[i]:=y[i]+a1×c1[i]+a2×c2[i];
                  if conv
                   then conv:=abs((x[i]-yi)/yi) < eps;
                  x[i]:=yi
                 end i;
                if conv
                 then
                 begin
                  maxit:=m;
                  go to exit
                 end conv
               end m;
              go to ns;
            exit:end sleS2
```

2. Method used. Let us represent the matrix A in the form

$$A = L+D+U, \tag{2}$$

where L and U are the lower and upper triangular matrices, respectively, with zeros on the main diagonal, and D is the diagonal matrix. Let the vectors $\varphi_1, \varphi_2, \ldots, \varphi_p$ $(p \leqslant n)$, $\varphi_j \in \boldsymbol{R}^n$, form an orthogonal system. We construct the sequence of vectors $x_0, x_1, \ldots, x_m, \ldots$ $(x_m \in \boldsymbol{R}^n)$ with x_0 given and

$$(L+D)x_m = b - U(x_{m-1}+a_m), \tag{3}$$

where

$$a_m = \Phi_p \delta_m, \tag{4}$$

$$\Phi_p = \sum_{j=1}^{p} \gamma_j^{-1} \varphi_j \varphi_j', \quad \gamma_j = \|\varphi_j\|^2 \overset{\mathrm{df}}{=} \varphi_j' \varphi_j, \qquad \delta_m = x_m - x_{m-1}.$$

The symbol $'$ denotes transposition. Calculate now x_m from (3). We have

$$x_m = Nb - NU\psi_p x_{m-1},$$

where $N = (L+D+U\Phi_p)^{-1}$, $\psi_p = E-\Phi_p$, and E denotes the unit matrix of order n. For method (3)-(4) to be convergent, it is necessary and sufficient that all eigenvalues of $NU\psi_p$ be absolutely less than one. It is easy to verify that

THEOREM. *If the sufficient condition for the convergence of Seidel's method is satisfied, i. e. if*

$$\|M\| < 1, \tag{5}$$

where $M = -(L+D)^{-1}U$, *and* $\|\ \|$ *denotes the spectral or Euclidean norm of a square matrix, then method* (3)-(4) *is convergent, and*

$$\|x^* - x_m\| \leqslant \frac{\|M\|}{1-\|M\|}\|\psi_p\,\delta_m\| \tag{6}$$

holds, where x^* *is the exact solution of* (1).

In fact, from (3) we have

$$(L+D)x_m = b - Ux_m + U\psi_p\,\delta_m.$$

Substracting sidewise this equation from

$$(L+D)x^* = b - Ux^*,$$

we obtain, after simple derivations,

$$x^* - x_m = (E-M)^{-1}M\psi_p\,\delta_m.$$

Hence and from (5), it follows (6).

Now, we shall prove that $\|\psi_p\,\delta_m\| \to 0$ for $m \to \infty$. Equation (3) can be written as

$$(L+D)x_m = b - U(\Phi_p x_m + \psi_p x_{m-1}).$$

Hence

$$\delta_m = M(\Phi_p\,\delta_m + \psi_p\,\delta_{m-1}). \tag{7}$$

Since $\Phi_p' = \Phi_p$ and $\Phi_p^2 = \Phi_p$, we have $w_1'\Phi_p'\psi_p w_2 = 0$ for every $w_1, w_2 \in \boldsymbol{R}^n$. Thus, $\|\delta_m\|^2 = \|\Phi_p\,\delta_m\|^2 + \|\psi_p\,\delta_m\|^2$ and, from (7), we obtain

$$\|\Phi_p\,\delta_m\|^2 + \|\psi_p\,\delta_m\|^2 \leqslant \|M\|^2(\|\Phi_p\,\delta_m\|^2 + \|\psi_p\,\delta_{m-1}\|^2)$$

which can be written as follows:

$$\|\psi_p\,\delta_m\|^2 \leqslant (\|M\|^2-1)\|\Phi_p\,\delta_m\|^2 + \|M\|^2\|\psi_p\,\delta_{m-1}\|^2.$$

By the assumption $\|M\| < 1$, i. e. $\|M\|^2-1 < 0$, we have

$$\|\psi_p\,\delta_m\| < \|M\|\cdot\|\psi_p\,\delta_{m-1}\|.$$

Therefore, $\|\psi_p\,\delta_m\| \to 0$ for $m \to \infty$. From (6) it follows also that $\|x^* - x_m\| \to 0$.

It has been noticed in [1] that the above-mentioned theorem can be obtained as a corollary from the theorem stating the sufficient condition of the convergence of Sokolov's method applied to $x = f+\lambda Ax$, where A denotes the linear operator in the Hilbert space H, $f, x \in H$, and λ is a complex number; our proof is based on the fundamental idea included in the proof of the theorem given in [1].

If $p = 0$, formula (3) gives Seidel's method, and formula (6) is the known inequality of Collatz.

For numerical purposes the new versions of formulae (3) and (4) are convenient:

$$x_m = s_m + \sum_{j=1}^{p} \beta_{mj} c_j, \tag{8}$$

where

$$s_m = D^{-1}(b - Ls_m - Ux_{m-1}),$$

$$\gamma_j \beta_{mj} - \sum_{i=1}^{p} \varphi_j' c_i \beta_{mi} = \varphi_j'(s_m - x_{m-1}), \qquad c_j = -D^{-1}(Lc_j + U\varphi_j) \tag{9}$$

$$(j = 1, 2, \ldots, p).$$

Procedure *sleS2* is based on (8) and (9) for $p = 2$. In this case, it is necessary to perform $n^2+4(n+1)$ multiplications and divisions in every iteration step, and, moreover, $n(2n+5)+7$ multiplications and divisions have to be performed in the initial step.

3. Certification. Procedure *sleS2* has been tested on the Odra 1204 computer for Pei matrices

$$\begin{bmatrix} d & 1 & \ldots & 1 \\ 1 & d & \ldots & 1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \ldots & d \end{bmatrix}$$

with solution vector $x = (1, 2, \ldots, n)'$ from which the vector b could be calculated. On entry, it was assumed $x[i] = 0$ $(i = 1, 2, \ldots, n)$, $fi1[i] = 1$, $fi2[i] = 0$ $(i = 1, 2, \ldots, [n/2])$, $fi1[i] = 0$, $fi2[i] = 1$ $(i = [n/2]+1, [n/2]+2, \ldots, n)$, and $eps = {}_{10}{-7}$.

The following maximum absolute errors of the solution components have been obtained:

n \ d	3	2	1.5	1.25
10		$5.62_{10}-8$	$3.76_{10}-7$	$4.66_{10}-7$
20	$1.06_{10}-7$	$2.38_{10}-7$	$5.41_{10}-7$	

The method was compared with that of Seidel. The iteration numbers I, needed to obtain the required accuracy, and the calculation times t in secs. were as follows:

d	n	*sleS2*		Seidel	
		I	t	I	t
3	20	29	38.0	99	105.2
2	10	26	10.2	75	20.9
2	20	58	74.1	229	243.4
1.5	10	43	16.6	154	42.8
1.5	20	124	156.7	300	319.2
1.25	10	84	31.9	315	87.5

For $d = 1.5$ and $n = 20$, the calculations for the Seidel method were stopped after the 300-th iteration; the maximum absolute error of the solution components was then equal to $1.13_{10}-3$.

4. Example of an application of Sokolov's method to the solution of boundary problems. Consider the differential problem:

$$\Delta u \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y) \quad (a < x < b,\ c < y < d),$$

$$(10) \qquad u(a, y) = g_1(y), \quad \left.\frac{\partial u}{\partial x}\right|_{x=b} = g_2(y),$$

$$u(x, c) = g_3(x), \quad \left.\frac{\partial u}{\partial y}\right|_{y=d} = g_4(x).$$

Denoting by u_{ik} the approximate value of u in the point $x_i = a+ih$, $y_k = c+kh_1$ ($i = 0, 1, \ldots, m+1$; $k = 0, 1, \ldots, n+1$; $h = (b-a)/(m+0.5)$, $h_1 = (d-c)/(n+0.5)$) and substituting in (10) the usual differential approximation, one gets the difference boundary problem

$$u_{i+1,k} + u_{i-1,k} + \gamma^2(u_{i,k+1} + u_{i,k-1}) - 2(1+\gamma^2)u_{ik} = h^2 f_{ik},$$

$$(11) \qquad u_{0k} = g_{1k}, \quad u_{m+1,k} - u_{mk} = g_{2k},$$

$$u_{i0} = g_{3i}, \quad u_{i,n+1} - u_{in} = g_{4i}$$

$$(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n),$$

where f_{ik}, g_{1k}, g_{2k}, g_{3i} and g_{4i} are given and $\gamma = h/h_1$. The eigenfunctions of (11) for $f_{ik} \equiv 0$ and homogeneous boundary conditions attain the form (see [2])

$$v_{ik}^{(q,r)} = 4\big((2m+1)(2n+1)\big)^{-1/2} \sin i\,\frac{2q-1}{2m+1}\, \sin k\,\frac{2r-1}{2n+1}$$

$$(i, q = 1, 2, \ldots, m;\ k, r = 1, 2, \ldots, n).$$

These functions form an orthonormal system.

System (11) has been solved by Sokolov's method ($p = 2$), assuming for φ_1 and φ_2 the functions $v^{(1,1)}$ and $v^{(2,1)}$, respectively, and by Seidel's method. For problem (10) with known exact solution

$$u = \cosh \pi y \sin \pi x / \cosh 0.475\pi \qquad (0 \leqslant x \leqslant 1,\ |y| \leqslant 0.475),$$

the results for $\varepsilon = 5_{10}-5$ are as follows (I denotes the number of performed iterations, t the calculation time in secs., and Δ the maximum absolute error of the solution):

m	n	Sokolov			Seidel		
		I	t	Δ	I	t	Δ
9	9	27	48.0	$1.09_{10}-2$	190	187.5	$9.21_{10}-3$
19	18	101	753.1	$3.83_{10}-3$	614	2571.5	$4.86_{10}-3$

It is worth noticing that the speed of convergence in Sokolov's method depends upon the choice of φ_1 and φ_2. For the last example, taking eigenfunctions, different from the above ones, leads, for $m = n = 9$, to an iteration number of at least 48.

References

[1] A. Yu. Lučka (А. Ю. Лучка), *Теория и применение метода осреднения функциональных поправок*, Киев 1963.

[2] I. N. Lyašenko (И. Н. Ляшенко), *Задачи на собственные значения для уравнений второго порядка в частных конечных разностях*, Киев 1970.

MATHEMATICAL INSTITUTE
UNIVERSITY OF WROCŁAW
50-384 WROCŁAW

Received on 19. 2. 1973

ALGORYTM 29

S. LEWANOWICZ (Wrocław)

ROZWIĄZYWANIE UKŁADU ALGEBRAICZNYCH RÓWNAŃ LINIOWYCH METODĄ SOKOŁOWA

STRESZCZENIE

Procedura *sleS2* rozwiązuje iteracyjną metodą Sokołowa (metodą uśredniania poprawek funkcjonalnych) [1] układ n równań liniowych $Ax = b$, gdzie A jest macierzą kwadratową n-tego stopnia, złożoną ze współczynników układu, natomiast $x, b \in \boldsymbol{R}^n$.

Dane:

n — stopień układu,

a[1 : n, 1 : n] — tablica elementów macierzy A,

b, x, fi1, fi2 [1 : n] — tablice składowych wektora b, początkowego przybliżenia rozwiązania x_0 oraz ortogonalnych wektorów φ_1 i φ_2; x_0 jest dowolne, φ_1 i φ_2 zaś muszą być tak wybrane, by spełniony był warunek konieczny zbieżności metody (zobacz § 2), jeśli spełnione jest założenie twierdzenia z § 2 (w szczególności, jeśli macierz A jest dodatnio określona), to φ_1 i φ_2 są w zasadzie dowolne (byle ortogonalne),

eps — liczba dodatnia ε, charakteryzująca błąd względny rozwiązania; obliczenia kończy się, gdy dla przybliżeń $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ oraz $x_{i-1} = (x_{i-1,1}, x_{i-1,2}, \ldots, x_{i-1,n})$ rozwiązania układu zachodzi nierówność

$$\max_{1 \leqslant k \leqslant n} |x_{ik} - x_{i-1,k}|/|x_{ik}| < \varepsilon,$$

maxit — zmienna, której wartością jest liczba ograniczająca liczbę wykonywanych iteracji.

Wyniki:

x[1 : n] — rozwiązanie układu,

maxit — liczba wykonanych iteracji.

Inne parametry:

ns — etykieta instrukcji (poza treścią procedury *sleS2*), do której następuje skok, gdy po wykonaniu *maxit* iteracji nie otrzymano rozwiązania z żądaną dokładnością.

Uwaga. Elementy występujące na przekątnej głównej macierzy A muszą być różne od zera.