MATHEMATICS OF COMPUTATION, VOLUME 27, NUMBER 124, OCTOBER 1973

An Extrapolated Gauss-Seidel Iteration for Hessenberg Matrices

By L. J. Lardy

Abstract. We show that for certain systems of linear equations with coefficient matrices of Hessenberg form it is possible to use Gaussian elimination to obtain an extrapolated version of the Gauss-Seidel iterative process where the iteration matrix has spectral radius zero. Computational aspects of the procedure are discussed.

1. Introduction. Let $A = (a_{ij})$ be a nonsingular $n \times n$ matrix. We write A in the form $A = D - E - F$ where $D = (d_{ij})$ is the diagonal matrix with $d_{ij} = a_{ij}$ for $i = j$ and $d_{ij} = 0$ for $i \neq j$, $E = (e_{ij})$ is the strictly lower triangular matrix with $e_{ij} = -a_{ij}$ for $i < j$ and $e_{ij} = 0$ for $i \geqq j$, and $F = (f_{ij})$ is the strictly upper triangular matrix with $f_{ij} = -a_{ij}$ for $i > j$ and $f_{ij} = 0$ for $i \leqq j$. Let C be a given $n \times 1$ matrix and consider the matrix equation

$$AX = C. \tag{1}$$

We examine here an iterative method for obtaining a numerical solution of such an equation when the coefficient matrix A has Hessenberg form. The procedure can be viewed as a generalization of the well-known extrapolated Gauss-Seidel or SOR method [3]. In that method, the iteration can be expressed in the form

$$X^{(r)} = M(\omega)X^{(r-1)} + K \tag{2}$$

where

$$M(\omega) = (D - \omega E)^{-1}((1 - \omega)D + \omega F), \tag{3}$$

$$K = (D - \omega E)^{-1}C, \tag{4}$$

and ω is a real nonzero relaxation parameter. The objective is then to choose a value for ω which minimizes the spectral radius of the matrices $M(\omega)$. In general, this optimal value of ω is not readily obtainable. The method discussed here involves an iteration of the form (2) with $M(\omega)$ expressed formally as in (3), except that the relaxation parameter ω is taken to be a fixed n-component real vector with no zero component, where, for the purposes of matrix manipulation, ω operates as a diagonal matrix. When A has Hessenberg form, an optimal relaxation vector yields a corresponding $M(\omega)$ which has spectral radius zero. Indeed, we shall show that in this case $M(\omega)$ is strictly upper triangular.

If D is nonsingular, then, for any vector ω, $(D - \omega E)$ is also nonsingular and $M(\omega)$ can be defined by (3). Multiplying both sides of the equation

Received June 12, 1972.

AMS (MOS) subject classifications. (1970) Primary 65F10.

Key words and phrases. Gaussian elimination, Gauss-Seidel iteration, Hessenberg matrices.

$$X = M(\omega)X + K \tag{5}$$

by $(D - \omega E)$ yields

$$\omega AX = \omega C. \tag{6}$$

Hence, if ω has no zero component, Eq. (6) is equivalent to Eq. (1). Thus, if the iteration defined by (2) yields a convergent sequence, then its limit is a solution of (1).

2. Analysis of the Method. Let the equation (1) be such that the nonsingular matrix A has Hessenberg form; that is, $a_{ij} = 0$ for $i \geqq j + 2$. Suppose that the associated diagonal matrix D is nonsingular and that Gaussian elimination can be applied in natural order to produce the matrices $A^{(1)} = A, \cdots, A^{(n)}$. Here and below we follow the notation in [2, p. 30]. Set

$$\omega_i = a_{ii}/a_{ii}^{(i)} \quad \text{for} \quad i = 1, \ldots, n. \tag{7}$$

The iteration (2) with this vector $\omega = (\omega_i)$ is then equivalent to the following iteration:

$$x_1^{(r)} = \left(c_1 - \sum_{j=2}^{n} a_{1j}x_j^{(r-1)}\right)\Big/ a_{11}, \tag{8a}$$

$$x_i^{(r)} = x_i^{(r-1)} + \left(c_i - \left(a_{i,i-1}x_{i-1}^{(r)} + \sum_{j=i}^{n} a_{ij}x_j^{(r-1)}\right)\right)\Big/ a_{ii}^{(i)} \quad \text{for } i = 2, \ldots, n. \tag{8b}$$

The following two lemmas are used to establish the form of the iteration matrix $M(\omega)$. Lemma 1 can be easily verified directly.

LEMMA 1. *Let $G = (g_{ij})$ be an $n \times n$ matrix with $g_{ii} \neq 0$ and $g_{ij} = 0$ when either $i < j$ or $i > j + 1$. Then $G^{-1} = (h_{ij})$ where*

$$\begin{aligned} h_{ij} &= 1/g_{ii} && \text{for} \quad i = j, \\ &= 0 && \text{for} \quad i < j, \\ &= (-1)^{i+j} \frac{g_{j+1,j} \cdots g_{i,i-1}}{g_{j,j} \cdots g_{i,i}} && \text{for} \quad i > j. \end{aligned} \tag{9}$$

LEMMA 2. *Let the matrix A satisfy the conditions given at the beginning of this section. For $2 \leqq i \leqq n$ and $1 \leqq j \leqq i - 1$, we have*

$$\sum_{k=j}^{i-1} (-1)^{k+i+1} \frac{a_{k+1,k}^{(j)} \cdots a_{i,i-1}^{(j)}}{a_{k,k}^{(k)} \cdots a_{i,i}^{(i)}} a_{k,i}^{(j)} = \frac{a_{ii}}{a_{ii}^{(i)}} - 1. \tag{10}$$

Proof. We fix i and proceed by induction. For $j = i - 1$ the summation in the left side of (10) reduces to a single term. Apply the definition of $a_{ii}^{(i)}$ (see [2, p. 30]), and use the fact that due to the Hessenberg form of A,

$$a_{pq}^{(j-1)} = a_{pq}^{(j)} \quad \text{for} \quad p > j \tag{11}$$

to observe that Eq. (10) holds for $j = i - 1$.

Suppose that (10) holds for a certain j, denote the sum on the left side of Eq. (10) by $S_i(j)$, and consider $S_i(j - 1)$. In $S_i(j - 1)$ separate the two terms obtained for $k = j - 1$ and $k = j$. Observe that the sum of these terms is a multiple of $a_{ji}^{(j)}$. Now apply (11) to see that the entire sum $S_i(j - 1)$ is precisely the same as $S_i(j)$. An application of the induction hypothesis now completes the proof.

THEOREM. *Let the matrix A satisfy the conditions given at the beginning of this section. Let ω be the vector determined by* (7) *and let $M(\omega)$ denote the iteration matrix defined as in* (3). *Then $M(\omega)$ has strictly upper triangular form. Thus $M(\omega)$ has spectral radius zero and $M(\omega)^n = 0$.*

Proof. Since $\omega_1 = 1$, the first column of the matrix $(1 - \omega)D + \omega F$ vanishes and it follows from the definition of $M(\omega)$ that its first column also vanishes. We show next that the diagonal entries of $M(\omega)$ vanish. Using the definitions of ω and $M(\omega)$ and then Lemma 1 to determine $(D - \omega E)^{-1}$, we find that for $i = 2, \dots, n$ the ith diagonal element of $M(\omega)$ can be written in the form

$$\begin{aligned} m_{ii} &= \sum_{k=1}^{i-1} (-1)^{i+k} \frac{\omega_{k+1}a_{k+1,k} \cdots \omega_i a_{i,i-1}}{a_{k,k} \cdots a_{i,i}} (-\omega_k a_{k,i}) + \frac{(1 - \omega_i)a_{ii}}{a_{ii}} \\ &= S_i(1) + (1 - \omega_i) \\ &= 0, \end{aligned}$$

where $S_i(1)$ denotes the sum appearing on the left in Eq. (10) and we used (7) and also Lemma 2.

If $i > j > 1$, then the element in the ith row and the jth column can be expressed as

$$\begin{aligned} m_{ij} &= \sum_{k=1}^{j-1} (-1)^{i+k} \frac{\omega_{k+1}a_{k+1,k} \cdots \omega_i a_{i,i-1}}{a_{k,k} \cdots a_{i,i}} (-\omega_k a_{k,j}) \\ &\quad + (1 - \omega_j)a_{jj}(-1)^{i+j} \frac{\omega_{j+1} \cdots \omega_i}{a_{jj} \cdots a_{ii}} \\ &= \frac{\omega_{j+1}a_{j+1,j} \cdots \omega_i a_{i,i-1}(-1)^{i+j}}{a_{jj} \cdots a_{ii}} S_j(1) + (1 - \omega_j) \\ &= 0, \end{aligned}$$

where again $S_j(1)$ denotes a sum as in (10) and we used (7) together with Lemma 2.

It is possible to view the above method as a special case of a more general class of iterative methods where the iteration matrix has spectral radius equal to zero. Suppose that A is a nonsingular matrix which, we shall assume, can be reduced to the upper triangular matrix $A^{(n)}$ by elementary row operations without the use of row interchanges. From the triangular reduction procedure, we obtain in the usual way a matrix T such that $TA = A^{(n)}$. If D_n is the diagonal matrix formed using the diagonal elements of $A^{(n)}$, then the equation

$$X = (I - D_n^{-1}TA)X + D_n^{-1}TC$$

is equivalent to the equation $AX = C$. Clearly, $I - D_n^{-1}TA$ is a strictly upper triangular matrix. Now, if A is a Hessenberg matrix, one can use Lemma 1 and induction on the rows of T to show that $(I - ED_n^{-1})^{-1} = T$, and then it is easy to deduce that $M(\omega) = I - D_n^{-1}TA$. Thus, the iteration (8) can be viewed as Gauss-Seidel iteration applied to the equation $A^{(n)}X = TC$.

3. Remarks on Numerical Aspects of the Method. To begin with, we comment on the number of operations required by the method to obtain the solution. It follows from the strictly upper triangular form of $M(\omega)$ that theoretically $x_n^{(1)}$ is exact, $x_{n-1}^{(2)}$ is exact, and in general $x_{n-r}^{(r+1)}$ is exact. Thus, the complete cycle indicated in (8) is not necessary and we might modify the method by successively terminating the calculations one step earlier. However, it is not difficult to convince oneself that even with this modification the number of multiplications required to obtain $x_n^{(1)}, \ldots, x_1^{(n)}$ after the relaxation vector has been calculated is not competetive with back substitution. Since it is necessary to reduce the matrix to triangular form in order to determine ω, it is then more economical to proceed with back substitution to obtain the solution.

Regarding accuracy, the method appears to be quite comparable with Gaussian elimination. Results from a limited number of experiments using the method for ill-conditioned systems suggest that with respect to both efficiency and accuracy it is preferable to use the partial step modification of the algorithm indicated above rather than the total step version. The computation of $x_n^{(1)}, x_{n-1}^{(2)}, \cdots, x_1^{(n)}$ will be called one cycle of the partial step method. Either version of the algorithm would appear to be natural for refining the solution obtained from back substitution. Here we have observed that for ill-conditioned matrices there is some improvement in the accuracy of the solution in the earlier iterates, but in order to produce substantial improvement it has been necessary to calculate the vector of iterates in double precision. Wilkinson [4] points out that for Hessenberg matrices, Gaussian elimination followed by back substitution is quite accurate, so that with single precision a substantial improvement in accuracy would probably not be expected. However, the total step version of the algorithm does not always sustain an improvement of accuracy throughout the later iterations as indicated by the computational results presented in Table II.

The computations presented below were performed on an IBM 370 computer using single precision unless otherwise indicated. The results in Table I and Table II were obtained using systems of the form

$$\begin{bmatrix} n & n-1 & \cdots & 2 & 1 \\ n-1 & n-1 & \cdots & 2 & 1 \\ 0 & n-2 & \cdots & 2 & 1 \\ \vdots & & & & \\ 0 & 0 & \cdots & 1 & 1 \end{bmatrix} X = \begin{bmatrix} n(n+1)/2 \\ (n-1)(n+2)/2 \\ (n-2)(n+1)/2 \\ \vdots \\ 2 \end{bmatrix}. \tag{12}$$

Thus all components of the exact solution are one.

In Table I we take $n = 6$ and use the total step version of the algorithm. The results in the upper rows are the computed iterates obtained with an initial vector zero. The results in the lower rows are the computed iterates obtained using the solution computed by back substitution as the initial vector.

When the total step version of the iterative scheme was applied to the system (12) with $n = 8$ and the solution computed by back substitution as initial approximation, the first iteration improved the accuracy but then the accuracy declined noticeably. The partial step version performed somewhat better. The first column of Table II contains the solution computed by back substitution, while the first and eighth iterates computed using this initial approximation and the total step version are given in the second and third columns respectively. One cycle of the partial step version was applied with the solution from back substitution as initial approximation and these results are given in column four of Table II.

TABLE I

$X^{(0)}$	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$	$X^{(4)}$	$X^{(5)}$	$X^{(6)}$
0	3.500000	−0.2333313	3.499901	0.0002154	1.166312	1.000377
0	2.999997	−0.9999781	1.999908	0.8001690	0.9997560	1.000212
0	2.500006	0.0000353	1.249947	1.000145	0.9997767	1.000022
0	1.999944	0.6666622	0.9998132	1.000061	1.000032	1.000144
0	1.500175	1.000123	1.000224	1.000131	0.9992709	0.9992094
0	0.9996924	1.000060	0.9994917	1.000247	1.001211	1.000370
0.9999879	0.9999898	1.000094	0.9995711	1.000547	0.9996770	1.000383
1.000074	1.000063	1.000176	0.9996626	1.000335	0.9999480	1.000410
0.9996946	0.9997684	0.9999889	0.9995621	0.9999018	0.9998327	1.000334
1.000913	1.000625	1.000609	1.000047	1.000251	1.000045	1.000539
0.9981725	0.9990394	0.9999549	0.9999534	0.9991996	0.9992082	0.9998519
1.001828	1.000093	0.9999977	1.000094	1.001506	1.000077	1.000218

TABLE II

$X^{(0)}$	$X^{(1)}$	$X^{(8)}$	$\hat{X}^{(8)}$
0.9999982	0.9999982	0.9746984	0.9999981
0.9999872	0.9999861	0.9869528	1.000000
1.000081	1.000072	0.9943312	1.000053
0.9995853	0.9996368	0.9976511	0.9997644
1.001658	1.001409	0.9983439	0.9977849
0.9950189	0.9960259	0.9976821	1.003881
1.009962	1.006941	1.001415	1.000993
0.9900382	0.9960770	0.9978753	0.9960770

A class of ill-conditioned tridiagonal matrices has been introduced by Dorr in [1]. We obtain an example from this class by taking $n = 20$, $\epsilon = .01$, and $h = .05$. Then for $1 \leqq i \leqq 10$, let $a_i = -\epsilon/h^2$ and $c_i = -\epsilon/h^2 + i - 10$, while for $11 \leqq i \leqq 20$, let $a_i = -\epsilon/h^2 + i - 10$ and $c_i = -\epsilon/h^2$. The tridiagonal coefficient matrix A is then obtained by defining the nonzero elements as follows:

$$\begin{aligned} a_{ii} &= -(a_i + c_i), && 1 \leqq i \leqq 20, \\ a_{i,i+1} &= c_i, && 1 \leqq i < 20, \\ a_{i+1,i} &= a_{i+1}, && 1 \leqq i < 20. \end{aligned}$$

We define the column vector C' so that again all components of the exact solution of $AX = C'$ are one. Thus, $c_1' = -a_1$, $c_i' = 0$, $1 < i < 20$, and $c_{20}' = -c_{20}$. With this system, Gaussian elimination, the calculation of the relaxation parameters, and back substitution were performed in single precision and the resulting solution was accurate to three digits in all components with a typical component being $x_{10} = 0.9995323$. Using this approximate solution as an initial vector, two cycles of the partial step algorithm were computed using a double-precision iteration vector. After the first cycle, each component had the value 0.9999998 and after the second cycle each component had the value 0.9999999.

Department of Mathematics
Syracuse University
Syracuse, New York 13210

1. F. W. Dorr, "An example of ill-conditioning in the numerical solution of singular perturbation problems," *Math. Comp.*, v. 25, 1971, pp. 271–283.

2. E. Isaacson & H. B. Keller, *Analysis of Numerical Methods*, Wiley, New York, 1966. MR **34** #924.

3. R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962. MR **28** #1725.

4. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965. MR **32** #1894.