A PROCESS FOR THE SELECTION OF OVER-RELAXATION FACTOR ω AND THE RELATED COMPUTATION FOR $A\vec{x}=\vec{b}$

By Manas Chanda* and Syamal Kumar Sen**

[Received : October 13, 1970]

Abstract

A procedure is presented for choosing over-relaxation factor ω when we do not possess any knowledge of it. This relaxation factor however does not represent the conventional scalar matrix in the matrix notation. Instead, it represents a diagonal matrix having the form diag (ω 1 1 . . . 1). A linear system $A\vec{x}=\vec{b}$ *is considered for illustration. A few examples worked out by one-step cyclic process (Gauss-Seidel method) and also by over-relaxation method (with the over-relaxation factor obtained in the aforesaid manner) indicate a much more rapid convergence characteristic of the latter process.*

1. Introductory Mathematics

Let the system of equations for the vector $\vec{x}\,(=x_1,\ x_2,\ .\ .\ .\ x_n)$, be given by

$$x_j=\phi_j\,(x_1,\ x_2,\ .\ .\ .\ x_n),\quad j=1,\ 2,\ .\ .\ .\ n. \qquad [1]$$

(′) indicates transpose.

One-step cyclic process (Gauss-Seidel[1,2] method) can be activated starting with an initial approximation $\vec{x}^{(0)}$ as follows:

$$x_j^{(k+1)}=\phi_j\,(x_1^{(k+1)},\ x_2^{(k+1)},\ .\ .\ .\ ,\ x_{j-1}^{(k+1)},\ x_j^{(k)},\ .\ .\ .\ x_n^{(k)}) \qquad [2]$$

$$j=1,\ 2,\ .\ .\ .\ ,\ n$$

$$k=0,\ 1,\ 2,\ .\ .\ .$$

$k+1$ indicates the iteration number.

* Department of Chemical Engineering, Indian Institute of Science, Bangalore-12, India.

** Central Instruments and Services Laboratory, Indian Institute of Science, Bangalore-12, India.

We consider as a special case, a linear system. We write, in matrix notation, the linear system as

$$A\vec{x}=\vec{b}$$

We put

$$A=A_L+A_R+D$$

where

$$A_L=(l_{ij})\ ; \qquad l_{ij}=\begin{cases} a_{ij} & i>j \\ 0 & i\leqslant j \end{cases}$$

$$A_R=(r_{ij})\ ; \qquad r_{ij}=\begin{cases} 0 & i\geqslant j \\ a_{ij} & i<j \end{cases}$$

$$D=(d_{ij})\ ; \qquad d_{ij}=\begin{cases} 0 & i>j \\ a_{ii} & i=j \\ 0 & i<j \end{cases}$$

The one-step cyclic process then becomes

$$(A_L+D)\,\vec{x}^{(k+1)}+A_R\,\vec{x}^{(k)}=\vec{b} \qquad [3]$$

$$k=0,\ 1,\ 2,\ \ldots$$

We write Eqn. [3] as

$$D\vec{x}^{(k+1)}=\vec{b}-A_L\vec{x}^{(k+1)}-A_R\vec{x}^{(k)}$$

Subtracting $D\vec{x}^{(k)}$ from both sides, we obtain the relation

$$D\vec{\delta}^{(k)}=\vec{d} \qquad [4]$$

where $\vec{\delta}^{(k)}\ (=\vec{x}^{(k+1)}-\vec{x}^{(k)})$ is called the differences and

$\vec{d}\,(=\vec{b}-A_L\vec{x}^{(k+1)}-(A_R+D)\,\vec{x}^{(k)}\,)$ is called the defect or residual.

Introducing a factor

$$W=\begin{bmatrix} \omega & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

where ω is called the relaxation factor, we write Eqn. [4] as

$$D\vec{\delta}^{(k)} = W\vec{d}, \quad 0 \leq \omega \leq 2 \qquad [5]$$

For $\omega = 1$, it is identical to Gauss-Seidel method. ω normally depends on k.

2. Choice of Over Relaxation Factor

We may not have pre-knowledge of ω. In such a case we take some value of ω (say $\omega_1 = 1.2$) satisfying the condition $0 \leq \omega \leq 2$ and obtain the corresponding defect or residual $\vec{d}_1$. Let us increase ω a little (say $\omega_2 = 1.5$), keeping it below 2, however, and see the residual $\vec{d}_2$. We then compare the two residuals $\vec{d}_1$ and $\vec{d}_2$ and determine a better ω value for which the residual becomes zero by linear interpolation. ω_{refined} is thus given by

$$\omega_{\text{refined}} = \omega_1 - \frac{\omega_1 - \omega_2}{q\,(|\vec{d}_1| - |\vec{d}_2|)} \cdot |\vec{d}_1| \qquad [6]$$

where $|\vec{d}_1|$ and $|\vec{d}_2|$ are Euclidean norms of $\vec{d}_1$ and $\vec{d}_2$, respectively, and q is a numerical factor $\geqq 1$. q depends on k. The actual role of q is to boost up the value of $(|\vec{d}_1| - |\vec{d}_2|)$ so that ω_{refined} does not go very much out of the closed interval [0,2] or it remains within [0,2]. We may take $q = 1$, 10, 10^2, 10^3 etc., so that the magnitude of $q\,(|\vec{d}_1| - |\vec{d}_2|)$ is of the same order as that of $(\omega_1 - \omega_2) \cdot |\vec{d}_1|$ or any other more suitable value that we may guess for q. If

$$|A_L \vec{x}^{(k+1)} + (A_R + D)\,\vec{x}^{(k)}| > |\vec{b}|$$

$|\vec{d}|$ has to be negatived. ω_{refined} mainly provides us an idea about the direction in which it moves. Quadratic or cubic interpolation is not of much use because besides more calculations they cannot produce any thing better for a new ω than the linear interpolation does. In fact, we obtain an idea of the direction in which ω moves through interpolation and not the actual ω. We should notice that ω_{refined} in [6] may produce a value either greater than 2 or less than 0 for a moderate initial approximation. We, however, put the new value of ω as 2 if $\omega_{\text{refined}} > 2$. If $\omega_{\text{refined}} < 0$, we put new $\omega = 0$. If ω_{refined} remains within 0 and 2, the new $\omega = \omega_{\text{refined}}$. The first iteration ends with obtaining the solution vector $\vec{x}^{(1)}$ from relation [5] with the newly found out ω. Identical is the situation for the second iteration, in which case we may take ω_1 as 1.2 and ω_2 as 1.5, or any two different values between 0 and 2 with the latest approximate vector $\vec{x}^{(1)}$.

3. CONVERSION OF THE SYSTEM $\vec{A}\vec{x}=\vec{b}$ TO $B\vec{x}=\vec{c}$

Keeping the solution vector $\vec{x}$ invariant we transform the matrix A and the column vector $\vec{b}$ to a matrix B and a column vector $\vec{c}$ such that the new system $B\vec{x}=\vec{c}$ produces a convergent sequence of the vectors $\vec{x}^{(k)}$ for any initial approximation $\vec{x}^{(0)}$. The following conversion normally achieves the aforesaid convergence.

We obtain Erhard-Schmidt's norm of A denoted by $\| A \|_{E.S}$, and given by

$$\left[\sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij})^2 \right]^{1/2} .$$

Let p be an arbitrary non-zero positive number. Then

$$B=\frac{A}{\| A \|_{E.S.}+p}$$

and

$$\vec{c}=\frac{\vec{b}}{\| A \|_{E.S.}+p}$$

It is easy to see that $\| B \|_{E.S.} < 1$. The optimum value of ω, *i.e.*,

$$\omega_b^{[8,9]}=1+\left\{\frac{\rho(B)}{1+[1-\rho^2(B)]^{\frac{1}{2}}}\right\}^2.$$

(where $\rho(B)$ is the spectral radius of B) for the maximum rate of convergence that depends on the spectral radius of the coefficient matrix B is tedious to obtain. On the contrary, new ω is easier to obtain. The actual difference between ω_b and ω_1 or ω_2 lies in the fact that ω_b is the actual over-relaxation factor for an iteration, whereas ω_1 and ω_2 usually give the direction for the actual over-relaxation factor.

For any initial vector $\vec{x}^{(0)}$ the new system $B\vec{x}=\vec{c}$ normally converges. The above convergence criterion is the result of the following theorem:

Theorem: Given any matrix norm $\|A\|$ which is consistent with a vector norm. The condition $\| A \| < 1$ is sufficient that for any initial vector $\vec{x}^{(0)}$, the vector $\vec{x}^{(k)}=A^k\vec{x}^{(0)}$ tends to a null vector, *i.e.*, A^k tends to a null matrix as k tends to ∞.

Since Erhard-Schmidt's norm is consistent with a vector norm (though not subordinate) we expect convergence mathematically for the new system $B\vec{x}=\vec{c}$.

We should not take p too large, since in that case the square-rooting operations for obtaining $|\vec{d}_1|$ and $|\vec{d}_2|$ will incur considerable amount of error. Moreover, the addition and subtraction error that depends entirely on the precision of the computer can mar the calculation. We may take p around 5.

4. Numerical Example

To illustrate the procedure for the choice of ω we work out below a few examples.

Let a linear system $A\vec{x}=\vec{b}$ be given as

$$\begin{bmatrix} 2 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The actual solution vector $\vec{x}$ is

$$\vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

We take $\|A\|_{E.S.}+p$ as 10, *i.e.*, p around 5. Then the new system $\mathbf{B}\vec{x}=\vec{c}$ becomes

$$\begin{pmatrix} .2 & .1 \\ .3 & .3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} .1 \\ 0 \end{pmatrix}$$

Hence

$$B_L = \begin{pmatrix} 0 & \\ .3 & 0 \end{pmatrix}, \; B_R = \begin{pmatrix} 0 & .1 \\ & 0 \end{pmatrix}, \; D_B = \begin{pmatrix} .2 & \\ & .3 \end{pmatrix}, \; \vec{c} = \begin{pmatrix} .1 \\ 0 \end{pmatrix}$$

The over-relaxation procedure is

$$D_B(\vec{x}^{(k+1)} - \vec{x}^{(k)}) = W(\vec{c} - B_L\vec{x}^{(k+1)} - (B_R + D_B)\vec{x}^{(k)}), \quad k=0,1,2, \ldots \qquad [7]$$

We take

$$\vec{x}^{(0)} = \begin{pmatrix} 2 \\ -2 \end{pmatrix} \text{ and } \omega_1 = 1.2$$

$k=0$ (*i.e., 1st iteration*)

From relation [7] we write

$$\begin{pmatrix} .2 & \\ & .3 \end{pmatrix} \begin{pmatrix} x_1^{(1)} & -2 \\ x_2^{(1)} & +2 \end{pmatrix} = \begin{pmatrix} 1.2 & 0 \\ 0 & 1 \end{pmatrix} \left\{ \begin{pmatrix} .1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 & \\ .3 & 0 \end{pmatrix} \begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} - \begin{pmatrix} .2 & .1 \\ & .3 \end{pmatrix} \begin{pmatrix} 2 \\ -2 \end{pmatrix} \right\}$$

$$= \begin{pmatrix} 1.2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -.1 \\ -.3x_1^{(1)} + .6 \end{pmatrix}.$$

Hence $x_1^{(1)} = 1.4$

We now take $\omega_2 = 1.5$ and consequently $x_1^{(1)} = 1.25$

Therefore

$$|\vec{d}_1| = (.01 + .0324)^{1/2} = .206$$

$$|\vec{d}_2| = (.01 + .050625)^{1/2} = .246$$

and

$$\omega_{\text{refined}} = \omega_1 + \frac{\omega_1 - \omega_2}{q(|\vec{d}_1| - |\vec{d}_2|)} \cdot |\vec{d}_1|$$

$$= 1.2 + \frac{1.2 - 1.5}{q(.206 - .246)} \times .206$$

$$= 2.745 \text{ taking } q = 1.$$

We take new ω (which is our actual relaxation factor) as 2.

Hence

$$x_1^{(1)} = 1\,; \quad x_2^{(1)} = -1$$

Thus the first iteration is over and the result obtained is exact. The one-step cyclic process with the same initial approximation

$$\vec{x}^{(0)} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

produce the following result as the 12th iteration

$$\vec{x}^{(12)} = \begin{bmatrix} 1.000244140625 \\ -1.000244140625 \end{bmatrix}$$

The result is correct just up to 3 decimal places. The above procedure achieved the exact result only in one iteration. Let us now consider a slight deviation of the actual solution vector for the aforesaid system changing the known column vector a little. We write the new system as

$$\begin{bmatrix} .2 & .1 \\ .3 & .3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} .09 \\ -.03 \end{bmatrix}$$

The Erhard-Schmidt's norm of the coefficient matrix is already less than 1. It is therefore not necessary to transform the matrix as before. The actual solution vector is

$$\vec{x} = \begin{bmatrix} 1 \\ -1.1 \end{bmatrix}$$

It we take an initial approximation

$$\vec{x}^{(0)} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

the over-relaxation method with the choice of ω based on $q=1$, $\omega_1=1.2$ and $\omega_2=1.5$ produces the exact result just in two iterations. The Gauss-Seidel procedure, on the other hand, requires 12 iterations with the identical initial approximation of $\vec{x}^{(0)}$ only to produce result correct upto 3 decimal places. The 12th iterated vector is

$$\vec{x}^{(12)} = \begin{bmatrix} 1.0002197265625 \\ -1.1002197265625 \end{bmatrix}$$

When we consider the first problem (that possesses positive definite coefficient matrix) with $\vec{x}^{(0)}=(2\ 2)'$ and $\omega_1=1.1$, $\omega_2=1.11$, we obtain ω_{refined} as $-.28$, taking $q=1$. We thus take new ω as 0. This new ω produces

$$x_1^{(1)}=2,\ x_2^{(1)}=-2$$

We note that sign reversal has taken place for the second component of the solution vector. The next iteration itself then produces the exact solution.

If the initial vector $\vec{x}^{(0)}$ is such that ω_{refined} always tends to be less than 0, or else, if at certain iteration the vector $\vec{x}^{(k)}$ is such that ω_{refined} is always less thad 0, the above procedure always takes new ω as 0 and normally produces necessary sign changes in the components of the trial solution vector. The last example illustrates this fact. When the initial approximation differs from the actual solution vector considerably, the ω_{refined} may become greater or much greater than 2. In such cases, the new ω that takes a value 2 $(\omega \not> 2)$ swings into action and very rapidly brings the approximate vector down to near the solution vector $\vec{x}$. The first two examples illustrate this fact. If ω_{refined} is found to have a value between 0 and 2, it is itself taken as the new ω. This is illustrated in the following example.

The system

$$\begin{bmatrix} 5 & 3 & 1 \\ -1 & 2 & 1 \\ 4 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -1 \\ 9 \end{bmatrix}$$

has the solution vector $\vec{x}=(1\quad 1\quad -2)'$

In order to apply the over-relaxation method, we first divide the elements of the coefficient matrix as also those of the b vector to form a new system in which the Erhard-Schmidt norm of the coefficient matrix is less than 1. Now, taking the initial approximation

$$x^{(0)} = (5 \quad 5 \quad -1)', \; q = 5.818, \; \omega_1 = 1.2 \text{ and } \omega_2 = 1.5$$

we obtain ω_{refined} as .753. Since it is within 0 and 2, we take new ω as .753 and calculate $\vec{x}^{(1)}$ at the first iteration, which is $(0.9518 \;\; .4759 \;\; -2.35845)'$. Using $\omega_1 = 1.2$ and $\omega_2 = 1.5$ in the second iteration also we obtain ω_{refined} as a negative quantity, taking $q = 1$. We therefor take $\omega = 0$ and obtain $\vec{x}^{(2)}$ as $(.9518 \;\; 1.155125 \;\; -2.0188375)'$. With the aforesaid initial approximation, ω_1 and ω_2, we obtain ω_{refined} as .94 with $q = 10$, and consequently, $\vec{x}^{(1)}$ becomes $(1.064 \;\; .532 \;\; -2.106)'$. In the second iteration, we obtain ω_{refined} as

$$1.2 - \frac{.077835}{q \times .02723}$$

which is negative if we allow q to be 1. We thus obtain $\vec{x}^{(2)} = (1.064 \;\; 1.085 \;\; -1.8295)'$ permitting new ω to be zero.

5. DISCUSSION

In commonly used forms[3, 6, 9, 10] of the over-relaxation formula, the relaxation factor is a scalar matrix. But in the method presented in this paper the relaxation factor is a diagonal matrix having the form diag $(\omega \;\; 1 \;\; 1 \; . \; . \; . \; 1)$. Consequently, the factor ω whose effect is injected in the first component x_1 of the solution vector, affects, in turn, the rest of the components $x_2, \; x_3, \; . \; . \; . \; , \; x_n$ of the vector $\vec{x}$ linearly. But in the commonly used methods referred to above, non-linear terms involving ω^2, ω^3, etc , creep into the components x_2, x_3, etc. of the solution vector $\vec{x}$. The relative effect of these two aspects on the rate of convergence remains to be explored.

The cited theorem defines only the sufficient condition for convergence for the system $A\vec{x} = \vec{b}$. It is, however, relevant to mention the necessary and sufficient condition proposed by Berry[11], *i e.*, $\lim_{m \to \infty} [(A_L + D)^{-1} A_R]^m = \Theta$ (Θ is the null matrix) for a general matrix A, which incidentally implies the following theorem[12]:

For any given square matrix A, the powers $A^p \to \Theta$ if and only if all eigenvalues λ_j of A have moduli that are less than 1.

6 Acknowledgment

The authors wish to express their gratitude to Dr. S. Dhawan, Director, for his constant encouragement and to Dr. E. V. Krishnamurthy for suggesting some valuable references related to this work The authors are also grateful to the referee for his kind suggestion of the paper of Clifford E Berry.

References

1 Young. D. .. .. *Trans. Am. math. Soc*, 1954, 76, 92.

2 Seidel, L. .. .. Abhandlungen der Bayerischen Akademie, 1873 **11**, (3), 81.

3 Southwell, R., ... " Relaxation Methods in Theoretical Physics" Oxford University Press, 1946.

4. ——— .. .. Proceedings of a Symposium on the construction and Application of Conformal Maps. National Bureau of Standards, Applied Mathematics Series, 1949, **18**, 239.

5. Stein, P and Rosenberg, R. .. *J. Lond. math Soc.*, 1948, **23**, 111.

6. Richardson, L F. .. .. *Phil Trans. R Soc*, *London*, 1910, **210A**.

7. Abarbandal, S. and G Zwas .. *Maths. Comput.*, 1969, **23**, (107), 549.

8 Smith, G. D. .. . " Numerical Solution of Partial Differential Equations ", Oxford University Press, 1965, 79.

9. Varga, R. S. . .. Iterative Solution of Matrix Equations, " The University of Michigan Engg. Summer Conferences (Numerical Analysis) " 1964.

10. Krishnamurthy, E. V. .. *Int. J. control*, 1965, **1** (5), 461.

11. Berry, C. E. .. .. *Annals. of Math. Statistics*, 1945, **16**, 398.

12. Collatz, L. . .. " Functional Analysis and Numerical Mathematics ", Academic Press, 1966, 183.