measure of the success of any article is provided not only by the number of important problems that it solves but also by the number of new questions that it opens for investigation. On the basis of both these criteria we must judge the present article to be a solid success.

REFERENCES

EUBANK, R. and GUNST, R. (1986). Diagnostics for penalized least-squares estimators. *Statist. Probab. Lett.* **4** 265–272.

EUBANK, R. and SPECKMAN, P. (1988). Curve fitting by polynomial-trigonometric regression. Unpublished manuscript.

NYCHKA, D. (1988). Bayesian confidence intervals for smoothing splines. *J. Amer. Statist. Assoc.* **83** 1134–1143.

RICE, J. (1986). Convergence rates for partially splined models. *Statist. Probab. Lett.* **4** 203–208.

SPECKMAN, P. (1988). Regression analysis for partially linear models. *J. Roy. Statist. Soc. Ser. B* **50** 413–436.

STONE, C. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.

WAHBA, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B* **45** 133–150.

DEPARTMENT OF STATISTICS
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

DEPARTMENT OF STATISTICS
UNIVERSITY OF MISSOURI
COLUMBIA, MISSOURI 65211

WALTER GANDER AND GENE H. GOLUB

Eidgenössische Technische Hochschule Zürich and Stanford University

The solution of linear algebraic equations arises in many situations in statistical computing. Most often the matrices are symmetric and positive definite and they may have some structure that can be taken advantage of; viz., Toeplitz matrices arise in time series and special algorithms are available for such problems (cf. [3]). It is unusual for matrices to be structured and nonsymmetric but this is the situation that arises in the paper by Buja, Hastie and Tibshirani. In addition, the system (19) the authors describe is singular though the nullspace can be determined without difficulty.

Very often for large structured systems, iterative methods are used. (We set aside the fact that $\hat{P}$ is singular at this time.) Thus one might split $\hat{P}$ and write

$$\hat{P} = M - N$$

and iterate as follows:

Given f_0
For $k = 0, 1, \ldots,$

$$Mf^{k+1} = Nf^k + \hat{Q}y \quad (\text{solve for } f^{k+1}).$$

It is important that solving the system

$$Mf^{k+1} = z^k \quad (\text{say})$$

be "easy," and, of course, one desires that

$$\|M^{-1}N\| < 1$$

for some $\|\cdots\|$. If the spectral radius of $\|M^{-1}N\|$ is close to 1, then the method will converge slowly and a method of acceleration will be required. Numerical analysts have studied such techniques extensively, mainly in connection with the numerical solution of partial differential equations. Note that assuming that

$$\begin{aligned} Mf &= Nf + \hat{Q}y, \\ f^k &= f + (M^{-1}N)^k(f^0 - f). \end{aligned} \tag{1}$$

We will see that (1) is the basis for accelerating the convergence of the process.

The choice of M obviously affects the convergence properties of the algorithm. Let

$$\hat{P} = D + L + U,$$

where D consists of the diagonal elements of $\hat{P}$; L is a lower triangular matrix (zeros on the diagonal and above the diagonal) and U is an upper triangular matrix. (The matrices D, L and U may consist of block elements.) Here are some splittings that have been extensively studied:

1. $M = D$, Jacobi method;
2. $M = D + L$, Gauss–Seidel method;
3. $M = (D + \omega L)$, successive over-relaxation for $\omega > 1$ (successive under-relaxation for $\omega < 1$).

The conditions under which each of these methods converge have been extensively studied and are fairly technical. One condition that is easy to verify for methods 1 and 2 is that the matrix be diagonally dominant (cf. [4]).

Successive over-relaxation has been thoroughly analyzed for symmetric positive-definite matrices which have property A: that is, any matrix which can be permuted to the form (cf. [6])

$$\hat{P} = \begin{pmatrix} I & S \\ S^T & I \end{pmatrix}. \tag{2}$$

The optimal choice of ω is given by

$$\hat{\omega} = \frac{2}{1 + \sqrt{1 - \|S\|_2^2}}.$$

It is easy to see that $1 < \hat{\omega} < 2$; the method converges, however, for all $0 < \omega < 2$.

The effect of this choice of ω can be very dramatic. If r is the number of iterations required for a certain accuracy when the Gauss–Seidel method is used, then for $\omega = \hat{\omega}$ the number of iterations would be roughly const. $\sqrt{r}$.

Now, we wish to study the situation when $p = 2$ so that

$$\hat{P} = \begin{pmatrix} I & S_1 \\ S_2 & I \end{pmatrix}$$

with $\|S_1\|_2 < 1$ and $\|S_2\|_2 < 1$. This situation has been recently studied in [2], and conditions are given for choosing the optimal ω. The choice of optimal parameter will depend upon eigenvalues of S_1S_2. For $p > 2$, we know of no classical theory for choosing the optimal parameter ω.

The situation which arises in the paper under review is that the matrix $\hat{P}$ is singular. That complication can be most easily handled by forcing the iterates to lie in the kernel of $\hat{P}$. Some adjustment must be made to the relaxation parameter when $p \doteq 2$ but that does not present any great technical difficulties [5]. A device which is often used for eliminating vectors in the nullspace is to take differences between successive iterates. For best numerical results, however, it is probably best to project successive iterates in the kernel of $\hat{P}$.

As we indicated above,

$$f^k = f + (M^{-1}N)^k(f^0 - f).$$

Let us assume that $M^{-1}N$ is diagonalizable, and we write

$$M^{-1}N = Q\Lambda Q^{-1}$$

where Λ is the diagonal matrix of eigenvalues and Q is the matrix of eigenvectors. Hence

$$f^k = f + \sum_{i=1}^{l} d_i\lambda_i^k q_i \equiv f + \sum_{i=1}^{l} \lambda_i^k w_i, \tag{3}$$

where $\{q_i\}$ are the eigenvectors and

$$d \equiv Q^{-1}(f^0 - f).$$

A particularly attractive algorithm for accelerating the convergence of the sequence f^k is the ε-algorithm of Wynn [1]. From successive iterates f_k one constructs the table

$$\begin{array}{llllll}
f_0 = \varepsilon_{00}, & & & & & \\
f_1 = \varepsilon_{10} & \varepsilon_{11}, & & & & \\
f_2 = \varepsilon_{20} & \varepsilon_{21} & \varepsilon_{22}, & & & \\
f_3 = \varepsilon_{30} & \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33}, & & \\
f_4 = \varepsilon_{40} & \varepsilon_{41} & \varepsilon_{42} & \varepsilon_{43} & \varepsilon_{44} & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{array}$$

using the relation

$$\varepsilon_{ij} = \varepsilon_{i-1,\,j-2} + (\varepsilon_{i,\,j-1} - \varepsilon_{i-1,\,j-1})^+,$$

where the pseudoinverse of a vector v is defined as

$$v^+ := \frac{1}{\|v\|} v^T.$$

Only the columns with even index in the ε table are meaningful and converge to the limit. The error expansion of every of these columns contains one term less of the sum (3) so that in theory then, from the $2l$ vectors $f^1, \ldots, f^{2l}$ and f^0, one is

able to determine precisely f. (When $l = 1$, the method is essentially the Aitken acceleration procedure.) Of course, one stops far short of $2l$ iterations to obtain a good approximation to f.

Therefore the extrapolation procedure consists of the following two steps:

1. Compute Gauss–Seidel iterations, $f^1, \ldots, f^{2p}$, forcing f^i to lie in the kernel of $\hat{P}$.
2. Extrapolate the iterates using the ε-algorithm to get an improved approximation to f.

We have used the above algorithm on various problems with relatively good success. A first data set which was given to us by the authors turned out to be incompatible and as a result of our calculations an error in a spline fitting program was detected! A second set of data was provided where Gauss–Seidel converged slowly and the application of the above algorithm yielded very satisfactory results.

Conclusion. Using the ε-algorithm to accelerate convergence of a basic iteration for linear systems with nonsymmetric matrices seems to be a very promising approach.

Acknowledgment. The authors wish to thank Mr. Will Sawyer for performing the indicated computations.

REFERENCES

[1] GEKELER, E. (1972). On the solution of systems of equations by the epsilon algorithm of Wynn. *Math. Comp.* **26** 472–436.

[2] GOLUB, G. H. and DE PILLIS, J. (1988). Toward an effective two-parameter SOR method. Presented at the Conference on Iterative Methods for Large Linear Systems, October 19–21, 1988, Austin, Texas.

[3] GOLUB, G. H. and VAN LOAN, C. F. (1983). *Matrix Computations*. Johns Hopkins Univ. Press, Baltimore, Md.

[4] ISAACSON, E. and KELLER, H. B. (1966). *Analysis of Numerical Methods*. Wiley, New York.

[5] KELLER, H. B. (1965). On the solution of singular and semidefinite linear systems by iteration. *SIAM J. Numer. Anal.* **B2** 281–290.

[6] VARGA, R. S. (1962). *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs, N.J.

INSTITUT FÜR WISSENSCHAFTLICHES RECHNEN
EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE
CH-8092 ZÜRICH
SWITZERLAND

COMPUTER SCIENCE DEPARTMENT
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

THEO GASSER AND ALOIS KNEIP

Zentralinstitut für Seelische Gesundheit and Universität Heidelberg

We would like to congratulate the authors for a stimulating paper. Additive models for approximating high-dimensional regression problems have been around for quite some time, but a number of important problems have remained