JOURNAL OF COMPUTATIONAL PHYSICS 19, 110–120 (1975)

Iterative Eigenvalue Algorithms Based on Convergent Splittings

AXEL RUHE

Department of Information Processing, University of Umeå, S-90187 Umeå, Sweden

Received February 12, 1975

Application of direct iterations, based on convergent splittings, to the eigenvalue problem of large sparse symmetric matrices is discussed. A general convergence proof is given, and it is shown how parameters should be chosen to give the best rate of convergence. As special examples are considered, SOR iteration and iteration based on the use of a fast direct Poisson solver. Numerical tests are reported.

1. INTRODUCTION

In the present contribution we set out to find the smallest eigenvalue and the corresponding eigenvector of the problem

$$(A - \lambda B)\, x = 0, \tag{1.1}$$

where A is symmetric and B is positive definite. We order the eigenvalues λ_i so that

$$\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$$

and denote the corresponding B-normalized eigenvectors

$$u_1 , u_2 , \ldots, u_n ,$$

$$(Bu_i , u_k) = 0, \qquad i \neq k,$$

$$\| u_i \|_B := (Bu_i , u_i)^{1/2} = 1.$$

We are interested in cases when the matrices are large and sparse so that transformation methods, such as the QR-method [4] or Rayleigh quotient iteration [7], are not conveniently applicable, and we have to rely upon some kind of direct iteration. Methods of this kind are generally not as powerful, since they have only a linear rate of convergence and do not give a complete set of solutions to (1.1), but on the other hand they do not destroy the sparsity of A and B and give short and simple programs.

A direct iteration is an extension of the simple power method which computes a sequence of vectors

$$x_1, x_2, \ldots, x_s \to C \cdot u_1, \tag{1.2}$$

where C is a real constant, and corresponding Rayleigh quotient approximations to the eigenvalue

$$\mu_s := \mu(x_s) = (Ax_s, x_s)/(Bx_s, x_s) \to \lambda_1. \tag{1.3}$$

The basic power method has been improved by iterating in subspaces [12], orthogonalizing the iterates in a systematic manner as in the Lanczos [6, 8] or c.g. methods [9, 10, 14], and finding a related matrix iteration with faster convergence as in SOR methods [13, 11]. Here, we concentrate on this last class of algorithms.

These algorithms all depend on a splitting of the matrix

$$A - \mu_s B = V_s - H_s, \tag{1.4}$$

where V_s is easy to invert and H_s has small norm. We will specially consider the case when A is a finite difference approximation to an elliptic partial differential equation, and study how the powerful direct methods applicable to this class of matrices [3, 2] can be utilized in eigenvalue calculations.

In Section 2 we show how the splitting (1.4) can be used to formulate an algorithm, and under which conditions the vectors x_s (1.2) and values μ_s (1.3) computed will converge toward u_1 and λ_1. We show in a few cases that it is relatively easy to make sure that these conditions are fulfilled. In Section 3 we study how the splitting (1.4) should be made to give the fastest possible rate of convergence. The theory is similar to that in the linear equations case [2, 15–17], but there are some interesting complications. We conclude in Section 4 by stating results on a numerical example. We have given several reports before on iterative methods [9–11], but the application of direct methods to difference-type matrices is new here. It is worth emphasizing, however, that the theoretical results of this paper are applicable to all algorithms based on a convergent splitting (1.4).

2. Formulation of the Algorithms

The algorithms we consider here will compute a sequence of vectors (1.2) which should converge to an eigenvector and a sequence of eigenvalue approximations (1.3).

The vectors are computed by means of the splitting (1.4) as

$$x_{s+1} = V_s^{-1} H_s x_s = x_s - V_s^{-1}(A - \mu_s B)\, x_s = x_s - p_s. \tag{2.1}$$

We have, to choose V_s so that it is easy to invert, while H_s has to have a small norm. We are mainly interested in the smallest eigenvalue of (1.1), and therefore we seek a condition for the sequence μ_s to decrease.

THEOREM 1. *If we choose the splitting* (1.4) *so that*

$$\lambda_{\min}((V_s + H_s) + (V_s + H_s)^H) = 2\delta > 0 \tag{2.2}$$

and

$$\| V_s \| < M, \tag{2.3}$$

then

$$\mu_{s+1} < \mu_s \tag{2.4}$$

and

$$\lim_{s\to\infty} (A - \mu_s B)\, x_s / \| x_s \|_B = 0. \tag{2.5}$$

This implies that when λ_1 is simple and x_1 is chosen so that $\mu_1 < \lambda_2$ we get convergence to the lowest eigenvalue of (1.1).

Proof (see [11]). From the definition (1.3) of the Rayleigh quotient we see that

$$\mu_{s+1} - \mu_s = ((A - \mu_s B)\, x_{s+1}\,,\, x_{s+1})/(Bx_{s+1}\,,\, x_{s+1}).$$

We can use the recurrence (2.1) to single out the cases when this difference is negative.

We denote

$$C_s = A - \mu_s B$$

and expand (note that $(C_s x_s\,,\, x_s) = 0$ and $C_s x_s = V_s p_s$):

$$\begin{aligned}
(C_s x_{s+1}\,,\, x_{s+1}) &= (C_s(x_s - p_s),\, x_s - p_s) \\
&= -(C_s p_s\,,\, x_s) - (C_s x_s\,,\, p_s) + (C_s p_s\,,\, p_s) \\
&= -(p_s\,,\, V_s p_s) - (V_s p_s\,,\, p_s) + ((V_s - H_s)\, p_s\,,\, p_s) \\
&= -\tfrac{1}{2}\{(p_s\,,\, (V_s + H_s)\, p_s) + ((V_s + H_s)\, p_s\,,\, p_s)\} \\
&\leqslant -\delta \| p_s \|^2 < 0,
\end{aligned}$$

provided that (2.2) is fulfilled.

Since $\mu_s \geqslant \lambda_1$ we can conclude that

$$\| p_s \|^2 / \| x_{s+1} \|_B^2 \to 0, \tag{2.6}$$

which implies that

$$(A - \mu_s B)\, x_s / \| x_s \|_B = V_s p_s / \| x_s \|_B \to 0,$$

since V_s are assumed to be bounded uniformly in s (2.3). Note that

$$\begin{aligned} \| p_s \| / \| x_s \|_B &= \| p_s \| / \| x_{s+1} + p_s \|_B \\ &\leqslant \| p_s \| / (\| x_{s+1} \|_B - \| p_s \|_B) < 2 \| p_s \| / \| x_{s+1} \|_B \end{aligned}$$

as soon as $\| p_s \|_B < \frac{1}{2} \| x_{s+1} \|_B$, which eventually will hold by (2.6).

Now consider some special examples of splittings.

EXAMPLE 1. *SOR splitting.* Here,

$$A - \mu_s B = D_s - E_s - F_s,$$

where E_s is strictly lower and F_s strictly upper triangular and D_s diagonal.

We take

$$\begin{aligned} V_s &= (1/\omega) D_s - E_s, \\ H_s &= [(1/\omega) - 1] D_s + F_s \end{aligned}$$

and see that the conditions for convergence are fulfilled if

$$\epsilon \leqslant \omega \leqslant 2 - \epsilon,$$

and the starting vector x_1 is chosen so that

$$\mu_1 < \min_i \mu(e_i) = \min_i | a_{ii} | / | b_{ii} |.$$

(See [11].)

EXAMPLE 2. *Poisson solver.* Now we restrict the class of matrices to

$$A - \mu_s B = -\Delta_h + P - \mu_s I, \tag{2.7}$$

where Δ_h is the five-point difference approximation to the Laplacian in two dimensions

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

over a rectangular region, and P is a diagonal matrix. For a range of values of the scalar K, it is possible to invert $-\Delta_h + KI$ [3] and thus we can take

$$V_s = -\Delta_h + K_s I, \tag{2.8}$$

$$H_s = K_s I - P + \mu_s I. \tag{2.9}$$

We can use the minimax property of the eigenvalues to get an interval for K_s where

we are assured of convergence. We note that the eigenvalues of our problem (2.7) are majorized by the eigenvalues of

$$-\Delta_h + \bar{P} - \lambda I = 0, \qquad \bar{P} = pI, \qquad \bar{p} = \max_i p_{ii}\,. \tag{2.10}$$

From this we can conclude that $H_s + V_s$ is positive definite whenever

$$K_s \geqslant \bar{p} - \lambda_1\,. \tag{2.11}$$

If furthermore, x_s has a smaller Rayleigh quotient (1.3) than p_s, an assumption that is natural to make, we can take

$$K_s \geqslant \bar{p} - \mu_s\,, \tag{2.12}$$

which has the advantage over (2.11) that it contains only computable quantities.

3. Rate of Convergence

We will now see how we shall choose parameters in the splitting to get the fastest rate of convergence locally. In most cases the algorithms have linear convergence, and then the rate is determined by

$$R = \limsup_{s\to\infty} \| \hat{r}_s \|^{1/s},$$

where $\hat{r}_s$ is the normalized residual

$$\hat{r}_s = (Ax_s - \mu_s B x_s)/\| x_s \|_2\,.$$

If we suppose that the splitting depends only on A, B, and μ_s so that

$$V_s = V_A - \mu_s V_B\,,$$
$$H_s = H_A - \mu_s H_B\,,$$

we can use the fact that the eigenvalue approximations μ_s converge much faster than the eigenvector approximations x_s (see, e.g., [9]), to state:

Theorem 2. *The asymptotic rate of convergence of* (2.1) *is determined by the convergence of the limiting linear iteration*

$$x_{s+1} = x_s - V^{-1}(A - \lambda_1 B)\, x_s = M x_s\,, \tag{3.1}$$

where

$$\lambda_1 = \lim \mu_s\,,$$
$$V = V_A - \lambda_1 V_B\,.$$

Proof. See the corresponding proof in [11] covering a slightly more general case.

We note that

$$Mu_1 = u_1 - V^{-1}(A - \lambda_1 B)\, u_1 = u_1\,,$$

so u_1 is an eigenvector of M corresponding to the eigenvalue 1. The rate of convergence is determined by the other eigenvalues so that

$$R = \max_{\lambda \neq 1} |\, \lambda_j(M)|.$$

Since

$$M = I - V^{-1}(V - H) = V^{-1}H,$$

we can in some situations use the fact that the eigenvalues ρ of M are eigenvalues of

$$(H - \rho V)\, x = 0. \tag{3.2}$$

Especially we have:

THEOREM 3. *Let H and V be symmetric, and V, furthermore, positive definite, and denote by h_i and v_i the eigenvalues of*

$$(H - h_i B)\, x = 0, \qquad h_1 \leqslant \cdots \leqslant h_n \geqslant 0;$$
$$(V - v_i B)\, x = 0, \qquad 0 < v_1 < v_2 \cdots \leqslant v_n\,.$$

If the spread of H is limited so that

$$h_n - h_1 < \lambda_2 - \lambda_1\,, \tag{3.3}$$

then the limiting iteration converges and the rate is bounded by

$$\frac{h_1}{\lambda_2 - \lambda_1 + h_1} \leqslant \rho_n < \cdots < \rho_2 \leqslant \frac{h_n}{\lambda_2 - \lambda_1 + h_1}\,. \tag{3.4}$$

Proof. Take a Rayleigh quotient of (3.2), and vary the vector over all vectors that are V-orthogonal to u_1,

$$\frac{(Hx, x)}{(Vx, x)} = \frac{(Hx, x)/(Bx, x)}{(Vx, x)/(Bx, x)} \leqslant \frac{h_n}{v_2} \leqslant \frac{h_n}{\lambda_2 - \lambda_1 + h_1}\,,$$

since $A - \lambda_1 B = V - H$, which implies that

$$h_1 + \lambda_k - \lambda_1 \leqslant v_k \leqslant h_n + \lambda_k - \lambda_1\,,$$

by the minimax principle. This proves the second inequality and the first is an immediate consequence, if we assume that the denominator is positive.

In this case it is possible to develop a theory parallel to that for the preconditioning of linear systems (see, e.g., [16]); that is, we shall choose the splitting so that the condition number

$$\kappa = (1 - p_n)/(1 - p_2) \tag{3.5}$$

is minimized. We can then either choose an iteration parameter τ in the damped iteration

$$x_{s+1} = x_s - \tau p_s$$

so that $\rho_2 = -\rho_n = R$, or use conjugate gradient or Chebyshev acceleration over the interval (ρ_n, ρ_2) (see [2, 16, 17]).

Let us consider the examples discussed in the foregoing section:

EXAMPLE 1. *SOR methods.* If $C = A - \lambda_1 B$ satisfies property A, then ω can be chosen so that $n - 2$ of the eigenvalues of M are situated at the circle $|z| = \omega - 1$, so that $R = \omega - 1$ for ω greater than

$$\omega_c = 2/(1 + (1 - \mu_2^2)^{1/2}),$$

where μ_k are the eigenvalues of the Jacobi iteration matrix corresponding to C. We have also made extensive tests of this algorithm for cases when property A is not fulfilled (see our earlier report [11]). Since the splitting is nonsymmetric we cannot apply Theorem 3.

EXAMPLE 2. *Poisson solver.* When we have a problem of the form (2.7) and split it as (2.8) and (2.9) the situation is rather similar to the case when a nonseparable elliptic equation is solved by means of a Poisson solver [2].

Choosing the shift

$$K_s = p - \mu_s \tag{3.6}$$

(cf. (2.12)), we see by (2.9) that

$$h_1 = p - \max_i p_{ii},$$
$$h_n = p - \min_i p_{ii},$$

and for a P with a limited spread (3.3) we can apply Theorem 3. Any choice of p in the range of p_{ii} is reasonable; we have used

$$p_1 = \tfrac{1}{2}(\max_i p_{ii} + \min_i p_{ii}) \tag{3.7}$$

and

$$p_2 = e^T P e / e^T e \tag{3.8}$$

on different occasions. We note that, exactly as in the nonseparable case [2, 17],

the estimates obtained are essentially independent of the discretization. The expected number of iterations depends only on the separation $\lambda_2 - \lambda_1$ of the eigenvalues, and the properties of P.

4. A Numerical Example

Tests of the SOR methods have been reported before, therefore we concentrate on the algorithm based on a Poisson solver. We have used a program published in [1] to solve the linear equations.

The test problem we consider is the Schrödinger equation in two dimensions,

$$-\Delta\phi + P\phi - \lambda\phi = 0,$$

where P is a potential function having negative values, in interesting cases so large that $\lambda_1 < 0$, which gives rise to bound states (see [5]).

For a rectangular region with Dirichlet boundary conditions on all sides, we used a 15×21 mesh, $h = 1/16$ ($n = 315$). In Table I we list results of the algorithm (2.7)–(2.8) with the shift chosen as in (3.6) and (3.7) for the case $P \equiv 0$, and in Table II with P having three holes, each with a diameter of $\frac{1}{8}$, and P varying between $p = -100$ and $\bar{p} = 0.0$. We note that the convergence is fast indeed, since each iteration of this algorithm is equivalent to a few SOR steps. For the case $P \equiv 0$ and a 15×21 mesh, SOR needed 54 iterations to converge.

The plots show the eigenvectors: Fig. 1 in the case $P \equiv 0$, Fig. 2 when P has three holes.

TABLE I

$-\Delta\phi - \lambda\phi = 0$ with $\lambda_1 = 0.05878656$

Iteration	$\dfrac{\|\phi_s - \phi_{s-1}\|_A}{\|\phi_{s-1} - \phi_{s-2}\|_A}$	$\|(A + \lambda I)\phi_s\|_2$
1		0.102
2	$0.712 \cdot 10^{-1}$	$0.515 \cdot 10^{-2}$
3	$0.239 \cdot 10^{-1}$	$0.114 \cdot 10^{-2}$
4	0.174	$0.288 \cdot 10^{-3}$
5	0.256	$0.759 \cdot 10^{-4}$
6	0.265	$0.202 \cdot 10^{-4}$
7	0.261	$0.543 \cdot 10^{-5}$
8	0.163	$0.146 \cdot 10^{-5}$
9	1.15	$0.392 \cdot 10^{-6}$

Note. $h = 1/16$, $\phi(x, y) = 0$ on the boundary
$\|\phi\|_A = \phi^T(A + k_s I)\phi$, $\phi_0{}^T = [1, 1, \ldots, 1]$.

TABLE II

$-\Delta\phi + P\phi - \lambda\phi = 0$ with $\lambda_1 = 0.02910283$

Iteration	$\frac{\|\phi_s - \phi_{s-1}\|_\Delta}{\|\phi_{s-1} - \phi_{s-2}\|_\Delta}$	$\|(\Delta - P + \lambda I)\phi_s\|_2$
1		0.215
2	$0.731 \cdot 10^{-1}$	$0.215 \cdot 10^{-1}$
3	$0.447 \cdot 10^{-1}$	$0.588 \cdot 10^{-2}$
4	0.336	$0.217 \cdot 10^{-2}$
5	0.372	$0.972 \cdot 10^{-3}$
6	0.415	$0.454 \cdot 10^{-3}$
7	0.456	$0.213 \cdot 10^{-3}$
8	0.467	$0.100 \cdot 10^{-3}$
9	0.468	$0.470 \cdot 10^{-4}$
10	0.469	$0.221 \cdot 10^{-4}$
11	0.458	$0.104 \cdot 10^{-4}$
12	0.455	$0.486 \cdot 10^{-5}$
13	0.227	$0.228 \cdot 10^{-5}$
14	1.48	$0.107 \cdot 10^{-5}$
15	0.302	$0.503 \cdot 10^{-6}$

Note. See Table I, Note.

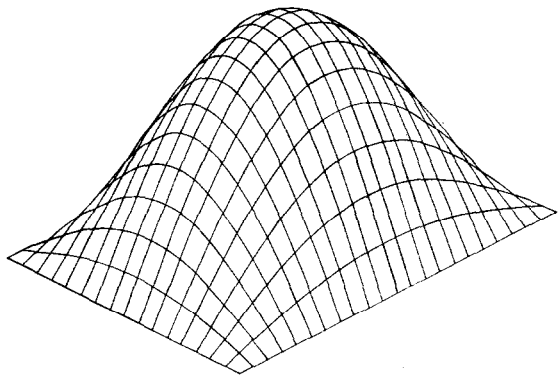

FIGURE 1

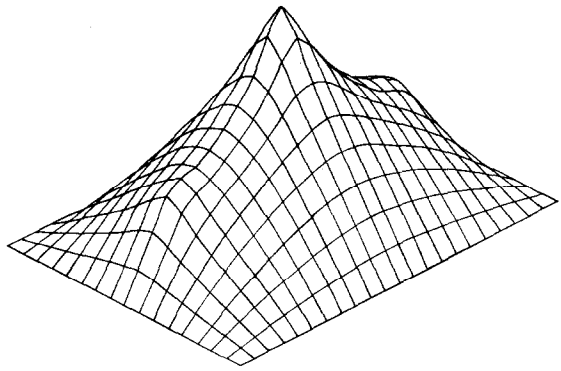

FIGURE 2

ACKNOWLEDGMENTS

The author is grateful to B. L. Buzbee for sending a copy of the program [1], and to Per Lindström for performing the computer tests. He has also had stimulating discussions with G. H. Golub and O. B. Widlund.

Part of this work was supported by the Swedish Natural Science Research Council.

REFERENCES

1. B. L. BUZBEE, A fast Poisson subroutine for $x - y$ and $z - r$ coordinates, *Comm. ACM*, to appear.
2. P. CONCUS AND G. GOLUB, *SIAM J. Numer. Anal.* **10** (1973), 1103–1120.
3. F. W. DORR, *SIAM Rev.* **12** (1970), 248–263.
4. J. G. F. FRANCIS, *Comput. J.* **4** (1961), 265–271, 332–345.
5. K. E. KHOR AND P. V. SMITH, *J. Phys. C: Solid State Phys.* **4** (1971), 2029–2040.
6. C. LANCZOS, *J. Res. Nat. Bur. Std.* **45** (1950), 255–282.
7. A. M. OSTROWSKI, *Arch. Rational Mech. Anal.* **1** (1958), 233–241; **2** (1958), 423–428; **3** (1958), 325–340, 341–347, 472–481; **4** (1958), 153–165.
8. C. C. PAIGE, *J. Inst. Math. Appl.* **10** (1972), 373–381.
9. A. RUHE AND T. WIBERG, *BIT* **12** (1972), 543–554.
10. A. RUHE, Iterative Eigenvalue Algorithms for Large Symmetric Matrices, in "Eigenwert probleme" (L. Collatz, Ed.), pp. 97–115, ISNM 24, Birkäuser, Basel–Stuttgart, 1974.
11. A. RUHE, *Math. Comp.* **28** (1974), 695–710.
12. H. RUTISHAUSER, Simultaneous Iteration Method for Symmetric Matrices, in "Handbook of Automatic Computation," Vol. 2, "Linear Algebra," pp. 284–302, Springer–Verlag, Berlin/New York, 1971.
13. H. R. SCHWARZ, *Comp. Meth. Appl. Mech. Engnrg.* **3** (1974), 11–28.
14. T. WIBERG, A Combined Lanczos and Conjugate Gradient Method for the Eigenvalue Problem of Large Sparse Matrices, Tech. Rept. UMINF-42.73, Department of Information Processing, Umeå University.

15. D. M. YOUNG, "Iterative Solution of Large Linear Systems," Academic Press, New York 1971.
16. O. AXELSSON, On Preconditioning and Convergence Acceleration in Sparse Matrix Problems, Report 74-10, Data Division, CERN, Geneva, 1974.
17. R. BARTELS AND J. W. DANIEL, A Conjugate Gradient Approach to Nonlinear Elliptic Boundary Value Problems in Irregular Regions, Tech. Rept. CNA-63, University of Texas, Austin, 1973.

Printed in Belgium