

## Estimating the Largest Eigenvalue of a Positive Definite Matrix

By Dianne P. O'Leary, G. W. Stewart\* and James S. Vandergraft

**Abstract.** The power method for computing the dominant eigenvector of a positive definite matrix will converge slowly when the dominant eigenvalue is poorly separated from the next largest eigenvalue. In this note it is shown that in spite of this slow convergence, the Rayleigh quotient will often give a good approximation to the dominant eigenvalue after a very few iterations—even when the order of the matrix is large.

Let  $A$  be a positive definite matrix of order  $n$  with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$  corresponding to the orthonormal system of eigenvectors  $x_1, x_2, \dots, x_n$ . In some applications, one must obtain an estimate of  $\lambda_1$  without going to the expense of computing the complete eigensystem of  $A$ . A simple technique that is applicable to a variety of problems is the power method. Starting with a vector  $u_0$  of Euclidean norm unity ( $\|u_0\|_2 = 1$ ), one iterates as follows:

```
1      : loop for  $k := 0, 1, 2, \dots$ 
1.1    :  $v_k := Au_k$ ;
1.2    :  $\rho_k := u_k^T v_k$ ;
1.3    :  $u_{k+1} := v_k / \|v_k\|_2$ ;
1      : end loop.
```

The theory of the method is well understood (e.g., see [4]). If  $\lambda_1 > \lambda_2$  and  $x_1^T u_0 \neq 0$ , then the vectors  $u_k$  converge linearly to  $x_1$  at a rate proportional to  $(\lambda_2/\lambda_1)^k$ . The Rayleigh quotients  $\rho_k$  converge to  $\lambda_1$  at a rate proportional to  $(\lambda_2/\lambda_1)^{2k}$ .

Convergence of the method can be hindered in two ways. First, if  $x_1^T u_0$  is pathologically small compared to some of the numbers  $x_i^T u_0$  ( $i > 1$ ), then it will take many iterations for  $u_k$  to become a good approximation to  $x_1$ . Second, if  $\lambda_2$  is very near  $\lambda_1$ , the final rate of convergence will be slow. We can do very little about the first problem, except to note that it is unlikely to occur with a randomly chosen starting vector  $u_0$ . Moreover, if our object is to compute the eigenvector  $x_1$ , the only way to accelerate slow convergence due to an unfavorable ratio  $\lambda_2/\lambda_1$  is to use more elaborate methods, such as simultaneous iteration [2], [3] or Lanczos tridiagonalization [1]. However, if we are only interested in a rough approximation to  $\lambda_1$ , it will

---

Received November 29, 1978.

AMS (MOS) subject classifications (1970). Primary 65F15.

\*This work was supported in part by the office of Naval Research under Contract No. N0014-76-C-0391.

be provided by  $\rho_k$  after a few iterations. The purpose of this note is to quantify this assertion with specific bounds.

To simplify the analysis, let the matrix  $A$  be scaled so that  $\lambda_1 = 1$ . If we set  $w_i = x_i^T u_0$ , then

$$(1) \quad u_0 = w_1 x_1 + w_2 x_2 + \dots + w_n x_n,$$

where  $w_1$  is assumed to be nonzero. Then it is easy to verify that

$$(2) \quad \rho_k = \frac{w_1^2 + \sum_{i=2}^n w_i^2 \lambda_i^{2k+1}}{w_1^2 + \sum_{i=2}^n w_i^2 \lambda_i^{2k}} \equiv \frac{\nu_k}{\delta_k}.$$

The number  $\rho_k$  satisfies  $0 < \rho_k \leq 1$ , and it will be a good approximation to  $\lambda_1 = 1$  if it is near one. Thus, we are led to investigate the worst case, when the eigenvalues of  $A$  are distributed to make  $\rho_k$  as small as possible. This distribution is easily determined by differentiating the expression (2) with respect to each  $\lambda_j$  and setting the results to zero. This gives

$$(2k + 1)w_j^2 \lambda_j^{2k} \delta_k - 2kw_j^2 \lambda_j^{2k-1} \nu_k = 0 \quad (j = 2, 3, \dots, n).$$

An inspection of (2) shows that if  $w_j \neq 0$  then no minimum of  $\rho_k$  can have  $\lambda_j = 0$ . Hence, we take

$$(3) \quad \lambda_j = \frac{2k}{2k + 1} \frac{\nu_k}{\delta_k} = \frac{2k}{2k + 1} \rho_k \quad (j = 2, 3, \dots, n).$$

Thus, if  $\tilde{\rho}_k$  denotes the minimum value of  $\rho_k$ , the minimizing distribution takes  $\lambda_2, \lambda_3, \dots, \lambda_n$  equal and slightly less than  $\tilde{\rho}_k$ . We may obtain an equation for  $\tilde{\rho}_k$  by substituting (3) into (2) to give

$$\tilde{\rho}_k = \frac{1 + \left(\frac{2k}{2k + 1}\right)^{2k+1} \tau^2 \tilde{\rho}_k^{2k+1}}{1 + \left(\frac{2k}{2k + 1}\right)^{2k} \tau^2 \tilde{\rho}_k^{2k}},$$

where

$$\tau^2 = (w_2^2 + w_3^2 + \dots + w_n^2)/w_1^2.$$

This expression may be simplified to the polynomial equation

$$(4) \quad f_k(\tilde{\rho}_k) \equiv \frac{c_k \tau^2}{2k + 1} \tilde{\rho}_k^{2k+1} + \tilde{\rho}_k - 1 = 0,$$

where

$$c_k = \left(\frac{2k}{2k + 1}\right)^{2k}.$$

The polynomial  $f_k(\rho)$  is increasing for  $\rho \geq 0$ . Since  $f_k(0) = -1$  and  $f_k(1) > 0$ ,  $f_k$  has a unique zero  $\tilde{\rho}_k \in (0, 1)$ .

Equation (4) allows us to give a qualitative description of the behavior of  $\tilde{\rho}_k$ . The positive zero of (4) decreases as the leading coefficient increases. Now  $c_k$  is well

behaved, having  $e^{-1}$  as a limit as  $k \rightarrow \infty$  and satisfying

$$.367 < e^{-1} \leq c_k \leq \frac{4}{9} < .445.$$

The number  $\tau^2$  can be written

$$\tau^2 = \tan^2 \theta,$$

where  $\theta$  is the angle between  $u_0$  and  $x_1$ . Thus, it reflects how good an approximation  $u_0$  is to  $x_1$ . In particular, if all the  $w_i$  are equal, then  $\tau^2 = n - 1$ . As  $\tau^2$  grows, the approximation becomes poorer and  $\tilde{\rho}_k$  decreases. On the other hand, as  $k$  increases, the leading coefficient of  $f_k$  decreases and  $\tilde{\rho}_k$  increases.

To obtain some quantitative results on the behavior of the power method, suppose that we wish to estimate  $\lambda_1$  to within a tolerance  $\alpha$ ; that is, for  $0 < \alpha < 1$  we wish the Rayleigh quotient  $\rho_k$  to satisfy  $\rho_k \geq \alpha \lambda_1$ . Then we must choose  $k$  so that

$$(5) \quad \frac{c_k \tau^2}{2k + 1} \alpha^{2k+1} + \alpha - 1 < 0.$$

This inequality simplifies if we set

$$x = -(2k + 1) \ln \alpha,$$

so that

$$(6) \quad \frac{e^{-x}}{x} < \frac{\alpha - 1}{c_k \tau^2 \ln \alpha}.$$

The following is a table of values of  $k$  satisfying (5) for  $\alpha = .9$  and various values of  $\tau^2$ .

$\tau^2 + 1$	5	10	15	25	50	75	100	250	500	1000
$k$	4	5	7	8	10	12	13	16	19	21

As we pointed out earlier, if the components of  $u_0$  along the eigenvectors of  $A$  are all equal, then  $\tau^2 + 1 = n$ . In this case the table gives  $k$  as a function of the order of  $A$ . What is significant is the extremely slow growth of  $k$ ; a tenfold increase in  $\tau^2 + 1$  increases  $k$  by about eight. The effect is even more marked when  $\alpha$  is taken to be 0.5, as it might be when one wants a rough estimate of the magnitude of  $\lambda_1$ . In this case  $k = 2$  for  $\tau^2 + 1 = 100$  and  $k = 3$  for  $\tau^2 + 1 = 1000$ .

We may get a crude approximation to  $k$  as a function of  $\tau^2$  by making the approximation

$$c_k \cong e^{-1}, \quad \frac{1 - \alpha}{\ln \alpha} \cong 1$$

and deleting the  $1/x$  term in (6). This gives

$$x > \ln \tau^2 - 1$$

or

$$(7) \quad k > \frac{\ln \tau^2 - 1}{2 |\ln \alpha|} - \frac{1}{2},$$

where  $\gtrsim$  indicates that the inequality is only approximate. Thus, the number of iterations required to obtain a given accuracy in  $\lambda_1$  increases as  $\ln \tau/|\ln \alpha|$ .

If one wishes to apply the power method with a fixed number of iterations, one must estimate  $\tau^2$ , which may not be easy to do. The slow growth of  $k$  with  $\tau^2$  suggests that underestimating  $\tau^2$ , even by an order of magnitude, will not affect things very much; a few extra iterations will wipe out the effect. This view is reinforced by the fact that our analysis assumes the worst possible distribution of eigenvalues.

We can make a probabilistic estimate of  $\tau^2$ , if we assume that the components of  $u_0$  are chosen to be independent normal random variables with mean zero and variance  $\sigma^2$ . In this case  $\tau^2/(n-1)$  has an  $F$  distribution with  $n-1$  and 1 degrees of freedom. It follows that at least 90% of the time,  $\tau^2 \leq 64(n-1)$ . If the approximation (7) is to be believed, this will add approximately  $-2/\ln \alpha$  iterations to those required for  $\tau^2 = (n-1)$ .

An alternative to fixing the number of iterations, is to terminate the process when  $\rho_k$  satisfies a convergence criterion such as  $|\rho_k - \rho_{k-1}| \leq (1-\alpha)\rho_k$ . The results of this note suggest that if  $\alpha$  is not too stringent, say  $\alpha \leq .9$ , then iteration will stop after a very few iterations.

Department of Computer Science  
University of Maryland  
College Park, Maryland 20742

1. B. N. PARLETT & D. S. SCOTT, "The Lanczos algorithm with selective or orthogonalization," *Math. Comp.*, v. 33, 1979, pp. 217-238.
2. H. RUTISHAUSER, "Simultaneous iteration method for symmetric matrices," *Numer. Math.*, v. 16, 1970, pp. 205-223.
3. G. W. STEWART, "Accelerating the orthogonal iteration for the eigenvalues of a Hermitian matrix," *Numer. Math.*, v. 13, 1969, pp. 362-376.
4. G. W. STEWART, *Introduction to Matrix Computations*, Academic Press, New York, 1974.