ON THE ACCURACY OF THE GERSCHGORIN CIRCLE THEOREM FOR BOUNDING THE SPREAD OF A SYMMETRIC MATRIX

David S. Scott

TR-83-19 September 1983

On the Accuracy of the Gerschgorin Circle Theorem for Bounding the Spread of a Symmetric Matrix

David S. Scott
Computer Sciences Department
University of Texas at Austin
Austin, TX 78712

Abstract The spread of a symmetric matrix is the difference between its largest and smallest eigenvalues. The Gerschgorin circle theorem can be used to bound the extreme eigenvalues of the matrix and hence its spread. This paper investigates how inaccurate this bound can be. It is shown that the ratio between the bound and the spread is bounded by sqrt(p+1), where p is the maximum number of offdiagonal nonzeroes in any row of the matrix. For full matrices this is just sqrt(n). This bound is not quite sharp for n greater than 2 but examples with ratios of sqrt(n-1) for all n are given. For banded matrices with m nonzero bands the maximum ratio is bounded by sqrt(m) independent of the size of n. This bound is sharp provided only that n is at least 2m. For sparse matrices, p may be quite small and the Gerschgorin bound may be surprisingly accurate.

1. Introduction

Let A be a symmetric matrix with eigenvalues

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

The spread of A is defined as $\lambda_n - \lambda_1$. The Gerschgorin circle theorem (see Varga [1] p. 16) gives bounds on the eigenvalues of A which in turn leads to a bound on the spread of A. Consider the matrix of dimension n:

$$A_n = \begin{bmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ . & . & . & . & & . \\ . & . & . & . & & . \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

The Gerschgorin circle theorem bounds λ_1 to be bigger than -(n-1) and λ_n to be smaller than n-1. Thus the Gerschgorin bound is 2(n-1). In fact A has $\pm$sqrt(n-1) as its only nonzero eigenvalues and so the spread of A is actually 2sqrt(n-1). Thus the ratio of the Gerschgorin bound to the spread is sqrt(n-1). This paper examines how large this ratio can be and shows that the matrix A_n given above is essentially the worst possible example. Section 2 investigates the case of $n = 2$, section 3 looks at the general full case, section 4 examines banded and sparse matrices, and section 5 gives some conclusions.

2. n = 2

Let **g(A)** denote the ratio of the Gerschgorin bound to the spread of A. If the spread is zero then A is just a scalar multiple of the identity matrix, the Gerschgorin bound is also zero and g(A) will be taken to be one, which is the minimum value over all other matrices. Let g_n be the maximum of g(A) over all symmetric matrices of dimension n. In general g(A), for a symmetric matrix of dimension n, is a very complicated function of n(n+1)/2 variables. Even for $n = 2$ this is three variables and finding the maximum g_2 would appear to be difficult. Fortunately for any real numbers δ and ϵ,

$$g(A) = g(\delta A + \epsilon I),$$

and so two of the variables can be eliminated. If A has zero offdiagonal elements then g(A) is one. Therefore for suitable choices of δ and ϵ the maximum value of g(A) is obtained by a matrix of the form

$$A(\alpha) = \begin{bmatrix} 0 & 1 \\ 1 & \alpha \end{bmatrix}.$$

Since $g(A(\alpha)) = g(A(-\alpha))$ it can be assumed that g(a) is positive. The Gerschgorin bound on the spread is $2+\alpha$. The characteristic polynomial of $A(\alpha)$ is $p(x) = x^2 - \alpha x - 1$ and the eigenvalues of $A(\alpha)$ are $(\alpha \pm \text{sqrt}(\alpha^2 + 4))/2$. Thus the spread of $A(a\alpha)$ is $\text{sqrt}(\alpha^2 + 4)$. The maximum value of g(A), which can be found by differentiating, is g_2=sqrt(2), obtained when $\alpha = 2$. This is larger (by a factor of sqrt(2)) than the ratio obtained using the matrix A_2 defined in the introduction.

3. The General Case

Let A be an nxn matrix which maximizes g(A). Let c_1 and R_1 be the center and radius of the Gerschgorin circle that gives the left endpoint of the bound on the spread. Let c_2 and R_2 be the center and radius of the corresponding righthand circle. Then Gerschgorin ratio of A is

$$g(A) = (c_2 - c_1 + R_1 + R_2)/(\lambda_n - \lambda_1). \qquad (1)$$

By Rayleigh's Principle λ_1 must be less than c_1 and λ_n must be greater than c_2. Even so it appears that g(A) might be made arbitrarily large when c_1 equals c_2 by making λ_1 and λ_n arbitrarily close together. This is false. The key observation is that the spread of A is bounded away from zero provided R_1 and R_2 are not both zero. (If they are both zero then r(A) is 1 which is the minimum possible value.) This can be shown as follows.

Symmetrically permute the rows and columns of A so that c_1 is in the upper left hand corner. Apply an orthogonal similarity (Householder reflection) to A which zeroes out all but the first two elements of the first row and column of the matrix. The transformed matrix will have a 2x2 leading principal minor of the form

$$H_1 = \begin{bmatrix} c_1 & r_1 \\ r_1 & \alpha_1 \end{bmatrix}$$

where c_1 is the same as before, r_1 is the two norm of the vector whose one norm is R_1, and α_1 is some real number. By the Cauchy interlace theorem the eigenvalues of H_1 are inner bounds on the eigenvalues of A. Thus the largest eigenvalue of H_1 is bounded above by λ_n and hence the smallest eigenvalue of H_1 cannot be larger than

$$c_1 + r_1^2/(c_1 - \lambda_n).$$

This approaches c_1 only as λ_n approaches infinity. The same argument works for bounding λ_n away from c_2.

Let μ_i and ν_i be the largest and smallest eigenvalues (respectively) of H_i, for i = 1 and 2. Then the spread of A is bounded below by

$$\gamma = \max(\mu_1,\mu_2) - \min(\nu_1,\nu_2)$$

which is a function of the six parameters c_1, r_1, c_2, r_2, α_1, and α_2. Four of the parameters (the c's and r's) are determined by the rows of the matrix which yield the Gerschgorin bound. For fixed r's and c's it is clear that g(A) is maximized when the α's are chosen to minimize γ. γ as a function of the α's is not smooth everywhere. The minimum cannot be found just by finding the zeroes of the partial derivatives. The minimum occurs at one of the singular points, when $\mu_1 = \mu_2$ and $\nu_1 = \nu_2$. Provided that c_2 is not equal to c_1, this leads to a value of γ of

$$\gamma = \mathrm{sqrt}(\,(c_2 - c_1 + (r_2^2 - r_1^2)/(c_2 - c_1))^2 + 4r_1^2).$$

If c_1 and c_2 are equal then r_1 and r_2 must also be equal and

$$\gamma = r_1 + r_2 \ \ (= 2r_1).$$

The numerator of eqn 1 depends on the R's while the denominator depends on the r's. For fixed r's it is clearly best to choose the R's as large as possible. The one norm of a vector cannot be more than sqrt(p) times the two norm of a vector where p is the number of nonzero entries in the vector. For full matrices this factor is sqrt(n-1). Furthermore g(A) is invariant under translations and scalar multiplications so, provided c_1 and c_2 are not equal, it may be assumed that $c_1 = 0$ and $c_2 = 2$. The bound, now just a function of r_1 and r_2, is

$$g(A) \leq (1 + \beta(r_1+r_2))/\mathrm{sqrt}((1+r_2^2-r_1^2)^2 + 4r_1^2),$$

where $\beta = \mathrm{sqrt}(n\text{-}1)$. This is maximized by

$$r_1 = r_2 = \beta,$$

with maximum

$$g_n \leq \mathrm{sqrt}(\beta^2 + 1) = \mathrm{sqrt}(n).$$

This is the value obtained in the previous section for n = 2. When c_1 and c_2 are equal then the bound to be maximized is just

$$g_n \leq \beta.$$

This bound is achieved for all n by the matrix A_n given in the introduction.

The bound sqrt(n) does not seem to be achievable for n greater than 2 since it is not possible to independently prescribe H_1 and H_2 in an optimal manner. On the other hand the achieved value of sqrt(n-1) is not a bound for n = 3 since g(A) = 1.5426 > sqrt(2) for the matrix

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} .$$

Further improvements can be made but all must lie between the achievable value of sqrt(n-1) and the bound sqrt(n). The value of the sharp bound will not be pursued further here.

4. Banded and Sparse Matrices

The analysis in the preceding section applies to band matrices as well with the exception of the value of the bound β. For band matrices with m nonzero bands, the row which generates R_1 has at most m-1 nonzeroes and so $\beta = \mathrm{sqrt}(m\text{-}1)$. This in turn leads to the bound

$$g(A) \leq \mathrm{sqrt}(m).$$

This bound is achievable for all n and m provided only that $n \geq 2m$ as shown by the following example.

Let m = 2k - 1 and let W be an mxm matrix which is zero except in the kth row and column. The kth row and column are all ones except the diagonal element which is two. Let X = 2I - W and let A be the direct sum of X, W, and a zero matrix of dimension n - 2m. Then A is an nxn matrix with m nonzero bands. The eigenvalues of A are just zeros and the eigenvalues of W and X. The extreme eigenvalues of both W and X are the same by construction and are $1 \pm \mathrm{sqrt}(m)$. Thus the spread of A is 2sqrt(m). The Gerschgorin bound of A is

$$2 + (m - 1) - (0 - (m - 1)) = 2m.$$

Thus g(A) = sqrt(m) as desired. The matrix A for m = 5 and n = 11 is shown below.

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

There is nothing special about being a band matrix. The key quantity is the maximum number of offdiagonal nonzeroes in a row. If A is any sparse matrix which has at most p offdiagonal nonzeroes of in any row then $g(A) \leq \text{sqrt}(p+1)$.

5. Conclusions

This paper has shown that the ratio between the Gerschgorin bound and the spread of the spectrum of a symmetric matrix A is bounded by sqrt(p+1) where p is the maximum number of offdiagonal nonzeroes in any row of the matrix. For full matrices of dimension bigger than 2 this bound is not quite sharp but the sharp bound must be at least sqrt(p). For banded and sparse matrices the bound is sharp. Furthermore p will be small and the Gerschgorin estimate will be adequate for most applications.

References

[1] Varga, R. S.
Matrix Iterative Analysis.
Prentice-Hall, 1962.