On a Theorem of Feingold

J. H. Wilkinson
Computer Science Department
Stanford University
Stanford, California 94305

ABSTRACT

We discuss an extension of a theorem which is apparently due to Feingold. Both the theorem and its extension follow *mutatis mutandis* from the corresponding material associated with the Bauer-Fike theorem. Feingold made no claims for his result and indeed referred to it as trivial. Although it proves to be rather disappointing, an understanding of the source of its weakness is instructive and leads to related techniques which are of greater value.

1. THE BAUER-FIKE THEOREM

The main topic of this paper is a result due to Feingold (oral communication) which is a simple generalization of the Bauer-Fike theorem [1]. Since we wish to make some comparisons with the latter, we begin with a brief description of it and of its limitations.

If A is an $n \times n$ matrix with linear elementary divisors, then there exists a nonsingular X such that

$$Y^H A X = X^{-1} A X = \text{diag}(\lambda_i). \tag{1.1}$$

The columns of X and Y are right-hand and left-hand eigenvectors respectively. We are interested in inclusion domains for the eigenvalues of $A + \eta E$ where $\|E\| = 1$ in some norm and $\eta \geqslant 0$. From (1.1) we have

$$X^{-1}(A + \eta E - \lambda I)X = \text{diag}(\lambda_i - \lambda) + \eta X^{-1} E X, \tag{1.2}$$

and for λ to be an eigenvalue of $A + \eta E$ the matrix on the right of (1.2) must be singular. There are two cases:

(i) $\lambda = \lambda_i$ for some i.
(ii) $\lambda \neq \lambda_i$ for all i.

In the second case $\mathrm{diag}(\lambda_i - \lambda)$ is nonsingular and singularity in (1.2) requires that

$$I + \eta \left[\mathrm{diag}(\lambda_i - \lambda) \right]^{-1} X^{-1} E X \tag{1.3}$$

be singular, giving

$$\min|\lambda_i - \lambda| \leqslant \eta \|X^{-1}\| \|X\| = \eta \kappa(X) \tag{1.4}$$

for any norm in which

$$\|\mathrm{diag}(a_i)\| = \max|a_i|. \tag{1.5}$$

This is true of the Hölder norms. The relation (1.4) is obviously satisfied in case (i), and hence in either case all eigenvalues lie in the union of the discs

$$|\lambda_i - \lambda| \leqslant \eta \kappa(X). \tag{1.6}$$

There is some freedom of choice in the columns of X, but when X has been chosen Y is determined uniquely from the relation $Y^H = X^{-1}$. Even when A has simple eigenvalues we can scale each column of X independently and arbitrarily; when A is derogatory there is an even greater freedom of choice. The proof shows that the inclusion domain is the union of the discs

$$|\lambda_i - \lambda| \leqslant \eta \min \kappa(X), \tag{1.7}$$

where the minimum is over all permissible X.

In [7] we obtain a bound for $\kappa(X)$ in terms of the sensitivity factors $1/s_i$ as follows. Let x_i and y_i be scaled so that

$$\|y_i\|_2 = \|x_i\|_2 = 1, \qquad s_i = y_i^H x_i > 0. \tag{1.8}$$

Now scale the columns of X and Y to give $\tilde{X}$ and $\tilde{Y}$ with

$$\tilde{x}_i = x_i / s_i^{1/2}, \qquad \tilde{y}_i = y_i / s_i^{1/2}, \tag{1.9}$$

and we have then

$$\kappa_2(\tilde{X}) = \|\tilde{X}\|_2 \|\tilde{Y}\|_2 \leqslant \left[\sum \frac{1}{s_i}\right]^{1/2} \left[\sum \frac{1}{s_i}\right]^{1/2} = \sum \frac{1}{s_i}. \qquad (1.10)$$

This gives as the inclusion domain the union of the discs

$$|\lambda_i - \lambda| \leqslant \eta \left(\sum \frac{1}{s_i}\right) \qquad (1.11)$$

A weakness of the result is that all the discs are of the same radius. If one of the λ_i is very sensitive, all of the discs are correspondingly large, even those associated with very insensitive eigenvalues. However, if we are interested in the smallest perturbation for which two discs coalesce, this is usually not too serious. Commonly coalescence first occurs for the two most sensitive eigenvalues, and for these the Bauer-Fike discs are not unduly large. Accordingly the theorem often provides a very realistic lower bound for coalescence. For example if A is of order 5 with

$$\lambda_1 = 1 - 10^{-3}, \quad \lambda_2 = 1, \quad \lambda_3 = 2, \quad \lambda_4 = 3, \quad \lambda_5 = 4, \qquad (1.12)$$

$$s_1 = 10^{-5}, \quad s_2 = 10^{-5}, \quad s_3 = 10^{-1}, \quad s_4 = 10^{-1}, \quad s_5 = 1, \qquad (1.13)$$

then

$$\sum \frac{1}{s_i} \approx 2 \times 10^5 \qquad (1.14)$$

and there is very little contribution from s_3 to s_5. In such examples the lower bound is almost *equal* to the minimum. Since the s_i are commonly the primary output of an eigenvalue procedure, the theorem has, in our experience, proved to be surprisingly useful.

The Bauer-Fike theorem is applicable only to matrices with linear elementary divisors. However, it is instructive to consider what happens in practice when one is given a matrix A (which happens to be defective) and attempts to determine its distance from the nearest defective matrix. Most likely the QR algorithm will be applied to determine a unitarily similar quasi-triangular matrix T. It may happen (though usually it will not) that the computed $\bar{T}$ itself has a multiple eigenvalue, in which case $\bar{T}$ is either defective or

arbitrarily close to a defective matrix. (The latter will be true if $\bar{T}$ is merely derogatory.) If η is the bound for the effect of rounding errors in the QR algorithm, the conclusion is that there is a matrix F (with $\|F\| \leqslant \eta$) such that $A + F$ is defective.

Because of the effect of rounding errors it is very likely that $\bar{T}$ will not have a multiple eigenvalue, and indeed it may not even have any pathologically close eigenvalues. In this case the Bauer-Fike theorem does apply to $\bar{T}$ and provides a lower bound η_1 for the perturbations giving a coalescence in its eigenvalues. If A really is defective, the computed η_1 must satisfy $\eta_1 \leqslant \eta$, thus giving zero as the lower bound for the perturbations in A inducing coalescence, since the lower bound cannot be negative. It might be felt that this is not a very impressive result. However, one would certainly regard the emergence of such a small η_1 as highly indicative. (It should be emphasized that one cannot *prove* that A is exactly defective via a transformation that involves rounding errors.)

2.   FEINGOLD'S THEOREM

Feingold has suggested a generalization of the Bauer-Fike theorem which can be applied to defective matrices. (This generalization follows naturally from the techniques developed by Feingold and Varga [5].) Suppose now we have

$$Y^H A X = X^{-1} A X = \mathrm{diag}(A_{ii}), \tag{2.1}$$

i.e. A is block-diagonalized via the matrix X. We may write

$$X = \begin{bmatrix} X_1 \mid X_2 \mid & \cdots & \mid X_s \end{bmatrix}, \qquad Y = \begin{bmatrix} Y_1 \mid Y_2 \mid & \cdots & \mid Y_s \end{bmatrix} \tag{2.2}$$

with conformal partitioning. Since $Y^H X = X Y^H = I$ we have

$$Y_i^H X_i = I, \qquad Y_i^H X_j = 0, \qquad I = \sum X_i Y_i^H. \tag{2.3}$$

We shall assume that the A_{ii} have disjoint spectra. Exactly the same argument as in Section 1 shows that for λ to be an eigenvalue of $A + \eta E$ we require either that λ must be an eigenvalue of some A_{ii} (and hence of A) or

$$I + \eta \, \mathrm{diag}(A_{ii} - \lambda I)^{-1} X^{-1} E X \tag{2.4}$$

must be singular, giving

$$\min \frac{1}{\left\| (A_{ii} - \lambda I)^{-1} \right\|} \leqslant \eta \kappa(X) \tag{2.5}$$

for any norm such that $\|\text{diag}(K_i)\| = \max \|K_i\|$. Hence all eigenvalues lie in the union of the domains

$$\frac{1}{\left\| (A_{ii} - \lambda I)^{-1} \right\|} \leqslant \eta \kappa(X). \tag{2.6}$$

This is Feingold's result, and from it one may deduce a result analogous to (1.11) as follows.

   The inequality (2.6) must hold for all acceptable matrices X. There is now a much wider choice for the X_i and Y_i. The columns of X_i form a basis for the right-hand invariant subspace associated with the spectrum of A_{ii}, while the columns of Y_i form a basis of the corresponding left-hand invariant subspace. If M_i $(i = 1, \ldots, s)$ are nonsingular matrices of the appropriate dimensions, then

$$\tilde{X} = X \text{diag}(M_i) \quad \text{and} \quad \tilde{Y} = Y \text{diag}(M_i^{-H}) \tag{2.7}$$

are such that

$$\tilde{Y}^H \tilde{X} = \text{diag}(M_i^{-1}) Y^H X \text{diag}(M_i) = I \tag{2.8}$$

and

$$\tilde{Y}^H A \tilde{X} = \text{diag}(M_i^{-1} A_{ii} M_i). \tag{2.9}$$

Hence corresponding to any set of M_i the corresponding $\tilde{Y}$ and $\tilde{X}$ give a block diagonalization of A which is conformal with $\text{diag}(A_{ii})$, the diagonal blocks being similar to the A_{ii}. Notice that

$$\tilde{X}_i \tilde{Y}_i^H = X_i M_i M_i^{-1} Y_i^H = X_i Y_i^H = P_i, \tag{2.10}$$

where P_i is the projector associated with the ith left-hand and right-hand subspaces.

The projector is, of course, independent of the bases. We shall be particularly interested in the case when one of each of the pairs X_i and Y_i is an orthonormal basis. According as X_i or Y_i is orthonormal we have

$$\|Y_i\|_2 = \|P_i\|_2 \quad \text{or} \quad \|X_i\|_2 = \|P_i\|_2. \tag{2.11}$$

When X_i is orthonormal we write

$$\tilde{X}_i = \|P_i\|_2^{1/2} X_i, \qquad \tilde{Y}_i = \|P_i\|_2^{-1/2} Y_i; \tag{2.12}$$

when Y_i is orthonormal we write

$$\tilde{X}_i = \|P_i\|_2^{-1/2} X_i, \qquad \tilde{Y}_i = \|P_i\|_2^{1/2} Y_i; \tag{2.13}$$

and in either case

$$\|\tilde{X}\|_2 \|\tilde{Y}\|_2 \leqslant \sum \|P_i\|_2. \tag{2.14}$$

Since only scalar factors are involved, $\tilde{X}_i, \tilde{Y}_i$ and X_i, Y_i give the same set of A_{ii}. Hence in the case when one of each of the pairs is orthonormal, our result becomes the following:

All eigenvalues of $A + \eta E$ lie in the union of the domains

$$\frac{1}{\left\|(A_{ii} - \lambda I)^{-1}\right\|} \leqslant \eta \sum \|P_i\|_2. \tag{2.15}$$

When the block form is strictly diagonal, this reduces to (1.11), since in that case $\|P_i\| = 1/s_i$. At the other extreme, when there is only one block, i.e. $A_{11} = A$, we have $X = Y = I$, $\|P\| = 1$, and the result gives as the simple inclusion domain for all n eigenvalues of A

$$\frac{1}{\left\|(A - \lambda I)^{-1}\right\|} \leqslant \eta. \tag{2.16}$$

Trivial though this is, I have found it the most useful of all results for dealing with illustrative examples. Its most impressive feature is its sheer economy. Not only are all eigenvalues of all $A + tE$ with $\|E\| = 1$ and $0 \leqslant t \leqslant \eta$ included in this domain, but every λ within it is an eigenvalue of some

$A + tE$ satisfying these conditions. Moreover, there is a rank-one perturbation which induces that λ as an eigenvalue. (The nature of these perturbations is discussed in detail in [9].)

At first sight the inequalities (2.15) are very attractive, especially in the important case when

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}. \tag{2.17}$$

It is then natural to take

$$X_1 = \begin{bmatrix} I \mid 0 \end{bmatrix}^H, \qquad Y_1^H = \begin{bmatrix} I \mid R \end{bmatrix},$$

$$X_2 = \begin{bmatrix} -R^H \mid I \end{bmatrix}^H, \qquad Y_2^H = \begin{bmatrix} 0 \mid I \end{bmatrix}, \tag{2.18}$$

where

$$A_{12} = A_{11}R - RA_{22}, \tag{2.19}$$

so that X_1 and Y_2 are automatically exactly orthonormal. [Equation (2.19) has a unique solution when the spectra of A_{11} and A_{22} are disjoint.] The corresponding block diagonal form is precisely $\mathrm{diag}(A_{ii})$. Obviously $\|P_1\|_2 = \|P_2\|_2 = \|P\|_2$ (say), and we have as our inclusion domain the union of the domains

$$\frac{1}{\left\| (A_{11} - \lambda I)^{-1} \right\|_2} \leqslant 2\|P\|_2 \eta \tag{2.20a}$$

and

$$\frac{1}{\left\| (A_{22} - \lambda I)^{-1} \right\|_2} \leqslant 2\|P\|_2 \eta \tag{2.20b}$$

(By an elegant argument Demmel [3] has shown that $2\|P\|_2$ may be replaced by $\|P\|_2 + (\|P\|_2^2 - 1)^{1/2}$, but this is not of great significance in the current context.)

From (2.16) we see that the domain defined by (2.20a) is the inclusion domain corresponding to all perturbations of norm $2\|P\|_2\eta$ in A_{11} itself, and a similar comment applies to (2.20b). The disappointing feature of this result

is that these inclusion domains are often far larger than necessary, and in a sense the bound itself is rather misleading.

3. TWO SIMPLE EXAMPLES

The weakness of the bounds given in (2.20) may be highlighted by two simple examples. We consider first the matrix A of order 5 defined by

$$A = \begin{bmatrix} \varepsilon & 1 & & & \\ & \varepsilon & 1 & & \\ & & \varepsilon & 1 & \\ & & & \varepsilon & 1 \\ \hline & & & & 0 \end{bmatrix}, \qquad \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \varepsilon, \quad \lambda_5 = 0, \quad (3.1)$$

where ε is small, and we think in terms of the leading 4×4 matrix A_{11} and the trailing 1×1 matrix 0. The inclusion domain corresponding to A itself is

$$\frac{1}{\left\| (A - \lambda I)^{-1} \right\|} \leqslant \eta, \qquad (3.2)$$

and using the l_∞ norm for simplicity, this gives

$$\frac{|\lambda - \varepsilon|^4 |\lambda|}{1 + |\lambda| \left(1 + |\varepsilon - \lambda| + |\varepsilon - \lambda|^2 + |\varepsilon - \lambda|^3 \right)} \leqslant \eta \qquad (3.3)$$

for all ε for which the first row sum of $|(A - \lambda I)^{-1}|$ is the maximum.

In order to induce an eigenvalue at $\lambda = \varepsilon/2$, for example, we need a perturbation which is $(\varepsilon/2)^5 [1 + O(\varepsilon)]$; indeed, to induce an eigenvalue at any point $k\varepsilon$ where $k = O(1)$ a perturbation of order ε^5 is required.

To apply (2.20) we require $\|P\|_2$, which is most simply determined via $\|P_2\|_2$. Since A_{22} is a one-by-one matrix, $\|P_2\| = 1/s_5$, and from

$$x_5^T = (1, -\varepsilon, \varepsilon^2, -\varepsilon^3, \varepsilon^4), \qquad y_5^T = (0, 0, 0, 0, 1) \qquad (3.4)$$

we have

$$s_5 = \frac{\varepsilon^4}{1 + O(\varepsilon^2)}, \qquad \|P\|_2 = \frac{1 + O(\varepsilon^2)}{\varepsilon^4}. \qquad (3.5)$$

Hence the inclusion domains given by (2.20) are

$$\frac{1}{\left\|\left(A_{11}-\lambda I\right)^{-1}\right\|} \leqslant 2\eta\,\frac{1+O(\varepsilon^2)}{\varepsilon^4} \quad\text{and}\quad |\lambda| \leqslant 2\eta\,\frac{1+O(\varepsilon^2)}{\varepsilon^4}. \quad (3.6)$$

The second of these is quite satisfactory. However, for values of λ in which we are interested the first domain is essentially

$$\frac{|\lambda-\varepsilon|^4}{1+O(\varepsilon)} \leqslant 2\eta\,\frac{1+O(\varepsilon^2)}{\varepsilon^4}. \quad (3.7)$$

If it were sharp, that would imply that an eigenvalue could be induced at $\varepsilon/2$ (say) by a perturbation of $\varepsilon^8/2^5$, whereas we already know that it requires a perturbation which is essentially $\varepsilon^5/2^5$. In fact a double eigenvalue can be induced at $\varepsilon/5$ by a perturbation of precisely $4^4\varepsilon^5/5^5 \approx 0.08192\varepsilon^5$ in the $(5,1)$ element. It is easy to see quite rigorously from (3.3) that if we are using the l_∞ norm and $\varepsilon \leqslant 10^{-3}$, for example, then it is not possible to make the eigenvalue at $\lambda = 0$ coalesce with one of the eigenvalues at $\lambda = \varepsilon$ until η is almost as large as this. When $\eta = 0.081\varepsilon^5$, for example, the domain containing $\lambda = 0$ has reached out to a point between 0.17ε and 0.18ε while the domain containing $\lambda = \varepsilon$ has reached out to a point between 0.23ε and 0.22ε. Notice that the second of the inequalities in (3.6) shows that *as long as η is smaller than the smallest value which makes the two domains coalesce*, the average speed of the eigenvalue starting at $\lambda = 0$ is bounded by twice its original maximum speed, $1/s_5$. However, the first domain is so absurdly large that this statement is weaker than it may seem. As is often the case, it is much easier to make sound deductions from the inclusion domain (2.16) for all n eigenvalues than to attempt to use the inclusion domains (2.20).

Our second example exhibits the weakness of the result even more cogently. Consider now the matrix

$$A = \begin{bmatrix} 2\varepsilon & 1 & & & \\ & \varepsilon & 1 & & \\ & & -\varepsilon & 1 & \\ & & & -2\varepsilon & 1 \\ \hline & & & & 0 \end{bmatrix}. \quad (3.8)$$

The eigenvalues are simple, and when ε is small they are all very sensitive. (We assume below that $\varepsilon \leqslant 10^{-3}$.) We have

$$s_1 \approx 24\varepsilon^4, \quad s_2 \approx 6\varepsilon^4, \quad s_3 \approx 6\varepsilon^4, \quad s_4 \approx 24\varepsilon^4, \quad s_5 \approx 4\varepsilon^4, \quad (3.9)$$

so that λ_5 is the most sensitive. Again for $\|P\|_2$ we have

$$\|P\|_2 = \frac{1}{s_5} \approx \frac{1}{4\varepsilon^4}. \tag{3.10}$$

Hence the inclusion domains are

$$\frac{1}{\left\|(A_{11} - \lambda I)^{-1}\right\|} \leqslant \frac{2\eta}{s_5}, \tag{3.11a}$$

$$|\lambda| \leqslant \frac{2\eta}{s_5} = \frac{2\eta}{4\varepsilon^4}, \tag{3.11b}$$

and again the second of these is satisfactory.

The first is the inclusion domain for eigenvalues of A_{11} subject to perturbations bounded by $2\eta/s_5$. Now suppose η is infinitesimally small. We know that the maximum perturbation of λ_i is η/s_i $(i = 1,\ldots,4)$. Let $1/\tilde{s}_i$ be the sensitivity factor of λ_i $(i = 1,\ldots,4)$ *regarded as an eigenvalue of* A_{11}. All we can deduce from (3.11a) is that the perturbation in λ_i is bounded by

$$\frac{2\eta}{s_5 \tilde{s}_i}. \tag{3.12}$$

However, we know that η/s_5 is itself larger than η/s_i, since λ_5 is the most sensitive eigenvalue. The alarming weakness of the bound is now apparent. The factor $2/s_5$ already more than accounts for the sensitivity of λ_i (the factor 2 is a gratuitous insult), and yet it is being reinforced by multiplication with the sensitivity of λ_i as an eigenvalue of A_{11}. In fact

$$2\eta/s_5 \tilde{s}_i = O(1/\varepsilon^7)\eta, \tag{3.13}$$

compared with the true result $O(1/\varepsilon^4)\eta$. The overestimate is strictly comparable with that in the previous case.

If we are interested in the smallest perturbation which makes $\lambda = 0$ coalesce with $\lambda = \pm \varepsilon$, the Bauer-Fike theorem gives a lower bound which is of the correct order of magnitude in ε. The inclusion domain (2.16) for all five eigenvalues of A when working with the l_∞ norm is

$$\frac{\left|(\lambda^2 - \varepsilon^2)(\lambda^2 - 4\varepsilon^2)\lambda\right|}{1 + |\lambda|\big[1 + |\lambda + 2\varepsilon| + |(\lambda + 2\varepsilon)(\lambda + \varepsilon)| + |(\lambda + 2\varepsilon)(\lambda^2 - \varepsilon^2)|\big]} \leqslant \eta \quad (3.14)$$

for values of λ in which we are interested. For small ε and $\lambda = O(\varepsilon)$ the denominator in (3.14) is essentially 1 and the domain is effectively

$$|(\lambda - \varepsilon)(\lambda + \varepsilon)(\lambda - 2\varepsilon)(\lambda + 2\varepsilon)\lambda| \leqslant \eta. \qquad (3.15)$$

For very small η this domain consists of five disjoint regions which are almost exactly circles centered on the λ_i. These spread out as η is increased, the one centered on $\lambda = 0$ expanding fastest (as we would expect from the s_i), and it reaches $\lambda = \varepsilon/2$ and $\lambda = -\varepsilon/2$ when $\eta \approx 45\varepsilon^5/32 \approx 1.41\varepsilon^5$. At this stage the domain centered on ε ($-\varepsilon$) has extended only to 0.59ε (-0.59ε). If we are measuring perturbations in the l_∞, l_1, or l_2 norms the domains coalesce at approximately $\lambda = \pm 0.54\varepsilon$ with a perturbation of approximately $\eta_c = 1.42\varepsilon^5$. (The value is actually very slightly different for the three norms.)

For the l_∞ norm the optimal perturbation to give coalescence at 0.54ε is

$$\eta_c E = \eta_c \big[e_1(-1,1,1,1,1)\big]^T, \qquad (3.16)$$

and this "direction" for E gives the optimal perturbation for inducing an eigenvalue anywhere between 0 and ε. For the l_1 norm the optimal perturbation is

$$\eta_c E = \eta_c e_5(1, -1,1,1,1), \qquad (3.17)$$

and again this direction of E gives the optimal perturbation for inducing an eigenvalue anywhere between 0 and ε. For the l_2 norm the optimal direction is not quite so simply expressed.

This topic is discussed in detail in [9]. It might seem surprising that the optimal E is so different in the l_∞ and l_1 norms. This is because it is the $e_5 e_1^T$ component which is achieving almost the whole effect. In the l_∞ norm, for example, the other components contribute only to the denominator in (3.14). If they are omitted the domain reduces to that given in (3.15). However, the l_∞ norm of E is not changed by including the other components. One has a slightly improved performance without increasing the norm. A similar comment applies to the l_1 norm.

It is salutory to remember that if the eigenvalues 0 and ε are to be moved to coalescence via a continuously varying perturbation, then one or other must pass through a point $\lambda = \varepsilon e^{i\theta}/2$ for some value of θ. To induce any λ as an eigenvalue the minimal perturbation in the l_∞ norm is given on the left-hand side of (3.14). Hence as a simple lower bound we have $\eta \approx 45\varepsilon^5/32$. Similar comments apply to the coalescence of any two neighboring eigenvalues of the five. The lower bounds obtained in this way are far superior to any

obtained from our extension of Feingold's result, and to obtain them we do not need the $(A - \lambda I)^{-1}$ in algebraic form.

4. THE SOURCE OF THE WEAKNESS

It is pertinent to ask, "What is the source of the weakness exhibited in these examples?" Essentially it is the fact that we replace an arbitrary perturbation in A itself by a related perturbation in each of the A_{ii}. Unfortunately we obtain only a bound for the norm of this related perturbation. Now this bound is, in fact, quite sharp, but unfortunately it may happen that however we choose the perturbation in A, the derived perturbation in A_{ii} is always special in that it is of a type to which the eigenvalues of A_{ii} are comparatively insensitive. Our second example makes this self-evident.

We now analyse the phenomenon more closely. Let us consider the matrix

$$A = \begin{bmatrix} \varepsilon & 1 \\ 0 & -\varepsilon \end{bmatrix}. \tag{4.1}$$

Perturbations ηE with η of order ε^2 can induce perturbations of order ε in each of the eigenvalues. Indeed, if we consider a *random* perturbation $\varepsilon^2 E$, then the eigenvalues *will* be changed by quantities of order ε unless the $2, 1$ component is poorly represented in E. However, $O(\varepsilon^2)$ perturbations in the $1, 1$ and $2, 2$ elements make $O(\varepsilon^2)$ changes in the eigenvalues, while $O(\varepsilon^2)$ perturbations in the $1, 2$ elements leave the eigenvalues unchanged. A realistic bound for the norm of the perturbation in A will not be a reliable guide to the perturbations in the eigenvalues when there is a danger that the direction of E is biased.

That this can happen is perhaps most easily seen by considering the effect of an infinitesimal perturbation ηE in A on the block diagonalization defined by Equations (2.1) to (2.3). A complicating factor is that the perturbation in A_{ii} is dependent on the form of "normalization" chosen for the perturbed invariant subspaces. The most convenient normalization for our present purpose is that embodied in the relation

$$X_i(\eta) = X_i + \sum_{j \neq i} X_j \left[\eta Z_{j1} + \eta^2 Z_{j2} + \cdots \right], \tag{4.2}$$

i.e., we hold fixed the component of $X_i(\eta)$ in the subspace X_i. (This is analogous to the normalization used in connection with individual eigenvectors in [7, pp. 68–70]. For some purposes it is more convenient to use a

normalization in which the perturbation in X_i is taken to lie in the orthogonal complement of X_i.) Equation (4.2) implies that

$$Y_i^H X_i(\eta) = I. \tag{4.3}$$

Writing

$$A_{ii}(\eta) = A_{ii} + \eta A_{ii}^{(1)} + \eta^2 A_{ii}^{(2)} + \cdots \tag{4.4}$$

and equating coefficients of η in the equation $(A + \eta E)X_i(\eta) = X_i(\eta)A_{ii}(\eta)$, we have

$$EX_i + A \sum_{j \neq i} X_j Z_{j1} = \sum_{j \neq i} X_j Z_{j1} A_{ii} + X_i A_{ii}^{(1)}. \tag{4.5}$$

Premultiplying with Y_i^H and remembering that $AX_j = X_j A_{jj}$, $Y_j^H X_i = 0$, and $Y_i^H X_i = I$, we have

$$Y_i^H E X_i = A_{ii}^{(1)}. \tag{4.6}$$

In conformity with our earlier policy we assume that X_i is orthonormal so that $\|Y_i^H\|_2 = \|P_i\|_2$, in which case we have

$$\|A_{ii}^{(1)}\| \leqslant \|Y_i^H\|_2 \|E\|_2 \|X_i\|_2 = \|P_i\|_2. \tag{4.7}$$

The bound is attained when

$$E = \frac{Y_i}{\|P_i\|_2} X_i^H. \tag{4.8}$$

(Similar results are obtained when Y_i is orthonormal and $\|X_i\|_2\|P_i\|_2$.) Hence it really is true that perturbations bounded by η in A are "equivalent to" perturbations bounded by $\eta\|P_i\|_2$ in A_{ii}. However, the E in ηE is truly arbitrary; the corresponding perturbation in A_{ii} is in the direction of $Y_i^H E X_i$. It can well happen that all such $Y_i^H E X_i$ are very special.

That is indeed true in both of the examples in Section 3. For the first example we have

$$X_1^H = [\, I \mid 0\,], \qquad Y_1^H = [\, I \mid r\,], \tag{4.9}$$

where

$$r^H = [\, \varepsilon^{-4}, -\varepsilon^{-3}, \varepsilon^{-2}, -\varepsilon^{-1}\,], \qquad \text{giving} \quad \|P_1\|_2 \approx 1/\varepsilon^4. \tag{4.10}$$

For all unit E the maximum order of magnitude of the i, jth element of $Y_1^H E X_1$ is ε^{-5+i}. It is only the elements in the first row which can be $O(\varepsilon^{-4})$, while the key element affecting the eigenvalues of A_{11} is that in the $(4, 1)$ position. However we choose E, this element is at most of order ε^{-1}.

A similar argument applies to the second example, and indeed the orders of magnitude of the elements of all the key matrices are the same for the two examples. In the second example, where all rates of change are finite, it is evident that the factor $\|P\| = 1/s_5$ already covers all the characteristics which make the other eigenvalues capable of "moving very fast." In the bound $2/s_5 \tilde{s}_i$ we are effectively counting some of these characteristics twice.

When A_{ii} is a 1×1 block (i.e. an eigenvalue λ_i), Equation (4.5) reduces to the well-known result [6]

$$\frac{d\lambda_i}{d\eta} = y_i^H E x_i, \qquad \left| \frac{d\lambda_i}{d\eta} \right| \leqslant \frac{1}{s_i} = \|P_i\|_2, \qquad (4.11)$$

where y_i and x_i are now the eigenvectors normalized so that $y_i^H x_i = 1$ in conformity with the normalization $Y_i^H X_i = I$. (In [6] we used the normalization $\|y_i\|_2 = \|x_i\|_2 = 1$ and $y_i^H x_i = s_i = 1/\|P_i\|_2$.) As in the general case, the bound is attained when $E = y_i x_i^H / s_i$. However, in the 1×1 case the perturbation in the "matrix" really is the perturbation in the eigenvalue and we cannot have a weakness of the type described above.

5.   THE SIGNIFICANCE OF $\|P\|_2$

We have seen that the presence of the factor $\|P\|_2$ in (2.20) is rather misleading. However, there is a further weakness in that $\|P\|_2$ itself is a somewhat capricious quantity. Consider the matrix

$$\begin{bmatrix} \gamma & 1 & & \\ & -\gamma & 1 & \\ & & \varepsilon & 1 \\ & & & -k\varepsilon \end{bmatrix}, \qquad (5.1)$$

where ε is small and $\gamma \ll \varepsilon$. We shall be concerned with two cases: (i) $k \approx \frac{1}{2}$, (ii) $k = 1$, the latter being rather special in that the eigenvalues are symmetric about their mean value. The sensitivity factors of the four eigenvalues are $1/s_i$ where

$$s_1 \approx 2k\varepsilon^2\gamma, \qquad s_2 \approx 2k\varepsilon^2\gamma, \qquad s_3 \approx (k+1)\varepsilon^3, \qquad s_4 \approx k^2(k+1)\varepsilon^3, \quad (5.2)$$

the first two eigenvalues being much the more sensitive. The lower bound given by the Bauer-Fike theorem for the perturbation needed to make λ_1 and λ_2 coalesce is almost exactly $k\varepsilon^2\gamma^2$. On the other hand, if γ and $-\gamma$ are to move to coalescence via a real path, one of them must pass through 0, and since $\|(A - 0 \cdot I)^{-1}\|^{-1} \approx k\varepsilon^2\gamma^2$, the perturbation required must be at least $k\varepsilon^2\gamma^2$. In fact a double eigenvalue may be induced at a point $\lambda = O(\lambda^2/\varepsilon) \ll \gamma$ (i.e. a point almost half way between $\pm\gamma$) by a perturbation $k\varepsilon^2\gamma^2[1 + o(1)]$ in the $(4, 1)$ element. These results may be deduced simply (and quite rigorously) via the simple inclusion domain $1/\|(A - \lambda I)^{-1}\| \leqslant \eta$. There is no essential difference between cases (i) and (ii), and indeed all comments remain valid for any k of order unity.

For partition into two 2×2 blocks we have

$$P_1 = \begin{bmatrix} I & R \\ 0 & 0 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0 & -R \\ 0 & I \end{bmatrix}, \tag{5.3}$$

and a rather tedious computation gives

$$R \approx \begin{bmatrix} -\dfrac{1}{\varepsilon^2} & \dfrac{1-k}{k^2\varepsilon^3} \\[2ex] -\dfrac{1}{\varepsilon} & -\dfrac{1}{k\varepsilon^2} \end{bmatrix} \qquad \text{when} \quad k \approx \tfrac{1}{2}. \tag{5.4}$$

(This is, of course, true for a very wide range of values of k.) When $k = 1$, however, the $(1, 2)$ term in R is *exactly* zero. Hence

$$\|P\|_2 = O(1/\varepsilon^3) \quad [\text{case (i)}]; \qquad \|P\|_2 = O(1/\varepsilon^2) \quad [\text{case (ii)}]. \tag{5.5}$$

This difference is not matched by any corresponding difference in the asymptotic behavior of the perturbations in the eigenvalues. Again the explanation is that the $(1, 2)$ element of R influences only the first row of $Y_1^H E X_1$; the critical $(2, 1)$ element is quite independent of it.

Perhaps it should be emphasized once again that easily the most effective way of dealing with coalescence in perturbed versions of the matrix in (3.1) is via the simple inclusion domain (2.16). For all λ in which we are interested the first row sum of $|(A - \lambda I)^{-1}|$ is the largest and the inclusion domain for the l_∞ norm is

$$\frac{|(\lambda - \eta)(\lambda + \eta)(\lambda - \varepsilon)(\lambda + k\varepsilon)|}{1 + o(1)} \leqslant \eta, \tag{5.6}$$

where the numerator comes from the $(1,4)$ element of $(A - \lambda I)^{-1}$ and the $o(1)$ in the denominator comes from the comparatively unimportant elements $(1,1)$, $(1,2)$, and $(1,3)$.

If, for example, $k = \frac{1}{2}$, $\eta = \varepsilon^2$, and $\varepsilon < 10^{-3}$, the minimal perturbation which makes λ_1 and λ_3 coalesce is of norm $0.177\varepsilon^4$; they coalesce at $\lambda = 0.721\varepsilon$. (As we would expect, λ_1 travels faster then λ_3.) A perturbation of norm $0.125\varepsilon^4$ in the optimal direction

$$E = \left[e_1(1,1,-1,-1) \right]^T \tag{5.7}$$

brings the eigenvalue λ_1 to 0.5ε; the same perturbation takes λ_3 only to 0.885ε. The direction E is optimal for inducing an eigenvalue anywhere between ε^2 and ε. The calculation to any required accuracy of the minimal value of η giving coalescence is a simple matter.

Notice that the average speed of λ_1 relative to that of λ_3 when the former has reached $\lambda = 0.5\varepsilon$ is higher than when they have reached coalescence. This is because $1/s_1 \gg 1/s_5$, so that λ_1 initially moves much faster, but when they are about to coalesce they are moving at equal and opposite (almost infinite) speeds. Considerations of this kind usually simplify the problem of making initial estimates of the values of λ giving equality in (5.6).

The bounds (2.20) do give reasonably precise information when $\|P\|_2 = O(1)$. They then show that perturbations restricted to the form of $\eta \operatorname{diag}(E_1, E_2)$ where E_1 and E_2 are conformal with A_{11} and A_{22} can be almost as effective as general perturbations in A. However, this comes as no great surprise. We would have expected *a priori* that in this case the eigenvalues of A_{11} and A_{22} would behave almost independently of each other.

6. RELATED RESULTS

In recent work Demmel [4] has presented a theorem which can be deduced directly from our version of Feingold's theorem, though it is expressed in terms in which the relationship is not immediately apparent. Demmel's theorem is specifically associated with the case when there are two blocks. It is expressed in terms of the l_2 norm and is therefore related to the bounds (2.20). Demmel's theorem is

$$\operatorname{diss}_2\!\left(\sigma_1(A_{11}), \sigma_2(A_{22})\right) \geq \frac{\operatorname{sep}_\lambda(A_{11}, A_{22})}{\|P\|_2 + \left(\|P\|_2^2 - 1\right)^{1/2}}. \tag{6.1}$$

Essentially $\mathrm{diss}_2(\sigma_1(A_{11}), \sigma_2(A_{22}))$ is the minimal perturbation in A (measured in the l_2 norm) needed to make the domain associated with the spectrum of A_{11} just intersect the domain associated with the spectrum of A_{22}. On the other hand $\mathrm{sep}_\lambda(A_{11}, A_{22})$ is the minimal perturbation of the form $\mathrm{diag}(E_{11}, E_{22})$ which achieves the same result. Hence our extension of Feingold's result expressed in these terms is

$$\mathrm{diss}_2(\sigma_1(A_{11}), \sigma_2(A_{22})) \geq \frac{\mathrm{sep}_\lambda(A_{11}, A_{22})}{2\|P\|_2}. \tag{6.2}$$

The factor $\|P\|_2 + (\|P\|_2^2 - 1)^{1/2}$ replaces the factor $2\|P\|_2$ in our result, as mentioned in Section 2. Except when $\|P\|_2 = O(1)$, this gives only a marginal improvement. Of course the weaknesses we have discussed are equally present in Demmel's theorem, and it is our opinion that expressing the result in terms of the highly implicit concepts $\mathrm{diss}_2(\sigma_1, \sigma_2)$ and $\mathrm{sep}_\lambda(A_{11}, A_{22})$ tends to obscure these weaknesses.

There is an earlier result of a related type due to Stewart [6] which may be expressed in the form

$$\mathrm{diss}_2 \geq \frac{\mathrm{sep}(A_{11}, A_{22})}{4\|P\|_2}. \tag{6.3}$$

Here $\mathrm{sep}(A_{11}, A_{22})$ is the smallest singular value of the matrix

$$A_{11} \otimes I - I \otimes A_{22}^T. \tag{6.4}$$

This is the matrix of the system of linear algebraic equations derived from (2.19). Demmel's result is always superior to that of Stewart and indeed often markedly so.

7. COMMENTS

Although Feingold's result proves to be rather disappointing, an appreciation of the source of its weakness is very instructive. The fact that it gives results based solely on *norms* of matrix perturbations is a serious shortcoming; the structure of a perturbation is of great importance. The bound

$$\left|\frac{d\lambda_i}{d\eta}\right| = \frac{1}{s_i} = \|P_i\|_2 \tag{7.1}$$

would not be nearly so useful were it not supported by the knowledge that the maximum rate of change is attained for a perturbation in the direction $e^{i\theta}y_i^H x_i$ for any value of θ. This simple result has made it, for us, one of the most useful for locating neighboring defective matrices and for determining the relevant perturbations.

The insight that can be gained from the parameter $\|P_i\|_2$ is somewhat limited, and its use tends to deflect consideration from the invariant subspaces themselves. Our formulation of the bounds (2.15) was originally in terms of X_i and Y_i, and indeed we have used $\|P_i\|_2$ only to facilitate comparison with related work of others. It is also worth noting that one cannot think in terms of X_i, Y_i, or P_i until one has made the decision to associate the relevant eigenvalues.

In a related paper [9] we discuss the coalescence problem in terms of the perturbation matrices themselves, and for this we have found the fundamental inclusion domain $\|(A - \lambda I)^{-1}\|^{-1} \leqslant \eta$ invaluable, particularly in association with the minimal rank-one perturbation needed to induce λ as an eigenvalue. As has been indicated in this paper, results in terms of the l_1 and l_∞ norms prove to be simpler (and often more instructive) than those in terms of the l_2 norm.

REFERENCES

1   F. L. Bauer and C. T. Fike, Norms and exclusion theorems, *Numer. Math.* 5:73–87 (1963).
2   J. W. Demmel, A numerical analyst's Jordan canonical form, Ph.D. Thesis, Univ. of California, Berkeley, 1983.
3   J. W. Demmel, The condition number of equivalence transformations that block diagonalize matrix pencils, *SIAM J. Numer. Anal.* 20:599–610 (1983).
4   J. W. Demmel, Computing stable eigendecompositions of matrices, *Linear Algebra Appl.*, 79:163–193 (1986).
5   D. G. Feingold and R. S. Varga, Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem, *Pacific J. Math.* 12:1241–1250 (1962).
6   G. W. Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems, *SIAM Rev.* 15:727–764 (1973).
7   J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon, Oxford, 1965.
8   J. H. Wilkinson, Sensitivity of eigenvalues, *Utilitas Math.* 25:5–76 (1984).
9   J. H. Wilkinson, Sensitivity of eigenvalues II, *Utilitas Math.*, 30:243–275 (1986).