UMIACS TR 90-31 February 1990
CS TR 2420

Eigenvalues of Graded Matrices and the Condition Numbers of a Multiple Eigenvalue

G. W. Stewart*
G. Zhang†

ABSTRACT

This paper concerns two closely related topics: the behavior of the eigenvalues of graded matrices and the perturbation of a nondefective multiple eigenvalue. We will show that the eigenvalues of a graded matrix tend to share the graded structure of the matrix and give precise conditions insuring that this tendency is realized. These results are then applied to show that the secants of the canonical angles between the left and right invariant subspaces of a multiple eigenvalue tend to characterize its behavior when its matrix is slightly perturbed.

*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the Air Force Office of Sponsored Research under Contract AFOSR-87-0188.
†Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742.

Eigenvalues of Graded Matrices and the Condition Numbers of a Multiple Eigenvalue

G. W. Stewart*
G. Zhang†

ABSTRACT

This paper concerns two closely related topics: the behavior of the eigenvalues of graded matrices and the perturbation of a nondefective multiple eigenvalue. We will show that the eigenvalues of a graded matrix tend to share the graded structure of the matrix and give precise conditions insuring that this tendency is realized. These results are then applied to show that the secants of the canonical angles between the left and right invariant subspaces of a multiple eigenvalue tend to characterize its behavior when its matrix is slightly perturbed.

1. Introduction

In this paper we will be concerned with the distribution of the eigenvalues of a graded matrix. The specific problem that gave rise to this investigation is that of explaining the behavior of a nondefective multiple eigenvalue of a general matrix when the matrix is slightly perturbed.[1] Under such circumstances, an eigenvalue of multiplicity m will typically spawn m simple eigenvalues, as might be expected. What requires explanation is that the new eigenvalues will be found at varying distances from the original eigenvalue, and these distances are more a characteristic of the matrix than of the perturbation. Thus, a multiple eigenvalue can have several condition numbers that reflect the different sensitivities of its progeny.

---

*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. This work was supported in part by the Air Force Office of Sponsored Research under Contract AFOSR-87-0188.

†Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742.

[1]There is a body of literature on the perturbation of multiple eigenvalues of Hermitian matrices or Hermitian pencils when the perturbation is an analytic function of one or more variables. For an entry into this literature, see [9].

As an illustration, consider the variation of the eigenvalues of the matrix

$$A = \frac{1}{\epsilon}\begin{pmatrix} 1 \\ 1 \\ \epsilon \end{pmatrix}(\epsilon \;\; \epsilon \;\; 1),$$

where ϵ is small. This matrix has a simple eigenvalue 3 and a double eigenvalue 0. Let $\epsilon = 10^{-5}$ and let

$$E = \begin{pmatrix} -6.7248\text{e}{-13} & -3.2031\text{e}{-11} & 1.6070\text{e}{-10} \\ -5.5392\text{e}{-11} & -1.0694\text{e}{-10} & 6.2824\text{e}{-12} \\ -4.0564\text{e}{-11} & -5.0153\text{e}{-11} & -1.6116\text{e}{-10} \end{pmatrix}.$$

Then the eigenvalues of $A+E$ are

$$\begin{aligned} \lambda_1 &= -3.244714710216396\text{e}{-12}, \\ \lambda_2 &= 3.023765334447618\text{e}{-06}, \\ \lambda_3 &= 2.999996975969131\text{e}{+00}. \end{aligned} \tag{1.1}$$

Both λ_2 and λ_3 come from the unperturbed eigenvalue 0, however, λ_2 is six orders of magnitude greater than λ_1. This difference is not an artifice of the perturbation E: almost any randomly chosen perturbation would cause the same behavior.

Some insight into this phenomenon may be obtained by choosing a suitable set of eigenvectors for A. Specifically let

$$X = \begin{pmatrix} 1.0000\text{e}{+00} & 1.0000\text{e}{+02} & 1.0000\text{e}{+02} \\ -1.0000\text{e}{+00} & 1.0000\text{e}{+02} & 1.0000\text{e}{+02} \\ 0 & -2.0000\text{e}{-03} & 1.0000\text{e}{-03} \end{pmatrix}$$

be a matrix of right eigenvectors of A. Then

$$X^{-1}(A+E)X = \begin{pmatrix} -1.0096\text{e}{-11} & 6.4813\text{e}{-09} & 6.4816\text{e}{-09} \\ -3.1963\text{e}{-09} & 3.0238\text{e}{-06} & 3.0239\text{e}{-06} \\ 3.1967\text{e}{-09} & -3.0239\text{e}{-06} & 3.0000\text{e}{+00} \end{pmatrix}.$$

Now from the theory of the perturbation of invariant subspaces, we know that up to terms of order 10^{-12} the eigenvalues of the leading principle matrix

$$\begin{pmatrix} -1.0096\text{e}{-11} & 6.4813\text{e}{-09} \\ -3.1963\text{e}{-09} & 3.0238\text{e}{-06} \end{pmatrix} \tag{1.2}$$

are eigenvalues of $A+E$. But this matrix is graded, and the theory to be developed in the next two sections will show that it must have a small eigenvalue of order 10^{-11} and a larger eigenvalue of order 10^{-6}. Hence, $A+E$ must have two eigenvalues of similar orders of magnitude. Since

$$X^{-1} = \begin{pmatrix} 5.0000\text{e}{-01} & -5.0000\text{e}{-01} & 0 \\ 1.6667\text{e}{-03} & 1.6667\text{e}{-03} & -3.3333\text{e}{+02} \\ 3.3333\text{e}{-03} & 3.3333\text{e}{-03} & 3.3333\text{e}{+02} \end{pmatrix}$$

it is easily seen that $X^{-1}(A+E)X$ has the same graded structure for almost any balanced perturbation E. Thus, the problem of assessing the effects of perturbations on the zero eigenvalues of A is reduced to the problem of characterizing the eigenvalues of graded matrices such as (1.2).

To investigate the distribution of the eigenvalues of a graded matrix, we need a characterization of graded matrices. At this point it is useful not to be too precise. We will call matrix A of order n a GRADED MATRIX if

$$A = DBD, \tag{1.3}$$

where

$$D = \text{diag}(\delta_1, \delta_2, \cdots, \delta_n)$$

with

$$\delta_1 \geq \delta_2 \geq \cdots \geq \delta_n > 0$$

and $B = (\beta_{ij})$ is "well behaved." The imprecision in this definition lies in the term "well behaved," which will be given specific meaning through hypotheses in the theorems of the next two sections. Note that this definition does not preclude some of the δ_i being equal, in which case the matrix is said to be BLOCK GRADED.

Throughout this paper, norm $\|\cdot\|$ denotes the vector 2-norm and the subordinate matrix operator norm. The magnitude of the largest element of B in (1.3) will be written

$$\beta_{\max} \stackrel{\text{def}}{=} \max_{i,j}\{|\beta_{ij}|\}.$$

We will also denote the ratio of δ_{i+1} to δ_i by

$$\rho_i \stackrel{\text{def}}{=} \frac{\delta_{i+1}}{\delta_i}.$$

In §2 of this paper, we give a lower bound on the largest eigenvalue of a graded matrix. In §3, we explore the relation between eigenvalues of a graded matrix and those of its Schur complements. These results are closely related to results obtained by one of the authors [7],[2] and more distantly to results by Barlow and Demmel [1] and Demmel and Veselić [3] for graded symmetric matrices. Here it should be stressed that our goal is not so much to derive tight bounds on the eigenvalues as to make statements about their magnitudes — as befits our intended application. Finally, in §4, we analyze the perturbation of a multiple eigenvalue and show that the secants of the canonical angles between its left and right invariant subspaces form a set of condition numbers for the eigenvalue.

2. The largest eigenvalue of a graded matrix

It is well known that the elements of a matrix can be arbitrarily larger than its spectral radius. In this section we will show that under appropriate conditions this is not true of graded matrices. Specifically, if β_{11} is not too small compared to $\beta_{\max}$, the graded matrix A has an eigenvalue that approximates $\alpha_{11} = \delta_1^2\beta_{11}$. The basic tool used to establish this result is Gerschgorin's theorem (see, e.g., [5, p. 341]). Related results may be found in [1].

The center of the Gerschgorin disk from the first row of A is $\delta_1^2\beta_{11}$, and its radius is bounded by $\beta_{\max}\delta_1(\delta_2 + \cdots + \delta_n)$. For each row other than the first, the sum of the absolute values of its elements is bounded by $\beta_{\max}\delta_2(\delta_1 + \cdots + \delta_n)$. From these facts and the Gerschgorin theorem we have the following result.

Theorem 2.1. *If*

$$\frac{|\beta_{11}|}{\beta_{\max}} \geq \frac{\delta_1(\delta_2 + \cdots + \delta_n) + \delta_2(\delta_1 + \cdots + \delta_n)}{\delta_1^2}, \tag{2.1}$$

then the largest eigenvalue $\lambda_{\max}$ of A is simple and satisfies

$$|\lambda_{\max} - \alpha_{11}| \leq \beta_{\max}\delta_1(\delta_2 + \cdots + \delta_n).$$

The other eigenvalues of A satisfy

$$|\lambda| \leq \beta_{\max}\delta_2(\delta_1 + \cdots + \delta_n) \leq |\alpha_{11}| - \beta_{\max}\delta_1(\delta_2 + \cdots + \delta_n) \leq |\lambda_{\max}|.$$

[2]Mention should also be made of two papers on low rank approximation [4, 2], whose results can be regarded as limiting cases of block scaling.

To gain more insight into the condition (2.1), suppose that A is UNIFORMLY GRADED in the sense that the ratios ρ_i are constant — say they are equal to ρ. Then the condition (2.1) is certainly satisfied if

$$\frac{|\beta_{11}|}{\beta_{\max}} \geq \frac{2\rho}{1-\rho}.$$

When $|\beta_{11}|/\beta_{\max} = 1$, this condition is satisfied for $\rho \leq \frac{1}{3}$. As $|\beta_{11}|/\beta_{\max}$ decreases, we must have

$$\rho \lesssim \frac{1}{2}\frac{|\beta_{11}|}{\beta_{\max}}.$$

Thus for the purpose of this section the "good behavior" of B means that the ratio $|\beta_{11}|/\beta_{\max}$ is near one. As this ratio grows smaller, the grading ratio ρ must decrease to compensate.

Theorem 2.1 is sufficient for assessing the magnitude of the largest eigenvalue. However, when the grading is strong, the bounds can be improved by the well-known technique of diagonal similarities. For example, ordinary Gerschgorin theory shows that the $(2,2)$-element of the matrix (1.2) is at least a three digit approximation to an eigenvalue. However, if we multiply its second row by 10^{-2} and its second column by 10^{2}, we obtain the matrix

$$\begin{pmatrix} -1.0096\text{e}{-11} & 6.4813\text{e}{-07} \\ -3.1963\text{e}{-11} & 3.0238\text{e}{-06} \end{pmatrix},$$

from which it is seen that the $(2,2)$-element approximates an eigenvalue to at least five digits. Note the agreement of this element with the second eigenvalue in the list (1.1).

Unfortunately, Gerschgorin theory can tell us little about the smaller eigenvalues. As an extreme example, the matrix

$$\begin{pmatrix} 1 & 10^{-2} & 10^{-4} \\ 10^{-2} & 10^{-4} & 10^{-6} \\ 10^{-4} & 10^{-6} & 10^{-8} \end{pmatrix}$$

has rank one, and hence its two smallest eigenvalues are zero. Nonetheless, it often happens that the eigenvalues of a graded matrix share its graded structure. In the next section we will show how this comes about.

3. Eigenvalues and Schur complements

The principal result of this section is that under appropriate hypotheses the $n-k$ smallest eigenvalues of a graded matrix A are approximated by the eigenvalues of the Schur complement of the $k \times k$ leading principal submatrix of A. Specifically, partition A in the form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} D_1 B_{11} D_1 & D_1 B_{12} D_2 \\ D_2 B_{21} D_1 & D_2 B_{22} D_2 \end{pmatrix},$$

where $D_1 = \mathrm{diag}(\delta_1, \delta_2, \cdots, \delta_k)$ and $D_2 = \mathrm{diag}(\delta_{k+1}, \cdots, \delta_n)$. Then the SCHUR COMPLEMENT of A_{11} is the matrix

$$A_{22} - A_{21} A_{11}^{-1} A_{12} = D_2 (B_{22} - B_{21} B_{11}^{-1} B_{12}) D_2.$$

Note that the Schur complement of A_{11} is the graded Schur complement of B_{11}. Consequently, if the Schur complement of B_{11} is well behaved, by the results of the last section it will have an eigenvalue of magnitude $\tilde{\beta}_{11}\delta_{k+1}^2$, which, under conditions given below, will approximate an eigenvalue of A.

The approach taken here is to determine an $(n-k) \times k$ matrix P such that

$$\begin{pmatrix} I \\ P \end{pmatrix}$$

spans an invariant subspace corresponding to the k largest eigenvalues of A. It then follows that the eigenvalues of the matrix

$$\tilde{A}_{22} = A_{22} - P A_{12} \tag{3.1}$$

are the eigenvalues associated with the complementary invariant subspace; i.e., the $n-k$ smallest eigenvalues of A. For details see [6] or [8, Ch. V].

It can be shown that the matrix P must satisfy the equation

$$P A_{11} - A_{22} P = A_{21} - P A_{12} P,$$

or in terms of B and D

$$P D_1 B_{11} D_1 - D_2 B_{22} D_2 P = D_2 B_{21} D_1 - P D_1 B_{12} D_2 P.$$

In this form, the equation is badly scaled. A better scaling is obtained by replacing P with

$$\hat{P} = D_2^{-1} P D_1, \tag{3.2}$$

which satisfies the equation

$$\hat{P} B_{11} D_1 - B_{22} D_2^2 \hat{P} D_1^{-1} = B_{21} D_1 - \hat{P} B_{12} D_2^2 \hat{P} D_1^{-1},$$

or

$$\hat{P} = B_{21} B_{11}^{-1} + B_{22} D_2^2 \hat{P} D_1^{-2} B_{11}^{-1} - \hat{P} B_{12} D_2^2 \hat{P} D_1^{-2} B_{11}^{-1}. \tag{3.3}$$

The following theorem gives a sufficient condition for the existence of a solution of (3.3) and a bound on $\hat{P}$.

Theorem 3.1. *If*

$$\gamma \stackrel{\text{def}}{=} (\|B_{12}\|\|B_{21}\|\|B_{11}^{-1}\| + \tfrac{1}{2}\|B_{22}\|)\|B_{11}^{-1}\|\rho_k^2 < \frac{1}{4} \tag{3.4}$$

then equation (3.3) has a solution satisfying

$$\|\hat{P}\| \le 2\|B_{21}\|\|B_{11}^{-1}\|. \tag{3.5}$$

Proof. Let $\hat{P}_0 = 0$, and for $k = 1, 2, \ldots$ let

$$\hat{P}_k = B_{21} B_{11}^{-1} + B_{22} D_2^2 \hat{P}_{k-1} D_1^{-2} B_{11}^{-1} - \hat{P}_{k-1} B_{12} D_2^2 \hat{P}_{k-1} D_1^{-2} B_{11}^{-1}.$$

We will show that if (3.4) is satisfied $\hat{P}_k$ converges to a solution of (3.3).[3] For brevity set

$$\eta_1 = \|B_{21}\|\|B_{11}^{-1}\|, \quad \eta_2 = \rho_k^2\|B_{22}\|\|B_{11}^{-1}\|, \quad \eta_3 = \rho_k^2\|B_{12}\|\|B_{11}^{-1}\|.$$

By a simple induction

$$\|\hat{P}_k\| \le \eta_1(1 + \eta_2(1 + s_{k-1}) + \eta_1\eta_3(1 + s_{k-1})^2) = \eta_1(1 + s_k), \tag{3.6}$$

where s_k satisfies the recursion

$$\begin{aligned} s_0 &= 0, \\ s_k &= \eta_2(1 + s_{k-1}) + \eta_1\eta_3(1 + s_{k-1})^2. \end{aligned}$$

[3] An alternative is to follow [6] and define $\hat{P}_k$ as the solution of

$$\hat{P}_k - B_{22} D_2^2 \hat{P}_k D_1^{-2} B_{11}^{-1} = B_{21} B_{11}^{-1} - \hat{P}_{k-1} B_{12} D_2^2 \hat{P}_{k-1} D_1^{-2} B_{11}^{-1}.$$

This approach gives a slightly less stringent condition for convergence but a slightly larger bound.

Using (3.4), we can show by induction that the sequence $\{s_k\}$ is monotonically increasing and bounded by one. Let $s \le 1$ be the limit of the sequence $\{s_k\}$. Then from (3.6),

$$\|\hat{P}_k\| \le \eta_1(1+s). \tag{3.7}$$

Now

$$\begin{aligned}\|\hat{P}_{k+1} - \hat{P}_k\| &\le \eta_2\|\hat{P}_k - \hat{P}_{k-1}\| + \eta_3\|\hat{P}_k - \hat{P}_{k-1}\|(\|\hat{P}_k\| + \|\hat{P}_{k-1}\|)\\ &\le \{\eta_2 + 2\eta_1\eta_3(1+s)\}\|\hat{P}_k - \hat{P}_{k-1}\|\\ &= \eta\|\hat{P}_k - \hat{P}_{k-1}\| \le \eta^k\eta_1.\end{aligned}$$

Since $\eta = \eta_2 + 2\eta_1\eta_3(1+s) < (\eta_2 + 2\eta_1\eta_3)(1+s) < 1$, the sequence $\{\hat{P}_k\}$ is a Cauchy sequence and has a limit $\hat{P}$, which by continuity must satisfy (3.3). The bound (3.5) follows from (3.7). ■

For purposes of discussion, let

$$\kappa_k = \|B\|\|B_{11}^{-1}\|.$$

Then the condition (3.4) will be satisfied if

$$(\kappa_k^2 + \tfrac{1}{2}\kappa_k)\rho_k^2 < \frac{1}{4}.$$

Moreover, it follows from (3.2) and (3.5) that

$$\|P\| \le 2\kappa_k\rho_k.$$

Thus, for the purposes of our theorem, B is "well behaved" if κ_k is near one. As κ_k grows, it must be compensated for by a larger grading ratio ρ_k.

If ρ_k is sufficiently small, then all but the first term on the right hand side of (3.3) are insignificant and

$$\hat{P} \cong B_{21}B_{11}^{-1}.$$

It follows from (3.1) that

$$\tilde{A}_{22} \cong D_2(B_{22} - B_{21}B_{11}^{-1}B_{12})D_2,$$

which is just the Schur complement of A_{11}, or equivalently the graded Schur complement of B_{11}. If the Schur complement is well behaved in the sense of the last section, then A must have an eigenvalue the size of the leading diagonal element of the Schur complement. The chief way in which ill behavior manifests itself is through cancellation producing a small leading diagonal element of the Schur complement of B_{11}.[4]

Since grading enters the condition (3.4) only through the grading ratio ρ_k, Theorem 3.1 applies to block graded matrices. More important for our purposes is the fact that the condition of the theorem can fail for one value of k and hold for another, larger value of k. Thus it is possible for the descending sequence of eigenvalues to stutter, with occasional groups of eigenvalues having the wrong magnitudes while the sequence itself remains generally correct. The following example exhibits this behavior.

Consider the matrix A whose elements are given in the following array.

-7.1532e-01	4.1271e-02	-2.0433e-03	1.7447e-03	-1.4459e-04	4.0821e-06
4.1745e-02	3.7412e-03	-1.4124e-03	3.1508e-04	-8.6632e-06	3.2593e-07
-3.2573e-03	-3.5565e-04	8.7329e-05	1.0717e-05	7.6451e-07	-2.7899e-08
-1.5509e-03	-1.2599e-04	9.1642e-06	6.8861e-07	-1.1000e-08	-1.7092e-09
-2.5920e-05	-2.3092e-06	5.8399e-07	-3.0490e-08	-2.5573e-09	5.2496e-10
-4.2303e-06	-1.0778e-07	-6.2901e-08	4.3068e-09	-6.5392e-10	1.2152e-11

The matrix was generated by uniformly grading a matrix of normal random numbers with mean zero and standard deviation one. The grading ratio is $\rho = 0.1$.

The eigenvalues of A are:

$$\begin{aligned}
\lambda_1 &= -7.1771\text{e}{-01} \\
\lambda_2 &= 6.2493\text{e}{-03} \\
\lambda_3 &= -1.3472\text{e}{-05} + 3.0630\text{e}{-05}i \\
\lambda_4 &= -1.3472\text{e}{-05} - 3.0630\text{e}{-05}i \\
\lambda_5 &= -6.3603\text{e}{-09} \\
\lambda_6 &= 2.6849\text{e}{-11}
\end{aligned}$$

The table below exhibits the value of γ from (3.4) and the first diagonal element of the Schur complement .

[4] We note in passing that the possibility of cancellation destroying the graded structure of a matrix is the bane of algorithmists who work with graded matrices.

k	γ	$\tilde{\alpha}_{11}$
1	0.10	6.1497e-03
2	0.35	-3.8747e-05
3	7.38	-2.9463e-05
4	0.05	-6.3180e-09
5	0.12	2.7030e-11

For $k = 2, 3$, the distribution stutters. The reasons for the failure of the above theory to predict the eigenvalues are different for the two values of k. For $k = 2$, the number γ can be made smaller than one-fourth by choosing a slightly different scaling matrix D. Thus, the failure is due to the cancellation in forming the leading diagonal element of the Schur complement of B_{11}. For $k = 3$, the number γ is always greater than one-fourth, and Theorem 3.1 fails. Even so, the leading diagonal element of the Schur complement gives a ball-park estimate of the size of the complex pair — something not predicted by our theory. For the other values of k the leading diagonal element predicts the size of the corresponding eigenvalue very well. In fact, when γ is small it approximates the eigenvalue itself.

4. The condition numbers of a multiple eigenvalue

Let us now return to the problem that initiated the investigations of the last two sections: the perturbation of a multiple eigenvalue. Let A have a nondefective eigenvalue λ of multiplicity m. Since we may replace A with $\lambda I - A$, we may assume without loss of generality that $\lambda = 0$.

Since zero is a nondefective eigenvalue of multiplicity m, there are m-dimensional subspaces $\mathcal{X}$ and $\mathcal{Y}$ such that $A\mathcal{X} = 0$ and $A^{\mathrm{H}}\mathcal{Y} = 0$: namely, the spaces of left and right eigenvectors corresponding to the eigenvalue zero. In §1 we saw that a judicious choice of eigenvectors led to a graded eigenvalue problem. The existence of a suitable choice for the general case is stated in the following theorem [6].

Theorem 4.1. *There are $n \times m$ matrices X and Y whose columns spaces are $\mathcal{X}$ and $\mathcal{Y}$ and which satisfy*

$$X^{\mathrm{H}}X = Y^{\mathrm{H}}Y = \mathrm{diag}(\sigma_1, \ldots, \sigma_m), \qquad \sigma_1 \geq \cdots \geq \sigma_m > 0,$$

and

$$Y^{\mathrm{H}}X = I.$$

The numbers σ_i may be defined sequentially as follows. First,

$$\sigma_1 = \max_{x\in X} \min_{y\in Y} \sec \angle(x, y). \tag{4.1}$$

If x_1 and y_1 are vectors for which the extrema are attained in (4.1), then

$$\sigma_2 = \max_{\substack{x\in X\\ x\perp x_1}} \min_{\substack{y\in Y\\ y\perp y_1}} \sec \angle(x, y). \tag{4.2}$$

If x_2 and y_2 are vectors for which the extrema are attained in (4.2), then

$$\sigma_3 = \max_{\substack{x\in X\\ x\perp x_1, x_2}} \min_{\substack{y\in Y\\ y\perp y_1, y_2}} \sec \angle(x, y), \tag{4.3}$$

and so on. The maximizing angles are called the CANONICAL ANGLES between $\mathcal{X}$ and $\mathcal{Y}$. For more details see [8, §I.5].

We must now relate this choice of basis to the eigenvalues of a perturbation $A + E$ of A. This is done in the following theorem [6].

Theorem 4.2. *Let*

$$C = Y^{\mathrm{H}} E X.$$

Then there is a matrix $\hat{C} = C + O(\|E\|^2)$ *whose eigenvalues are eigenvalues of* $A + E$.

Since the eigenvalues of C approach zero, they must approximate the m eigenvalues spawned by the zero eigenvalue of A. Now the (i, j)-element of C has the form $y_i^{\mathrm{H}} E x_j$. Hence,

$$|c_{ij}| \le \sqrt{\sigma_i \sigma_j}\|E\|.$$

Thus, unless E has special structure, C will tend to be a graded matrix with grading constants $\delta_i = \sqrt{\sigma_i}$, and by the characterizations (4.1)–(4.3) the grading will tend to be maximal. From the results of the last two sections, we know that the magnitudes of the eigenvalues of C will tend to be around $\sigma_i\|E\|$. For the example of §1, we calculated that $\|E\|$ is of order 10^{-10}, and σ_1 and σ_2 are of order one and 10^{+4}, respectively. This explains the behavior of the double eigenvalue 0.

It is unfortunate that we cannot say with complete rigor that the σ_i are condition numbers for λ. In the first place, without precise knowledge of E we cannot assert that C is graded. And even when C is graded, the example of the last section shows that the eigenvalues need not behave as we would like them to. But the phenomenon is no less real for having exceptions; and if we recognize that we are speaking about what is likely to be instead of what has to be, there can be no objection to calling the numbers σ_i, the condition numbers of the multiple eigenvalue λ.

References

[1] J. Barlow and J. Demmel (1988). "Computig Accurate Eigensystems of Scaled Diagonally Dominant Matrices." Technical Report 421, Computer Science Department, Courant Institute. Cited in [3]. To appear in *SIAM Journal on Numerical Analysis*.

[2] J. Demmel (1987). "The Smallest Perturbation of a Submatrix which Lowers the Rank and Constrained Total Least Squares Problems." *SIAM Journal on Numerical Analysis*, **24**, 199–206.

[3] J. Demmel and K. Veselić (1989). "Jacobi's Method is More Accurate Than QR." Technical Report 468, Computer Science Department, New York University.

[4] G. Golub, A. Hoffman, and G. W. Stewart (1987). "A Generalization of the Eckart-Young Matrix Approximation Theorem." *Linear Algebra and Its Applications*, **88/89**, 317–327.

[5] G. H. Golub and C. F. Van Loan (1989). *Matrix Computations* (2nd ed.). Johns Hopkins University Press, Baltimore, Maryland.

[6] G. W. Stewart (1973). "Error and Perturbation Bounds for Subspaces Associated with Certain Eigenvalue Problems." *SIAM Review*, **15**, 727–764.

[7] G. W. Stewart (1984). "On the Asymptotic Behavior of Scaled Singular Value and QR Decompostions." *Mathematics of Computation*, **43**, 483–489.

[8] G. W. Stewart and Ji guang Sun (1990). *Matrix Perturbation Theory.* Academic Press, Boston. In production.

[9] J.-G. Sun (1989). "A Note on Local Behavior of Multiple Eigenvalues." *SIAM Journal on Matrix Analysis and Applications*, **10**, 533–541.