Chapter 6
Singular Value and Eigenvalue Decompositions

The SVD was used in the discussion of the Gauss–Markov linear model and the construction of its solution. The SVD has broader implications for least squares and related computations that are based upon the properties and generalizations given below.

We also include a discussion of perturbation theory and a discussion of the SVDs of two important special matrices: bidiagonal matrices and broken arrow matrices.

6.1. The Minimax, Interlace, and Inertia Theorems

The connection between the SVD and the symmetric eigenvalue problem is very useful in the development of algorithms and perturbation theory.

We make use of two notational conventions. For a symmetric matrix A, its eigenvalues are given by

$$\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A).$$

For an $m \times n$ matrix X, $m \geq n$, its singular values are given by

$$\psi_1(X) \geq \psi_2(X) \geq \ldots \geq \psi_n(X).$$

We now present three important theorems for the symmetric eigenvalue problem, the Courant-Fischer minimax theorem, the Cauchy interlace theorem, and Sylvester's law of inertia. The first two have important interpretations in terms of the SVD, the third is useful in the development and analysis of algorithms for the symmetric eigenvalue problem. Any of these three theorems can be proven from any of the others, thus they are essentially equivalent. We choose to prove the minimax theory, and then use it to prove the interlace theorem and the inertia theorem. For an interesting discussion of the mathematical relationship among these three theorems, see Ikebe et al [**?**, 1987] or Parlett [**?**, 1998,Chapter 10].

The matrix version of the theorem is attributed to Fischer [11, 1905], but it extends to linear operators in more general inner product spaces as discussed by Courant and Hilbert [6, 1953,pp.405–407].

Theorem 6.1 (Courant–Fischer Minimax Characterization) *Let* $A \in \mathrm{R}^{n\times n}$ *be symmetric and let* $\mathcal{S}$ *be a subspace of* R^n. *Then*

$$\lambda_k(A) = \max_{\dim(\mathcal{S})=k} \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}}, \quad k = 1,\ldots,n.$$

Proof. Let A have the orthonormal eigenvectors $\mathbf{z}_1,\ldots,\mathbf{z}_n$. In this proof, let $\lambda_k = \lambda_k(A)$. Let $\mathbf{v}$ be a nonzero vector in R^n and define

$$f_j = \frac{\mathbf{z}_j^T\mathbf{v}}{\|\mathbf{v}\|_2}, \quad j = 1,\ldots,n.$$

Then the f_j satisfy

$$\frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \sum_{j=1}^{n} \lambda_j f_j^2, \quad \sum_{j=1}^{n} f_j^2 = 1.$$

For $k = 1,\ldots,n$, define the linear spaces

$$\mathcal{Z}_k = \mathrm{span}\{\mathbf{z}_1,\ldots,\mathbf{z}_k\}, \quad \mathcal{Z}_k^{\perp} = \mathrm{span}\{\mathbf{z}_{k+1},\ldots,\mathbf{z}_n\}.$$

Let $\mathcal{S}$ be any subspace of dimension k. Since $\dim(\mathcal{S}) + \dim(\mathcal{Z}_{k-1}^{\perp}) = n+1$, there exists a non-zero vector $\mathbf{v} \in \mathcal{S} \bigcap \mathcal{Z}_{k-1}^{\perp}$.

For such a $\mathbf{v}$ we may state,

$$\frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \sum_{j=k}^{n} \lambda_j f_j^2 \leq \lambda_k \sum_{j=k}^{n} f_j^2 = \lambda_k.$$

Therefore

$$\min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}} \leq \lambda_k.$$

Since $\mathcal{S}$ is arbitrary,

$$\max_{\dim(\mathcal{S})=k} \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}} \leq \lambda_k.$$

To show equality, choose $\mathcal{S} = \mathcal{Z}_k$, then for all $\mathbf{v} \in \mathcal{S}$,

$$\frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \sum_{j=1}^{k} \lambda_j f_j^2 \geq \lambda_k \sum_{j=1}^{k} f_j^2 = \lambda_k.$$

□

The following corollary is an interpretation of the minimax theorem for the SVD.

Corollary 6.2 *Let $X \in \mathrm{R}^{m\times n}$, $m \geq n$, let $\mathcal{S}$ be a subspace of R^n, and let $\mathcal{T}$ be a subspace of R^m. Then for $k = 1, \ldots, n$,*

$$\psi_k(X) = \max_{\dim(\mathcal{S})=k} \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\|X\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \tag{6.1}$$

$$= \max_{\dim(\mathcal{T})=k} \min_{0\neq\mathbf{u}\in\mathcal{T}} \frac{\|X^T\mathbf{u}\|_2}{\|\mathbf{u}\|_2}. \tag{6.2}$$

Proof. Characterizations (6.1) and (6.2) result from applying Theorem 6.1 to $X^T X$ and $X\ X^T$ respectively. □

The second theorem is Cauchy's interlace theorem [5, 1821].

Theorem 6.3 (The Cauchy Interlace Theorem) *Let $A \in \mathrm{R}^{n\times n}$ be symmetric, let $P \in \mathrm{R}^{n\times n}$ be a permutation matrix, and let k and ℓ be integers such that $k + \ell = n$. Suppose that*

$$PAP^T = \begin{matrix} \\ k \\ \ell \end{matrix} \begin{matrix} \begin{matrix} k & \ell \end{matrix} \\ \begin{pmatrix} A_k & B^T \\ B & C \end{pmatrix} \end{matrix}, \tag{6.3}$$

then

$$\lambda_j(A) \geq \lambda_j(A_k) \geq \lambda_{j+\ell}(A), \quad j = 1, \ldots, k. \tag{6.4}$$

Proof. The proof results from a proper choice of subspace for application of the Theorem 6.1. Let A_k have the eigenvectors $\mathbf{w}_1, \ldots, \mathbf{w}_k$, and let

$$\mathbf{y}_j = \begin{matrix} k \\ \ell \end{matrix} \begin{pmatrix} \mathbf{w}_j \\ 0 \end{pmatrix}$$

For each j, choose

$$\mathcal{S}_j = \mathrm{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_j\}, \quad j = 1, \ldots, n$$

and

$$\tilde{\mathcal{S}}_j = \mathrm{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_j\}.$$

Then from Theorem 6.1,

$$\lambda_j(A_k) = \min_{\mathbf{v}\in\mathcal{S}_j} \frac{\mathbf{v}^T A_k \mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \min_{\mathbf{u}\in\tilde{\mathcal{S}}_j} \frac{\mathbf{u}^T A \mathbf{u}}{\mathbf{u}^T\mathbf{u}} \leq \lambda_j(A).$$

To obtain the lower bound, note that the eigenvalues of $-A$ and $-A_k$ are

$$-\lambda_1(A) \leq \cdots \leq -\lambda_n(A),$$

and

$$-\lambda_1(A_k) \leq \cdots \leq -\lambda_k(A_k),$$

so that $-\lambda_{j+\ell}(A) \geq -\lambda_j(A_k)$ by the above argument. Thus $\lambda_{j+\ell}(A) \leq \lambda_j(A_k)$. □

The following corollary for the SVD can be proven by just applying Theorem 6.3 to $A = X^T X$.

Corollary 6.4 *Let $X \in \mathrm{R}^{m\times n}$ have the singular values $\psi_1 \geq \cdots \geq \psi_n$, let $P \in \mathrm{R}^{n\times n}$ be a permutation matrix, let k and ℓ be integers such that $k+\ell = n$, and let*

$$XP = \begin{array}{cc} k & \ell \\ \multicolumn{2}{c}{\big(} \end{array} \begin{array}{cc} X_k & Y \end{array} \big).$$

then

$$\psi_j(X) \geq \psi_j(X_k) \geq \psi_{j+\ell}(X), \quad j = 1, \ldots, k.$$

The last theorem for this section is Sylvester's law of inertia. It is important for characterizing many transformations that are used in computing eigenvectors and singular vectors. The *inertia* of a symmetric matrix A is a triple (n_A, z_A, p_A) such that $n_A + z_A + p_A = n$ where n_A,z_A, and p_A are, respectively, the number of negative, zero, and positive eigenvalues of A.

Theorem 6.5 *Let $A, B \in \mathrm{R}^{n\times n}$ be symmetric and let $Y \in \mathrm{R}^{n\times n}$ be nonsingular. If $B = Y^T A Y$ then A and B have the same inertia.*

Proof. We use the minimax characterization to show that $\mathrm{sign}(\lambda_k(B)) = \mathrm{sign}(\lambda_k(A)), k = 1, \ldots, n$. Without loss of generality, assume that $\lambda_k(A) \geq 0$. From Theorem 6.1,

$$\lambda_k(B) = \max_{\dim(\mathcal{S})=k} \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T B \mathbf{v}}{\mathbf{v}^T\mathbf{v}}$$

$$= \max_{\dim(\mathcal{S})=k} \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T Y^T A Y \mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \max_{\dim(\mathcal{S})=k} \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T Y^T A Y \mathbf{v}}{\mathbf{v}^T Y^T Y \mathbf{v}} \frac{\mathbf{v}^T Y^T Y \mathbf{v}}{\mathbf{v}^T\mathbf{v}}$$

Suppose that $\lambda_k(A) \geq 0$, then for a fixed subspace S, we have

$$\begin{aligned}\min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T B\mathbf{v}}{\mathbf{v}^T\mathbf{v}} &= \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T Y^T AY\mathbf{v}}{\mathbf{v}^T Y^T Y\mathbf{v}} \frac{\mathbf{v}^T Y^T Y\mathbf{v}}{\mathbf{v}^T\mathbf{v}} \\ &\leq \min_{0\neq\mathbf{v}\in\mathcal{S}} \lambda_k(A) \frac{\mathbf{v}^T Y^T Y\mathbf{v}}{\mathbf{v}^T\mathbf{v}} \\ &\leq \lambda_k(A)\sigma_1^2(Y).\end{aligned}$$

Since $\mathcal{S}$ is arbitrary, we have

$$\lambda_k(B) = max_{\dim(\mathcal{S})=k} \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\mathbf{v}^T B\mathbf{v}}{\mathbf{v}^T\mathbf{v}} \leq \lambda_k(A)\psi_1^2(Y).$$

By a similar argument, if $\lambda_k(B) < 0$, then

$$\lambda_k(B) \leq \lambda_k(A)\psi_n^2(Y).$$

If we apply this argument to $-A$ and $-B$, then we get that either

$$\lambda_k(A) = \lambda_k(B) = 0$$

or

$$\psi_n^2(Y) \leq \frac{\lambda_k(B)}{\lambda_k(A)} \leq \psi_1^2(Y).$$

Clearly, $\lambda_k(A)$ and $\lambda_k(B)$ must have the same sign, the A and B must have the same inertia. □

The proof of Theorem 6.5 actually proves the following stronger result.

Theorem 6.6 *Let $A, B \in \mathrm{R}^{n\times n}$ be symmetric and let $Y \in \mathrm{R}^{n\times n}$ be nonsingular. If $B = Y^T AY$ then the eigenvalues of A and B satisfy either*

$$\lambda_k(B) = \lambda_k(A) = 0$$

or

$$\psi_n(Y)^2 \leq \frac{\lambda_k(B)}{\lambda_k(A)} \leq \psi_1(Y)^2.$$

Although Theorem 6.5 does not have an SVD analogue, Theorem 6.6 does. It is a special case of a result in Horn and Johnson [**?**, pp.423–424,problem 18]. We lift the assumption that Y is nonsingular.

Corollary 6.7 *Let $X \in \mathrm{R}^{m\times n}$ where $m \geq n$, and let $Y \in \mathrm{R}^{n\times n}$. Then for $k = 1, \ldots, n$,*

$$\psi_k(X)\psi_n(Y) \leq \psi_k(XY) \leq \psi_k(X)\psi_1(Y).$$

Proof. We note that from Corollary 6.2

$$\psi_k(XY) = \max_{\dim(\mathcal{T})=k} \min_{0\neq\mathbf{u}\in\mathcal{T}} \frac{\|Y^T X^T \mathbf{u}\|_2}{\|\mathbf{u}\|_2}$$

$$\leq \max_{\dim(\mathcal{T})=k} \min_{0\neq\mathbf{u}\in\mathcal{T}} \psi_1(Y)\frac{\|X^T\mathbf{u}\|_2}{\|\mathbf{u}\|_2} = \psi_1(Y)\psi_k(X).$$

A similar argument yields the lower bound

$$\psi_k(XY) \geq \psi_k(X)\psi_n(Y).$$

□

The theorems in this section are a foundation for our understanding and developing algorithms for computing symmetric eigenvalue decomposition and the singular value decomposition.

In the next section, we give a basic perturbation theory for the SVD. In §6.3, we characterize the best rank k of a matrix using the Schmidt-Mirsky theorem which is proven from the interlace theorem for the SVD.

The interlace theorem plays an important role in two special singular value problems, that for bidiagonal matrices as discussed in §6.4, and that for broken arrow matrices as discussed in §6.5.

6.2. Perturbation Theory for SVDs and the Symmetric Eigenvalue Problems

Since all SVD algorithms are approximate, it is necessary to know at least the most elementary results about the conditioning of singular values and vectors. For more details on this fascinating area of computational mathematics, see the Notes and References section at the end of this chapter.

For the purposes of this section, We also let $X, \delta X \in \mathrm{R}^{m\times n}$ where $m \geq n$. (If $m < n$, we apply our theorems to X^T and δX^T.) The kth singular value of X is given by $\psi_k(X)$.

We begin with a simple perturbation theorem due to Weyl [**?**, 1912].

Theorem 6.8 (Weyl's Monotonicity Theorem for Singular Values) *Let $X, \delta X \in \mathrm{R}^{m\times n}$ where $m \geq n$. Then*

$$|\psi_k(X+\delta X) - \psi_k(X)| \leq \|\delta X\|_2, \quad k = 1, \ldots, n. \tag{6.5}$$

Proof. Equation (6.5) is a result of the Corollary 6.2. We have that

$$\psi_k(X+\delta X) = \max_{\dim(S)=k} \min_{0\neq\mathbf{v}\in\mathcal{S}} \frac{\|(X+\delta X)\mathbf{v}\|_2}{\|\mathbf{v}\|_2}$$

$$\leq \max_{\dim(S)=k} \min_{0\neq \mathbf{v}\in\mathcal{S}} \frac{\|X\mathbf{v}\|_2 + \|\delta X\mathbf{v}\|_2}{\|\mathbf{v}\|_2}$$

$$\leq \max_{\dim(S)=k} \min_{0\neq \mathbf{v}\in\mathcal{S}} \frac{\|X\mathbf{v}\|_2}{\|\mathbf{v}\|_2} + \|\delta X\|_2 = \psi_k(X) + \|\delta X\|_2.$$

Reversing the roles of X and $X + \delta X$ yields,

$$\psi_k(X + \delta X) \geq \psi_k(X) - \|\delta X\|_2$$

thereby establishing the inequality. □

The Wielandt-Hoffman theorem yields a similar inequality for the Frobenius norm. The proof of the eigenvalue inequality is long and given in Wilkinson [16, pp.104–108] and Horn and Johnson [**?**, 1985,p.368]. The singular value version is stated in Horn and Johnson [**?**, 1985,p.419].

Theorem 6.9 (Wielandt–Hoffman Theorem) *Let $X, \delta X \in \mathrm{R}^{m\times n}$ where $m \geq n$. Then*

$$\sum_{i=1}^{n} (\psi_i(X + \delta X) - \psi_i(X))^2 \leq \|\delta X\|_F^2.$$

Theorems 6.8 and 6.9 both show us that singular values are perfectly conditioned. That is a change of size ϵ in the matrix causes a change of size no more than ϵ in any and all of the singular values. In §6.4, and in §6.5, we will show that even these bounds are sometimes pessimistic.

Singular vectors tend to be more sensitive to perturbation as shown in the next two theorems.

Theorem 6.10 *Let $X, \delta X \in \mathrm{R}^{m\times n}$ where $m \geq n$. Let X have the singular triplets $(\psi_k, \mathbf{u}_k, \mathbf{v}_k), k = 1, \ldots, n$ and let $X + \delta X$ have the singular triplets $(\tilde{\psi}_k, \tilde{\mathbf{u}}_k, \tilde{\mathbf{v}}_k)$. Then for all $j, k = 1 = n, \ldots, , j \neq k$, we have*

$$|\mathbf{u}_k^T \tilde{\mathbf{u}}_j|, |\mathbf{v}_k^T \tilde{\mathbf{v}}_j| \leq \frac{\|\delta X\|_2}{|\tilde{\psi}_j - \psi_k|} \tag{6.6}$$

Proof. We have that

$$(X + \delta X)^T (X + \delta X)\tilde{\mathbf{v}}_j = \tilde{\psi}_j^2 \tilde{\mathbf{v}}_j.$$

$$X^T X \mathbf{v}_k = \psi_j^2 \mathbf{v}_k.$$

Multiplying the first equation by $\mathbf{v}_k^T$ and the second by $\tilde{\mathbf{v}}_j^T$ and subtracting yields

$$\mathbf{v}_k^T (\delta X)^T (X + \delta X)\tilde{\mathbf{v}}_j + \tilde{\mathbf{v}}_j^T (\delta X)^T X \mathbf{v}_k = (\tilde{\psi}_j^2 - \psi_k^2)\mathbf{v}_k^T \tilde{\mathbf{v}}_j.$$

Since

$$(X + \delta X)\tilde{\mathbf{v}}_j = \tilde{\psi}_j \tilde{\mathbf{u}}_j, \quad X\mathbf{v}_k = \tilde{\psi}_k \mathbf{u}_k$$

we have

$$\tilde{\psi}_j \mathbf{v}_k^T (\delta X)^T \tilde{\mathbf{u}}_j + \psi_k \tilde{\mathbf{v}}_j^T (\delta X)^T \mathbf{u}_k = (\tilde{\psi}_j^2 - \psi_k^2)\mathbf{v}_k^T \tilde{\mathbf{v}}_j.$$

Thus

$$|\tilde{\psi}_j^2 - \psi_k^2||\mathbf{v}_k^T \tilde{\mathbf{v}}_j| \leq \tilde{\psi}_j |\mathbf{v}_k^T (\delta X)^T \tilde{\mathbf{u}}_j| + \psi_k |\tilde{\mathbf{v}}_j^T (\delta X)^T \mathbf{u}_k| \leq (\tilde{\psi}_j + \psi_k)\|\delta X\|_2.$$

Solving for $|\mathbf{v}_k^T \tilde{\mathbf{v}}_j|$ yields the bound on the right singular vectors in (6.6). The bound on left singular vectors comes from considering X^T. □

These simple bounds use only the fact that $\|\delta X\|_2$ or $\|\delta X\|_F$ is bounded. Very often, singular value perturbations are structured.

If we make the simplifying assumption that X and $X + \delta X$ both have full column rank, we can depict the relative error in the singular values in an elegant fashion. A relaxation of the non-singularity assumption is possible, but makes the result more difficult to characterize, see Barlow and Slapničar [**?**, 2000].

From Chapter **??**, we use the parameters

$$\eta = \|(\delta X)X^\dagger\|_2, \quad \eta_F = \|(\delta X)X^\dagger\|_F \tag{6.7}$$

We assume that $\eta < 1$ as before and define also

$$\hat{\eta} = \max\{\eta, \|\delta X(X + \delta X)^\dagger\|_2\}, \quad \hat{\eta}_F = \max\{\eta_F, \|\delta X(X + \delta X)^\dagger\|_F\}. \tag{6.8}$$

Assuming that $\eta < 1$, we have that

$$\hat{\eta} \leq \frac{\eta}{1 - \eta}.$$

Two simple lemmas state the difference between the singular values and vectors of X and Y in terms of η. This result is proved in Demmel and Veselić [9, 1992].

Lemma 6.11 *Let $X, \delta X \in \mathrm{R}^{m\times n}$ be such that* $\mathrm{rank}(X) = n$. *Let η be given by* (6.7) *and satisfy $\eta < 1$. Then the singular values of X and $X + \delta X$ satisfy*

$$\frac{|\psi_k(X) - \psi_k(X + \delta X)|}{\psi_k(X)} \leq \eta.$$

Proof. The proof follows from the minimax characterization. We have that

$$\psi_k(X + \delta X) = \max_{\dim(S)=k} \min_{\mathbf{v}\in S} \frac{\|(X + \delta X)\mathbf{v}\|_2}{\|\mathbf{v}\|_2}$$

$$\leq \max_{\dim(S)=k} \min_{\mathbf{v}\in S} \frac{\|(I+(\delta X)X^\dagger)X\mathbf{v}\|_2}{\|\mathbf{v}\|_2}$$

$$\leq \max_{\dim(S)=k} \min_{\mathbf{v}\in S} \frac{\|I+(\delta X)X^\dagger\|_2\|X\mathbf{v}\|_2}{\|X\mathbf{v}\|_2}$$

$$\leq (1+\eta) \max_{\dim(S)=k} \min_{\mathbf{v}\in S} \frac{\|X\mathbf{v}\|_2}{\|\mathbf{v}\|_2} = (1+\eta)\psi_k(X).$$

A similar argument yields

$$\psi_k(X+\delta X) \geq (1-\eta)\psi_k(X)$$

Combining both bounds yields the desired result. □

We now bound the error in the vectors in $\hat{\eta}$. This is a specific case of a bound in Barlow and Demmel [2, 1990] and an elaboration of a bound in Demmel and Veselić [9, 1992]. The analysis is very similar to that used to bound error in subspaces in ULV decomposition by Barlow, Yoon, and Zha [3, 1996]. This particular proof is in Barlow [**?**, 2000].

Lemma 6.12 *Assume the hypothesis and terminology of Lemma* 6.11. *Let* $(\psi_i, \mathbf{u}_i, \mathbf{v}_i)$ *denote the ith singular triplet of* X *and let* $(\tilde{\psi}_i, \tilde{\mathbf{u}}_i, \tilde{\mathbf{v}}_i)$ *denote the ith singular triplet of* $X+\delta X$ *for* $i = 1, \ldots, n$. *Then for* $i \neq j$ *we have*

$$|\tilde{\mathbf{v}}_j^T \mathbf{v}_i| \leq \hat{\eta}\frac{2\psi_i\tilde{\psi}_j}{|\psi_i^2 - \tilde{\psi}_j^2|}, \tag{6.9}$$

and

$$|\tilde{\mathbf{u}}_j^T \mathbf{u}_i| \leq \hat{\eta}\frac{\psi_i^2 + \tilde{\psi}_j^2}{|\psi_i^2 - \tilde{\psi}_j^2|}. \tag{6.10}$$

Proof. To prove (6.9), we simply use the fact that $\mathbf{v}_i$ is an eigenvector of X^TX and $\tilde{\mathbf{v}}_j$ is an eigenvector of X^TX. Thus

$$\mathbf{v}_i^T(X+\delta X)^T(X+\delta X)\tilde{\mathbf{v}}_j = \tilde{\psi}_j^2\mathbf{v}_i^T\tilde{\mathbf{v}}_j.$$

which leads to

$$(\psi_i^2 - \tilde{\psi}_j^2)\mathbf{v}_i^T\tilde{\mathbf{v}}_j = -\mathbf{v}_i^T(\delta X)^T(X+\delta X)\tilde{\mathbf{v}}_j - \mathbf{v}_i^TX^T\delta X\tilde{\mathbf{v}}_j.$$

The use of the Cauchy-Schwarz inequality and the definition of $\hat{\eta}$ in (6.8) yields

$$|\psi_i^2 - \tilde{\psi}_j^2||\mathbf{v}_i^T\tilde{\mathbf{v}}_j| \leq \|\delta X\mathbf{v}_i\|_2\|(X+\delta X)\tilde{\mathbf{v}}_j\|_2 + \|X\mathbf{v}_i\|_2\|\delta X\tilde{\mathbf{v}}_j\|_2.$$

$$\leq 2\hat{\eta}\psi_i\tilde{\psi}_j.$$

Thus we have (6.9). A nearly identical derivation from $(X+\delta X)\ (X+\delta X)^T$ yields

$$|\psi_i^2 - \tilde{\psi}_j^2||\mathbf{u}_i^T\tilde{\mathbf{u}}_j| \leq \hat{\eta}(\psi_i^2 + \tilde{\psi}_j^2).$$

Thus (6.10). □

We note that the bound in the error of the right singular vectors is stronger than that for left singular vectors since

$$2\psi_i\tilde{\psi}_j \leq \psi_i^2 + \tilde{\psi}_j^2.$$

The singular vector bound depends upon having a gap between the ith singular value ψ_i of X and the jth singular value $\tilde{\psi}_j$ of $X+\delta X$. If that gap is zero or is very small, it may be difficult to distinguish between the ith and jth left and right singular vectors of X.

Let X and $X+\delta X$ have the partitioned SVDs

$$X = \begin{array}{c} \begin{matrix} k & m-k \end{matrix} \\ \begin{pmatrix} U_1 & U_2 \end{pmatrix} \end{array} \operatorname{diag}(\Psi_1, \Psi_2) \begin{array}{l} k \\ n-k \end{array} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix}, \tag{6.11}$$

$$X+\delta X = \begin{array}{c} \begin{matrix} k & m-k \end{matrix} \\ \begin{pmatrix} \tilde{U}_1 & \tilde{U}_2 \end{pmatrix} \end{array} \operatorname{diag}(\tilde{\Psi}_1, \tilde{\Psi}_2) \begin{array}{l} k \\ n-k \end{array} \begin{pmatrix} \tilde{V}_1^T \\ \tilde{V}_2^T \end{pmatrix}. \tag{6.12}$$

We now prove the following theorem

Theorem 6.13 *Let X and $X+\delta X$ have the SVDs partitioned as above. Suppose that either $\psi_k > \tilde{\psi}_{k+1}$ or $\tilde{\psi}_k > \psi_{k+1}$. Then*

$$\|V_1^T\tilde{V}_2\|_F = \|V_2^T\tilde{V}_1\|_F \leq 2\hat{\eta}_F/\gamma, \tag{6.13}$$

where

$$\gamma = \max\left\{\frac{\psi_k\tilde{\psi}_{k+1}}{\psi_k^2 - \tilde{\psi}_{k+1}^2}, \frac{\tilde{\psi}_k\psi_{k+1}}{\tilde{\psi}_k^2 - \psi_{k+1}^2}\right\}.$$

Proof. As in the proof of Lemma 6.12, we note that $\tilde{V}_2$ solves the eigenproblem

$$(X+\delta X)^T\ (X+\delta X)\tilde{V}_2 = \tilde{V}_2\tilde{\Psi}_2^2,$$

and V_1 solves the eigenproblem

$$V_1^T X^T X = \Psi_1^2 V_1^T.$$

Multiplying V_1^T times the first equation from the left, and $\tilde{V}_2$ times the second equation from the right and subtracting we have

$$-V_1^T[\delta X^T(X+\delta X)+X^T\delta X]\tilde{V}_2 = \Psi_1^2 V_1^T\tilde{V}_2 - V_1^T\tilde{V}_2\tilde{\Psi}_2^2. \tag{6.14}$$

Since for $X \in \mathrm{R}^{m\times n}$, $m \geq n$ with $\mathrm{rank}(X) = n$, we have

$$X^\dagger X = (X+\delta X)^\dagger(X+\delta X) = I,$$

we can rewrite (6.14) as

$$-V_1^T[(X^\dagger X)^T\delta X^T(X+\delta X)+X^T\delta X(X+\delta X)^\dagger(X+\delta X)]\tilde{V}_2 = \Psi_1^2V_1^T\tilde{V}_2-V_1^T\tilde{V}_2\tilde{\Psi}_2^2$$

which becomes

$$-\Psi_1^T U_1^T[(\delta X X^\dagger)^T + \delta X(X+\delta X)^\dagger]\tilde{U}_2\tilde{\Psi}_2 = \Psi_1^2V_1^T\tilde{V}_2 - V_1^T\tilde{V}_2\tilde{\Psi}_2^2. \tag{6.15}$$

Equation (6.15) can be written

$$\Psi_1^T Z\tilde{\Psi}_2 = \Psi_1^2 E - E\tilde{\Psi}_2^2$$

where

$$Z = -U_1^T[(\delta X X^\dagger)^T + \delta X(X+\delta X)^\dagger]\tilde{U}_2,$$
$$E = V_1^T\tilde{V}_2.$$

Clearly,

$$\|Z\|_F \leq \|\delta X X^\dagger\|_F + \|\delta X(X+\delta X)^\dagger\|_F \leq 2\hat{\eta}_F.$$

Moreover the components of Z and E satisfy

$$e_{ij} = z_{ij}\frac{\psi_i\tilde{\psi}_{j+k}}{\psi_i^2 - \tilde{\psi}_{j+k}^2}.$$

For $a > b$, the function $\frac{ab}{a^2-b^2}$ is strictly decreasing in a and strictly increasing in b, thus,

$$|e_{ij}| \leq |z_{ij}|\frac{\psi_k\tilde{\psi}_{k+1}}{\psi_k^2 - \tilde{\psi}_{k+1}^2}$$

for all appropriate i and j. Thus

$$\|E\|_F \leq \|Z\|_F\frac{\psi_k\tilde{\psi}_{k+1}}{\psi_k^2 - \tilde{\psi}_{k+1}^2}. \tag{6.16}$$

Inserting the definition of E and the bound on $\|Z\|_F$ we have

$$\|V_1^T\tilde{V}_2\|_F \leq 2\hat{\eta}_F\frac{\psi_k\tilde{\psi}_{k+1}}{\psi_k^2 - \tilde{\psi}_{k+1}^2}. \tag{6.17}$$

By Corollary **??**, we have

$$\|V_1^T \tilde{V}_2\|_F = \|\tilde{V}_1^T V_2\|_F.$$

Reversing the roles of X and $X + \delta X$ in the argument above yields

$$\|\tilde{V}_1^T V_2\|_F \leq 2\hat{\eta}_F \frac{\tilde{\psi}_k \psi_{k+1}}{\tilde{\psi}_k^2 - \psi_{k+1}^2}. \tag{6.18}$$

Taking the minimum of (6.17) and (6.18) yields (6.13). □

Many algorithms lead to bounds based upon residuals. We now give a simple one for SVDs.

Theorem 6.14 *Let $X \in \mathrm{R}^{m \times n}$, let $\phi \in \mathrm{R}$, and let X have the singular triples $(\psi_k, \mathbf{u}_k, \mathbf{v}_k), k = 1, \ldots, n$ ordered so that*

$$|\psi_1 - \phi| \leq |\psi_2 - \phi| \leq \ldots \leq |\psi_n - \phi|.$$

Let $\mathbf{w}, \mathbf{r} \in \mathrm{R}^n$ and $\mathbf{y} \in \mathrm{R}^m$ satisfy

$$X\mathbf{w} = \phi\mathbf{y}, \tag{6.19}$$

$$X^T\mathbf{y} = \phi\mathbf{w} + \mathbf{r}. \tag{6.20}$$

Then

1. $\mathbf{w}^T\mathbf{r} = 0$.
2. $|\psi_1 - \phi| \leq \|\mathbf{r}\|_2$
3. *If* $\cos\theta = \mathbf{v}_1^T\mathbf{w}$, *then*

$$|\tan\theta| \leq \frac{\||\phi|\mathbf{r}\|_2}{|\psi_2^2 - \phi^2|}$$

Proof. To show that $\mathbf{w}^T\mathbf{r} = 0$, note that

$$\mathbf{w}^T X^T \mathbf{y} = \phi\mathbf{w}^T\mathbf{w} + \mathbf{w}^T\mathbf{r}.$$

Since $\mathbf{w}^T X^T = \phi\mathbf{w}^T$, we have

$$\mathbf{w}^T\mathbf{r} = \phi(\mathbf{w}^T\mathbf{w} - \mathbf{y}^T\mathbf{y}) = 0.$$

To show the singular value bound, we note that $(\phi, \mathbf{y}, \mathbf{w})$ is an exact singular triplet of

$$\tilde{X} = X - \mathbf{y}\mathbf{r}^T.$$

Thus from Theorem 6.8 ,ψ_1, the closest singular value of X to ϕ, must satisfy

$$|\psi_1 - \phi| \leq \|\mathbf{y}\mathbf{r}^T\|_2 = \|\mathbf{r}\|_2.$$

Let $V_2 = (\mathbf{v}_2, \ldots, \mathbf{v}_n)$ and let $U_2 = (\mathbf{u}_2, \ldots, \mathbf{u}_m)$. Then

$$V_2^T \tilde{X}^T \tilde{X} \mathbf{w} = \phi^2 V_2^T \mathbf{w}$$

which implies that

$$\begin{aligned}(\Psi_2^2 - \phi^2 I) V_2^T \mathbf{w} &= -V_2^T \mathbf{r}\mathbf{y}^T \tilde{X}\mathbf{w} - V_2^T X^T \mathbf{y}\mathbf{r}^T \mathbf{w} \\ &= -\phi V_2^T \mathbf{r}.\end{aligned}$$

Thus,

$$|\sin\theta| = \|V_2^T \mathbf{w}\|_2 \leq |\phi|(\psi_2^2 - \phi^2)^{-1} \|V_2^T \mathbf{r}\|_2 \leq |\phi|(\psi_2^2 - \phi^2)^{-1} |\cos\theta| \|\mathbf{r}\|_2.$$

Which leads to the bound

$$|\tan\theta| \leq |\phi|(\psi_2^2 - \phi^2)^{-1} \|\mathbf{r}\|_2.$$

□

6.3. The Schmidt–Mirsky Approximation Theorem

In many least squares applications, it is desirable to reduce the number of variables by identifying linear dependencies. For that, we need to characterize when a matrix $X \in \mathrm{R}^{m\times n}$ is close to a rank k matrix X_k. Thus given X and a norm $\|\cdot\|$, we want to find $X_k \in \mathrm{R}^{m\times n}$ that satisfies

$$\|X - X_k\| = \min_{\mathrm{rank}(Y)=k} \|X - Y\|$$

For the Frobenius norm, the solution of this problem was first proven by Schmidt [15, 1907], and later stated without proof by Eckart and Young [10, 1936]. For the two–norm, the solution was given by Mirsky [14, 1960]. We state both results as one theorem.

Theorem 6.15 (Schmidt–Mirsky Approximation Theorem) *Let $X \in \mathrm{R}^{m\times n}$ have the SVD given by (??). Assume that* $\mathrm{rank}(X) > k$. *Let*

$$X_k = U_1 \Psi_1 V_1^T, \tag{6.21}$$

where $U_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$, $\Psi = \mathrm{diag}(\psi_1, \ldots, \psi_k)$, $V_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_k)$. Then

$$\|X - X_k\|_2 = \min_{\mathrm{rank}(Y)=k} \|X - Y\|_2 = \psi_{k+1} \tag{6.22}$$

$$\|X - X_k\|_F = \min_{\mathrm{rank}(Y)=k} \|X - Y\|_F = \|(\psi_{k+1}, \ldots, \psi_n)^T\|_2. \tag{6.23}$$

Proof. We note that

$$X - X_k = U_2 \Psi_2 V_2^T$$

where $U_2 = (\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m)$, $V_2 = (\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n)$, and $\Psi_2 = \mathrm{diag}(\psi_{k+1}, \ldots, \psi_n)$. Thus clearly,

$$\|X - X_k\|_2 = \psi_k, \quad \|X - X_k\|_F = \|(\psi_{k+1}, \ldots, \psi_n)^T\|_2.$$

Let $Y \in \mathrm{R}^{m \times n}$ have rank k and let it have the SVD

$$Y = Z \Theta W^T, \quad Z \in \mathrm{R}^{m \times m}, W \in \mathrm{R}^{n \times n}, \text{ orthogonal}$$

where

$$\begin{array}{c} \\ Z = \end{array} \begin{array}{c} \begin{array}{cc} k & m-k \end{array} \\ \begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \end{array}, \quad \begin{array}{c} \\ W = \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} W_1 & W_2 \end{pmatrix} \end{array},$$

$$\Theta = \begin{array}{c} \\ k \\ m-k \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} \Theta_1 & 0 \\ 0 & 0 \end{pmatrix} \end{array}.$$

By orthogonal invariance,

$$\|X - Y\|_2 = \|(X - Y)W\|_2.$$

Applying Corollary 6.4 to $(X - Y)W$ yields

$$\|(X - Y)W\|_2 \geq \|(X - Y)W_2\|_2.$$

However, since $YW_2 = 0$, we have

$$\|X - Y\|_2 \geq \|XW_2\|_2.$$

Let $\hat{X} = XW$ and be partitioned

$$\begin{array}{c} \\ \hat{X} = \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} \hat{X}_1 & \hat{X}_2 \end{pmatrix} \end{array}, \quad \hat{X}_i = XW_i, \quad i = 1, 2.$$

Let $\phi_1, \ldots, \phi_{n-k}$ be the singular values of $\hat{X}_2$. By Corollary 6.4,

$$\|\hat{X}_2\|_2 = \phi_1 \geq \psi_{k+1}. \tag{6.24}$$

Thus $\|X - Y\|_2 \geq \psi_{k+1}$, which proves (6.22).

To show (6.23), observe that

$$\|X - Y\|_F^2 = \|(X - Y)W\|_F^2 = \|(X - Y)W_1\|_F^2 + \|(X - Y)W_2\|_F^2$$
$$\geq \|\hat{X}_2\|_F^2 = \|(\phi_1, \ldots, \phi_{n-k})^T\|_2^2.$$

Again, Corollary 6.4 yields the bound

$$\|(\phi_1, \ldots, \phi_{n-k})^T\|_2 \geq \|(\psi_{k+1}, \ldots, \psi_n)^T\|_2.$$

Thus we obtain (6.23). □

Example 6.1 Consider $X \in \mathrm{R}^{5\times 5}$ given by

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 5.000000e-01 & 2.886751e-01 & 0 & 0 & 0 \\ 3.333333e-01 & 2.886751e-01 & 7.453560e-02 & 0 & 0 \\ 2.500000e-01 & 2.598076e-01 & 1.118034e-01 & 1.889822e-02 & 0 \\ 2.000000e-01 & 2.309401e-01 & 1.277753e-01 & 3.779645e-02 & 4.761905e-03 \end{pmatrix}.$$

It has the SVD

$$X = U\Psi V^T$$

where

$$U = \begin{pmatrix} 7.678547e-01 & -6.018715e-01 & 2.142136e-01 & -4.716181e-02 & 6.173863e-03 \\ 4.457911e-01 & 2.759134e-01 & -7.241021e-01 & 4.326673e-01 & -1.166927e-01 \\ 3.215783e-01 & 4.248766e-01 & -1.204533e-01 & -6.673504e-01 & 5.061637e-01 \\ 2.534389e-01 & 4.439030e-01 & 3.095740e-01 & -2.330245e-01 & -7.671912e-01 \\ 2.098226e-01 & 4.290134e-01 & 5.651934e-01 & 5.575999e-01 & 3.762455e-01 \end{pmatrix},$$

$$\Psi = \mathrm{diag}(1.251819e{+}00, 4.566555e{-}01, 1.068059e{-}01, 1.748994e{-}02, 1.813265e{-}03),$$

$$V = \begin{pmatrix} 9.612151e-01 & -2.748479e-01 & 2.287927e-02 & -8.248572e-04 & 1.119485e-05 \\ 2.682675e-01 & 9.125182e-01 & -3.075360e-01 & 2.764268e-02 & -7.523761e-04 \\ 6.319965e-02 & 2.980707e-01 & 9.161588e-01 & -2.599660e-01 & 1.517757e-02 \\ 1.016129e-02 & 5.387904e-02 & 2.547866e-01 & 9.532078e-01 & -1.532075e-01 \\ 7.981629e-04 & 4.473659e-03 & 2.519897e-02 & 1.518151e-01 & 9.880772e-01 \end{pmatrix}.$$

Its Schmidt–Mirsky approxmation of rank three is given by

$$X_3 = \begin{pmatrix} 9.999993e-01 & 2.280969e-05 & -2.146048e-04 & 7.879754e-04 & 1.141644e-04 \\ 5.000062e-01 & 2.884658e-01 & 1.970459e-03 & -7.245652e-03 & -9.397627e-04 \\ 3.333237e-01 & 2.889985e-01 & 7.148737e-02 & 1.126638e-02 & 8.651083e-04 \\ 2.499967e-01 & 2.599192e-01 & 1.107650e-01 & 2.256997e-02 & 1.993270e-03 \\ 2.000080e-01 & 2.306710e-01 & 1.303002e-01 & 2.860492e-02 & 2.607246e-03 \end{pmatrix}$$

$$= U\,\mathrm{diag}(1.251819e+00, 4.566555e-01, 1.068059e-01, 0, 0)V^T.$$

The Schmidt–Mirsky theorem is probably the most ubiquitous theorem concerning the SVD. In least squares, it arises in any application where rank is an issue.

6.4. Bidiagonal Matrices and Related Tridiagonal Matrices

6.4.1. Definitions for Bidiagonal Matrices

An *upper bidiagonal* matrix $B \in \mathrm{R}^{n\times n}$ is one of the form

$$B = \begin{pmatrix} \gamma_1 & \phi_1 & \cdot & \cdot & \cdot & 0 \\ 0 & \gamma_2 & \phi_2 & \cdot & \cdot & \cdot \\ & & \vdots & \vdots & \vdots & \vdots \\ & & & \vdots & \vdots & \vdots \\ & & & & \gamma_{n-1} & \phi_{n-1} \\ & & & & 0 & \gamma_n \end{pmatrix}. \tag{6.25}$$

We denote it by the shorthand

$$B = \mathrm{ubidiag}(\gamma_1, \ldots, \gamma_n; \phi_1, \ldots, \phi_{n-1})$$

or

$$B = \mathrm{ubidiag}(\boldsymbol{\gamma}; \boldsymbol{\phi}), \qquad \begin{array}{l} \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^T \\ \boldsymbol{\phi} = (\phi_1, \ldots, \phi_{n-1})^T \end{array}.$$

Likewise a *lower bidiagonal* matrix $C \in \mathrm{R}^{n\times n}$ is just the transpose of an upper bidiagonal matrix. Thus we write

$$C = \mathrm{lbidiag}(\boldsymbol{\rho}; \boldsymbol{\zeta}), \qquad \begin{array}{l} \boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)^T \\ \boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{n-1})^T \end{array}$$

which also means

$$C = \mathrm{ubidiag}(\boldsymbol{\rho}; \boldsymbol{\zeta})^T.$$

To be able to calculate the singular values and vectors of B and to understand certain mathematical properties of them, we relate the singular value problem for B to equivalent eigenvalue problems.

6.4.2. Related Symmetric Tridiagonal Matrices

A symmetric tridiagonal matrix $T \in \mathrm{R}^{n\times n}$ has the form

$$T = \begin{pmatrix} \alpha_1 & \beta_1 & 0 & & \cdot & \cdot & \cdot \\ \beta_1 & \alpha_2 & \beta_2 & & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdot & \beta_{n_2} & \alpha_{n-1} & \beta_{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \beta_{n-1} & \alpha_n \end{pmatrix}. \tag{6.26}$$

We will use the shorthand

$$T = \text{tridiag}(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_{n-1}) = \text{tridiag}(\boldsymbol{\alpha}; \boldsymbol{\beta})$$

where

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T, \quad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_{n-1})^T. \tag{6.27}$$

There are three special tridiagonal matrices that are assoiciated with the singular value problem for bidiagonal matrices. For
$B = \text{ubidiag}(\gamma_1, \ldots, \gamma_n; \phi_1, \ldots, \phi_{n-1})$, we have

$$T_1 = B^T B, \tag{6.28}$$

which is

$$T_1 = \text{tridiag}(\gamma_1^2, \gamma_2^2 + \phi_1^2, \ldots, \gamma_n^2 + \phi_{n-1}^2; \gamma_1\phi_1, \ldots, \gamma_{n-1}\phi_{n-1}).$$

There is also

$$T_2 = B\,B^T = \text{tridiag}(\gamma_1^2 + \phi_1^2, \ldots, \gamma_{n-1}^2 + \phi_{n-1}^2, \gamma_n^2; \gamma_2\phi_1, \ldots, \gamma_n\phi_{n-1}). \tag{6.29}$$

Both T_1 in (6.28) and T_2 in (6.29) have the eigenvalues $\psi_1^2, \psi_2^2, \ldots, \psi_n^2$. Working with the these matrices has the same numerical disadvantages with regard to the singular value problem as for least squares problems.

The third tridiagonal matrix was first used by Golub and Kahan [12, 1965] and comes from considering

$$H = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}.$$

If we let P be the permutation matrix representing the perfect shuffle $1, n+1, 2, n+2, \ldots, n, 2n$ then the matrix $T_3 = PHP^T$ is the tridiagonal matrix

$$T_3 = \text{tridiag}(0, \ldots, 0; \gamma_1, \phi_1, \gamma_2, \phi_2, \ldots, \gamma_{n-1}, \phi_{n-1}, \gamma_n). \tag{6.30}$$

For this we will use the notation

$$T = \text{Btridiag}(B) = \text{Btridiag}(\boldsymbol{\gamma}; \boldsymbol{\phi}).$$

The matrix T_3 has the eigenvalues $\psi_1, \ldots, \psi_n, -\psi_1, \ldots, -\psi_n$ and both the left and right singular vectors can be recovered from its eigenvectors. However, it has double the dimension of the tridiagonal matrices in (6.28) and (6.29).

6.4.3. Norms and Condition Numbers of Bidiagonals

For any bidiagonal matrix $B = \text{ubidiag}(\gamma_1, \ldots, \gamma_n; \phi_1, \ldots, \phi_{n-1})$, there are $O(n)$ algorithms to compute

$$\kappa(B) = \|B\|\|B^{-1}\|$$

for any of the three norms $\|\cdot\|_1, \|\cdot\|_\infty$, and $\|\cdot\|_F$. We can obtain a tight bound for $\kappa(B)$ for the two–norm as well.

We take B to be upper bidiagonal as in (6.25) and for simplicity, let $\phi_0 = \phi_n = 0$. For lower bidiagonal matrices, we just consider the transpose. Without loss of generality, we may assume that $\gamma_k \neq 0$ for all k. Otherwise B is singular and $\|B^{-1}\|$ is not meaningful for any norm.

First, for any Hölder norm, we have

$$\|B\mathbf{e}_i\|_p = \left(|\gamma_i|^p + |\phi_{i-1}|^p\right)^{1/p}, \quad 1 \leq p < \infty, \tag{6.31}$$

$$\|B^T\mathbf{e}_i\|_p = \left(|\gamma_i|^p + |\phi_i|^p\right)^{1/p}, \quad 1 \leq p < \infty. \tag{6.32}$$

Thus

$$\|B\|_1 = \max_{1\leq i\leq n} |\gamma_i| + |\phi_{i-1}|, \tag{6.33}$$

$$\|B\|_\infty = \max_{1\leq i\leq n} |\gamma_i| + |\phi_{i-1}|, \tag{6.34}$$

and

$$\|B\|_F = \left(\|\boldsymbol{\gamma}\|_2^2 + \|\boldsymbol{\phi}\|_2^2\right)^{1/2} \tag{6.35}$$

are straightforward computations.

An upper bound for $\|B\|_2$ uses just the inequality

$$\|B\|_2 \leq \min\{\|B\|_F, \sqrt{\|B\|_1\|B\|_\infty}\}. \tag{6.36}$$

A lower bound for $\|B\|_2$ comes from

$$\|B\|_2 \geq \max\{\max_{1\leq i\leq n} \|B\mathbf{e}_i\|_2, \max_{1\leq i\leq n} \|B^T\mathbf{e}_i\|_2\}. \tag{6.37}$$

The upper and lower bounds in (6.36)–(6.37) are always within a factor of $\sqrt{2}$ of one another.

To compute $\|B^{-1}\|$ for $\|\cdot\|_1$, $\|\cdot\|_\infty$, and $\|\cdot\|_F$, consider the computations of

$$\|B^{-1}\mathbf{e}_k\|_p, \quad k = 1, \ldots, n.$$

From the upper triangular form of B, we may conclude that the solution of

$$B\mathbf{y}_k = \mathbf{e}_k$$

has the form

$$\mathbf{y}_k = \begin{matrix} k \\ n-k \end{matrix} \begin{pmatrix} \tilde{\mathbf{y}}_k \\ 0 \end{pmatrix}.$$

Let

$$r_k^{(p)} = \|B^{-1}\mathbf{e}_k\| = \|\tilde{\mathbf{y}}_k\|_p.$$

The following algorithm computes all of the $\tilde{\mathbf{y}}_k$:

$$\tilde{\mathbf{y}}_1 = (\gamma_1^{-1}),$$

$$\tilde{\mathbf{y}}_k = \gamma_k^{-1} \begin{pmatrix} -\phi_{k-1}\tilde{\mathbf{y}}_k \\ 1 \end{pmatrix}, \quad k = 2, \ldots, n.$$

Simply taking norms an algorithm for computing r_k^p is given by

$$r_1^{(p)} = |\gamma_1|^{-1}, \quad 1 \le p \le \infty, \tag{6.38}$$

$$r_k^{(p)} = (|\phi_{k-1} r_{k-1}^{(p)}|^p + 1)^{1/p} / |\gamma_k|. \tag{6.39}$$

For our purpose $p = 1, 2$, so the recurrence (6.42)–(6.43) computes all $r_k^{(p)}$ in about $5n$ flops for $p = 2$ and $3n$ flops for $p = 1$. Thus, $\|B^{-1}\|_1$ and $\|B^{-1}\|_F$ are recovered from

$$\|B^{-1}\|_1 = \max_{1 \le k \le n} r_k^{(1)}, \tag{6.40}$$

$$\|B^{-1}\|_F = \|(r_1^{(2)}, \ldots, r_n^{(2)})^T\|_2. \tag{6.41}$$

A recurrence may also be developed for

$$s_k^{(p)} = \|B^{-T}\mathbf{e}_k\|_p.$$

It is given by

$$s_n^{(p)} = |\gamma_n|^{-1}, \tag{6.42}$$

$$s_k^{(p)} = (|\phi_k s_{k+1}^{(p)}|^p + 1)^{1/p} / |\gamma_k|. \tag{6.43}$$

Equations (6.42)–(6.43) allow us to recover

$$\|B^{-1}\|_\infty = \|B^{-T}\|_1 = \max_{1 \le k \le n} s_k^{(1)}. \tag{6.44}$$

Upper and lower bounds for $\|B^{-1}\|_2$ are given by

$$\max_{1 \le k \le n} \max\{r_k^{(2)}, s_k^{(2)}\} \le \|B^{-1}\|_2 \le \min\{\|B^{-1}\|_F, \sqrt{\|B^{-1}\|_1 \|B^{-1}\|_\infty}\}. \tag{6.45}$$

The ratio between the upper and lower bounds in (6.45) can be no larger than $\sqrt{n}$.

Thus, the condition numbers of bidiagonal matrices can be computed in $O(n)$ flops in the one , infinity, and Frobenius norms. They can approximated to within a factor of $\sqrt{2n}$ in the two–norm. In the next section, we show how to make our bound on the condition number in the two–norm more precise.

6.4.4. Eigenvalues of Symmetric Tridiagonal Matrices

Symmetric tridiagonal matrices are very useful in eigenvalue computations and have useful special structure. Moreover, these matrices can be used to understand the SVD of bidiagonals.

We first consider the eigenvalue problem for the general symmetric tridiagonal matrix

$$T = \text{tridiag}(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_{n-1}) = \text{tridiag}(\boldsymbol{\alpha}; \boldsymbol{\beta})$$

where

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T, \quad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_{n-1})^T. \tag{6.46}$$

The matrix T is said to be *unreduced* if $\beta_i \neq 0$, $i = 1, \ldots, n-1$ and is *reduced* otherwise.

If T is reduced, that is, if $\beta_k = 0$ for some k then T can be decoupled into

$$T = \begin{array}{c} \\ k \\ n-k \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \end{array}.$$

The set of eigenvalues of T are simply the union of the sets of eigenvalues of T_1 and T_2. Likewise, λ is an eigenvalue of T_1, with eigenvector $\mathbf{x} \in \mathrm{R}^k$, then $\tilde{\mathbf{x}} = (\mathbf{x}^T, 0)^T \in \mathrm{R}^n$ is an eigenvector of T corresponding to λ. There is an analogous result for T_2. Thus, we may assume that T is unreduced.

There is a well–known recurrence for computing the characteristic polynomial of T. For $r = 1, \ldots, n$, let T_r be the matrix

$$T_r = \text{tridiag}(\boldsymbol{\alpha}(1\!:\!r); \boldsymbol{\beta}(1\!:\!r-1)). \tag{6.47}$$

This is called the $r \times r$ *leading principal submatrix* of T. These matrices can be given recursively by

$$T_1 = (\alpha_1),$$

$$T_r = \begin{array}{c} \\ r-1 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} r-1 & 1 \end{array} \\ \begin{pmatrix} T_{r-1} & \beta_{r-1}\mathbf{e}_{r-1} \\ \beta_{r-1}\mathbf{e}_{r-1}^T & \alpha_r \end{pmatrix} \end{array}, \quad r = 2, \ldots, n.$$

Let

$$p_r(\lambda) = \det(T_r - \lambda I)$$

be the *characteristic polynomial* of T_r. The use of expansion by minors yields the recurrence

$$p_0(\lambda) \equiv 1, \tag{6.48}$$

$$p_1(\lambda) = \alpha_1 - \lambda, \tag{6.49}$$

$$p_r(\lambda) = (\alpha_r - \lambda)p_{r-1}(\lambda) - \beta_{r-1}^2 p_{r-2}(\lambda). \tag{6.50}$$

Since

$$p_n(\lambda) = \det(T - \lambda I)$$

we can compute the characteristic polynomial of T at any point in about $5n$ flops. The equations (6.48)–(6.50) tell us much more about the eigenvalues of T.

First, we define two functions *polyval*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda, n)$ and *sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda, n)$.

Function 6.16 (Polyval)

function $pr =$ *polyval*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda)$
$n =$**length**$(\boldsymbol{\alpha})$;
% pr denotes $p_r(\lambda)$, $prm1$ denotes $p_{r-1}(\lambda)$
% $prm2$ denotes $p_{r-2}(\lambda)$.
$pr \leftarrow \alpha_1 - \lambda$; $prm1 \leftarrow 1$;
for $k = 2{:}\,n$
    $prm2 \leftarrow prm1$; $prm1 \leftarrow pr$;
    $pr \leftarrow (\alpha_k - \lambda) * prm1 - \beta_{k-1}^2 * prm2$
end for
end *polyval*

polyval merely evaluates the characteristic polynomial $p_n(\lambda)$.

Function 6.17 (Sign Count)

function $k =$*sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda)$
% pr denotes $p_r(\lambda)$, $prm1$ denotes $p_{r-1}(\lambda)$
% $prm2$ denotes $p_{r-2}(\lambda)$.
$k = 0$; $pr = \alpha_1 - \lambda$; $prm1 = 1$;
for $r = 2{:}\,n$
    $prm2 = prm1$; $prm1 = pr$;
    $pr = (\alpha_r - \lambda) * prm1 - \beta_{r-1}^2 * prm2$
    if $sign(pr) \neq$ sign$(prm1)$
        if $prm1 \neq 0$
            $k = k + 1$;
        end if
    end if
end for
end *sign_count*

The function *sign_count* counts the number of sign changes in the sequence

$$\{p_0(\lambda), p_1(\lambda), \ldots, p_n(\lambda)\}. \tag{6.51}$$

Since sign(0) = 0, it counts $p_{r-1}(\lambda) \neq 0$ with $p_r(\lambda) = 0$ as a sign change (but not $p_{r-1}(\lambda) = 0$ with $p_r(\lambda) \neq 0$). This sign count is valuable for finding individual eigenvalues.

Theorem 6.18 (Sturm Sequence Theorem) *Let $T \in \mathrm{R}^{n\times n}$ be given by $T = \mathrm{tridiag}(\boldsymbol{\alpha}; \boldsymbol{\beta})$ where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are as in* (6.46) *and T is unreduced. Let T_r be as in* (6.47). *Then for $r = 2, \ldots, n$,*

$$\lambda_{i+1}(T_r) < \lambda_i(T_{r-1}) < \lambda_i(T_r), \quad i = 1, \ldots, r-1. \tag{6.52}$$

*Moreover, $k =$*sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda)$ if exactly k eigenvalues of T are less than λ.*

Proof. First, we show (6.52). Suppose that

$$\mu = \lambda_i(T_r) = \lambda_i(T_{r-1}).$$

Then

$$p_r(\mu) = p_{r-1}(\mu) = 0.$$

By (6.50), we then have

$$-\beta_{r-1}^2 p_{r-2}(\mu) = 0.$$

Since $\beta_{r-1} \neq 0$, then $p_{r-2}(\mu) = 0$. If $\beta_i \neq 0, i = 1, \ldots, r-1$, then $p_0(\mu) = \ldots = p_r(\mu) = 0$. Since $p_0(\mu) = 1$, this is a contradiction. Thus, $\beta_i = 0$ for some i. The same argument holds if $\mu = \lambda_i(T_r) = \lambda_{i+1}(T_{r-1})$.

An induction argument works for the sign change property. For $n = 1$, clearly λ is less than or equal to the only eigenvalue of T (namely α_1) only if $p_1(\lambda) < 0$, which is a sign change from $p_0(\mu) = 1$.

Suppose that the theorem holds for $n-1$. Thus if *sign_count*$(\boldsymbol{\alpha}(1\colon n-1), \boldsymbol{\beta}(1\colon n-2), \lambda) = k$, then k eigenvalues of $T(1\colon n-1, 1\colon n-1)$ are less than or equal to λ. Then *sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda)$ is either k or $k+1$.

There are three cases. The first case is when $p_{n-1}(\lambda) = 0$. Since *sign_count*$(\boldsymbol{\alpha}(1\colon n-1), \boldsymbol{\beta}(1\colon n-2), \lambda) = k$, then λ is the kth smallest eigenvalue of T_{n-1}. From (6.52), we have that $p_n(\lambda) \neq 0$ and λ is larger than exactly k eigenvalues of T. From the algorithm *sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) = k$.

The second case is when $p_n(\lambda) = 0$. This counts as sign change and Function 6.17 yields *sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) = k+1$. Again (6.52) states that λ must be the $(k+1)st$ smallest eigenvalue of T.

The third case is when neither $p_{n-1}(\lambda)$ nor $p_n(\lambda)$ is zero. Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of T and let $\mu_1, \ldots, \mu_{n-1}$ denote the eigenvalues of T_{n-1}. The polynomials $p_n(\lambda)$ and $p_{n-1}(\lambda)$ may be written

$$p_n(\lambda) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda), \tag{6.53}$$

$$p_{n-1}(\lambda) = (\mu_1 - \lambda) \cdots (\mu_{n-1} - \lambda). \tag{6.54}$$

From (6.52), λ is greater than or equal to either exactly k eigenvalues of T or exactly $k+1$. From the form of $p_{n-1}(\lambda)$, we have that

$$\text{sign}(p_{n-1}(\lambda)) = (-1)^k,$$

since k factors in (6.54) will be negative. By the same reasoning,

$$\text{sign}(p_n(\lambda)) = (-1)^{k+1},$$

if and only if $k+1$ eigenvalues of T are less than λ. Thus, *sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda,)$= $k+1$ if and only if $k+1$ eigenvalues of T are less than λ. □

The sequence $\{p_0(\lambda), p_1(\lambda), \ldots, p_n(\lambda)\}$ is sometimes subject to overflow. Thus it is common to compute the nonlinear recurrence for

$$q_r(\lambda) = \frac{p_r(\lambda)}{p_{r-1}(\lambda)}.$$

That recurrence is given by

$$q_1(\lambda) = \alpha_1 - \lambda, \quad q_r(\lambda) = \alpha_r - \lambda - \frac{\beta_r^2}{q_{r-1}(\lambda)} \quad r = 2, \ldots, n.$$

Here the number of eigenvalues smaller than λ will be the number of negative terms in the sequence $\{q_r(\lambda)\}$. From this point, however, we will discuss the sequence $\{p_r(\lambda)\}$.

The above theorem leads to a procedure to produce an interval $[\mu_{min}, \mu_{max}]$ that contains λ_k, the *kth* eigenvalue of T. This procedure simply uses bisection to narrow the interval to be one that contain only the eigenvalue λ_k and then narrows the interval further that brackets the eigenvalue in an interval of a specified size *tol*. It is assume that the use will supply routines that give upper and lower bounds on the eigenvalue.

Procedure 6.19

function $[\mu_{min}, \mu_{max}]$ =*bisect_eig*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k, tol,)$
% *bisecteig* produces an interval $[\mu_{min}, \mu_{max}]$ that
% contains the *kth* eigenvalue of T = tridiag$(\boldsymbol{\alpha}, \boldsymbol{\beta})$

% *bisect_eig*calls the procedure *brack_eig*which computes
% an initial interval $[\mu_{min}, \mu_{max}]$ that contains
%the *kth* eigenvalue and no others.

$[\mu_{min}, \mu_{max}] =$*brack_eig*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k)$; $n =$**length**$(\boldsymbol{\alpha})$;
$pval_{min} =$*polyval*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k, \mu_{min})$; $pval_{max} =$*polyval*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k, \mu_{max})$;
if $pval_{min} = 0$ **then**
  $\mu_{max} = \mu_{min}$;
else
  if $pval_{max} = 0$ **then**
    $\mu_{min} = \mu_{max}$;
  end if
end if
while $\mu_{max} - \mu_{min} > tol$ **do**
  $\mu_{mid} = (\mu_{max} + \mu_{min})/2$;
  $poly_{mid} =$*polyval*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k, \mu_{mid})$;
  if $\mu_{mid} = 0$ **then**
    $\mu_{max} = \mu_{min} \leftarrow \mu_{mid}$
  else
    if $poly_{mid} * poly_{max} < 0$ **then**
      $\mu_{min} = \mu_{mid}$
      $poly_{min} = poly_{mid}$
    else
      $\mu_{max} = \mu_{mid}$
    end if
  end if
end while
end *bisect_eig*

function $[\mu_{min}, \mu_{max}] =$ *brack_eig*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k)$
Choose $\mu_{max} \geq \lambda_k$
  $\mu_{min} \leq \lambda_k$
  $\mu_{mid} = (\mu_{max} + \mu_{min})/2$;
  $k_{min} =$*sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k, \mu_{min})$; $k_{max} =$*sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k, \mu_{max})$;
  $k_{mid} =$*sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k, \mu_{mid})$;
while $k_{max} > n - k + 1$ **or** $k_{min} < n - k$ **do**
  if $k_{mid} \geq n - k$ **then**
    $\mu_{max} = \mu_{mid}$; $k_{max} = k_{mid}$
  else
    $\mu_{min} = \mu_{mid}$; $k_{min} = k_{mid}$
  end
  $\mu_{mid} = (\mu_{max} + \mu_{min})/2$; $k_{mid} =$*sign_count*$(\boldsymbol{\alpha}, \boldsymbol{\beta}, k, \mu_{mid})$;
end while
end *bracket_eig*

We now give some techniques for supplying initial upper and lower bounds μ_{min} and μ_{max} to *brack_eig*.

From [**?**, 1996,p.440], for an arbitrary symmetric matrix we can choose

$$\mu_{max} = \min_{1 \le j \le n} \alpha_j + |\beta_j| + |\beta_{j-1}|$$

and

$$\mu_{min} = \min_{1 \le j \le n} \alpha_j - |\beta_j| + |\beta_{j-1}|$$

with the convention that $\beta_0 = 0$. When looking for singular values the bound μ_{min} is not of much use, since it can be negative for all of $T_i, i = 1, 2, 3$ in equations (6.28)–(6.30). Moreover, the upper bound μ_{max} can also be made tighter if we know we are looking for singular values.

To get the most accurate singular values of B in (6.25), we take $T = T_3$ from (6.30). This is a tridiagonal of dimension $2n$. The recurrence (6.48)–(6.50) becomes

$$p_0(\lambda) = 1, \quad p_1(\lambda) = -\lambda, \tag{6.55}$$

$$p_{2r}(\lambda) = -\lambda p_{2r-1}(\lambda) - \gamma_r^2 p_{2r-2}(\lambda), \quad r = 1, \ldots, n, \tag{6.56}$$

$$p_{2r+1}(\lambda) = -\lambda p_{2r}(\lambda) - \phi_{r+1}^2 p_{2r}(\lambda), \quad r = 1, \ldots, n-1, \tag{6.57}$$

We need only the positive eigenvalues of T, since they correspond to the singular values of B.

Clear choices for μ_{min} and μ_{max} are

$$\mu_{min} = \max\{\|B^{-1}\|_F^{-1}, (\|B^{-1}\|_1 \|B^{-1}\|_\infty)^{-1/2}\}, \quad \mu_{max} = \max\{\|B\|_F, (\|B\|_1 \|B\|_\infty)^{1/2}\}.$$

If we want the largest singular value from this procedure, we can specify a better value of μ_{min}. The value

$$\mu_{min} = \max_{1 \le i \le n} \max\{\|B\mathbf{e}_i\|_2, \|B^T\mathbf{e}_i\|_2\}$$

is a suitable lower bound for the largest singular value of B.

If the smallest singular value is desired, a better value of μ_{max} is

$$\mu_{max} = \min_{1 \le i \le n} \min\{\|B^{-1}\mathbf{e}_i\|_2^{-1}, \|B^{-T}\mathbf{e}_i\|_2^{-1}\}.$$

If our goal is to estimate $\|B\|_2$ and $\|B^{-1}\|_2$ to only a few digits, a few iterations of *bisect_eig* procedure applied to (6.48)–(6.50) will usually be satisfactory. If our goal is to estimate $\|B\|_2$ and $\|B^{-1}\|_2$ to a few significant digits, then a few iterations of **bisect_eig** applied to (6.55)–(6.57) is likely to be satisfactory. As a method to get all of the singular values of a bidiagonal matrix, it is very reliable and can compute them with almost all digits correct [8, 2, **?**]. Other faster methods for finding the singular values of bidiagonals will be discussed in Chapter **??**.

6.4.5. Perturbations Theory for the Bidiagonal SVD

A very clever bound on componentwise perturbations in the singular values of bidiagonal and arrow matrices follows from this characterization. This lemma was originally proven by Kahan [13, 1966] and was first published by Demmel and Kahan [**?**, 1990]. The original proof used a clever Sturm sequence argument. The proof given here is similar to that in Barlow and Demmel [2, 1990] is based upon Lemma **??**.

Theorem 6.20 *Let* $B = \text{ubidiag}(\gamma(1{:}n); \phi(2{:}n)) \in \mathrm{R}^{n\times n}$, *let* $\tilde{B} \equiv B + \delta B = \text{ubidiag}(\tilde{\gamma}(1{:}n); \tilde{\phi}(2{:}n)) \in \mathrm{R}^{n\times n}$, *and let* $\zeta \geq 1$. *Using the convention that* $0/0 = 1$, *if*

$$\frac{1}{\zeta} \leq \frac{\tilde{\gamma}_j}{\gamma_j}, \frac{\tilde{\phi}_i}{\phi_i} \leq \zeta, \qquad \begin{matrix} i = 2, \ldots, n \\ j = 1, 2, \ldots, n, \end{matrix}$$

then

$$\frac{1}{\zeta^{2n-1}} \leq \frac{\psi_j(\tilde{B})}{\psi_j(B)} \leq \zeta^{2n-1}, \quad j = 1, 2, \ldots, n.$$

Proof. The hypothesis of this lemma may be written

$$\tilde{\gamma}_j = \gamma_j \alpha_j, \quad \tilde{\phi}_i = \phi_i \beta_i$$

where

$$\frac{1}{\zeta} \leq \alpha_j, \beta_i \leq \zeta, \qquad \begin{matrix} i = 2, \ldots, n \\ j = 1, 2, \ldots, n, \end{matrix} .$$

Let $B_\ell, \ell = 0, 1, 2, \ldots, 2n-1$ be the sequence of matrices such that

$$B_0 = B,$$

$$B_{2j-1} = \text{bidiag}(\tilde{\gamma}_1, \ldots, \tilde{\gamma}_j, \gamma_j, \ldots, \gamma_n; \tilde{\phi}_2, \ldots, \tilde{\phi}_j, \phi_{j+1}, \ldots, \phi_n)$$
$$j = 1, \ldots, n$$
$$B_{2j} = \text{bidiag}(\tilde{\gamma}_1, \ldots, \tilde{\gamma}_j, \gamma_j, \ldots, \gamma_n; \tilde{\phi}_2, \ldots, \tilde{\phi}_j, \tilde{\phi}_{j+1}, \phi_{j+2}, \ldots, \phi_n)$$
$$j = 1, \ldots, n-1$$

Thus,

$$B_{2j-1} = \text{diag}(I_{j-1}, \alpha_j I_{n-j+1}) B_{2j-2} \, \text{diag}(I_j, \alpha_j^{-1} I_{n-j}).$$

$$B_{2j} = \text{diag}(I_j, \beta_j^{-1} I_{n-j+1}) B_{2j-1} \, \text{diag}(I_j, \beta_j I_{n-j}).$$

A simple application of Corollary 6.7 yields

$$\psi_k(B_{2j-1}) \leq \psi_1[(\text{diag}(I_{j-1}, \alpha_j I_{n-j+1})]\psi_1[(I_j, \alpha_j^{-1})]\psi_k(B_{2j-2}) \leq \zeta\psi_k(B_{2j-2}),$$

$$\psi_k(B_{2j}) \le \psi_1[(\mathrm{diag}(I_j, \beta_j^{-1} I_{n-j+1})]\psi_1[(I_j, \alpha_j^{-1})]\psi_k(B_{2j-2}) \le \zeta\psi_k(B_{2j-2}),$$

and also produces the lower bounds

$$\psi_k(B_{2j-1}) \ge \psi_n[(\mathrm{diag}(I_{j-1}, \alpha_j I_{n-j+1})]\psi_n[(I_j, \alpha_j^{-1})]\psi_k(B_{2j-2}) \ge \zeta^{-1}\psi_k(B_{2j-2}),$$

$$\psi_k(B_{2j}) \le \psi_n[(\mathrm{diag}(I_j, \beta_j^{-1} I_{n-j+1})]\psi_n[(I_j, \alpha_j^{-1})]\psi_k(B_{2j-2}) \ge \zeta^{-1}\psi_k(B_{2j-2}).$$

An induction argument shows that

$$\zeta^{-(2n-1)}\psi_k(B) \ge \psi_k(B + \delta B) = \psi_k(B_{2n-1}) \le \zeta^{2n-1}\psi_k(B) \tag{6.58}$$

which is the desired result. □

The following theorem gives two criteria for setting a superdiagonal element ϕ_k to zero, one based on the "absolute change" in the singular values $|\tilde{\psi}_k - \psi_k|$, the other in the "relative change $|\tilde{\psi}_k - \psi_k|/\psi_k$.

Theorem 6.21 *Let* $B = \mathrm{ubidiag}(\boldsymbol{\gamma}(1{:}n); \boldsymbol{\phi}(2{:}n))$ *with* $\gamma_k \ne 0, k = 1, \ldots, n$. *For fixed* $j \in \{2, \ldots, n\}$ *let* $B_1 = \mathrm{ubidiag}(\boldsymbol{\gamma}(1{:}j-1); \boldsymbol{\phi}(2{:}j-1))$, *let* $B_2 = \mathrm{ubidiag}(\boldsymbol{\gamma}(j{:}n); \boldsymbol{\phi}(j+1{:}n)$ *and let* $\tilde{B} = \mathrm{diag}(B_1, B_2)$, *that is,* $\tilde{B}$ *is* B *with* ϕ_j *set to zero. Then*

$$\sum_{k=1}^{n} [\psi_k(B) - \psi_k(\tilde{B})]^2 \le \phi_j^2. \tag{6.59}$$

and

$$\frac{|\psi_k(B) - \psi_k(\tilde{B})|}{\psi_k(B)} \le |\phi_j| \min\{\|B_1^{-1}\mathbf{e}_{j-1}\|_2, \|B_2^{-T}\mathbf{e}_1\|_2\} \tag{6.60}$$

Proof. Equation (6.59) is just Theorem **??** applied to the rank-one perturbation

$$\delta B = B - \tilde{B} = \phi_j \mathbf{e}_{j-1}\mathbf{e}_j^T.$$

To get (6.60), we note that from Lemma 6.11

$$\frac{|\psi_k - \tilde{\psi}_k|}{\psi_k} \le \eta = \|(\delta B)B^{-1}\|_2 = |\phi_j|\|\mathbf{e}_1^T B_2^{-1}\|_2$$

$$= |\phi_j|\|B_2^{-T}\mathbf{e}_1\|_2.$$

Applying Lemma 6.11 to B^T and $(\delta B)^T$ yields

$$\frac{|\psi_k - \tilde{\psi}_k|}{\psi_k} \le |\phi_j|\|B_1^{-1}\mathbf{e}_{j-1}\|_2.$$

Taking the minimum of the last two bounds yields (6.60). □

Many methods for finding the singular values of B rely upon similarity transformations that are a designed to set off-diagonal elements to zero. If our criterion for setting ϕ_j to zero is so that

$$\sum_{k=1}^{n}[\psi_k(B) - \psi_k(\tilde{B})]^2 \leq \epsilon^2$$

then set ϕ_j to zero if $|\phi_j| \leq \epsilon$. If we want

$$\frac{|\psi_k(B) - \psi_k(\tilde{B})|}{\psi_k(B)} \leq \epsilon$$

then set ϕ_j to zero if

$$|\phi_j| \min\{\|B_1^{-1}\mathbf{e}_{j-1}\|_2, \|B_2^{-T}\mathbf{e}_1\|_2\} \leq \epsilon.$$

The latter criterion is used by LAPACK [1, 1992] routines for computing the bidiagonal SVD.

6.5. Rank–One Modifications of Eigenvalue and Singular Value Problems

In least squares computation, the addition and deletion of rows or columns are common. Thus it is important to quantify the effect of such additions and deletions on the singular values. First, we discuss the change in the eigenvalues of a symmetric matrix after adding a row and a column.

Consider the symmetric matrix $A \in \mathrm{R}^{(n+1)\times(n+1)}$ given by

$$A = \begin{array}{c} \\ n \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} n & 1 \end{array} \\ \begin{pmatrix} \bar{A} & \mathbf{y} \\ \mathbf{y}^T & \alpha \end{pmatrix} \end{array}. \tag{6.61}$$

Suppose that $\bar{A}$ has the eigendecomposition

$$\bar{A} = \bar{Q}D\bar{Q}^T \tag{6.62}$$

where $\bar{Q} \in \mathrm{R}^{n\times n}$ is orthogonal and

$$D = \mathrm{diag}(d_1, \ldots, d_n) \tag{6.63}$$

is the eigenvalue matrix of $\bar{A}$ with $d_1 \geq \cdots \geq d_n$.

If we let $\mathbf{z} = \bar{Q}^T\mathbf{y}$, then the *arrow matrix* T given by

$$T = \begin{pmatrix} \bar{Q}^T & 0 \\ 0 & 1 \end{pmatrix} A \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} D & \mathbf{z} \\ \mathbf{z}^T & \alpha \end{pmatrix} \tag{6.64}$$

has the same eigenvalues as A. Special cases regarding T are considered in the following two lemmas.

Lemma 6.22 *Let T be as in (6.64). Suppose that for some $k \geq 1$ and $i \leq n-k$, we have*

$$d_i = d_{i+1} = \ldots = d_{i+k}.$$

Then there is an orthogonal transformation $H \in \mathrm{R}^{(n+1)\times(n+1)}$ such that

$$\hat{T} = HTH^T$$

where

$$\hat{T} = \begin{pmatrix} D & \hat{\mathbf{z}} \\ \hat{\mathbf{z}}^T & \alpha \end{pmatrix}, \quad \hat{z}_j = \begin{cases} z_j, & j \neq i, \ldots, i+k \\ \pm\|\mathbf{z}(i\!:\!i+k)\|_2 & j = i \\ 0 & j = i+1, \ldots, i+k, \end{cases} \tag{6.65}$$

and $\lambda_{i+1} = d_{i+1} = \ldots = d_{i+k}$ is an eigenvalue of T of multiplicity at least k.

Proof. Let $\tilde{H} \in \mathrm{R}^{k+1\times k+1}$ be a Householder transformation such that

$$\tilde{H}\mathbf{z}(i\!:\!i+k) = \pm\|\mathbf{z}(i\!:\!i+k)\|_2\mathbf{e}_1.$$

Then the orthogonal matrix

$$H = \mathrm{diag}(I_{i-1}, \tilde{H}, I_{n-k-1})$$

produces $\hat{T}$ in (6.65). By inspection, $\lambda = d_{i+1}$ is an eigenvalue of $\hat{T}$ with multiplicity k, the matrix $E_{i,k} = (\mathbf{e}_{i+1}, \ldots, \mathbf{e}_{i+k})$ is basis for the corresponding invariant subspace of $\hat{T}$, and $HE_{i,k}$ is the corresponding invariant subspace of T. □

There is a very precise relationship between the eigenvalues of T, $\lambda_1, \ldots, \lambda_{n+1}$ and components of $\mathbf{z}$. Versions of this result date back to Löwner [**?**, 1934]. This particular theorem is due to Boley and Golub [4, 1977].

Theorem 6.23 *Let $D = \mathrm{diag}(d_1, \ldots, d_n)$ satisfy $d_1 > \cdots > d_n$ and let $\lambda_1, \ldots, \lambda_{n+1}$ be any numbers satisfying the interlacing property*

$$\lambda_i \geq d_i \geq \lambda_{i+1}, \quad i = 1, \ldots, n.$$

Then $\lambda_1, \ldots, \lambda_{n+1}$ are eigenvalues of any arrow matrix T defined by $\mathbf{z} \in \mathrm{R}^n$ and $\alpha \in \mathrm{R}$ such that

$$z_i^2 = (d_i - \lambda_{n+1})\frac{\prod_{j=1}^{n}(\lambda_j - d_i)}{\prod_{j\neq i}(d_j - d_i)} \tag{6.66}$$

and

$$\alpha = \lambda_{n+1} + \sum_{j=1}^{n}(\lambda_j - d_j). \tag{6.67}$$

Proof. Let

$$\hat{D} = \mathrm{diag}(d_1, \ldots, d_{i-1}, d_{i+1}, \ldots, d_n), \quad \hat{z} = \mathrm{diag}(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n).$$

Let P be a permutation matrix that maps rows $1, 2, \ldots, n$ to $1, 2, \ldots, i-1, i+1, \ldots, n, i$. Then

$$\hat{T} = P(T - d_i I)P^T = \begin{array}{c} \\ n-1 \\ 1 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} n-1 & 1 & 1 \end{array} \\ \begin{pmatrix} \hat{D} - d_i I & \hat{\mathbf{z}} & 0 \\ \hat{\mathbf{z}}^T & \alpha - d_i & z_i \\ 0 & z_i & 0 \end{pmatrix} \end{array}. \tag{6.68}$$

Since the $d_1 > \cdots > d_n$, $D - d_i I$ is nonsingular, so $\hat{T}$ may be factored into the L–R factorization of $\hat{T}$ in (6.68) is given by

$$\hat{T} = \begin{pmatrix} I_{n-1} & 0 & 0 \\ \hat{\mathbf{z}}^T(\hat{D} - d_i I)^{-1} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{D} - d_i I & \hat{\mathbf{z}} & 0 \\ 0 & \hat{f}(d_i) & z_i \\ 0 & z_i & 0 \end{pmatrix},$$

where $\hat{f}(d_i) = \alpha - \lambda - \mathbf{z}^T(\hat{D} - d_i I)^{-1}\hat{\mathbf{z}}$. Taking determinants, we have that the characteristic polynomial $p(\lambda)$ evaluated at d_i is given by

$$p(d_i) = \det(T - d_i I) = \det(\hat{D} - d_i I) \det \begin{pmatrix} \hat{f}(d_i) & z_i \\ z_i & 0 \end{pmatrix} = -\det(\hat{D} - d_i I) z_i^2 = -z_i^2 \prod_{j \neq i} (d_j - d_i).$$

But, $p(d_i)$ is also given by

$$p(d_i) = \prod_{j=1}^{n+1} (\lambda_j - d_i).$$

Thus,

$$z_i^2 = -\frac{p(d_i)}{\prod_{j \neq i}(d_j - d_i)} = (d_i - \lambda_{n+1}) \frac{\prod_{j=1}^{n}(\lambda_j - d_i)}{\prod_{j \neq i}(d_j - d_i)}$$

That establishes (6.66).

The expression for α in (6.67) comes from the fact that

$$\sum_{j=1}^{n+1} t_{jj} = \sum_{j=1}^{n+1} \lambda_j.$$

Thus

$$\sum_{j=1}^{n} d_j + \alpha = \sum_{j=1}^{n+1} \lambda_j.$$

Solving for α yields (6.67). □

The Cauchy Interlace Theorem (Theorem 6.3) states that $\lambda_i \geq d_i \geq \lambda_{i+1}$ for $i = 1, \ldots, n$. The next proposition tells us exactly when that inequality is strict.

Corollary 6.24 *Let T have the form (6.64) with $d_1 > \cdots > d_n$. For $i = 1, 2 \ldots, n$, d_i is an eigenvalue of T if and only if $z_i = 0$ and $\mathbf{e}_i$ is a corresponding eigenvector.*

Proof. It is obvious from the form of T that if $z_i = 0$ then $(d_i, \mathbf{e}_i)$ is an eigenpair. Moreover, if d_i is equal to any eigenvalue of T then from (6.66), we have $z_i = 0$. □

Lemma 6.22 and Corollary 6.24 allow us to assume without loss of generality that $d_1 > \cdots > d_n$, that for all $i = 1, 2, \ldots, n$, $z_i \neq 0$ and d_i is not an eigenvalue of T.

Using these two assumptions, we can write an expression for the characteristic polynomial of T. If λ is an eigenvalue of T then $D - \lambda I$ is nonsingular. Thus, the L–R factorization of T is

$$T - \lambda I = \begin{pmatrix} I_n & 0 \\ \mathbf{z}^T (D - \lambda I)^{-1} & 1 \end{pmatrix} \begin{pmatrix} D - \lambda I & \mathbf{z} \\ 0 & f(\lambda) \end{pmatrix}. \tag{6.69}$$

where

$$f(\lambda) = \alpha - \lambda - \mathbf{z}^T (D - \lambda I)^{-1} \mathbf{z}. \tag{6.70}$$

Using (6.69), we may infer that the characteristic polynomial $p(\lambda)$ is

$$p(\lambda) = \det(T - \lambda I) = \det(D - \lambda I) f(\lambda)$$

so that $p(\lambda) = 0$ if and only if $f(\lambda) = 0$.

Thus finding the eigenvalues of T is equivalent to finding the zeros of $f(\lambda)$ in (6.70). The eigenvalues of T can be recovered using rootfinding procedures as given in §??.

The function $f(\cdot)$ defined by (6.70) is called a spectral function or Pick function. The equation $f(\lambda) = 0$ is called a secular equation.

Once the eigenvalue λ_i is found, its corresponding eigenvector $\mathbf{v}_i$ is given by

$$\mathbf{v}_i = \frac{1}{|f'(\lambda_i)|^{-1/2}} \begin{pmatrix} (D - \lambda_i I)^{-1} \mathbf{z} \\ -1 \end{pmatrix}, \tag{6.71}$$

where $f'(\lambda) = -1 - \mathbf{z}^T (D - \lambda I)^{-2} \mathbf{z}$ is the first derivative of the spectral function $f(\lambda)$.

Lemma 6.25 *Let T have the form (6.64) with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{n+1}$. Then*

$$\lambda_1 \leq \frac{1}{2}[\alpha + d_1 + \left((\alpha - d_1)^2 + 4\|\mathbf{z}\|_2^2\right)^{1/2}], \tag{6.72}$$

$$\lambda_{n+1} \geq \frac{1}{2}[\alpha + d_n - \left((\alpha - d_n)^2 + 4\|\mathbf{z}\|_2^2\right)^{1/2}]. \tag{6.73}$$

Proof. First, we assume that $\lambda_1 > d_1$, otherwise (6.72) is a trivial consequence of the interlace theorem. Let $\lambda \geq \lambda_1$ and let $f(\lambda)$ be given by (6.70), since $D - \lambda I$ is nonsingular, we can factor $T - \lambda I$ into

$$T - \lambda I = SJS^T$$

where

$$S = \begin{array}{c} \\ n \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} n & 1 \end{array} \\ \begin{pmatrix} I & 0 \\ \mathbf{z}^T(D - \lambda I)^{-1} & 1 \end{pmatrix} \end{array}, \quad J = \operatorname{diag}(D - \lambda I, f(\lambda)).$$

All of the eigenvalues of $T - \lambda I$ are less than or equal to zero. By Theorem 6.5, the same must be true of J, thus $f(\lambda) \leq 0$.

Thus

$$f(\lambda) = \alpha - \lambda - \mathbf{z}^T(D - \lambda I)^{-1}\mathbf{z} \leq 0.$$

For $\lambda > d_1$, we have that

$$f(\lambda) \leq \alpha - \lambda - \mathbf{z}^T\mathbf{z}/(d_1 - \lambda).$$

Hence, for $\lambda \geq d_1$, if

$$\alpha - \lambda - \mathbf{z}^T\mathbf{z}/(d_1 - \lambda) \leq 0 \tag{6.74}$$

so is $f(\lambda)$. Equality holds for (6.74) when

$$\lambda = \frac{1}{2}[\alpha + d_1 + \left((\alpha - d_1)^2 + 4\|\mathbf{z}\|_2^2\right)^{1/2}].$$

Thus we have (6.72). Equation (6.73) results from applying (6.72) to $-T$. □

More accurate bounds on the roots of $f(\lambda)$ are possible. See Melman [**?**, 1999].

The applications in least squares come from adding or rows to the matrices defining the least squares problem. We now show the problem of adding a column and how it leads to two different arrow matrix eigenvalue problems.

Let $C \in \mathrm{R}^{m\times(n+1)}$, $X \in \mathrm{R}^{m\times n}$, and $\mathbf{b} \in \mathrm{R}^m$ satisfy

$$C = \begin{pmatrix} X & \mathbf{b} \end{pmatrix}. \tag{6.75}$$

We assume that $m \geq n$ and that X has the known SVD

$$X = U \begin{pmatrix} \Psi \\ 0 \end{pmatrix} V^T, \qquad \begin{matrix} U \in \mathrm{R}^{m\times m} \\ V \in \mathrm{R}^{n\times n} \end{matrix}$$

$$\begin{matrix} & n & m-n \\ U = & (U_1 & U_2 \;), \end{matrix} \quad \Psi = \mathrm{diag}(\Psi_1, \ldots, \Psi_n).$$

We can then reduce C to

$$\bar{F} = U^T C \begin{pmatrix} V & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \Psi & \mathbf{g} \\ 0 & \mathbf{h} \end{pmatrix} \tag{6.76}$$

Since for any orthogonal matrix $H \in \mathrm{R}^{(m-n)\times(m-n)}$, $\tilde{U}_2 = U_2 H$ also yields the SVD of F, we can choose U_2 such that $\mathbf{h} = U_2^T \mathbf{b} = \rho \mathbf{e}_1$, $\rho \in \mathrm{R}$.

Thus we may consider the SVD of the $(n+1) \times (n+1)$ matrix

$$F = \begin{pmatrix} \Psi & \mathbf{g} \\ 0 & \rho \end{pmatrix} = Q \Phi P^T \tag{6.77}$$

where

$$Q = (\mathbf{q}_1, \ldots, \mathbf{q}_{n+1}), \quad P = (\mathbf{p}_1, \ldots, \mathbf{p}_{n+1}), \tag{6.78}$$

are orthogonal and

$$\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_{n+1}), \quad \phi_1 \geq \cdots \geq \phi_{n+1} \tag{6.79}$$

is matrix of singular values.

The SVD of F can be recovered one of two arrow matrix eigenvalue problems. The first is to compute

$$T = F^T F = \begin{pmatrix} \Psi^2 & \Psi \mathbf{g} \\ \mathbf{g}^T \Psi & \rho^2 + \mathbf{g}^T \mathbf{g} \end{pmatrix}$$

which is simply (6.64) with $D = \Psi^2$,$\mathbf{z} = \Psi \mathbf{g}$, and $\alpha = \rho^2 + \mathbf{g}^T \mathbf{g}$.

Lemmas **??** and **??** generate the following corollaries for F.

Corollary 6.26 *Let $F \in \mathrm{R}^{(n+1)\times(n+1)}$ be as in (6.77) with the SVD Let i and k be positive integers such that $i + k \leq n$. If $\psi_i = \ldots = \psi_{i+k}$, then there is an orthogonal matrix H such that*

$$\hat{F} = H F H^T$$

where

$$\hat{F} = \begin{pmatrix} \Psi & \hat{\mathbf{g}} \\ 0 & \rho \end{pmatrix}, \quad \hat{g}_j = \begin{cases} g_j & j \neq i, \ldots, i+k \\ \pm \|\mathbf{g}(i\!:\!i+k)\|_2 & j = i \\ 0 & j = i+1, \ldots, i+k. \end{cases}$$

Moreover, $\psi_{i+1} = \phi_{i+1}$ is a singular value of F with multiplicity at least k.

The other circumstance where ϕ_i may be a singular value of F is when $g_i = 0$.

Corollary 6.27 *Let $F \in \mathrm{R}^{(n+1)\times(n+1)}$ be as in (6.77) and assume that $\psi_1 > \cdots > \psi_n$. For $i = 1, \ldots, n$, ψ_i is a singular value of F if and only if $\psi_i = 0$ or $g_i = 0$. If $g_i = 0$, the associated left and right singular vectors are $\mathbf{q}_i = \mathbf{p}_i = \mathbf{e}_i$, if $g_i \neq 0$ and $\psi_i = 0$, then an associated left singular vector is $\mathbf{q}_i = (\rho \mathbf{e}_i - g_i \mathbf{e}_{n+1})/\gamma$ where $\gamma = \|(\rho, g_i)^T\|_2$, and an associated right singular vector is $\mathbf{p}_i = \mathbf{e}_i$.*

Another formulation of the SVD of F can be found by considering the eigenvalue problem for

$$M = \begin{pmatrix} 0 & F \\ F^T & 0 \end{pmatrix} = \begin{array}{c} \\ n \\ 1 \\ n \\ 1 \end{array} \begin{array}{c} \begin{array}{cccc} n & 1 & n & 1 \end{array} \\ \begin{pmatrix} 0 & 0 & \Psi & \mathbf{g} \\ 0 & 0 & 0 & \rho \\ \Psi & 0 & 0 & 0 \\ \mathbf{g}^T & \rho & 0 & 0 \end{pmatrix} \end{array}.$$

If we perform the permutation $1, n+2, 2, n+3, \ldots, n+1, 2(n+1)$ on the rows and columns of M, we obtain

$$PMP^T = \begin{array}{c} \\ 2n+1 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 2n+1 & 1 \end{array} \\ \begin{pmatrix} \bar{M} & \bar{\mathbf{y}} \\ \bar{\mathbf{y}}^T & 0 \end{pmatrix} \end{array},$$

where

$$\bar{M} = diag(\bar{M}_1, \ldots, \bar{M}_n, 0), \quad \bar{M}_i = \begin{pmatrix} 0 & \psi_i \\ \psi_i & 0 \end{pmatrix}$$

$$\bar{\mathbf{y}} = (g_1, 0, g_2, 0, \ldots, g_n, 0, \rho)^T.$$

Each $\bar{M}$ has eigendecomposition

$$\bar{M}_i = G \begin{pmatrix} \psi_i & 0 \\ 0 & -\psi_i \end{pmatrix} G^T \quad G = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

thus if we let

$$\bar{G} = \mathrm{diag}(\overbrace{G, \ldots, G}^{n \text{ times}}, 1)$$

then

$$T = GPMP^T G^T = \begin{pmatrix} D & \mathbf{z} \\ \mathbf{z}^T & 0 \end{pmatrix} \tag{6.80}$$

where

$$D = diag(\psi_1, -\psi_1, \psi_2, -\psi_2, \ldots, \psi_n, -\psi_n, 0),$$

$$\mathbf{z} = \sqrt{0.5}(g_1, g_1, g_2, g_2, \ldots, g_n, g_n, \sqrt{2}\rho)^T.$$

The singular values of F, $\phi_1 \geq \cdots \geq \phi_{n+1}$ are the zeros of the spectral function

$$f(\phi) = -\phi + \rho^2/\phi - 0.5(\mathbf{g}^T(\Psi - \phi I)^{-1}\mathbf{g} - \mathbf{g}^T(\Psi + \phi I)^{-1}\mathbf{g}) - \phi + \rho^2/\phi - \phi\|\mathbf{g}\|_2^2/(\psi_1^2 - \phi^2).$$

The upper bound on ϕ_1 from Lemma 6.25 reads

$$\phi_1 \leq 0.5(\phi_1 + \left(\phi_1 + 4(\|\mathbf{g}\|_2^2 + \rho^2)\right)^{1/2}.$$

The singular values of F are very stable. Small in the components of F result in small changes in the singular values as shown be below. This perturbation theorem is similar to a special case of a result due Demmel and Gragg [7, 1992].

Theorem 6.28 *Let F have the form (6.77 and let $\tilde{F}$ be matrix*

$$\tilde{F} = \begin{pmatrix} \Psi(I + E_1) & (I + E_2)\mathbf{g} \\ 0 & (1 + \epsilon_3)\rho \end{pmatrix}$$

where for some $\tau < 1$, $\max\{\|E_1\|_2, \|E_2\|_2, \epsilon_3\} \leq \tau$ and E_2 is diagonal. If $\tilde{F}$ has the singular values $\tilde{\phi}_1 \geq \cdots \geq \tilde{\phi}_{n+1}$, then for $k = 1, \ldots, n+1$ either $\tilde{\phi}_k = \phi_{=}0$ or

$$\frac{(1-\tau)^2}{1+\tau}\phi_k \leq \tilde{\phi}_k \leq \frac{(1+\tau)^2}{1-\tau}\phi_k. \tag{6.81}$$

Proof. Since $E_2 \leq \tau < 1$, $I + E_2$ is nonsingular. Thus $\tilde{F}$ may be written

$$\tilde{F} = D_1 F D_2$$

where

$$D_1 = \text{diag}(I + E_2, 1 + \epsilon_3), \quad D_2 = \text{diag}((I + E_2)^{-1}(I + E_1), 1).$$

It is straightforward to show that

$$\psi_n(D_1) \geq 1 - \tau, \quad \psi_n(D_2) \geq (1 - \tau)/(1 + \tau),$$

$$\psi_1(D_1) \geq 1 + \tau, \quad \psi_n(D_2) \geq (1 + \tau)/(1 - \tau),$$

By Corollary 6.7, for $k = 1, \ldots, n$,

$$\psi_n(D_1)\psi_n(D_2)\psi_k(F) \leq \psi_k(\tilde{F}) \leq \psi_1(D_1)\psi_1(D_2)\psi_k(F).$$

Thus we obtain (6.81). □

We now give circumstances where elements of the vector $\mathbf{g}$ may be set to zero.

Corollary 6.29 *Assume the hypothesis and terminology of Theorem* 6.28 *and let* $\tilde{F}$ *be* F *with the entry* g_i *set to zero. Then*

$$\sum_{k=1}^{n+1}(\phi_k - \tilde{\phi}_k)^2 \leq g_i^2.$$

Moreover, if $\tau = |g_i|/\psi_i$, *then*

$$(1-\tau)\phi_k \leq \tilde{\phi}_k \leq (1+\tau)\phi_k, \quad k = 1, \ldots, n+1. \tag{6.82}$$

Proof. The first inequality is just an application of the Wielandt-Hoffman Theorem (Theorem **??**. The matrix F and $\tilde{F}$ are related by

$$\tilde{F} = FL$$

where

$$L = \begin{array}{c} \\ n \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} n & 1 \end{array} \\ \begin{pmatrix} I & 0 \\ -g_i/\phi_i \mathbf{e}_i^T & 1 \end{pmatrix} \end{array}.$$

Clearly the singular values of L satsify

$$1 - \tau \leq \psi_k(L) \leq 1 + \tau,$$

thus the use of Corollary 6.7 yields (6.82). □

A more sophisticated method for setting g_i to zero involves finding a Givens rotation $G_1 = J(j, i, \theta_1)$ such that $c_1 = \cos\,\theta_1$ and $s_1 = \sin\,\theta_1$ satisfy

$$\begin{pmatrix} c_1 & s_1 \\ -s_1 & c_1 \end{pmatrix} \begin{pmatrix} g_j \\ g_i \end{pmatrix} = \begin{pmatrix} \tilde{g}_j \\ 0 \end{pmatrix} \tag{6.83}$$

We also need a second rotation $G = J(j, i, \theta_2), c_2 = \cos\,\theta_2, s_2 = \sin\,\theta_2$ such that

$$\begin{pmatrix} c_2 & s_2 \\ -s_2 & c_2 \end{pmatrix} \begin{pmatrix} c_1\psi_j \\ s_1\psi_i \end{pmatrix} = \begin{pmatrix} \tilde{\psi}_j \\ 0 \end{pmatrix} \tag{6.84}$$

If we apply these two rotations to F, we get

$$G_1 F G_2 = \tilde{F} + \delta_{ji} e_i e_j^T$$

where

$$\tilde{F} = \begin{pmatrix} \tilde{\Psi} & \tilde{\mathbf{g}} \\ 0 & \rho \end{pmatrix}. \tag{6.85}$$

Here $\tilde{\mathbf{g}} = G_1^T \mathbf{g}$ and $\tilde{F}$ is identical to F except that the jth and ith row and column have the form

$$\begin{pmatrix} \tilde{\psi}_j & 0 \\ \delta_{ji} & \tilde{\psi}_i \end{pmatrix} \tag{6.86}$$

where

$$\tilde{\psi}_j = \left(c_1^2 \psi_j^2 + s_2^2 \psi_i^2\right)^{1/2}, \quad \tilde{\psi}_i = \psi_i \psi_j / \tilde{\psi}_j$$

and

$$\delta_{ji} = -c_1 s_1 (\psi_j^2 - \psi_i^2)/\tilde{\psi}_j.$$

The issue here is whether or not δ_{ji} may be set to zero. For that criterion, we have the following theorem.

Theorem 6.30 *Assume the hypothesis and terminology of Theorem* 6.28 *and let* $\tilde{F}$ *be given by* (6.85). *Let*

$$\zeta_{ji} = \delta_{ji}/\tilde{\psi}_i = -\frac{c_1 s_1 (\psi_j^2 - \psi_i^2)}{\psi_j \psi_i}.$$

If $\tau = |\zeta_{ji}|$, *then*

$$(1+\tau)^{-1} \phi_k \leq \tilde{\phi}_k \leq (1-\tau)^{-1} \phi_k, \quad k = 1, \ldots, n. \tag{6.87}$$

Proof. We have that

$$G_1 F G_2 = \begin{pmatrix} \tilde{\Psi}(I + E_1) & \tilde{\mathbf{g}} \\ 0 & \rho \end{pmatrix} = \tilde{F} D_2$$

where

$$E_1 = I + \zeta_{ji} \mathbf{e}_i \mathbf{e}_j^T, \quad D_2 = \mathrm{diag}((I + E_1), 1).$$

Using orthogonal equivalence and Corollary 6.7, we have

$$\psi_{n+1}(D_2) \tilde{\phi}_k \leq \phi_k \leq \psi_1(D_2) \tilde{\phi}_k$$

which becomes the inequality (6.87). □

Needed here are a couple more theorems relating perturbations to total least squares problems.

Bibliography

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. DuCroz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK User's Guide*. SIAM Publications, Philadelphia, 1992.

[2] J.L. Barlow and J.W. Demmel. Computing accurate eigensystems of scaled diagonally dominant matrices. *SIAM J. Numer. Anal.*, 27:762–791, 1990.

[3] J.L. Barlow, P.A. Yoon, and H. Zha. An algorithm and a stability theory for downdating the ULV decomposition. *BIT*, 36:14–40, 1996.

[4] D. Boley and G.H. Golub. Inverse problems for band matrices. In G.A. Watson, editor, *Numerical Analysis 1977*, pages 23–31, New York, 1977. Springer–Verlag.

[5] A. Cauchy. Cours D'Analyse de L'Ecole Polytechnique. In *Oeuvres Complètes*, volume 2 and 3, 1821.

[6] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience, New York, 1953. First English Edition.

[7] J.W. Demmel and W.B. Gragg. On computing accurate singular values and eigenvalues of matrices with acyclic graphs. *Linear Algebra and Its Applications*, 185:203–217, 1993.

[8] J.W. Demmel and W.H. Kahan. Accurate singular values of bidiagonal matrices. *SIAM J. Sci. Stat. Computing*, 11:873–912, 1990.

[9] J.W. Demmel and K. Veselić. Jacobi's method is more accurate than QR. *SIAM J. Matrix Anal. Appl.*, 13:1204–1243, 1992.

[10] G. Eckart and G. Young. The approximation of one matrix by one of lower rank. *Psychometrica*, 1:211–218, 1936.

[11] E. Fischer. Concerning quadratic forms with real coefficients. *Monatsh. Math. Phys.*, 16:234–249, 1905.

[12] G.H. Golub and W.M. Kahan. Calculating the singular values and pseudoinverse of a matrix. *SIAM J. Num. Anal. Ser. B*, 2:205–224, 1965.

[13] W. Kahan. Numerical linear algebra. *Canad. Math. Bull.*, 9:757–801, 1966.

[14] L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *Quart. J. Math. Oxford*, 11:50–59, 1960.

[15] E. Schmidt. *Math. Annalen*, 63:433–476, 1907.

[16] J.H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, London, 1965.