Jim Lambers
MAT 610
Summer Session 2009-10
Lecture 13 Notes

These notes correspond to Sections 7.2 and 8.1 in the text.

The Eigenvalue Problem: Perturbation Theory

The Unsymmetric Eigenvalue Problem

Just as the problem of solving a system of linear equations $A\mathbf{x} = \mathbf{b}$ can be sensitive to perturbations in the data, the problem of computing the eigenvalues of a matrix can also be sensitive to perturbations in the matrix. We will now obtain some results concerning the extent of this sensitivity.

Suppose that A is obtained by perturbing a diagonal matrix D by a matrix F whose diagonal entries are zero; that is, $A = D + F$. If λ is an eigenvalue of A with corresponding eigenvector $\mathbf{x}$, then we have

$$(D - \lambda I)\mathbf{x} + F\mathbf{x} = \mathbf{0}.$$

If λ is not equal to any of the diagonal entries of A, then $D - \lambda I$ is nonsingular and we have

$$\mathbf{x} = -(D - \lambda I)^{-1} F\mathbf{x}.$$

Taking ∞-norms of both sides, we obtain

$$\|\mathbf{x}\|_\infty = \|(D - \lambda I)^{-1} F\mathbf{x}\|_\infty \leq \|(D - \lambda I)^{-1} F\|_\infty \|\mathbf{x}\|_\infty,$$

which yields

$$\|(D - \lambda I)^{-1} F\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1, j \neq i}^{n} \frac{|f_{ij}|}{|d_{ii} - \lambda|} \geq 1.$$

It follows that for some i, $1 \leq i \leq n$, λ satisfies

$$|d_{ii} - \lambda| \leq \sum_{j=1, j \neq i}^{n} |f_{ij}|.$$

That is, λ lies within one of the *Gerschgorin circles* in the complex plane, that has center a_{ii} and radius

$$r_i = \sum_{j=1, j \neq i}^{n} |a_{ij}|.$$

This is result is known as the *Gerschgorin Circle Theorem*.

Example The eigenvalues of the matrix

$$A = \begin{bmatrix} -5 & -1 & 1 \\ -2 & 2 & -1 \\ 1 & -3 & 7 \end{bmatrix}$$

are

$$\lambda(A) = \{6.4971, 2.7930, -5.2902\}.$$

The Gerschgorin disks are

$$D_1 = \{z \in \mathbb{C} \,|\, |z - 7| \leq 4\}, \quad D_2 = \{z \in \mathbb{C} \,|\, |z - 2| \leq 3\}, \quad D_3 = \{z \in \mathbb{C} \,|\, |z + 5| \leq 2\}.$$

We see that each disk contains one eigenvalue. □

It is important to note that while there are n eigenvalues and n Gerschgorin disks, it is not necessarily true that each disk contains an eigenvalue. The Gerschgorin Circle Theorem only states that all of the eigenvalues are contained within the *union* of the disks.

Another useful sensitivity result that applies to diagonalizable matrices is the *Bauer-Fike Theorem*, which states that if

$$X^{-1}AX = \text{diag}(\lambda_1, \ldots, \lambda_n),$$

and μ is an eigenvalue of a perturbed matrix $A + E$, then

$$\min_{\lambda \in \lambda(A)} |\lambda - \mu| \leq \kappa_p(X)\|E\|_p.$$

That is, μ is within $\kappa_p(X)\|E\|_p$ of an eigenvalue of A. It follows that if A is "nearly non-diagonalizable", which can be the case if eigenvectors are nearly linearly dependent, then a small perturbation in A could still cause a large change in the eigenvalues.

It would be desirable to have a concrete measure of the sensitivity of an eigenvalue, just as we have the condition number that measures the sensitivity of a system of linear equations. To that end, we assume that λ is a simple eigenvalue of an $n \times n$ matrix A that has Jordan canonical form $J = X^{-1}AX$. Then, $\lambda = J_{ii}$ for some i, and $\mathbf{x}_i$, the ith column of X, is a corresponding right eigenvector.

If we define $Y = X^{-H} = (X^{-1})^H$, then $\mathbf{y}_i$ is a left eigenvector of A corresponding to λ. From $Y^H X = I$, it follows that $\mathbf{y}^H\mathbf{x} = 1$. We now let A, λ, and $\mathbf{x}$ be functions of a parameter ϵ that satisfy

$$A(\epsilon)\mathbf{x}(\epsilon) = \lambda(\epsilon)\mathbf{x}(\epsilon), \quad A(\epsilon) = A + \epsilon F, \quad \|F\|_2 = 1.$$

Differentiating with respect to ϵ, and evaluating at $\epsilon = 0$, yields

$$F\mathbf{x} + A\mathbf{x}'(0) = \lambda\mathbf{x}'(0) + \lambda'(0)\mathbf{x}.$$

Taking the inner product of both sides with $\mathbf{y}$ yields

$$\mathbf{y}^H F\mathbf{x} + \mathbf{y}^H A\mathbf{x}'(0) = \lambda\mathbf{y}^H\mathbf{x}'(0) + \lambda'(0)\mathbf{y}^H\mathbf{x}.$$

Because $\mathbf{y}$ is a left eigenvector corresponding to λ, and $\mathbf{y}^H\mathbf{x} = 1$, we have

$$\mathbf{y}^H F\mathbf{x} + \lambda\mathbf{y}^H\mathbf{x}'(0) = \lambda\mathbf{y}^H\mathbf{x}'(0) + \lambda'(0).$$

We conclude that

$$|\lambda'(0)| = |\mathbf{y}^H F\mathbf{x}| \leq \|\mathbf{y}\|_2\|F\|_2\|\mathbf{x}\|_2 \leq \|\mathbf{y}\|_2\|\mathbf{x}\|_2.$$

However, because θ, the angle between $\mathbf{x}$ and $\mathbf{y}$, is given by

$$\cos\theta = \frac{\mathbf{y}^H\mathbf{x}}{\|\mathbf{y}\|_2\|\mathbf{x}\|_2} = \frac{1}{\|\mathbf{y}\|_2\|\mathbf{x}\|_2},$$

it follows that

$$|\lambda'(0)| \leq \frac{1}{|\cos\theta|}.$$

We define $1/|\cos\theta|$ to be the *condition number* of the simple eigenvalue λ. We require λ to be simple because otherwise, the angle between the left and right eigenvectors is not unique, because the eigenvectors themselves are not unique.

It should be noted that the condition number is also defined by $1/|\mathbf{y}^H\mathbf{x}|$, where $\mathbf{x}$ and $\mathbf{y}$ are normalized so that $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$, but either way, the condition number is equal to $1/|\cos\theta|$. The interpretation of the condition number is that an $O(\epsilon)$ perturbation in A can cause an $O(\epsilon/|\cos\theta|)$ perturbation in the eigenvalue λ. Therefore, if $\mathbf{x}$ and $\mathbf{y}$ are nearly orthogonal, a large change in the eigenvalue can occur. Furthermore, if the condition number is large, then A is close to a matrix with a multiple eigenvalue.

Example The matrix

$$A = \begin{bmatrix} 3.1482 & -0.2017 & -0.5363 \\ 0.4196 & 0.5171 & 1.0888 \\ 0.3658 & -1.7169 & 3.3361 \end{bmatrix}$$

has a simple eigenvalue $\lambda = 1.9833$ with left and right eigenvectors

$$\mathbf{x} = \begin{bmatrix} 0.4150 & 0.6160 & 0.6696 \end{bmatrix}^T, \quad \mathbf{y} = \begin{bmatrix} -7.9435 & 83.0701 & -70.0066 \end{bmatrix}^T$$

such that $\mathbf{y}^H\mathbf{x} = 1$. It follows that the condition number of this eigenvalue is $\|\mathbf{x}\|_2\|\mathbf{y}\|_2 = 108.925$. In fact, the nearby matrix

$$B = \begin{bmatrix} 3.1477 & -0.2023 & -0.5366 \\ 0.4187 & 0.5169 & 1.0883 \\ 0.3654 & -1.7176 & 3.3354 \end{bmatrix}$$

has a double eigenvalue that is equal to 2. □

We now consider the sensitivity of repeated eigenvalues. First, it is important to note that while the eigenvalues of a matrix A are continuous functions of the entries of A, they are not necessarily differentiable functions of the entries. To see this, we consider the matrix

$$A = \begin{bmatrix} 1 & a \\ \epsilon & 1 \end{bmatrix},$$

where $a > 0$. Computing its characteristic polynomial

$$\det(A - \lambda I) = \lambda^2 - 2\lambda + 1 - a\epsilon$$

and computings its roots yields the eigenvalues $\lambda = 1 \pm \sqrt{a\epsilon}$. Differentiating these eigenvalues with respect to ϵ yields

$$\frac{d\lambda}{d\epsilon} = \pm\sqrt{\frac{a}{\epsilon}},$$

which is undefined at $\epsilon = 0$. In general, an $O(\epsilon)$ perturbation in A causes an $O(\epsilon^{1/p})$ perturbation in an eigenvalue associated with a $p \times p$ Jordan block, meaning that the "more defective" an eigenvalue is, the more sensitive it is.

We now consider the sensitivity of eigenvectors, or, more generally, invariant subspaces of a matrix A, such as a subspace spanned by the first k Schur vectors, which are the first k columns in a matrix Q such that $Q^H AQ$ is upper triangular. Suppose that an $n \times n$ matrix A has the Schur decomposition

$$A = QTQ^H, \quad Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}, \quad T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix},$$

where Q_1 is $n \times r$ and T_{11} is $r \times r$. We define the *separation* between the matrices T_{11} and T_{22} by

$$\text{sep}(T_{11}, T_{22}) = \min_{X \neq 0} \frac{\|T_{11}X - XT_{22}\|_F}{\|X\|_F}.$$

It can be shown that an $O(\epsilon)$ perturbation in A causes a $O(\epsilon/\text{sep}(T_{11}, T_{22}))$ perturbation in the invariant subspace Q_1.

We now consider the case where $r = 1$, meaning that Q_1 is actually a vector $\mathbf{q}_1$, that is also an eigenvector, and T_{11} is the corresponding eigenvalue, λ. Then, we have

$$\begin{aligned} \text{sep}(\lambda, T_{22}) &= \min_{X \neq 0} \frac{\|\lambda X - XT_{22}\|_F}{\|X\|_F} \\ &= \min_{\|\mathbf{y}\|_2 = 1} \|\mathbf{y}^H(T_22 - \lambda I)\|_2 \\ &= \min_{\|\mathbf{y}\|_2 = 1} \|(T_22 - \lambda I)^H \mathbf{y}\|_2 \\ &= \sigma_{\min}((T_{22} - \lambda_I)^H) \\ &= \sigma_{\min}(T_{22} - \lambda I), \end{aligned}$$

since the Frobenius norm of a vector is equivalent to the vector 2-norm. Because the smallest singular value indicates the distance to a singular matrix, $\text{sep}(\lambda, T_{22})$ provides a measure of the separation of λ from the other eigenvalues of A. It follows that eigenvectors are more sensitive to perturbation if the corresponding eigenvalues are clustered near one another. That is, eigenvectors associated with nearby eigenvalues are "wobbly".

It should be emphasized that there is no direct relationship between the sensitivity of an eigenvalue and the sensitivity of its corresponding invariant subspace. The sensitivity of a simple eigenvalue depends on the angle between its left and right eigenvectors, while the sensitivity of an invariant subspace depends on the clustering of the eigenvalues. Therefore, a sensitive eigenvalue, that is nearly defective, can be associated with an insensitive invariant subspace, if it is distant from other eigenvalues, while an insensitive eigenvalue can have a sensitive invariant subspace if it is very close to other eigenvalues.

The Symmetric Eigenvalue Problem

In the symmetric case, the Gerschgorin circles become Gerschgorin intervals, because the eigenvalues of a symmetric matrix are real.

Example The eigenvalues of the 3×3 symmetric matrix

$$A = \begin{bmatrix} -10 & -3 & 2 \\ -3 & 4 & -2 \\ 2 & -2 & 14 \end{bmatrix}$$

are

$$\lambda(A) = \{14.6515, 4.0638, -10.7153\}.$$

The Gerschgorin intervals are

$$D_1 = \{x \in \mathbb{R} \,|\, |x - 14| \leq 4\}, \quad D_2 = \{x \in \mathbb{R} \,|\, |x - 4| \leq 5\}, \quad D_3 = \{x \in \mathbb{R} \,|\, |x + 10| \leq 5\}.$$

We see that each intervals contains one eigenvalue. □

The characterization of the eigenvalues of a symmetric matrix as constrained maxima of the Rayleight quotient lead to the following results about the eigenvalues of a perturbed symmetric matrix. As the eigenvalues are real, and therefore can be ordered, we denote by $\lambda_i(A)$ the ith largest eigenvalue of A.

Theorem (Wielandt-Hoffman) If A and $A + E$ are $n \times n$ symmetric matrices, then

$$\sum_{i=1}^{n} (\lambda_i(A + E) - \lambda_i(A))^2 \leq \|E\|_F^2.$$

It is also possible to bound the distance between individual eigenvalues of A and $A + E$.

Theorem If A and $A+E$ are $n \times n$ symmetric matrices, then

$$\lambda_n(E) \leq \lambda_k(A+E) - \lambda_k(A) \leq \lambda_1(E).$$

Furthermore,

$$|\lambda_k(A+E) - \lambda_k(A)| \leq \|E\|_2.$$

The second inequality in the above theorem follows directly from the first, as the 2-norm of the symmetric matrix E, being equal to its spectral radius, must be equal to the larger of the absolute value of $\lambda_1(E)$ or $\lambda_n(E)$.

Theorem (Interlacing Property) If A is an $n \times n$ symmetric matrix, and A_r is the $r \times r$ leading principal minor of A, then, for $r = 1, 2, \ldots, n-1$,

$$\lambda_{r+1}(A_{r+1}) \leq \lambda_r(A_r) \leq \lambda_r(A_{r+1}) \leq \cdots \leq \lambda_2(A_{r+1}) \leq \lambda_1(A_r) \leq \lambda_1(A_{r+1}).$$

For a symmetric matrix, or even a more general normal matrix, the left eigenvectors and right eigenvectors are the same, from which it follows that every simple eigenvalue is "perfectly conditioned"; that is, the condition number $1/|\cos\theta|$ is equal to 1 because $\theta = 0$ in this case. However, the same results concerning the sensitivity of invariant subspaces from the nonsymmetric case apply in the symmetric case as well: such sensitivity increases as the eigenvalues become more clustered, even though there is no chance of a defective eigenvalue. This is because for a nondefective, repeated eigenvalue, there are infinitely many possible bases of the corresponding invariant subspace. Therefore, as the eigenvalues approach one another, the eigenvectors become more sensitive to small perturbations, for any matrix.

Let Q be an $n \times r$ matrix with orthonormal columns, meaning that $Q_1^T Q_1 = I_r$. If it spans an invariant subspace of an $n \times n$ symmetric matrix A, then $AQ_1 = Q_1 S$, where $S = Q_1^T A Q_1$. On the other hand, if range(Q_1) is *not* an invariant subspace, but the matrix

$$AQ_1 - Q_1 S = E_1$$

is small for any given $r \times r$ symmetric matrix S, then the columns of Q_1 define an *approximate* invariant subspace.

It turns out that $\|E_1\|_F$ is minimized by choosing $S = Q_1^T A Q_1$. Furthermore, we have

$$\|AQ_1 - S_1 S\|_F = \|P_1^\perp A Q_1\|_F,$$

where $P_1^\perp = I - Q_1 Q_1^T$ is the orthogonal projection into $(\text{range}(Q_1))^\perp$, and there exist eigenvalues $\mu_1, \ldots, \mu_r \in \lambda(A)$ such that

$$|\mu_k - \lambda_k(S)| \leq \sqrt{2}\|E_1\|_2, \quad k = 1, \ldots, r.$$

That is, r eigenvalues of A are close to the eigenvalues of S, which are known as *Ritz values*, while the corresponding eigenvectors are called *Ritz vectors.* If $(\theta_k, \mathbf{y}_k)$ is an eigenvalue-eigenvector pair, or an *eigenpair* of S, then, because S is defined by $S = Q_1^T A Q_1$, it is also known as a *Ritz pair.* Furthermore, as θ_k is an approximate eigenvalue of A, $Q_1\mathbf{y}_k$ is an approximate corresponding eigenvector.

To see this, let σ_k (not to be confused with singular values) be an eigenvalue of S, with eigenvector $\mathbf{y}_k$. We multiply both sides of the equation $S\mathbf{y}_k = \sigma_k\mathbf{y}_k$ by Q_1:

$$Q_1 S\mathbf{y}_k = \sigma_k Q_1\mathbf{y}_k.$$

Then, we use the relation $AQ_1 - Q_1 S = E_1$ to obtain

$$(AQ_1 - E_1)\mathbf{y}_k = \sigma_k Q_1\mathbf{y}_k.$$

Rearranging yields

$$A(Q_1\mathbf{y}_k) = \sigma_k(Q_1\mathbf{y}_k) + E_1\mathbf{y}_k.$$

If we let $\mathbf{x}_k = Q_1\mathbf{y}_k$, then we conclude

$$A\mathbf{x}_k = \sigma_k\mathbf{x}_k + E_1\mathbf{y}_k.$$

Therefore, $\|E_1\|$ is small in some norm, $Q_1\mathbf{y}_k$ is nearly an eigenvector.