5. Calculation of eigenvalues and eigenvectors

Given an $n \times n$ matrix A, we consider the problem of finding scalars λ and vectors $x \neq 0$, such that $Ax = \lambda x$. There are a number of methods for solving this problem, and the best choice will depend on the type of matrix and whether we want some or all of the eigenvalues and/or eigenvectors.

Generally speaking, we do not compute the eigenvalues of A by finding the solution of the characteristic polynomial $\det(A - \lambda I) = 0$. For $n > 4$, there is no closed-form solution of polynomial equations, so we need an approximation method. Also, the roots of a polynomial can often be very sensitive to small changes in the coefficients, so roundoff errors in the computation could lead to bad approximations of the eigenvalues. Instead, we use methods that reduce the matrix A to a matrix that has the same eigenvalues, but one where the eigenvalues are easy to compute.

5.1. **Canonical forms of matrices.** We begin by recalling the following definition.

Definition: Two matrices A and B are similar if $A = C^{-1}BC$ for some nonsingular matrix C.

Lemma 2. *If A and B are similar, then they have the same characteristic polynomial (and hence the same eigenvalues).*

Proof. Let $A = C^{-1}BC$. Then

$$\begin{aligned}\det(A - \lambda I) = \det(C^{-1}BC - \lambda I) = \det[C^{-1}(B - \lambda I)C] &= \det(C^{-1}) \cdot \det(B - \lambda I) \cdot \det C \\ &= [\det C]^{-1} \cdot \det(B - \lambda I) \cdot \det C = \det(B - \lambda I).\end{aligned}$$

□

The importance of this result is that we can use similarity transformations to reduce A to a simple form from which the eigenvalues are easily determined. We note that if $A = C^{-1}BC$ and $Ax = \lambda x$, then $C^{-1}BCx = \lambda x$ and so $BCx = \lambda Cx$. Hence, if λ is an eigenvalue of A with eigenvector x, then λ is an eigenvalue of B with eigenvector Cx.

We next consider the question: For what types of matrices are the eigenvalues easily determined? Clearly, this is the case for triangular matrices, since the eigenvalues are just the diagonal entries. This leads to the obvious question: What types of matrices can be reduced to these forms by similarity transformations?

Definition: An $n \times n$ matrix A is called *defective* if it has an eigenvalue of multiplicity k having fewer than k linearly independent eigenvectors.

Example: The matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has a double eigenvalue $\lambda = 1$, but only one linearly independent eigenvector $[1, 0]^T$, so is defective. By contrast, the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ also has a double eigenvalue $\lambda = 1$, but has two linearly independent eigenvectors $[1, 0]^T$ and $[0, 1]^T$.

Theorem 13. *Let A be an $n \times n$ matrix with complex entries. Then A is nondefective if and only if there is a nonsingular matrix X such that $X^{-1}AX = \Lambda = diag(\lambda_1, \dots, \lambda_n)$, where the λ_i are the eigenvalues of A and the ith column of X is an eigenvector of A corresponding to λ_i.*

There are a few problems with this result. The first is that X may be ill-conditioned so that $X^{-1}AX$ may be inaccurate. The second problem is that given an X, we must calculate X^{-1}. Instead, we consider what simplifications can be accomplished using the class of unitary transformations, i.e., $U^{-1} = U^* \equiv \bar{U}^T$.

Theorem 14. *Let A be an $n \times n$ matrix with complex entries. If A has eigenvalues $\lambda_1, \dots, \lambda_n$, then there is a unitary matrix U such that U^*AU is upper triangular with diagonal elements $\lambda_1, \dots, \lambda_n$.*

Corollary 2. *If A is normal, i.e., $A^*A = AA^*$, then U^*AU will be diagonal.*

Note that if A is Hermitian, $A = A^*$, then A is normal.

If A is real, we do not want to use general unitary transformations to triangularize A, since U^*AU will then have complex elements. We then confine ourselves to real unitary transformations, i.e., orthogonal matrices Q satisfying $Q^{-1} = Q^T$. In this case, we will not generally get Q^TAQ to be a triangular matrix. Instead, we have the following result.

Theorem 15. *Let A be a real $n \times n$ matrix. Then there exists an orthogonal matrix Q such that Q^TAQ is quasi-triangular (i.e., A is block triangular with each diagonal block of order ≤ 2). Moreover, Q may be chosen so that any 2×2 diagonal block of Q^TAQ has only complex eigenvalues (which must be complex conjugates – since a polynomial with real coefficients can only have complex roots in conjugate pairs).*

A quasi-triangular matrix will have the form

$$\begin{pmatrix} A_1 & A_2 & A_3 \\ 0 & A_4 & A_5 \\ 0 & 0 & A_6 \end{pmatrix},$$

where the matrices A_1, A_4, and A_6 are at most 2×2.

Another terminology that is used is to call A upper Hessenberg if $a_{ij} = 0$ when $i > j+1$. So if A is real, then there exists an orthogonal matrix Q such that Q^TAQ is upper Hessenberg.

Corollary 3. *If A is real and symmetric, then there exists an orthogonal matrix Q such that Q^TAQ is tridiagonal.*

Proof.

$$(Q^T AQ)^T = Q^T A^T Q = Q^T AQ,$$

so $Q^T AQ$ is symmetric. Hence $a_{ij} = 0$ when $i > j + 1$ and when $j > i + 1$. □

In fact, since A has only real eigenvalues in this case, $Q^T AQ = D$, where D is a diagonal matrix containing the eigenvalues of A. Then $AQ = QD$, so Q is a matrix of corresponding orthogonal eigenvectors.

We will use these results as a basis for numerical approximation schemes.

5.2. **Perturbation theory for eigenvalues and eigenvectors.** The numerical schemes we study will produce approximate eigenvalues and eigenvectors. If λ is an approximate eigenvalue of a matrix A and x a corresponding approximate eigenvector, we can ask whether if $\|Ax - \lambda x\|$ is small, will λ and/or x be good approximations to some eigenvalue and corresponding eigenvector of A.

The following perturbation result gives a partial answer to that question.

Theorem 16. *(i) If A has n linearly independent eigenvectors x^i corresponding to eigenvalues λ_i, then for any scalar λ and vector $x \neq 0$,*

$$\min_i |\lambda_i - \lambda| \leq \|P^{-1}\| \|P\| \|Ax - \lambda x\| / \|x\|,$$

where P is a matrix whose ith column is the eigenvector x^i and $\|\cdot\|$ is the 1,2, or ∞ norm.
(ii) If A is normal, then

$$\min_i |\lambda_i - \lambda| \leq \|Ax - \lambda x\|_2 / \|x\|_2.$$

Proof. Let $r = (A - \lambda I)x$. By the definition of P, $P^{-1}AP = \Lambda$, where Λ is a diagonal matrix with the λ_i on the diagonal. Hence, $A = P\Lambda P^{-1}$ and so

$$P(\Lambda - \lambda I)P^{-1}x = (A - \lambda I)x = r.$$

If λ is an eigenvalue of A then the estimate is trivially true. If not, then $(\Lambda - \lambda I)^{-1}$ exists, so $x = P(\Lambda - \lambda I)^{-1}P^{-1}r$. Hence,

$$\|x\| \leq \|P\| \|(\Lambda - \lambda I)^{-1}\| \|P^{-1}\| \|r\|.$$

Now

$$\|(\Lambda - \lambda I)^{-1}\| = \max_i |\lambda_i - \lambda|^{-1}$$

for the $1, 2, \infty$ norms. Hence,

$$\|x\| \leq \frac{1}{\min_i |\lambda_i - \lambda|} \|P^{-1}\| \|P\| \|Ax - \lambda x\|,$$

which implies that

$$\min_i |\lambda_i - \lambda| \leq \|P^{-1}\| \|P\| \|Ax - \lambda x\| / \|x\|.$$

If A is normal, then P is a unitary matrix, so $\|P\|_2 = \|P^{-1}\|_2 = 1$ and the result follows immediately. □

So for normal matrices, a small residual does guarantee a good approximation, while for more general matrices, we could get magnification of the error by the condition number of the matrix P.

We next ask whether a small residual guarantees an accurate approximation to an eigenvector. We shall see by the following example that the answer is no, even for symmetric matrices. Let

$$A = \begin{pmatrix} a & \epsilon \\ \epsilon & a \end{pmatrix}, \qquad \lambda = a, \qquad x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Then $Ax - \lambda x = (0, \epsilon)^T$ so $\|Ax - \lambda x\|_2 = |\epsilon|$. An easy calculation shows that the true eigenvalues of A are given by $a \pm \epsilon$ with corresponding orthonormal eigenvectors $(1/\sqrt{2}, 1/\sqrt{2})^T$ and $(1/\sqrt{2}, -1/\sqrt{2})^T$. Obviously, x is not a good approximation to either of these. The problem occurs when the eigenvalues are close. However, we are able to obtain the following error bound.

Theorem 17. *Suppose $A = A^*$ and let λ be a scalar and $x \neq 0$ a unit vector. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A and $x^1, \dots, x^n$ corresponding orthonormal eigenvectors, Suppose that*

$$|\lambda_i - \lambda| \leq \|Ax - \lambda x\|_2, \quad i = 1, 2, \dots, r, \qquad |\lambda_i - \lambda| \geq d, \quad i = r+1, \dots, n.$$

Then

$$\min_{a_1, \dots, a_r} \|x - \sum_{i=1}^r a_i x^i\|_2 \leq \|Ax - \lambda x\|_2 / d.$$

Note: by a previous theorem, there exist some λ_i such that $|\lambda_i - \lambda| \leq \|Ax - \lambda x\|_2$. We are considering the situation when there might be other close eigenvalues that have some separation from the remaining eigenvalues.

Proof. Since the $\{x^i\}$ are a basis for $\mathbb{R}^n$, there exist constants b_i such that $x = \sum_{i=1}^n b_i x^i$. Using the fact the $\{x^i\}$ are an orthonormal basis, we have

$$\min_{a_1, \dots, a_r} \|x - \sum_{i=1}^r a_i x^i\|_2 \leq \|x - \sum_{i=1}^r b_i x^i\|_2 = \|\sum_{i=r+1}^n b_i x^i\|_2 = \sum_{i=r+1}^n |b_i|^2.$$

Now

$$\begin{aligned} \|Ax - \lambda x\|_2^2 = \|(A - \lambda I) \sum_{i=1}^n b_i x^i\|_2^2 = \|\sum_{i=1}^n b_i (\lambda_i - \lambda) x^i\|_2^2 &= \sum_{i=1}^n b_i^2 (\lambda_i - \lambda)^2 \\ &\geq \sum_{i=r+1}^n b_i^2 (\lambda_i - \lambda)^2 \geq d^2 \sum_{i=r+1}^n b_i^2. \end{aligned}$$

Combining these results, we get

$$\min_{a_1, \dots, a_r} \|x - \sum_{i=1}^r a_i x^i\|_2 \leq \|Ax - \lambda x\|_2^2 / d^2,$$

and the result follows by taking square roots of both sides of the inequality. ☐