
Basic Theory of Algebraic Eigenvalue Problems: a quick review

---

1 Introduction

This lecture covers basic theory of the algebraic eigenvalue problem and a brief invitation to the first order asymptotic error expansion of eigenvalues and eigenvectors for sensitivity analysis.

2 Essentials

Definition 2.1. *Let $A \in \mathbb{C}^{n\times n}$.*

1. *A scalar λ is an* ***eigenvalue*** *of an $n \times n$ matrix A and a nonzero vector $x \in \mathbb{C}^n$ is a corresponding* ***(right) eigenvector*** *if*

$$Ax = \lambda x.$$

   $\mathcal{L}_{A,\lambda} \stackrel{\text{def}}{=} \{x : Ax = \lambda x\}$ *is an* ***eigenspace*** *of A.*

2. *A nonzero vector y such that*

$$y^H A = \lambda y^H$$

   is a ***left eigenvector****.*

3. *The set of all eigenvalues of A is called the* ***spectrum*** *of A.*

4. $p_A(\lambda) \stackrel{\text{def}}{=} \det(\lambda I - A)$*, a polynomial of degree n, is called the* ***characteristic polynomial*** *of A.*

The following is a list of properties straightforwardly from Definition 2.1.

1. λ is A's eigenvalue $\Leftrightarrow$ $\lambda I - A$ is singular $\Leftrightarrow$ $\det(\lambda I - A) = 0 \Leftrightarrow p_A(\lambda) = 0$.
2. There is at least one eigenvector x associated with A's eigenvalue λ; in the other word, the dimension $\dim(\mathcal{L}_{A,\lambda}) \geq 1$.
3. $\mathcal{L}_{A,\lambda}$ is a linear subspace, i.e., it has the following two properties:
   (a) $x \in \mathcal{L}_{A,\lambda} \Rightarrow \alpha x \in \mathcal{L}_{A,\lambda}$ for all $\alpha \in \mathbb{C}$.
   (b) $x_1, x_2 \in \mathcal{L}_{A,\lambda} \Rightarrow x_1 + x_2 \in \mathcal{L}_{A,\lambda}$.
4. Suppose A is real. λ is A's eigenvalue $\Leftrightarrow$ conjugate $\bar{\lambda}$ is also A's eigenvalue.
5. A is singular $\Leftrightarrow$ 0 is A's eigenvalue.
6. If A is upper (or lower) triangular, then its eigenvalues consist of its diagonal entries.

Theorem 2.1. *Let $Ax_i = \lambda_i x_i$, $x_i \neq 0$ for $i = 1, 2, \ldots k$, and $\lambda_i \neq \lambda_j$ for $i \neq j$. Then $x_1, x_2, \ldots, x_k$ are linearly independent.*

Proof. By induction. □

Definition 2.2. *$A \in \mathbb{C}^{n\times n}$ is* ***simple*** *if it has n linearly independent eigenvectors; otherwise it is* ***defective****.*

When A is simple, there is a basis of $\mathbb{C}^n$ consisting of eigenvectors of A.

Example 2.1. Simple and defective matrices.

1. I and any diagonal matrices is simple. $e_1, e_2, \ldots, e_n$ *are* n *linearly independent eigenvectors.*

2. $\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}$ is simple. It has two different eigenvalues -1 and 5. By Theorem 2.1 and by the fact that each eigenvalue corresponds to at least one eigenvector, it must have 2 linearly independent eigenvectors.

3. If $A \in \mathbb{C}^{n\times n}$ has n different eigenvalues, then A is simple.

4. $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ is defective. It has two repeated eigenvalues 2, but only one eigenvector $e_1 = (1,0)^T$.

Definition 2.3. *An* ***invariant subspace*** *of A is a subspace $\mathcal{V}$ of $\mathbb{R}^n$, with the property that $v \in \mathcal{V}$ implies that $Av \in \mathcal{V}$. We also write this as $A\mathcal{V} \subseteq \mathcal{V}$.*

Example 2.2.
(1) The simplest, one-dimensional invariant subspace is the set $\mathrm{span}(x)$ of all scalar multiples of an eigenvector x.
(2) Let $x_1, x_2, \ldots, x_m$ be any set of independent eigenvectors associated with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. Then $\mathcal{X} = \mathrm{span}(\{x_1, x_2, \ldots, x_m\})$ is an invariant subspace.

Proposition 2.1. *Let A be n-by-n, $V = [v_1, v_2, \ldots, v_m]$ any n-by-m matrix with linearly independent columns, and let $\mathcal{V} = \mathrm{span}(V)$, the m-dimensional space spanned by the columns of V. Then $\mathcal{V}$ is an invariant subspace if and only if there is an m-by-m matrix B such that*

$$AV = VB.$$

In this case the m eigenvalues of B are also eigenvalues of A.

Definition 2.4. *$n \times n$ matrices A and B are* ***similar*** *if there is an $n \times n$ non-singular matrix X such that $B = X^{-1}AX$. We also say A is similar to B, and likewise B is similar to A; X is a similarity transformation. A is unitarily similar to B if X is unitary.*

Proposition 2.2. *Suppose that A and B are similar: $B = X^{-1}AX$. Then*

1. *A and B have the same eigenvalues. In fact $p_A(\lambda) \equiv p_B(\lambda)$.*

2. *$Ax = \lambda x \Rightarrow B(X^{-1}x) = \lambda(X^{-1}x)$.*

3. *$Bw = \lambda w \Rightarrow A(Xw) = \lambda(Xw)$.*

Let $X_1 \in \mathbb{C}^{n\times k}$, $k < n$, be a matrix with linearly independent columns that represents a nontrivial invariant subspace of A, i.e., $AX_1 = X_1 B_{11}$ for some $k \times k$ matrix B_{11}. Then let X_2 be a matrix such that $X = (X_1, X_2)$ is nonsingular, then it's easy to verify that

$$B = X^{-1}AX = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix}.$$

In other words, we can use X_1 to build a similarity transformation to block trianguarlize A. Therefore, the basic task of eigenvalue computations is to find a nontrivial invariant subspace.

The following theorem shows that we can use a unitary similarity transformation to trianguarlize A.

Theorem 2.2 (Schur decomposition or Schur canonical form). *For any $A \in \mathbb{C}^{n\times n}$, there is an $n \times n$ unitary matrix U ($U^H U = I$) such that*

$$A = UTU^H,$$

where T is upper triangular. By appropriate choice of U, the eigenvalues of A, which are the diagonal elements of T, can be made to appear in any given order.

Proof. We prove it by induction. The conclusion is true for $n = 1$. Suppose it is true for $n = k$. Consider $n = k + 1$. Let (μ, x) be an eigenpair of A and $x^H x = 1$. Extend x to a unitary matrix $Q = (x\ Q_2)$. Then

$$Q^H AQ = \begin{pmatrix} x^H Ax & x^H AQ_2 \\ Q_2^H Ax & Q_2^H AQ_2 \end{pmatrix} = \begin{pmatrix} \mu & x^H AQ_2 \\ 0 & Q_2^H AQ_2 \end{pmatrix},$$

and $\operatorname{eig}(A) = \{\mu\} \cup \operatorname{eig}(Q_2^H AQ_2)$. Since $Q_2^H AQ_2$ is $k \times k$, by the induction assumption there is a $k \times k$ unitary matrix $\hat{U}$ such that $\hat{U}^H (Q_2^H AQ_2)\hat{U} = T_2$ and the eigenvalues of $Q_2^H AQ_2$ can be made to appear in any given order as the diagonal entries of T_1. Finally, we have

$$U = Q\begin{pmatrix} 1 & \\ & \hat{U} \end{pmatrix}, \quad U^H AU = T \equiv \begin{pmatrix} \mu & x^H AQ_2\hat{U} \\ & T_2 \end{pmatrix}.$$

It is clear the eigenvalues of A can be made to appear in any given order as the diagonal entries of T. This completes the proof. □

The Schur decomposition is one of the most important tools in theoretical and computational linear algebra! A real square matrix A may have complex eigenvalues. Which means U and T in its Schur canonical form may have to be complex. From the computational point of view, this is bad news because it takes twice as much space to store a complex number as to store a real number and 6 flops to multiply two complex numbers *vs* 1 flop to multiply two real numbers. Fortunately with a little compromise on the shape of T, there is an analogue of the real Schur canonical form of a real matrix. See Exercise 2.2.

If A is simple, then there exists an invertible matrix X such that

$$A = X\Lambda X^{-1}, \quad \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \tag{2.1}$$

where Λ is diagonal. It can be seen that the columns of X are the n linear independent eigenvectors of A. The process of computing such a decomposition is also called ***diagonalizing*** A.

If A is a $n \times n$ real symmetric matrix, $A^T = A$, then it can be transformed into diagonal form by means of an orthogonal transformation, namely X in (2.1) can be chosen to be an orthogonal matrix Q, where $Q^T Q = I_n$.

An ***eigen-decomposition*** of a square matrix A often refers to any one of the diagonalization form (2.1), Schur decomposition, and Jordan canonical decomposition.

Lemma 2.1. *A real symmetric matrix A is positive definite if and only if all of its eigenvalues of A are positive.*

Exercises

2.1. Prove Theorem 2.1.

2.2. This exercise leads to an analogue of the Schur decomposition for a real matrix.

1. Let the columns of X form an orthonormal basis for an invariant subspace of A. Let $AX = XL$, and let (X, Y) be unitary. Show that
$$\begin{pmatrix} X^H \\ Y^H \end{pmatrix} A(\, X \; Y \,) = \begin{pmatrix} L & H \\ 0 & M \end{pmatrix}$$

2. Let A be real, and let λ be a complex eigenvalue of A with eigenvector $x + iy$. Show that the space spanned by x and y is a 2-dimensional invariant subspace of A.

3. Real Schur Decomposition. Show that if A is real, there is an orthogonal matrix Q such that $Q^T AQ$ is block triangular with 1×1 and 2×2 blocks on its diagonal. The 1×1 blocks contain the real eigenvalues of A, and the eigenvalues of the 2×2 blocks are the complex eigenvalues of A.

2.3. Prove Lemma 2.1. Find a necessary and sufficient condition in terms of eigenvalues for a real symmetric matrix A being positive semidefinite.

3 Sensitivity of eigenvalue problems

3.1 Preliminaries

An eigenvalue λ_i of A is called ***simple*** if it is a simple root of A's characteristic polynomial, i.e.,
$$p_A(\lambda) \equiv (\lambda - \lambda_i) \times q(\lambda),$$
where $q(\lambda)$ is a polynomial of degree $n-1$ and $q(\lambda_i) \neq 0$.

Theorem 3.1. *Let A be an $n \times n$ matrix, and let λ be A's eigenvalue with corresponding (right) eigenvector x and left eigenvector y.*

1. *Let μ be another eigenvalue of A and w the corresponding (right) eigenvector (i.e., $Aw = \mu w$). If $\lambda \neq \mu$, then $y^H w = 0$.*

2. *If λ is simple, then $y^H x \neq 0$.*

Proof. We prove *1* first. Notice that
$$Aw = \mu w \Rightarrow y^H Aw = \mu y^H w \quad \text{and} \quad y^H A = \lambda y^H \Rightarrow y^H Aw = \lambda y^H w,$$
which lead to $\mu y^H w = \lambda y^H w$, and thus $(\lambda - \mu) y^H w = 0$. Since $\mu \neq \lambda$, we must have $y^H w = 0$.

Now we prove *2*. By the Schur decomposition theorem, there is a unitary matrix U such that
$$U^H AU = T \equiv \begin{array}{cc} & \begin{array}{cc} 1 & n-1 \end{array} \\ \begin{array}{c} 1 \\ n-1 \end{array} & \begin{pmatrix} \lambda & t^H \\ 0 & T_1 \end{pmatrix} \end{array}.$$
Since λ is simple, λ is not T_1's eigenvalue. This means $\lambda I - T_1$ is nonsingular. Let $z^H = -t^H(\lambda I - T_1)^{-1}$, and set
$$Z = \begin{array}{cc} & \begin{array}{cc} 1 & n-1 \end{array} \\ \begin{array}{c} 1 \\ n-1 \end{array} & \begin{pmatrix} 1 & z^H \\ 0 & I \end{pmatrix} \end{array}, \quad \text{and then} \quad Z^{-1} = \begin{array}{cc} & \begin{array}{cc} 1 & n-1 \end{array} \\ \begin{array}{c} 1 \\ n-1 \end{array} & \begin{pmatrix} 1 & -z^H \\ 0 & I \end{pmatrix} \end{array}.$$
It can be verified that
$$Z^{-1} U^H AUZ = \begin{pmatrix} 1 & -z^H \\ 0 & I \end{pmatrix} \begin{pmatrix} \lambda & t^H \\ 0 & T_1 \end{pmatrix} \begin{pmatrix} 1 & z^H \\ 0 & I \end{pmatrix} = \begin{pmatrix} \lambda & \lambda z^H + t^H - z^H T_1 \\ 0 & T_1 \end{pmatrix}.$$

By the definition of z, $\lambda z^H + t^H - z^H T_1 = t^H + z^H(\lambda I - T_1) = 0$. So we have

$$AUZ = UZ\begin{pmatrix} \lambda & 0 \\ 0 & T_1 \end{pmatrix} \quad \text{and} \quad Z^{-1}U^H A = \begin{pmatrix} \lambda & 0 \\ 0 & T_1 \end{pmatrix} Z^{-1}U^H.$$

Write $X = UZ$ and $Y^H = Z^{-1}U^H$. We see that Xe_1 is A's (right) eigenvector corresponding to λ and that Ye_1 is A's left eigenvector corresponding to λ. Now since λ is simple, A's (left and right) eigenspace corresponding to λ is 1-dimensional (Why?). So

$$\alpha x = Xe_1 = Ue_1 \quad \text{and} \quad \beta y = Ye_1 = U\begin{pmatrix} 1 \\ z \end{pmatrix},$$

where $\alpha \neq 0 \neq \beta$; therefore

$$y^H x = \frac{1}{\bar{\beta}\alpha}(U\begin{pmatrix} 1 \\ z \end{pmatrix})^H(Ue_1) = \frac{1}{\bar{\beta}\alpha} \neq 0,$$

as required. □

Example 3.1 (repeated eigenvalue case)**.** $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has a repeated eigenvalue 1 with corresponding (right) eigenvector $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and left eigenvector $y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. It can be verified that $y^H x = 0$.

Example 3.2 (distinct eigenvalue case)**.** $A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ whose eigenvalues and eigenvectors are

$$\lambda_1 = 1,\ x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ y_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \text{and} \quad \lambda_2 = 2,\ x_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ y_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

It can be verified that

$$y_1^H x_1 = 1 \neq 0, \quad y_1^H x_2 = 0, \quad y_2^H x_2 = 1 \neq 0, \quad y_2^H x_1 = 0.$$

Theorem 3.2 (Gershgorin Circle Theorem)**.** *Let $A = (a_{ij})$ be an $n \times n$ matrix. Then the eigenvalues of A are located in the union of the n disks:*

$$\{\lambda : |\lambda - a_{kk}| \leq \sum_{j \neq k} |a_{kj}|\}, \quad \textit{for} \quad k = 1, 2, \ldots, n,$$

*so-called **Gershgorin disks**.*

Proof. For any $\lambda \in \mathrm{eig}(A)$, let $x \neq 0$ be the associated eigenvector, i.e., $Ax = \lambda x$. Let $1 = \|x\|_\infty = x_k$ for some k by scaling x if necessary. (It also means that $|x_j| \leq 1$ for $j \neq k$.) Then from the kth row of the equation $Ax = \lambda x$, we have $\sum_{j=1}^N a_{kj}x_j = \lambda x_k = \lambda$, so $\lambda - a_{kk} = \sum_{j \neq k} a_{kj}x_j$. Hence

$$|\lambda - a_{kk}| \leq \sum_{j \neq k} |a_{kj}x_j| \leq \sum_{j \neq k} |a_{kj}|.$$

That is λ is in the kth disk. Since $\lambda \in \mathrm{eig}(A)$ is arbitrary, the theorem is proved. □

Furthermore, it can also be shown that if the ith Gershgorin disk is isolated from the other disks, then it contains precisely one of A's eigenvalue. Its proof uses the fact that an eigenvalue of a matrix is continuous a function of the elements of the matrix.

3.2 Eigenvalue sensitivity

Theorem 3.3. *Let A be an $n \times n$ matrix, and let λ be A's simple eigenvalue with corresponding (right) eigenvector x and left eigenvector y. Suppose A is perturbed to $\widetilde{A} \equiv A + \delta A$, and consequently λ is perturbed to $\widetilde{\lambda} \equiv \lambda + \delta\lambda$. If $\|\delta A\|_2 = \epsilon$ is sufficiently small, then*

$$\delta\lambda = \frac{y^H(\delta A)x}{y^H x} + \mathcal{O}(\epsilon^2).$$

This implies

$$|\delta\lambda| \le \frac{\|y\|_2\|x\|_2}{|y^H x|}\|\delta A\|_2 + \mathcal{O}(\epsilon^2).$$

Proof. Let $\widetilde{x} \equiv x + \delta x$ be $\widetilde{A}$'s eigenvector corresponding to $\widetilde{\lambda}$. We have

$$\widetilde{A}\widetilde{x} = \widetilde{\lambda}\widetilde{x} \quad \Rightarrow \quad (A + \delta A)(x + \delta x) = (\lambda + \delta\lambda)(x + \delta x),$$

expanding which leads to

$$Ax + A\,\delta x + \delta A\,x + \delta A\,\delta x = \lambda x + \lambda\,\delta x + \delta\lambda\,x + \delta\lambda\,\delta x.$$

Ignoring second order terms – $\delta A\,\delta x$ and $\delta\lambda\,\delta x$ – and noticing that $Ax = \lambda x$, we have

$$A\,\delta x + \delta A\,x = \lambda\,\delta x + \delta\lambda\,x + \mathcal{O}(\epsilon^2),$$

pre-multiplying the equation by y^H and noticing $y^H A = y^H \lambda$ to get

$$y^H \delta A\,x = \delta\lambda\,y^H x + \mathcal{O}(\epsilon^2),$$

as required. $\square$

Definition 3.1. *In Theorem 3.3, define*

$$s_\lambda \stackrel{\text{def}}{=} \frac{|y^H x|}{\|x\|_2\|y\|_2} \quad \textit{and} \quad \text{cond}(\lambda) \stackrel{\text{def}}{=} \frac{1}{s_\lambda}.$$

$\text{cond}(\lambda)$ *is called λ's* ***individual condition number***.

Example 3.3. Consider

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 4.001 \end{pmatrix} \quad \text{is perturbed by} \quad \delta A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.001 & 0 & 0 \end{pmatrix}.$$

A's eigenvalues are easily read and its eigenvectors can be computed:

$$\lambda_1 = 1,\ x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ y_1 = \begin{pmatrix} 0.8285 \\ -0.5523 \\ 0.0920 \end{pmatrix},$$

$$\lambda_2 = 4,\ x_2 = \begin{pmatrix} 0.5547 \\ 0.8321 \\ 0 \end{pmatrix},\ y_2 = \begin{pmatrix} 0 \\ 0.0002 \\ -1.0000 \end{pmatrix},$$

$$\lambda_3 = 4.001,\ x_3 = \begin{pmatrix} 0.5547 \\ 0.8321 \\ 0.0002 \end{pmatrix},\ y_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

On the other hand, eigenvalues of $\widetilde{A} = A + \delta A$ computed by MATLAB's `eig` are

$$\widetilde{\lambda}_1 = 1.0001, \quad \widetilde{\lambda}_2 = 3.9427, \quad \widetilde{\lambda}_3 = 4.0582.$$

The following table compares $|\widetilde{\lambda}_i - \lambda_i|$ with its first order upper bound $\text{cond}(\lambda_i)\|\delta A\|_2$.

i	s_{λ_i}	$\text{cond}(\lambda_i)$	$\text{cond}(\lambda_i)\|\delta A\|_2$	$\lvert\widetilde{\lambda}_i - \lambda_i\rvert$
1	0.8285	1.2070	0.0012	0.0001
2	$1.7 \cdot 10^{-4}$	$6.0 \cdot 10^3$	6.0	0.057
3	$1.7 \cdot 10^{-4}$	$6.0 \cdot 10^3$	6.0	0.057

Theorem 3.3 is useful only for sufficiently small $\|\delta A\|$. Such kind of error bound is called an ***asymptotic error bound***. Sometimes we can remove the $\mathcal{O}(\epsilon^2)$ term and get an error bound which can be applied for any size perturbation $\|\delta A\|$. It is called a ***global error bound***. Under the condition that A is diagonalizable, we have

Theorem 3.4 (Bauer-Fike)**.** *Suppose A is $n \times n$ and diagonalizable, i.e., it has n linearly independent eigenvectors. Its eigentriplets are denoted by (λ_i, x_i, y_i), normalized so that $\|x_i\|_2 = \|y_i\|_2 = 1$ and $y_i^H x_j = 0$ for $i \neq j$. Then eigenvalues of $A + \delta A$ lie in the union of the disk $\mathscr{D}_i$ for $1 \leq i \leq n$, where $\mathscr{D}_i$ has center λ_i and radius $n\frac{\|\delta A\|_2}{|y_i^H x_i|}$.*

Proof. It can be verified that

$$X^{-1} = \begin{pmatrix} y_1^H/(y_1^H x_1) \\ y_2^H/(y_2^H x_2) \\ \vdots \\ y_n^H/(y_n^H x_n) \end{pmatrix}, \quad A = X\Lambda X^{-1}, \quad \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Since $A + \delta A = X[\Lambda + X^{-1}(\delta A)X]X^{-1}$, the eigenvalues of $A + \delta A$ are the same as those of $\Lambda + X^{-1}(\delta A)X$. The ith Gerschgorin disk for the latter is specified by

$$|\lambda - [\lambda_i + (X^{-1}(\delta A)X)_{(i,i)}]| \leq \sum_{j \neq i} |(X^{-1}(\delta A)X)_{(i,j)}|$$

which implies

$$\begin{aligned} |\lambda - \lambda_i| &\leq \|e_i^T(X^{-1}(\delta A)X)\|_\infty \\ &\leq \sqrt{n}\|e_i^T X^{-1}(\delta A)X\|_2 \\ &\leq \sqrt{n}\|y_i^H/(y_i^H x_i)\|_2\|\delta A\|_2\|X\|_F \\ &= n\frac{\|\delta A\|_2}{|y_i^H x_i|}, \end{aligned}$$

as expected. $\square$

3.3 Eigenvector sensitivity

Eigenvectors may not be uniquely determined in the case of a repeated eigenvalue. Let us start with a couple of examples.

Example 3.4. Consider

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{is perturbed to} \quad \widetilde{A} = A + \delta A = \begin{pmatrix} 1+\epsilon & 0 \\ 0 & 1-\epsilon \end{pmatrix},$$

where $\epsilon \neq 0$ and small. Any vector is A's eigenvector; while $\widetilde{A}$'s eigenvectors are $e_1 = (1,0)^T$ and $e_2 = (0,1)^T$. (Why?) If we unfortunately choose an eigenvector of A, say $x = (1,1)^T$, we would find it has nothing to do with any of $\widetilde{A}$'s eigenvector even though $\widetilde{A}$ and A could be made arbitrarily close.

This example alerts us that care must be taken when talking about changes in eigenvectors of a repeated eigenvalue.

Example 3.5 (Watkins)**.** Consider

$$A = \begin{pmatrix} 1+\epsilon & 0 \\ 0 & 1-\epsilon \end{pmatrix} \quad \text{is perturbed to} \quad \widetilde{A} = A + \delta A = \begin{pmatrix} 1+\epsilon & \sqrt{3}\epsilon \\ \sqrt{3}\epsilon & 1-\epsilon \end{pmatrix},$$

where $\epsilon \neq 0$ and small.

A eigenvalues	A eigenvectors	$\widetilde{A}$ eigenvalues	$\widetilde{A}$ eigenvectors
$1+\epsilon,\ 1-\epsilon$	$e_1,\ e_2$	$1+2\epsilon,\ 1-2\epsilon$	$\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2,\ \frac{\sqrt{3}}{2}e_1 - \frac{1}{2}e_2$

What we see here is that the perturbation δA makes small changes to A's eigenvalues but its eigenvectors suffer enormous changes. This is due to *the closeness of A's eigenvalues* as will be clear from the next theorem which roughly says *the changes in an eigenvector is proportional to the reciprocal of the distance between the corresponding eigenvalue and the rest of A's eigenvalues*. Note this theorem assumes that A has distinct eigenvalues. It does get more complicated otherwise.

Theorem 3.5. *Suppose $n \times n$ matrix A has n distinct eigenvalues with corresponding (left and right) eigenvectors:*

$$\begin{aligned} \textit{eigenvalues:} &\quad \lambda_1, \quad \lambda_2, \quad \cdots, \quad \lambda_n, \\ \textit{(right) eigenvalues:} &\quad x_1, \quad x_2, \quad \cdots, \quad x_n, \\ \textit{left eigenvalues:} &\quad y_1, \quad y_2, \quad \cdots, \quad y_n. \end{aligned}$$

Normalize x_i and y_i such that $\|x_i\|_2 = 1 = \|y_i\|_2$ for all i. Suppose A is perturbed to $\widetilde{A} = A + \delta A$. If $\|\delta A\|_2 = \epsilon$ is sufficiently small, then for $k = 1, 2, \cdots, n$

$$\widetilde{x}_k = x_k + \sum_{j \neq k} \frac{y_j^H (\delta A) x_k}{(\lambda_k - \lambda_j) y_j^H x_j} + \mathcal{O}(\epsilon^2).$$

These equations imply that for $k = 1, 2, \cdots, n$

$$\|\widetilde{x}_k - x_k\|_2 \leq \left(\sum_{j \neq k} \frac{\operatorname{cond}(\lambda_j)}{|\lambda_k - \lambda_j|} \right) \epsilon + \mathcal{O}(\epsilon^2).$$

Proof. The conditions of this theorem imply that each λ_j is a simple eigenvalue of A. Thus by Theorem 3.1,

$$y_j^H x_k = 0 \quad \text{if } k \neq j, \text{ and} \quad y_k^H x_k \neq 0. \tag{3.2}$$

Notice also that A's eigenvectors $x_1, x_2, \cdots, x_n$ form a basis of the n-dimensional space since A's eigenvalues are pairwise distinct. (Why?)

Denote that under the perturbation δA, λ_k is changed to $\widetilde{\lambda}_k \equiv \lambda_k + \delta\lambda_k$ and x_k to $\widetilde{x}_k \equiv x_k + \delta x_k$. Write

$$\delta x_k = \sum_{i=1}^{n} c_i x_i,$$

where $c_1, c_2, \cdots, c_n$ are small coefficients. Thus $x_k + \delta x_k = (1 + c_k)x_k + \sum_{i \neq k} c_i x_i$. We may normalize $x_k + \delta x_k$ by setting $c_k = 0$, since *eigenvectors are determined only up to a scalar multiple* and if $c_k \neq 0$ we can always consider

$$\frac{(1 + c_k)x_k + \sum_{i \neq k} c_i x_i}{1 + c_k} = x_k + \sum_{i \neq k} \frac{c_k}{1 + c_k} x_i \stackrel{\text{def}}{=} x_k + \sum_{i \neq k} c_k' \, x_i$$

instead. Thus

$$\delta x_k = \sum_{i \neq k} c_i x_i. \tag{3.3}$$

The task is to find those c_i's. To this end, we expand

$$(A + \delta A)(x_k + \delta x_k) = (\lambda_k + \delta \lambda_k)(x_k + \delta x_k)$$

as before to get

$$A \delta x_k + \delta A\, x_k = \lambda_k\, \delta x_k + \delta \lambda_k\, x_k + \mathcal{O}(\epsilon^2),$$

substituting (3.3) into which gives

$$\sum_{i \neq k} c_i \lambda_i x_i + \delta A x_k = \lambda_k \sum_{i \neq k} c_i x_i + \delta \lambda_k\, x_k + \mathcal{O}(\epsilon^2),$$

since $A x_i = \lambda_i x_i$. We have

$$\sum_{i \neq k} c_i (\lambda_k - \lambda_i) x_i = \delta A x_k - \delta \lambda_k\, x_k + \mathcal{O}(\epsilon^2).$$

Pre-multiplying the equation by y_j^H ($j \neq k$), together with (3.2), yield

$$c_j (\lambda_k - \lambda_j) y_j^H x_j = y_j^H \delta A x_k - 0 + \mathcal{O}(\epsilon^2),$$

i.e.,

$$c_j = \frac{y_j^H \delta A x_k}{(\lambda_k - \lambda_j) y_j^H x_j} + \mathcal{O}(\epsilon^2),$$

as required. $\square$

This theorem suggests that the sensitivity of x_k depends upon eigenvalue sensitivity and the separation of λ_k from the other eigenvalues.

Exercises

3.1. Another version of the Bauer-Fike theorem reads as follows. Suppose A is $n \times n$ and diagonalizable and its eigendecomposition is given by (2.1). Then the eigenvalues of $A + \delta A$ are in the union of disks

$$\{\lambda \,:\, |\lambda - \lambda_i| \leq \|X^{-1} \cdot \delta A \cdot X\|_2\}.$$

Prove this version of the Bauer-Fike theorem. Find out what other matrix norms can be used in the place of $\|\cdot\|_2$ here.

4 Symmetric eigenvalue problems

Throughout this section A is a real symmetric matrix.

First, it's easy to see that the eigen-decomposition (the Schur Decomposition) of a real symmetric matrix A is given by

$$A = Q \Lambda Q^T,$$

where $Q = [q_1, q_2, \ldots, q_n]$ is orthogonal, and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Furthermore, all λ_i are real, and without loss of generality, assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Eigenpairs are denoted as $\{\lambda_i, q_i\}$ and $\|A\|_2 = |\lambda_1|$.

Definition 4.1. *The **inertia** of a symmetric matrix A is the triplet of integers:*

$$\mathit{Inertia}(A) \equiv (\nu, \zeta, \pi),$$

where

$$\begin{aligned} \nu &= \text{the number of negative eigenvalues of } A, \\ \zeta &= \text{the number of zero eigenvalues of } A, \\ \pi &= \text{the number of positive eigenvalues of } A. \end{aligned}$$

Recall that $X^T AX$ is called the ***congruence transformation*** of A.

Theorem 4.1 (Sylvester's Law of Inertia)**.** *Let A be symmetric and X be nonsingular. Then A and $X^T AX$ have the same inertia.*

Definition 4.2. *The **Rayleigh Quotient** of a symmetric matrix A and a nonzero vector u is*

$$\rho(u, A) = \frac{u^T Au}{u^T u}$$

Sometimes $\rho(u, A)$ is also written as $\rho(u)$, if the matrix A is understood from the context.

Basic properties of Rayleigh quotient (verify!):

- If q is an eigenvalue of A, then $\rho(q)$ is the corresponding eigenvalue.
- Given any vector $u \neq 0$, then

$$\min_{\sigma} \|Au - \sigma u\|_2 = \|Au - \rho(u)u\|_2$$

- $\lambda_n \leq \rho(u) \leq \lambda_1$ for any $u \neq 0$
- $\lambda_1 = \max_u \rho(u)$ and $\lambda_n = \min_u \rho(u)$.

More generally, we have

Theorem 4.2 (Courant-Fischer Minimax/Maximin Theorem)**.**

$$\max_{\mathbb{S}^j} \min_{0 \neq u \in \mathbb{S}^j} \rho(u, A) = \lambda_j = \min_{\mathbb{S}^{n-j+1}} \max_{0 \neq u \in \mathbb{S}^{n-j+1}} \rho(u, A)$$

where $\mathbb{S}^j$ denotes a j-dimensional subspace of $\mathbb{R}^n$.

Remark **4.1.**

1. The maximum in the first expression for λ_j is over all j-dimensional subspaces $\mathbb{S}^j$ of $\mathbb{R}^n$, and the subsequent minimum is over all nonzero vector u in the subspace. The maximum is attained for $\mathbb{S}^j = \text{span}(q_1, q_2, \ldots, q_j)$, and a minimizing u is $u = q_j$.
2. The minimum in the second expression for λ_j is over all $(n-j+1)$-dimensional subspaces $\mathbb{S}^{n-j+1}$ of $\mathbb{R}^n$, and the subsequent maximum is over all nonzero vector s in the subspace. The minimum is attained for $\mathbb{S}^{n-j+1} = \text{span}(q_j, q_{j+1}, \ldots, q_n)$, and a maximizing u is $u = q_j$.

4.1 Sensitivity of eigenvalues

Let A and E be n-by-n symmetric matrices, and let $\{\lambda_i\}$ and $\{\widetilde{\lambda}_i\}$ be the eigenvalues of A and $A+E$, respectively. Then note that $\text{cond}(\lambda_i) = 1$ for all i. From the first order perturbation analysis of a general matrix A, we immediately have

$$|\lambda_j - \widetilde{\lambda}_i| \leq \|E\|_2 + O(\|E\|_2^2)$$

The above result is weak in three aspects: 1) it does not say $j = i$, and 2) it is under the assumption that sufficiently small $\|E\|_2$ is sufficiently small, and 3) it is an asymptotical bound. The next theorem eliminates all three weaknesses.

Theorem 4.3 (Weyl). *Let A and E be n-by-n real symmetric matrices, and let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and $\widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq \cdots \geq \widetilde{\lambda}_n$ be the eigenvalues of A and $A+E$, respectively, then*

$$|\lambda_i - \widetilde{\lambda}_i| \leq \|E\|_2.$$

4.2 Sensitivity of eigenvectors

Let $A = Q\Lambda Q^T = Q\,\text{diag}(\lambda_i)Q^T$ and $\widetilde{A} = A + E = \widetilde{Q}\widetilde{\Lambda}\widetilde{Q}^T = \widetilde{Q}\text{diag}(\widetilde{\lambda}_i)\widetilde{Q}^T$ be the eigen-decompositions of A and $A+E$, respectively. Write

$$Q = (q_1, q_2, \ldots, q_n) \quad \text{and} \quad \widetilde{Q} = (\widetilde{q}_1, \widetilde{q}_2, \ldots, \widetilde{q}_n).$$

Note q_i and $\widetilde{q}_i$ are the original and perturbed eigenvectors, respectively.

Theorem 4.4 (Davis-Kahan). *Let θ_i denote the acute angle between q_i and $\widetilde{q}_i$. Then*

$$\frac{1}{2}\sin 2\theta_i \leq \frac{\|E\|_2}{\min_{j\neq i}|\lambda_j - \lambda_i|}, \tag{4.4}$$

provided that $\min_{j\neq i}|\lambda_j - \lambda_i| > 0$.

Remark **4.2.**

1. When $\theta_i \ll 1$, then $\frac{1}{2}\sin 2\theta_i \approx \sin\theta_i \approx \theta_i$.

2. $\min_{j\neq i}|\lambda_j - \lambda_i| > 0$ is called the ***gap*** between the eigenvalue λ_i and the rest of the spectrum. It is sometimes written as $\text{gap}(i, A)$ or $\text{gap}(i)$ if A is understood from the context. The upper bound (4.4) is often written as

$$\frac{1}{2}\sin 2\theta_i \leq \frac{\|E\|_2}{\text{gap}(i, A)}.$$

3. By considering $A+E$ as the unperturbed matrix and $A = (A+E) - E$ as the perturbed matrix. then the upper bound (4.4) can also be written

$$\frac{1}{2}\sin 2\theta_i \leq \frac{\|E\|_2}{\min_{j\neq i}|\widetilde{\lambda}_j - \widetilde{\lambda}_i|} \equiv \frac{\|E\|_2}{\text{gap}(i, A+E)}.$$

   The attraction of stating the bound in terms of $\text{gap}(i, A+E)$ is that frequently we know only the eigenvalues of $A+E$, since they are typically the output of an eigenvalue algorithm that we have used. In this case, it is straightforward to evaluate $\text{gap}(i, A+E)$, whereas we can only estimate $\text{gap}(i, A)$.

Exercises

4.1. Show that for any scalar σ and any nonzero vector x, there is an eigenvalue of λ of A satisfying

$$|\lambda - \sigma| \leq \|Ax - x\sigma\|_2 / \|x\|_2.$$

5 Generalized eigenvalue problems

The standard eigenvalue problem asks for which scalars λ the matrix $A - \lambda I$ is singular; these scalars are the eigenvalues. This notion generalizes in several important ways:

Definition 5.1. *Given $m \times n$ matrices A and B, $A - \lambda B$ is called a **matrix pencil**. Here λ is an indeterminate, not a particular numerical value.*

Definition 5.2. *Let A and B be $n \times n$.*

- *If A and B are square and $\det(A - \lambda B)$ is not identically zero, then pencil $A - \lambda B$ is called **regular**. Otherwise it is called **singular**.*
- *When $A - \lambda B$ is regular, $p(\lambda) = \det(A - \lambda B)$ is called the **characteristic polynomial** of $A - \lambda B$.*
- *The **eigenvalues** of $A - \lambda B$ are defined to be*
  1. *the roots of $p(\lambda) = 0$*
  2. *∞ (with algebraic multiplicity $n - deg(p)$) if $deg(p) < n$.*
- *The set of the eigenvalues of $A - \lambda B$ is denoted as $\mathrm{eig}(A, B)$.*

Example 5.1.

1. For $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathrm{eig}(A, B) = \{1, 3\}$
2. For $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\mathrm{eig}(A, B) = \{1, \infty\}$.
3. For $A = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathrm{eig}(A, B) = \{\infty, \infty\}$.
4. For $A = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $p(\lambda) = \det(A - \lambda B) \equiv 0$ for any λ. It is a singular pencil.

Definition 5.3. *Let λ be a finite eigenvalue of the regular pencil $A - \lambda B$. We say $x \neq 0$ is a **right eigenvector** if*

$$Ax = \lambda Bx.$$

*If $\lambda = \infty$ is an eigenvalue and $Bx = 0$, then x is a right eigenvector. A **left eigenvector** of $A - \lambda B$ is a right eigenvector of $(A - \lambda B)^H$, i.e.*

$$y^H A = \lambda y^H B$$

The following proposition relates the eigenvalues of a regular pencil $A - \lambda B$ to the eigenvalues of a single matrix.

Proposition 5.1. *Suppose $A - \lambda B$ is an $n \times n$ regular pencil.*

1. *If B is nonsingular, all eigenvalues of $A - \lambda B$ are finite and the same as the eigenvalues of AB^{-1} or $B^{-1}A$.*

2. *If B is singular, $A - \lambda B$ has eigenvalue ∞ with geometric multiplicity $n - \mathrm{rank}(B)$.*

3. *If A is nonsingular, the eigenvalues of $A - \lambda B$ are the same as reciprocals of the eigenvalues of $A^{-1}B$ or BA^{-1}, where a zero eigenvalue of $A^{-1}B$ corresponds to an infinite eigenvalue of $A - \lambda B$.*

Recall that all of our theory and algorithms for the eigenvalue problem of a single matrix depend on finding a similarity transformation $S^{-1}AS$ of A that is in a "simpler" form. The next definition shows how to generalize the notion of similarity to matrix pencils.

Definition 5.4. *Let Y and X be nonsingular matrices. Then pencils $A - \lambda B$ and $YAX - \lambda YBX$ are called **equivalent**.*

Proposition 5.2.

- *The equivalent regular pencils $A - \lambda B$ and $YAX - \lambda YBX$ have the same eigenvalues.*

- *The vector x is a right eigenvector of $A - \lambda B$ if and only if $X^{-1}x$ is a right eigenvector of $YAX - \lambda YBX$.*

- *The vector y is a left eigenvector of $A - \lambda B$ if and only if $Y^{-H}x$ is a left eigenvector of $YAX - \lambda YBX$.*

Theorem 5.1 (Generalized Schur Decomposition)**.** *For any $A, B \in \mathbb{C}^{n \times n}$, there exist unitary matrices Q and Z so that*

$$Q^H AZ = T \quad \textit{and} \quad Q^H BZ = S$$

are both upper triangular. The eigenvalues of $A - \lambda B$ are then ratios of the diagonal entries of T and S.

REMARK **5.1.**

1. If A and B are real, there is a generalized real Schur form, too: Q and Z are (real) orthogonal, and that $Q^T AZ$ is quasi-upper triangular and $Q^T BZ$ is upper triangular.

2. ***Weierstrass canonical form*** for a regular matrix pencil is a generalization of Jordan canonical form. ***Kronecker canonical form*** is for a singular pencil.

Definition 5.5. *Matrix pencil $A - \lambda B$ with Hermitian A and B such that*

$$\min_{\substack{x \in \mathbb{C}^n \\ \|x\|_2 = 1}} |x^H (A + iB)x| > 0.$$

*is called a **definite pencil**.*

Example 5.2. A special case that often arises in practice is matrix pencil $A - \lambda B$ with real symmetric A and B and B positive definite. Such a pencil is evidently a definite pencil.

Theorem 5.2. *Let $A - \lambda B$ be a definite pencil. Then there is a nonsingular matrix X so that*

$$X^H AX = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n) \quad \textit{and} \quad X^H BX = \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_n).$$

In particular, all the eigenvalues of $A - \lambda B$ are real, and they are finite if B is nonsingular.

6 Further reading

For more studies of perturbation theory, see

- G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press, Boston, 1990.
- J.-G. Sun, *Matrix Perturbation Analysis* (Second Edition), Science Press, Beijing, China, 2001 (in Chinese)