THE MATRIX EIGENVALUE PROBLEM

Find scalars λ and vectors $x \neq 0$ for which

$$Ax = \lambda x$$

The form of the matrix affects the way in which we solve this problem, and we also have variety as to what is to be found.

- A symmetric and real (or Hermitian and complex). This is the most common case. In some cases we want only the eigenvalues (and perhaps only some of them); and in other cases, we also want the eigenvectors. There are special classes of such A, e.g. banded, positive definite, sparse, and others.

- A non-symmetric, but with a diagonal Jordan canonical form. This means there is a nonsingular matrix P for which

$$P^{-1}AP = D = \text{diag}[\lambda_1, ..., \lambda_n]$$

Then $AP = PD$ and the columns of P are the eigenvectors of A.
As we see later, these matrix eigenvalue problems may be ill-conditioned. There are special subclasses of problems, as with the symmetric case. Note that when A is real, the complex eigenvalues (if any) must occur in conjugate pairs.

- A non-symmetric and the Jordan canonical form is not diagonal. These are very difficult problems, especially when calculating the eigenvectors.

<u>GENERAL APPROACH</u>. Begin by finding the eigenvalues. Then find the eigenvectors, if they are needed.

<u>Finding the eigenvalues</u>. Proceed in two steps.
(1) Reduce A to a simpler form T, usually using orthogonal similarity transformations.
(2) Apply some method to finding the eigenvalues of T.

GERSCHGORIN'S CIRCLE THEOREM

Where are the eigenvalues of A located? We know that for any matrix norm $\|\cdot\|$,

$$\max_{\lambda\in\sigma(A)} |\lambda| \leq \|A\|$$

How can this be improved? Let

$$r_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} \left|a_{i,j}\right|, \quad i = 1, ..., n$$

$$Z_i = \left\{z \in \mathbb{C} \mid \left|z - a_{i,i}\right| \leq r_i\right\}, \quad i = 1, ..., n$$

The set Z_i is a circle with center $a_{i,i}$ and radius r_i. Then the eigenvalues of A are located in the union of the circles Z_i:

$$\lambda \in \sigma(A) \;\Rightarrow\; \lambda \in \bigcup_{i=1}^{n} Z_i$$

Moreover, break this union into disjoint components, say $C_1, ..., C_m$. Then each such component contains exactly as many eigenvalues as circles Z_i.

PROOF. Let $Ax = \lambda x$, $x \neq 0$. Let k be an index for which

$$\|x\|_\infty = |x_k| > 0$$

From $Ax = \lambda x$,

$$\sum_{j=1}^{n} a_{i,j} x_j = \lambda x_i, \quad i = 1, ..., n$$

Solve equation k for x_k:

$$\left(\lambda - a_{k,k}\right) x_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{k,j} x_j$$

Taking absolute values,

$$\left|\lambda - a_{k,k}\right| |x_k| \leq \sum_{\substack{j=1 \\ j \neq k}}^{n} \left|a_{k,j}\right| \left|x_j\right| \leq \sum_{\substack{j=1 \\ j \neq k}}^{n} \left|a_{k,j}\right| |x_k|$$

Cancel $|x_k|$ to get

$$\left|\lambda - a_{k,k}\right| \leq \sum_{\substack{j=1 \\ j \neq k}}^{n} \left|a_{k,j}\right| = r_k$$

EXAMPLE

Recall the matrix

$$A = \begin{bmatrix} c & 1 & 0 & 0 & 0 \\ 1 & c & 1 & 0 & 0 \\ 0 & 1 & c & 1 & 0 \\ 0 & 0 & 1 & c & 1 \\ 0 & 0 & 0 & 1 & c \end{bmatrix}$$

which is used as an example at the end of Chapter 7. In this case,

$$r_1 = r_5 = 1; \quad r_i = 2, \quad i = 2, 3, 4$$

The centers of the circles are all $a_{i,i} = c$. Then the union of the circles Z_i is the circle

$$\{z \mid |z - c| \leq 2\}$$

The matrix A is real and symmetric, and thus all eigenvalues are real. Thus the eigenvalues λ must be located in the interval

$$c - 2 \leq \lambda \leq c + 2$$

BAUER-FIKE THEOREM

Assume A has a diagonal Jordan canonical form, meaning there is a nonsingular matrix P for which

$$P^{-1}AP = D = \text{diag}[\lambda_1, ..., \lambda_n]$$

Assume we are using a matrix norm for which

$$\|D\| = \max_{1 \le i \le n} |\lambda_i|$$

Then consider the eigenvalues λ of the perturbed matrix $A + E$. For such λ, we have

$$\min_{1 \le i \le n} |\lambda - \lambda_i| \le \|P\| \left\|P^{-1}\right\| \|E\|$$

PROOF. Write

$$\begin{aligned} (A + E)\, x &= \lambda x, \quad x \ne 0 \\ (\lambda I - A)\, x &= Ex \end{aligned}$$

Assume $\lambda \neq \lambda_1, ..., \lambda_n$, as otherwise the theorem is easily true. Substitute $A = PDP^{-1}$,

$$\begin{aligned}\left(\lambda I - PDP^{-1}\right) x &= Ex \\ (\lambda I - D)\left(P^{-1}x\right) &= \left(P^{-1}EP\right)\left(P^{-1}x\right) \\ P^{-1}x &= (\lambda I - D)^{-1}\left(P^{-1}EP\right)\left(P^{-1}x\right)\end{aligned}$$

Take norms of both sides,

$$\left\|P^{-1}x\right\| \leq \left\|(\lambda I - D)^{-1}\right\| \left\|P^{-1}EP\right\| \left\|P^{-1}x\right\|$$

Cancel $\left\|P^{-1}x\right\|$,

$$1 \leq \left\|(\lambda I - D)^{-1}\right\| \left\|P^{-1}EP\right\|$$

Also note that by our assumption on the matrix norm,

$$\left\|(\lambda I - D)^{-1}\right\| = \max_i \frac{1}{|\lambda - \lambda_i|} = \frac{1}{\min_i |\lambda - \lambda_i|}$$

Then

$$\min_i |\lambda - \lambda_i| \leq \left\|P^{-1}EP\right\| \leq \|P\| \left\|P^{-1}\right\| \|E\|$$

This completes the proof.

Consider the case in which A is symmetric and real. Then the matrix P can be chosen to be orthogonal, and $P^{-1} = P^T$. If we use the matrix norm $\|\cdot\|_2$ induced by the Euclidean vector norm $\|\cdot\|_2$, then from Problem 13 of Chapter 7, $\|P\|_2 = 1$. Thus for this particular matrix norm,

$$\min_i |\lambda - \lambda_i| \leq \left\|P^{-1}EP\right\|_2 \leq \|E\|_2 = \text{sqrt}\left[r_\sigma\left(E^T E\right)\right]$$

Thus small changes in the matrix lead to small changes in the eigenvalues.

We can also use the inequality

$$\min_i |\lambda - \lambda_i| \leq \|P\| \left\|P^{-1}\right\| \|E\|$$

to define a condition number for the eigenvalue problem. For it, we would use

$$\text{cond}\,(A) = \inf_{P^{-1}AP=D} \|P\| \left\|P^{-1}\right\|$$

Then

$$\min_i |\lambda - \lambda_i| \leq \text{cond}\,(A)\,\|E\|$$

This says the changes in the eigenvalues are small. But there may still be a large relative change.

From the book, consider the 3×3 Hilbert matrix and its version rounded to four decimal digits.

$$H_3 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

$$\overline{H}_3 = \begin{bmatrix} 1.000 & .5000 & .3333 \\ .5000 & .3333 & .2500 \\ .3333 & .2500 & .2000 \end{bmatrix}$$

In this case, $A = H_3$ and $A + E = \overline{H}_3$, or $E = \overline{H}_3 - H_3$. Using the matrix norm $\|\cdot\|_2$, the Bauer-Fike result says that for each eigenvalue λ of $\overline{H}_3$,

$$\min_i |\lambda - \lambda_i| \le \|E\|_2 = 3.3 \times 10^{-5}$$

In fact the true eigenvalues of H_3 are

$$\lambda_1 = 1.408319, \quad \lambda_2 = .1223271, \quad \lambda_3 = .002687340$$

and the true eigenvalues of $\overline{H}_3$ are

$$\widehat{\lambda}_1 = 1.408294, \quad \widehat{\lambda}_2 = .1223415, \quad \widehat{\lambda}_3 = .002664489$$

For the errors,

$$\begin{aligned} \lambda_1 - \widehat{\lambda}_1 &= 2.49 \times 10^{-5} \\ \lambda_2 - \widehat{\lambda}_2 &= -1.44 \times 10^{-5} \\ \lambda_3 - \widehat{\lambda}_3 &= 2.29 \times 10^{-5} \end{aligned}$$

which is in agreement with

$$\min_i |\lambda - \lambda_i| \le \|E\|_2 = 3.3 \times 10^{-5}$$

For the relative errors,

$$\text{Rel}\left(\widehat{\lambda}_1\right) = 1.77 \times 10^{-5}, \quad \text{Rel}\left(\widehat{\lambda}_2\right) = -1.18 \times 10^{-4}$$

$$\text{Rel}\left(\widehat{\lambda}_3\right) = 8.5 \times 10^{-3}$$

EXAMPLE

$$A = \begin{bmatrix} 101 & -90 \\ 110 & -98 \end{bmatrix}, \quad A + E = \begin{bmatrix} 100.999 & -90.001 \\ 110 & -98 \end{bmatrix}$$

For A, the eigenvalues are $1, 2$. For $A + E$, the eigenvalues are

$$\lambda \doteq 1.298,\ 1.701$$

This is a very significant change in eigenvalues for a very small change in the matrix. It is illustrative of what can happen with the non-symmetric eigenvalue problem.

WIELANDT-HOFFMAN THEOREM

Let A and E be real and symmetric, and let $\widehat{A} = A + E$. Let the eigenvalues of A be

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

and let those of $\widehat{A}$ be

$$\widehat{\lambda}_1 \leq \widehat{\lambda}_2 \leq \cdots \leq \widehat{\lambda}_n$$

Then

$$\left[\sum_{j=1}^{n} \left(\lambda_j - \widehat{\lambda}_j\right)^2\right]^{\frac{1}{2}} \leq F(E) \equiv \left[\sum_{i=1}^{n} \sum_{j=1}^{n} \left|E_{i,j}\right|^2\right]^{\frac{1}{2}}$$

EXAMPLE - NONSYMMETRIC

Consider the $n \times n$ matrix

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 0 & & 0 \\ & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \\ & & & & 1 & 1 \\ 0 & & \cdots & & 0 & 1 \end{bmatrix}$$

Its characteristic polynomial is

$$f(\lambda) = (1 - \lambda)^n$$

Its only eigenvalue is $\lambda = 1$; and there is only a one-dimensional family of eigenvectors, all multiples of

$$x = [1, 0, \cdots, 0]^T$$

Now perturb the matrix to

$$A(\epsilon) = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 0 & & 0 \\ & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \\ 0 & & & & 1 & 1 \\ \epsilon & 0 & \cdots & & 0 & 1 \end{bmatrix}$$

Its characteristic polynomial is

$$f_\epsilon(\lambda) = (1-\lambda)^n - (-1)^n \epsilon$$

Its roots, and the eigenvalues of $A(\epsilon)$, are

$$\lambda_k(\epsilon) = 1 + \omega_k \epsilon^{1/n}, \quad k = 1, ..., n$$

with $\{\omega_k\}$ the n^{th} roots of unity,

$$\omega_k = e^{2\pi ki/n}, \quad k = 1, ..., n$$

Thus

$$|\lambda_k - \lambda_k(\epsilon)| = |\epsilon|^{1/n}$$

For $n = 10$ and $\epsilon = 10^{-10}$, $|\lambda_k - \lambda_k(\epsilon)| = 0.1$.

STABILITY FOR NONSYMMETRIC MATRICES

Assume the matrix A has a diagonal Jordan canonical form:

$$P^{-1}AP = D = \text{diag}[\lambda_1, ..., \lambda_n]$$

Let $P = [u_1, ..., u_n]$. Then $AP = PD$ implies

$$Au_i = \lambda_i u_i, \quad i = 1, ..., n$$

and the vectors $\{u_1, ..., u_n\}$ form a basis of $\mathbb{C}^n$.

To see some of the nonuniqueness in the choice of P, let F be an arbitrary nonsingular diagonal matrix,

$$F = \text{diag}[f_1, ..., f_n]$$

Then

$$\begin{aligned} F^{-1}P^{-1}APF &= F^{-1}DF \\ (PF)^{-1}A(PF) &= D \end{aligned}$$

The matrix PF is another nonsingular matrix; and since F is diagonal,

$$PF = [u_1, ..., u_n]\, F = [f_1 u_1, ..., f_n u_n]$$

The vectors $f_i u_i$ are again eigenvectors of A. Therefore, we assume that P has been so chosen that the vectors u_i all have Euclidean length 1:

$$u_i^* u_i = 1, \quad i = 1, ..., n$$

Note that because the eigenvalues can be complex, we must now work in $\mathbb{C}^n$; and we also allow A to be complex.

Form the complex conjugate transpose of $P^{-1}AP = D$:

$$P^* A^* \left(P^*\right)^{-1} = D^* = \text{diag}[\overline{\lambda}_1, ..., \overline{\lambda}_n]$$

Write

$$\left(P^*\right)^{-1} = [w_1, ..., w_n]$$

Then as before with A, we have

$$\begin{aligned} A^* w_i &= \overline{\lambda}_i w_i, \quad i = 1, ..., n \\ w_i^* A &= \lambda_i w_i^* \end{aligned}$$

The vectors w_i^* are sometimes called *left eigenvectors* of A. Taking the transpose of

$$(P^*)^{-1} = [w_1, ..., w_n]$$

$$P^{-1} = \begin{bmatrix} w_1^* \\ \vdots \\ w_n^* \end{bmatrix}$$

Write out $P^{-1}P = I$ to get

$$\begin{bmatrix} w_1^* \\ \vdots \\ w_n^* \end{bmatrix} [u_1, ..., u_n] = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

$$w_i^* u_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Normalize the eigenvectors $\{w_i\}$ by

$$v_i = \frac{w_i}{\|w_i\|_2}, \quad i = 1, ..., n$$

giving eigenvectors of A^* of length 1.

Define

$$s_i = v_i^* u_i = \frac{1}{\|w_i\|_2}, \quad i = 1, ..., n$$

We can write

$$(P^*)^{-1} = \left[\frac{v_1}{s_1}, \cdots, \frac{v_n}{s_n}\right]$$

and also

$$A^* v_i = \overline{\lambda}_i v_i, \quad \|v_i\|_2 = 1, \quad i = 1, ..., n$$

With these tools, we can now do a stability analysis for isolated eigenvalues of A. In particular, assume the eigenvalue λ_k is a simple eigenvalue of A. Consider what happens to it with a perturbation of the matrix A, namely

$$A(\epsilon) = A + \epsilon B, \quad \epsilon > 0$$

Let $\lambda_1(\epsilon), ..., \lambda_n(\epsilon)$ denote the perturbed eigenvalues for $A(\epsilon)$. We want to estimate $\lambda_k(\epsilon) - \lambda_k$.

Using the matrix P,

$$P^{-1}A(\epsilon)P = P^{-1}(A + \epsilon B)\, P = D + \epsilon C$$

with

$$C = P^{-1}BP = \begin{bmatrix} \dfrac{v_1^*}{s_1} \\ \vdots \\ \dfrac{v_n^*}{s_n} \end{bmatrix} B\,[u_1, ..., u_n]$$

$$c_{i,j} = \frac{1}{s_i} v_i^* B u_j, \quad 1 \le i, j \le n$$

We want to prove that

$$\lambda_k(\epsilon) = \lambda_k + \frac{\epsilon}{s_k} v_k^* B u_k + O\left(\epsilon^2\right)$$

The argument for this is given on page 598 of the text, which I omit here.

Using the vector and matrix 2-norms,

$$|\lambda_k(\epsilon) - \lambda_k| \le \frac{\epsilon}{s_k} \|B\|_2 + O\left(\epsilon^2\right)$$

since $\|u\|_2 = \|v\|_2 = 1$. Thus the size of s_k is of crucial importance in determining the stability of λ_k.

EXAMPLE

Consider

$$A = \begin{bmatrix} 101 & -90 \\ 110 & -98 \end{bmatrix}, \quad P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

$$P = \begin{bmatrix} \frac{9}{\text{sqrt}(181)} & \frac{-10}{\text{sqrt}(221)} \\ \frac{10}{\text{sqrt}(181)} & \frac{-11}{\text{sqrt}(221)} \end{bmatrix}$$

$$P^{-1} = \begin{bmatrix} -11\,\text{sqrt}\,(181) & 10\,\text{sqrt}\,(181) \\ -10\,\text{sqrt}\,(221) & 9\,\text{sqrt}\,(221) \end{bmatrix}$$

This defines the vectors $\{u_1, u_2\}$ and $\{w_1, w_2\}$, and thus

$$s_1 = v_1^* u_1 = \frac{1}{\|w_1\|_2} = \frac{1}{\text{sqrt}\,(221 \cdot 181)} \doteq .005$$

$$\begin{aligned} |\lambda_1(\epsilon) - \lambda_1| &\leq \frac{\epsilon}{s_1}\|B\|_2 + O\left(\epsilon^2\right) \\ &\doteq 200\epsilon\,\|B\|_2 + O\left(\epsilon^2\right) \end{aligned}$$

ORTHOGONAL TRANSFORMATIONS

Suppose we transform A using an orthogonal similarity transformation,

$$\hat{A} = U^* A U, \qquad U^* U = I$$

What are the vectors $\widehat{u}_i, \widehat{v}_i, \widehat{w}_i$ for this new matrix? Tranform

$$P^{-1} A P = D = \text{diag}[\lambda_1, ..., \lambda_n]$$

to

$$P^{-1} U \left(U^* A U \right) U^* P = D$$

$$\left(U^* P \right)^{-1} \widehat{A} \left(U^* P \right) = D$$

This says that the columns $\{\widehat{u}_i\}$ are obtained from

$$U^* P = U^* \left[u_1, ..., u_n \right] = \left[U^* u_1, ..., U^* u_n \right]$$

$$\widehat{u}_i = U^* u_i, \quad i = 1, ..., n$$

Similarly,

$$\widehat{v}_i = U^* v_i, \quad i = 1, ..., n$$

Now consider the numbers s_i which measure the sensitivity of the eigenvalues (provided they are simple). For the new matrix, call these numbers $\widehat{s}_i$. Then

$$\begin{aligned} \widehat{s}_i &= \widehat{v}_i^* \widehat{u}_i \\ &= (U^* v_i)^* (U^* u_i) \\ &= v_i^* U U^* u_i \\ &= s_i \end{aligned}$$

Thus an orthogonal similarity transformation of A does not change these numbers $\{s_i\}$, and thus the conditioning of the eigenvalue problem is not changed. This is a major reason for using orthogonal transformations.