• Recall that if $X$ is a non-singular matrix of eigenvectors of a matrix $A$ and $D$ is the corresponding diagonal matrix of eigenvalues, then

$$X^{-1}AX = D.$$  

• Recall also that the degree of singularity of a matrix $A$ is given by its condition number, $\|A\|\|A^{-1}\|$.

• Today we will see that the number $\|X\|_p\|X^{-1}\|_p$ plays a similar role for the eigenvalue problem as the condition number plays for a linear system with coefficient matrix $A$.

• Specifically we have the Bauer-Fike Theorem, which is Theorem 7.2.2 in Golub/van Loan.

**Theorem.** If $\mu$ is an eigenvalue of $A+E$ and

$$X^{-1}AX = D$$

then

$$\min_{\lambda \text{ eigenvalue of } A} |\lambda - \mu| \leq \|X\|_p\|X^{-1}\|_p\|E\|_p.$$

• This is very similar to what we saw for linear systems. The error in the solution of the perturbed system may be as large as the (norm...
of the) perturbation multiplied with the condition number.

- Also: If the matrix is defective then the matrix of eigenvectors, listed with proper multiplicities, is singular.

- To see why the Bauer-Fike theorem is true we need two facts:

**Fact 1.** If $D$ is diagonal then

$$
\|D\|_p = \max_{\|x\|_p=1} \| Dx \|_p = \max_{i=1,\ldots,n} |d_{ii}|.
$$

- To show this is an easy **exercise**.

**Fact 2.** if $\|F\|_p < 1$ then $I - F$ is non-singular.

- To see this suppose $\|F\|_p < 1$ and $I - F$ is singular. Then there exists a non-zero vector $x$ such that

$$(I - F)x = 0$$

or

$$x =Fx.$$ 

Taking norms on both sides and using the fact that for induced matrix norms the norm of the product does not exceed the product of the norms we get

$$
\|x\|_p \leq \|F\|_p \|x\|_p < \|x\|_p
$$

which is a contradiction.
Back to
\[ \min_{\lambda \text{ eigenvalue of } A} |\lambda - \mu| \leq \|X\|_p \|X^{-1}\|_p \|E\|_p. \]
Clearly this is true if \( \mu \) is in fact an eigenvalue of \( A \) since we can pick \( \lambda = \mu \).

So suppose \( \mu \) is not an eigenvalue of \( A \). Note first that \( A + E - \mu I \) is singular since \( \mu \) is an eigenvalue of \( A + E \).

So is
\[ X(A + E - \mu I)X^{-1} = D + X(E - \mu I)X^{-1} = D - \mu I + XEX^{-1} \]

By multiplying with \( (D - \mu I)^{-1} \) (noting that \( D - \mu I \) is non-singular) we get that
\[ (D - \mu I)^{-1}(D - \mu I) + (D - \mu I)^{-1}XEX^{-1} = I + (D - \mu I)^{-1}XEX^{-1} \]
is singular.

By fact 2 we know that
\[ \|(D - \mu I)^{-1}XEX^{-1}\|_p \geq 1 \]

This implies that
\[ \|(D - \mu I)^{-1}\|_p \|XEX^{-1}\|_p \geq 1 \]

By Fact 1 this gives
\[ \frac{1}{\min |\lambda - \mu|} \geq \frac{1}{\|XEX^{-1}\|_p} \]
which gives our desired result:

$$\min |\lambda - \mu| \leq \|XE X^{-1}\|_p \leq \|X\|_p \|X^{-1}\|_p \|E\|_p.$$ 

- Exercise: verify that this inequality is sharp.
- It is also useful to look at the conditioning of a single eigenvalue.
- We’ll see that what matters is the angle between left and corresponding right eigenvector.
- We continue to focus on the real case.
- The following analysis introduces a new idea: dependence on a parameter and implicit differentiation.
- Suppose \(\lambda\) is a simple eigenvalue (of multiplicity 1) and let

\[
\|F\|_2 = \|x(\epsilon)\|_2 = 1, \quad x(0) = x, \quad \lambda(0) = \lambda, \quad Ax = \lambda x
\]

and

\[
(A + \epsilon F)x(\epsilon) = \lambda(\epsilon)x(\epsilon)
\]

(1)

- Thus we think of \(x(\epsilon)\) and \(\lambda(\epsilon)\) as functions of \(\epsilon\) defined implicitly by the equation (1).
- Also suppose that \(y\) is the normalized corresponding left eigenvector of \(A\), i.e.,

\[
y^T A = \lambda y^T \quad \text{and} \quad \|y\|_p = 1.
\]
\[(A + \epsilon F)x(\epsilon) = 2(\epsilon)x(\epsilon)\]

- Denote differentiation with respect to epsilon with a dot.
- Differentiating in (1) gives

\[Fx(\epsilon) + (A + \epsilon F)\dot{x}(\epsilon) = \dot{\lambda}(\epsilon)x(\epsilon) + \lambda(\epsilon)\dot{x}(\epsilon)\]

- Setting \(\epsilon = 0\) turns this into

\[Fx + Ax = \dot{\lambda}x + \lambda\dot{x}\]

- Multiply with \(y^T\) to get

\[y^TFx + y^TA\dot{x} = \dot{\lambda}y^Tx + \lambda y^T\dot{x}\]

- Noting that \(y^TA\dot{x} = \lambda y^T\dot{x}\) this simplifies to

\[\dot{\lambda}y^Tx = y^TFx\]

which becomes

\[\dot{\lambda} = \frac{y^TFx}{y^Tx} = \frac{\gamma^TFx}{\cos \alpha} \tag{2}\]

- Recall that

\[\|x\|_2 = \|y\|_2 = \|F\|_2 = 1\]

Hence (2) turns into

\[|\dot{\lambda}| \leq \frac{\|y\|_p \|F\|_p \|x\|_p}{\|x\|_p \|y\|_p \cos \alpha} = \frac{1}{\cos \alpha} \tag{3}\]

\[p = 2\]
where $\alpha$ is the angle formed by $x$ and $y$.

- The derivative of $\lambda$ measures the sensitivity of $\lambda$ with respect to $\epsilon$. According to (3) that sensitivity is the larger the larger the angle between the eigenvector $x$ and the corresponding left eigenvector $y$. The sensitivity becomes infinite when $x$ and $y$ are orthogonal.

- Caveat: the angle in question is unique only if the left eigenvector corresponding to $\lambda$ is unique. If it isn’t we pick the eigenvector giving the smallest angle. For example, for the identity matrix every non-zero vector is a left and also a right eigenvector, and we pick pairs of left and right eigenvectors that are identical.
• Example 1:

\[
A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}
\]

\[
A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + y \\ y \end{bmatrix} \quad \Rightarrow \quad \gamma = 0
\]

\[
A + \varepsilon F = \begin{bmatrix} 1 & 1 \\ \varepsilon & 1 \end{bmatrix}
\]

\[
|A + \varepsilon F - \mu I| = \left| \begin{bmatrix} 1 - \mu & 1 \\ \varepsilon & 1 - \mu \end{bmatrix} \right| = (1 - \mu)^2 - \varepsilon
\]

\[
x = \begin{bmatrix} 1 \\ z \end{bmatrix} \quad 1 - \mu = \pm \sqrt{\varepsilon}
\]

\[
\begin{bmatrix} 1 \\ \varepsilon \\ 1 \end{bmatrix} \begin{bmatrix} 1 + z \\ \varepsilon + z \end{bmatrix} = (1 \pm \sqrt{\varepsilon}) \begin{bmatrix} 1 \\ z \end{bmatrix}
\]

\[
y = \begin{bmatrix} \pm \sqrt{\varepsilon} \\ 1 \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ \pm \sqrt{\varepsilon} \end{bmatrix}
\]

\[
\frac{x^T y}{\|x\| \|y\|} = \frac{\pm 2 \sqrt{\varepsilon}}{1 + \varepsilon} = \cos \alpha
\]
• Example 2:

\[ A(\epsilon) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \epsilon \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & \epsilon \end{bmatrix}. \]

• \( A(\epsilon) \) is defective if \( \epsilon = 0 \) and non-defective otherwise.

Exercise
Summary: Correspondence between Singularity and Defectiveness

Property: \[ Ax = b \quad Ax = \lambda x \]

perfect: orthogonal symmetric

infinitely bad: singular defective

condition number: \[ \|A\| \|A^{-1}\| \quad \|X\| \|X^{-1}\| \]

single evector: \[ x^T y \]