Lecture 22: Diagonalizable matrices

Tuesday, November 10, 2015 9:30 AM

Reading: Meyer 7.5 Normal matrices
7.6 Positive semi-definite matrices

WHEN IS A MATRIX DIAGONALIZABLE?

When does it have a complete set of eigenvectors?

Exercise: Which of these matrices can be diagonalized?

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Answer:

A and C.

A: Since A is triangular, you can read its eigenvalues off the diagonal: A's eigenvalues are 1 and 2.
Two different e-values $\Rightarrow$ two independent e-vectors, and in $\mathbb{R}^2$ that's all there's room for. ✓

B: B has eigenvalue 1 with multiplicity 2, but the associated eigenspace $N(B-I) = N\left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right)$ is only one-dimensional. B does not have a complete set of e-vectors.

C: $C = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \end{pmatrix}$. It is proportional to the projection $\vec{v}\vec{v}^T$ for $\vec{x} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. The orthogonal direction is $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Eigenvector	Eigenvalue
$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	2
$\begin{pmatrix} -1 \\ 1 \end{pmatrix}$	0

A matrix is diagonalizable when

- for each eigenvalue, the dimension of the associated eigenspace equals the multiplicity of the eigenvalue.
- In other words, if

$$\det(A-\lambda I) = (\lambda-\lambda_1)^{\alpha_1}\cdot(\lambda-\lambda_2)^{\alpha_2}\cdot\dots\cdot(\lambda-\lambda_k)^{\alpha_k}$$

distinct e-values $\lambda_1,\dots,\lambda_k$, with multiplicities $\alpha_1,\dots,\alpha_k$

$\alpha_1+\alpha_2+\dots+\alpha_k = n$

$$\dim N(A-\lambda_j I) = \alpha_j \text{ for all } j$$

$$\Rightarrow A = \underbrace{\begin{pmatrix} \boxed{\begin{matrix}\text{basis for}\\ N(A-\lambda_1 I)\end{matrix}} & \boxed{\begin{matrix}\text{basis for}\\ N(A-\lambda_2 I)\end{matrix}} & \cdots \end{pmatrix}}_{U} \begin{pmatrix} \underbrace{\lambda_1 \ddots \lambda_1}_{\substack{\alpha_1 \\ \shortparallel \\ \dim N(A-\lambda_1 I) \\ \text{times}}} & & & \\ & \overbrace{\lambda_2 \ddots \lambda_2}^{\alpha_2 \text{ times}} & & \\ & & \lambda_3 & \\ & & & \ddots \end{pmatrix} U^{-1}$$

<u>Corollary</u>: If an $n\times n$ matrix A has n distinct eigenvalues, then A must be diagonalizable.

But not every diagonalizable matrix has n distinct eigenvalues.

e.g., $I = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & \ddots & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & & & & & 0 \\ & 1 & & & & \\ & & 1 & & & \\ & & & 2 & & \\ & & & & 2 & \\ 0 & & & & & 2 \end{pmatrix}$.

e-value 1, multiplicity n $\qquad$ $\lambda_1 = 1$, multiplicity $\alpha_1 = 3$; $\lambda_2 = 2$, $\alpha_2 = 3$

Even if a matrix can be diagonalized, its eigenvectors might not be orthogonal.

E.g., $\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$

<u>Evalue</u>	<u>E-vector</u>
2	$(0,1)$
1	$(1,-1)$

not orthogonal!

$$N\left(\begin{pmatrix}1&0\\1&2\end{pmatrix} - \begin{pmatrix}1&0\\0&1\end{pmatrix}\right) = N\left(\begin{pmatrix}0&0\\1&1\end{pmatrix}\right) = \text{Span}\left(\begin{pmatrix}1\\-1\end{pmatrix}\right)$$

Why doesn't Gram-Schmidt help?

- Performed on eigenvectors with different eigenvalues, it will output orthogonal vectors spanning the same space, but they won't (in general) still be eigenvectors.

E.g., $\left\{\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right\} \xrightarrow[\text{Gram-}]{} \left\{\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}\right\}$

E.g., $\left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \xrightarrow[\text{Gram-Schmidt}]{} \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$

not an eigenvector of $\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$!

Example:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \neq AA^T = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

but A is diagonalizable!

$\lambda_1 = 1$

$\lambda_2 = 0$

(e.g., since $\mathrm{Trace}(A) = 1 = \lambda_1 + \lambda_2$
and $\mathrm{Det}(A) = 0 = \lambda_1 \cdot \lambda_2$)

All distinct eigenvalues $\Rightarrow$ diagonalizable.

Today: Lots of matrices are diagonalizable, with orthonormal eigenvectors.

For example, all symmetric matrices ($A = A^T$).

(And real symmetric matrices even have real eigenvalues)

Recall: Adjoint = conjugate transpose

$$\begin{pmatrix} a+bi & c+di \\ e+fi & g+hi \end{pmatrix}^\dagger = \begin{pmatrix} a-bi & e-fi \\ c-di & g-hi \end{pmatrix}$$

(same as transpose for real matrices).

THEOREM: A has a complete, orthogonal set of eigenvectors

$$\Updownarrow$$

$$A^\dagger A = AA^\dagger.$$

(**Definition:** A is "normal" $\Leftrightarrow A^\dagger A = AA^\dagger$.)

Proof:

$\Downarrow$: One direction is trivial. Assume A has a complete, orthonormal set of eigenvectors. Letting

$$U = \left(\text{those e-vectors} \right),$$

$$A = U D U^{-1},$$

where D is a diagonal matrix of the eigenvalues.

Since U is unitary, $U^{-1} = U^{+}$. Hence

$$\begin{aligned} A A^{+} &= (U D U^{+})(U D U^{+})^{+} \\ &= U D U^{+} U D^{+} U^{+} \\ &= U D D^{+} U^{+} \\ A^{+} A &= (U D U^{+})^{+} (U D U^{+}) \\ &= U D^{+} U^{+} U D U^{+} \\ &= U D^{+} D U^{+} \end{aligned}$$

These are equal since $D D^{+} = D^{+} D$ both just have the squared magnitudes $|\lambda_i|^2$ along the diagonal. ✓

(Every diagonal matrix is normal.)

The other direction (A normal $\Rightarrow A = U D U^{-1}$) is much more interesting. First let me prove two claims:

Claim 1: A normal $\Rightarrow$ $R(A) = R(A^T)$.

(**Note**: If $R(A) \neq R(A^T)$, then there is no hope of finding a basis of eignvectors.

A maps rowspace, $R(A^T)$, to columnspace, $R(A)$

If these spaces are different, then A does more than just scale some vectors.)

Proof:

We have seen already that

$$N(A) = N(A^{+} A)$$

and applying this to A^{+} gives

$$N(A^{+}) = N(A A^{+})$$

(since $(A^{+})^{+} = A$) Hence

$$R(A^T) \underset{\text{rank-nullity}}{=} N(A)^\perp = N(A^\dagger A)^\perp \underset{\text{normal}}{=} N(AA^\dagger)^\perp = N(A^\dagger)^\perp \underset{\text{rank-nullity}}{=} R(A)$$

✓ $\square$

Claim 2: A normal $\Rightarrow$ there exists a unitary matrix $\mathcal{U}$ such that

$$\mathcal{U}^\dagger A \mathcal{U} = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}$$

for some nonsingular matrix C (a $\text{rank}(A) \times \text{rank}(A)$ matrix)

Proof: By Claim 1, $R(A) = R(A^\dagger)$, and $N(A) = N(A^\dagger) = R(A)^\perp$.

Let

$$\mathcal{U} = \begin{pmatrix} \text{basis for } R(A) & \text{basis for } N(A) \end{pmatrix}.$$

Since its columns are orthonormal, $\mathcal{U}$ is unitary: $\mathcal{U}^\dagger = \mathcal{U}^{-1}$.

$$\mathcal{U}^\dagger A \mathcal{U} = \begin{pmatrix} \text{basis for } R(A) \\ \text{basis for } N(A^\dagger) \end{pmatrix} A \begin{pmatrix} \text{basis for } R(A^\dagger) & \text{basis for } N(A) \end{pmatrix}$$

$$= \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}$$

since $A \cdot \begin{pmatrix} \text{vector} \\ \text{in } R(A^\dagger) \end{pmatrix} = \begin{pmatrix} \text{vector} \\ \text{in } R(A) \end{pmatrix} \perp N(A^\dagger)$

and $A \cdot \begin{pmatrix} \text{vector} \\ \text{in } N(A) \end{pmatrix} = 0$

$$C = \begin{pmatrix} \text{basis for } R(A) \end{pmatrix} A \begin{pmatrix} \text{basis for } R(A^\dagger) \end{pmatrix}$$

$\square$

Now we're ready to prove the interesting direction:

Theorem: $A^\dagger A = AA^\dagger \Rightarrow A = \underset{\text{unitary}}{\mathcal{U}} \underset{\text{diagonal}}{D} \mathcal{U}^\dagger$ using **Claim 2:** $A^\dagger A = AA^\dagger \Rightarrow \mathcal{U}^\dagger A \mathcal{U} = \begin{pmatrix} \underset{\text{nonsingular}}{C} & 0 \\ 0 & 0 \end{pmatrix}$

Proof:

Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be A's distinct eigenvalues.

A normal $\Rightarrow$ $A - \lambda_1 I$ normal

$$\left((A - \lambda_1 I)(A - \lambda_1 I)^\dagger = (A - \lambda_1 I)^\dagger (A - \lambda_1 I) \right)$$

$\underset{\text{Claim 2}}{\Longrightarrow}$ $\exists\ U_1$ s.t.

$$U_1^\dagger (A - \lambda_1 I) U_1 = \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Rightarrow U_1^\dagger A U_1 = \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix} + \lambda_1 \underbrace{U_1^\dagger U_1}_{I}$$

$$= \begin{pmatrix} C_1 + \lambda_1 I & 0 \\ 0 & \lambda_1 I \end{pmatrix}$$

Let $A_1 = C_1 + \lambda_1 I$.

<u>Observe</u>:

- λ_1 is not an eigenvalue of A_1 (or $A_1 - \lambda_1 I$ would be singular)
- $\lambda_2, \ldots, \lambda_k$ are still eigenvalues of A_1 (since conjugating A by U_1 does not change the set of e-values, and $\lambda_2, \ldots, \lambda_k$ are definitely not e-values of the second block $\lambda_1 I$)
- A_1 is normal! (b/c $U_1^\dagger A U_1 = \begin{pmatrix} A_1 & 0 \\ 0 & \lambda_1 I \end{pmatrix}$ is normal — conjugating by a unitary does not change normality. Since the matrix is block-diagonal, each block must be normal separately.)

Therefore we can just recurse: find a unitary U_2 so

$$U_2^\dagger (A_1 - \lambda_2 I) U_2 = \begin{pmatrix} \underbrace{C_2}_{\text{nonsingular}} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Rightarrow U_2^\dagger A_1 U_2 = \begin{pmatrix} \underbrace{C_2 + \lambda_2 I}_{A_3!\ \text{etc.}} & 0 \\ 0 & \lambda_2 I \end{pmatrix}$$

Putting everything together, we find that for

$$U^\dagger = \cdots \begin{pmatrix} \begin{matrix} U_3^\dagger & 0 \\ 0 & I \end{matrix} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} U_2^\dagger & 0 \\ 0 & \underbrace{I}_{\text{to leave the } \lambda_1 I \text{ term}} \end{pmatrix} U_1^\dagger ,$$

$$U^{+} A U = \begin{pmatrix} \overset{\text{unchanged}}{\lambda_k I} & & & & 0 \\ & \lambda_{k-1} I & & & \\ & & \ddots & & \\ & & & \lambda_2 I & \\ 0 & & & & \lambda_1 I \end{pmatrix} \quad \checkmark$$

$\square$

Important: The theorem ($A^{+}A = AA^{+} \Rightarrow$ unitarily diagonalizable) is very important. So is the proof technique: Find one eigenspace, split it off, and recurse with the remainder.

Example: How can we use the power method to find the **second**-largest magnitude eigenvalue and the corresponding eigenvector?

One approach, in Matlab:

```
% using the power method to find the second-largest-magnitude eigenvalue
eigenvector
n = 100;
A = randn(n,n);
A = A + A';           % symmetric matrix => normal matrix

% first find the principal eigenvector using the power method
x = randn(n,1);
for j = 1:10000
  x = A * x;
  x = x / norm(x);
  x';
end
A*x ./ x    % using component-wise division, check that we've found an e-vector

% this starts with a vector perpendicular to the principal eigenvector,
%  but numerical errors explode, causing it to be pushed parallel to the
%  principal eigenvector
y = randn(n,1);
y = y - (x'*y)*x;
for j = 1:10000
  y = A * y;
  y = y / norm(y);
  y';
end
A * y ./ y

% to get the power method to work, we need to project orthogonal to the
%  principal eigenvector after every step (or at least occasionally)
y = randn(n,1);
y = y - (x'*y)*x;
for j = 1:10000
  y = A * y;
  y = y - (x'*y)*x;
  y = y / norm(y);
  y';
end
A * y ./ y

% we can also find the k largest-magnitude eigenvalues simultaneously, using
the Gram-Schmidt procedure at every step of the power method
```