Lecture 22: Diagonalizable matrices

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?

\[ A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 1 \\ \end{pmatrix} \]

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 \( N(B-I) = N(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}) \) is only one-dimensional. B does not have a complete set of e-vectors.

C: C = \begin{pmatrix} 1 \\ 1 \\ \end{pmatrix}. It is proportional to the projection \( v'v^T \) for \( v = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \). The orthogonal direction is \( (-1) \).

\[ \begin{array}{c|c}
\text{Eigenvector} & \text{Eigenvalue} \\
\hline
\begin{pmatrix} 1 \\ 1 \end{pmatrix} & 2 \\
\begin{pmatrix} 1 \\ -1 \end{pmatrix} & 0 \\
\end{array} \]

A matrix is diagonalizable when

\[ \text{for each eigenvalue, the dimension of the associated eigenspace equals the multiplicity of the eigenvalue.} \]

\[ \text{In other words, if} \]

---

EE 441 Page 1
\[
\det(A - \lambda I) = (\lambda - \lambda_1)^{d_1} \cdot (\lambda - \lambda_2)^{d_2} \cdot \cdots \cdot (\lambda - \lambda_k)^{d_k}
\]

- distinct \( \lambda_i \) with multiplicities \( d_i \)

- \( \dim N(A - \lambda_j I) = d_j \) for all \( j \)

\[
\Psi A = \begin{pmatrix}
\text{basis for } N(A - \lambda_1 I) \\
\text{basis for } N(A - \lambda_2 I) \\
\vdots \\
\text{basis for } N(A - \lambda_k I)
\end{pmatrix} \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
0 & 0 & \cdots & \lambda_k
\end{pmatrix} \begin{pmatrix}
\text{dim } N(A - \lambda_1 I) \\
\text{dim } N(A - \lambda_2 I) \\
\vdots \\
\text{dim } N(A - \lambda_k I)
\end{pmatrix}
\]

**Corollary:** 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.**

\[
e.g., \quad I = \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
1 & 0 \\
0 & 2
\end{pmatrix}
\]

- \( \lambda_1 = 1 \), multiplicity \( n \)
- \( \lambda_1 = 1 \), multiplicity \( d_1 = 3 \)
- \( \lambda_2 = 2 \), \( d_2 = 3 \)

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

\[
e.g., \quad \begin{pmatrix}
1 & 0 \\
1 & 2
\end{pmatrix} \quad \text{Eigenvalue} \quad 2 \quad \text{E-vector} \quad \begin{pmatrix}
0, 1
\end{pmatrix} \quad \text{not orthogonal!} \quad \begin{pmatrix}
0, 1
\end{pmatrix} \quad \begin{pmatrix}
1, -1
\end{pmatrix}
\]

\[
N(\begin{pmatrix}
1 & 0 \\
1 & 2
\end{pmatrix} - \begin{pmatrix}
0 & 0 \\
0 & 0
\end{pmatrix}) = N(\begin{pmatrix}
0 & 0 \\
0 & 0
\end{pmatrix}) = \text{Span}(\begin{pmatrix}
0, 1
\end{pmatrix})
\]

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.

\[
e.g., \quad \{ (0), (1) \} \xrightarrow{\text{Gram-Schmidt}} \{ (0), (1) \}
\]
E.g., \( \{(0), (-1)\} \xrightarrow{\text{Gram-Schmidt}} \{(0), (1)\} \) not an eigenvector of \((1, 2)^T)!

Example:
\[
A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}
\]
\[
A^TA = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}
\]
but \(A\) is diagonalizable!
\[
AA^T = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}
\]
\[\lambda_1 = 1, \quad \lambda_2 = 0\]
\[\text{eg., since } \text{Trace}(A) = 1 = \lambda_1 + \lambda_2 \]
\[\text{and } \text{Det}(A) = 0 = \lambda_1 \lambda_2\]
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
\[
A^\dagger A = AA^\dagger.
\]
(Definition: \(A\) is “normal” \(\iff\) \(A^\dagger A = AA^\dagger\).)

Proof:
\[\uparrow\downarrow\]
One direction is trivial. Assume \(A\) has a complete, orthonormal set of eigenvectors. Letting
\[ U = \left( \begin{array}{c|c|c|c} \text{e-vectors} \end{array} \right), \]

\[ A = U D U^*, \]

where \( D \) is a diagonal matrix of the eigenvalues.

Since \( U \) is unitary, \( U^* = U^\dagger \). Hence

\[
AA^\dagger = (U D U^\dagger)(U D U^\dagger)^\dagger
\]
\[
= U D U^\dagger U D^\dagger U^\dagger
\]
\[
= U D D^\dagger U^\dagger
\]
\[
A^\dagger A = (U D U^\dagger)^\dagger(U D U^\dagger)
\]
\[
= U D^\dagger U^\dagger U D U^\dagger
\]
\[
= U D D U^\dagger
\]

These are equal since \( D D^\dagger = D^\dagger D \) both just have the squared magnitudes \( |\lambda_i|^2 \) along the diagonal.

(\text{Every diagonal matrix is normal.}) \checkmark

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

\textbf{Claim 1:} \( A \text{ normal} \Rightarrow R(A) = R(A^\dagger). \)

\textbf{Note:} If \( R(A) \neq R(A^\dagger) \), then there is no hope of finding a basis of eigenvectors.

\( A \) maps row space, \( R(A^\dagger) \), to column space, \( R(A) \).

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

\textbf{Proof:}

We have seen already that

\[ N(A) = N(A^\dagger A) \]

and applying this to \( A^\dagger \) gives

\[ N(A^\dagger) = N(A A^\dagger) \]

(since \( (A^\dagger)^\dagger = A \)) Hence
\[ R(A^T) = N(A)^+ = N(AA^T)^+ = N(AA^*)^+ = N(A^*)^+ = R(A) \]

Claim 2: A normal \( \Rightarrow \) there exists a unitary matrix \( U \) such that
\[ U^*AU = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \]
for some nonsingular matrix \( C \) (a \( \text{rank}(A) \times \text{rank}(A) \) matrix).

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

Let
\[ U = \begin{pmatrix} \text{basis for } R(A) \\ \text{basis for } N(A) \end{pmatrix} \]

Since its columns are orthonormal, \( U \) is unitary: \( U^* = U^{-1} \).

\[ U^*AU = \begin{pmatrix} \text{basis for } R(A) \\ \text{basis for } N(A) \end{pmatrix}^* A \begin{pmatrix} \text{basis for } R(A) \\ \text{basis for } N(A) \end{pmatrix} = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \]

since \( A \cdot (\text{vector in } R(A)) = (\text{vector in } R(A))^\perp \cap N(A) \)
and \( A \cdot (\text{vector in } N(A)) = 0 \)

\[ C = \begin{pmatrix} \text{basis for } R(A) \\ \text{basis for } R(A) \end{pmatrix} A \begin{pmatrix} \text{basis for } N(A) \\ \text{basis for } N(A) \end{pmatrix} \]

Now we're ready to prove the interesting direction:

**Theorem:** \( A^*A = AA^* \)
\( \Rightarrow A = UDU^* \) using \( U^*AU = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \)

**Claim 2:** \( A^*A = AA^* \)

Proof:

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

A normal \( \Rightarrow A - \lambda I \) normal
\[ (A - \lambda I)(A - \lambda I)^* = (A - \lambda I)^*(A - \lambda I) \]
\[ \Rightarrow \exists U_1 \text{ s.t.} \]
\[ U_1^T (A - \lambda_1 I) U_1 = \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix} \]
\[ \Rightarrow U_1^T A U_1 = \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix} + \lambda_1 U_1^T U_1 \]
\[ = \begin{pmatrix} C_1 + \lambda_1 I, 0 \\ 0, \lambda_1 I \end{pmatrix} \]

Let \( A_1 = C_1 + \lambda_1 I \).

**Observe:**
- \( \lambda_1 \) is not an eigenvalue of \( A_1 \) (or \( A_1 - \lambda_1 I \) would be singular).
- \( \lambda_2, ..., \lambda_k \) are still eigenvalues of \( A_1 \) since conjugating \( A \) by \( U_1 \) does not change the set of e-values, and \( \lambda_2, ..., \lambda_k \) are definitely not e-values of the second block \( \lambda_1 I \).

- \( A_1 \) is normal because \( U_1^T A U_1 = \begin{pmatrix} C_1 & 0 \\ 0 & \lambda_1 I \end{pmatrix} \) is normal.

Therefore, we can just recurse: find a unitary \( U_2 \) so
\[ U_2^T (A_1 - \lambda_2 I) U_2 = \begin{pmatrix} C_2 & 0 \\ 0 & 0 \end{pmatrix} \]
\[ \Rightarrow U_2^T A_1 U_2 = \begin{pmatrix} C_2 + \lambda_2 I, 0 \\ 0, \lambda_2 I \end{pmatrix} \]
\[ A_3 \text{! etc.} \]

Putting everything together, we find that for
\[ U^T = \ldots \left( \begin{pmatrix} U_3^T & 0 \\ 0 & I \end{pmatrix} \right) \left( \begin{pmatrix} U_2^T & 0 \\ 0 & I \end{pmatrix} \right) U_1^T, \]
\[ \text{to leave the } \lambda_1 I \text{ term.} \]
Important: The theorem \((A^*A = AA^* \Rightarrow \text{unitarily diagonalizable})\) 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