Lecture 12: Diagonalizing Matrices

Recall that when \( \vec{v} \in \mathbb{R}^n \) is nonzero, \( \lambda \in \mathbb{R} \), and 
\[
A \vec{v} = \lambda \vec{v}, \\
(A = n \times n matrix)
\]
\( \lambda \) (resp. \( \vec{v} \)) is an eigenvalue (resp. eigenvector) of \( A \).

- To find eigenvalues: solve \( \det(A - \lambda \mathbb{I}_n) = 0 \)
- To find eigenvectors: for each eigenvalue \( \lambda_0 \), find (a basis for) 
  \[
  E_{\lambda_0} = \operatorname{Nul} (A - \lambda_0 \mathbb{I}_n)
  \]
  by row-reduction. Its dimension is \( n - \operatorname{rk}(A - \lambda_0 \mathbb{I}_n) \), by \( \mathbb{R} + \mathbb{N} \).

- To check if \( \vec{v}_0 \) is an eigenvector: apply \( A \) to \( \vec{v} \)
- To check if \( \lambda_0 \) is an eigenvalue: see if \( \operatorname{rk}(A - \lambda_0 \mathbb{I}_n) < n \) by using row-reduction.

- The eigenvalues of an upper or lower-triangular matrix are the diagonal entries.

\[\]

Now by the Fundamental Theorem of Algebra, the characteristic polynomial factors

\[
\det(A - \lambda \mathbb{I}_n) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)
\]

Where in general the \( \{\lambda_i\} \) may be non-real (i.e. complex numbers) and may not be distinct. Assume for now that they are real.
**Definition:** The *multiplicity* of an eigenvalue of $A$ is the number of times it appears in (**$\ast$**). (If all multiplicities are 1, then $A$ has $n$ distinct eigenvalues.)

**Lemma:** If $\vec{v}_1, ..., \vec{v}_k$ are $k$ eigenvectors of $A$ with distinct eigenvalues $\lambda_1, ..., \lambda_k$, then they are linearly independent.

**Proof:** Use induction: this is clear for $k=1$, since $\vec{v}_1 \neq \vec{0}$ by definition. Assume it holds for $k-1$ eigenvectors with distinct eigenvalues, i.e., that $\vec{v}_1, ..., \vec{v}_{k-1}$ are independent, and let $\vec{v}_k$ be an eigenvector with a "new" eigenvalue.

Suppose $\vec{0} = c_1 \vec{v}_1 + ... + c_k \vec{v}_k$. (We must show the $c_i$'s are all 0.) On one hand (multiplying by $\lambda_k$)

(1) \[ \vec{0} = c_1 \lambda_k \vec{v}_1 + ... + c_k \lambda_k \vec{v}_k. \]

On the other hand (applying $A$)

(2) \[ \vec{0} = c_1 A \vec{v}_1 + ... + c_k A \vec{v}_k = c_1 \lambda_1 \vec{v}_1 + ... + c_k \lambda_k \vec{v}_k. \]

Subtracting (2) - (2) gives

\[ \vec{0} = c_1 (\lambda_k - \lambda_1) \vec{v}_1 + ... + c_{k-1} (\lambda_k - \lambda_{k-1}) \vec{v}_{k-1} + c_k (\lambda_k - \lambda_k) \vec{v}_k = \vec{0} \]

\[ \implies c_1 = ... = c_{k-1} = 0 \quad (\text{since } \vec{v}_1, ..., \vec{v}_{k-1} \text{ are L.I.}). \]

But then the original equation reads $\vec{0} = c_k \vec{v}_k \implies c_k = 0$. \[\square\]

**Theorem:** If the eigenvalues of $A$ are distinct (and real), then a basis of $\mathbb{R}^n$ consisting of eigenvectors of $A$ exists, called the "$A$-eigenbasis".

**Proof:** For each of the $n$ eigenvalues, there's an eigenvector. Apply the Lemma. \[\square\]
Remark: The existence of an $A$-eigenvector is crucial for being able to write a given vector as a sum of eigenvectors of $A$, as part of solving systems of differential equations, etc.

Given an eigenvector $\vec{v}_i, \ldots, \vec{v}_n$, write $P = (\vec{v}_1 \ldots \vec{v}_n)$ and compute $AP = (A\vec{v}_1 \ldots A\vec{v}_n) = (\lambda_1 \vec{v}_1 \ldots \lambda_n \vec{v}_n)$; assemble the eigenvalues into a diagonal matrix $D = (\lambda_1 \ldots \lambda_n)$, so that $PD = (\lambda_1 \vec{v}_1 \ldots \lambda_n \vec{v}_n)$. Hence $AP = PD$ and

$$A = PDP^{-1}$$

(Equivalently, $P^{-1}AP = D$). We have diagonalized $A$.

**Corollary 1:** A matrix with $n$ distinct eigenvalues can be diagonalized (or "is diagonalizable").

**Ex 1 / Diagonalize** $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. From lecture 11,

$A$ has

- eigenvalues: $0, 0, 3$ (multiplicity 2) — recall characteristic polynomial was $\lambda^2(\lambda - 3)$.
- eigenvectors: $(1, 0, 0)$, $(1, 0, 0)$, $(1, 0, 0)$

So Lemma doesn’t apply — in this case, check independence.
Therefore \( A = \begin{pmatrix}
-1 & 1 & 0 \\
1 & 0 & 1 \\
0 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 3 \\
\end{pmatrix}
\begin{pmatrix}
\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\end{pmatrix} \begin{pmatrix}
P & D & P^{-1}
\end{pmatrix}
.

So you don’t need to have distinct eigenvalues in order to diagonalize.

Ex 2/ Is \( A = \begin{pmatrix}
5 & -8 & -21 \\
0 & 0 & 7 \\
0 & 0 & -2 \\
\end{pmatrix}
\) diagonalizable? If so, do it.

Eigenvalues: \( 5, 0, -2 \) (upper triangular) \( \Rightarrow \) distinct \( \Rightarrow \) diagonalizable.

Eigenvectors: row-reduce (if necessary) to find vector in null spaces of
\[
A - 5I = \begin{pmatrix}
0 & -8 & -21 \\
-5 & 7 \\
-7 \\
\end{pmatrix} \rightarrow \vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}
\]

\[
A - 0I = \begin{pmatrix}
5 & -8 & -21 \\
0 & 7 \\
-2 \\
\end{pmatrix} \rightarrow \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}
\]

\[
A + 2I = \begin{pmatrix}
7 & -8 & -21 \\
2 & 7 \\
0 \\
\end{pmatrix} \rightarrow \vec{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}
\]

Ex 3/ What about \( A = \begin{pmatrix}
5 & 0 & 0 \\
0 & 0 & 0 \\
0 & 2 & 0 \\
\end{pmatrix}
\)?

Eigenvalues: \( 2, \frac{3}{3} \) (upper triangular, w/multiplicity 2)

Find eigenvectors:
\[
A - 2I = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0 \\
\end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ span } E_2
\]

\[
A - 3I = \begin{pmatrix}
0 & 0 & 0 \\
0 & -1 & 0 \\
1 & 0 & 0 \\
\end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \text{ span } E_3
\]

Up to scale, there are only 2 eigenvectors. So there’s no \( A \)-eigensetasis of \( \mathbb{R}^3 \), and \( A \) isn’t diagonalizable.
1. The determinant of a diagonalizable matrix is the product of its eigenvalues.

\[ \det A = \lambda_1 \lambda_2 \cdots \lambda_n. \]

Why? \[ \det A = \det (PDP^{-1}) = \det P \cdot \det D \cdot \det P^{-1} \]
\[ = \det D = \det \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{pmatrix} = \Pi \lambda_i \quad (i=1,\ldots, n). \]

(In fact, even if \( A \) is not diagonalizable, \( \det A \) is the product of the roots [with multiplicity] of the characteristic polynomial \( p(\lambda) = \det(\lambda I - A) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n) = \lambda^n - \cdots + (-1)^n \lambda_1 \cdots \lambda_n. \) This is simply because setting \( \lambda = 0 \) gives \((-1)^n \lambda_1 \cdots \lambda_n = \det (0I - A) = \det (-A) = (-1)^n \det A. \)

2. Compute \( \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}^{10} \) by diagonalizing \( A \): 

\( A = PDP \) where \( D = \begin{pmatrix} -2 & 0 \\ 0 & 5 \end{pmatrix}, \quad P = \begin{pmatrix} -1 & 3 \\ 1 & 4 \end{pmatrix} \)

\( \Rightarrow A^{10} = (PDP^{-1})^{10} = PDP^{-1} \cdot PDP^{-1} \cdots PDP^{-1} \)
\[ = PD^{10}P^{-1} = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} (\begin{pmatrix} -2^{10} & 0 \\ 0 & 5^{10} \end{pmatrix} \begin{pmatrix} \frac{9}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}) \]
\[ = \frac{1}{7} \begin{pmatrix} -4.2^{10} & 3.2^{10} \\ -3.5^{10} & -3.5^{10} \end{pmatrix}. \]
Stochastic matrices

Let \( A \) be an \( n \times n \) matrix with all positive entries, whose columns sum to 1. (This is called a regular stochastic matrix.) These occur when iterating "conditional probabilities."

- \( A \) has a steady-state vector, i.e., eigenvector with eigenvalue 1.

Since columns of \( A \) sum to 1, columns of \( A - I_n \) sum to 0 and hence belong to an \((n-1)\)-diml subspace of \( \mathbb{R}^n \) and cannot all be independent \( \implies \) \( \text{rank} (A - I_n) \leq n - 1 \implies \text{nullity} (A - I_n) \geq 1 \). In fact, by a careful examination of \( A - I \) one can deduce that \( \text{dim} E_1 (= \text{nullity} (A - I_n)) \geq 1 \), so the steady-state vector is unique (provided we scale it to have its entries sum to 1). That is, the multiplicity of the eigenvalue 1, is 1.

- Any eigenvector with a different eigenvalue than 1
  - (a) must lie in the plane \( x_1 + \cdots + x_n = 0 \)
  - (b) must have eigenvalue \( \in (-1, 1) \)

(b) is important for dynamical systems/Markov chains — if the initial state is \( \hat{x}_0 = c_1 \hat{v}_1 + c_2 \hat{v}_2 + \cdots \) and \( \hat{v}_1 \) is steady-state vector, then

\[
\hat{x}(t) = A^t \hat{x}_0 = c_1 \hat{v}_1 + c_2 \left( \lambda_2 \right)^t \hat{v}_2 + \cdots \quad \xrightarrow{t \to \infty} \quad c_1 \hat{v}_1
\]

\( 0 \leq a \leq |\lambda_2| < 1 \)