Chapter 7: Spectral Theory

Greg Knapp

April 3, 2024

1 Eigenvalues and Eigenvectors

- Let's look at the linear map given by
$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & -2 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$
- We can play around with it for a bit, and find that $A(1,0) = (1,0)$. Also, $A(2,0) = (2,0)$. And $A(-1,0) = (-1,0)$.
- Weird!
- Question: are there any more vectors like that?
- Not really, but we get kinda close in the fourth quadrant.
- There, we find that $A(2,-1) = (4,-2)$. And $A(-2,1) = (-4,2)$. And $A(-4,2) = (-8,4)$.
- We can explore around for a little while, but we won't find anything else like this.
- So what happened here?
- In each case, we have a setting where Av was a scalar multiple of v.
- When we notice a phenomenon in mathematics, we like to give it a name, and then explore it.

Def: Suppose that $v \in \mathbb{R}^n$ is a nonzero vector and that A is an $n \times n$ matrix. Suppose also that $\lambda \in \mathbb{R}$ and that $Av = \lambda v$. Then we call v an <u>eigenvector</u> and λ an <u>eigenvalue</u> of A. The set of eigenvalues of A is called the <u>spectrum</u> of A.

- Aside: where do the words eigenvector and eigenvalue come from?
  - In modern German, the prefix "eigen-" roughly translates into "self-"
  - So an eigenvector is a self-vector.
  - Not particularly enlightening.
  - We knew about eigenvectors and eigenvalues long before they had those names: Cauchy called them "characteristic roots."
  - Hilbert gave us the name "eigenvalue" long after Cauchy had come up with them.
  - He used "eigenschaften" as a translation into German of Cauchy's "caractéristique."
  - He then kept the prefix of "eigenschaften" to refer to the various numbers and vectors that Cauchy had discovered.
  - We then took Hilbert's notation, but only translated half of the word, and we got eigenvalues and eigenvectors.
  - See Eric Tou's article: `https://maa.org/press/periodicals/convergence/math-origins-eigenvectors-and-`

Ex: The matrix A has:

- Eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = 2$.
- Eigenvectors: $(1,0), (2,0), (-\pi, 0), \ldots$ and $(2,-1), (-2,1), (4,-2), (-2\ln(2), \ln(2)), \ldots$.

- In the above example, even though there are infinitely many eigenvectors, there are essentially only two classes of eigenvectors: those associated with λ_1 (scalar multiples of $(1,0)$)and those associated with λ_2 (scalar multiples of $(-2,1)$).

Ex: Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the function which reflects a vector across the line $y = 2x$. What are the eigenvalues and eigenvectors of T?

- Any vector on the line $y = 2x$ is an eigenvector with eigenvalue 1.
- These are the scalar multiples of $(1,2)$.
- Any vector perpendicular to $y = 2x$ is an eigenvector with eigenvalue -1.
- These are the scalar multiples of $(-2,1)$.
- No other vector is an eigenvector.

Q: Given a matrix, how do we find the eigenvectors and eigenvalues without appealing to geometry?

- Quick fact: If A is an invertible matrix, and $Av = 0$, then $v = 0$.
  - Why?
  - Multiply both sides of $Av = 0$ by A^{-1}.
- Equivalent fact: If $Av = 0$ for a nonzero vector v, then A is not invertible.
- Now we can find eigenvectors and values.
- Eigenvalues first:
  - Suppose we have an eigenvalue λ of a matrix A.
  - Then there is a nonzero vector v so that $Av = \lambda v$
  - Rewrite to get $(\lambda I_n - A)v = 0$.
  - Now v is not zero.
  - So $\lambda I_n - A$ is noninvertible.
  - I.e. $\det(\lambda I_n - A) = 0$.
- It turns out that the reverse fact is true: if $\det(\lambda I_n - A) = 0$, then λ is an eigenvalue.
- Summary: the eigenvalues of a matrix A are exactly the roots of the polynomial $\det(xI_n - A)$.

Def: The <u>characteristic polynomial</u> of a $n \times n$ matrix A is $\det(xI_n - A)$.

Ex: Find the eigenvalues of $A = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right)$.

- Solution: $\frac{5 \pm \sqrt{33}}{2}$.
- Hm, we didn't get great numbers, but we did get some numbers I guess.
- To be honest, though, the eigenvalues aren't the interesting thing.
- We started our discussion by discovering eigenvectors, then eigenvalues were a necessary evil to describe the interesting phenomenon.

Ex: What are the eigenvectors and eigenvalues of $A = \left(\begin{smallmatrix} 1 & -2 \\ 0 & 2 \end{smallmatrix}\right)$?

- Eigenvalues first: $\lambda_1 = 1$ and $\lambda_2 = 2$.

  - Now we want to find solutions to $Av = v$ and $Av = 2v$.
  - I.e. we want to find (x, y) so that $(I_2 - A)(x, y) = 0$, and $(2I_2 - A)(x, y) = 0$
  - Hm, these look like some systems of equations.
  - $\lambda_1 = 1$:
    * $I_2 - A = \left(\begin{smallmatrix} 0 & 2 \\ 0 & -1 \end{smallmatrix}\right)$
    * The RREF of the augmented matrix for the corresponding system is:

$$\begin{pmatrix} [cc|c]0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

    * We assign parameters to columns without leading 1s, so any vector of the form $(t, 0) = t(1, 0)$ where $t \neq 0$ is an eigenvector with eigenvalue $\lambda_1 = 1$.
    * Cool, this is what we found earlier.
  - $\lambda_2 = 2$:
    * $2I_2 - A = \left(\begin{smallmatrix} 1 & 2 \\ 0 & 0 \end{smallmatrix}\right)$.
    * This is already in RREF, so we find that (assigning $y = s$) any vector of the form $(-2s, s) = s(-2, 1)$ is an eigenvector with eigenvalue $\lambda_2 = 2$.
    * Cool, this is also what we found earlier.

- Observation: The eigenvalues of a triangular matrix are its diagonal entries.

Ex: Find the eigenvectors and eigenvalues of

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{pmatrix}.$$

  - Eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = -3$.
  - $\lambda_1 = 1$:
    * The RREF of the relevant system is

$$\begin{pmatrix} [ccc|c]0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

      meaning that we get $x = s$, $y = t$, and $z = 0$.
    * So the eigenvectors with eigenvalue 1 are the nonzero vectors of the form

$$s \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

  - $\lambda_2 = -3$:
    * The RREF of the relevant system is

$$\begin{pmatrix} [ccc|c]1 & 0 & \frac{1}{4} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

      and setting $z = p$, we get that the eigenvectors with eigenvalue -3 are the nonzero vectors of the form

$$p \begin{pmatrix} -\frac{1}{4} \\ 0 \\ 1 \end{pmatrix}.$$

- Observation: when finding eigenvectors, we're solving systems of linear equations that are guaranteed to have parameters. This means that we'll have "basic solutions" and we "get" to give these basic solutions a special name: basic eigenvectors.
- Observation: some eigenvalues count for more than just one (whatever this means). The eigenvalue $\lambda_1 = 1$ counted "twice" since the characteristic polynomial factored as $(x-1)^2(x+3)$.

Def: Suppose that A is a square matrix with eigenvalues $\lambda_1, \dots, \lambda_k$. Suppose that the characteristic polynomial of A factors as $(x-\lambda_1)^{n_1} \dots (x-\lambda_k)^{n_k}$. Then the multiplicity of λ_i is n_i.

Ex: Find the eigenvalues and eigenvectors of $A = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

- $xI_2 - A = \left(\begin{smallmatrix} x & 1 \\ -1 & x \end{smallmatrix}\right)$.
- Characteristic polynomial: x^2+1.
- This has roots, I guess...$x = \pm i$.
- $\lambda_1 = i$:
  - $iI_2 - A = \left(\begin{smallmatrix} i & 1 \\ -1 & i \end{smallmatrix}\right)$
  - Relevant RREF is
  $$\begin{pmatrix} [cc|c]1 & -i & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$
  - So I guess the eigenvectors with eigenvalue i are any of the nonzero vectors of the form $t(i,1)$?
- $\lambda_2 = -i$:
  - $-iI_2 - A = \left(\begin{smallmatrix} -i & 1 \\ -1 & -i \end{smallmatrix}\right)$.
  - Relevant RREF is
  $$\begin{pmatrix} [cc|c]1 & i & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$
  - So the eigenvectors with eigenvalue $-i$ are any of the nonzero vectors of the form $s(-i,1)$.

- What's going on here? Remember that A represents rotation: this can't have (real) eigenvectors.

Ex: Find the eigenvalues and eigenvectors of $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

- Eigenvalues are $\lambda_1 = 1$ and that's it.
- To find eigenvectors, relevant RREF is
$$\begin{pmatrix} [cc|c]0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$
- Now we get that any vector of the form $t(1,0)$ is an eigenvector with eigenvalue 1.

- Observation: the number of basic eigenvectors of λ does not have to equal the multiplicity of λ.
- Fact: number of basic eigenvectors of λ is less than or equal to the multiplicity of λ.

2 Diagonalization

Ex: Let $A = \left(\begin{smallmatrix} 4 & -2 \\ -1 & 3 \end{smallmatrix}\right)$. Compute A^{100}.

- Here's a magic observation: $A = PDP^{-1}$ where
$$P = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix} \qquad P^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}.$$
- We'll talk about where this comes from later, but for now, we'll just make the observation.

- Why is it helpful?
-

$$A^{100} = (PDP^{-1})^{100} = PD^{100}P^{-1} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2^{100} & 0 \\ 0 & 5^{100} \end{pmatrix} \cdot \frac{1}{3} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}.$$

- Lesson: if $A = PDP^{-1}$ for a diagonal matrix D, then A^n is easy to compute.

Def: A square matrix A is diagonalizable if there exists an invertible matrix P and a diagonal matrix D so that $A = PDP^{-1}$.

- How the heck could we ever hope to do this?

Def: If A and B are $n \times n$ matrices, they are said to be similar if there exists an invertible $n \times n$ matrix P for which $A = PBP^{-1}$.

- Fact: If A and B are similar, then they have the same eigenvalues.
  - Suppose $A = PBP^{-1}$ and λ is an eigenvalue of A with eigenvector v.
  - Then we claim that λ is an eigenvalue of B with eigenvector $P^{-1}v$.
  - Note that $Av = \lambda v$ implies that $PBP^{-1}v = \lambda v$, i.e. $B(P^{-1}v) = \lambda(P^{-1}v)$.
  - The reverse is also true: any eigenvalue of B is an eigenvalue of A.
  - So A and B have the same eigenvalues.
- Now, if we hope to write $A = PDP^{-1}$ where D is diagonal.
- This means that A and D had better have the same eigenvalues.
- But the eigenvalues of D are the diagonal entries.
- So the diagonal entries of D had better be the eigenvalues of A.
- The matrix P is a little harder to figure out. Here's the theorem:

Thm: Suppose that A is an $n \times n$ matrix. Then A is diagonalizable if and only if A has n eigenvectors $v_1, \ldots, v_n$ so that the matrix

$$P = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \ldots & v_n \\ | & | & & | \end{pmatrix}$$

is invertible. If A is diagonalizable, then $A = PDP^{-1}$ where $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ where λ_i is the eigenvalue for v_i.

Ex: Diagonalize the matrix $A = \left(\begin{smallmatrix} 4 & -2 \\ -1 & 3 \end{smallmatrix}\right)$, if possible.

- This is the matrix from the beginning of this section.

Ex: Diagonalize the matrix

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{pmatrix}$$

if possible.

- In the last section, we saw that A has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -3$.
- The eigenvectors for $\lambda_1 = 1$ have the form $(s, t, 0)$ for s and t not both zero.
- The eigenvectors for $\lambda_2 = -3$ have the form $(-p/4, 0, p)$ for $p \neq 0$.
- We want to pick three eigenvectors which make $P = (v_1 v_2 v_3)$ invertible.
- Always safe choice: pick basic eigenvectors for v_1, v_2, v_3.

- Talk about nonuniqueness here: choose different ordering, and choose different eigenvectors.

Ex: Diagonalize the matrix $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ if possible.

- Before, we saw that A has only one eigenvalue $\lambda = 1$.
- Every eigenvector of A has the form $(t, 0)$ for $t \neq 0$.
- So if we try to form a matrix of eigenvectors $P = \left(\begin{smallmatrix} t_1 & t_2 \\ 0 & 0 \end{smallmatrix}\right)$, we can't make P invertible.
- This means that A is not diagonalizable.

- Fact: A is diagonalizable if and only if for every eigenvalue λ, the number of basic eigenvalues equals the multiplicity of λ.

Ex: Suppose that A is a 3×3 matrix with eigenvalues $1, 2, 3$. Is A diagonalizable?

- The characteristic polynomial of A has degree 3, and since it has three different roots, it must be $(x-1)(x-2)(x-3)$.
- Now the multiplicity of each eigenvalue is 1.
- The number of basic eigenvectors of each eigenvalue must be at least 1, otherwise we wouldn't have an eigenvalue.
- But the number of basic eigenvectors of each eigenvalue must is no more than the multiplicity, i.e. 1.
- So the number of basic eigenvectors equals the multiplicity of each eigenvalue.
- Hence, A is diagonalizable.

- Note: "dimension of the λ-eigenspace" means "number of basic λ-eigenvectors."
- Note: "basis of the λ-eigenspace" means "list of basic λ-eigenvectors."
- Note: $A = PDP^{-1}$ is equivalent to $P^{-1}AP = D$.

3 Applications of Spectral Theory

- Now, we're going to see some powerful examples of linear algebra!
- One application is mathematical, and the other is intensely practical.
- First up, we'll look at linear recurrence sequences.

Def: A linear recurrence sequence is a sequence of numbers $x_0, x_1, \ldots$ for which there exist constants $c_1, \ldots, c_d$ so that
$$x_n = c_1 x_{n-1} + c_2 x_{n-2} + \cdots + c_d x_{n-d}.$$

Ex: The Fibonacci sequence $0, 1, 1, 2, 3, 5, 8, \ldots$ is a linear recurrence sequence because it satisfies $F_n = F_{n-1} + F_{n-2}$.

- An important question we might ask: Can we describe F_n as a function of n? Can I find F_{100} without finding $F_0, \ldots, F_{99}$?
  - We can do this by considering vectors of the form (F_k, F_{k+1}).
  - For example, we have $(0,1)$, $(1,1)$, $(1,2)$, $(2,3)$, $(3,5)$, $(5,8)$, etc.
  - How do we get from one vector to the next?
  - If you know that $(144, 233)$ is one such vector, the next must be $(233, 144 + 233)$.
  - To get from (F_k, F_{k+1}) to (F_{k+1}, F_{k+2}), we take F_{k+1} to be the second entry in the previous vector and we take F_{k+2} to be the sum of the entries of the previous vector.

- This sounds awfully linear...
- Notice that we have

$$\begin{pmatrix} F_{k+1} \\ F_{k+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} F_k \\ F_{k+1} \end{pmatrix}$$

- This means that

$$\begin{pmatrix} F_k \\ F_{k+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^k \begin{pmatrix} F_0 \\ F_1 \end{pmatrix}.$$

- But now we're interested in raising a matrix to a power! This means we have to diagonalize:
- The eigenvalues of $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ are the roots of $x(x-1)-1 = x^2 - x - 1$, i.e.

$$\frac{1 \pm \sqrt{5}}{2}.$$

- An eigenvector for $\lambda_1 = \frac{1+\sqrt{5}}{2}$ is

$$\begin{pmatrix} \frac{-1+\sqrt{5}}{2} \\ 1 \end{pmatrix}.$$

- An eigenvector for $\lambda_2 = \frac{1-\sqrt{5}}{2}$ is

$$\begin{pmatrix} \frac{-1-\sqrt{5}}{2} & 1 \end{pmatrix}.$$

- Hence,

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^k = \begin{pmatrix} \frac{1-\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}^k \begin{pmatrix} \frac{1-\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix}^{-1}.$$

- Hence,

$$\begin{aligned} \begin{pmatrix} F_k \\ F_{k+1} \end{pmatrix} &= \begin{pmatrix} \frac{1-\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}^k \begin{pmatrix} \frac{1-\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k \right) \\ \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2}\right)^{k+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{k+1} \right) \end{pmatrix}. \end{aligned}$$

- Letting $\varphi_1 = \frac{1+\sqrt{5}}{2}$ and $\varphi_2 = \frac{1-\sqrt{5}}{2}$, we have the nice form

$$F_k = \frac{1}{\sqrt{5}} (\varphi_1^k - \varphi_2^k).$$

Ex: Find a closed-form expression for the sequence with $x_0 = 0$, $x_1 = 0$, $x_2 = 1$, and $x_n = 6x_{n-1} - 11x_{n-2} + 6x_{n-3}$.

- We'll do this problem by first, writing down the recurrence relation as a matrix:

$$\begin{pmatrix} x_n \\ x_{n+1} \\ x_{n+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{pmatrix} \begin{pmatrix} x_{n-1} \\ x_n \\ x_{n+1} \end{pmatrix}$$

- This means that

$$\begin{pmatrix} x_n \\ x_{n+1} \\ x_{n+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{pmatrix}^n \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

– Now we diagonalize the coefficient matrix:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix}^{-1}$$

– Then large powers of the coefficient matrix are:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{pmatrix}^k = \frac{1}{2} \begin{pmatrix} -6(2^k - 3^{k-1} - 1) & 2^{k+3} - 3^{k+1} - 5 & -2^{k+1} + 3^k + 1 \\ -6(2^{k+1} - 3^k - 1) & 2^{k+4} - 3^{k+2} - 5 & -2^{k+2} + 3^{k+1} + 1 \\ -6(2^{k+2} - 3^{k+1} - 1) & 2^{k+5} - 3^{k+3} - 5 & -2^{k+3} + 3^{k+2} + 1 \end{pmatrix}$$

– Finally, multiply by the initial condition vector $(0, 0, 1)$ to get

$$x_k = \frac{-2^{k+1} + 3^k + 1}{2}.$$

- The other type of application is very practical and relevant.
- Markov chains are/were commonly used in predictive text algorithms.
- Amusing example from around the time that I was graduating from college: `https://kingjamesprogramming.tumblr.com/`
- Another example is the type of predictive text algorithm that your phone/Google/Apple/Microsoft uses to predict what words you want to use when writing an email or text.
- Modern predictive text algorithms (e.g. ChatGPT) use much more sophisticated algorithms.
- But we can do a toy example to see the basic idea.
- You have a language consisting of only three words: "every," "student," and "learns."
- You also have a large book filled with sentences consisting of those three words.
- After doing some statistical analysis on the book, you find that:
  - If you find the word "every," the next word is "every" 10% of the time, the next word is "student" 60% of the time, and the next word is "learns" 30% of the time.
  - If you find the word "student," the next word is "every" 30% of the time, the next word is "student" 20% of the time, and the next word is "learns" 50% of the time.
  - If you find the word "learns," the next word is "every" 50% of the time, the next word is "student" 40% of the time, and the next word is "learns" 10% of the time.

Ex: The first word in the book is "Every." What is the probability that

1. ...the word "learns" appears next?
2. ...the word "learns" appears 2 words after the word "every?"
3. ...the word "learns" appears n words after the word "every?"

1. The probability that the second word is "learns" is 0.3, just reading off of the information given.
2. Draw a tree where the the first level is the word "every," the second level is each of the words "every," "student," and "learns" with the edges weighted appropriately, and the third level is a bunch of copies of the word "learns," with the edges weighted appropriately. The outcome is

$$0.1 \cdot 0.3 + 0.6 \cdot 0.5 + 0.3 \cdot 0.1 = 0.36.$$

So there's a 36% chance that the third word is "learns." Hm, something looks linear here...

3. This one is trickier. A matrix is a helpful tool here!
   - We're going to set up the matrix of probabilities: the column will correspond to the "current" word and the row will correspond to the "next" word:

$$M = \begin{pmatrix} 0.1 & 0.3 & 0.5 \\ 0.6 & 0.2 & 0.4 \\ 0.3 & 0.5 & 0.1 \end{pmatrix}$$

   - Quick observation: the columns each sum to 1.
   - Let's think about the matrix M^2. The computations are a little gnarly, but

$$M^2 = \begin{pmatrix} 0.34 & 0.34 & 0.22 \\ 0.3 & 0.42 & 0.42 \\ 0.36 & 0.24 & 0.36 \end{pmatrix}.$$

   - That 0.36 in the lower left (corresponding to the "every" column and the "learns" row) was exactly the same computation we did before.
   - General fact: The entry in the "word1" column and "word2" row of M^n is the probability that "word2" appears n words after "word1."
   - By computing M^2, we can see that the probability that "every" is two words after "every" is 0.34, while the probability that "student" is two words after "every" is 0.3.
   - To get M^n, we need to diagonalize.
   - Use a computer to do this: find eigenvalues and eigenvectors. One eigenvalue is $\lambda_1 = 1$ and an eigenvector is $(52/172, 66/172, 54/172)$. The other two eigenvalues are complex: $\lambda_2 = \frac{-3+i\sqrt{3}}{10}$ and $\lambda_3 = \frac{-3-i\sqrt{3}}{10}$.
   - Use a computer to do this: the probability that "learns" appears n words after the word "every" is:

$$\frac{1}{172}\left((-27+31\sqrt{3}i)\left(\frac{-3+i\sqrt{3}}{10}\right)^n + (-27-31\sqrt{3}i)\left(\frac{-3-i\sqrt{3}}{10}\right)^n + 54\right).$$

   - Observation: as n grows, the two exponential terms get closer and closer to zero.
   - You can apply this observation to every entry of the matrix. As n grows, the entries of M^n get closer and closer to:

$$\frac{1}{172}\begin{pmatrix} 52 & 52 & 52 \\ 66 & 66 & 66 \\ 54 & 54 & 54 \end{pmatrix}$$

   - Interpretation: the probability that a given word is "every" is 52/172. The probability that a given word is "student" is 66/172. The probability that a given word is "learns" is 54/172.

Def: A matrix M is called a <u>Markov/migration/transition matrix</u> if:

- It is square.
- All of its entries are nonnegative.
- Each of its columns sums to 1.

- We think of the columns of the matrix as representing "current states" and the rows as representing "future states."
- Fact: Any Markov matrix has 1 has an eigenvalue with only 1 basic eigenvector.
- Fact: Suppose M is a Markov matrix and suppose there is a k for which the entries of M^k are all strictly greater than 0. Then as n grows, the columns of M^n approach the 1-eigenvector whose entries sum to 1.

- Note: your book misses the fact the requirement that M^k has strictly positive entries for some k. However, this will be true for all examples that you actually have to worry about for this course.

- For a counterexample to your book's claim that the state vector always approaches the appropriate 1-eigenvector, use the Markov matrix $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and the initial state $(1, 0)$.

Ex: There are three towns, A, B, and C. Each year, 10% of town A's residents move to town B and 20% of town A's residents move to town C. Each year, 40% of town B's residents move to town A and 10% move to town C. Each year, 20% of town C's residents move to town A and 20% move to town B. Initially, town A has 300 people, town B has 100 people, and town C has 200 people. How many people live in each town after a long time?

- Set up the transition matrix:
$$M = \begin{pmatrix} 0.7 & 0.4 & 0.2 \\ 0.1 & 0.5 & 0.2 \\ 0.2 & 0.1 & 0.6 \end{pmatrix}.$$
- After t years, we can tell how many residents of each town there are by looking at
$$\begin{pmatrix} 0.7 & 0.4 & 0.2 \\ 0.1 & 0.5 & 0.2 \\ 0.2 & 0.1 & 0.6 \end{pmatrix}^t \begin{pmatrix} 300 \\ 100 \\ 200 \end{pmatrix}.$$
- The entries of M are strictly positive.
- Hence, as t grows, we can tell what M^t looks like by finding the 1-eigenvector of M whose entries sum to 1.
- After doing some work, you find that the eigenvectors for eigenvalue 1 have the form $(18t/11, 8t/11, t)$.
- Getting those numbers to add up to 1 requires $t = 11/37$, so you find that
$$M^t \to \frac{1}{37}\begin{pmatrix} 18 & 18 & 18 \\ 8 & 8 & 8 \\ 11 & 11 & 11 \end{pmatrix}$$
- Now multiply by $(300, 100, 200)$ to get about $(292, 130, 178)$.

Def: For a Markov matrix M and an initial condition v, the steady-state vector is the vector which $M^n v$ approaches as n grows, if such a vector exists.

- Fact: The steady state vector is the 1-eigenvector whose entries sum to the sum of entries of v.

Ex: Suppose that M is a Markov matrix with positive entries. Suppose that M has 1-eigenvectors of the form $t(2, 2, 1)$ for $t \neq 0$. Suppose that $(1, 3, 9)$ is an initial condition. What does $M^n \cdot (1, 3, 9)$ approach as n grows?

- We want a vector of the form $t(2, 2, 1)$ whose entries add up to $1 + 3 + 9 = 13$.
- This means that we need $t = 13/5$.
- The steady state vector is then $(26/5, 26/5, 13/5)$.

Ex: You are battling some Pokémon against your friend. In a battle, you each typically bring 6 Pokémon, then you battle, and you see who wins. You are better than your friend, so to make things more competitive, if you win a match, you bring one fewer Pokémon to the next battle, down to a minimum of four. Your probabilities of winning are as follows:

Number of Pokémon you bring	Probability you win
6	0.8
5	0.6
4	0.3

What proportion of games do you win in the long run?

- We want to set up a Markov matrix for this situation.
- We need to have some states, so we'll choose three states:
  * You win zero games in a row
  * You win one game in a row
  * You win two or more games in a row
- From here, we can set up a transition matrix describing the various probabilities of transitioning from state to state:
$$M = \begin{pmatrix} 0.2 & 0.4 & 0.7 \\ 0.8 & 0 & 0 \\ 0 & 0.6 & 0.3 \end{pmatrix}$$
- Not hard to check: M^2 has strictly positive entries, so the columns of M^n converge to the 1-eigenvector.
- Now we want to find the steady state vector for this system:
$$I_3 - M = \begin{pmatrix} 0.8 & -0.4 & -0.7 \\ -0.8 & 1 & 0 \\ 0 & -0.6 & 0.7 \end{pmatrix}$$
- Putting this in RREF yields:
$$\begin{pmatrix} 1 & 0 & -35/24 \\ 0 & 1 & -28/24 \\ 0 & 0 & 0 \end{pmatrix}$$
- So we get eigenvectors of the form: $s(35/24, 28/24, 1)$.
- The steady state vector is the 1-eigenvector whose coordinates add up to 1, which is $(35/87, 28/87, 24/87)$.
- Hence, 35/87 of the the time, we're in the "you win zero in a row" state
- 28/87 of the the time, we're in the "you win one in a row" state
- And 24/87 of the time, we're in the "you win two or more in a row" state.
- The long term proportion of games that you win is the proportion of events that occur in either of the last two states, which is 52/87.