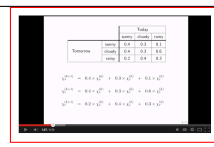
Week 12

Eigenvalues, Eigenvectors, and Diagonalization

12.1 Opening Remarks

12.1.1 Predicting the Weather, Again



☛ **View at edX**

Let us revisit the example from Week 4, in which we had a simple model for predicting the weather. Again, the following table tells us how the weather for any day (e.g., today) predicts the weather for the next day (e.g., tomorrow):

		Today		
		sunny	cloudy	rainy
Tomorrow	sunny	0.4	0.3	0.1
	cloudy	0.4	0.3	0.6
	rainy	0.2	0.4	0.3

This table is interpreted as follows: If today is rainy, then the probability that it will be cloudy tomorrow is 0.6, etc.

We introduced some notation:

- Let $\chi_s^{(k)}$ denote the probability that it will be sunny k days from now (on day k).
- Let $\chi_c^{(k)}$ denote the probability that it will be cloudy k days from now.
- Let $\chi_r^{(k)}$ denote the probability that it will be rainy k days from now.

We then saw that predicting the weather for day $k+1$ based on the prediction for day k was given by the system of linear equations

$$\begin{array}{rcrcrcr}
\chi_s^{(k+1)} & = & 0.4 \times \chi_s^{(k)} & + & 0.3 \times \chi_c^{(k)} & + & 0.1 \times \chi_r^{(k)} \\
\chi_c^{(k+1)} & = & 0.4 \times \chi_s^{(k)} & + & 0.3 \times \chi_c^{(k)} & + & 0.6 \times \chi_r^{(k)} \\
\chi_r^{(k+1)} & = & 0.2 \times \chi_s^{(k)} & + & 0.4 \times \chi_c^{(k)} & + & 0.3 \times \chi_r^{(k)}.
\end{array}$$

which could then be written in matrix form as

$$x^{(k)} = \begin{pmatrix} \chi_s^{(k)} \\ \chi_c^{(k)} \\ \chi_r^{(k)} \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 0.4 & 0.3 & 0.1 \\ 0.4 & 0.3 & 0.6 \\ 0.2 & 0.4 & 0.3 \end{pmatrix}.$$

so that

$$\begin{pmatrix} \chi_s^{(k+1)} \\ \chi_c^{(k+1)} \\ \chi_r^{(k+1)} \end{pmatrix} = \begin{pmatrix} 0.4 & 0.3 & 0.1 \\ 0.4 & 0.3 & 0.6 \\ 0.2 & 0.4 & 0.3 \end{pmatrix} \begin{pmatrix} \chi_s^{(k)} \\ \chi_c^{(k)} \\ \chi_r^{(k)} \end{pmatrix}$$

or $x^{(k+1)} = Px^{(k)}$.

Now, if we start with day zero being cloudy, then the predictions for the first two weeks are given by

Day #	Sunny	Cloudy	Rainy
0	0.	1.	0.
1	0.3	0.3	0.4
2	0.25	0.45	0.3
3	0.265	0.415	0.32
4	0.2625	0.4225	0.315
5	0.26325	0.42075	0.316
6	0.263125	0.421125	0.31575
7	0.2631625	0.4210375	0.3158
8	0.26315625	0.42105625	0.3157875
9	0.26315813	0.42105188	0.31579
10	0.26315781	0.42105281	0.31578938
11	0.26315791	0.42105259	0.3157895
12	0.26315789	0.42105264	0.31578947
13	0.2631579	0.42105263	0.31578948
14	0.26315789	0.42105263	0.31578947

What you notice is that eventually

$$x^{(k+1)} \approx Px^{(k)}.$$

What this means is that there is a vector x such that $Px = x$. Such a vector (if it is non-zero) is known as an eigenvector. In this example, it represents the long-term prediction of the weather. Or, in other words, a description of "typical weather": approximately 26% of the time it is sunny, 42% of the time it is cloudy, and 32% of the time rainy.

The question now is: How can we compute such vectors?

Some observations:

- $Px = x$ means that $Px - x = 0$ which in turn means that $(P - I)x = 0$.
- This means that x is a vector in the null space of $P - I$: $x \in \mathcal{N}(P - I)$.
- But we know how to find vectors in the null space of a matrix. You reduce a system to row echelon form, identify the free variable(s), etc.
- But we also know that a nonzero vector in the null space is not unique.
- In this **particular** case, we know two more pieces of information:
  - The components of x must be nonnegative (a negative probability does not make sense).
  - The components of x must add to one (the probabilities must add to one).

The above example can be stated as a more general problem:

$$Ax = \lambda x,$$

which is known as the (algebraic) eigenvalue problem. Scalars λ that satisfy $Ax = \lambda x$ for nonzero vector x are known as **eigenvalues** while the nonzero vectors are known as **eigenvectors**.

From the table above we can answer questions like "what is the typical weather?" (Answer: Cloudy). An approach similar to what we demonstrated in this unit is used, for example, to answer questions like "what is the most frequently visited webpage on a given topic?"

12.1.2 Outline

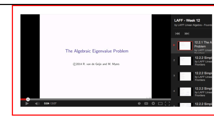
12.1.3 What You Will Learn

Upon completion of this unit, you should be able to

- Determine whether a given vector is an eigenvector for a particular matrix.
- Find the eigenvalues and eigenvectors for small-sized matrices.
- Identify eigenvalues of special matrices such as the zero matrix, the identity matrix, diagonal matrices, and triangular matrices.
- Interpret an eigenvector of A, as a direction in which the "action" of A, Ax, is equivalent to x being scaled without changing its direction. (Here scaling by a negative value still leaves the vector in the same direction.) Since this is true for any scalar multiple of x, it is the direction that is important, not the length of x.
- Compute the characteristic polynomial for 2×2 and 3×3 matrices.
- Know and apply the property that a matrix has an inverse if and only if its determinant is nonzero.
- Know and apply how the roots of the characteristic polynomial are related to the eigenvalues of a matrix.
- Recognize that if a matrix is real valued, then its characteristic polynomial has real valued coefficients but may still have complex eigenvalues that occur in conjugate pairs.
- Link diagonalization of a matrix with the eigenvalues and eigenvectors of that matrix.
- Make conjectures, reason, and develop arguments about properties of eigenvalues and eigenvectors.
- Understand practical algorithms for finding eigenvalues and eigenvectors such as the power method for finding an eigenvector associated with the largest eigenvalue (in magnitude).

12.2 Getting Started

12.2.1 The Algebraic Eigenvalue Problem



☛ **View at edX**

The *algebraic eigenvalue problem* is given by

$$Ax = \lambda x.$$

where $A \in \mathbb{R}^{n \times n}$ is a square matrix, λ is a scalar, and x is a nonzero vector. Our goal is to, given matrix A, compute λ and x. It must be noted from the beginning that λ may be a complex number and that x will have complex components if λ is complex valued. If $x \neq 0$, then λ is said to be an *eigenvalue* and x is said to be an eigenvector associated with the eigenvalue λ. The tuple (λ, x) is said to be an *eigenpair*.

Here are some equivalent statements:

- $Ax = \lambda x$,, where $x \neq 0$.
  This is the statement of the (algebraic) eigenvalue problem.
- $Ax - \lambda x = 0$, where $x \neq 0$.
  This is merely a rearrangement of $Ax = \lambda x$.
- $Ax - \lambda Ix = 0$, where $x \neq 0$.
  Early in the course we saw that $x = Ix$.
- $(A - \lambda I)x = 0$, where $x \neq 0$.
  This is a matter of fractoring' x out.
- $A - \lambda I$ is singular.
  Since there is a vector $x \neq 0$ such that $(A - \lambda I)x = 0$.
- $\mathcal{N}(A - \lambda I)$ contains a nonzero vector x.
  This is a consequence of there being a vector $x \neq 0$ such that $(A - \lambda I)x = 0$.
- $\dim(\mathcal{N}(A - \lambda I)) > 0$.
  Since there is a nonzero vector in $\mathcal{N}(A - \lambda I)$, that subspace must have dimension greater than zero.

If we find a vector $x \neq 0$ such that $Ax = \lambda x$, it is certainly not unique.

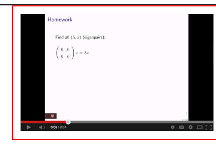
- For any scalar α, $A(\alpha x) = \lambda(\alpha x)$ also holds.
- If $Ax = \lambda x$ and $Ay = \lambda y$, then $A(x + y) = Ax + Ay = \lambda x + \lambda y = \lambda(x + y)$.

We conclude that the set of all vectors x that satisfy $Ax = \lambda x$ is a subspace.

It is not the case that the set of all vectors x that satisfy $Ax = \lambda x$ is the set of all eigenvectors associated with λ. After all, the zero vector is in that set, but is not considered an eigenvector.

It is important to think about eigenvalues and eigenvectors in the following way: If x is an eigenvector of A, then x is a direction in which the "action" of A (in other words, Ax) is equivalent to x being scaled in length without changing its direction other than changing sign. (Here we use the term "length" somewhat liberally, since it can be negative in which case the direction of x will be exactly the opposite of what it was before.) Since this is true for any scalar multiple of x, it is the direction that is important, not the magnitude of x.

12.2.2 Simple Examples



☛ View at edX

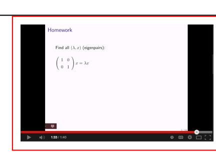
In this unit, we build intuition about eigenvalues and eigenvectors by looking at simple examples.

Homework 12.2.2.1 Which of the following are eigenpairs (λ, x) of the 2×2 zero matrix:

$$\left(\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array} \right) x = \lambda x,$$

where $x \neq 0$.
(Mark all correct answers.)

1. $(1, \left(\begin{array}{c} 0 \\ 0 \end{array} \right))$.
2. $(0, \left(\begin{array}{c} 1 \\ 0 \end{array} \right))$.
3. $(0, \left(\begin{array}{c} 0 \\ 1 \end{array} \right))$.
4. $(0, \left(\begin{array}{r} -1 \\ 1 \end{array} \right))$.
5. $(0, \left(\begin{array}{c} 1 \\ 1 \end{array} \right))$.
6. $(0, \left(\begin{array}{c} 0 \\ 0 \end{array} \right))$.



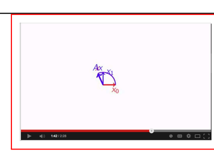
☛ View at edX

Homework 12.2.2.2 Which of the following are eigenpairs (λ, x) of the 2×2 zero matrix:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x = \lambda x,$$

where $x \neq 0$.
(Mark all correct answers.)

1. $(1, \begin{pmatrix} 0 \\ 0 \end{pmatrix})$.
2. $(1, \begin{pmatrix} 1 \\ 0 \end{pmatrix})$.
3. $(1, \begin{pmatrix} 0 \\ 1 \end{pmatrix})$.
4. $(1, \begin{pmatrix} -1 \\ 1 \end{pmatrix})$.
5. $(1, \begin{pmatrix} 1 \\ 1 \end{pmatrix})$.
6. $(-1, \begin{pmatrix} 1 \\ -1 \end{pmatrix})$.



☛ View at edX

Homework 12.2.2.3 Let $A = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$.

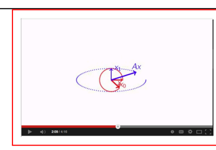
- $\begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ so that $(3, \begin{pmatrix} 1 \\ 0 \end{pmatrix})$ is an eigenpair.

  True/False

- The set of all eigenvectors associated with eigenvalue 3 is characterized by (mark all that apply):
  - All vectors $x \neq 0$ that satisfy $Ax = 3x$.
  - All vectors $x \neq 0$ that satisfy $(A - 3I)x = 0$.
  - All vectors $x \neq 0$ that satisfy $\begin{pmatrix} 0 & 0 \\ 0 & -4 \end{pmatrix} x = 0$.
  - $\left\{ \begin{pmatrix} \chi_0 \\ 0 \end{pmatrix} \middle| \chi_0 \text{ is a scalar} \right\}$

- $\begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -1 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ so that $(-1, \begin{pmatrix} 0 \\ 1 \end{pmatrix})$ is an eigenpair.

  True/False

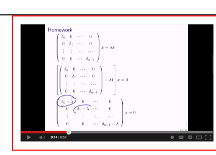


☛ View at edX

Homework 12.2.2.4 Consider the diagonal matrix $\begin{pmatrix} \delta_0 & 0 & \cdots & 0 \\ 0 & \delta_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_{n-1} \end{pmatrix}$.

Eigenpairs for this matrix are given by $(\delta_0, e_0), (\delta_1, e_1), \cdots, (\delta_{n-1}, e_{n-1})$, where e_j equals the jth unit basis vector.

Always/Sometimes/Never



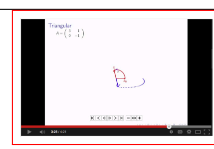
☛ View at edX

Homework 12.2.2.5 Which of the following are eigenpairs (λ, x) of the 2×2 triangular matrix:

$$\begin{pmatrix} 3 & 1 \\ 0 & -1 \end{pmatrix} x = \lambda x,$$

where $x \neq 0$.
(Mark all correct answers.)

1. $(-1, \begin{pmatrix} -1 \\ 4 \end{pmatrix})$.
2. $(1/3, \begin{pmatrix} 1 \\ 0 \end{pmatrix})$.
3. $(3, \begin{pmatrix} 1 \\ 0 \end{pmatrix})$.
4. $(-1, \begin{pmatrix} 1 \\ 0 \end{pmatrix})$.
5. $(3, \begin{pmatrix} -1 \\ 0 \end{pmatrix})$.
6. $(-1, \begin{pmatrix} 3 \\ -1 \end{pmatrix})$.

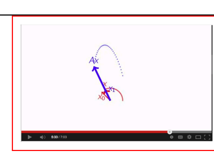


☛ **View at edX**

Homework 12.2.2.6 Consider the upper triangular matrix $U = \begin{pmatrix} \upsilon_{0,0} & \upsilon_{0,1} & \cdots & \upsilon_{0,n-1} \\ 0 & \upsilon_{1,1} & \cdots & \upsilon_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \upsilon_{n-1,n-1} \end{pmatrix}$.

The eigenvalues of this matrix are $\upsilon_{0,0}, \upsilon_{1,1}, \ldots, \upsilon_{n-1,n-1}$.

Always/Sometimes/Never



☛ **View at edX**

Below, on the left we discuss the general case, side-by-side with a specific example on the right.

General	Example
Consider $Ax = \lambda x$.	$\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \lambda \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix}.$
Rewrite as $Ax - \lambda x$	$\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} - \lambda \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$
Rewrite as $Ax - \lambda Ix = 0$.	$\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$
Now $[A - \lambda I]x = 0$	$\left[\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$
$A - \lambda I$ is the matrix A with λ subtracted from its diagonal elements.	$\begin{pmatrix} 1-\lambda & -1 \\ 2 & 4-\lambda \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$

Now $A - \lambda I$ has a nontrivial vector x in its null space if that matrix does *not* have an inverse. Recall that

$$\begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} \\ \alpha_{1,0} & \alpha_{1,1} \end{pmatrix}^{-1} = \frac{1}{\alpha_{0,0}\alpha_{1,1} - \alpha_{1,0}\alpha_{0,1}} \begin{pmatrix} \alpha_{1,1} & -\alpha_{0,1} \\ -\alpha_{1,0} & \alpha_{0,0} \end{pmatrix}.$$

Here the scalar $\alpha_{0,0}\alpha_{1,1} - \alpha_{1,0}\alpha_{0,1}$ is known as the determinant of 2×2 matrix A, $\det(A)$.

This turns out to be a general statement:

> Matrix A has an inverse if and only if its determinant is nonzero.
> We have not yet defined the determinant of a matrix of size greater than 2.

So, the matrix $\begin{pmatrix} 1-\lambda & -1 \\ 2 & 4-\lambda \end{pmatrix}$ does not have an inverse if and only if

$$\det(\begin{pmatrix} 1-\lambda & -1 \\ 2 & 4-\lambda \end{pmatrix}) = (1-\lambda)(4-\lambda) - (2)(-1) = 0.$$

But

$$(1-\lambda)(4-\lambda) - (2)(-1) = 4 - 5\lambda + \lambda^2 + 2 = \lambda^2 - 5\lambda + 6$$

This is a quadratic (second degree) polynomial, which has at most two district roots. In particular, by examination,

$$\lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3) = 0$$

so that this matrix has two eigenvalues: $\lambda = 2$ and $\lambda = 3$.

If we now take $\lambda = 2$, then we can determine an eigenvector associated with that eigenvalue:

$$\begin{pmatrix} 1-(2) & -1 \\ 2 & 4-(2) \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

or

$$\begin{pmatrix} -1 & -1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

By examination, we find that $\begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is a vector in the null space and hence an eigenvector associated with the eigenvalue $\lambda = 2$. (This is not a unique solution. Any vector $\begin{pmatrix} \chi \\ -\chi \end{pmatrix}$ with $\chi \neq 0$ is an eigenvector.)

Similarly, if we take $\lambda = 3$, then we can determine an eigenvector associated with that second eigenvalue:

$$\begin{pmatrix} 1-(3) & -1 \\ 2 & 4-(3) \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

or

$$\begin{pmatrix} -2 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

By examination, we find that $\begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ is a vector in the null space and hence an eigenvector associated with the eigenvalue $\lambda = 3$. (Again, this is not a unique solution. Any vector $\begin{pmatrix} \chi \\ -2\chi \end{pmatrix}$ with $\chi \neq 0$ is an eigenvector.)

The above discussion identifies a systematic way for computing eigenvalues and eigenvectors of a 2×2 matrix:

- Compute
$$\det\left(\begin{pmatrix} (\alpha_{0,0}-\lambda) & \alpha_{0,1} \\ \alpha_{1,0} & (\alpha_{1,1}-\lambda) \end{pmatrix}\right) = (\alpha_{0,0}-\lambda)(\alpha_{1,1}-\lambda) - \alpha_{0,1}\alpha_{1,0}.$$
- Recognize that this is a second degree polynomial in λ.
- It is called the *characteristic polynomial* of the matrix A, $p_2(\lambda)$.
- Compute the coefficients of $p_2(\lambda)$ so that
$$p_2(\lambda) = -\lambda^2 + \beta\lambda + \gamma.$$
- Solve
$$-\lambda^2 + \beta\lambda + \gamma = 0$$
for its roots. You can do this either by examination, or by using the quadratic formula:
$$\lambda = \frac{-\beta \pm \sqrt{\beta^2 + 4\gamma}}{-2}.$$
- For each of the roots, find an eigenvector that satisfies
$$\begin{pmatrix} (\alpha_{0,0}-\lambda) & \alpha_{0,1} \\ \alpha_{1,0} & (\alpha_{1,1}-\lambda) \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
The easiest way to do this is to subtract the eigenvalue from the diagonal, set one of the components of x to 1, and then solve for the other component.
- Check your answer! It is a matter of plugging it into $Ax = \lambda x$ and seeing if the computed λ and x satisfy the equation.

A 2×2 matrix yields a characteristic polynomial of degree at most two, and has at most two distinct eigenvalues.

Homework 12.2.2.7 Consider $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$

- The eigenvalue *largest in magnitude* is
- Which of the following are eigenvectors associated with this largest eigenvalue (in magnitude):
  - $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$
  - $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
  - $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$
  - $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$
- The eigenvalue *smallest in magnitude* is
- Which of the following are eigenvectors associated with this largest eigenvalue (in magnitude):
  - $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$
  - $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
  - $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$
  - $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$

Homework 12.2.2.8 Consider $A = \begin{pmatrix} -3 & -4 \\ 5 & 6 \end{pmatrix}$

- The eigenvalue *largest in magnitude* is
- The eigenvalue *smallest in magnitude* is

Example 12.1 Consider the matrix $A = \begin{pmatrix} 3 & -1 \\ 2 & 1 \end{pmatrix}$. To find the eigenvalues and eigenvectors of this matrix, we form $A - \lambda I = \begin{pmatrix} 3-\lambda & -1 \\ 2 & 1-\lambda \end{pmatrix}$ and check when the characteristic polynomial is equal to zero:

$$\det\left(\begin{pmatrix} 3-\lambda & -1 \\ 2 & 1-\lambda \end{pmatrix}\right) = (3-\lambda)(1-\lambda) - (-1)(2) = \lambda^2 - 4\lambda + 5.$$

When is this equal to zero? We will use the quadratic formula:

$$\lambda = \frac{-(-4) \pm \sqrt{(-4)^2 - 4(5)}}{2} = 2 \pm i.$$

Thus, this matrix has complex valued eigenvalues in form of a conjugate pair: $\lambda_0 = 2+i$ and $\lambda_1 = 2-i$. To find the corresponding eigenvectors:

$\lambda_0 = 2+i$:

$$\begin{aligned} A - \lambda_0 I &= \begin{pmatrix} 3-(2+i) & -1 \\ 2 & 1-(2+i) \end{pmatrix} \\ &= \begin{pmatrix} 1-i & -1 \\ 2 & -1-i \end{pmatrix}. \end{aligned}$$

Find a nonzero vector in the null space:

$$\begin{pmatrix} 1-i & -1 \\ 2 & -1-i \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

By examination,

$$\begin{pmatrix} 1-i & -1 \\ 2 & -1-i \end{pmatrix} \begin{pmatrix} 1 \\ 1-i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Eigenpair: $(2+i, \begin{pmatrix} 1 \\ 1-i \end{pmatrix})$.

$\lambda_0 = 2-i$:

$$\begin{aligned} A - \lambda_1 I &= \begin{pmatrix} 3-(2-i) & -1 \\ 2 & 1-(2-i) \end{pmatrix} \\ &= \begin{pmatrix} 1+i & -1 \\ 2 & -1+i \end{pmatrix}. \end{aligned}$$

Find a nonzero vector in the null space:

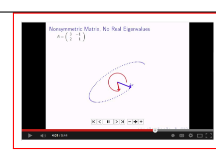
$$\begin{pmatrix} 1+i & -1 \\ 2 & -1+i \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

By examination,

$$\begin{pmatrix} 1+i & -1 \\ 2 & -1+i \end{pmatrix} \begin{pmatrix} 1 \\ 1+i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Eigenpair: $(2-i, \begin{pmatrix} 1 \\ 1+i \end{pmatrix})$.

If A is real valued, then its characteristic polynomial has real valued coefficients. However, a polynomial with real valued coefficients may still have complex valued roots. Thus, the eigenvalues of a real valued matrix may be complex.



☛ **View at edX**

Homework 12.2.2.9 Consider $A = \begin{pmatrix} 2 & 2 \\ -1 & 4 \end{pmatrix}$. Which of the following are the eigenvalues of A:

- 4 and 2.
- $3+i$ and 2.
- $3+i$ and $3-i$.
- $2+i$ and $2-i$.

12.2.3 Diagonalizing

☛ View at edX

Diagonalizing a square matrix $A \in \mathbb{R}^{n\times n}$ is closely related to the problem of finding the eigenvalues and eigenvectors of a matrix. In this unit, we illustrate this for some simple 2×2 examples. A more thorough treatment then follows when we talk about the eigenvalues and eigenvectors of $n\times n$ matrix, later this week.

In the last unit, we found eigenpairs for the matrix

$$\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}.$$

Specifically,

$$\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 2\begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix} = 3\begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

so that eigenpairs are given by

$$(2, \begin{pmatrix} -1 \\ 1 \end{pmatrix}) \quad \text{and} \quad 3\begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

Now, let's put our understanding of matrix-matrix multiplication from Weeks 4 and 5 to good use:

	Comment (A here is 2×2)
$\underbrace{\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 2\begin{pmatrix} -1 \\ 1 \end{pmatrix}; \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix} = 3\begin{pmatrix} -1 \\ 2 \end{pmatrix}}$	$Ax_0 = \lambda_0 x_0; Ax_1 = \lambda_1 x_1$
$\underbrace{\left(\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} \middle\vert \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix}\right) = \left(2\begin{pmatrix} -1 \\ 1 \end{pmatrix} \middle\vert 3\begin{pmatrix} -1 \\ 2 \end{pmatrix}\right)}$	$\left(\, Ax_0 \mid Ax_1 \,\right) = \left(\, \lambda_0 x_0 \mid \lambda_1 x_1 \,\right)$
$\underbrace{\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}}_{A} \underbrace{\left(\begin{array}{c\|c} -1 & -1 \\ 1 & 2 \end{array}\right)}_{X} = \underbrace{\left(\begin{array}{c\|c} -1 & -1 \\ 1 & 2 \end{array}\right)}_{X} \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}}_{\Lambda}$	$A \underbrace{\left(\, x_0 \mid x_1 \,\right)}_{X} = \underbrace{\left(\, x_0 \mid x_1 \,\right)}_{X} \underbrace{\begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix}}_{\Lambda}$
$\underbrace{\left(\begin{array}{c\|c} -1 & -1 \\ 1 & 2 \end{array}\right)^{-1}}_{X^{-1}} \underbrace{\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}}_{A} \underbrace{\left(\begin{array}{c\|c} -1 & -1 \\ 1 & 2 \end{array}\right)}_{X} = \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}}_{\Lambda}$	$\underbrace{\left(\, x_0 \mid x_1 \,\right)^{-1}}_{X^{-1}} A \underbrace{\left(\, x_0 \mid x_1 \,\right)}_{X} = \underbrace{\begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix}}_{\Lambda}$

What we notice is that if we take the two eigenvectors of matrix A, and create with them a matrix X that has those eigenvectors as its columns, then $X^{-1}AX = \Lambda$, where Λ is a diagonal matrix with the eigenvalues on its diagonal. The matrix X is said to *diagonalize* matrix A.

Defective matrices

Now, it is *not* the case that for every $A \in \mathbb{R}^{n\times n}$ there is a nonsingular matrix $X \in \mathbb{R}^{n\times n}$ such that $X^{-1}AX = \Lambda$, where Λ is diagonal. Matrices for which such a matrix X does not exists are called *defective* matrices.

Homework 12.2.3.1 The matrix

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

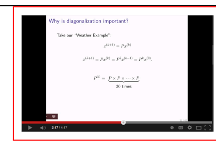
can be diagonalized.

True/False

The matrix

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

is a simple example of what is often called a *Jordan block*. It, too, is defective.



☛ **View at edX**

Homework 12.2.3.2 In Homework 12.2.2.7 you considered the matrix

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

and computed the eigenpairs

$$(4, \begin{pmatrix} 1 \\ 1 \end{pmatrix}) \quad \text{and} \quad (-2, \begin{pmatrix} 1 \\ -1 \end{pmatrix}).$$

- Matrix A can be diagonalized by matrix $X =$. (Yes, this matrix is not unique, so please use the info from the eigenpairs, in order...)
- $AX =$
- $X^{-1} =$
- $X^{-1}AX =$

12.2.4 Eigenvalues and Eigenvectors of 3×3 Matrices

☛ **View at edX**

Homework 12.2.4.1 Let $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Then which of the following are true:

- $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector associated with eigenvalue 3.

  True/False

- $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ is an eigenvector associated with eigenvalue -1.

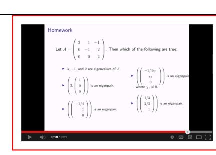
  True/False

- $\begin{pmatrix} 0 \\ \chi_1 \\ 0 \end{pmatrix}$, where $\chi_1 \neq 0$ is a scalar, is an eigenvector associated with eigenvalue -1.

  True/False

- $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is an eigenvector associated with eigenvalue 2.

  True/False



☛ View at edX

Homework 12.2.4.2 Let $A = \begin{pmatrix} \alpha_{0,0} & 0 & 0 \\ 0 & \alpha_{1,1} & 0 \\ 0 & 0 & \alpha_{2,2} \end{pmatrix}$. Then which of the following are true:

- $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector associated with eigenvalue $\alpha_{0,0}$.

  True/False

- $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ is an eigenvector associated with eigenvalue $\alpha_{1,1}$.

  True/False

- $\begin{pmatrix} 0 \\ \chi_1 \\ 0 \end{pmatrix}$ where $\chi_1 \neq 0$ is an eigenvector associated with eigenvalue $\alpha_{1,1}$.

  True/False

- $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is an eigenvector associated with eigenvalue $\alpha_{2,2}$.

  True/False

Homework 12.2.4.3 Let $A = \begin{pmatrix} 3 & 1 & -1 \\ 0 & -1 & 2 \\ 0 & 0 & 2 \end{pmatrix}$. Then which of the following are true:

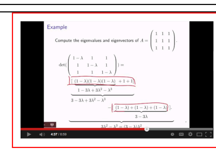
- 3, -1, and 2 are eigenvalues of A.
- $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector associated with eigenvalue 3.

  True/False
- $\begin{pmatrix} -1/4 \\ 1 \\ 0 \end{pmatrix}$ is an eigenvector associated with eigenvalue -1.

  True/False
- $\begin{pmatrix} -1/4\chi_1 \\ \chi_1 \\ 0 \end{pmatrix}$ where $\chi_1 \neq 0$ is an eigenvector associated with eigenvalue -1.

  True/False
- $\begin{pmatrix} 1/3 \\ 2/3 \\ 1 \end{pmatrix}$ is an eigenvector associated with eigenvalue 2.

  True/False



☛ **View at edX**

Homework 12.2.4.4 Let $A = \begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} \\ 0 & \alpha_{1,1} & \alpha_{1,2} \\ 0 & 0 & \alpha_{2,2} \end{pmatrix}$. Then the eigenvalues of this matrix are $\alpha_{0,0}$, $\alpha_{1,1}$, and $\alpha_{2,2}$.

True/False

When we discussed how to find the eigenvalues of a 2×2 matrix, we saw that it all came down to the determinant of $A - \lambda I$, which then gave us the characteristic polynomial $p_2(\lambda)$. The roots of this polynomial were the eigenvalues of the matrix.

Similarly, there is a formula for the determinant of a 3×3 matrix:

$$\det\left(\begin{pmatrix} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} \end{pmatrix}\right) =$$

$$\underbrace{(\alpha_{0,0}\alpha_{1,1}\alpha_{2,2} + \alpha_{0,1}\alpha_{1,2}\alpha_{2,0} + \alpha_{0,2}\alpha_{1,0}\alpha_{2,1})}_{\begin{array}{ccc|cc} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,0} & \alpha_{0,1} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,0} & \alpha_{1,1} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,0} & \alpha_{2,1} \end{array}} - \underbrace{(\alpha_{2,0}\alpha_{1,1}\alpha_{0,2} + \alpha_{2,1}\alpha_{1,2}\alpha_{0,0} + \alpha_{2,2}\alpha_{1,0}\alpha_{0,1})}_{\begin{array}{ccc|cc} \alpha_{0,0} & \alpha_{0,1} & \alpha_{0,2} & \alpha_{0,0} & \alpha_{0,1} \\ \alpha_{1,0} & \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,0} & \alpha_{1,1} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,0} & \alpha_{2,1} \end{array}}.$$

Thus, for a 3×3 matrix, the characteristic polynomial becomes

$$p_3(\lambda) = \det\left(\begin{pmatrix} \alpha_{0,0}-\lambda & \alpha_{0,1} & \alpha_{0,2} \\ \alpha_{1,0} & \alpha_{1,1}-\lambda & \alpha_{1,2} \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2}-\lambda \end{pmatrix}\right) =$$

$$[(\alpha_{0,0}-\lambda)(\alpha_{1,1}-\lambda)(\alpha_{2,2}-\lambda) + \alpha_{0,1}\alpha_{1,2}\alpha_{2,0} + \alpha_{0,2}\alpha_{1,0}\alpha_{2,1}]$$
$$-[\alpha_{2,0}(\alpha_{1,1}-\lambda)\alpha_{0,2} + \alpha_{2,1}\alpha_{1,2}(\alpha_{0,0}-\lambda) + (\alpha_{2,2}-\lambda)\alpha_{1,0}\alpha_{0,1}].$$

Multiplying this out, we get a third degree polynomial. The roots of this cubic polynomial are the eigenvalues of the 3×3 matrix. Hence, a 3×3 matrix has at most three distinct eigenvalues.

Example 12.2 Compute the eigenvalues and eigenvectors of $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

$$\det\left(\begin{pmatrix} 1-\lambda & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{pmatrix}\right) = \underbrace{\underbrace{[\underbrace{(1-\lambda)(1-\lambda)(1-\lambda)}_{1-3\lambda+3\lambda^2-\lambda^3} + 1 + 1]}_{3-3\lambda+3\lambda^2-\lambda^3} - [\underbrace{(1-\lambda)+(1-\lambda)+(1-\lambda)}_{3-3\lambda}]}_{3\lambda^2-\lambda^3 = (3-\lambda)\lambda^2}.$$

So, $\lambda = 0$ is a double root, while $\lambda = 3$ is the third root.

$\lambda_2 = 3$:	$\lambda_0 = \lambda_1 = 0$:
$A - \lambda_2 I = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$	$A - 0I = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$
We wish to find a nonzero vector in the null space: $\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}\begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.	Reducing this to row-echelon form gives us the matrix $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$,
By examination, I noticed that $\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.	for which we find vectors in the null space $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.
Eigenpair: $(3, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix})$.	Eigenpairs: $(0, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix})$ and $(0, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix})$

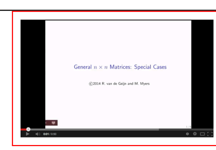
What is interesting about this last example is that $\lambda = 0$ is a double root and yields two linearly independent eigenvectors.

Homework 12.2.4.5 Consider $A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$. Which of the following is true about this matrix:

- $(1, \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix})$ is an eigenpair.
- $(0, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix})$ is an eigenpair.
- $(0, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix})$ is an eigenpair.
- This matrix is defective.

12.3 The General Case

12.3.1 Eigenvalues and Eigenvectors of $n \times n$ matrices: Special Cases

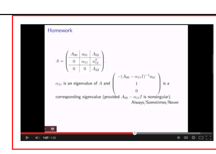


☛ **View at edX**

We are now ready to talk about eigenvalues and eigenvectors of arbitrary sized matrices.

Homework 12.3.1.1 Let $A \in \mathbb{R}^{n \times n}$ be a diagonal matrix: $A = \begin{pmatrix} \alpha_{0,0} & 0 & 0 & \cdots & 0 \\ 0 & \alpha_{1,1} & 0 & \cdots & 0 \\ 0 & 0 & \alpha_{2,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_{n-1,n-1} \end{pmatrix}$. Then e_i is an eigenvector associated with eigenvalue $\alpha_{i,i}$.

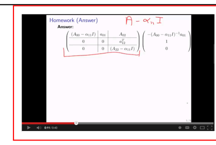
True/False



☛ **View at edX**

Homework 12.3.1.2 Let $A = \left(\begin{array}{c|c|c} A_{00} & a_{01} & A_{02} \\ \hline 0 & \alpha_{11} & a_{12}^T \\ \hline 0 & 0 & A_{22} \end{array}\right)$, where A_{00} is square. Then α_{11} is an eigenvalue of A and $\left(\begin{array}{c} -(A_{00}-\alpha_{11}I)^{-1}a_{01} \\ 1 \\ 0 \end{array}\right)$ is a corresponding eigenvalue (provided $A_{00}-\alpha_{11}I$ is nonsingular).

True/False

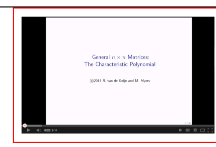


☛ **View at edX**

Homework 12.3.1.3 The eigenvalues of a triangular matrix can be found on its diagonal.

True/False

12.3.2 Eigenvalues of $n \times n$ Matrices



☛ **View at edX**

There is a formula for the determinant of a $n \times n$ matrix, which is a "inductively defined function", meaning that the formula for the determinant of an $n \times n$ matrix is defined in terms of the determinant of an $(n-1) \times (n-1)$ matrix. Other than as a theoretical tool, the determinant of a general $n \times n$ matrix is not particularly useful. We restrict our discussion to some facts and observations about the determinant that impact the characteristic polynomial, which is the polynomial that results when one computes the determinant of the matrix $A - \lambda I$, $\det(A - \lambda I)$.

Theorem 12.3 *A matrix* $A \in \mathbb{R}^{n \times n}$ *is nonsingular if and only if* $\det(A) \neq 0$.

Theorem 12.4 *Given* $A \in \mathbb{R}^{n \times n}$,

$$p_n(\lambda) = \det(A - \lambda I) = \lambda^n + \gamma_{n-1}\lambda^{n-1} + \cdots + \gamma_1\lambda + \gamma_0.$$

for some coefficients $\gamma_1, \ldots, \gamma_{n-1} \in \mathbb{R}$.

Since we don't give the definition of a determinant, we do not prove the above theorems.

Definition 12.5 *Given* $A \in \mathbb{R}^{n \times n}$, $p_n(\lambda) = \det(A - \lambda I)$ *is called the characteristic polynomial.*

Theorem 12.6 *Scalar* λ *satisfies* $Ax = \lambda x$ *for some nonzero vector* x *if and only if* $\det(A - \lambda I) = 0$.

Proof: This is an immediate consequence of the fact that $Ax = \lambda x$ is equivalent to $(A - \lambda I)x =$ and the fact that $A - \lambda I$ is singular (has a nontrivial null space) if and only if $\det(A - \lambda I) = 0$.

Roots of the characteristic polynomial

Since an eigenvalue of A is a root of $p_n(A) = \det(A - \lambda I)$ and vise versa, we can exploit what we know about roots of nth degree polynomials. Let us review, relating what we know to the eigenvalues of A.

- The characteristic polynomial of $A \in \mathbb{R}^{n \times n}$ is given by $p_n(\lambda) = \det(A - \lambda I) = \gamma_0 + \gamma_1 \lambda + \cdots + \gamma_{n-1} \lambda^{n-1} + \lambda^n$
- Since $p_n(\lambda)$ is an nth degree polynomial, it has n roots, counting multiplicity. Thus, matrix A has n eigenvalues, counting multiplicity.
  - Let k equal the number of *distinct* roots of $p_n(\lambda)$. Clearly, $k \leq n$. Clearly, matrix A then has k distinct eigenvalues.
  - The set of all roots of $p_n(\lambda)$, which is the set of all eigenvalues of A, is denoted by $\Lambda(A)$ and is called the *spectrum* of matrix A.
  - The characteristic polynomial can be factored as $p_n(\lambda) = \det(A - \lambda I) = (\lambda - \lambda_0)^{n_0} (\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_{k-1})^{n_{k-1}}$, where $n_0 + n_1 + \cdots + n_{k-1} = n$ and n_j is the root λ_j, which is known as the (algebraic) multiplicity of eigenvalue λ_j.
- If $A \in \mathbb{R}^{n \times n}$, then the coefficients of the characteristic polynomial are real ($\gamma_0, \ldots, \gamma_{n-1} \in \mathbb{R}$), but
  - Some or all of the roots/eigenvalues may be complex valued and
  - Complex roots/eigenvalues come in "conjugate pairs": If $\lambda = \mathcal{R}\rceil(\lambda) + iI\Updownarrow(\lambda)$ is a root/eigenvalue, so is $\lambda = \mathcal{R}\rceil(\lambda) - iI\Updownarrow(\lambda)$

An inconvenient truth

Galois theory tells us that for $n \geq 5$, roots of arbitrary $p_n(\lambda)$ cannot be found in a finite number of computations.

Since we did not tell you how to compute the determinant of $A - \lambda I$, you will have to take the following for granted: For every $n$the degree polynomial

$$p_n(\lambda) = \gamma_0 + \gamma_1 \lambda + \cdots + \gamma_{n-1} \lambda^{n-1} + \lambda^n,$$

there exists a matrix, C, called the companion matrix that has the property that

$$p_n(\lambda) = \det(C - \lambda I) = \gamma_0 + \gamma_1 \lambda + \cdots + \gamma_{n-1} \lambda^{n-1} + \lambda^n.$$

In particular, the matrix

$$C = \begin{pmatrix} -\gamma_{n-1} & -\gamma_{n-2} & \cdots & -\gamma_1 & -\gamma_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

is the companion matrix for $p_n(\lambda)$:

$$p_n(\lambda) = \gamma_0 + \gamma_1 \lambda + \cdots + \gamma_{n-1} \lambda^{n-1} + \lambda^n = \det(\begin{pmatrix} -\gamma_{n-1} & -\gamma_{n-2} & \cdots & -\gamma_1 & -\gamma_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} - \lambda I).$$

Homework 12.3.2.1 If $A \in \mathbb{R}^{n \times n}$, then $\Lambda(A)$ has n distinct elements.

True/False

Homework 12.3.2.2 Let $A \in \mathbb{R}^{n \times n}$ and $\lambda \in \Lambda(A)$. Let S be the set of all vectors that satisfy $Ax = \lambda x$. (Notice that S is the set of all eigenvectors corresponding to λ *plus* the zero vector.) Then S is a subspace.

True/False

12.3.3 Diagonalizing, Again

☛ **View at edX**

We now revisit the topic of diagonalizing a square matrix $A \in \mathbb{R}^{n\times n}$, but for general n rather than the special case of $n = 2$ treated in Unit 12.2.3.

Let us start by assuming that matrix $A \in \mathbb{R}^{n\times n}$ has n eigenvalues, $\lambda_0, \ldots, \lambda_{n-1}$, where we simply repeat eigenvalues that have algebraic multiplicity greater than one. Let us also assume that x_j equals the eigenvector associated with eigenvalue λ_j and, importantly, that $x_0, \ldots, x_{n-1}$ are linearly independent. Below, we generalize the example from Unit 12.2.3.

$$
\begin{array}{l}
Ax_0 = \lambda_0 x_0; Ax_1 = \lambda_1 x_1; \cdots ; Ax_{n-1} = \lambda_{n-1} x_{n-1} \\
\qquad \text{if and only if} \qquad < \text{two matrices are equal if their columns are equal} > \\
\left(\begin{array}{c|c|c|c} Ax_0 & Ax_1 & \cdots & Ax_{n-1} \end{array} \right) = \left(\begin{array}{c|c|c|c} \lambda_0 x_0 & \lambda_1 x_1 & \cdots & \lambda_{n-1} x_{n-1} \end{array} \right) \\
\qquad \text{if and only if} \qquad < \text{partitioned matrix-matrix multiplication} > \\
A \left(\begin{array}{c|c|c|c} x_0 & x_1 & \cdots & x_{n-1} \end{array} \right) = \left(\begin{array}{c|c|c|c} \lambda_0 x_0 & \lambda_1 x_1 & \cdots & \lambda_{n-1} x_{n-1} \end{array} \right) \\
\qquad \text{if and only if} \qquad < \text{multiplication on the right by a diagonal matrix} > \\
A \left(\begin{array}{c|c|c|c} x_0 & x_1 & \cdots & x_{n-1} \end{array} \right) = \left(\begin{array}{c|c|c|c} x_0 & x_1 & \cdots & x_{n-1} \end{array} \right) \left(\begin{array}{cccc} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} \end{array} \right) \\
\qquad \text{if and only if} \qquad < \text{multiplication on the right by a diagonal matrix} > \\
AX = X\Lambda \text{ where } X = \left(\begin{array}{c|c|c|c} x_0 & x_1 & \cdots & x_{n-1} \end{array} \right) \text{ and } \Lambda = \left(\begin{array}{cccc} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} \end{array} \right) \\
\qquad \text{if and only if} \qquad < \text{columns of } X \text{ are linearly independent} > \\
X^{-1}AX = \Lambda \text{ where } X = \left(\begin{array}{c|c|c|c} x_0 & x_1 & \cdots & x_{n-1} \end{array} \right) \text{ and } \Lambda = \left(\begin{array}{cccc} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} \end{array} \right)
\end{array}
$$

The above argument motivates the following theorem:

Theorem 12.7 *Let $A \in \mathbb{R}^{n\times n}$. Then there exists a nonsingular matrix X such that $X^{-1}AX = \Lambda$ if and only if A has n linearly independent eigenvectors.*

If X is invertible (nonsingular, has linearly independent columns, etc.), then the following are equivalent

$$
\begin{array}{rcl}
X^{-1}\,A\,X & = & \Lambda \\
A\,X & = & X\,\Lambda \\
A & = & X\,\Lambda\,X^{-1}
\end{array}
$$

If Λ is in addition diagonal, then the diagonal elements of Λ are eigenvectors of A and the columns of X are eigenvectors of A.

Recognize that $\Lambda(A)$ denotes the spectrum (set of all eigenvalues) of matrix A while here we use it to denote the matrix Λ, which has those eigenvalues on its diagonal. This possibly confusing use of the same symbol for two different but related things is commonly encountered in the linear algebra literature. For this reason, you might as well get use to it!

Defective (deficient) matrices

We already saw in Unit 12.2.3, that it is *not* the case that for every $A \in \mathbb{R}^{n\times n}$ there is a nonsingular matrix $X \in \mathbb{R}^{n\times n}$ such that $X^{-1}AX = \Lambda$, where Λ is diagonal. In that unit, a 2×2 example was given that did not have two linearly independent eigenvectors.

In general, the $k \times k$ matrix $J_k(\lambda)$ given by

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

has eigenvalue λ of algebraic multiplicity k, but *geometric multiplicity* one (it has only one linearly independent eigenvector). Such a matrix is known as a Jordan block.

Definition 12.8 *The geometric multiplicity of an eigenvalue λ equals the number of linearly independent eigenvectors that are associated with λ.*

The following theorem has theoretical significance, but little practical significance (which is why we do not dwell on it):

Theorem 12.9 *Let $A \in \mathbb{R}^{n\times n}$. Then there exists a nonsingular matrix $X \in \mathbb{R}^{n\times n}$ such that $A = XJX^{-1}$, where*

$$J = \begin{pmatrix} J_{k_0}(\lambda_0) & 0 & 0 & \cdots & 0 \\ 0 & J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & 0 & J_{k_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & J_{k_{m-1}}(\lambda_{m-1}) \end{pmatrix}$$

where each $J_{k_j}(\lambda_j)$ is a Jordan block of size $k_j \times k_j$.

The factorization $A = XJX^{-1}$ is known as the Jordan Canonical Form of matrix A.

A few comments are in order:

- It is *not* the case that $\lambda_0, \lambda_1, \ldots, \lambda_{m-1}$ are distinct. If λ_j appears in multiple Jordan blocks, the number of Jordan blocks in which λ_j appears equals the geometric multiplicity of λ_j (and the number of linearly independent eigenvectors associated with λ_j).
- The sum of the sizes of the blocks in which λ_j as an eigenvalue appears equals the algebraic multiplicity of λ_j.
- If each Jordan block is 1×1, then the matrix is diagonalized by matrix X.
- If any of the blocks is not 1×1, then the matrix cannot be diagonalized.

Homework 12.3.3.1 Consider $A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$.

- The algebraic multiplicity of $\lambda = 2$ is
- The geometric multiplicity of $\lambda = 2$ is
- The following vectors are linearly independent eigenvectors associated with $\lambda = 2$:

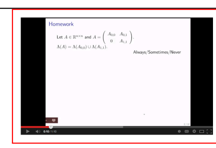
$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

True/False

Homework 12.3.3.2 Let $A \in \mathbb{A}^{n \times n}$, $\lambda \in \Lambda(A)$, and S be the set of all vectors x such that $Ax = \lambda x$. Finally, let λ have algebraic multiplicity k (meaning that it is a root of multiplicity k of the characteristic polynomial).
The dimension of S is k ($\dim(S) = k$).

Always/Sometimes/Never

12.3.4 Properties of Eigenvalues and Eigenvectors



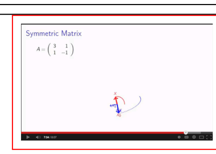
☛ View at edX

In this unit, we look at a few theoretical results related to eigenvalues and eigenvectors.

Homework 12.3.4.1 Let $A \in \mathbb{R}^{n \times n}$ and $A = \begin{pmatrix} A_{0,0} & A_{0,1} \\ 0 & A_{1,1} \end{pmatrix}$, where $A_{0,0}$ and $A_{1,1}$ are square matrices.

$\Lambda(A) = \Lambda(A_{0,0}) \cup \Lambda(A_{1,1})$.

Always/Sometimes/Never



☛ View at edX

The last exercise motives the following theorem (which we will not prove):

Theorem 12.10 *Let $A \in \mathbb{R}^{n \times n}$ and*

$$A = \begin{pmatrix} A_{0,0} & A_{0,1} & \cdots & A_{0,N-1} \\ 0 & A_{1,1} & \cdots & A_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{N-1,N-1} \end{pmatrix}$$

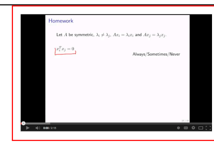
where all $A_{i,i}$ are a square matrices. Then $\Lambda(A) = \Lambda(A_{0,0}) \cup \Lambda(A_{1,1}) \cup \cdots \cup \Lambda(A_{N-1,N-1})$.

Homework 12.3.4.2 Let $A \in \mathbb{R}^{n \times n}$ be symmetric, $\lambda_i \neq \lambda_j$, $Ax_i = \lambda_i x_i$ and $Ax_j = \lambda_j x_j$.
$x_i^T x_j = 0$

Always/Sometimes/Never

The following theorem requires us to remember more about complex arithmetic than we have time to remember. For this reason, we will just state it:

Theorem 12.11 *Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then its eigenvalues are real valued.*



☛ **View at edX**

Homework 12.3.4.3 If $Ax = \lambda x$ then $AAx = \lambda^2 x$. (AA is often written as A^2.)

Always/Sometimes/Never

Homework 12.3.4.4 Let $Ax = \lambda x$ and $k \geq 1$. Recall that $A^k = \underbrace{AA \cdots A}_{k \text{ times}}$.

$A^k x = \lambda^k x$.

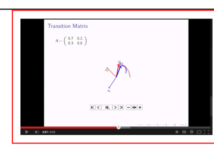
Always/Sometimes/Never

Homework 12.3.4.5 $A \in \mathbb{R}^{n \times n}$ is nonsingular if and only if $0 \notin \Lambda(A)$.

True/False

12.4 Practical Methods for Computing Eigenvectors and Eigenvalues

12.4.1 Predicting the Weather, One Last Time



☛ **View at edX**

If you think back about how we computed the probabilities of different types of weather for day k, recall that

$$x^{(k+1)} = Px^{(k)}$$

where $x^{(k)}$ is a vector with three components and P is a 3×3 matrix. We also showed that

$$x^{(k)} = P^k x^{(0)}.$$

We noticed that eventually

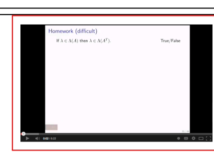
$$x^{(k+1)} \approx Px^{(k)}$$

and that therefore, eventually, $x^{(k+1)}$ came arbitrarily close to an eigenvector, x, associated with the eigenvalue 1 of matrix P:

$$Px = x.$$

Homework 12.4.1.1 If $\lambda \in \Lambda(A)$ then $\lambda \in \Lambda(A^T)$.

True/False



☛ **View at edX**

Homework 12.4.1.2 $\lambda \in \Lambda(A)$ if and only if $\lambda \in \Lambda(A^T)$.

True/False

Ah! It seems like we may have stumbled upon a possible method for computing an eigenvector for this matrix:

- Start with a first guess $x^{(0)}$.
- **for** $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
  - $x^{(k+1)} := Px^{(k)}$.

Can we use what we have learned about eigenvalues and eigenvectors to explain this? In the video, we give one explanation. Below we give an alternative explanation that uses diagonalization.

Let's assume that P is diagonalizable:

$$P = V \Lambda V^{-1}, \quad \text{where } \Lambda = \begin{pmatrix} \lambda_0 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}.$$

Here we use the letter V rather than X since we already use $x^{(k)}$ in a different way.

Then we saw before that

$$\begin{aligned} x^{(k)} = P^k x^{(0)} &= (V\Lambda V^{-1})^k x^{(0)} = V \Lambda^k V^{-1} x^{(0)} \\ &= V \begin{pmatrix} \lambda_0 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}^k V^{-1} x^{(0)} \\ &= V \begin{pmatrix} \lambda_0^k & 0 & 0 \\ 0 & \lambda_1^k & 0 \\ 0 & 0 & \lambda_2^k \end{pmatrix} V^{-1} x^{(0)}. \end{aligned}$$

Now, let's assume that $\lambda_0 = 1$ (since we noticed that P has one as an eigenvalue), and that $|\lambda_1| < 1$ and $|\lambda_2| < 1$. Also, notice that $V = \begin{pmatrix} v_0 & v_1 & v_2 \end{pmatrix}$

where v_i equals the eigenvector associated with λ_i. Finally, notice that V has linearly independent columns and that therefore there exists a vector w such that $Vw = x^{(0)}$.

Then

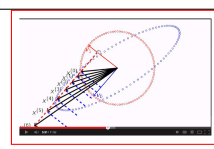
$$\begin{aligned} x^{(k)} &= V \begin{pmatrix} \lambda_0^k & 0 & 0 \\ 0 & \lambda_1^k & 0 \\ 0 & 0 & \lambda_2^k \end{pmatrix} V^{-1} x^{(0)} \\ &= V \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda_1^k & 0 \\ 0 & 0 & \lambda_2^k \end{pmatrix} V^{-1} V w \\ &= \begin{pmatrix} v_0 & v_1 & v_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda_1^k & 0 \\ 0 & 0 & \lambda_2^k \end{pmatrix} w \\ &= \begin{pmatrix} v_0 & v_1 & v_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda_1^k & 0 \\ 0 & 0 & \lambda_2^k \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \end{pmatrix}. \end{aligned}$$

Now, what if k gets very large? We know that $\lim_{k\to\infty}\lambda_1^k = 0$, since $|\lambda_1| < 1$. Similarly, $\lim_{k\to\infty}\lambda_2^k = 0$. So,

$$\begin{aligned}
\lim_{k\to\infty} x^{(k)} &= \lim_{k\to\infty}\left[\begin{pmatrix} v_0 & v_1 & v_2 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda_1^k & 0 \\ 0 & 0 & \lambda_2^k \end{pmatrix}\begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \end{pmatrix}\right] \\
&= \begin{pmatrix} v_0 & v_1 & v_2 \end{pmatrix}\lim_{k\to\infty}\left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda_1^k & 0 \\ 0 & 0 & \lambda_2^k \end{pmatrix}\right]\begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \end{pmatrix} \\
&= \begin{pmatrix} v_0 & v_1 & v_2 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \lim_{k\to\infty}\lambda_1^k & 0 \\ 0 & 0 & \lim_{k\to\infty}\lambda_2^k \end{pmatrix}\begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \end{pmatrix} \\
&= \begin{pmatrix} v_0 & v_1 & v_2 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \end{pmatrix} \\
&= \begin{pmatrix} v_0 & v_1 & v_2 \end{pmatrix}\begin{pmatrix} \omega_0 \\ 0 \\ 0 \end{pmatrix} = \omega_0 v_0.
\end{aligned}$$

Ah, so $x^{(k)}$ eventually becomes arbitrarily close (converges) to a multiple of the eigenvector associated with the eigenvalue 1 (provided $\omega_0 \neq 0$).

12.4.2 The Power Method



☛ **View at edX**

So, a question is whether the method we described in the last unit can be used in general. The answer is yes. The resulting method is known as the Power Method.

First, let's make some assumptions. Given $A \in \mathbb{R}^{n\times n}$,

- Let $\lambda_0, \lambda_1, \ldots, \lambda_{n-1} \in \Lambda(A)$. We list eigenvalues that have algebraic multiplicity k multiple (k) times in this list.
- Let us assume that $|\lambda_0| > |\lambda_1| \geq |\lambda_2| \geq \cdots \geq \lambda_{n-1}$. This implies that λ_0 is real, since complex eigenvalues come in conjugate pairs and hence there would have been two eigenvalues with equal greatest magnitude. It also means that there is a real valued eigenvector associated with λ_0.
- Let us assume that $A \in \mathbb{R}^{n\times n}$ is diagonalizable so that

$$A = V\Lambda V^{-1} = \begin{pmatrix} v_0 & v_1 & \cdots & v_{n-1} \end{pmatrix}\begin{pmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} \end{pmatrix}\begin{pmatrix} v_0 & v_1 & \cdots & v_{n-1} \end{pmatrix}^{-1}.$$

This means that v_i is an eigenvector associated with λ_i.

These assumptions set the stage.

Now, we start with some vector $x^{(0)} \in \mathbb{R}^n$. Since V is nonsingular, the vectors $v_0, \ldots, v_{n-1}$ form a linearly independent bases for $\mathbb{R}^n$. Hence,

$$x^{(0)} = \gamma_0 v_0 + \gamma_1 v_1 + \cdots + \gamma_{n-1} v_{n-1} = \left(\begin{array}{cccc} v_0 & v_1 & \cdots & v_{n-1} \end{array} \right) \left(\begin{array}{c} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-1} \end{array} \right) = Vc.$$

Now, we generate

$$\begin{array}{rcl} x^{(1)} & = & Ax^{(0)} \\ x^{(2)} & = & Ax^{(1)} \\ x^{(3)} & = & Ax^{(2)} \\ & \vdots & \end{array}$$

The following algorithm accomplishes this

for $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
 $x^{(k+1)} := Ax^{(k)}$
endfor

Notice that then

$$x^{(k)} = Ax^{(k-1)} = A^2 x^{(k-2)} = \cdots = A^k x^{(0)}.$$

But then

$$\begin{array}{rcl} A^k x^{(0)} & = & A^k (\underbrace{\gamma_0 v_0 + \gamma_1 v_1 + \cdots + \gamma_{n-1} v_{n-1}}_{Vc}) \\ & = & A^k \gamma_0 v_0 + A^k \gamma_1 v_1 + \cdots + A^k \gamma_{n-1} v_{n-1} \\ & = & \gamma_0 A^k v_0 + \gamma_1 A^k v_1 + \cdots + \gamma_{n-1} A^k v_{n-1} \\ & = & \underbrace{\gamma_0 \lambda_0^k v_0 + \gamma_1 \lambda_1^k v_1 + \cdots + \gamma_{n-1} \lambda_{n-1}^k v_{n-1}}_{\underbrace{\left(\begin{array}{cccc} v_0 & v_1 & \cdots & v_{n-1} \end{array} \right) \left(\begin{array}{cccc} \lambda_0^k & 0 & \cdots & 0 \\ 0 & \lambda_1^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1}^k \end{array} \right) \left(\begin{array}{c} \gamma_0 \\ \gamma_1 \\ \cdots \\ \gamma_{n-1} \end{array} \right)}_{V\Lambda^k c}} \end{array}$$

Now, $\boxed{\textbf{if}}$ $\lambda_0 = 1$, then $|\lambda_j| < 1$ for $j > 0$ and hence

$$\lim_{k\to\infty} x^{(k)} = \underbrace{\lim_{k\to\infty}\left(\gamma_0 v_0 + \gamma_1 \lambda_1^k v_1 + \cdots + \gamma_{n-1}\lambda_{n-1}^k v_{n-1}\right)}_{\underbrace{\lim_{k\to\infty}\left(\begin{array}{cccc} v_0 & v_1 & \cdots & v_{n-1}\end{array}\right)\left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & \lambda_1^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1}^k \end{array}\right)\left(\begin{array}{c} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-1}\end{array}\right)}_{\underbrace{\left(\begin{array}{cccc} v_0 & v_1 & \cdots & v_{n-1}\end{array}\right)\left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right)\left(\begin{array}{c} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-1}\end{array}\right)}_{\underbrace{\left(\begin{array}{cccc} v_0 & 0 & \cdots & 0\end{array}\right)\left(\begin{array}{c} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-1}\end{array}\right)}_{\gamma_0 v_0}}}} = \gamma_0 v_0$$

which means that $x^{(k)}$ eventually starts pointing towards the direction of v_0, the eigenvector associated with the eigenvalue that is largest in magnitude. (Well, as long as $\gamma_0 \neq 0$.)

Homework 12.4.2.1 Let $A \in \mathbb{R}^{n\times n}$ and $\mu \neq 0$ be a scalar. Then $\lambda \in \Lambda(A)$ if and only if $\lambda/\mu \in \Lambda(\frac{1}{\mu}A)$.
True/False

What this last exercise shows is that if $\lambda_0 \neq 1$, then we can instead iterate with the matrix $\frac{1}{\lambda_0}A$, in which case

$$1 = \frac{\lambda_0}{\lambda_0} > \left|\frac{\lambda_1}{\lambda_0}\right| \geq \cdots \geq \left|\frac{\lambda_{n-1}}{\lambda_0}\right|.$$

The iteration then becomes

$$\begin{array}{rcl} x^{(1)} & = & \frac{1}{\lambda_0} A x^{(0)} \\ x^{(2)} & = & \frac{1}{\lambda_0} A x^{(1)} \\ x^{(3)} & = & \frac{1}{\lambda_0} A x^{(2)} \\ & \vdots & \end{array}$$

The following algorithm accomplishes this

for $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
$x^{(k+1)} := Ax^{(k)}/\lambda_0$
endfor

It is not hard to see that then

$$\lim_{k\to\infty} x^{(k)} = \underbrace{\underbrace{\underbrace{\underbrace{\lim_{k\to\infty}\left(\gamma_0\left(\frac{\lambda_0}{\lambda_0}\right)^k v_0 + \gamma_1\left(\frac{\lambda_1}{\lambda_0}\right)^k v_1 + \cdots + \gamma_{n-1}\left(\frac{\lambda_{n-1}}{\lambda_0}\right)^k v_{n-1}\right)}_{\lim_{k\to\infty}\left(\begin{array}{cccc} v_0 & v_1 & \cdots & v_{n-1}\end{array}\right)\left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & (\lambda_1/\lambda_0)^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\lambda_{n-1}/\lambda_0)^k \end{array}\right)\left(\begin{array}{c}\gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-1}\end{array}\right)}}_{\left(\begin{array}{cccc} v_0 & v_1 & \cdots & v_{n-1}\end{array}\right)\left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right)\left(\begin{array}{c}\gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-1}\end{array}\right)}}_{\left(\begin{array}{cccc} v_0 & 0 & \cdots & 0\end{array}\right)\left(\begin{array}{c}\gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-1}\end{array}\right)}}_{\gamma_0 v_0} = \gamma_0 v_0$$

So, it seems that we have an algorithm that always works as long as

$$|\lambda_0| > |\lambda_1| \geq \cdots \geq |\lambda_{n-1}|.$$

Unfortunately, we are cheating... If we knew λ_0, then we could simply compute the eigenvector by finding a vector in the null space of $A - \lambda_0 I$. The key insight now is that, in $x^{(k+1)} = Ax^{(k)}/\lambda_0$, dividing by λ_0 is merely meant to keep the vector $x^{(k)}$ from getting progressively larger (if $|\lambda_0| > 1$) or smaller (if $|\lambda_0| < 1$). We can alternatively simply make $x^{(k)}$ of length one at each step, and that will have the same effect without requiring λ_0:

for $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
  $x^{(k+1)} := Ax^{(k)}$
  $x^{(k+1)} := x^{(k+1)}/\|x^{(k+1)}\|_2$
endfor

This last algorithm is known as the *Power Method* for finding an eigenvector associated with the largest eigenvalue (in magnitude).

Homework 12.4.2.2 We now walk you through a simple implementation of the Power Method, referring to files in directory `LAFF-2.0xM/Programming/Week12`.
We want to work with a matrix A for which we know the eigenvalues. Recall that a matrix A is diagonalizable if and only if there exists a nonsingular matrix V and diagonal matrix Λ such that $A = V\Lambda V^{-1}$. The diagonal elements of Λ then equal the eigenvalues of A and the columns of V the eigenvectors.
Thus, given eigenvalues, we can create a matrix A by creating a diagonal matrix with those eigenvalues on the diagonal and a random nonsingular matrix V, after which we can compute A to equal $V\Lambda V^{-1}$. This is accomplished by the function

[A, V] = CreateMatrixForEigenvalueProblem(eigs)

(see file `CreateMatrixForEigenvalueProblem.m`).
The script in `PowerMethodScript.m` then illustrates how the Power Method, starting with a random vector, computes an eigenvector corresponding to the eigenvalue that is largest in magnitude, and via the Rayleigh quotient (a way for computing an eigenvalue given an eigenvector that is discussed in the next unit) an approximation for that eigenvalue.
To try it out, in the Command Window type

```
>> PowerMethodScript
input a vector of eigenvalues. e.g.: [ 4; 3; 2; 1 ]
[ 4; 3; 2; 1 ]
```

The script for each step of the Power Method reports for the current iteration the length of the component orthogonal to the eigenvector associated with the eigenvalue that is largest in magnitude. If this component becomes small, then the vector lies approximately in the direction of the desired eigenvector. The Rayleigh quotient slowly starts to get close to the eigenvalue that is largest in magnitude. The slow convergence is because the ratio of the second to largest and the largest eigenvalue is not much smaller than 1.
Try some other distributions of eigenvalues. For example, `[ 4; 1; 0.5; 0.25]`, which should converge faster, or `[ 4; 3.9; 2; 1 ]`, which should converge much slower.
You may also want to try `PowerMethodScript2.m`, which illustrates what happens if there are two eigenvalues that are equal in value and both largest in magnitude (relative to the other eigenvalues).

12.4.3 In Preparation for this Week's Enrichment

In the last unit we introduce a practical method for computing an eigenvector associated with the largest eigenvalue in magnitude. This method is known as the Power Method. The next homework shows how to compute an eigenvalue associated with an eigenvector. Thus, the Power Method can be used to first approximate that eigenvector, and then the below result can be used to compute the associated eigenvalue.

Given $A \in \mathbb{R}^{n\times n}$ and nonzero vector $x \in \mathbb{R}^n$, the scalar $x^T Ax/x^T x$ is known as the *Rayleigh quotient*.

Homework 12.4.3.1 Let $A \in \mathbb{R}^{n\times n}$ and x equal an eigenvector of A. Assume that x is real valued as is the eigenvalue λ with $Ax = \lambda x$.
$\lambda = \frac{x^T Ax}{x^T x}$ is the eigenvalue associated with the eigenvector x.

Always/Sometimes/Never

Notice that we are carefully avoiding talking about complex valued eigenvectors. The above results can be modified for the case where x is an eigenvector associated with a complex eigenvalue and the case where A itself is complex valued. However, this goes beyond the scope of this course.

The following result allows the Power Method to be extended so that it can be used to compute the eigenvector associated with the smallest eigenvalue (in magnitude). The new method is called the Inverse Power Method and is discussed in this week's enrichment section.

Homework 12.4.3.2 Let $A \in \mathbb{R}^{n\times n}$ be nonsingular, $\lambda \in \Lambda(A)$, and $Ax = \lambda x$. Then $A^{-1}x = \frac{1}{\lambda}x$.

True/False

The Inverse Power Method can be accelerated by "shifting" the eigenvalues of the matrix, as discussed in this week's enrichment, yielding the Rayleigh Quotient Iteration. The following exercise prepares the way.

Homework 12.4.3.3 Let $A \in \mathbb{R}^{n \times n}$ and $\lambda \in \Lambda(A)$. Then $(\lambda - \mu) \in \Lambda(A - \mu I)$.

True/False

12.5 Enrichment

12.5.1 The Inverse Power Method

The Inverse Power Method exploits a property we established in Unit 12.3.4: If A is nonsingular and $\lambda \in \Lambda(A)$ then $1/\lambda \in \Lambda(A^{-1})$.

Again, let's make some assumptions. Given nonsingular $A \in \mathbb{R}^{n \times n}$,

- Let $\lambda_0, \lambda_1, \ldots, \lambda_{n-2}, \lambda_{n-1} \in \Lambda(A)$. We list eigenvalues that have algebraic multiplicity k multiple (k) times in this list.
- Let us assume that $|\lambda_0| \geq |\lambda_1| \geq \cdots \geq |\lambda_{n-2}| > |\lambda_{n-1}| > 0$. This implies that λ_{n-1} is real, since complex eigenvalues come in conjugate pairs and hence there would have been two eigenvalues with equal smallest magnitude. It also means that there is a real valued eigenvector associated with λ_{n-1}.
- Let us assume that $A \in \mathbb{R}^{n \times n}$ is diagonalizable so that

$$A = V \Lambda V^{-1} = \begin{pmatrix} v_0 & v_1 & \cdots & v_{n-2} & v_{n-1} \end{pmatrix} \begin{pmatrix} \lambda_0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-2} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_{n-1} \end{pmatrix} \begin{pmatrix} v_0 & v_1 & \cdots & v_{n-2} & v_{n-1} \end{pmatrix}^{-1}.$$

  This means that v_i is an eigenvector associated with λ_i.

These assumptions set the stage.

Now, we again start with some vector $x^{(0)} \in \mathbb{R}^n$. Since V is nonsingular, the vectors $v_0, \ldots, v_{n-1}$ form a linearly independent bases for $\mathbb{R}^n$. Hence,

$$x^{(0)} = \gamma_0 v_0 + \gamma_1 v_1 + \cdots + \gamma_{n-2} v_{n-2} + \gamma_{n-1} v_{n-1} = \begin{pmatrix} v_0 & v_1 & \cdots & v_{n-2} & v_{n-1} \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1} \end{pmatrix} = Vc.$$

Now, we generate

$$\begin{array}{rcl} x^{(1)} & = & A^{-1} x^{(0)} \\ x^{(2)} & = & A^{-1} x^{(1)} \\ x^{(3)} & = & A^{-1} x^{(2)} \\ & \vdots & \end{array}$$

The following algorithm accomplishes this

for $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
    Solve $Ax^{(k+1)} := x^{(k)}$
endfor

(In practice, one would probably factor A once, and reuse the factors for the solve.) Notice that then

$$x^{(k)} = A^{-1}x^{(k-1)} = (A^{-1})^2 x^{(k-2)} = \cdots = (A^{-1})^k x^{(0)}.$$

But then

$$\begin{aligned}
(A^{-1})^k x^{(0)} &= (A^{-1})^k (\underbrace{\gamma_0 v_0 + \gamma_1 v_1 + \cdots + \gamma_{n-2} v_{n-2} + \gamma_{n-1} v_{n-1}}_{Vc}) \\
&= (A^{-1})^k \gamma_0 v_0 + (A^{-1})^k \gamma_1 v_1 + \cdots + (A^{-1})^k \gamma_{n-2} v_{n-2} + (A^{-1})^k \gamma_{n-1} v_{n-1} \\
&= \gamma_0 (A^{-1})^k v_0 + \gamma_1 (A^{-1})^k v_1 + \cdots + \gamma_{n-2} (A^{-1})^k v_{n-2} + \gamma_{n-1} (A^{-1})^k v_{n-1} \\
&= \underbrace{\underbrace{\gamma_0 \left(\frac{1}{\lambda_0}\right)^k v_0 + \gamma_1 \left(\frac{1}{\lambda_1}\right)^k v_1 + \cdots + \gamma_{n-2} \left(\frac{1}{\lambda_{n-2}}\right)^k v_{n-2} + \gamma_{n-1} \left(\frac{1}{\lambda_{n-1}}\right)^k v_{n-1}}_{\left(\begin{array}{cccc} v_0 & \cdots & v_{n-2} & v_{n-1} \end{array}\right) \left(\begin{array}{cccc} \left(\frac{1}{\lambda_0}\right)^k & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \left(\frac{1}{\lambda_{n-2}}\right)^k & 0 \\ 0 & \cdots & 0 & \left(\frac{1}{\lambda_{n-1}}\right)^k \end{array}\right) \left(\begin{array}{c} \gamma_0 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1} \end{array}\right)}}_{V(\Lambda^{-1})^k c}
\end{aligned}$$

Now, $\boxed{\text{if}}$ $\lambda_{n-1} = 1$, then $\left|\frac{1}{\lambda_j}\right| < 1$ for $j < n-1$ and hence

$$\lim_{k\to\infty} x^{(k)} = \underbrace{\underbrace{\underbrace{\lim_{k\to\infty} \left(\gamma_0 \left(\frac{1}{\lambda_0}\right)^k v_0 + \cdots + \gamma_{n-2} \left(\frac{1}{\lambda_{n-2}}\right)^k v_{n-2} + \gamma_{n-1} v_{n-1}\right)}_{\lim_{k\to\infty} \left(\begin{array}{cccc} v_0 & \cdots & v_{n-2} & v_{n-1} \end{array}\right) \left(\begin{array}{cccc} \left(\frac{1}{\lambda_0}\right)^k & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \left(\frac{1}{\lambda_{n-2}}\right)^k & 0 \\ 0 & \cdots & 0 & 1 \end{array}\right) \left(\begin{array}{c} \gamma_0 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1} \end{array}\right)}}_{\left(\begin{array}{cccc} v_0 & \cdots & v_{n-2} & v_{n-1} \end{array}\right) \left(\begin{array}{cccc} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{array}\right) \left(\begin{array}{c} \gamma_0 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1} \end{array}\right)}}_{\underbrace{\left(\begin{array}{cccc} 0 & \cdots & 0 & v_{n-1} \end{array}\right) \left(\begin{array}{c} \gamma_0 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1} \end{array}\right)}_{\gamma_{n-1} v_{n-1}}} = \gamma_{n-1} v_{n-1}$$

which means that $x^{(k)}$ eventually starts pointing towards the direction of v_{n-1}, the eigenvector associated with the eigenvalue that is smallest in magnitude. (Well, as long as $\gamma_{n-1} \neq 0$.)

Similar to before, we can instead iterate with the matrix $\lambda_{n-1} A^{-1}$, in which case

$$\left|\frac{\lambda_{n-1}}{\lambda_0}\right| \leq \cdots \leq \left|\frac{\lambda_{n-1}}{\lambda_{n-2}}\right| < \left|\frac{\lambda_{n-1}}{\lambda_{n-1}}\right| = 1.$$

The iteration then becomes

$$\begin{array}{rcl} x^{(1)} & = & \lambda_{n-1}A^{-1}x^{(0)} \\ x^{(2)} & = & \lambda_{n-1}A^{-1}x^{(1)} \\ x^{(3)} & = & \lambda_{n-1}A^{-1}x^{(2)} \\ & \vdots & \end{array}$$

The following algorithm accomplishes this

for $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
  Solve $Ax^{(k+1)} := x^{(k)}$
  $x^{(k+1)} := \lambda_{n-1}x^{(k+1)}$
endfor

It is not hard to see that then

$$\lim_{k\to\infty} x^{(k)} = \underbrace{\underbrace{\underbrace{\lim_{k\to\infty}\left(\gamma_0\left(\frac{\lambda_{n-1}}{\lambda_0}\right)^k v_0 + \cdots + \gamma_{n-2}\left(\frac{\lambda_{n-1}}{\lambda_{n-2}}\right)^k v_{n-2} + \gamma_{n-1}v_{n-1}\right)}_{\lim_{k\to\infty}\left(\begin{array}{cccc} v_0 & \cdots & v_{n-2} & v_{n-1}\end{array}\right)\left(\begin{array}{cccc} (\lambda_{n-1}/\lambda_0)^k & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & (\lambda_{n-1}/\lambda_{n-2})^k & 0 \\ 0 & \cdots & 0 & 1 \end{array}\right)\left(\begin{array}{c}\gamma_0 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1}\end{array}\right)}}_{\left(\begin{array}{cccc} v_0 & \cdots & v_{n-2} & v_{n-1}\end{array}\right)\left(\begin{array}{cccc} 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{array}\right)\left(\begin{array}{c}\gamma_0 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1}\end{array}\right)}}_{\underbrace{\left(\begin{array}{cccc} 0 & \cdots & 0 & v_{n-1}\end{array}\right)\left(\begin{array}{c}\gamma_0 \\ \vdots \\ \gamma_{n-2} \\ \gamma_{n-1}\end{array}\right)}_{\gamma_{n-1}v_{n-1}}} = \gamma_{n-1}v_{n-1}$$

So, it seems that we have an algorithm that always works as long as

$$|\lambda_0| \geq \cdots \geq |\lambda_{n-1}| > |\lambda_{n-1}|.$$

Again, we are cheating... If we knew λ_{n-1}, then we could simply compute the eigenvector by finding a vector in the null space of $A - \lambda_{n-1}I$. Again, the key insight is that, in $x^{(k+1)} = \lambda_{n-1}Ax^{(k)}$, multiplying by λ_{n-1} is merely meant to keep the vector $x^{(k)}$ from getting progressively larger (if $|\lambda_{n-1}| < 1$) or smaller (if $|\lambda_{n-1}| > 1$). We can alternatively simply make $x^{(k)}$ of length one at each step, and that will have the same effect without requiring λ_{n-1}:

for $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
  Solve $Ax^{(k+1)} := x^{(k)}$
  $x^{(k+1)} := x^{(k+1)}/\|x^{(k+1)}\|_2$
endfor

This last algorithm is known as the *Inverse Power Method* for finding an eigenvector associated with the smallest eigenvalue (in magnitude).

Homework 12.5.1.1 The script in `InversePowerMethodScript.m` illustrates how the Inverse Power Method, starting with a random vector, computes an eigenvector corresponding to the eigenvalue that is smallest in magnitude, and (via the Rayleigh quotient) an approximation for that eigenvalue.
To try it out, in the Command Window type

```
>> InversePowerMethodScript
input a vector of eigenvalues. e.g.: [ 4; 3; 2; 1 ]
[ 4; 3; 2; 1 ]
```

If you compare the script for the Power Method with this script, you notice that the difference is that we now use A^{-1} instead of A. To save on computation, we compute the LU factorization once, and solve $LUz = x$, overwriting x with z, to update $x := A^{-1}x$. You will notice that for this distribution of eigenvalues, the Inverse Power Method converges faster than the Power Method does.
Try some other distributions of eigenvalues. For example, `[ 4; 3; 1.25; 1 ]`, which should converge slower, or `[ 4; 3.9; 3.8; 1 ]`, which should converge faster.

Now, it is possible to accelerate the Inverse Power Method if one has a good guess of λ_{n-1}. The idea is as follows: Let μ be close to λ_{n-1}. Then we know that $(A-\mu I)x = (\lambda-\mu)x$. Thus, an eigenvector of A is an eigenvector of A^{-1} is an eigenvector of $A-\mu I$ is an eigenvector of $(A-muI)^{-1}$. Now, if μ is close to λ_{n-1}, then (hopefully)

$$|\lambda_0-\mu| \geq |\lambda_1-\mu| \geq \cdots \geq |\lambda_{n-2}-\mu| > |\lambda_{n-1}-\mu|.$$

The important thing is that if, as before,

$$x^{(0)} = \gamma_0 v_0 + \gamma_1 v_1 + \cdots + \gamma_{n-2} v_{n-2} + \gamma_{n-1} v_{n-1}$$

where v_j equals the eigenvector associated with λ_j, then

$$\begin{array}{rcl} x^{(k)} & = & (\lambda_{n-1}-\mu)(A-\mu I)^{-1}x^{(k-1)} = \cdots = (\lambda_{n-1}-\mu)^k((A-\mu I)^{-1})^k x^{(0)} = \\ & = & \gamma_0(\lambda_{n-1}-\mu)^k((A-\mu I)^{-1})^k v_0 + \gamma_1(\lambda_{n-1}-\mu)^k((A-\mu I)^{-1})^k v_1 + \cdots \\ & & \quad + \gamma_{n-2}(\lambda_{n-1}-\mu)^k((A-\mu I)^{-1})^k v_{n-2} + \gamma_{n-1}(\lambda_{n-1}-\mu)^k((A-\mu I)^{-1})^k v_{n-1} \\ & = & \gamma_0 \left|\frac{\lambda_{n-1}-\mu}{\lambda_0-\mu}\right|^k v_0 + \gamma_1 \left|\frac{\lambda_{n-1}-\mu}{\lambda_1-\mu}\right|^k v_1 + \cdots + \gamma_{n-2}\left|\frac{\lambda_{n-1}-\mu}{\lambda_{n-2}-\mu}\right|^k v_{n-2} + \gamma_{n-1} v_{n-1} \end{array}$$

Now, how fast the terms involving $v_0, \ldots, v_{n-2}$ approx zero (become negligible) is dictated by the ratio

$$\left|\frac{\lambda_{n-1}-\mu}{\lambda_{n-2}-\mu}\right|.$$

Clearly, this can be made arbitrarily small by picking arbitrarily close to λ_{n-1}. Of course, that would require knowning λ_{n-1}...

The practical algorithm for this is given by

for $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
  Solve $(A-\mu I)x^{(k+1)} := x^{(k)}$
  $x^{(k+1)} := x^{(k+1)}/\|x^{(k+1)}\|_2$
endfor

which is referred to as the Shifted Inverse Power Method. Obviously, we would want to only factor $A-\mu I$ once.

Homework 12.5.1.2 The script in `ShiftedInversePowerMethodScript.m` illustrates how shifting the matrix can improve how fast the Inverse Power Method, starting with a random vector, computes an eigenvector corresponding to the eigenvalue that is smallest in magnitude, and (via the Rayleigh quotient) an approximation for that eigenvalue.
To try it out, in the Command Window type

```
>> ShiftedInversePowerMethodScript
input a vector of eigenvalues. e.g.: [ 4; 3; 2; 1 ]
[ 4; 3; 2; 1 ]

<bunch of output>

enter a shift to use: (a number close to the smallest eigenvalue) 0.9
```

If you compare the script for the Inverse Power Method with this script, you notice that the difference is that we now iterate with $(A - sigmaI)^{-1}$, where σ is the shift, instead of A. To save on computation, we compute the LU factorization of $A - \sigma I$ once, and solve $LUz = x$, overwriting x with z, to update $x := (A^{-1} - \sigma I)x$. You will notice that if you pick the shift close to the smallest eigenvalue (in magnitude), this Shifted Inverse Power Method converges faster than the Inverse Power Method does. Indeed, pick the shift very close, and the convergence is very fast. See what happens if you pick the shift exactly equal to the smallest eigenvalue. See what happens if you pick it close to another eigenvalue.

12.5.2 The Rayleigh Quotient Iteration

In the previous unit, we explained that the Shifted Inverse Power Method converges quickly if only we knew a scalar μ close to λ_{n-1}.

The observation is that $x^{(k)}$ eventually approaches v_{n-1}. If we knew v_{n-1} but not λ_{n-1}, then we could compute the Rayleigh quotient:

$$\lambda_{n-1} = \frac{v_{n-1}^T A v_{n-1}}{v_{n-1}^T v_{n-1}}.$$

But we know an approximation of v_{n-1} (or at least its direction) and hence can pick

$$\mu = \frac{x^{(k)T} A x^{(k)}}{x^{(k)T} x^{(k)}} \approx \lambda_{n-1}$$

which will become a progressively better approximation to λ_{n-1} as k increases.

This then motivates the Rayleigh Quotient Iteration:

for $k = 0, \ldots,$ until $x^{(k)}$ doesn't change (much) anymore
  $\mu := \frac{x^{(k)T} A x^{(k)}}{x^{(k)T} x^{(k)}}$
  Solve $(A - \mu I)x^{(k+1)} := x^{(k)}$
  $x^{(k+1)} := x^{(k+1)} / \|x^{(k+1)}\|_2$
endfor

Notice that if $x^{(0)}$ has length one, then we can compute $\mu := x^{(k)T} A x^{(k)}$ instead, since $x^{(k)}$ will always be of length one.

The disadvantage of the Rayleigh Quotient Iteration is that one cannot factor $(A - \mu I)$ once before starting the loop. The advantage is that it converges dazingly fast. Obviously "dazingly" is not a very precise term. Unfortunately, quantifying how fast it converges is beyond this enrichment.

Homework 12.5.2.1 The script in `RayleighQuotientIterationScript.m` illustrates how shifting the matrix by the Rayleigh Quotient can greatly improve how fast the Shifted Inverse Power Method, starting with a random vector, computes an eigenvector. It could be that the random vector is close to an eigenvector associated with any of the eigenvalues, in which case the method will start converging towards an eigenvector associated with that eigenvalue. Pay close attention to how many digit are accurate from one iteration to the next.
To try it out, in the Command Window type

```
>> RayleighQuotientIterationScript
input a vector of eigenvalues. e.g.: [ 4; 3; 2; 1 ]
[ 4; 3; 2; 1 ]
```

12.5.3 More Advanced Techniques

The Power Method and its variants are the bases of algorithms that compute all eigenvalues and eigenvectors of a given matrix. Details, presented with notation similar to what you have learned in this class, can be found in LAFF: Notes on Numerical Linear Algebra.

Consider this unit a "cliff hanger". You will want to take a graduate level course on Numerical Linear Algebra next!

12.6 Wrap Up

12.6.1 Homework

No additional homework this week.

12.6.2 Summary

The algebraic eigenvalue problem

The *algebraic eigenvalue problem* is given by

$$Ax = \lambda x.$$

where $A \in \mathbb{R}^{n\times n}$ is a square matrix, λ is a scalar, and x is a nonzero vector.

- If $x \neq 0$, then λ is said to be an *eigenvalue* and x is said to be an eigenvector associated with the eigenvalue λ.
- The tuple $(\lambda, x\ )$ is said to be an *eigenpair*.
- The set of all vectors that satisfy $Ax = \lambda x$ is a subspace.

Equivalent statements:

- $Ax = \lambda x$,, where $x \neq 0$.
- $(A - \lambda I)x = 0$, where $x \neq 0$.
  This is a matter of fractoring' x out.
- $A - \lambda I$ is singular.
- $\mathcal{N}(A - \lambda I)$ contains a nonzero vector x.
- $\dim(\mathcal{N}(A - \lambda I)) > 0$.
- $\det(A - \lambda I) = 0$.

If we find a vector $x \neq 0$ such that $Ax = \lambda x$, it is certainly not unique.

- For any scalar α, $A(\alpha x) = \lambda(\alpha x)$ also holds.
- If $Ax = \lambda x$ and $Ay = \lambda y$, then $A(x + y) = Ax + Ay = \lambda x + \lambda y = \lambda(x + y)$.

We conclude that the set of all vectors x that satisfy $Ax = \lambda x$ is a subspace.

Simple cases

- The eigenvalue of the zero matrix is the scalar $\lambda = 0$. All nonzero vectors are eigenvectors.
- The eigenvalue of the identity matrix is the scalar $\lambda = 1$. All nonzero vectors are eigenvectors.
- The eigenvalues of a diagonal matrix are its elements on the diagonal. The unit basis vectors are eigenvectors.
- The eigenvalues of a triangular matrix are its elements on the diagonal.
- The eigenvalues of a 2×2 matrix can be found by finding the roots of $p_2(\lambda) = \det(A - \lambda I) = 0$.
- The eigenvalues of a 3×3 matrix can be found by finding the roots of $p_3(\lambda) = \det(A - \lambda I) = 0$.

For 2×2 matrices, the following steps compute the eigenvalues and eigenvectors:

- Compute
$$\det\left(\begin{pmatrix} (\alpha_{0,0} - \lambda) & \alpha_{0,1} \\ \alpha_{1,0} & (\alpha_{1,1} - \lambda) \end{pmatrix}\right) = (\alpha_{0,0} - \lambda)(\alpha_{1,1} - \lambda) - \alpha_{0,1}\alpha_{1,0}.$$
- Recognize that this is a second degree polynomial in λ.
- It is called the *characteristic polynomial* of the matrix A, $p_2(\lambda)$.
- Compute the coefficients of $p_2(\lambda)$ so that
$$p_2(\lambda) = -\lambda^2 + \beta\lambda + \gamma.$$
- Solve
$$-\lambda^2 + \beta\lambda + \gamma = 0$$
for its roots. You can do this either by examination, or by using the quadratic formula:
$$\lambda = \frac{-\beta \pm \sqrt{\beta^2 + 4\gamma}}{-2}.$$
- For each of the roots, find an eigenvector that satisfies
$$\begin{pmatrix} (\alpha_{0,0} - \lambda) & \alpha_{0,1} \\ \alpha_{1,0} & (\alpha_{1,1} - \lambda) \end{pmatrix} \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
The easiest way to do this is to subtract the eigenvalue from the diagonal, set one of the components of x to 1, and then solve for the other component.
- Check your answer! It is a matter of plugging it into $Ax = \lambda x$ and seeing if the computed λ and x satisfy the equation.

General case

Theorem 12.12 *A matrix $A \in \mathbb{R}^{n \times n}$ is nonsingular if and only if* $\det(A) \neq 0$.

Theorem 12.13 *Given $A \in \mathbb{R}^{n \times n}$,*
$$p_n(\lambda) = \det(A - \lambda I) = \lambda^n + \gamma_{n-1}\lambda^{n-1} + \cdots + \gamma_1\lambda + \gamma_0.$$
for some coefficients $\gamma_1, \ldots, \gamma_{n-1} \in \mathbb{R}$.

Definition 12.14 *Given $A \in \mathbb{R}^{n \times n}$, $p_n(\lambda) = \det(A - \lambda I)$ is called the characteristic polynomial.*

Theorem 12.15 *Scalar λ satisfies $Ax = \lambda x$ for some nonzero vector x if and only if* $\det(A - \lambda I) = 0$.

- The characteristic polynomial of $A \in \mathbb{R}^{n \times n}$ is given by
$$p_n(\lambda) = \det(A - \lambda I) = \gamma_0 + \gamma_1\lambda + \cdots + \gamma_{n-1}\lambda^{n-1} + \lambda^n.$$

- Since $p_n(\lambda)$ is an nth degree polynomial, it has n roots, counting multiplicity. Thus, matrix A has n eigenvalues, counting multiplicity.
  - Let k equal the number of *distinct* roots of $p_n(\lambda)$. Clearly, $k \leq n$. Clearly, matrix A then has k distinct eigenvalues.
  - The set of all roots of $p_n(\lambda)$, which is the set of all eigenvalues of A, is denoted by $\Lambda(A)$ and is called the *spectrum* of matrix A.
  - The characteristic polynomial can be factored as

$$p_n(\lambda) = \det(A - \lambda I) = (\lambda - \lambda_0)^{n_0}(\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_{k-1})^{n_{k-1}},$$

    where $n_0 + n_1 + \cdots + n_{k-1} = n$ and n_j is the root λ_j, which is known as that (algebraic) multiplicity of eigenvalue λ_j.
- If $A \in \mathbb{R}^{n\times n}$, then the coefficients of the characteristic polynomial are real ($\gamma_0, \ldots, \gamma_{n-1} \in \mathbb{R}$), but
  - Some or all of the roots/eigenvalues may be complex valued and
  - Complex roots/eigenvalues come in "conjugate pairs": If $\lambda = \mathrm{Re}(\lambda) + i\mathrm{Im}(\lambda)$ is a root/eigenvalue, so is $\lambda = \mathrm{Re}(\lambda) - i\mathrm{Im}(\lambda)$

Galois theory tells us that for $n \geq 5$, roots of arbitrary $p_n(\lambda)$ cannot be found in a finite number of computations.
For every $n$the degree polynomial

$$p_n(\lambda) = \gamma_0 + \gamma_1\lambda + \cdots + \gamma_{n-1}\lambda^{n-1} + \lambda^n,$$

there exists a matrix, C, called the companion matrix that has the property that

$$p_n(\lambda) = \det(C - \lambda I) = \gamma_0 + \gamma_1\lambda + \cdots + \gamma_{n-1}\lambda^{n-1} + \lambda^n.$$

In particular, the matrix

$$C = \begin{pmatrix} -\gamma_{n-1} & -\gamma_{n-2} & \cdots & -\gamma_1 & -\gamma_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

is the companion matrix for $p_n(\lambda)$:

$$p_n(\lambda) = \gamma_0 + \gamma_1\lambda + \cdots + \gamma_{n-1}\lambda^{n-1} + \lambda^n = \det(\begin{pmatrix} -\gamma_{n-1} & -\gamma_{n-2} & \cdots & -\gamma_1 & -\gamma_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} - \lambda I).$$

Diagonalization

Theorem 12.16 *Let $A \in \mathbb{R}^{n\times n}$. Then there exists a nonsingular matrix X such that $X^{-1}AX = \Lambda$ if and only if A has n linearly independent eigenvectors.*

If X is invertible (nonsingular, has linearly independent columns, etc.), then the following are equivalent

$$\begin{array}{rcl} X^{-1}\,A\,X & = & \Lambda \\ A\,X & = & X\,\Lambda \\ A & = & X\,\Lambda\,X^{-1} \end{array}$$

If Λ is in addition diagonal, then the diagonal elements of Λ are eigenvalues of A and the columns of X are eigenvectors of A.

Defective matrices

It is *not* the case that for every $A \in \mathbb{R}^{n \times n}$ there is a nonsingular matrix $X \in \mathbb{R}^{n \times n}$ such that $X^{-1}AX = \Lambda$, where Λ is diagonal. In general, the $k \times k$ matrix $J_k(\lambda)$ given by

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

has eigenvalue λ of algebraic multiplicity k, but *geometric multiplicity* one (it has only one linearly independent eigenvector). Such a matrix is known as a Jordan block.

Definition 12.17 *The geometric multiplicity of an eigenvalue λ equals the number of linearly independent eigenvectors that are associated with λ.*

Theorem 12.18 *Let $A \in \mathbb{R}^{n \times n}$. Then there exists a nonsingular matrix $X \in \mathbb{R}^{n \times n}$ such that $A = XJX^{-1}$, where*

$$J = \begin{pmatrix} J_{k_0}(\lambda_0) & 0 & 0 & \cdots & 0 \\ 0 & J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & 0 & J_{k_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & J_{k_{m-1}}(\lambda_{m-1}) \end{pmatrix}$$

where each $J_{k_j}(\lambda_j)$ is a Jordan block of size $k_j \times k_j$.

The factorization $A = XJX^{-1}$ is known as the Jordan Canonical Form of matrix A.

In the above theorem

- It is *not* the case that $\lambda_0, \lambda_1, \ldots, \lambda_{m-1}$ are distinct. If λ_j appears in multiple Jordan blocks, the number of Jordan blocks in which λ_j appears equals the geometric multiplicity of λ_j (and the number of linearly independent eigenvectors associated with λ_j).
- The sum of the sizes of the blocks in which λ_j as an eigenvalue appears equals the algebraic multiplicity of λ_j.
- If each Jordan block is 1×1, then the matrix is diagonalized by matrix X.
- If any of the blocks is not 1×1, then the matrix cannot be diagonalized.

Properties of eigenvalues and eigenvectors

Definition 12.19 *Given $A \in \mathbb{R}^{n \times n}$ and nonzero vector $x \in \mathbb{R}^n$, the scalar $x^T Ax/x^T x$ is known as the Rayleigh quotient.*

Theorem 12.20 *Let $A \in \mathbb{R}^{n \times n}$ and x equal an eigenvector of A. Assume that x is real valued as is the eigenvalue λ with $Ax = \lambda x$. Then $\lambda = \frac{x^T Ax}{x^T x}$ is the eigenvalue associated with the eigenvector x.*

Theorem 12.21 *Let $A \in \mathbb{R}^{n \times n}$, β be a scalar, and $\lambda \in \Lambda(A)$. Then $(\beta\lambda) \in \Lambda(\beta A)$.*

Theorem 12.22 *Let $A \in \mathbb{R}^{n \times n}$ be nonsingular, $\lambda \in \Lambda(A)$, and $Ax = \lambda x$. Then $A^{-1}x = \frac{1}{\lambda}x$.*

Theorem 12.23 *Let $A \in \mathbb{R}^{n \times n}$ and $\lambda \in \Lambda(A)$. Then $(\lambda - \mu) \in \Lambda(A - \mu I)$.*