DSC 140B

Representation Learning

Lecture 04 | Part 1

News

News

- Quiz 02 tonight.
- See Campuswire post re: Peter Chi's study on data science education.
  - **You get a $10 gift card to the bookstore**.
- Waitlist

Live QA

Exercise

Are you planning on taking the quiz tonight?

True = Yes

False = No

DSC 140B
Representation Learning

Lecture 04 | Part 2

The Spectral Theorem

Eigenvectors

- Let A be an $n \times n$ matrix. An **eigenvector** of A with **eigenvalue** λ is a nonzero vector $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$.

Eigenvectors (of Linear Transformations)

- Let $\vec{f}$ be a linear transformation. An **eigenvector** of $\vec{f}$ with **eigenvalue** λ is a nonzero vector $\vec{v}$ such that $\vec{f}(\vec{v}) = \lambda\vec{v}$.

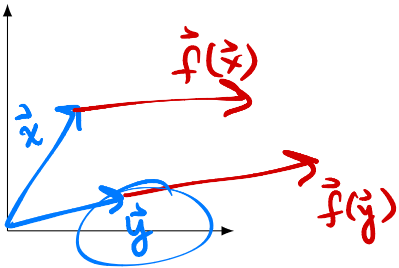
Importance

- We will see why eigenvectors are important in the next part.
- For now: what are they?

Geometric Interpretation

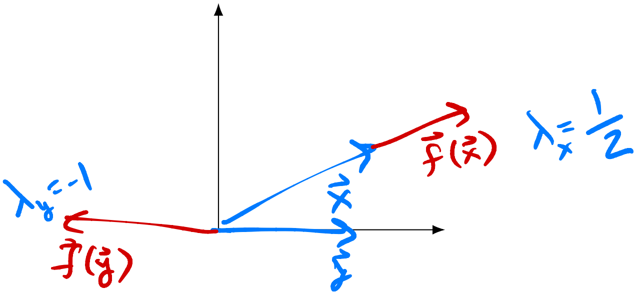
- Recall: $\vec{v}$ is an **eigenvector** if $\vec{f}(\vec{v}) = \lambda \vec{v}$.

- Meaning: when $\vec{f}$ is applied to one of its eigenvectors, $\vec{f}$ simply scales it.



Geometric Interpretation

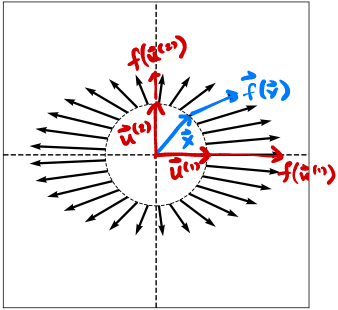
- The **eigenvalue**, λ, tells us how much the eigenvector is scaled.
  - If $\lambda > 1$, the eigenvector is stretched.
  - If $0 < \lambda < 1$, the eigenvector is shrunk.
  - If $\lambda < 0$, the eigenvector is flipped and scaled.



Exercise

Draw as many (linearly independent) eigenvectors as you can.

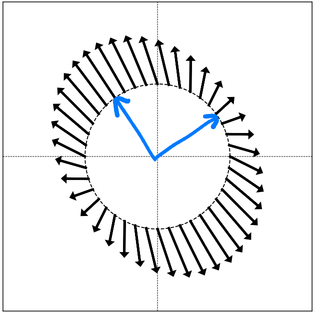
$$A = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}$$



Exercise

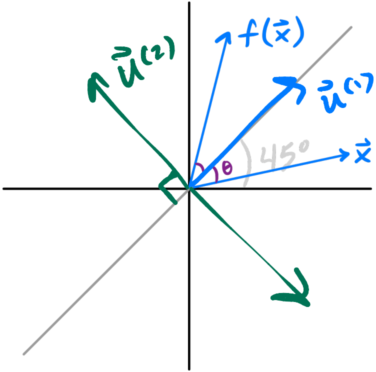
Draw as many (linearly independent) eigenvectors as you can.

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$$



Exercise

Consider the linear transformation which mirrors its input over the line of 45°. Give two orthogonal eigenvectors of the transformation.



$\vec{f}(\vec{u}^{(1)}) = \vec{u}^{(1)}$ ← $\lambda_1 = 1$

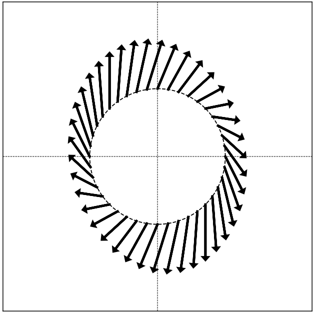
$\vec{f}(\vec{u}^{(2)}) = -\vec{u}^{(2)}$ ← $\lambda_2 = -1$

Exercise

Draw as many (linearly independent) eigenvectors as you can.

None

$$A = \begin{pmatrix} 5 & 5 \\ -10 & 12 \end{pmatrix}$$

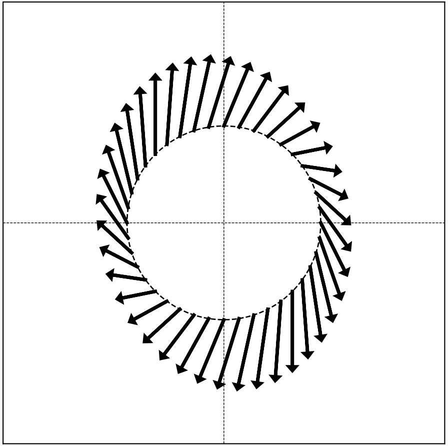


Caution!

- Not all matrices have even one eigenvector![1]
- When does a matrix have multiple (linearly independent) eigenvectors?

[1]That is, with a *real-valued* eigenvalue.

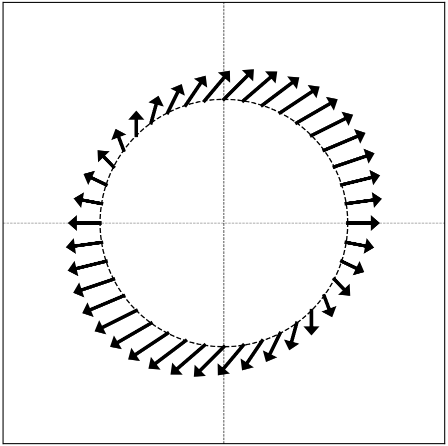
Example Linear Transformation



None

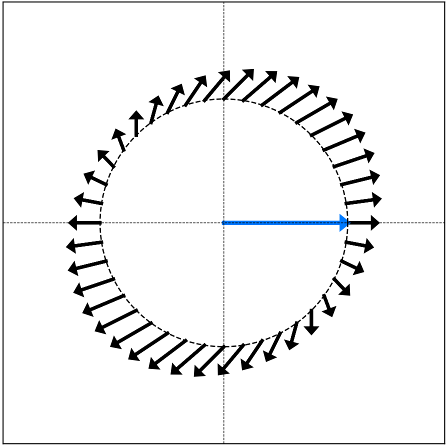
$$A = \begin{pmatrix} 5 & 5 \\ -10 & 12 \end{pmatrix}$$

Example Linear Transformation



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

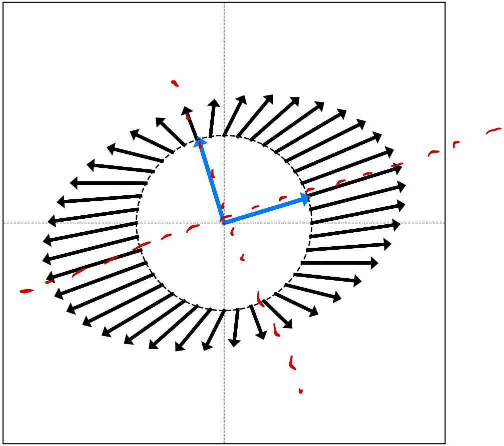
Example Linear Transformation



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

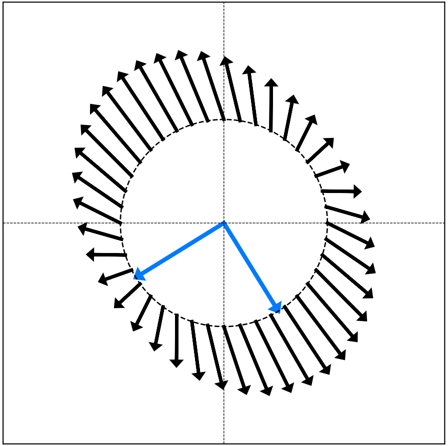
$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Example Symmetric Linear Transformation



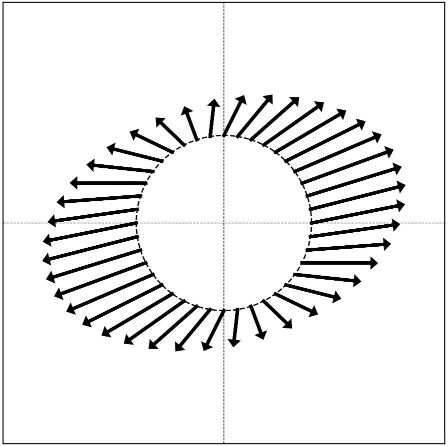
$$A = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$$

Example **Symmetric** Linear Transformation



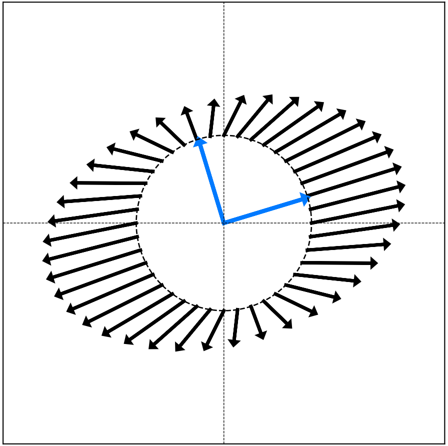
$$A = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$$

Example Symmetric Linear Transformation



$$A = \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}$$

Example Symmetric Linear Transformation



$$A = \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}$$

Observation

- It seems that there is something special about **symmetric** matrices...

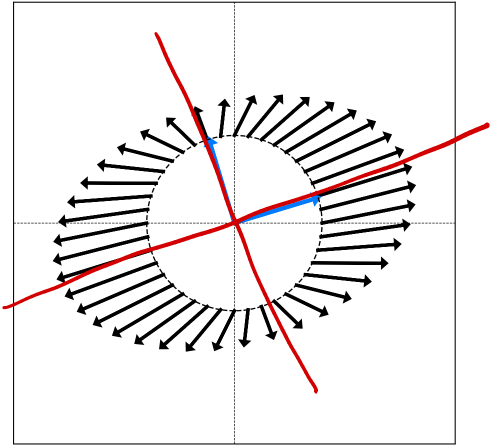
Symmetric Matrices

- Recall: a matrix A is **symmetric** if $A^T = A$.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}$$

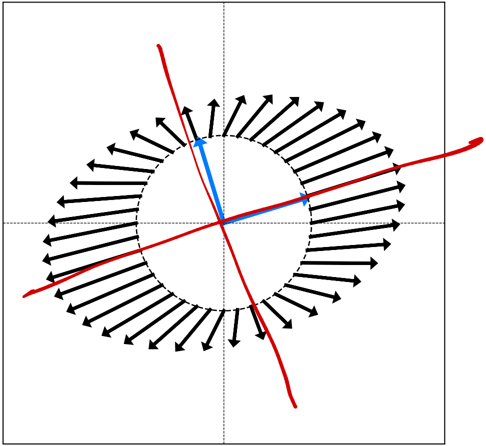
- A linear transformation $\vec{f}$ is **symmetric** if its matrix representation is symmetric.

Observation #1



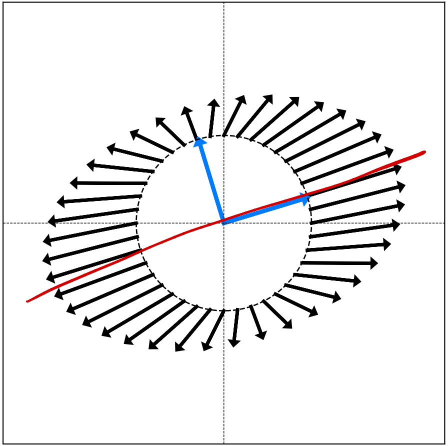
- Symmetric linear transformations have **axes of symmetry**.
  - One for each dimension.

Observation #2



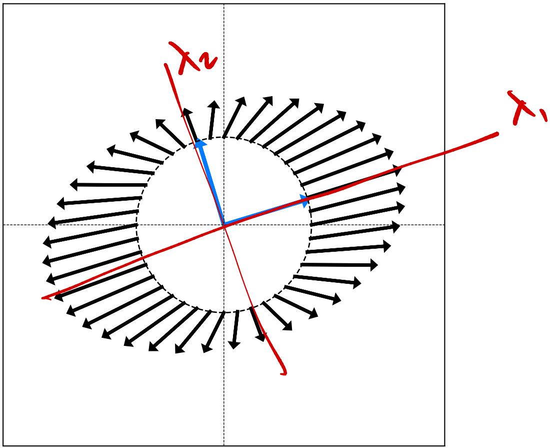
- The axes of symmetry are **orthogonal** to one another.

Observation #3



- The action of $\vec{f}$ along an axis of symmetry is simply to **scale** its input.
- That is, the **eigenvectors** point along the axes of symmetry.

Observation #4



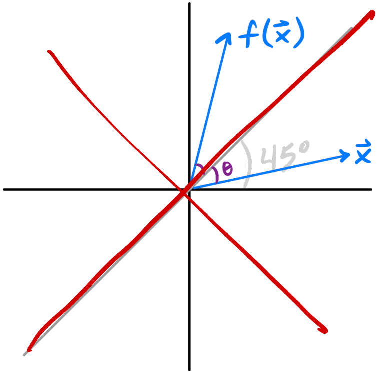
- The size of this scaling can be different for each axis.

Main Idea

The **eigenvectors** of a symmetric linear transformation (matrix) are its axes of symmetry. The **eigenvalues** describe how much each axis of symmetry is scaled.

Exercise

Consider the linear transformation which mirrors its input over the line of 45°. Give two orthogonal eigenvectors of the transformation.



How many?

- The symmetric 2×2 matrices we saw all had 2 orthogonal eigenvectors.
- Does a 3×3 symmetric matrix have 3 orthogonal eigenvectors?
- What about $n \times n$ symmetric matrices?

The Spectral Theorem[2]

Theorem

Let A be an $n \times n$ **symmetric** matrix. Then you can always find n eigenvectors of A which are all mutually orthogonal.

[2]for symmetric matrices

Careful!

- The spectral theorem *does not* say that an $n \times n$ matrix has n eigenvectors!

Live Q&A

Exercise

Consider the 2×2 identity matrix. How many (unit) eigenvectors does it have?

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

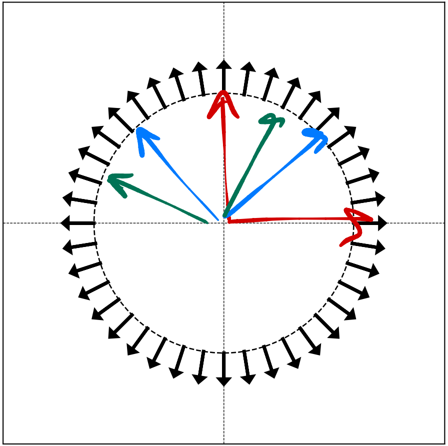
A. 0

B. 1

C. 2

D. ∞

Solution



- Infinitely many!
- *Every* (nonzero) vector is an eigenvector with eigenvalue 1.

Solution

- It would be incorrect to say that the identity matrix has just 2 orthogonal eigenvectors.

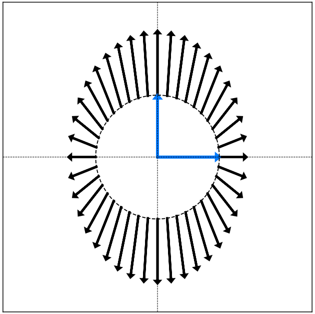
- Instead, the spectral theorem says: "You can find 2 different orthogonal eigenvectors of I."

- There are infinitely-many ways to do this!
  - $(1, 0)^T$ and $(0, 1)^T$
  - $(1/\sqrt{2}, 1/\sqrt{2})^T$ and $(-1/\sqrt{2}, 1/\sqrt{2})^T$
  - etc.

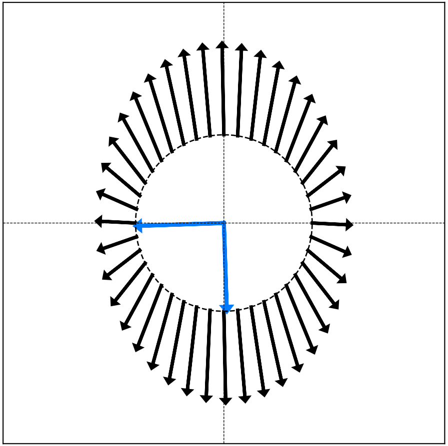
Diagonal Matrices

- If A is diagonal, its eigenvectors are simply the standard basis vectors.

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}$$

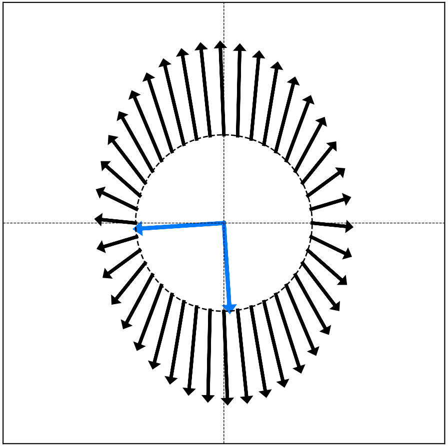


Off-diagonal elements



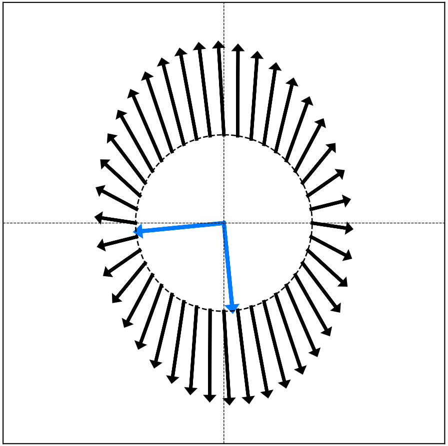
$$A = \begin{pmatrix} 2 & -0.1 \\ -0.1 & 5 \end{pmatrix}$$

Off-diagonal elements



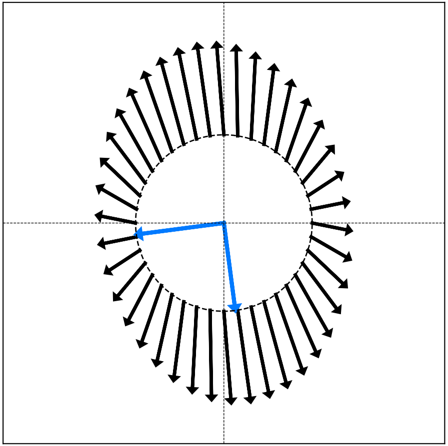
$$A = \begin{pmatrix} 2 & -0.2 \\ -0.2 & 5 \end{pmatrix}$$

Off-diagonal elements



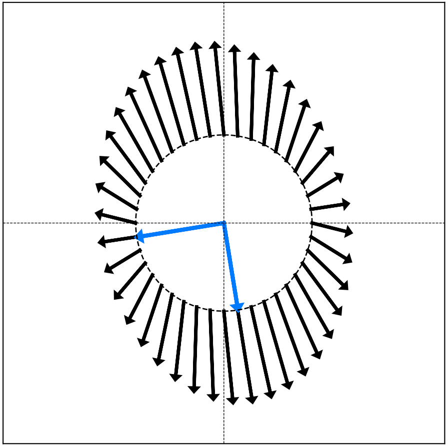
$$A = \begin{pmatrix} 2 & -0.3 \\ -0.3 & 5 \end{pmatrix}$$

Off-diagonal elements



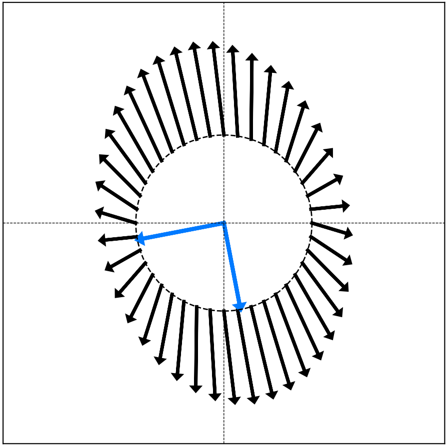
$$A = \begin{pmatrix} 2 & -0.4 \\ -0.4 & 5 \end{pmatrix}$$

Off-diagonal elements



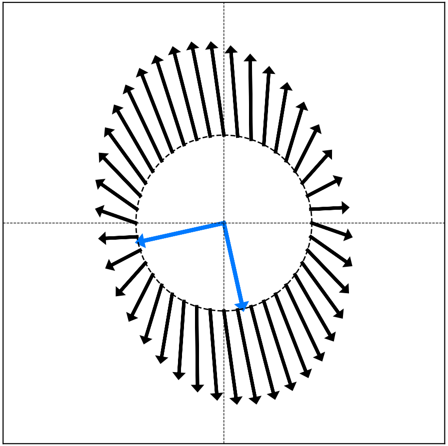
$$A = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 5 \end{pmatrix}$$

Off-diagonal elements



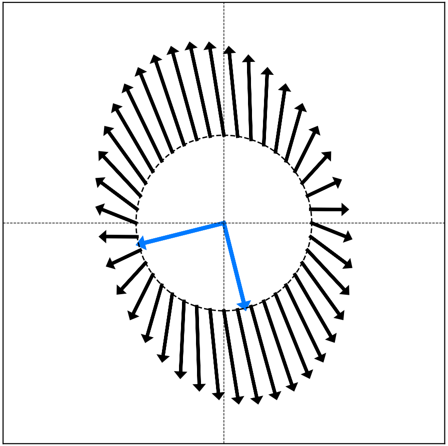
$$A = \begin{pmatrix} 2 & -0.6 \\ -0.6 & 5 \end{pmatrix}$$

Off-diagonal elements



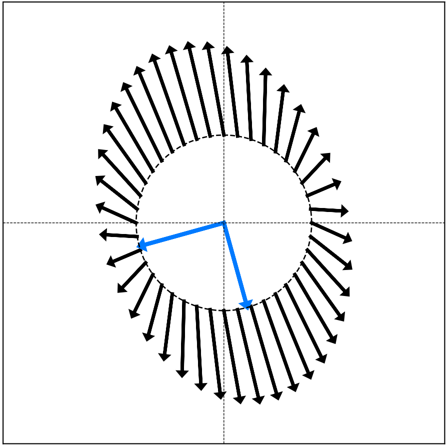
$$A = \begin{pmatrix} 2 & -0.7 \\ -0.7 & 5 \end{pmatrix}$$

Off-diagonal elements



$$A = \begin{pmatrix} 2 & -0.8 \\ -0.8 & 5 \end{pmatrix}$$

Off-diagonal elements



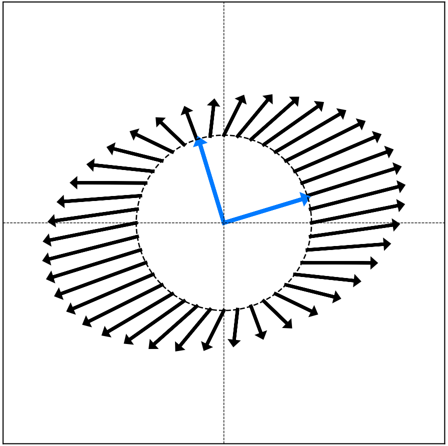
$$A = \begin{pmatrix} 2 & -0.9 \\ -0.9 & 5 \end{pmatrix}$$

Non-Diagonal Symmetric Matrices

- When a symmetric matrix is not diagonal, its eigenvectors are not the standard basis vectors.

- But they are still orthogonal!

Computing Eigenvectors



$$\begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$$

Use `np.linalg.eigh`[a]:

```
>>> A = np.array([[2, -1], [-1, 3]])
>>> np.linalg.eigh(A)
(array([1.38196601, 3.61803399]),
 array([[-0.85065081, -0.52573111],
        [-0.52573111,  0.85065081]]))
```

[a] if the input is *symmetric*

DSC 140B

Representation Learning

Lecture 04 | Part 3

Why are eigenvectors useful?

OK, but why are eigenvectors[3] useful?

1. Eigenvectors are natural **basis vectors**.

2. Eigenvectors are **equilibria**.

3. Eigenvectors are **maximizers** (or minimizers).

[3] of symmetric matrices

OK, but why are eigenvectors[3] useful?

1. Eigenvectors are natural **basis vectors**.

2. Eigenvectors are **equilibria**.

3. Eigenvectors are **maximizers** (or minimizers).

---

[3] of symmetric matrices

Recall: Spectral Theorem

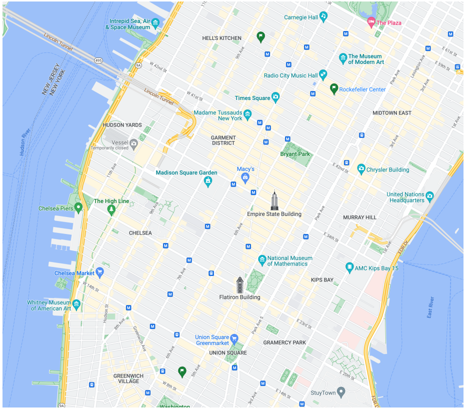
Theorem

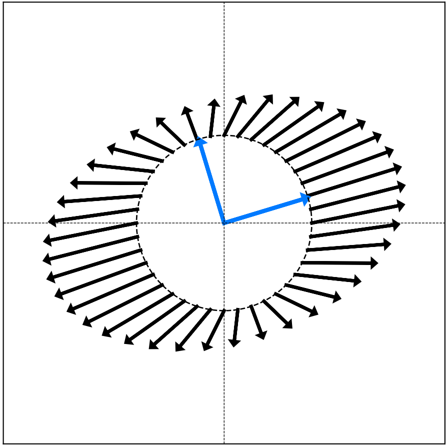
Let A be a symmetric $d \times d$ matrix. Then you can find d orthonormal eigenvectors $\hat{u}^{(1)}, \ldots, \hat{u}^{(d)}$ of A.

- In other words, you can make an orthonormal basis out of eigenvectors of A.

“Nice” Bases


N
E


Intrepid Sea, Air & Space Museum
Carnegie Hall
The Plaza
HELL'S KITCHEN
The Museum of Modern Art
Radio City Music Hall
Rockefeller Center
Times Square
Madame Tussauds New York
MIDTOWN EAST
HUDSON YARDS
GARMENT DISTRICT
Vessel
Temporarily closed
Bryant Park
Madison Square Garden
Macy's
Chrysler Building
United Nations Headquarters
The High Line
Chelsea Piers
Empire State Building
MURRAY HILL
CHELSEA
National Museum of Mathematics
AMC Kips Bay 15
Chelsea Market
KIPS BAY
Flatiron Building
Whitney Museum of American Art
Union Square Greenmarket
GRAMERCY PARK
UNION SQUARE
GREENWICH VILLAGE
StuyTown
NEW JERSEY
NEW YORK
Hudson River
East River
Lincoln Tunnel



Using the Eigenbasis

- When we work in the **eigenbasis** of A, many things become simpler.

Example

- Consider the symmetric matrix A.
- If we change basis, A changes.
- What does it look like if we change to the eigenbasis of A?

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix}$$

$f(\hat{e}^{(1)})$ $f(\hat{e}^{(2)})$ $f(\hat{e}^{(3)})$

Example

- Consider the symmetric matrix A.
- If we change basis, A changes.
- What does it look like if we change to the eigenbasis of A?

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix} \xrightarrow{\text{eigenbasis}} [A]_{\mathcal{U}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{pmatrix}$$

- It becomes diagonal!

Example

- Evaluating the linear transformation becomes easier, too.
- Suppose $\vec{x} = (3, 2, 1)^T$. Before:

$$\begin{aligned}\vec{f}(\vec{x}) &= A\vec{x} \\ &= \begin{pmatrix} 4 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \times 3 + 2 \times 2 + 1 \times 1 \\ 2 \times 3 + 3 \times 2 + 2 \times 1 \\ 1 \times 3 + 2 \times 2 + 4 \times 1 \end{pmatrix} = \begin{pmatrix} 17 \\ 14 \\ 11 \end{pmatrix}\end{aligned}$$

Example

- In the eigenbasis, $\vec{x}$'s coordinates are:

$$[\vec{x}]_{\mathcal{U}} = (0, \sqrt{2}, 2\sqrt{3})^T$$

- So:

$$[A]_{\mathcal{U}}[\vec{x}]_{\mathcal{U}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{pmatrix} \begin{pmatrix} 0 \\ \sqrt{2} \\ 2\sqrt{3} \end{pmatrix} = \begin{pmatrix} 0 \\ 3\sqrt{2} \\ 14\sqrt{3} \end{pmatrix}$$

OK, but why are eigenvectors[4] useful?

1. Eigenvectors are natural **basis vectors**.

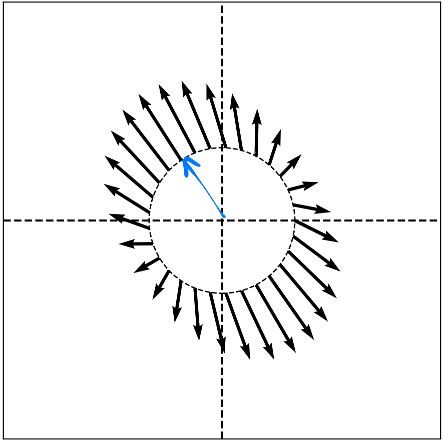
2. Eigenvectors are **equilibria**.

3. Eigenvectors are **maximizers** (or minimizers).

---

[4] of symmetric matrices

Eigenvectors are Equilibria



- $\vec{f}(\vec{x})$ rotates $\vec{x}$ towards the "top" eigenvector $\vec{v}$.
- $\vec{v}$ is an equilibrium.

Use Case: The Power Method

- Method for computing the top eigenvector/value of A.
- Initialize $\vec{x}^{(0)}$ randomly
- Repeat until convergence:
  - Set $\vec{x}^{(i+1)} = A\vec{x}^{(i)} / \|A\vec{x}^{(i)}\|$

OK, but why are eigenvectors[5] useful?

1. Eigenvectors are natural **basis vectors**.
2. Eigenvectors are **equilibria**.
3. Eigenvectors are **maximizers** (or minimizers).

[5] of symmetric matrices

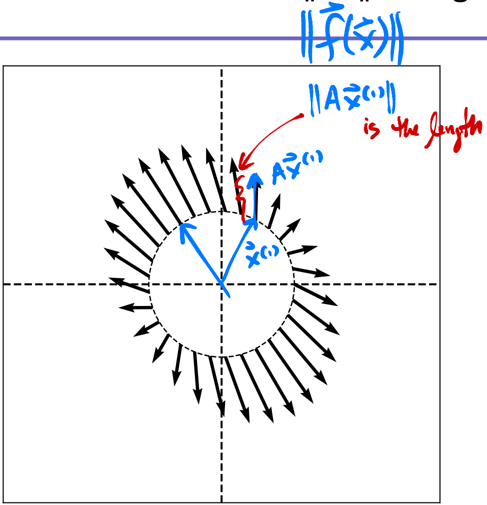
Eigenvectors as Optimizers

- Eigenvectors are the solutions to certain common optimization problems involving matrices / linear transformations.

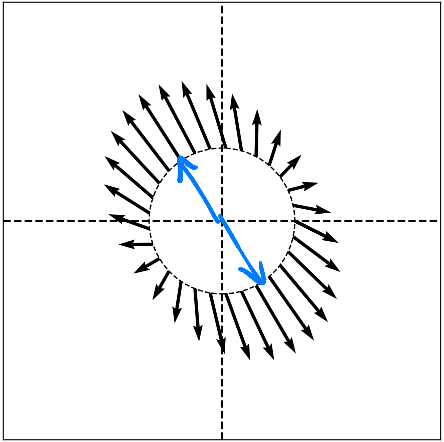
- This might be **the** most important reason why eigenvectors are useful in **data science**.

Exercise

Draw a unit vector $\vec{x}$ such that $\|A\vec{x}\|$ is largest.



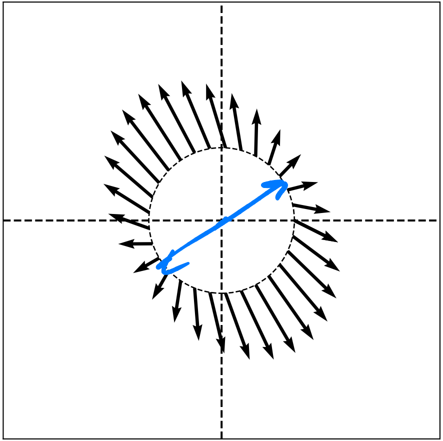
Observation #1



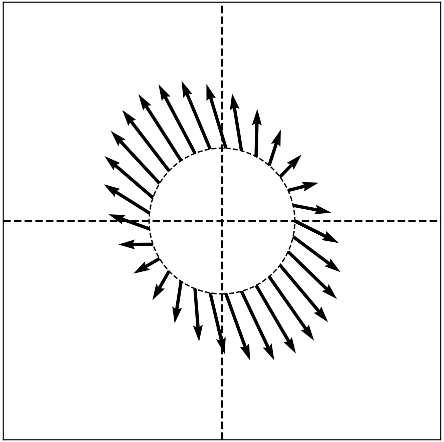
- $\vec{f}(\vec{x})$ is longest along the "main" axis of symmetry.
  - In the direction of the eigenvector with largest eigenvalue.

Exercise

Draw a unit vector $\vec{x}$ such that $\|A\vec{x}\|$ is smallest.



Observation #2



- $\vec{f}(\vec{x})$ is smallest along the "minor" axis of symmetry.
  - In the direction of the eigenvector with smallest eigenvalue.

Main Idea

Suppose A is a symmetric matrix.

To maximize $\|A\vec{x}\|$ over unit vectors, pick $\vec{x}$ to be a top eigenvector of A. That is, an eigenvector with the largest eigenvalue (in abs. value).

To minimize $\|A\vec{x}\|$, pick $\vec{x}$ to be a bottom eigenvector. That is, an eigenvector with the smallest eigenvalue (in abs. value).

Main Idea

Suppose $\vec{f}$ is a symmetric linear transformation.

To maximize $\|\vec{f}(\vec{x})\|$ over unit vectors, pick $\vec{x}$ to be a top eigenvector of $\vec{f}$.

To minimize $\|\vec{f}(\vec{x})\|$ over unit vectors, pick $\vec{x}$ to be a bottom eigenvector.

Also true...

- $\vec{x} \cdot A\vec{x}$ is called a **quadratic form**.

Theorem

Let A be a symmetric matrix.

To maximize $\vec{x} \cdot A\vec{x}$ over unit vectors, pick $\vec{x}$ to be a top eigenvector of A.

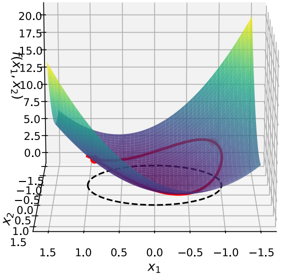
To minimize $\vec{x} \cdot A\vec{x}$ over unit vectors, pick $\vec{x}$ to be a bottom eigenvector of A.

By the way...

- We'll walk you through the proofs in the homework.

Example

- **Problem:** Maximize $f(x_1, x_2) = 4x_1^2 + 2x_2^2 + 3x_1x_2$ subject to $x_1^2 + x_2^2 = 1$



Solution

- **Problem:** Maximize $f(x_1, x_2) = 4x_1^2 + 2x_2^2 + 3x_1x_2$ subject to $x_1^2 + x_2^2 = 1$

- You can write $f(x_1, x_2)$ as $f(\vec{x}) = \vec{x} \cdot A\vec{x}$ where

$$A = \begin{pmatrix} 4 & 1.5 \\ 1.5 & 2 \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

- Top eigenvector of A is approximately: $(0.88, 0.47)^T$

- Solution: maximized at $x_1 = 0.88$, $x_2 = 0.47$

Next time...

- Change of basis matrices, diagonalization.
- Dimensionality reduction (**actual ML!**)