

# DSC 140B

## Representation Learning

Lecture 04 | Part 1

**Matrices**

# Matrices?

- ▶ I thought this week was supposed to be about linear algebra... Where are the matrices?

# Matrices?

- ▶ I thought this week was supposed to be about linear algebra... Where are the matrices?
- ▶ What is a matrix, anyways?

# What is a matrix?

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

# Recall: Linear Transformations

- ▶ A **transformation**  $\vec{f}(\vec{x})$  is a function which takes a vector as input and returns a vector of the same dimensionality.
- ▶ A transformation  $\vec{f}$  is **linear** if

$$\vec{f}(\alpha\vec{u} + \beta\vec{v}) = \alpha\vec{f}(\vec{u}) + \beta\vec{f}(\vec{v})$$

# Recall: Linear Transformations

- ▶ **Key** consequence of **linearity**: to compute  $\vec{f}(\vec{x})$ , only need to know what  $\vec{f}$  does to basis vectors.
- ▶ Example:

$$\vec{x} = 3\hat{e}^{(1)} - 4\hat{e}^{(2)} = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$$

$$\vec{f}(\hat{e}^{(1)}) = -\hat{e}^{(1)} + 3\hat{e}^{(2)}$$

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

$$\vec{f}(\vec{x}) =$$

# Matrices

- ▶ **Idea:** Since  $\vec{f}$  is defined by what it does to basis, place  $\vec{f}(\hat{e}^{(1)})$ ,  $\vec{f}(\hat{e}^{(2)})$ , ... into a table as columns
- ▶ This is the **matrix** representing<sup>1</sup>  $\vec{f}$

$$\begin{aligned}\vec{f}(\hat{e}^{(1)}) &= -\hat{e}^{(1)} + 3\hat{e}^{(2)} = \begin{pmatrix} -1 \\ 3 \end{pmatrix} \\ \vec{f}(\hat{e}^{(2)}) &= 2\hat{e}^{(1)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}\end{aligned}\qquad \begin{pmatrix} -1 & 2 \\ 3 & 0 \end{pmatrix}$$

---

<sup>1</sup>with respect to the standard basis  $\hat{e}^{(1)}, \hat{e}^{(2)}$

## Exercise

Write the matrix representing  $\vec{f}$  with respect to the standard basis, given:

$$\vec{f}(\hat{e}^{(1)}) = (1, 4, 7)^T$$

$$\vec{f}(\hat{e}^{(2)}) = (2, 5, 7)^T$$

$$\vec{f}(\hat{e}^{(3)}) = (3, 6, 9)^T$$



## Exercise

Suppose  $\vec{f}$  has the matrix below:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

Let  $\vec{x} = (-2, 1, 3)^T$ . What is  $\vec{f}(\vec{x})$ ?

## Main Idea

A square ( $n \times n$ ) matrix can be interpreted as a compact representation of a linear transformation  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ .

# What is matrix multiplication?

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} \phantom{0} \\ \phantom{0} \\ \phantom{0} \end{pmatrix}$$

## A low-level definition

$$(A\vec{x})_i = \sum_{j=1}^n A_{ij}x_j$$

## A low-level interpretation

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} = -2 \begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} + 1 \begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix} + 3 \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix}$$

**In general...**

$$\begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{a}^{(1)} & \vec{a}^{(2)} & \vec{a}^{(3)} \\ \downarrow & \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \vec{a}^{(1)} + x_2 \vec{a}^{(2)} + x_3 \vec{a}^{(3)}$$

# Matrix Multiplication

$$\vec{x} = x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)} + x_3 \hat{e}^{(3)} = (x_1, x_2, x_3)^T$$
$$\vec{f}(\vec{x}) = x_1 \vec{f}(\hat{e}^{(1)}) + x_2 \vec{f}(\hat{e}^{(2)}) + x_3 \vec{f}(\hat{e}^{(3)})$$

$$A = \begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{f}(\hat{e}^{(1)}) & \vec{f}(\hat{e}^{(2)}) & \vec{f}(\hat{e}^{(3)}) \\ \downarrow & \downarrow & \downarrow \end{pmatrix}$$
$$A\vec{x} = \begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{f}(\hat{e}^{(1)}) & \vec{f}(\hat{e}^{(2)}) & \vec{f}(\hat{e}^{(3)}) \\ \downarrow & \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$
$$= x_1 \vec{f}(\hat{e}^{(1)}) + x_2 \vec{f}(\hat{e}^{(2)}) + x_3 \vec{f}(\hat{e}^{(3)})$$

# Matrix Multiplication

- ▶ Matrix  $A$  represents a linear transformation  $\vec{f}$ 
  - ▶ With respect to the standard basis
  - ▶ If we use a different basis, the matrix changes!
  
- ▶ Matrix multiplication  $A\vec{x}$  **evaluates**  $\vec{f}(\vec{x})$



## **What are they, *really*?**

- ▶ Matrices are sometimes just tables of numbers.
- ▶ But they often have a deeper meaning.

## Main Idea

A square ( $n \times n$ ) matrix can be interpreted as a compact representation of a linear transformation  $\vec{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ .

What's more, if  $A$  represents  $\vec{f}$ , then  $A\vec{x} = \vec{f}(\vec{x})$ ; that is, multiplying by  $A$  is the same as evaluating  $\vec{f}$ .

# Example

$$\vec{x} = 3\hat{e}^{(1)} - 4\hat{e}^{(2)} = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$$

$$A =$$

$$\vec{f}(\hat{e}^{(1)}) = -\hat{e}^{(1)} + 3\hat{e}^{(2)}$$

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

$$\vec{f}(\vec{x}) =$$

$$A\vec{x} =$$

## Note

- ▶ All of this works because we assumed  $\vec{f}$  is **linear**.
- ▶ If it isn't, evaluating  $\vec{f}$  isn't so simple.

## Note

- ▶ All of this works because we assumed  $\vec{f}$  is **linear**.
- ▶ If it isn't, evaluating  $\vec{f}$  isn't so simple.
- ▶ Linear algebra = simple!

# Matrices in Other Bases

- ▶ The matrix of a linear transformation wrt the **standard basis**:

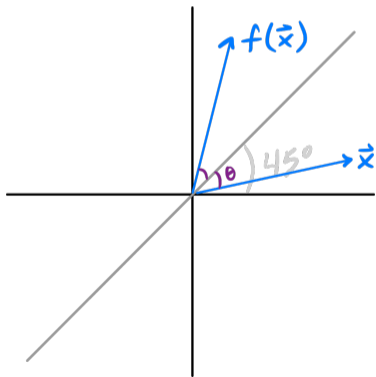
$$\begin{pmatrix} \uparrow & \uparrow & \uparrow & \\ \vec{f}(\hat{e}^{(1)}) & \vec{f}(\hat{e}^{(2)}) & \dots & \vec{f}(\hat{e}^{(d)}) \\ \downarrow & \downarrow & \downarrow & \end{pmatrix}$$

- ▶ With respect to basis  $\mathcal{U}$ :

$$\begin{pmatrix} \uparrow & \uparrow & \uparrow & \\ [\vec{f}(\hat{u}^{(1)})]_{\mathcal{U}} & [\vec{f}(\hat{u}^{(2)})]_{\mathcal{U}} & \dots & [\vec{f}(\hat{u}^{(d)})]_{\mathcal{U}} \\ \downarrow & \downarrow & \downarrow & \end{pmatrix}$$

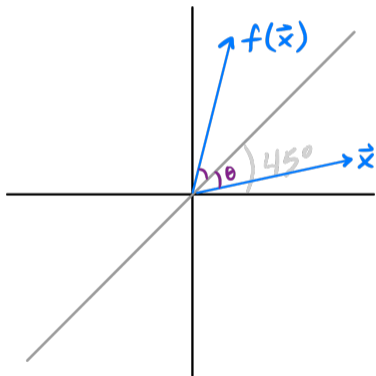
# Matrices in Other Bases

- ▶ Consider the transformation  $\vec{f}$  which “mirrors” a vector over the line of  $45^\circ$ .



- ▶ What is its matrix in the standard basis?

# Matrices in Other Bases



- ▶ Let  $\hat{u}^{(1)} = \frac{1}{\sqrt{2}}(1, 1)^T$
- ▶ Let  $\hat{u}^{(2)} = \frac{1}{\sqrt{2}}(-1, 1)^T$
- ▶ What is  $[\vec{f}(\hat{u}^{(1)})]_{\mathcal{U}}$ ?
- ▶  $[\vec{f}(\hat{u}^{(2)})]_{\mathcal{U}}$ ?
- ▶ What is the matrix?



# 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**  $\lambda$  is a nonzero vector  $\vec{v}$  such that  $A\vec{v} = \lambda\vec{v}$ .

# Eigenvectors (of Linear Transformations)

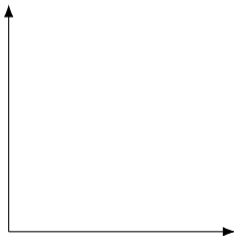
- ▶ Let  $\vec{f}$  be a linear transformation. An **eigenvector** of  $\vec{f}$  with **eigenvalue**  $\lambda$  is a nonzero vector  $\vec{v}$  such that  $f(\vec{v}) = \lambda\vec{v}$ .

# Importance

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

# Geometric Interpretation

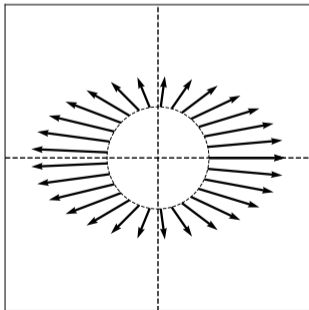
- ▶ When  $\vec{f}$  is applied to one of its eigenvectors,  $\vec{f}$  simply scales it.
  - ▶ Possibly by a negative amount.



## Exercise

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

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



# Finding Eigenvectors

- ▶ We typically compute the eigenvectors of a matrix with a computer.
- ▶ But it can help our understanding to find them “graphically”.

# Procedure

Given a matrix  $A$  (or transformation  $\vec{f}$ ), to find an eigenvector “graphically”.

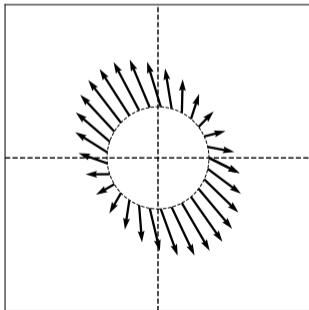
1. Think about (or draw) the output of  $\vec{f}$  for a handful of unit vector inputs.
  - ▶ Linear transformations are continuous so you can “interpolate”.
2. Find place(s) where the input vector and the output vector are parallel.



## Exercise

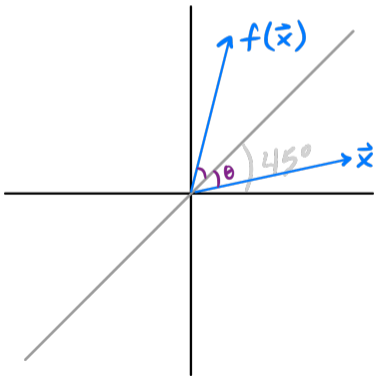
Draw as many (linearly independent) eigenvectors as you can:

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



## Exercise

Consider the linear transformation which mirrors its input over the line of  $45^\circ$ . Give two orthogonal eigenvectors of the transformation.



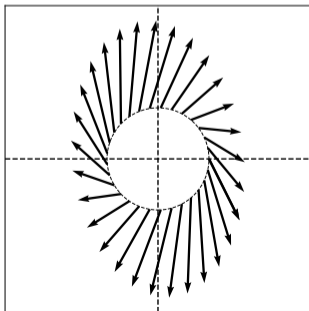
# Alternate Procedure: Guess and Check

1. Guess a vector  $\vec{x}$ .
2. Check that  $\vec{f}(\vec{x}) = \lambda\vec{x}$ .

## Exercise

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

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



## Caution!

- ▶ Not all matrices have even one eigenvector!<sup>2</sup>
- ▶ When does a matrix have multiple (linearly independent) eigenvectors?

---

<sup>2</sup>That is, with a *real-valued* eigenvalue.

# Symmetric Matrices

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

# The Spectral Theorem<sup>3</sup>

- ▶ **Theorem:** Let  $A$  be an  $n \times n$  *symmetric* matrix. Then there exist  $n$  eigenvectors of  $A$  which are all mutually orthogonal.

---

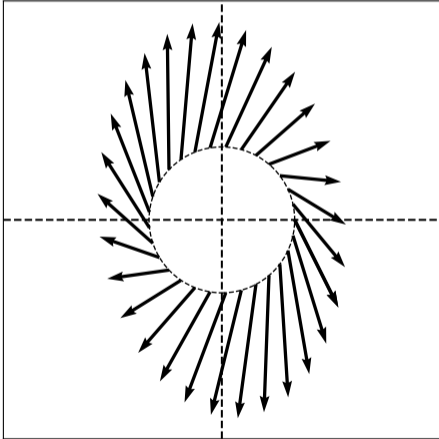
<sup>3</sup>for symmetric matrices

# What?

- ▶ What does the spectral theorem mean?
- ▶ What is an eigenvector, really?
- ▶ Why are they useful?

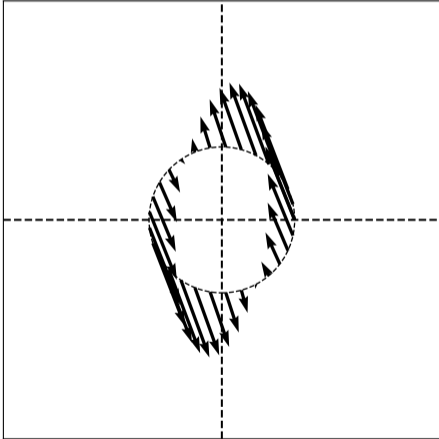


# Example Linear Transformation



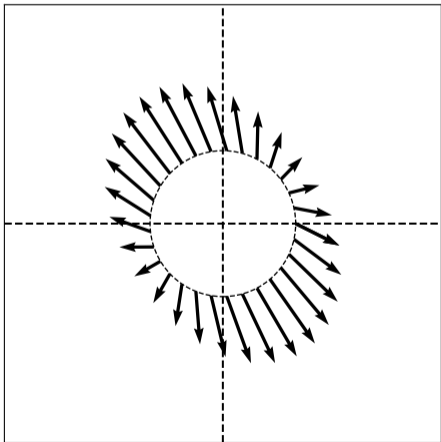
$$A = \begin{pmatrix} 5 & 5 \\ -10 & 12 \end{pmatrix}$$

# Example Linear Transformation



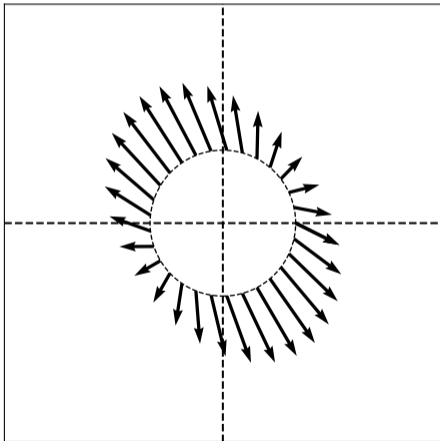
$$A = \begin{pmatrix} -2 & -1 \\ -5 & 3 \end{pmatrix}$$

# Example Symmetric Linear Transformation



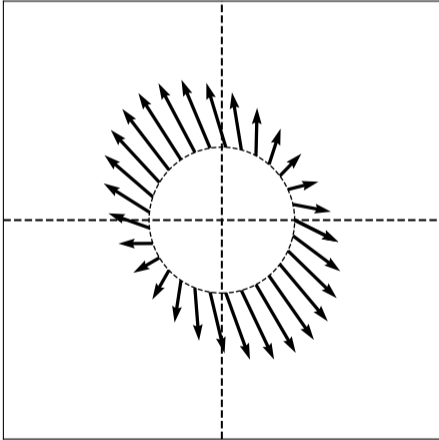
$$A = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$$

# Observation #1



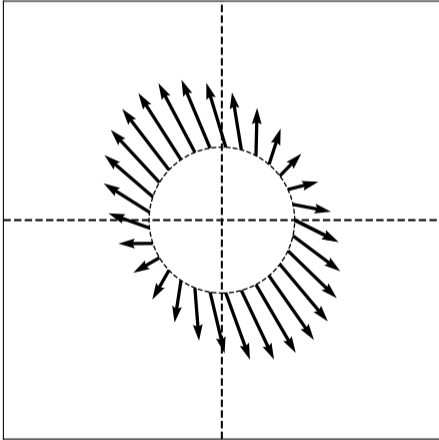
- ▶ Symmetric linear transformations have **axes of symmetry.**

## Observation #2



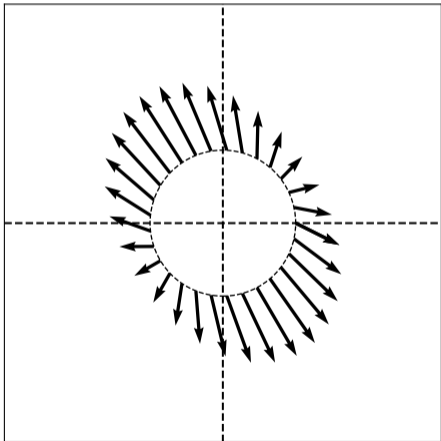
- ▶ The axes of symmetry are **orthogonal** to one another.

## Observation #3



- ▶ The action of  $\vec{f}$  along an axis of symmetry is simply to scale its input.

## Observation #4



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

## Main Idea

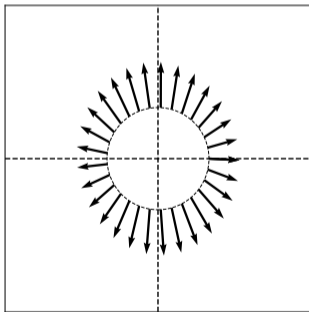
The **eigenvectors** of a symmetric linear transformation (matrix) are its axes of symmetry. The **eigenvalues** describe how much each axis of symmetry is scaled.



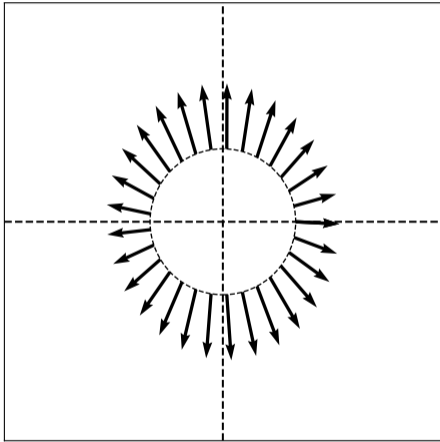
# Diagonal Matrices

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

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

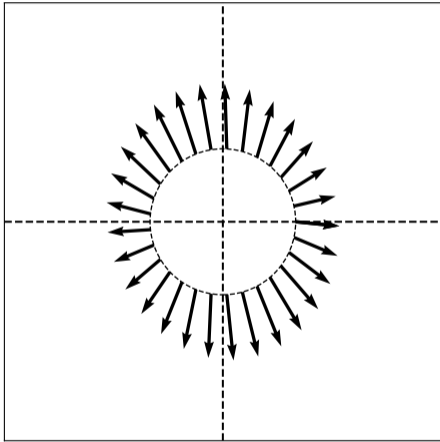


# Off-diagonal elements



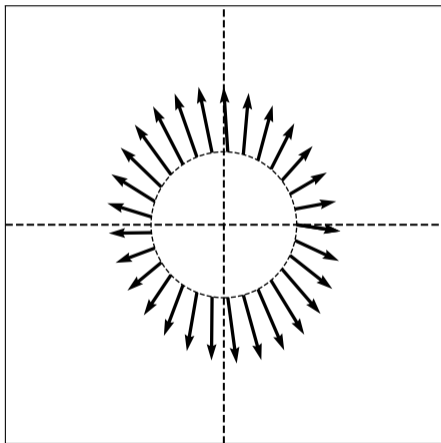
$$A = \begin{pmatrix} 2 & -0.1 \\ -0.1 & 5 \end{pmatrix}$$

# Off-diagonal elements



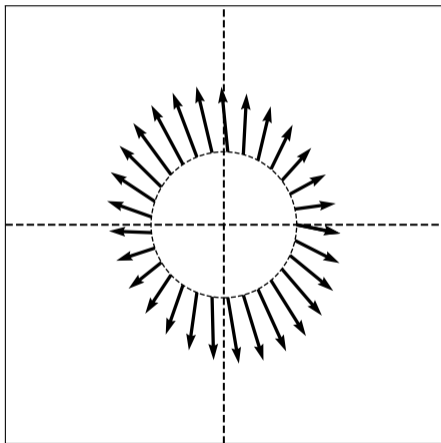
$$A = \begin{pmatrix} 2 & -0.2 \\ -0.2 & 5 \end{pmatrix}$$

# Off-diagonal elements



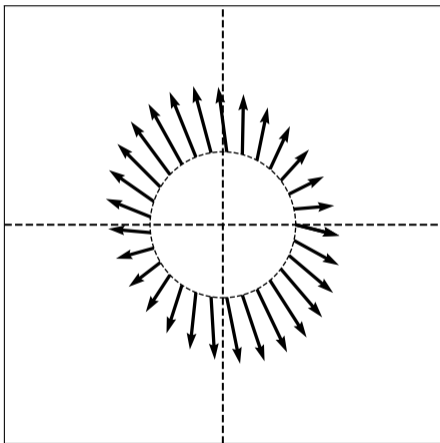
$$A = \begin{pmatrix} 2 & -0.3 \\ -0.3 & 5 \end{pmatrix}$$

# Off-diagonal elements



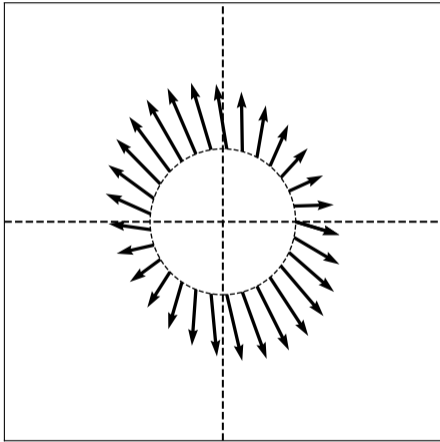
$$A = \begin{pmatrix} 2 & -0.4 \\ -0.4 & 5 \end{pmatrix}$$

# Off-diagonal elements



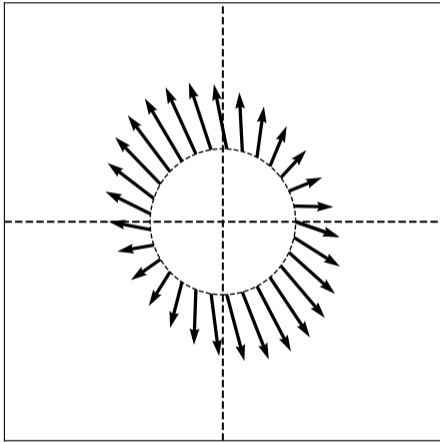
$$A = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 5 \end{pmatrix}$$

# Off-diagonal elements



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

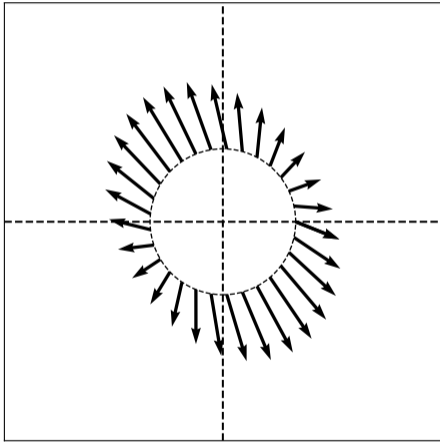
# Off-diagonal elements



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

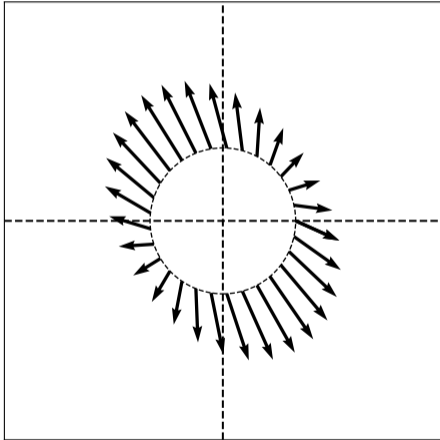


# Off-diagonal elements



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

# Off-diagonal elements



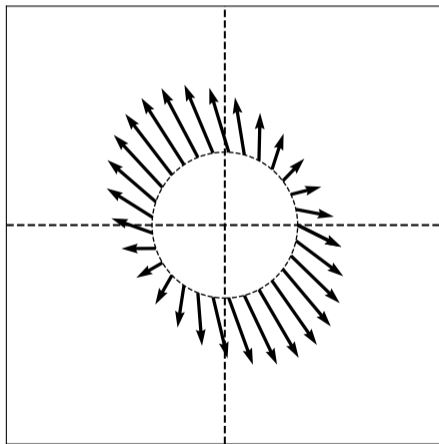
$$A = \begin{pmatrix} 2 & -0.9 \\ -0.9 & 5 \end{pmatrix}$$

# Non-Diagonal Symmetric Matrices

- ▶ When a symmetric matrix is not diagonal, its eigenvectors are not the standard basis vectors.
- ▶ But they can be used to form an orthonormal basis!

# The Spectral Theorem<sup>4</sup>

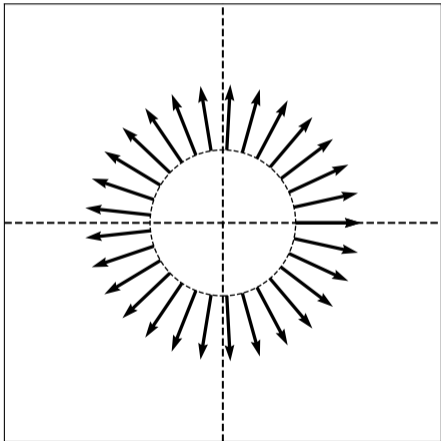
- ▶ **Theorem:** Let  $A$  be an  $n \times n$  symmetric matrix. Then there exist  $n$  eigenvectors of  $A$  which are all mutually orthogonal.



---

<sup>4</sup>for symmetric matrices

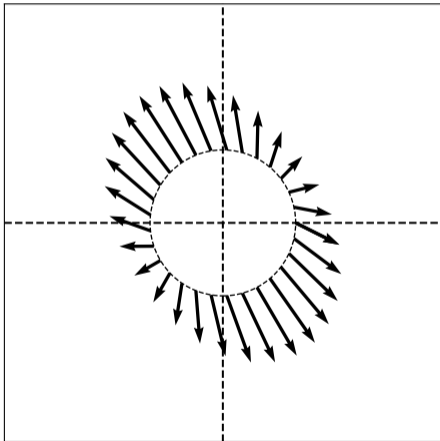
# What about total symmetry?



- ▶ Every vector is an eigenvector.

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

# Computing Eigenvectors



```
»» 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]]))
```