DSC 140B

Representation Learning

Lecture 06 | Part 1

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)

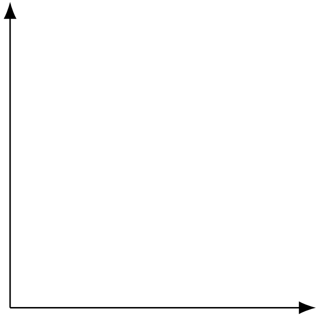
- Let $\vec{f}$ be a linear transformation. An **eigenvector** of $\vec{f}$ with **eigenvalue** λ is a nonzero vector $\vec{v}$ such that $f(\vec{v}) = \lambda\vec{v}$.

Importance

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

Geometric Interpretation

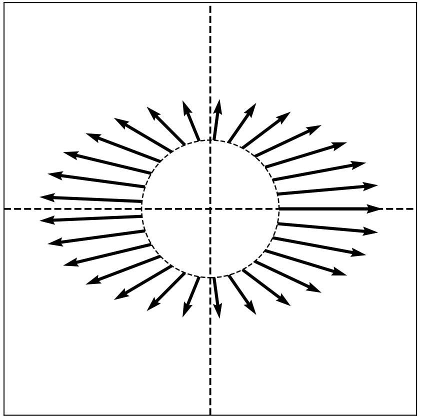
- When $\vec{f}$ is applied to one of its eigenvectors, $\vec{f}$ simply scales it.
  - Possibly by a negative amount.



Exercise

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

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



Finding Eigenvectors

numpy
eigh(·)

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

Procedure

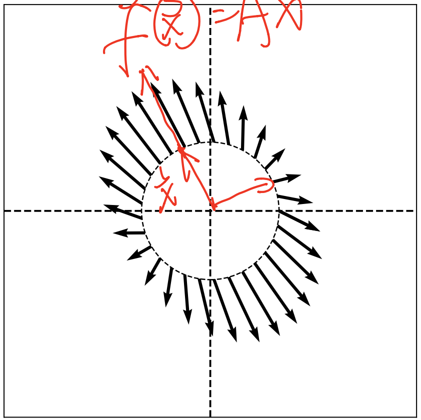
Given a matrix A (or transformation $\vec{f}$), to find an eigenvector “graphically”.

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:

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



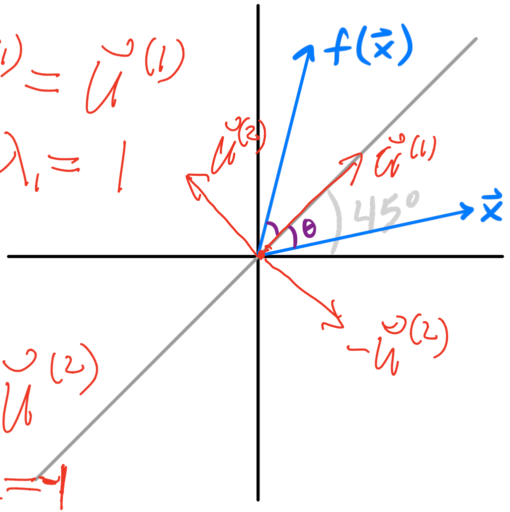
Exercise

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

$$A\vec{x} = \lambda \vec{x}$$

$$\vec{u}^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \vec{u}^{(2)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$A\vec{u}^{(1)} = \vec{u}^{(1)}, \quad \lambda_1 = 1$$



$$A\vec{u}^{(2)} = -\vec{u}^{(2)}, \quad \lambda_2 = -1$$

$$\vec{u}^{(1)} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \vec{u}^{(2)} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

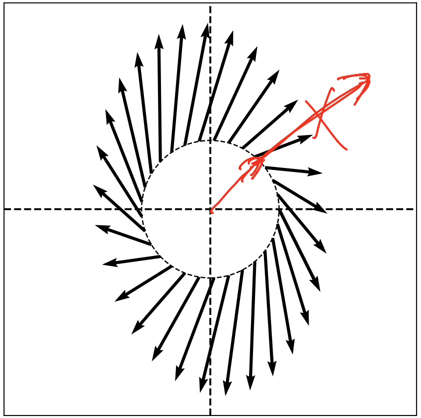
Alternate Procedure: Guess and Check

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![1]
- When does a matrix have multiple (linearly independent) eigenvectors?

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

Symmetric Matrices

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

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} d & b \\ c & a \end{pmatrix} \checkmark$$

$c = b$

The Spectral Theorem[2]

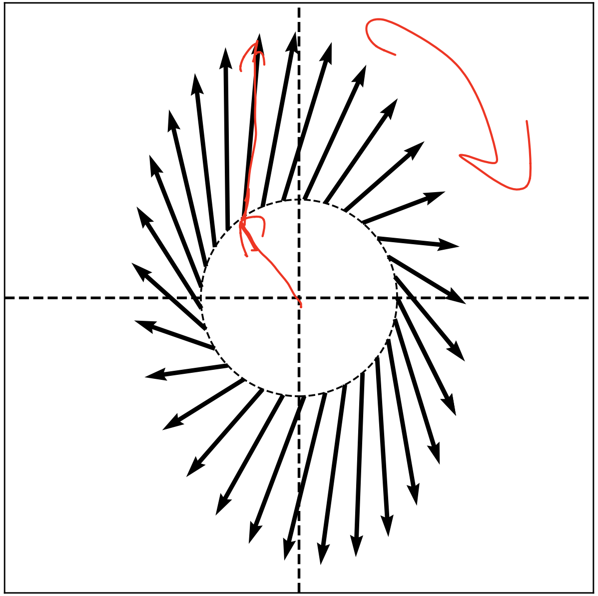
- **Theorem:** Let A be an $n \times n$ *symmetric* matrix. Then there exist n eigenvectors of A which are all mutually orthogonal.

[2] for symmetric matrices

What?

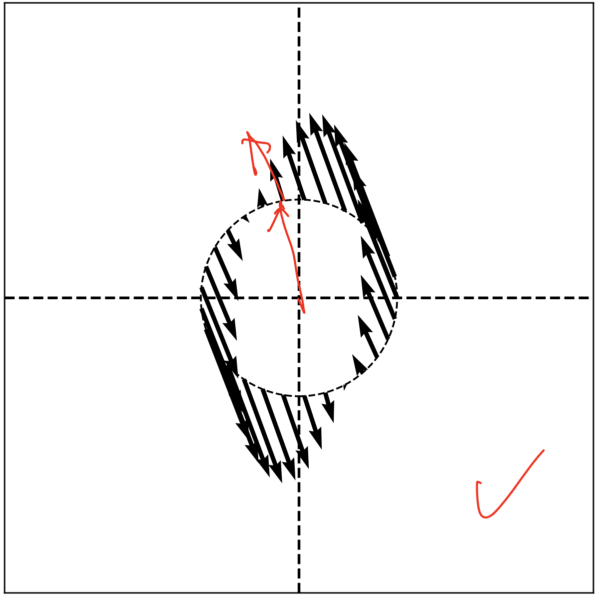
- What does the spectral theorem mean?
- What is an eigenvector, really?
- Why are they useful?

Example Linear Transformation



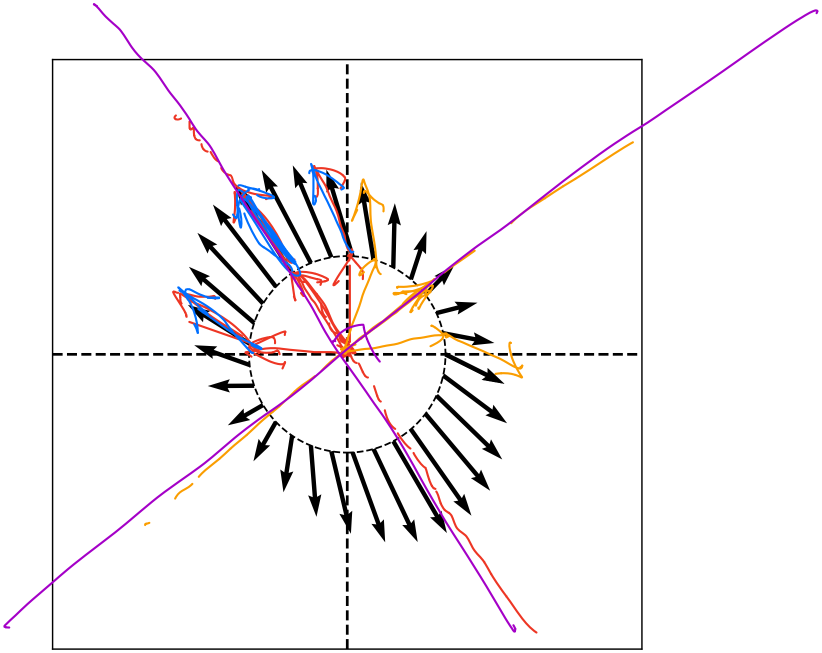
$$A = \begin{pmatrix} 5 & 5 \\ -10 & 12 \end{pmatrix}$$

Example Linear Transformation



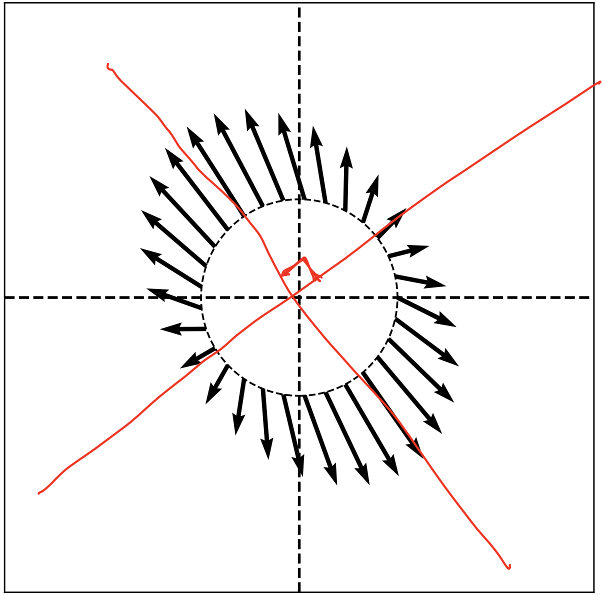
$$A = \begin{pmatrix} -2 & -1 \\ -5 & 3 \end{pmatrix}$$

Example Symmetric Linear Transformation



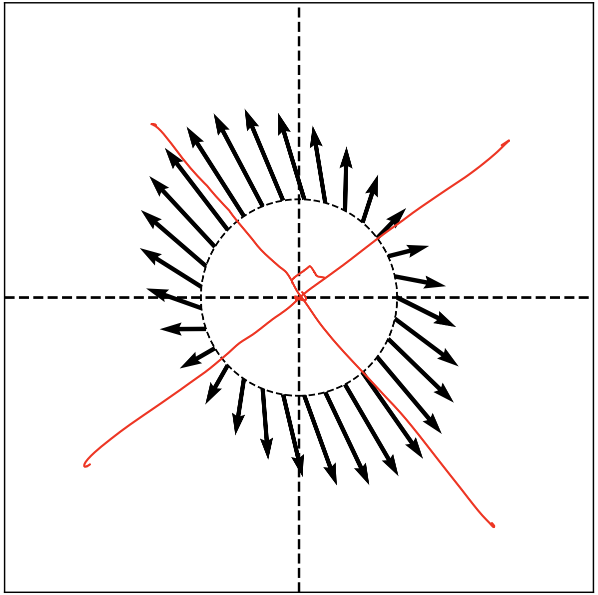
$$A = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$$

Observation #1



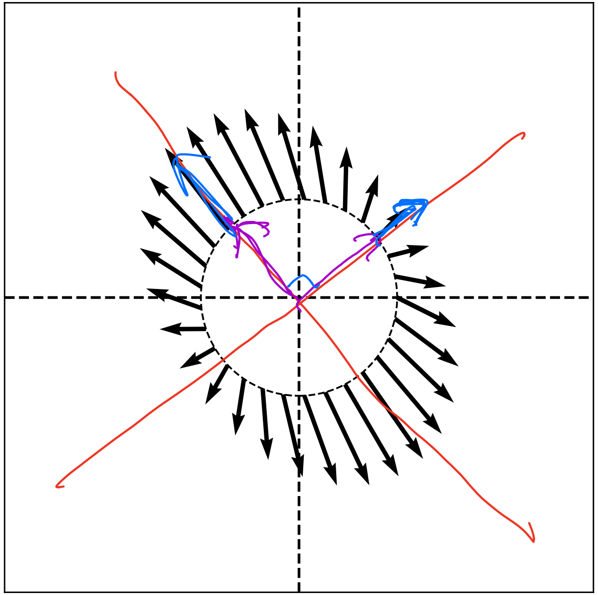
- Symmetric linear transformations have **axes of symmetry**.

Observation #2



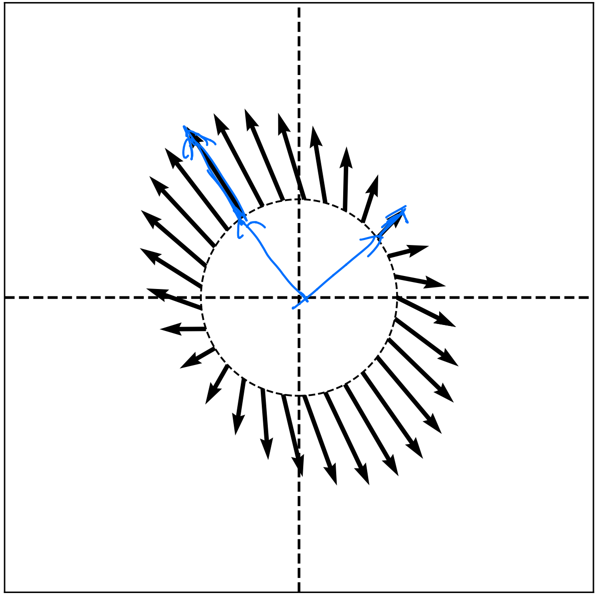
- The axes of symmetry are **orthogonal** to one another.

Observation #3

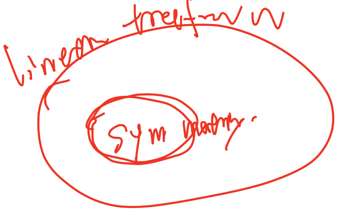


- The action of $\vec{f}$ along an axis of symmetry is simply to scale its input.

Observation #4



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



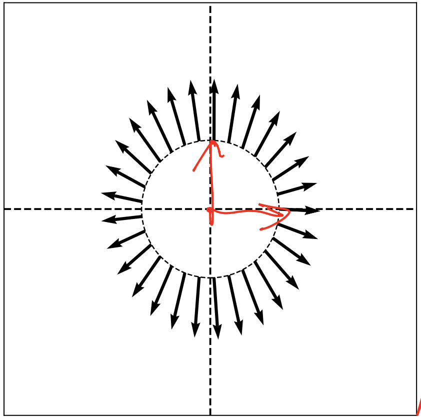
Main Idea

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

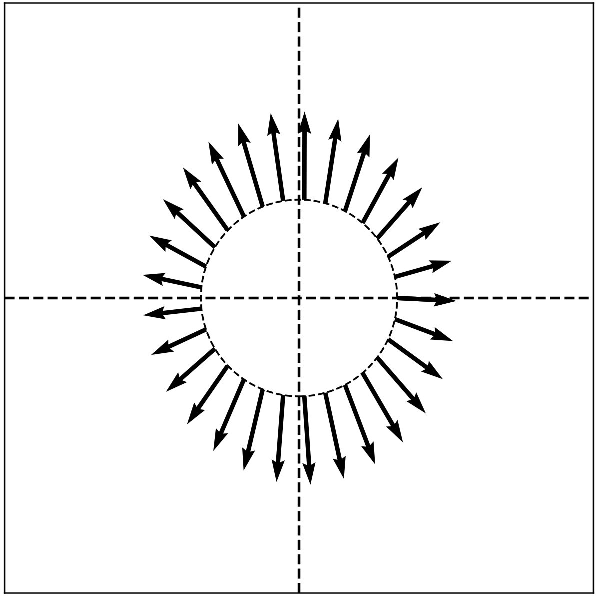
Diagonal Matrices

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

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

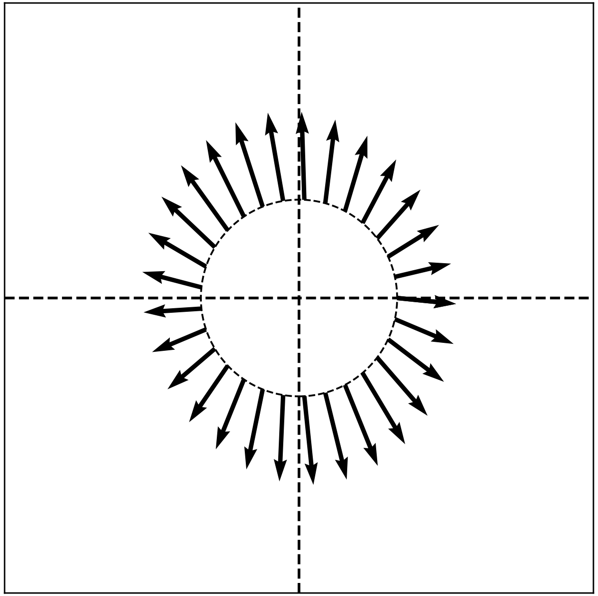


Off-diagonal elements



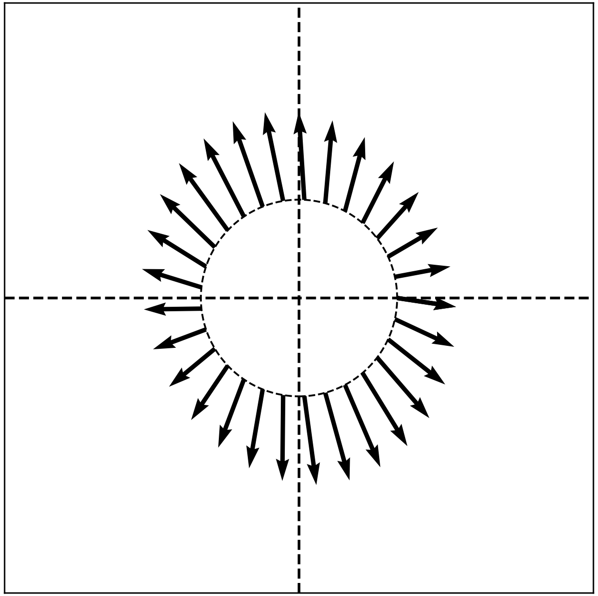
$$A = \begin{pmatrix} 2 & -0.1 \\ -0.1 & 5 \end{pmatrix}$$

Off-diagonal elements



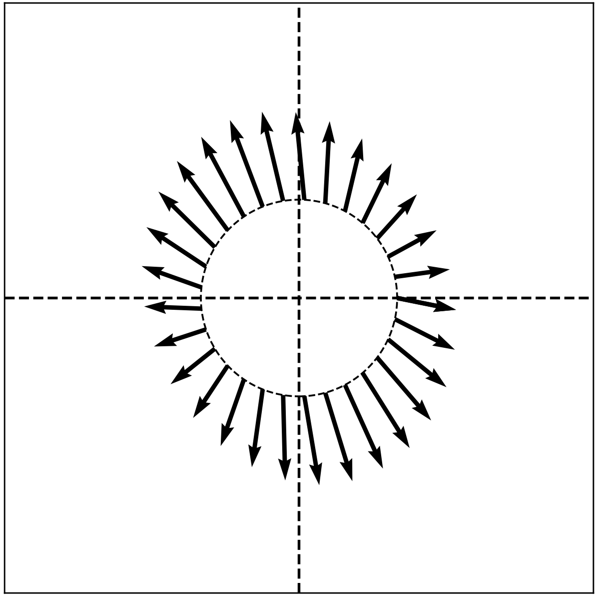
$$A = \begin{pmatrix} 2 & -0.2 \\ -0.2 & 5 \end{pmatrix}$$

Off-diagonal elements



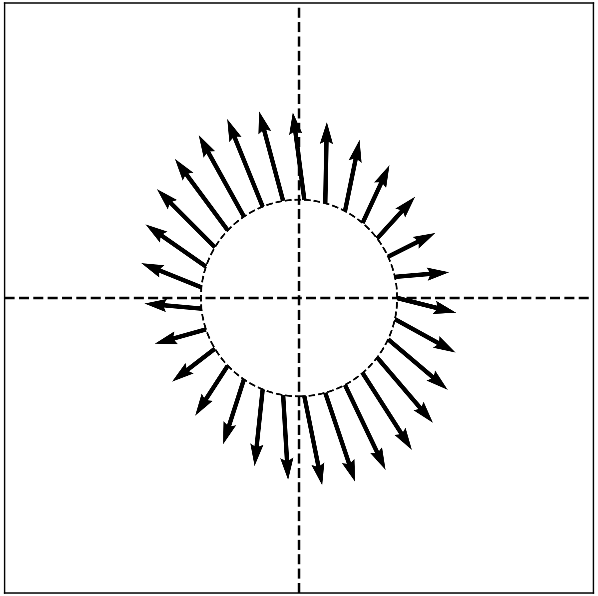
$$A = \begin{pmatrix} 2 & -0.3 \\ -0.3 & 5 \end{pmatrix}$$

Off-diagonal elements



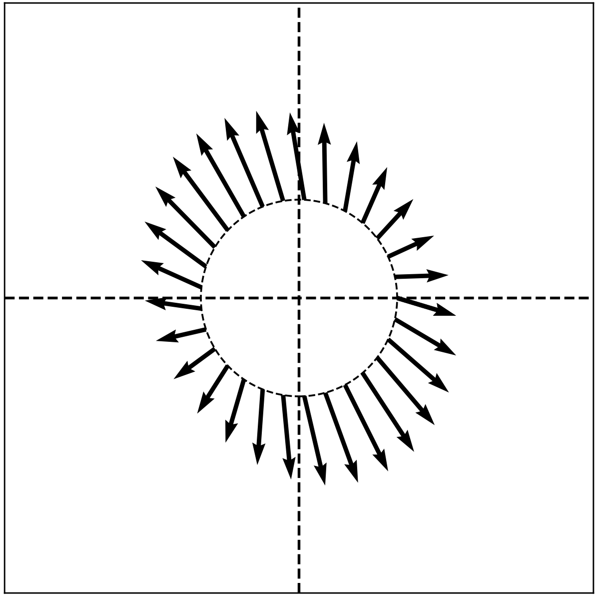
$$A = \begin{pmatrix} 2 & -0.4 \\ -0.4 & 5 \end{pmatrix}$$

Off-diagonal elements



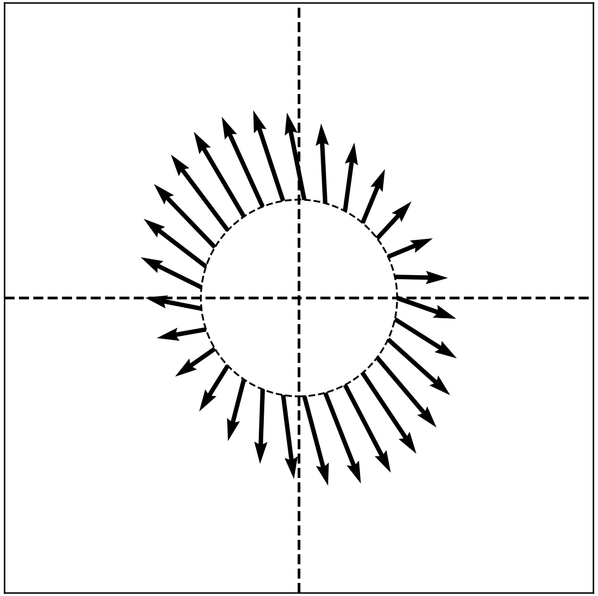
$$A = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 5 \end{pmatrix}$$

Off-diagonal elements



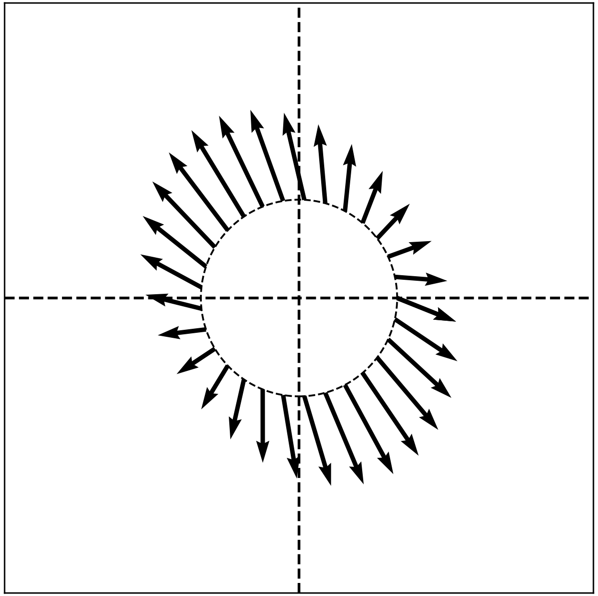
$$A = \begin{pmatrix} 2 & -0.6 \\ -0.6 & 5 \end{pmatrix}$$

Off-diagonal elements



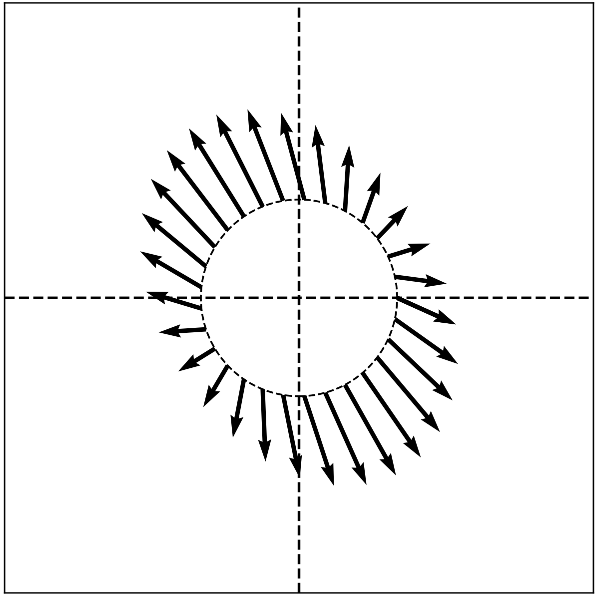
$$A = \begin{pmatrix} 2 & -0.7 \\ -0.7 & 5 \end{pmatrix}$$

Off-diagonal elements



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

Off-diagonal elements



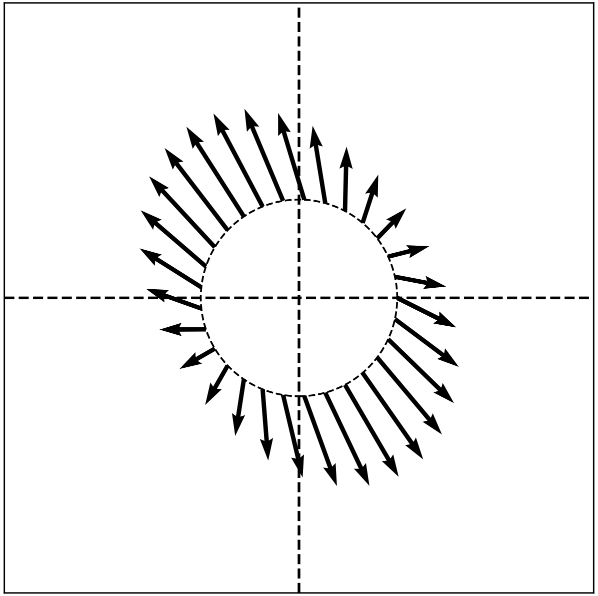
$$A = \begin{pmatrix} 2 & -0.9 \\ -0.9 & 5 \end{pmatrix}$$

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!

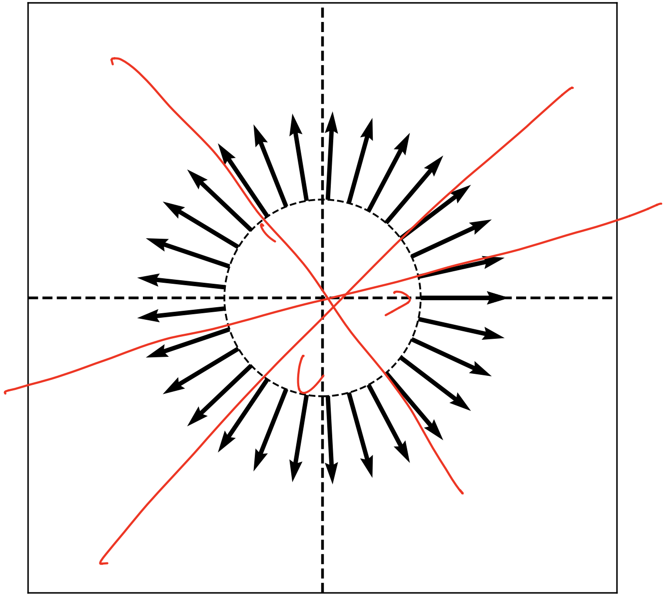
The Spectral Theorem[3]

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



[3] for symmetric matrices

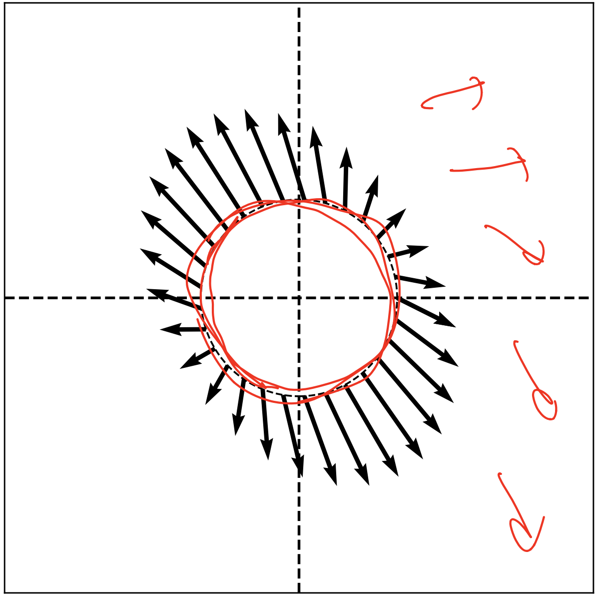
What about total symmetry?



- 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]]))
```

DSC 140B

Representation Learning

Lecture 06 | Part 2

Why are eigenvectors useful?

OK, but why are eigenvectors[4] useful?

1. Eigenvectors are nice “building blocks” (basis vectors).
2. Eigenvectors are **maximizers** (or minimizers). PCA “best”
3. Eigenvectors are **equilibria**.

[4] of symmetric matrices

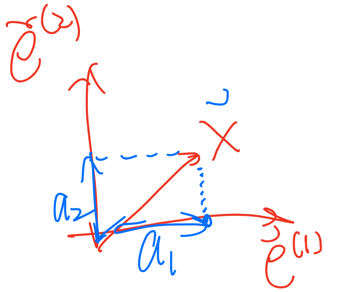
Vector Decomposition

- We can always "decompose" a vector $\vec{x}$ in terms of the basis vectors.

- With respect to the standard basis:

$$\vec{x} = a_1\hat{e}^{(1)} + a_2\hat{e}^{(2)} + \ldots + a_d\hat{e}^{(d)}$$

$$\begin{pmatrix} a_1 \\ \vdots \\ a_d \end{pmatrix}$$



$$\vec{x} = a_1\vec{e}^{(1)} + a_2\vec{e}^{(2)}$$

Eigendecomposition

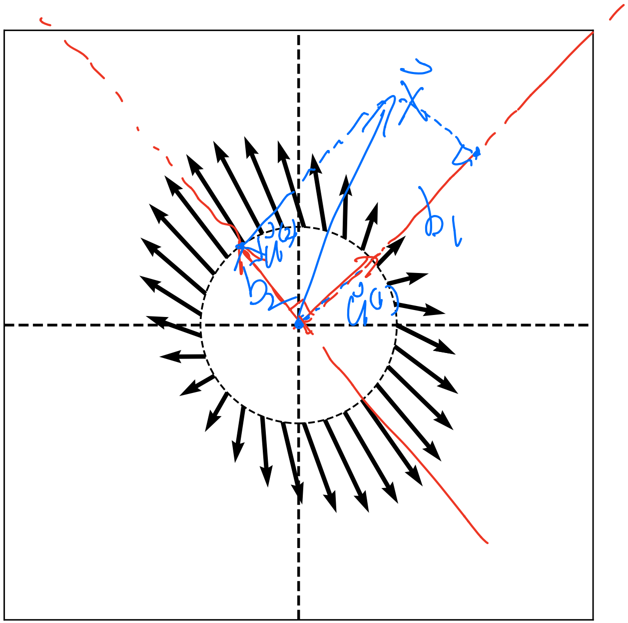
- If A is a symmetric matrix, we can pick d of its eigenvectors $\hat{u}^{(1)}, \ldots, \hat{u}^{(d)}$ to form an orthonormal basis.

- Any vector $\vec{x}$ can be written in terms of this basis.

- This is called its **eigendecomposition**:

$$\vec{x} = b_1 \hat{u}^{(1)} + b_2 \hat{u}^{(2)} + \ldots + b_d \hat{u}^{(d)}$$

Eigendecomposition



$$\vec{x} = b_1 \vec{u}^{(1)} + b_2 \vec{u}^{(2)}$$

Why?

- Compare working in the standard basis decomposition:

$$\begin{aligned} A\vec{x} &= A(a_1\hat{e}^{(1)} + a_2\hat{e}^{(2)} + \ldots + a_d\hat{e}^{(d)}) \\ &= a_1A\hat{e}^{(1)} + a_2A\hat{e}^{(2)} + \ldots + a_dA\hat{e}^{(d)} \end{aligned}$$

- To working with the eigendecomposition:

$A\vec{u} = \lambda\vec{u}$

$$\begin{aligned} A\vec{x} &= A(b_1\hat{u}^{(1)} + b_2\hat{u}^{(2)} + \ldots + b_d\hat{u}^{(d)}) \\ &= b_1A\hat{u}^{(1)} + b_2A\hat{u}^{(2)} + \ldots + b_dA\hat{u}^{(d)} \\ &= \lambda_1b_1\hat{u}^{(1)} + \lambda_2b_2\hat{u}^{(2)} + \ldots + \lambda_db_d\hat{u}^{(d)} \end{aligned}$$

Main Idea

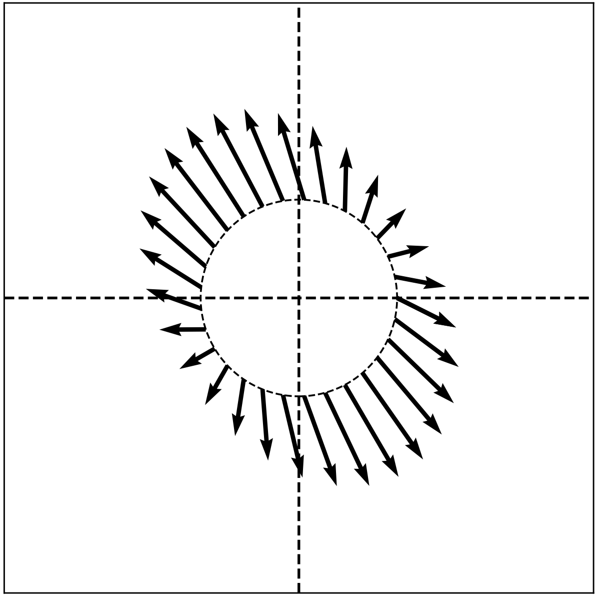
If A is a symmetrix matrix, an eigenbasis formed from its eigenvectors is an especially natural basis.

Eigenvectors as Optimizers

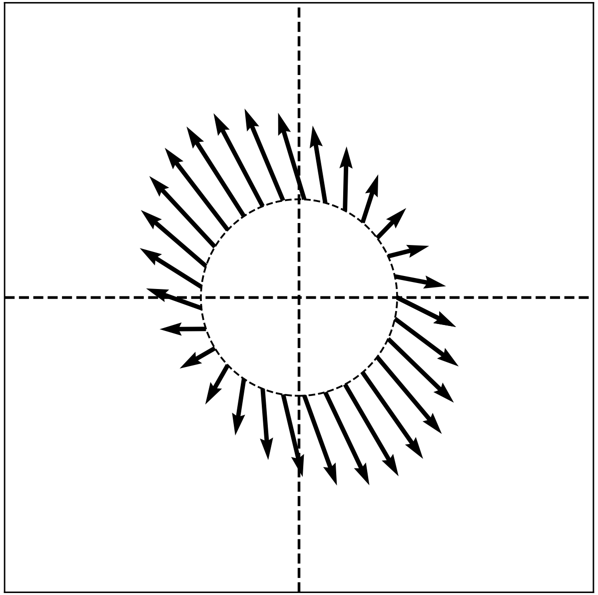
- Eigenvectors are the solutions to certain common optimization problems involving matrices / linear transformations.

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.

Main Idea

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

Main Idea

To **minimize** $\|\vec{f}(\vec{x})\|$ over unit vectors, pick $\vec{x}$ to be an eigenvector of $\vec{f}$ with the smallest eigenvalue (in abs. value).

Proof

Show that the maximizer of $\|A\vec{x}\|$ s.t., $\|\vec{x}\| = 1$ is the top eigenvector of A.

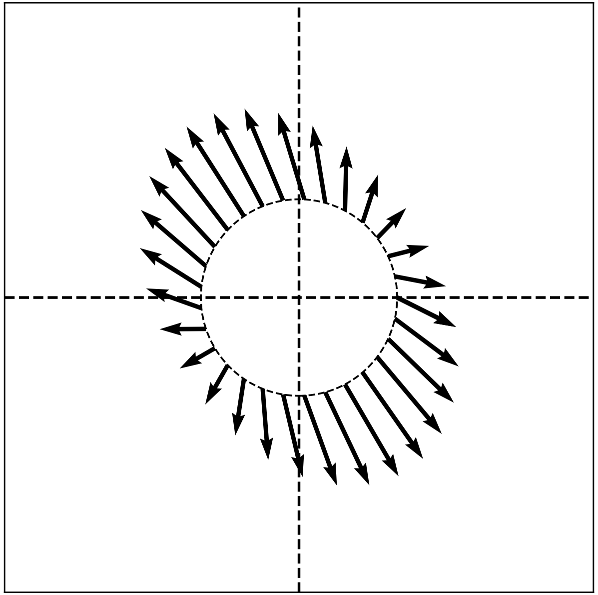
Corollary

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

Example

- Maximize $4x_1^2 + 2x_2^2 + 3x_1x_2$ subject to $x_1^2 + x_2^2 = 1$

Observation #2



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

The Power Method

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