# DSC 1408 Representation Learning

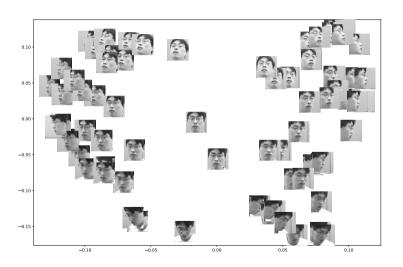
Lecture 02 | Part 1

Why Linear Algebra?

### **Last Time**



# **Last Time**



# **Dimensionality Reduction**

- This is an example of dimensionality reduction:
  - ► Input: vectors in  $\mathbb{R}^{10,000}$ .
  - ▶ Output: vectors in  $\mathbb{R}^2$ .
- The method which produced this result is called Laplacian Eigenmaps.
- How does it work?

## A Preview of Laplacian Eigenmaps

To reduce dimensionality from d to d':

- 1. Create an undirected similarity graph G
  - ightharpoonup Each vector in  $\mathbb{R}^d$  becomes a node in the graph.
  - ightharpoonup Make edge (u, v) if u and v are "close"
- 2. Form the graph Laplacian matrix, L:
  - Let A be the adjacency matrix, D be the degree matrix.
  - ▶ Define the graph Laplacian matrix, L = D A.
- 3. Compute d' eigenvectors of L.
  - Each eigenvector gives one new feature.

# Why eigenvectors?

- We will cover Laplacian Eigenmaps in much greater detail.
- For now: why do eigenvectors appear here?
  - What are eigenvectors?
  - How are they useful?
  - Why is linear algebra important in ML?

# DSC 1408 Representation Learning

Lecture 02 | Part 2

**Coordinate Vectors** 

#### **Coordinate Vectors**

We can write a vector  $\vec{x} \in \mathbb{R}^d$  as a coordinate vector:

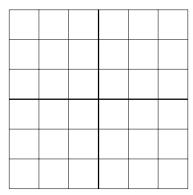
$$\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

# **Example**

Χ̈́	=	$\begin{pmatrix} 2 \\ -3 \end{pmatrix}$
ÿ	=	$\begin{pmatrix} 0 \\ 2 \end{pmatrix}$

### **Standard Basis**

- Writing a vector in coordinate form requires choosing a basis.
- ► The "default" is the **standard basis**:  $\hat{e}^{(1)},...,\hat{e}^{(d)}$ .



#### **Standard Basis**

When we write  $\vec{x} = (x_1, ..., x_d)^T$ , we mean that  $\vec{x} = x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)} + ... x_d \hat{e}^{(d)}$ .

Example:  $\vec{x} = (3, -2)^T$ 

#### **Standard Basis Coordinates**

► In coordinate form:

$$\hat{e}^{(i)} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

where the 1 appears in the *i*th place.

#### **Exercise**

Let  $\vec{x} = (3, 7, 2, -5)^T$ . What is  $\vec{x} \cdot \hat{e}^{(4)}$ ?

#### **Recall: the Dot Product**

► The **dot product** of  $\vec{u}$  and  $\vec{v}$  is defined as:

$$\vec{u} \cdot \vec{v} = ||\vec{u}|| ||\vec{v}|| \cos \theta$$

where  $\theta$  is the angle between  $\vec{u}$  and  $\vec{v}$ .

 $\vec{u} \cdot \vec{v} = 0$  if and only if  $\vec{u}$  and  $\vec{v}$  are orthogonal

## **Dot Product (Coordinate Form)**

► In terms of coordinate vectors:

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v}$$

$$= \begin{pmatrix} u_1 & u_2 & \cdots & u_d \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \cdots \\ v_d \end{pmatrix}$$

$$=$$

This definition assumes the standard basis.

# **Example**

$$\begin{pmatrix} 3 \\ 7 \\ 2 \\ -5 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} =$$

## What does ChatGPT say?



I will ask you a simple vector math question. Respond with the answer, and nothing else.

Do not provide an explanation. Your answer should be a number.

Let  $\ x = (3,7,2,-5)^T$ . What is  $\ x \in x \cdot (4)$ ?



-5



#### **Other Bases**

- The standard basis is not the only basis.
- Sometimes more convenient to use another.

# **Example**



#### **Orthonormal Bases**

- Orthonormal bases are particularly nice.
- A set of vectors  $\hat{u}^{(1)}$ , ...,  $\hat{u}^{(d)}$  forms an **orthonomal** basis  $\mathcal{U}$  for  $R^d$  if:
  - They are mutually orthogonal:  $\hat{u}^{(i)} \cdot \hat{u}^{(j)} = 0$ .
  - ► They are all unit vectors:  $\|\hat{u}^{(i)}\| = 1$ .

# **Example**

#### **Coordinate Vectors**

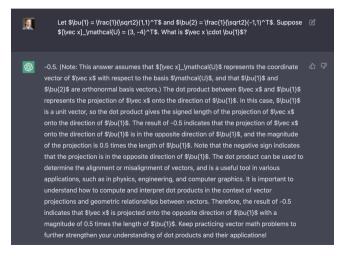
- A vector's coordinates depend on the basis used.
- If we are using the basis  $\mathcal{U} = \{\hat{u}^{(1)}, \hat{u}^{(2)}\}\$ , then  $\vec{x} = (x_1, x_2)^T$  means  $\vec{x} = x_1 \hat{u}^{(1)} + x_2 \hat{u}^{(2)}$ .
- We will write  $[\vec{x}]_{\mathcal{U}} = (x_1, ..., x_d)^T$  to denote that the coordinates are with respect to the basis  $\mathcal{U}$ .

#### **Exercise**

Let 
$$\hat{u}^{(1)} = \frac{1}{\sqrt{2}} (1, 1)^T$$
 and  $\hat{u}^{(2)} = \frac{1}{\sqrt{2}} (-1, 1)^T$ . Suppose

Let 
$$\hat{u}^{(1)} = \frac{1}{\sqrt{2}} (1,1)^T$$
 and  $\hat{u}^{(2)} = \frac{1}{\sqrt{2}} (-1,1)^T$ . Suppose  $[\vec{x}]_{\mathcal{U}} = (3,-4)^T$ . What is  $\vec{x} \cdot \hat{u}^{(1)}$ ?

# What did ChatGPT say?

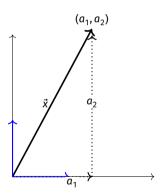


#### **Exercise**

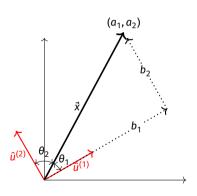
Consider 
$$\vec{x} = (2,2)^T$$
 and let  $\hat{u}^{(1)} = \frac{1}{2}(1,1)^T$  and  $\hat{u}^{(2)} = \frac{1}{2}(1,1)^T$ 

Consider 
$$\vec{x} = (2,2)^T$$
 and let  $\hat{u}^{(1)} = \frac{1}{\sqrt{2}}(1,1)^T$  and  $\hat{u}^{(2)} = \frac{1}{\sqrt{2}}(-1,1)^T$ . What is  $[\vec{x}]_{\mathcal{U}}$ ?

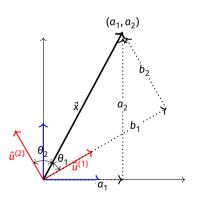
- ► How do we compute the coordinates of a vector in a new basis, U?
- Some trigonometry is involved.
- **Key Fact**:  $\vec{a} \cdot \vec{b} = ||\vec{a}|| ||\vec{b}|| \cos \theta$



- Suppose we know  $\vec{x} = (a_1, a_2)^T$  w.r.t. standard basis.
- Then  $\vec{x} = a_1 \hat{e}^{(1)} + a_2 \hat{e}^{(2)}$



- Want to write:  $\vec{x} = b_1 \hat{u}^{(1)} + b_2 \hat{u}^{(2)}$
- Need to find  $b_1$  and  $b_2$ .



- Exercise: Solve for  $b_1$ , writing the answer as a dot product.
- Hint: cos θ = adjacent/hypotenuse

- Let  $\mathcal{U} = {\hat{u}^{(1)}, ..., \hat{u}^{(d)}}$  be an orthonormal basis.
- ▶ The coordinates of  $\vec{x}$  w.r.t.  $\mathcal{U}$  are:

$$[\vec{x}]_{\mathcal{U}} = \begin{pmatrix} \vec{x} \cdot \hat{u}^{(1)} \\ \vec{x} \cdot \hat{u}^{(2)} \\ \vdots \\ \vec{x} \cdot \hat{u}^{(d)} \end{pmatrix}$$

#### **Exercise**

Suppose 
$$\vec{x} = (2, 1)^T$$
 and let  $\hat{\mu}^{(1)} = \frac{1}{2} (1, 1)^T$  and  $\hat{\mu}^{(2)}$ 

Suppose 
$$\vec{x} = (2, 1)^T$$
 and let  $\hat{u}^{(1)} = \frac{1}{\sqrt{2}}(1, 1)^T$  and  $\hat{u}^{(2)} = \frac{1}{\sqrt{2}}(-1, 1)^T$ . What is  $[\vec{x}]_{\mathcal{U}}$ ?

#### **Exercise**

What is  $[\vec{x}]_{i,i}$ ?

Let 
$$\vec{x} = (-1, 4)^T$$
 and suppose:

et 
$$x = (-1, 4)^n$$
 and suppose:

 $\hat{u}^{(1)} \cdot \hat{e}^{(2)} = -2$ 

$$\hat{q}^{(1)} \cdot \hat{e}^{(1)} = 3$$
  $\hat{q}^{(2)} \cdot \hat{e}^{(1)} = -1$ 

$$\hat{u}^{(2)}$$

$$u \cdot v \cdot e \cdot v = -1$$

$$\hat{\boldsymbol{u}}^{(2)}\cdot\hat{\boldsymbol{e}}^{(2)}=5$$

$$(2) = 5$$

# DSC 1408 Representation Learning

Lecture 02 | Part 3

**Functions of a Vector** 

#### **Functions of a Vector**

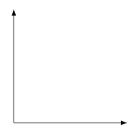
- In ML, we often work with functions of a vector:  $f: \mathbb{R}^d \to \mathbb{R}^{d'}$ .
- Example: a prediction function,  $H(\vec{x})$ .
- Functions of a vector can return:
  - ightharpoonup a number:  $f: \mathbb{R}^d \to \mathbb{R}^1$
  - ightharpoonup a vector  $\vec{f}: \mathbb{R}^d \to \mathbb{R}^{d'}$
  - something else?

#### **Transformations**

- A transformation  $\vec{f}$  is a function that takes in a vector, and returns a vector of the same dimensionality.
- ▶ That is,  $\vec{f} : \mathbb{R}^d \to \mathbb{R}^d$ .

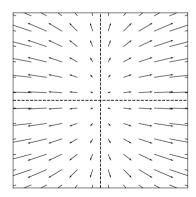
## **Visualizing Transformations**

- A transformation is a vector field.
  - Assigns a vector to each point in space.
  - ► Example:  $\vec{f}(\vec{x}) = (3x_1, x_2)^T$



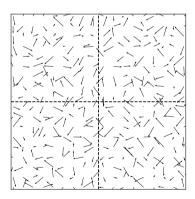
## **Example**

$$\vec{f}(\vec{x}) = (3x_1, x_2)^T$$



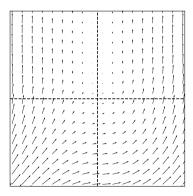
## **Arbitrary Transformations**

Arbitrary transformations can be quite complex.



## **Arbitrary Transformations**

Arbitrary transformations can be quite complex.



#### **Linear Transformations**

- Luckily, we often<sup>1</sup> work with simpler, linear transformations.
- ► A transformation *f* is linear if:

$$\vec{f}(\alpha \vec{x} + \beta \vec{y}) = \alpha \vec{f}(\vec{x}) + \beta \vec{f}(\vec{y})$$

<sup>&</sup>lt;sup>1</sup>Sometimes, just to make the math tractable!

## **Checking Linearity**

► To check if a transformation is linear, use the definition.

**Example:**  $\vec{f}(\vec{x}) = (x_2, -x_1)^T$ 

#### **Exercise**

Let  $\vec{f}(\vec{x}) = (x_1 + 3, x_2)$ . Is  $\vec{f}$  a linear transformation?

## **Implications of Linearity**

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

$$\begin{split} \vec{f}(\vec{x}) &= \vec{f}(x_1 \hat{e}^{(1)} + x_2 \hat{e}^{(2)}) \\ &= x_1 \vec{f}(\hat{e}^{(1)}) + x_2 \vec{f}(\hat{e}^{(2)}) \end{split}$$

▶ I.e.,  $\vec{f}$  is **totally determined** by what it does to the basis vectors.

# The Complexity of Arbitrary Transformations

- Suppose f is an arbitrary transformation.
- ► I tell you  $\vec{f}(\hat{e}^{(1)}) = (2,1)^T$  and  $\vec{f}(\hat{e}^{(2)}) = (-3,0)^T$ .
- $\vdash \text{I tell you } \vec{x} = (x_1, x_2)^T.$
- ▶ What is  $\vec{f}(\vec{x})$ ?

# The Simplicity of Linear Transformations

- Suppose f is a linear transformation.
- ► I tell you  $\vec{f}(\hat{e}^{(1)}) = (2,1)^T$  and  $\vec{f}(\hat{e}^{(2)}) = (-3,0)^T$ .
- $\vdash \text{I tell you } \vec{x} = (x_1, x_2)^T.$
- ▶ What is  $\vec{f}(\vec{x})$ ?

#### **Exercise**

- Suppose f is a linear transformation.
- I tell you  $\vec{f}(\hat{e}^{(1)}) = (2,1)^T$  and  $\vec{f}(\hat{e}^{(2)}) = (-3,0)^T$ . I tell you  $\vec{x} = (3,-4)^T$ .
- ▶ What is  $\vec{f}(\vec{x})$ ?

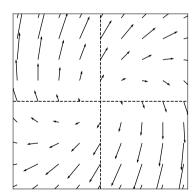
### **Key Fact**

- Linear functions are determined **entirely** by what they do on the basis vectors.
- I.e., to tell you what f does, I only need to tell you  $\vec{f}(\hat{e}^{(1)})$  and  $\vec{f}(\hat{e}^{(2)})$ .
- This makes the math easy!



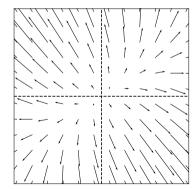
## **Example Linear Transformation**

$$\vec{f}(\vec{x}) = (x_1 + 3x_2, -3x_1 + 5x_2)^T$$



# Another Example Linear Transformation

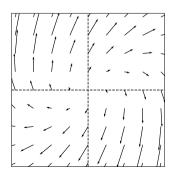
$$\vec{f}(\vec{x}) = (2x_1 - x_2, -x_1 + 3x_2)^T$$

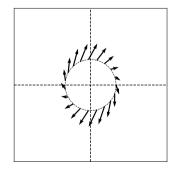


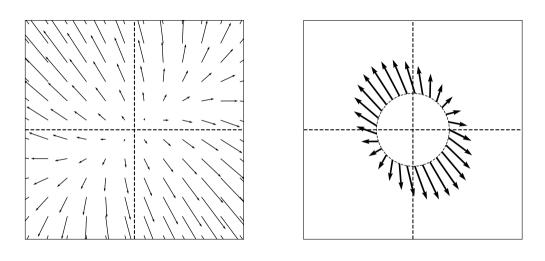
#### **Note**

Because of linearity, along any given direction  $\vec{f}$  changes only in scale.

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







### **Linear Transformations and Bases**

We have been writing transformations in coordinate form. For example:

$$\vec{f}(\vec{x}) = (x_1 + x_2, x_1 - x_2)^T$$

- ► To do so, we assumed the **standard basis**.
- If we use a different basis, the formula for  $\vec{f}$  changes.

#### **Example**

- Suppose that in the standard basis,  $\vec{f}(\vec{x}) = (x_1 + x_2, x_1 x_2)^T$ .
- Let  $\hat{u}^{(1)} = \frac{1}{\sqrt{2}} (1,1)^T$  and  $\hat{u}^{(2)} = \frac{1}{\sqrt{2}} (-1,1)^T$ .
- ► Write  $[\vec{x}]_{t/t} = (z_1, z_2)^T$ .
- ▶ What is  $[\vec{f}(\vec{x})]_{\mathcal{U}}$  in terms of  $z_1$  and  $z_2$ ?

## DSC 140B Representation Learning

Lecture 02 | Part 4

**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.
- ightharpoonup 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  $\vec{f}$

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

$$\vec{f}(\hat{e}^{(2)}) = 2\hat{e}^{(1)} = \begin{pmatrix} 2\\0 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 2\\3 & 0 \end{pmatrix}$$

<sup>&</sup>lt;sup>2</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  $f: \mathbb{R}^n \to \mathbb{R}^n$ .

## What is matrix multiplication?

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

## 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...

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

## **Matrix Multiplication**

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

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

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

$$A = \begin{pmatrix} \uparrow & \uparrow & \uparrow & \uparrow \\ \vec{f}(\hat{e}^{(1)}) & \vec{f}(\hat{e}^{(2)}) & \vec{f}(\hat{e}^{(3)}) \end{pmatrix}$$

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

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

## **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  $\vec{f}: \mathbb{R}^n \to \mathbb{R}^n$ .

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**

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

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

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

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

$$A\vec{x} =$$

#### **Note**

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

#### **Note**

- ightharpoonup 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**:

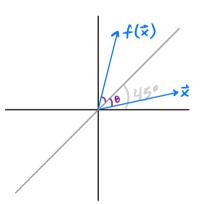
$$\begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{f}(\hat{e}^{(1)}) & \vec{f}(\hat{e}^{(2)}) & \cdots & \vec{f}(\hat{e}^{(d)}) \\ \downarrow & \downarrow & \downarrow \end{pmatrix}$$

ightharpoonup With respect to basis  $\mathcal{U}$ :

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

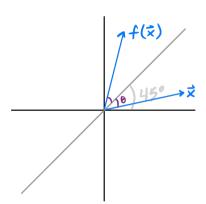
### **Matrices in Other Bases**

Consider the transformation  $\vec{f}$  which "mirrors" a vector over the line of 45°.



What is its matrix in the standard basis?

#### **Matrices in Other Bases**



Let 
$$\hat{u}^{(1)} = \frac{1}{\sqrt{2}}(1,1)$$

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

- $\vdash [\vec{f}(\hat{u}^{(2)})]_{i,j}$ ?
- What is the matrix?