CSE 455 Section 1 - Linear Algebra Primer

Nishat Anjum Khan

Outline

- Vectors and matrices
 - Basic Matrix Operations
 - Determinants, norms, trace
 - Special Matrices
- Transformation Matrices
 - Homogeneous coordinates
 - Translation
- Matrix inverse
- Matrix rank
- Matrix Calculus

Outline

- Vectors and matrices
 - Basic Matrix Operations
 - Determinants, norms, trace
 - Special Matrices
- Transformation Matrices
 - Homogeneous coordinates
 - Translation
- Matrix inverse
- Matrix rank
- Matrix Calculus

Vectors and matrices are just collections of ordered numbers that represent something: movements in space, scaling factors, pixel brightness, etc. We'll define some common uses and standard operations on them.

Vector

ullet A column vector $\mathbf{v} \in \mathbb{R}^{n imes 1}$ where

$$\mathbf{v} = egin{bmatrix} v_1 \ v_2 \ dots \ v_n \end{bmatrix}$$

• A row vector $\mathbf{v}^T \in \mathbb{R}^{1 \times n}$ where

$$\mathbf{v}^T = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$$

 ${\cal T}$ denotes the transpose operation

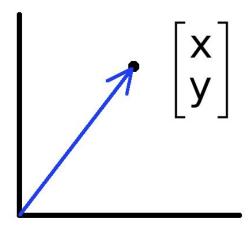
Vector

We'll default to column vectors in this class

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

- You'll want to keep track of the orientation of your vectors when programming in python
- You can transpose a vector V in python by writing V.t. (But in class materials, we will **always** use V^T to indicate transpose, and we will use V' to mean "V prime")

Vectors have two main uses



- Data (pixels, gradients at an image keypoint, etc) can also be treated as a vector.
- 0 255 178 122 217 34

- Vectors can represent an offset in 2D or 3D space.
- Points are just vectors from the origin.
- Such vectors don't have a geometric interpretation, but calculations like "distance" can still have value.

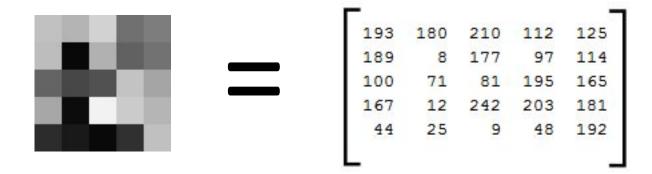
Matrix

• A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is an array of numbers with size by , i.e. m rows and n columns.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

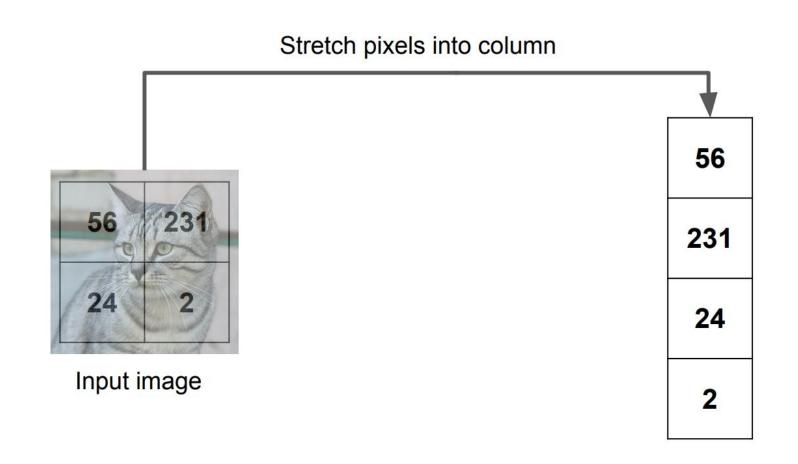
• If m = n, we say that ${\bf A}$ is square.

Images



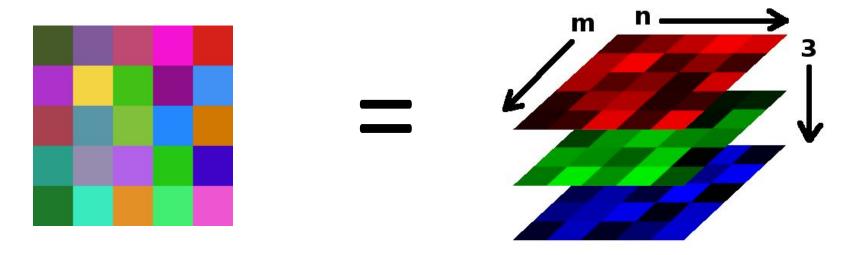
- Python represents an image as a matrix of pixel brightnesses
- Note that the upper left corner is [y,x] = (0,0)

Images as both a matrix as well as a vector



Color Images

- Grayscale images have one number per pixel, and are stored as an m × n matrix.
- Color images have 3 numbers per pixel red, green, and blue brightnesses (RGB)
- Stored as an m × n × 3 matrix



Basic Matrix Operations

- We will discuss:
 - Addition
 - Scaling
 - Dot product
 - Multiplication
 - Transpose
 - Inverse / pseudoinverse
 - Determinant / trace

Addition

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} a+1 & b+2 \\ c+3 & d+4 \end{bmatrix}$$

 Can only add a matrix with matching dimensions, or a scalar.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + 7 = \begin{bmatrix} a+7 & b+7 \\ c+7 & d+7 \end{bmatrix}$$

Scaling

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times 3 = \begin{bmatrix} 3a & 3b \\ 3c & 3d \end{bmatrix}$$

Vectors

• Norm
$$||x||_2 = \sqrt{\sum_{i=1}^n x_i^2}$$
.

- More formally, a norm is any function $f: \mathbb{R}^n \to \mathbb{R}$ that satisfies 4 properties:
- Non-negativity: For all $x \in \mathbb{R}^n$, $f(x) \geq 0$
- **Definiteness**: f(x) = 0 if and only if x = 0.
- Homogeneity: For all $x \in \mathbb{R}^n$, $t \in \mathbb{R}$, f(tx) = |t| f(x)
- Triangle inequality: For all $x, y \in \mathbb{R}^n, \ f(x+y) \leq f(x) + f(y)$

Norms

Example Norms

$$||x||_1 = \sum_{i=1}^n |x_i|$$

$$||x||_{\infty} = \max_{i} |x_{i}|.$$

• General ℓ_p norms:

$$||x||_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$$

- Inner product (dot product) of vectors
 - Multiply corresponding entries of two vectors and add up the result
 - $x \cdot y$ is also |x||y|Cos(the angle between x and y)

$$\mathbf{x}^T \mathbf{y} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i \quad \text{(scalar)}$$

The product of two matrices

$$A \in \mathbb{R}^{m \times n}$$

$$C = AB \in \mathbb{R}^{m \times p}$$

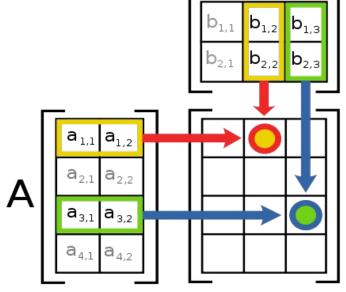
$$B \in \mathbb{R}^{n \times p}$$

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

$$C = AB = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} \begin{bmatrix} \mid & \mid & & \mid \\ b_1 & b_2 & \cdots & b_p \\ \mid & \mid & & \mid \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}.$$

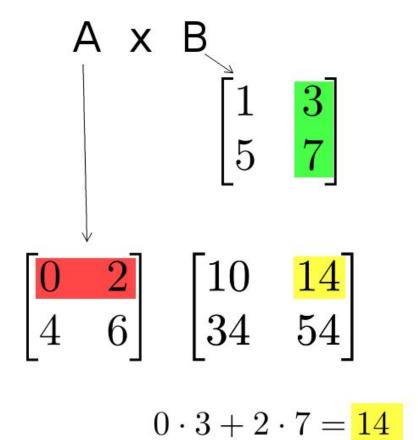
• Multiplication

The product AB is:



- Each entry in the result is (that row of A) dot product with (that column of B)
- Many uses, which will be covered later

Multiplication example:



Each entry of the matrix
 product is made by taking the
 dot product of the
 corresponding row in the left
 matrix, with the corresponding
 column in the right one.

The product of two matrices

Matrix multiplication is associative: (AB)C = A(BC).

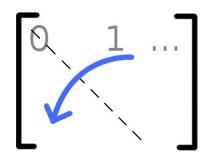
Matrix multiplication is distributive: A(B+C) = AB + AC.

Matrix multiplication is, in general, not commutative; that is, it can be the case that $AB \neq BA$. (For example, if $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times q}$, the matrix product BA does not even exist if m and q are not equal!)

Powers

- By convention, we can refer to the matrix product AA as A^2 , and AAA as A^3 , etc.
- Obviously only square matrices can be multiplied that way

 Transpose – flip matrix, so row 1 becomes column 1



$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}^T = \begin{bmatrix} 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix}$$

A useful identity:

$$(ABC)^T = C^T B^T A^T$$

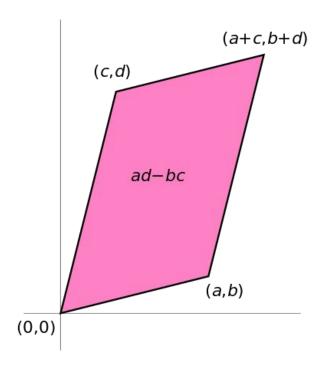
Determinant

- $-\det(\mathbf{A})$ returns a scalar
- Represents area (or volume) of the parallelogram described by the vectors in the rows of the matrix

- For
$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
, $\det(\mathbf{A}) = ad - bc$

- Properties:

$$det(\mathbf{AB}) = det(\mathbf{BA})$$
$$det(\mathbf{A}^{-1}) = \frac{1}{det(\mathbf{A})}$$
$$det(\mathbf{A}^{T}) = det(\mathbf{A})$$
$$det(\mathbf{A}) = 0 \Leftrightarrow \mathbf{A} \text{ is singular}$$



Trace

 $\operatorname{tr}(\mathbf{A}) = \operatorname{sum of diagonal elements}$ $\operatorname{tr}(\begin{bmatrix} 1 & 3 \\ 5 & 7 \end{bmatrix}) = 1 + 7 = 8$

- Invariant to a lot of transformations, so it's used sometimes in proofs. (Rarely in this class though.)
- Properties:

$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$

 $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$

Vector Norms

$$||x||_1 = \sum_{i=1}^n |x_i|$$
 $||x||_{\infty} = \max_i |x_i|$

$$||x||_2 = \sqrt{\sum_{i=1}^n x_i^2}.$$
 $||x||_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$

 Matrix norms: Norms can also be defined for matrices, such as

$$||A||_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} = \sqrt{\operatorname{tr}(A^T A)}.$$

Special Matrices

- Identity matrix I
 - Square matrix, 1's along diagonal, 0's elsewhere
 - I ⋅ [another matrix] = [that matrix]

Diagonal matrix

- Square matrix with numbers along diagonal, 0's elsewhere
- A diagonal [another matrix] scales the rows of that matrix

[1	0	0
$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	1	0
0	0	1

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 2.5 \end{bmatrix}$$

Special Matrices

Symmetric matrix

$$\mathbf{A}^T = \mathbf{A}$$

Skew-symmetric matrix

$$\mathbf{A}^T = -\mathbf{A}$$

$$\begin{bmatrix} 1 & 2 & 5 \\ 2 & 1 & 7 \\ 5 & 7 & 1 \end{bmatrix}$$

$$\mathbf{A}^T = -\mathbf{A} \begin{bmatrix} 0 & -2 & -5 \\ 2 & 0 & -7 \\ 5 & 7 & 0 \end{bmatrix}$$

Outline

- Vectors and matrices
 - Basic Matrix Operations
 - Determinants, norms, trace
 - Special Matrices
- Transformation Matrices
 - Homogeneous coordinates
 - Translation
- Matrix inverse
- Matrix rank
- Matrix Calculus

Matrix multiplication can be used to transform vectors. A matrix used in this way is called a transformation matrix.

Transformation

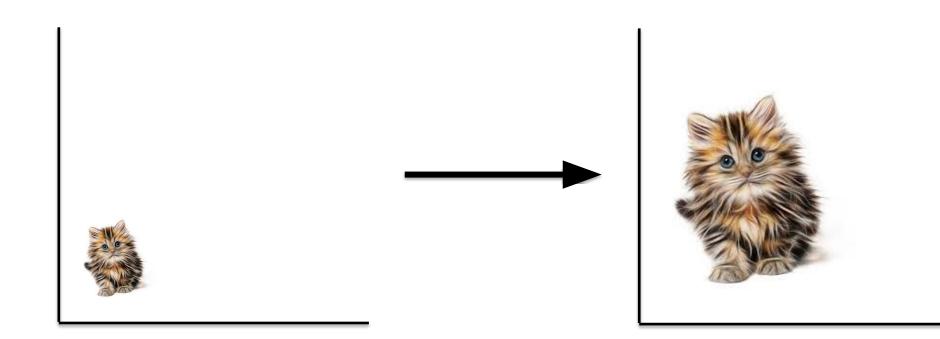
- Matrices can be used to transform vectors in useful ways, through multiplication: x'= Ax
- Simplest is scaling:

$$\begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} s_x x \\ s_y y \end{bmatrix}$$

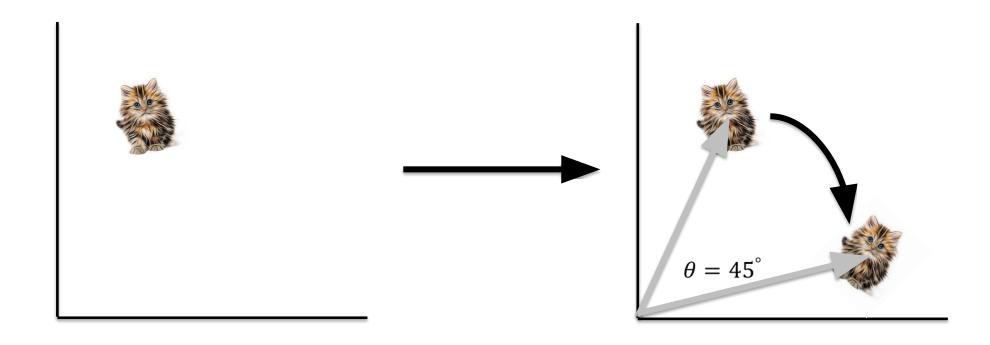
(Verify to yourself that the matrix multiplication works out this way)

Transformation

$$\begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} s_x x \\ s_y y \end{bmatrix}$$

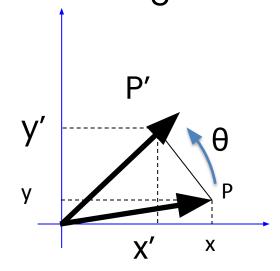


Rotation



2D Rotation Matrix Formula

Counter-clockwise rotation by an angle $\boldsymbol{\theta}$



$$x' = \cos \theta x - \sin \theta y$$
$$y' = \cos \theta y + \sin \theta x$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$P' = R P$$

Transformation Matrices

 Multiple transformation matrices can be used to transform a point:

$$p'=R_2R_1Sp$$

- The effect of this is to apply their transformations one after the other, from right to left.
- In the example above, the result is (R₂ (R₁ (S p)))
- The result is exactly the same if we multiply the matrices first, to form a single transformation matrix:

$$p' = (R_2 R_1 S) p$$

Homogeneous system

 In general, a matrix multiplication lets us linearly combine components of a vector

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

- This is sufficient for scale, rotate, skew transformations.
- But notice, we can't add a constant!

Homogeneous system

– The (somewhat hacky) solution? Stick a "1" at the end of every vector:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax + by + c \\ dx + ey + f \\ 1 \end{bmatrix}$$

- Now we can rotate, scale, and skew like before, AND translate (note how the multiplication works out, above)
- This is called "homogeneous coordinates"

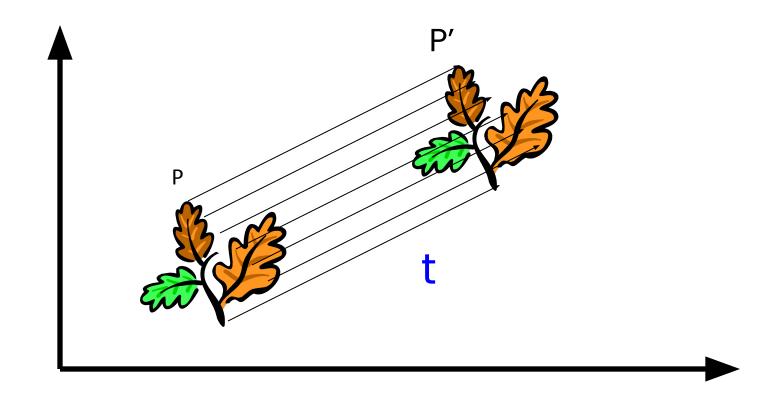
Homogeneous system

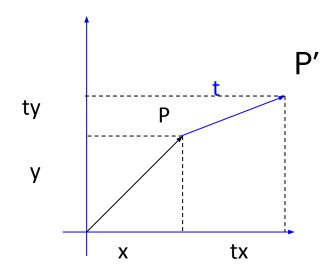
 In homogeneous coordinates, the multiplication works out so the rightmost column of the matrix is a vector that gets added.

$$\begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} ax + by + c \\ dx + ey + f \\ 1 \end{bmatrix}$$

 Generally, a homogeneous transformation matrix will have a bottom row of [0 0 1], so that the result has a "1" at the bottom too.

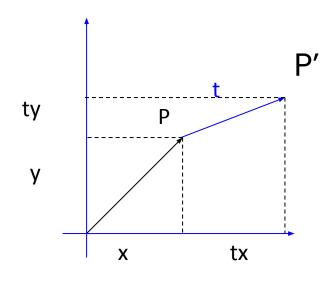
2D Translation





$$\mathbf{P} = (x, y) \to (x, y, 1)$$
$$\mathbf{t} = (t_x, t_y) \to (t_x, t_y, 1)$$

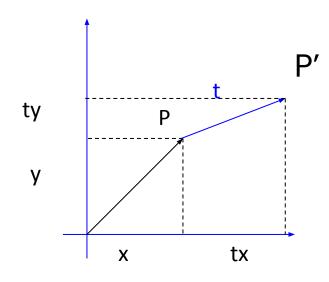
$$\mathbf{P'} \to \begin{bmatrix} x + t_x \\ y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} y \\ y \\ 1 \end{bmatrix}$$



$$\mathbf{P} = (x, y) \to (x, y, 1)$$

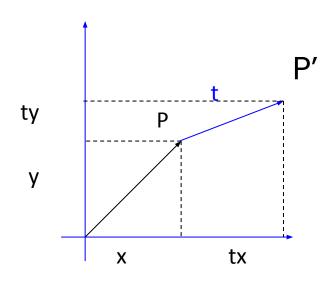
$$\mathbf{t} = (t_x, t_y) \to (t_x, t_y, 1)$$

$$\mathbf{P'} \to \begin{bmatrix} x + t_x \\ y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ y \\ 1 \end{bmatrix}$$



$$\mathbf{P} = (x, y) \to (x, y, 1)$$
$$\mathbf{t} = (t_x, t_y) \to (t_x, t_y, 1)$$

$$\mathbf{P'} \to \begin{bmatrix} x + t_x \\ y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ y \\ 1 \end{bmatrix}$$

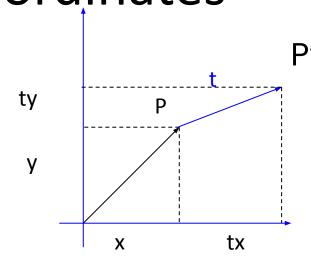


$$\mathbf{P} = (x, y) \to (x, y, 1)$$
$$\mathbf{t} = (t_x, t_y) \to (t_x, t_y, 1)$$

$$\mathbf{P} = (x, y) \to (x, y, 1)$$

$$\mathbf{t} = (t_x, t_y) \to (t_x, t_y, 1)$$

$$\mathbf{P'} \to \begin{bmatrix} x + t_x \\ y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ y & 1 \end{bmatrix}$$



$$\mathbf{P} = (x, y) \to (x, y, 1)$$

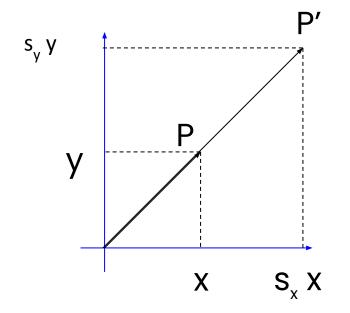
$$\mathbf{t} = (t_x, t_y) \to (t_x, t_y, 1)$$

$$\mathbf{P'} \to \begin{bmatrix} x + t_x \\ y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \cdot \mathbf{P} = \mathbf{T} \cdot \mathbf{P}$$

Scaling

Scaling Equation

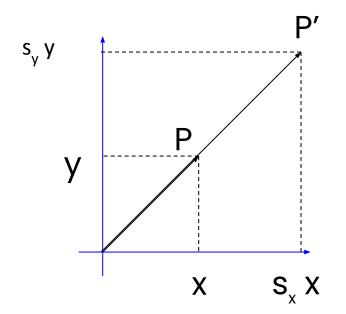


$$\mathbf{P} = (\mathbf{x}, \mathbf{y}) \rightarrow \mathbf{P'} = (\mathbf{s}_{\mathbf{x}} \mathbf{x}, \mathbf{s}_{\mathbf{y}} \mathbf{y})$$

$$\mathbf{P} = (x, y) \to (x, y, 1)$$

$$\mathbf{P'} = (s_x x, s_y y) \rightarrow (s_x x, s_y y, 1)$$

Scaling Equation



$$\mathbf{P} = (\mathbf{x}, \mathbf{y}) \rightarrow \mathbf{P'} = (\mathbf{s}_{\mathbf{x}} \mathbf{x}, \mathbf{s}_{\mathbf{y}} \mathbf{y})$$

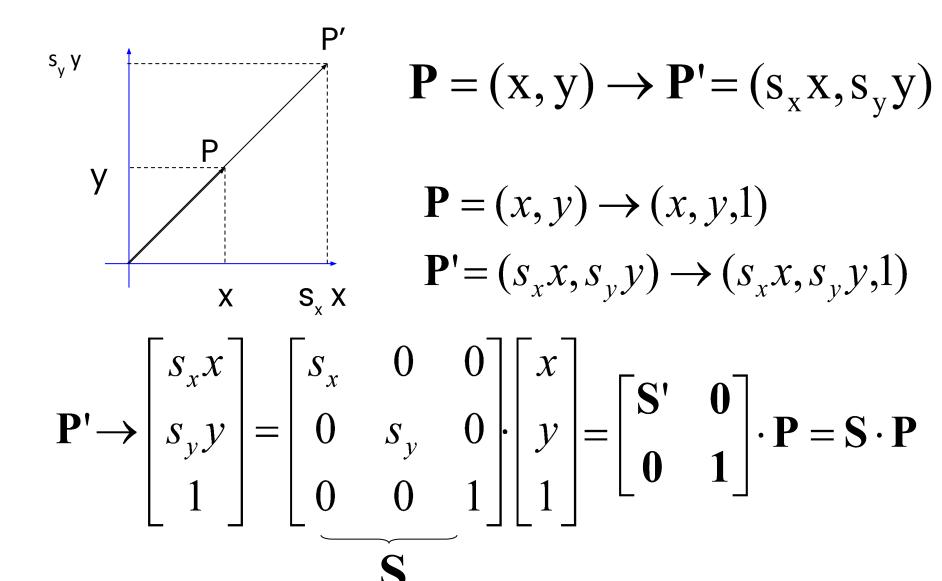
$$\mathbf{P} = (x, y) \to (x, y, 1)$$

$$\mathbf{P'} = (s_x x, s_y y) \rightarrow (s_x x, s_y y, 1)$$

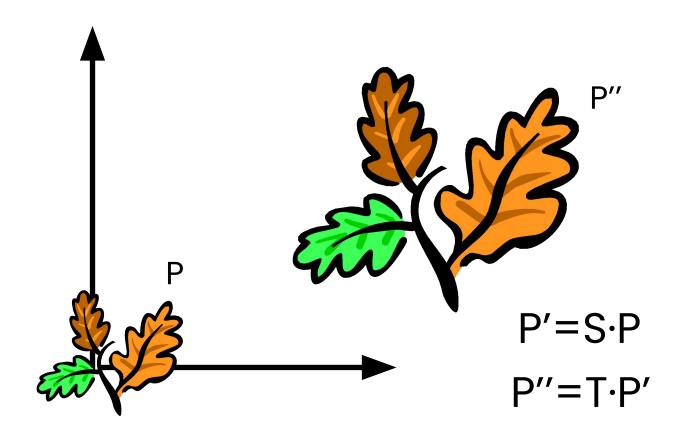
$$\mathbf{P'} \to \begin{bmatrix} s_x x \\ s_y y \\ 1 \end{bmatrix} = \begin{bmatrix} \\ \\ \end{bmatrix}$$

$$\begin{vmatrix} x \\ y \\ 1 \end{vmatrix}$$

Scaling Equation



Scaling & Translating



$$P''=T \cdot P'=T \cdot (S \cdot P)=T \cdot S \cdot P$$

Scaling & Translating

$$\mathbf{P''} = \mathbf{T} \cdot \mathbf{S} \cdot \mathbf{P} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Scaling & Translating

$$\mathbf{P''} = \mathbf{T} \cdot \mathbf{S} \cdot \mathbf{P} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & t_x \\ 0 & s_y & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_x x + t_x \\ s_y y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} s_x x + t_x \\ s_y y + t_y \\ 1 \end{bmatrix}$$

Translating & Scaling != Scaling & Translating

$$\mathbf{P'''} = \mathbf{T} \cdot \mathbf{S} \cdot \mathbf{P} = \begin{bmatrix} 1 & 0 & t_{x} \\ 0 & 1 & t_{y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_{x} & 0 & 0 \\ 0 & s_{y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_{x} & 0 & t_{x} \\ 0 & s_{y} & t_{y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_{x}x + t_{x} \\ s_{y}y + t_{y} \\ 1 \end{bmatrix}$$

Translating & Scaling != Scaling & Translating

$$\mathbf{P'''} = \mathbf{T} \cdot \mathbf{S} \cdot \mathbf{P} = \begin{bmatrix} 1 & 0 & t_{x} \\ 0 & 1 & t_{y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_{x} & 0 & 0 \\ 0 & s_{y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_{x} & 0 & t_{x} \\ 0 & s_{y} & t_{y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_{x}x + t_{x} \\ s_{y}y + t_{y} \\ 1 \end{bmatrix}$$

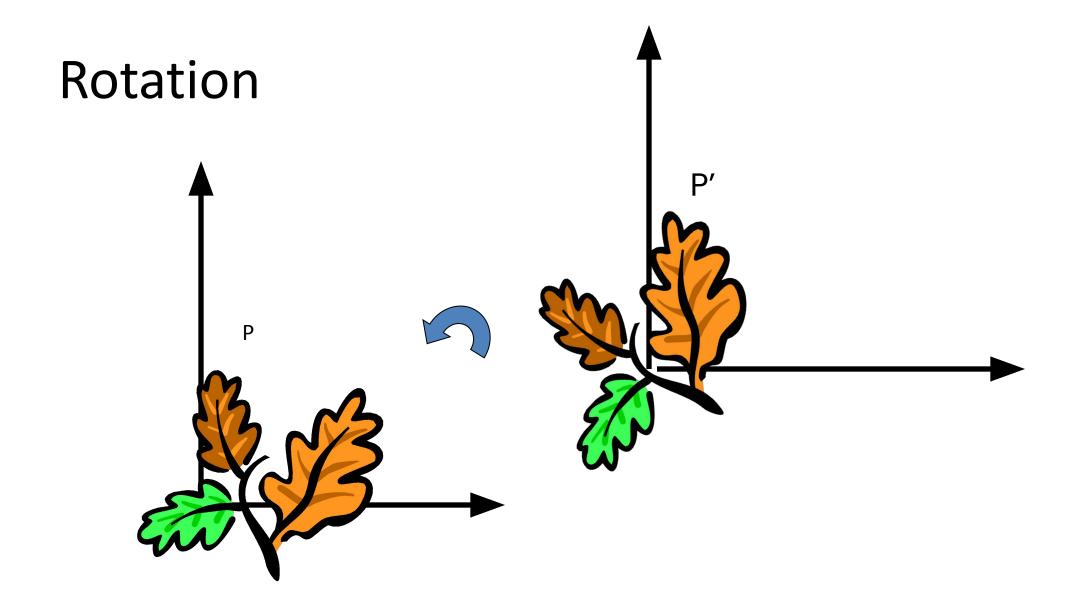
$$\mathbf{P'''} = \mathbf{S} \cdot \mathbf{T} \cdot \mathbf{P} = \begin{bmatrix} \mathbf{s}_{\mathbf{x}} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{t}_{\mathbf{x}} & \mathbf{x} \\ \mathbf{0} & \mathbf{s}_{\mathbf{y}} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{t}_{\mathbf{y}} & \mathbf{y} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \end{bmatrix} =$$

Translating & Scaling != Scaling & Translating

$$\mathbf{P'''} = \mathbf{T} \cdot \mathbf{S} \cdot \mathbf{P} = \begin{bmatrix} 1 & 0 & t_{x} \\ 0 & 1 & t_{y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_{x} & 0 & 0 \\ 0 & s_{y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_{x} & 0 & t_{x} \\ 0 & s_{y} & t_{y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_{x}x + t_{x} \\ s_{y}y + t_{y} \\ 1 \end{bmatrix}$$

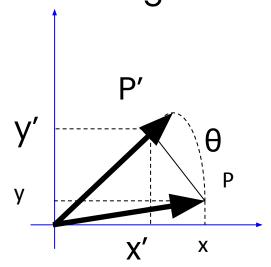
$$\mathbf{P'''} = \mathbf{S} \cdot \mathbf{T} \cdot \mathbf{P} = \begin{bmatrix} \mathbf{s}_{\mathbf{x}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{s}_{\mathbf{y}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{t}_{\mathbf{x}} \\ \mathbf{0} & \mathbf{1} & \mathbf{t}_{\mathbf{y}} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{1} \end{bmatrix} = \mathbf{0}$$

$$= \begin{bmatrix} \mathbf{s}_{\mathbf{x}} & \mathbf{0} & \mathbf{s}_{\mathbf{x}} \mathbf{t}_{\mathbf{x}} \\ \mathbf{0} & \mathbf{s}_{\mathbf{y}} & \mathbf{s}_{\mathbf{y}} \mathbf{t}_{\mathbf{y}} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_{\mathbf{x}} \mathbf{x} + \mathbf{s}_{\mathbf{x}} \mathbf{t}_{\mathbf{x}} \\ \mathbf{s}_{\mathbf{y}} \mathbf{y} + \mathbf{s}_{\mathbf{y}} \mathbf{t}_{\mathbf{y}} \\ \mathbf{1} \end{bmatrix}$$



Rotation Equations

Counter-clockwise rotation by an angle θ



$$x' = \cos \theta x - \sin \theta y$$
$$y' = \cos \theta y + \sin \theta x$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$P' = R P$$

Rotation Matrix Properties

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

A 2D rotation matrix is 2x2

Note: R belongs to the category of *normal* matrices and satisfies many interesting properties:

$$\mathbf{R} \cdot \mathbf{R}^{\mathrm{T}} = \mathbf{R}^{\mathrm{T}} \cdot \mathbf{R} = \mathbf{I}$$
$$\det(\mathbf{R}) = 1$$

Rotation Matrix Properties

 Transpose of a rotation matrix produces a rotation in the opposite direction

$$\mathbf{R} \cdot \mathbf{R}^{\mathrm{T}} = \mathbf{R}^{\mathrm{T}} \cdot \mathbf{R} = \mathbf{I}$$
$$\det(\mathbf{R}) = 1$$

- The rows of a rotation matrix are always mutually perpendicular (a.k.a. orthogonal) unit vectors
 - (and so are its columns)

Scaling + Rotation + Translation

$$\mathbf{P'} = \mathbf{T} \cdot \mathbf{R} \cdot \mathbf{S} \cdot \mathbf{P} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & t_x$$

$$= \begin{bmatrix} \cos \theta & -\sin \theta & t_x \\ \sin \theta & \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} =$$

$$= \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} R S & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

This is the form of the general-purpose transformation matrix

Outline

- Vectors and matrices
 - Basic Matrix Operations
 - Determinants, norms, trace
 - Special Matrices
- Transformation Matrices
 - Homogeneous coordinates
 - Translation
- Matrix inverse
- Matrix rank
- Matrix Calculate

The inverse of a transformation matrix reverses its effect

Inverse

• Given a matrix A, its inverse A^{-1} is a matrix such that $AA^{-1} = A^{-1}A = I$

• E.g.
$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$

- Inverse does not always exist. If A⁻¹ exists, A is
 invertible or non-singular. Otherwise, it's singular.
- Useful identities, for matrices that are invertible:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$
$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$
$$\mathbf{A}^{-T} \triangleq (\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}$$

Matrix Operations

Pseudoinverse

- Fortunately, there are workarounds to solve AX=B in these situations. And python can do them!
- Instead of taking an inverse, directly ask python to solve for X in AX=B, by typing np.linalg.solve(A, B)
- Python will try several appropriate numerical methods (including the pseudoinverse if the inverse doesn't exist)
- Python will return the value of X which solves the equation
 - If there is no exact solution, it will return the closest one
 - If there are many solutions, it will return the smallest one

Matrix Operations

Python example:

$$AX = B$$

$$A = \begin{bmatrix} 2 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

```
>> import numpy as np
>> x = np.linalg.solve(A,B)
x =
    1.0000
    -0.5000
```

Outline

- Vectors and matrices
 - Basic Matrix Operations
 - Determinants, norms, trace
 - Special Matrices
- Transformation Matrices
 - Homogeneous coordinates
 - Translation
- Matrix inverse
- Matrix rank
- Matrix Calculate

The rank of a transformation matrix tells you how many dimensions it transforms a vector to.

Linear independence

- Suppose we have a set of vectors $v_1, ..., v_n$
- If we can express \mathbf{v}_1 as a linear combination of the other vectors $\mathbf{v}_2...\mathbf{v}_n$, then \mathbf{v}_1 is linearly *dependent* on the other vectors.
 - The direction \mathbf{v}_1 can be expressed as a combination of the directions $\mathbf{v}_2 \dots \mathbf{v}_n$. (E.g. $\mathbf{v}_1 = .7 \ \mathbf{v}_2 .7 \ \mathbf{v}_4$)

Linear independence

- Suppose we have a set of vectors $v_1, ..., v_n$
- If we can express \mathbf{v}_1 as a linear combination of the other vectors $\mathbf{v}_2...\mathbf{v}_n$, then \mathbf{v}_1 is linearly dependent on the other vectors.
 - The direction \mathbf{v}_1 can be expressed as a combination of the directions $\mathbf{v}_2 \dots \mathbf{v}_n$. (E.g. $\mathbf{v}_1 = .7 \ \mathbf{v}_2 .7 \ \mathbf{v}_4$)
- If no vector is linearly dependent on the rest of the set, the set is linearly *independent*.
 - Common case: a set of vectors $\mathbf{v_1}$, ..., $\mathbf{v_n}$ is always linearly independent if each vector is perpendicular to every other vector (and non-zero)

Matrix rank

Column/row rank

Column rank always equals row rank

Matrix rank

$$rank(\mathbf{A}) \triangleq col\text{-}rank(\mathbf{A}) = row\text{-}rank(\mathbf{A})$$

Matrix rank

- For transformation matrices, the rank tells you the dimensions of the output
- E.g. if rank of A is 1, then the transformation

maps points onto a line.

• Here's a matrix with rank 1:

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+y \\ 2x+2y \end{bmatrix} - \text{All points get mapped to the line y=2x}$$

Matrix rank

- If an m x m matrix is rank m, we say it's "full rank"
 - Maps an m x 1 vector uniquely to another m x 1 vector
 - An inverse matrix can be found
- If rank < m, we say it's "singular"
 - At least one dimension is getting collapsed. No way to look at the result and tell what the input was
 - Inverse does not exist
- Inverse also doesn't exist for non-square matrices

Outline

- Vectors and matrices
 - Basic Matrix Operations
 - Determinants, norms, trace
 - Special Matrices
- Transformation Matrices
 - Homogeneous coordinates
 - Translation
- Matrix inverse
- Matrix rank
- Matrix Calculus

Matrix Calculus – The Gradient

- Let a function $f: \mathbb{R}^{m \times n} \to \mathbb{R}$ take as input a matrix A of size $m \times n$ and return a real value.
- Then the gradient of f:

$$\nabla_{A} f(A) \in \mathbb{R}^{m \times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \dots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \dots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \dots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

Matrix Calculus – The Gradient

- Every entry in the matrix is: $\nabla_A f(A))_{ij} = \frac{\partial f(A)}{\partial A_{ij}}$.
- the size of $\nabla_A f(A)$ is always the same as the size of A. So if A is just a vector x:

$$abla_x f(x) = \left[egin{array}{c} rac{\partial f(x)}{\partial x_1} \ rac{\partial f(x)}{\partial x_2} \ dots \ rac{\partial f(x)}{\partial x_n} \end{array}
ight]$$

Exercise

• Example:

For $x \in \mathbb{R}^n$, let $f(x) = b^T x$ for some known vector $b \in \mathbb{R}^n$

$$f(x) = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}^T egin{bmatrix} x_1 \ x_2 \ dots \ x_n \end{bmatrix}$$

• Find:

$$\frac{\partial f(x)}{\partial x_k} = ?$$

$$\nabla_x f(x) = ?$$

Exercise

• Example:

For $x \in \mathbb{R}^n$, let $f(x) = b^T x$ for some known vector $b \in \mathbb{R}^n$

$$f(x) = \sum_{i=1}^{n} b_i x_i$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n b_i x_i = b_k.$$

• From this we can conclude that:

$$\nabla_x b^T x = b$$

Matrix Calculus – The Gradient

Properties

•
$$\nabla_x(f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x)$$
.

• For $t \in \mathbb{R}$, $\nabla_x(t f(x)) = t\nabla_x f(x)$.

• The Hessian matrix with respect to x, written $\nabla_x^2 f(x)$ or simply as H:

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

The Hessian of n-dimensional vector is the n × n matrix.

• Each entry can be written as: $\nabla_x^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$

Exercise: Why is the Hessian always symmetric?

• Each entry can be written as:

$$\nabla_x^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

• The Hessian is always symmetric, because

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}.$$

 This is known as Schwarz's theorem: The order of partial derivatives don't matter as long as the second derivative exists and is continuous.

 Note that the hessian is not the gradient of whole gradient of a vector (this is not defined). It is actually the gradient of every entry of the gradient of the vector.

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

• Eg, the first column is the gradient of $\frac{\partial f(x)}{\partial x_1}$

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

• Example:

consider the quadratic function $f(x) = x^T A x$

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right]$$

Divide the summation into 3 parts depending on whether:

- i == k or
- j == k

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right]$$

$$= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right]$$

$$= \left[\sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k \right]$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right]$$

$$= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right]$$

$$= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right]$$

$$= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right]$$

$$= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k$$

$$= \sum_{i=1}^n A_{ik} x_i + \sum_{j=1}^n A_{kj} x_j = 2 \sum_{i=1}^n A_{ki} x_i,$$

$$f(x) = x^{T} A x$$

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_{i} x_{j}$$

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_\ell} = \frac{\partial}{\partial x_k} \left[\frac{\partial f(x)}{\partial x_\ell} \right] = \frac{\partial}{\partial x_k} \left[\sum_{i=1}^n A_{\ell i} x_i \right]$$

$$f(x) = x^{T} A x$$

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_{i} x_{j}$$

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_\ell} = \frac{\partial}{\partial x_k} \left[\frac{\partial f(x)}{\partial x_\ell} \right] = \frac{\partial}{\partial x_k} \left[\sum_{i=1}^n A_{\ell i} x_i \right]$$
$$= 2A_{\ell k} = 2A_{k\ell}.$$

$$f(x) = x^{T} A x$$

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_{i} x_{j}$$

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_\ell} = \frac{\partial}{\partial x_k} \left[\frac{\partial f(x)}{\partial x_\ell} \right] = \frac{\partial}{\partial x_k} \left[\sum_{i=1}^n 2A_{\ell i} x_i \right]$$
$$= 2A_{\ell k} = 2A_{k\ell}.$$

$$\nabla_x^2 f(x) = 2A$$

What we have learned

- Vectors and matrices
 - Basic Matrix Operations
 - Special Matrices
- Transformation Matrices
 - Homogeneous coordinates
 - Translation
- Matrix inverse
- Matrix rank
- Eigenvalues and Eigenvectors
- Matrix Calculate