Linear Algebra Primer

Note: the slides are based on CS131 (Juan Carlos et al) and EE263 (by Stephen Boyd et al) at Stanford. Reorganized, revised, and typed by Hao Su

Matrix

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

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

- if $m = n$, we say that A is square.

Vector

- A column vector $v \in \mathbb{R}^{n \times 1}$ where

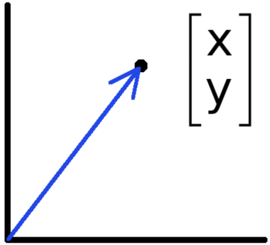
$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

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

$$v^T = [v_1 v_2 \dots v_n]$$

  T denotes the **transpose** operation

Vectors have two main uses



- Vectors can represent an offset in 2D or 3D space
- Points are just vectors from the origin

- Data (pixels, gradients at an image keypoint, etc) can also be treated as a vector
- Such vectors do not have a geometric interpretation, but calculations like "distance" can still have value

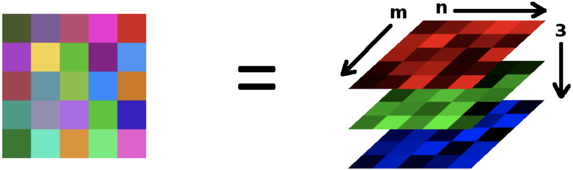
Images

$$= \begin{bmatrix} 193 & 180 & 210 & 112 & 125 \\ 189 & 8 & 177 & 97 & 114 \\ 100 & 71 & 81 & 195 & 165 \\ 167 & 12 & 242 & 203 & 181 \\ 44 & 25 & 9 & 48 & 192 \end{bmatrix}$$

- Python represents an image as a matrix of pixel brightness

Color Images

- Grayscale images have one number per pixel, and are stored as an $m \times n$ matrix
- Color images have 3 numbers per pixel – red, green, and blue brightness (RGB)
- stored as an $m \times n \times 3$ matrix



Matrix Operations

- 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

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

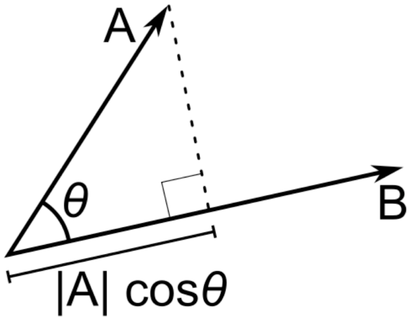
Vector Operations

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

$$x^T y = [x_1 \ldots x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i \qquad \text{(scalar)}$$

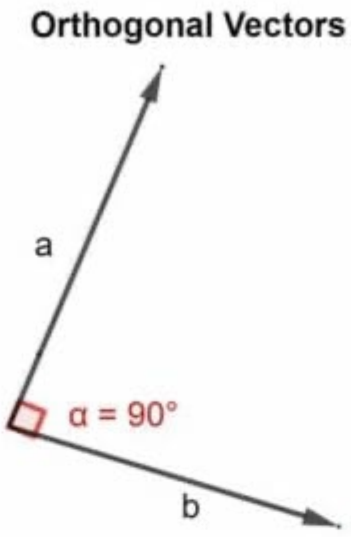
Vector Operations

- Inner product (dot product) of vectors
  - If B is a unit vector, then $A \cdot B$ gives the length of A, which lies in the direction of B



Vector Operations

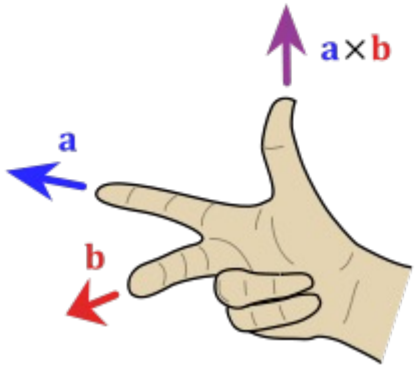
- Orthogonality
  - We say that 2 vectors are orthogonal if they are perpendicular to each other
  - In terms of the dot product, if the two vectors are orthogonal to each other, then their dot product is 0.
    - For some vectors u and v, if they are orthogonal, then the cosine of the angle between them is 0 so their dot product is also 0
    - $|u||v|\cos(90) = 0$



https://www.storyofmathematics.com/orthogona-vector/

Vector Operations

- Cross Product
  - For two vectors a,b in R^3, a cross product is defined as a vector c that is orthogonal to both a and b



https://en.wikipedia.org/wiki/Cross_product

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 $x\mathbb{R}^n$, $t \in \mathbb{R}$, $f(tx) = |t| f(x)$
  - Triangle inequality: For all $x, y \in \mathbb{R}^n$, $f(x + y) \leq f(x) + f(y)$

Vector Operations

- Example norms

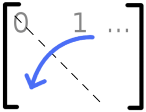
$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|x\|_\infty = \max_i |x_i|$$

- General ℓ_p norms:

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

Matrix Operations

- 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:

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

Matrix Operations

- The product of two matrices

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

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

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

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

Matrix Operations

Multiplication example:

A x B

$$\begin{bmatrix} 1 & \colorbox{green}{3} \\ 5 & \colorbox{green}{7} \end{bmatrix}$$

$$\begin{bmatrix} \colorbox{red}{0} & \colorbox{red}{2} \\ 4 & 6 \end{bmatrix} \quad \begin{bmatrix} 10 & \colorbox{yellow}{14} \\ 34 & 54 \end{bmatrix}$$

$$0 \cdot 3 + 2 \cdot 7 = \colorbox{yellow}{14}$$

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.

Matrix Operations

- 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!)

Matrix Operations

- 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

Matrix Operations

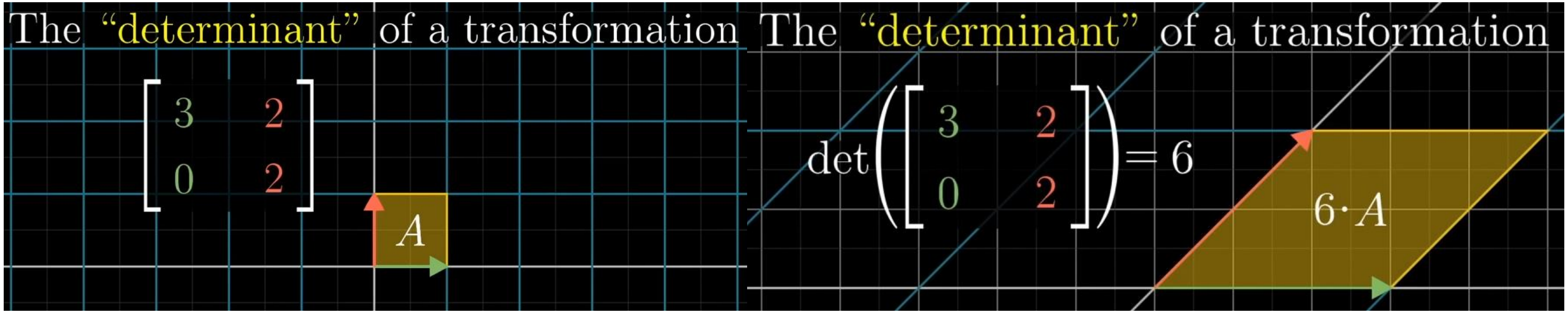
- Determinant
  - det(A) returns a scalar
  - Represents the scaling factor and orientation of a region after going through the transformation described by A
  - For A = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det(A) = ad - bc$
  - Properties:

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

https://www.youtube.com/watch?v=Ip3X9LOh2dk

Matrix Operations

- Determinant
  - Geometrically, it looks something like this



https://www.youtube.com/watch?v=Ip3X9LOh2dk

Matrix Operations

- Trace
  - $\text{trace}(A) = \text{sum of diagonal elements}$

$$\text{tr}(\begin{bmatrix} 1 & 3 \\ 5 & 7 \end{bmatrix}) = 1 + 7 = 8$$

  - Properties:

$$\begin{aligned} \text{tr}(AB) &= \text{tr}(BA) \\ \text{tr}(A + B) &= \text{tr}(A) + \text{tr}(B) \\ \text{tr}(ABC) &= \text{tr}(BCA) = \text{tr}(CAB) \end{aligned}$$

Special Matrices

- Identity matrix I

$$I_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad AI = ?$$

- Diagonal matrix

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

Special Matrices

- Identity matrix I

$$I_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad AI = A$$

- Diagonal matrix

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

Special Matrices

- Symmetric matrix: $A^T = A$

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

- Skew-symmetric matrix: $A^T = -A$

$$\begin{bmatrix} 0 & -2 & -5 \\ 2 & 0 & -7 \\ 5 & 7 & 0 \end{bmatrix}$$

Transformation

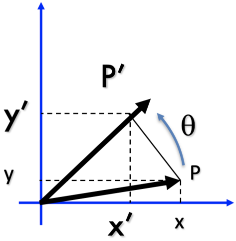
- Matrices can be used to transform vectors in useful ways, through multiplication: $x' = Ax$
- Simplest is scaling:

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

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

Rotation (2D case)

Counter-clockwise rotation by an angle θ



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

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

$$P' = RP$$

Transformation Matrices

- Multiple transformation matrices can be used to transform a point:

$$p' = R_2 R_1 S p$$

- The effect of this is to apply their transformations one after the other, from **right to left**
- In the example above, the result is

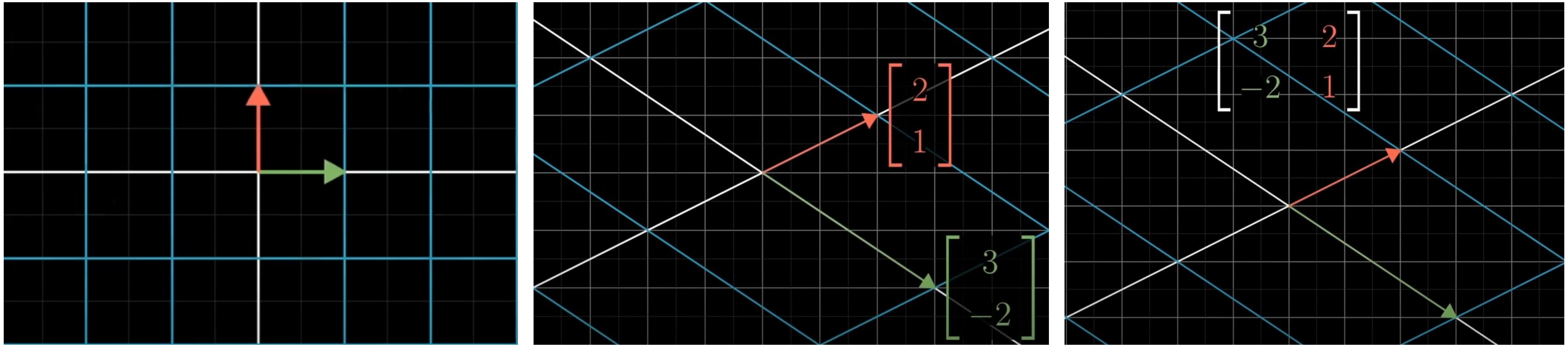
$$(R_2(R_1(Sp)))$$

- The result is exactly the same if we multiply the matrices first, to form a single transformation matrix:

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

Matrices and Linear Transformations

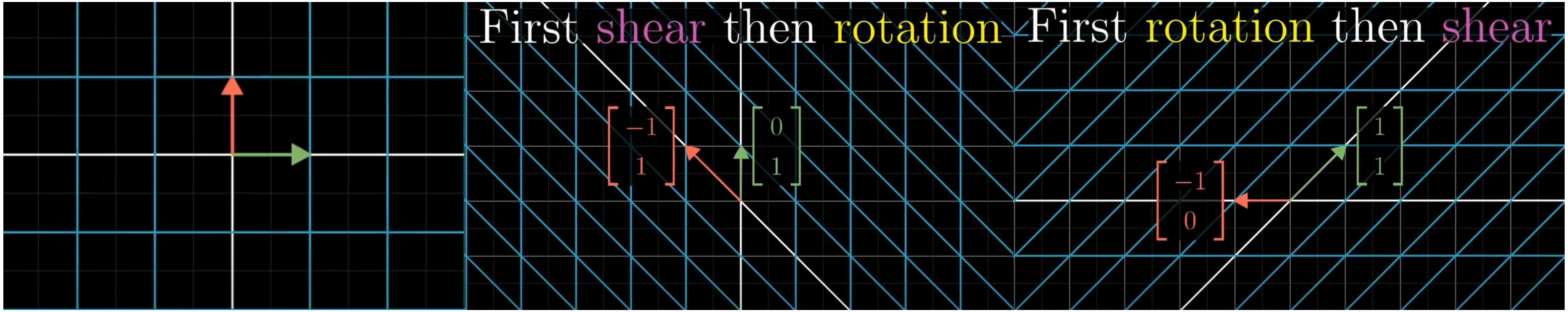
- Matrices as linear transformations
  - One way to think about matrices is that they describe a linear transformation



https://www.youtube.com/watch?v=Ip3X9LOh2dk

Matrices and Linear Transformations

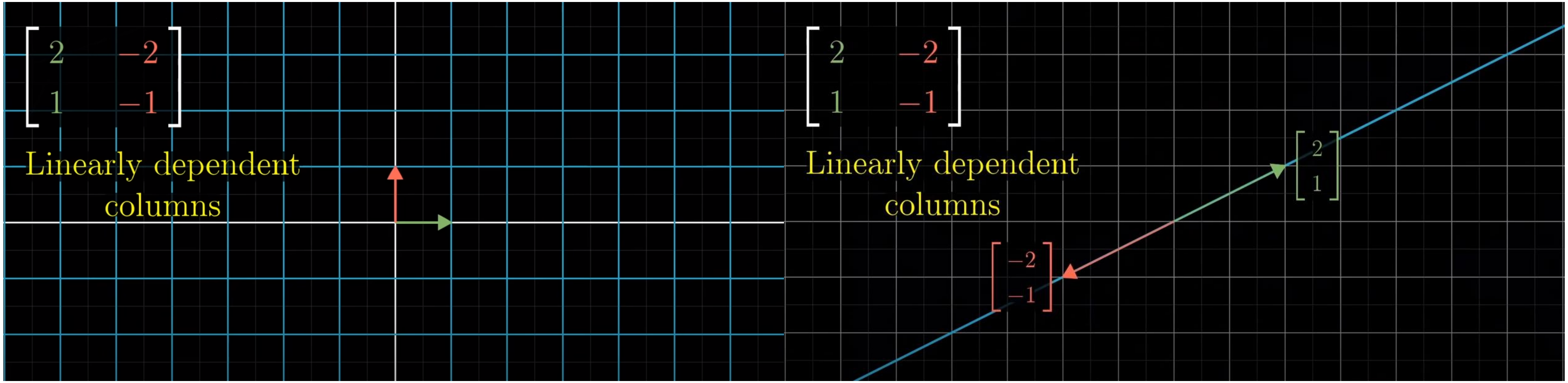
- Matrices as linear transformations
  - As previously discussed, the order in which you apply the transformations often matters (not commutative)



https://www.youtube.com/watch?v=Ip3X9LOh2dk

Matrices and Linear Transformations

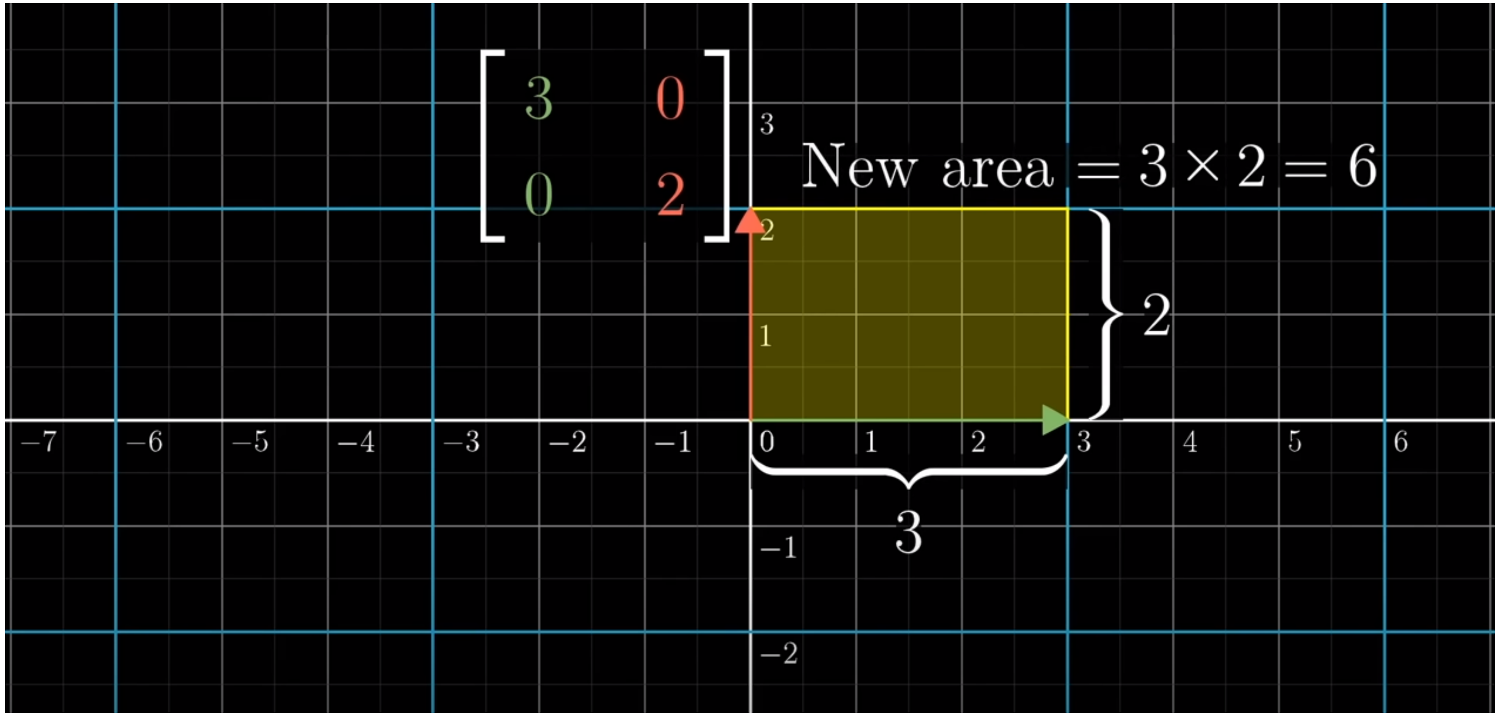
- Linear independence and dependence
  - If our matrix is singular (zero determinant), then the transformation squishes space into a lower dimension



https://www.youtube.com/watch?v=Ip3X9LOh2dk

Matrices and Linear Transformations

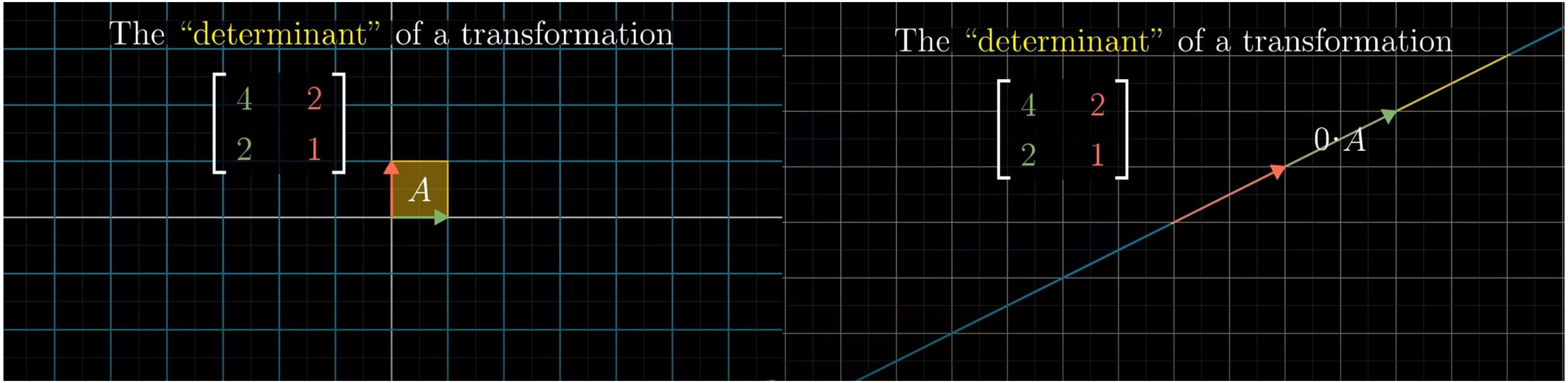
- Linear dependence and determinants
  - Non-singular matrix



https://www.youtube.com/watch?v=Ip3X9LOh2dk

Matrices and Linear Transformations

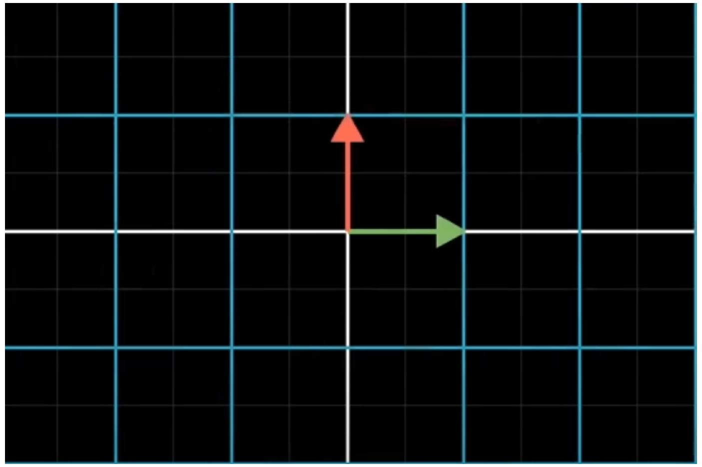
- Linear dependence and determinants
  - If our matrix is singular, then the determinant is 0 since our space is squished into a lower dimension.



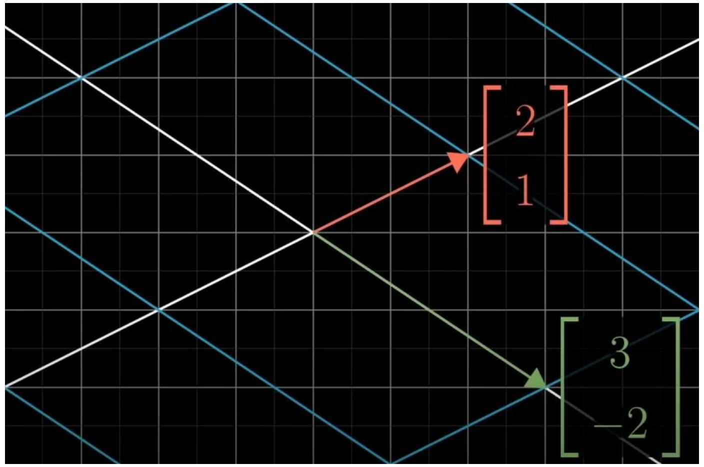
https://www.youtube.com/watch?v=Ip3X9LOh2dk

Matrices and Linear Transformations

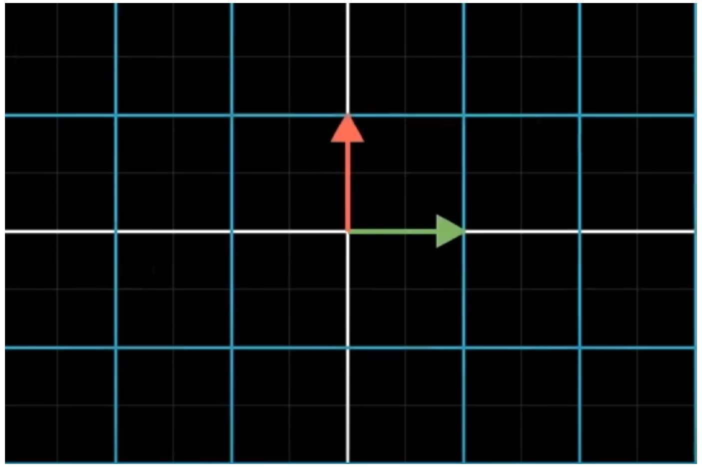
- Inverses
  - If our matrix is not singular, that means it is invertible
  - Applying the inverse of a transformation is like “undoing” the transformation.



$\mathrm{I}x$



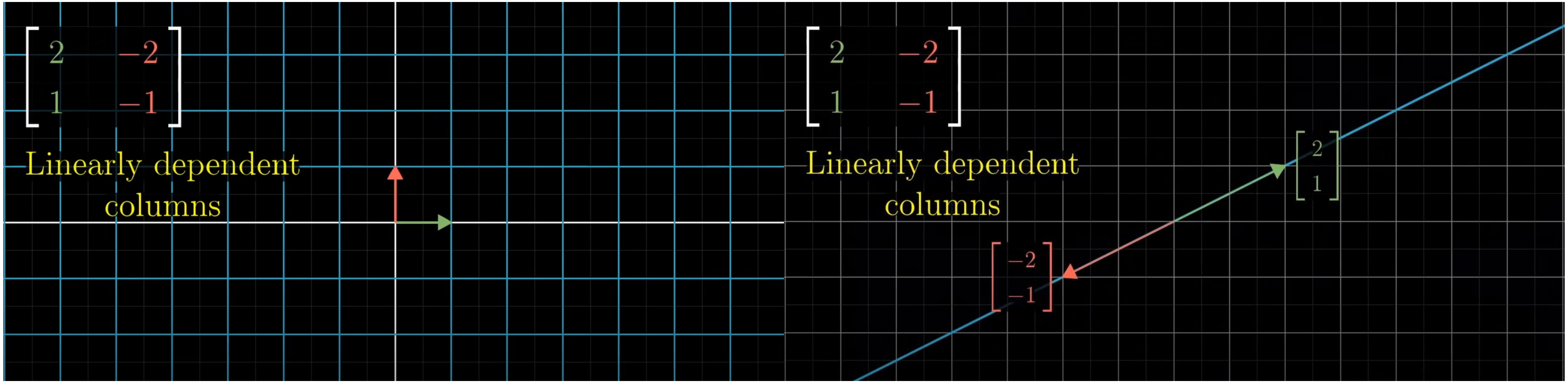
Ax



$A^{-1}Ax$

Matrices and Linear Transformations

- Inverses
  - If our matrix is singular, that means it is not invertible
  - There is no way to "undo" the transformation since a single vector would need to map to multiple vectors



https://www.youtube.com/watch?v=Ip3X9LOh2dk

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 cannot 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

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

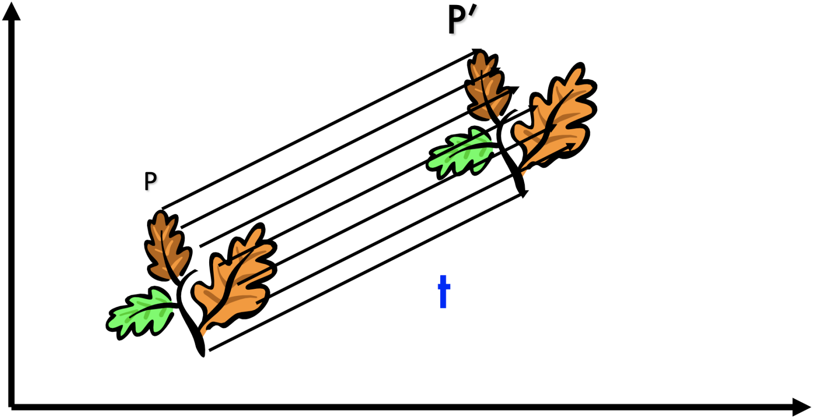
- Generally, a homogeneous transformation matrix will have a bottom row of [0 0 1], so that the result has a "1" at the bottom, too.

Homogeneous System

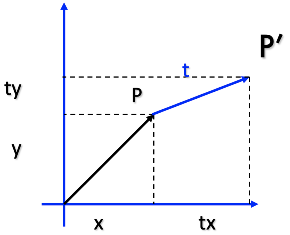
- One more thing we might want: to divide the result by something:
  - Matrix multiplication cannot actually divide
  - So, **by convention**, in homogeneous coordinates, we'll divide the result by its last coordinate after doing a matrix multiplication

$$\begin{bmatrix} x \\ y \\ 7 \end{bmatrix} \Rightarrow \begin{bmatrix} x/7 \\ y/7 \\ 1 \end{bmatrix}$$

2D Transformation using Homogeneous Coordinates



2D Transformation using Homogeneous Coordinates



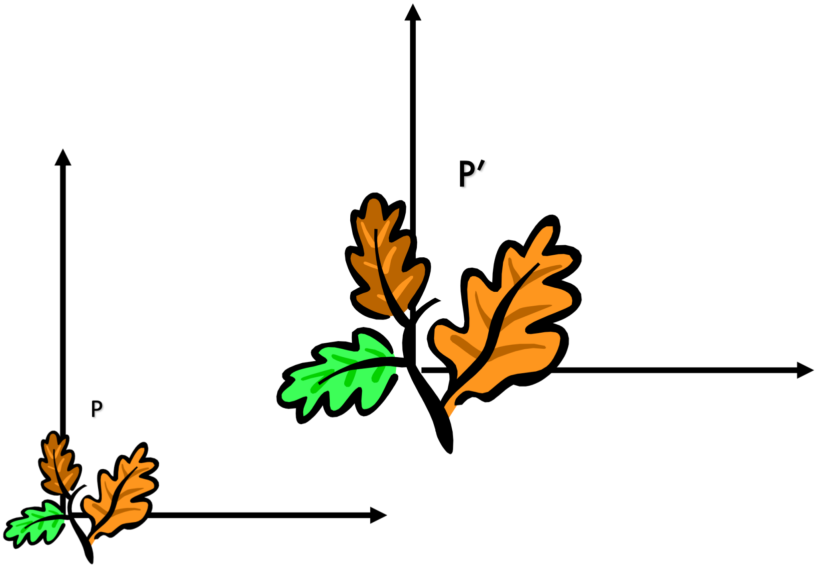
$$\mathbf{P} = (x, y) \rightarrow (x, y, 1)$$

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

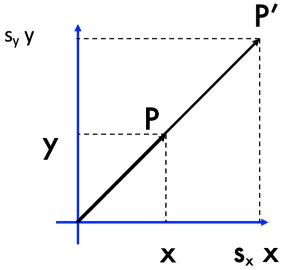
$$\mathbf{P}' \rightarrow \begin{bmatrix} x + t_x \\ y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

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

Scaling



Scaling Equation



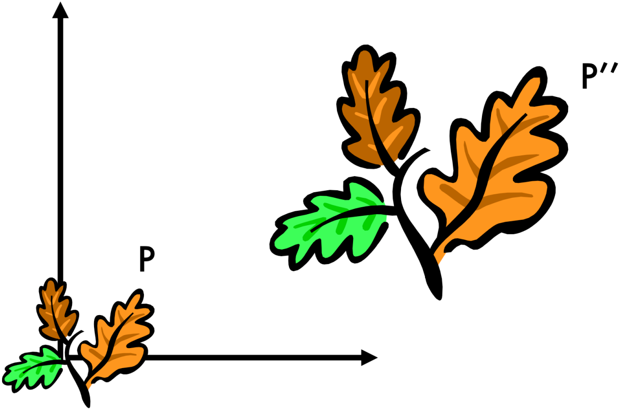
$$\mathbf{P} = (\mathrm{x}, \mathrm{y}) \rightarrow \mathbf{P}' = (\mathrm{s}_\mathrm{x}\mathrm{x}, \mathrm{s}_\mathrm{y}\mathrm{y})$$

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

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

$$\mathbf{P}' \rightarrow \begin{bmatrix} s_x x \\ s_y y \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{S}' & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \cdot \mathbf{P} = \mathbf{S} \cdot \mathbf{P}$$

Scaling & Translating



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

Scaling & Translating

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

$$= \begin{bmatrix} s_x & 0 & t_x \\ 0 & s_y & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s_x x + t_x \\ s_y y + t_y \\ 1 \end{bmatrix} = \begin{bmatrix} S & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Translation & Scaling versus Scaling & Translating

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

$$= \begin{bmatrix} s_x x + t_x \\ s_y y + t_y \\ 1 \end{bmatrix}$$

Translation & Scaling $\neq$ Scaling & Translating

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

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

Translation & Scaling $\neq$ Scaling & Translating

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

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