CS60010: Deep Learning

Spring 2023

Sudeshna Sarkar

Module 1 Part B

Linear Algebra

Sudeshna Sarkar

5 Jan 2023

Scalars

- A scalar is a single number
- Integers, real numbers, rational numbers, etc.

$$a, n, x$$

Vectors

- A vector is a 1-D array of numbers:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}$$

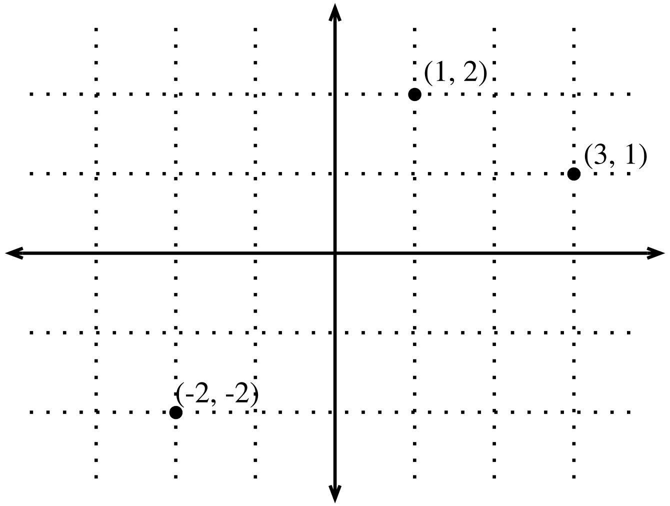
- Can be real, binary, integer, etc.
- Example notation for type and size:

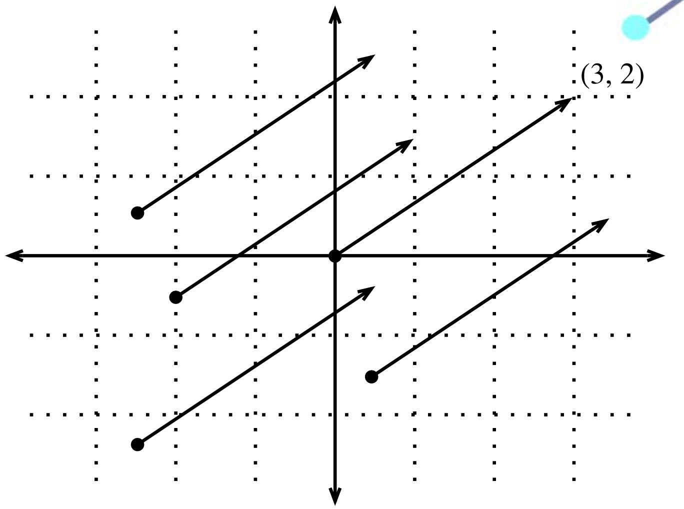
$$\mathbb{R}^n$$

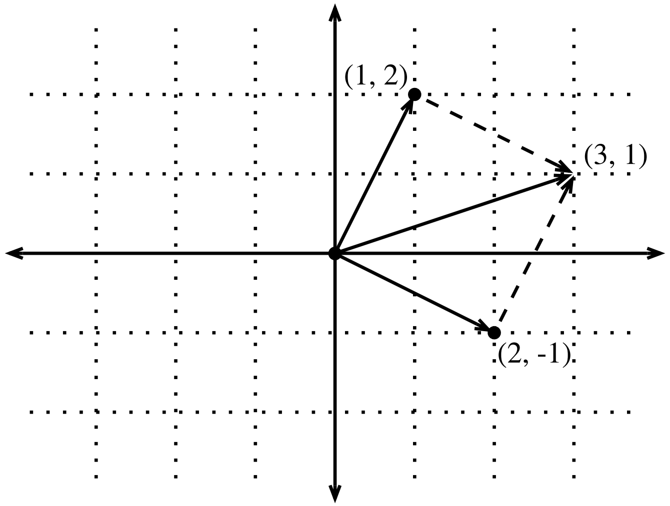
v=[2, 1, 5, 0]

Geometric interpretation of Vectors

1. a point in space
2. direction in space







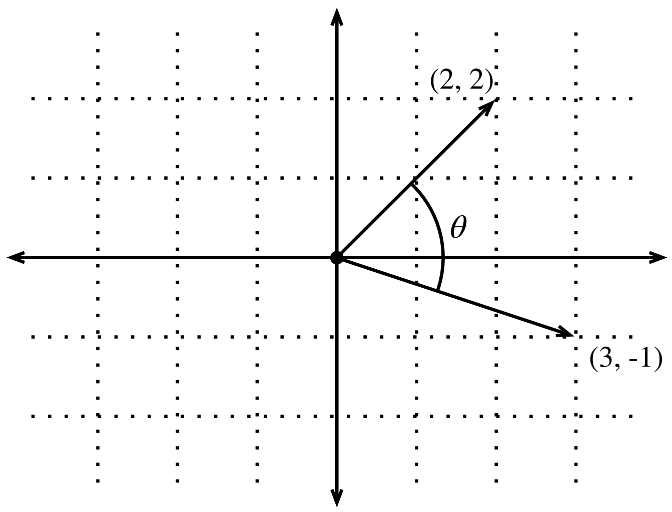
Dot Products and Angles

$$u^T v = \sum_i u_i \cdot v_i$$

$$u \cdot v = u^T v = v^T u$$

$$u \cdot v = \|u\| \|v\| \cos \theta$$

$$\theta = \cos^{-1}\left(\frac{u \cdot v}{\|u\| \|v\|}\right)$$



Cosine Similarity

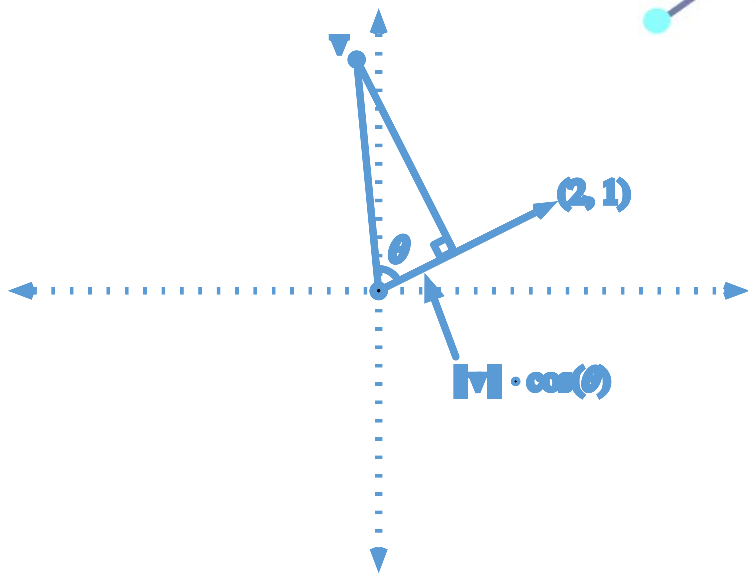
- $\cos\theta = \frac{u \cdot v}{\|u\| \|v\|}$

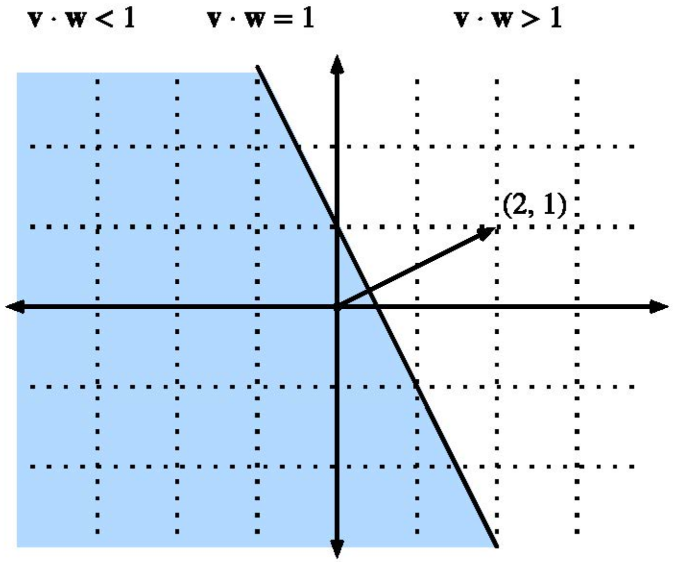
Hyperplanes

- In a d-dimensional vector space, a hyperplane has d−1 dimensions and divides the space into two half-spaces.

Hyperplanes

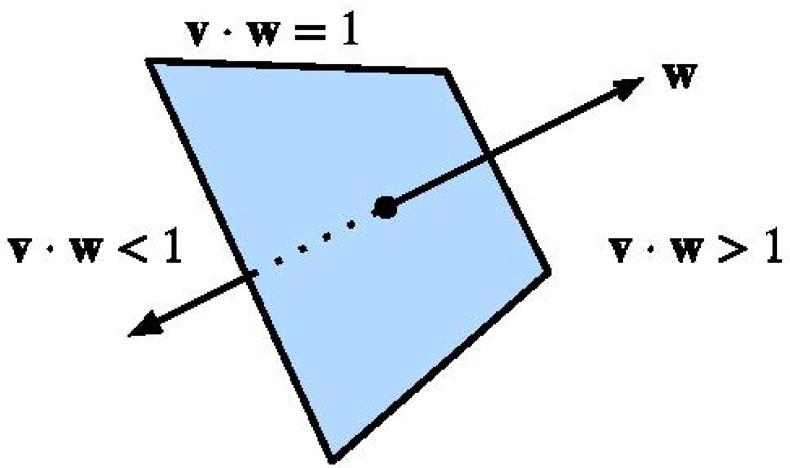
- Consider the column vector $w = [2,1]^T$
- what are the points v with w·v=1?
- $\|v\|\|w\| \cos\theta = 1 \Leftrightarrow \|v\| \cos\theta = \frac{1}{\|w\|} = \frac{1}{\sqrt{5}}$
- Geometric interpretation: the length of the projection of v onto the direction of w is exactly 1/||w||
- The set of all points where this is true is a line at right angles to the vector w.





Hyperplanes

- Consider w=[1,2,3]T
- What are the points in three dimensions with w·v=1?
    - we obtain a plane at right angles to the given vector w
- Hyperplanes in any dimension separate the space into two halves.



Matrices

- A matrix is a 2-D array of numbers:

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix}$$

- Example notation for type and shape:

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

- Symmetric Matrix: $AA = A^T$
- Orthogonal Matrix: $AA^T = A^T A = I$ and $A^{-1} = A^T$
- Diagonal Matrix: Non-zero entries only in the diagonals

Tensors

- A tensor is an array of numbers, that may have
  - zero dimensions, and be a scalar
  - one dimension, and be a vector
  - two dimensions, and be a matrix
  - or more dimensions.

Matrix Transpose

Transpose: Transpose of a matrix is the mirror image of the matrix across the diagonal line, called the main diagonal of the matrix.

- The transpose of a matrix $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^T$,

$$(\boldsymbol{A}^\top)_{i,j} = A_{j,i}.$$

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow \boldsymbol{A}^\top = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

$$(AB)^T = B^T A^T$$

Matrix Operations

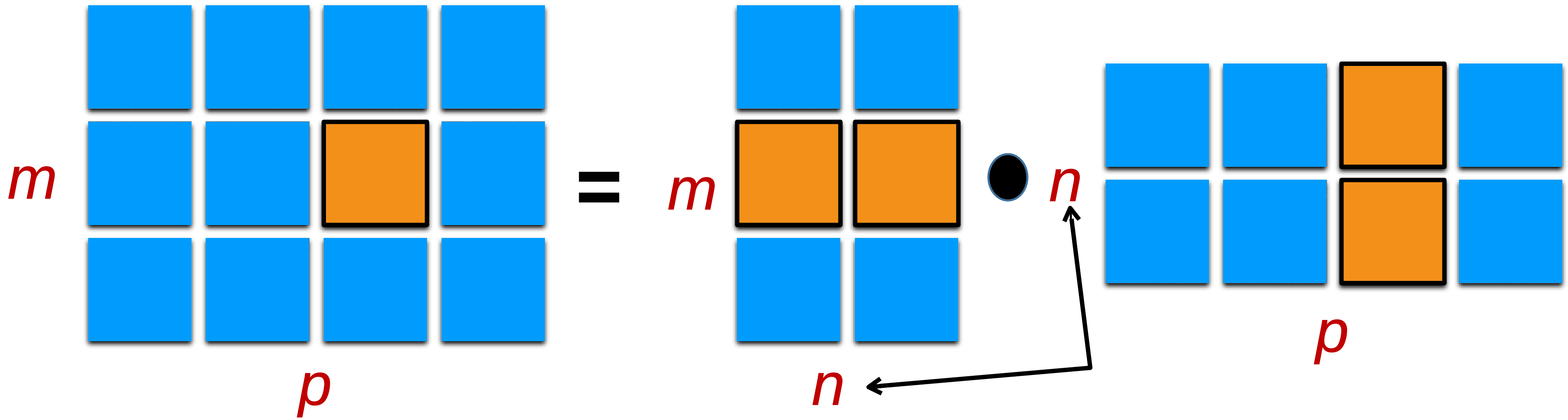
- **Addition**: Matrices can be added as long as they have the same shape, by adding their corresponding elements.
  - $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$, where $\boldsymbol{C}_{i,j} = \boldsymbol{A}_{i,j} + \boldsymbol{B}_{i,j}$
- **Multiplication**: In order for the product of the two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ to be defined, $\boldsymbol{A}$ must have the same number of columns as that of the rows of $\boldsymbol{B}$. If $\boldsymbol{A}$ is of shape $m \times n$ and $\boldsymbol{B}$ is of shape $n \times p$ then $\boldsymbol{C}$ is of shape $m \times p$, the product operation $\boldsymbol{C} = \boldsymbol{AB}$ is defined by

$$\boldsymbol{C}_{i,j} = \sum_k \boldsymbol{A}_{i,k} \boldsymbol{B}_{k,j}$$

Matrix (Dot) Product

$$\boldsymbol{C} = \boldsymbol{AB}$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}$$



Matrix product is not commutative ($\boldsymbol{AB} = \boldsymbol{BA}$ does not always hold).

Matrix Operations

- **Elementwise or Hadamard Product**: It's a matrix containing the product of the individual elements. It is denoted as $\boldsymbol{A} \odot \boldsymbol{B}$

Geometry of Linear Transformations

Linear transformations represented by matrices. Consider matrix A and vector v.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad \mathrm{v} = [x, y]^T$$

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

$$= x \begin{bmatrix} a \\ c \end{bmatrix} + y \begin{bmatrix} b \\ d \end{bmatrix}$$

$$= x \left\{ A \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} + y \left\{ A \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

Basis Vectors
$[1, 0]^T$ and $[0, 1]^T$

Matrices are incapable of distorting some parts of space differently than others.

All they can do is take the original coordinates on our space and skew, rotate, and scale them.

Identity Matrix

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\forall x \in \mathbb{R}^n, \qquad I_n x = x$$

Column space of A / All combinations of columns

$$Ax = \begin{bmatrix} 1 & 4 & 5 \\ 3 & 2 & 5 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} x_1 + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} x_2 + \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} x_3$$

$=$ linear combination of columns of A

- Column space of A = C(A)= all vectors Ax

  = all linear combinations of the columns

What is the column space of this example ?

Basis for the column space / Basis for the row space

- Include column 1 $= \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ in C. Include column 2 $= \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$ in C
- DO NOT INCLUDE COLUMN 3 $= \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$ IT IS NOT INDEPENDENT

$A = CR = \begin{bmatrix} 1 & 4 \\ 3 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ Row rank = Column rank = r = 2

The rows of R are a basis for the row space

A=CR

1. The r columns of C are independent (by their construction)
2. Every column of A is a combination of those r columns (because A=CR)
3. The r rows of R are independent (they contain the r by r matrix I)
4. Every row of A is a combination of those r rows (because A = CR)

Key facts

- The r columns of C are a basis for the column space of A: dimension r
  The r rows of R are a basis for the row space of A: dimension r

Basis for the column space / Basis for the row space

- Include column 1 $= \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ in C.   Include column 2 $= \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$ in C
- DO NOT INCLUDE COLUMN 3 $= \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$ IT IS NOT INDEPENDENT

- Basis has 2 vectors   A has rank r =2   n − r =3 − 2=1
- Counting Theorem   Ax = 0 has one solution x = (1, 1,−1)

There are n − r independent solutions to Ax =0

Rank

- If we have a general $n \times n$ matrix, it is reasonable to ask what dimension space the matrix maps into.
- The rank of a matrix A is the largest number of linearly independent columns amongst all subsets of columns.

Matrix Representation of Linear Functions

- A linear function (or map or transformation)

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

can be represented by a matrix A, $A \in \mathbb{R}^{m \times n}$, such that

$$f(x) = Ax = y, \qquad \forall x \in \mathbb{R}^n, y \in \mathbb{R}^m$$

- span($A_{:,1}, \cdots, A_{:,n}$) is called the column space of A
- rank(A)= dim(span($A_{:,1}, \cdots, A_{:,n}$))

Systems of Equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

expands to

$$\boldsymbol{A}_{1,:}\boldsymbol{x} = b_1$$

$$\boldsymbol{A}_{2,:}\boldsymbol{x} = b_2$$

$$\dots$$

$$\boldsymbol{A}_{m,:}\boldsymbol{x} = b_m$$

Null space of A

- If Ax = 0 then

$$\begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} & \cdots & \boldsymbol{A}_{1,n} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} & \cdots & \boldsymbol{A}_{2,n} \\ . & . & . & . \\ \boldsymbol{A}_{m,1} & \boldsymbol{A}_{m,2} & \cdots & \boldsymbol{A}_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ . \\ b_m \end{bmatrix}$$

- Every x in the nullspace of A is orthogonal to the row space of A
- Every *y* in the nullspace of A^{T} is orthogonal to the column space of *A*

$$N(A) \perp C(A^T) \qquad N(A^T) \perp C(A)$$

Dimensions $\quad n-r \quad r \qquad m-r \quad r$

System of Linear Equations

$$\boldsymbol{A}_{1,1}x_1 + \boldsymbol{A}_{1,2}x_2 + \cdots + \boldsymbol{A}_{1,n}x_n = b_1$$
$$\boldsymbol{A}_{2,1}x_1 + \boldsymbol{A}_{2,2}x_2 + \cdots + \boldsymbol{A}_{2,n}x_n = b_2$$
$$\cdots$$
$$\boldsymbol{A}_{m,1}x_1 + \boldsymbol{A}_{m,2}x_2 + \cdots + \boldsymbol{A}_{m,n}x_n = b_m$$

We can write these a

$$\begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} & \cdots & \boldsymbol{A}_{1,n} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} & \cdots & \boldsymbol{A}_{2,n} \\ . & . & . & . \\ \boldsymbol{A}_{m,1} & \boldsymbol{A}_{m,2} & \cdots & \boldsymbol{A}_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ . \\ b_m \end{bmatrix}$$

$$\boldsymbol{Ax} = \boldsymbol{b}$$

Solving Systems of Equations

$\boldsymbol{Ax} = \boldsymbol{b}$

$\boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{x} \in \mathbb{R}^{n}, \boldsymbol{b} \in \mathbb{R}^{m}$

- A linear system of equations can have
  - No solution - $m > n$ - overdetermined problem (No. of equations > No. of variables).
  - Many solutions: $m < n$ - underdetermined problem (No. of equations < No. of variables). Infinitely many solutions.
  - Exactly one solution: this means multiplication by the matrix is an invertible function. $m = n$ and $\det(\boldsymbol{A}) \neq 0$, the solution is unique, $\boldsymbol{x} = \boldsymbol{A}^{-1} b$.

System of Linear Equations

- Given A and b, solve x in $Ax = b$
- What kind of A that makes Ax = b always have a solution?
  - Since $Ax = \sum_i x_i A_{:,i}$, the column space of A must contain $\mathbb{R}^m$, that is, $\mathbb{R}^m \subseteq span(A_{:,1}, \ldots, A_{:,n})$
  - Implies $n \geq m$
- When does Ax = b always have exactly one solution?
  - A has at most m columns; otherwise there is more than one x parametrizing each b
  - Implies n = m and the columns of A are linear independent with each other
  - A^{-1} exists at this time, and $x = A^{-1}b$

Matrix Inversion

- Matrix inverse:

$$A^{-1}A = I_n$$

- Solving a system using an inverse:

$$\boldsymbol{Ax} = \boldsymbol{b}$$
$$\boldsymbol{A}^{-1}\boldsymbol{Ax} = \boldsymbol{A}^{-1}\boldsymbol{b}$$
$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

Numerically unstable, but useful for abstract analysis

Invertibility

- Matrix can’t be inverted if…
  - More rows than columns
  - More columns than rows
  - Redundant rows/columns (“linearly dependent”, “low rank”)

Norms

- Functions that measure how “large” a vector is
- Similar to a distance between zero and the point represented by the vector
  - $f(x) = 0 \Rightarrow x = 0$
  - $f(x + y) \leq f(x) + f(y)$ the triangle inequality
  - $\forall \alpha \in \mathbb{R}, f(\alpha x) = |\alpha| f(x)$

Norms

- L^p norm

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

- Most popular norm: L^2 norm
- L^1 norm $\|x\|_1 = \sum_i |x_i|$
- Max norm, infinite p:

$$\|x\|_\infty = \max_i |x_i|$$

- $x^T y = ||x|| ||y|| \cos\theta$, where θ is the angle between x and y.
- x and y are orthonormal iff
  - $x^T y = 0$ (orthogonal) and
  - $||x|| = ||y|| = 1$ (unit vectors)

Matrix Norms

- Frobenius norm

$$\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$$

- An orthogonal matrix is a square matrix whose column (resp. rows) are mutually orthonormal, i.e.,

$$A^T A = AA^T = I.$$

Implies $$A^{-1} = A^T$$

Special Matrices and Vectors

- Unit vector:

$$\|x\|_2 = 1.$$

- Symmetric Matrix:

$$A = A^T$$

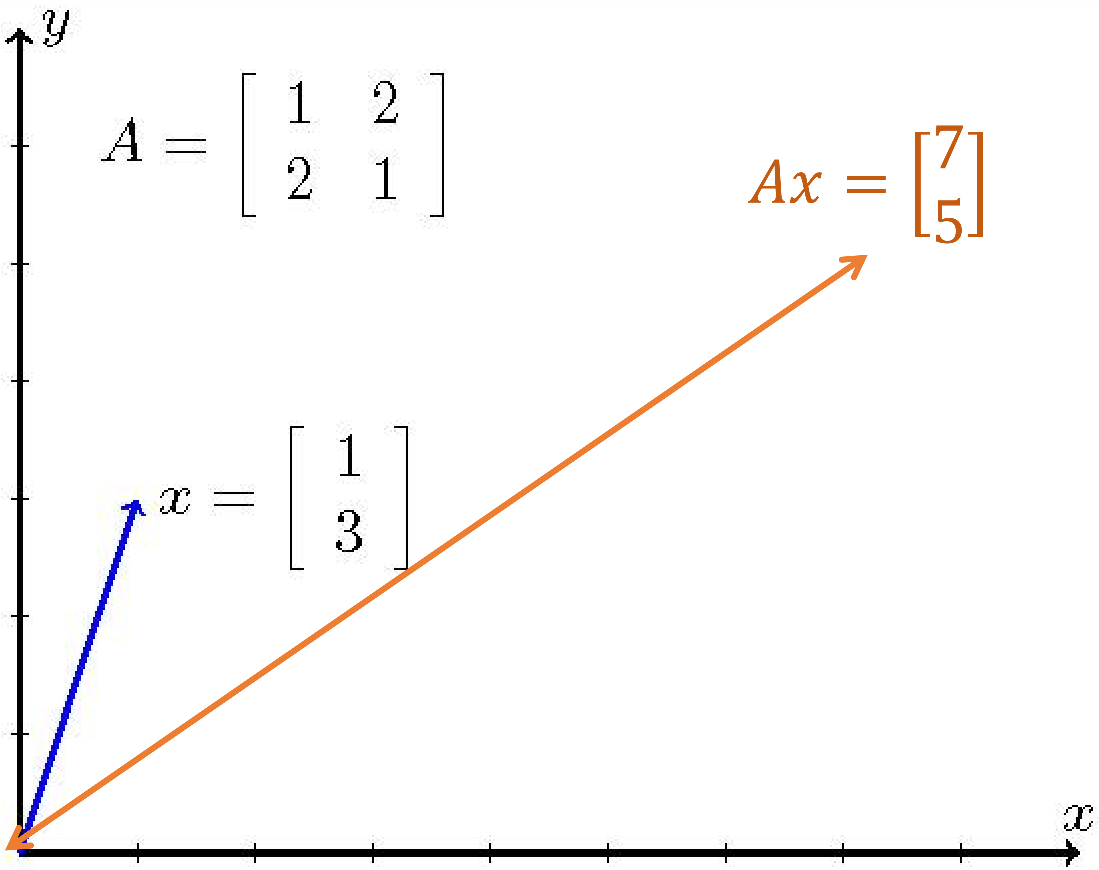
- Orthogonal matrix:

$$A^T A = AA^T = I.$$

$$A^{-1} = A^T$$

Eigendecomposition

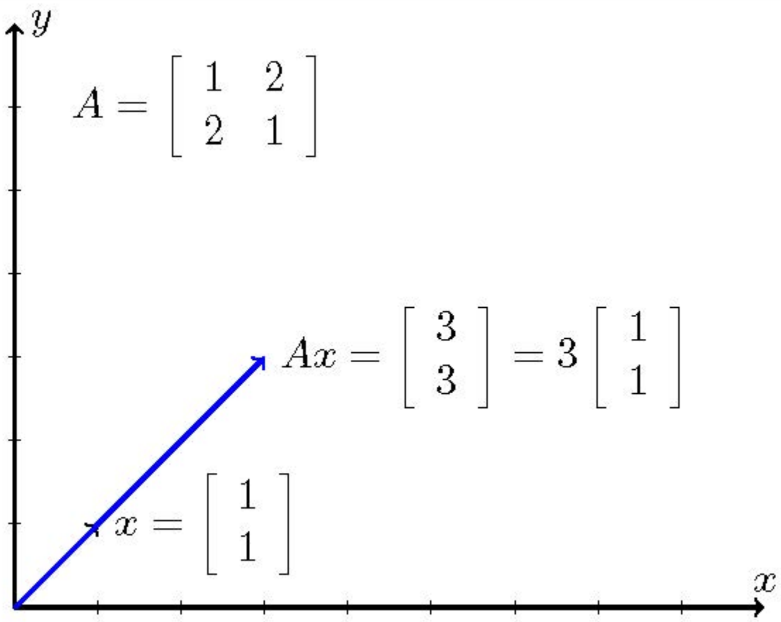
- Decomposition:
- Integers can be decomposed into prime factors
  - E.g., $12 = 2 \times 2 \times 3$
  - Helps identify useful properties, e.g., 12 is not divisible by 5
- Can we decompose matrices to identify information about their functional properties more easily?



- What happens when a matrix hits a vector?
  - The vector gets transformed into a new vector
  - The vector may also get scaled

Eigenvectors

For a given square matrix A, there exist special vectors which refuse to stray from their path.



Eigenvectors and Eigenvalues

- An eigenvector of a square matrix A is a non-zero vector v such that multiplication by A alters only the scale of v :

$$\boldsymbol{A}v = \lambda v$$

- where $\lambda \in \mathbb{R}$ is called the eigenvalue corresponding to this eigenvector
- If v is an eigenvector, so is any its scaling cv, $c \in \mathbb{R}$, $c \neq 0$
  - cv has the same eigenvalue
  - Thus, we usually look for unit eigenvectors

Eigenvalues and Eigenvectors *example*

- Example: Let

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$$

  - $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an eigenvector corresponding to the eigenvalue $\lambda_1 = 2$ because

$$A\mathbf{x}_1 = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_1 \mathbf{x}_1$$

  - $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is also an eigenvector corresponding to the eigenvalue $\lambda_2 = -1$ because

$$A\mathbf{x}_2 = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} = -1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \lambda_2 \mathbf{x}_2$$

Eigenvalues and Eigenvectors, *cont'd*

- Example: Let

$$A = \begin{bmatrix} 92 & -32 & -15 \\ -64 & 34 & 39 \\ 176 & -68 & -99 \end{bmatrix}$$

  - Matrix A has several eigenvectors $\mathbf{x}$ and corresponding eigenvalues λ:

λ	$\mathbf{x}$	$A\mathbf{x}$	$\lambda\mathbf{x}$
88.519	$\begin{bmatrix} 1 \\ -0.399 \\ 1.083 \end{bmatrix}$	$\begin{bmatrix} 88.519 \\ -35.319 \\ 95.889 \end{bmatrix}$	$\begin{bmatrix} 88.519 \\ -35.319 \\ 95.889 \end{bmatrix}$
-70.791	$\begin{bmatrix} 1 \\ -16.639 \\ 46.349 \end{bmatrix}$	$\begin{bmatrix} -70.791 \\ 1177.895 \\ -3281.123 \end{bmatrix}$	$\begin{bmatrix} -70.791 \\ 1177.895 \\ -3281.123 \end{bmatrix}$
9.272	$\begin{bmatrix} 1 \\ 2.584 \\ 0.003 \end{bmatrix}$	$\begin{bmatrix} 9.272 \\ 23.96 \\ 0.025 \end{bmatrix}$	$\begin{bmatrix} 9.272 \\ 23.96 \\ 0.025 \end{bmatrix}$

Uses of Eigenvalues and Eigenvectors

- Eigenvalues and eigenvectors are used to “decompose” a matrix into its constituent parts in order to simplify complex operations.
  - “eigendecomposition”

- A major example of its use is in performing principal component analysis (PCA) .
  - We used PCA to reduce the four features of the Iris dataset down to two features so we could draw a 2-D graph that showed the clusters.

Let $u_1, u_2, \dots, u_n$ be the eigenvectors of a matrix A and let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the corresponding eigenvalues.

Consider a matrix U whose columns are $u_1, u_2, \dots, u_n$.

Now

$$AU = A\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \dots & u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ Au_1 & Au_2 & \dots & Au_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$= \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \lambda_1 u_1 & \lambda_2 u_2 & \dots & \lambda_n u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$= \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \dots & u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} = U\Lambda$$

where Λ is a diagonal matrix whose diagonal elements are the eigenvalues of A.

- $AU = U\Lambda$
- If U^{-1} exists, we can write

  $A = U\Lambda U^{-1}$ [eigenvalue decomposition]

  $U^{-1}AU = \Lambda$ [diagonalization of A]

Eigendecomposition

$$\boldsymbol{A}v = \lambda v$$

- Eigendecomposition of a diagonalizable matrix:

$$\boldsymbol{A} = \boldsymbol{U}\,\mathrm{diag}(\lambda)\,\boldsymbol{U}^{-1}$$

- If a matrix $\mathbf{A}$ has n linearly independent eigenvectors $\{\mathbf{u}^1, \dots, \mathbf{u}^n\}$ with corresponding eigenvalues $\{\lambda_1, \dots, \lambda_n\}$, the eigen decomposition of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{U\Lambda U}^{-1}$$

  - Columns of the matrix $\mathbf{U}$ are the eigenvectors, i.e., $\mathbf{U} = [\mathbf{u}^1, \dots, \mathbf{u}^n]$
  - $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues, i.e., $\mathbf{\Lambda} = [\lambda_1, \dots, \lambda_n]$

Eigendecomposition

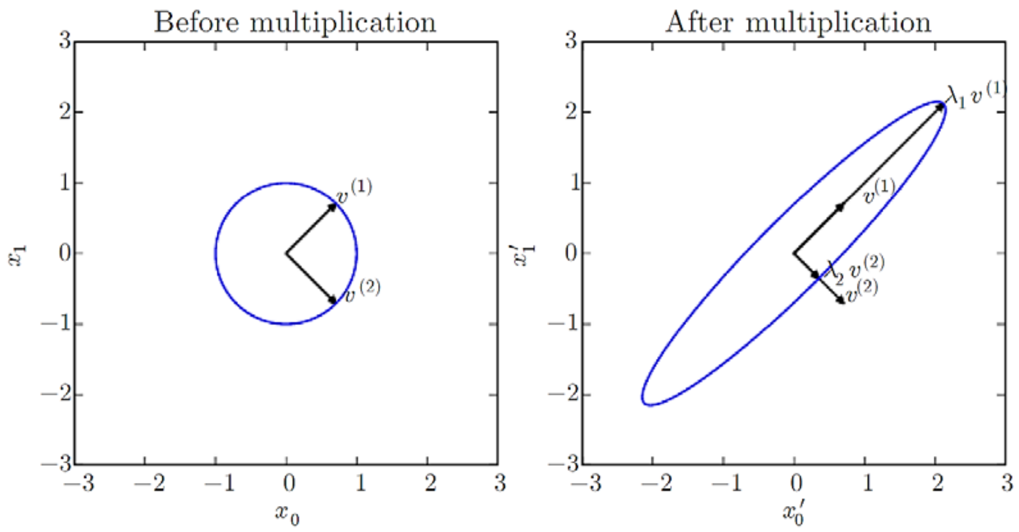
- If A is a real symmetric matrix
  - The eigenvectors are orthogonal
- is guaranteed to have an eigen decomposition , where $\mathbf{Q}$ is an orthogonal matrix

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$$

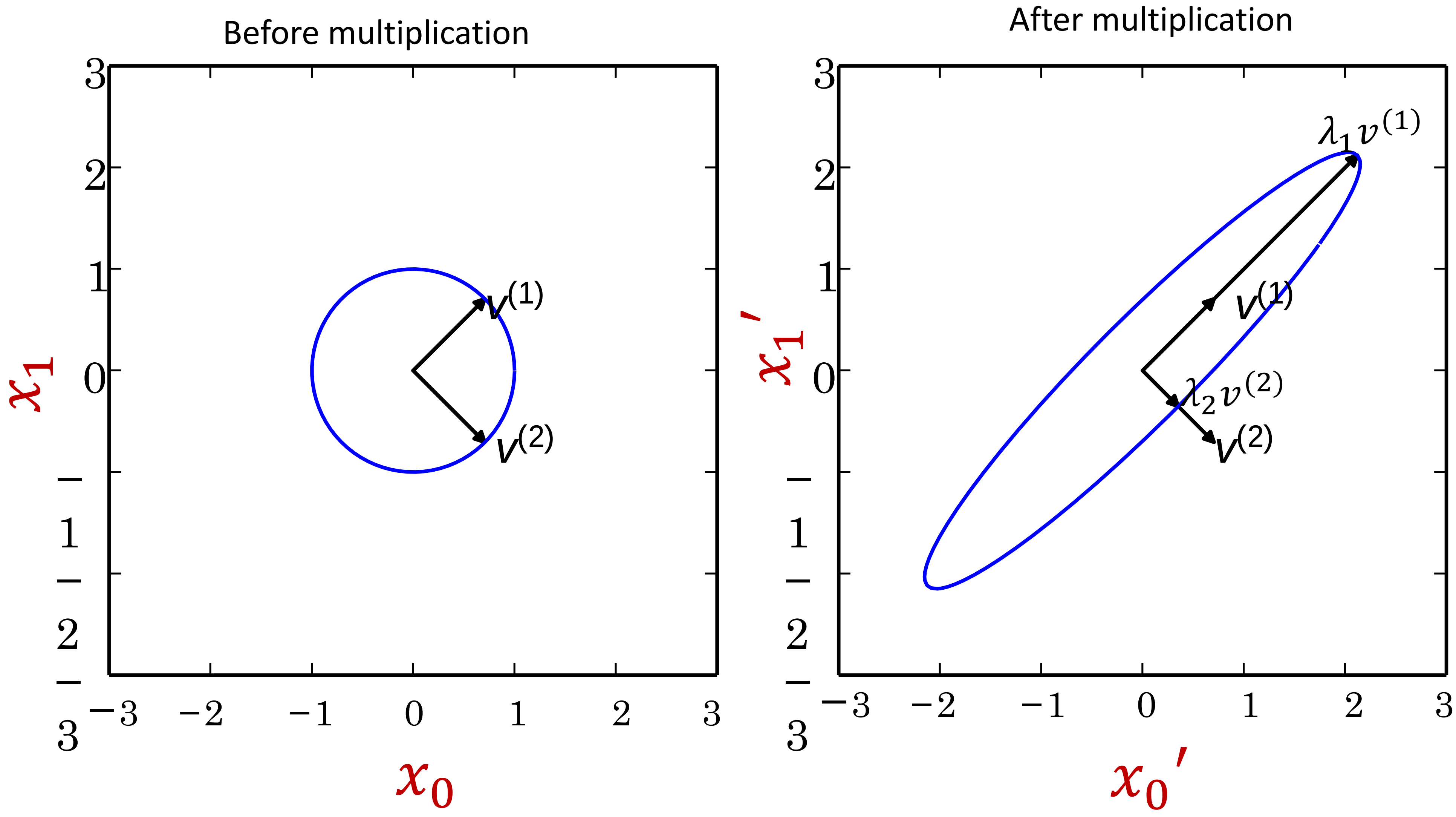
Geometric interpretation of the eigenvalues and eigenvectors

they allow to stretch the space in specific directions

- Left figure: the two eigenvectors $\mathbf{v}^1$ and $\mathbf{v}^2$ are shown for a matrix, where the two vectors are unit vectors (i.e., they have a length of 1)
- Right figure: the vectors $\mathbf{v}^1$ and $\mathbf{v}^2$ are multiplied with the eigenvalues λ_1 and λ_2
  - We can see how the space is scaled in the direction of the larger eigenvalue λ_1
- E.g., this is used for dimensionality reduction with PCA (principal component analysis) where the eigenvectors corresponding to the largest eigenvalues are used for extracting the most important data dimensions



Because Q = [v(1), ⋯ , v(n)] is an orthogonal matrix, we can think of A as scaling space by λ_i in direction $v^{(i)}$



- Let $v_1, v_2, \cdots, v_n$ be the eigen vectors of A and let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be corresponding eigen values

$$Av_1 = \lambda_1 v_1, Av_2 = \lambda_2 v_2, \cdots, Av_n = \lambda_n v_n$$

- If a vector x in $\mathbb{R}^n$ is represented using $v_1, v_2, \cdots, v_n$ as basis then

$$x = \sum_{i=1}^{n} \alpha_i v_i$$

$$\text{Now, } Ax = \sum_{i=1}^{n} \alpha_i A v_i = \sum_{i=1}^{n} \alpha_i \lambda_i v_i$$

- The matrix multiplication reduces to a scalar multiplication if the eigen vectors of A are used as a basis.

Singular Value Decomposition

- More general matrix need not be square.
- Every real matrix $A \in \mathbb{R}^{m \times n}$ has a singular value decomposition:

$$A = UDV^T$$

where U $\in \mathbb{R}^{m \times m}$, $D \in \mathbb{R}^{m \times n}$ and V$\in \mathbb{R}^{n \times n}$

  - U and V are orthogonal matrices, and their columns are called the left and right-singular vectors respectively
  - Elements along the diagonal of D are called the singular values
- Left-singular vectors of A are eigenvectors of AA^T
- Right-singular vectors of A are eigenvectors of A^TA
- Non-zero singular values of A are square roots of eigenvalues of AA^T (or)

$S = S^T$ Real Eigenvalues and Orthogonal Eigenvectors

$S = S^T$ has orthogonal eigenvectors $x^T y = 0$

- Proof:

$$Sx = \lambda x \qquad Sy = \lambda y \qquad \lambda \neq \alpha \qquad S^T = S$$

1. Transpose to $x^T S^T = \lambda x^T$ and use $S^T = S$ $\qquad x^T S y = \lambda x^T y$
2. Multiply $Sy = \alpha y$ by x^T $\qquad x^T S y = \alpha x^T y$
3. Now $\alpha x^T y = \lambda x^T y$. Since $\lambda \neq \alpha$, $x^T y$ must be 0.

Eigenvectors of S go into Orthogonal Matrix Q

$$S[q_1 \ \ldots q_n] = [\lambda_1 q_1 \ \cdots \lambda_n q_n] = [q_1 \ \ldots q_n]\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

That says $SQ = Q\Lambda$ $\quad S = Q\Lambda Q^{-1} = Q\Lambda Q^T$

$S = Q\Lambda Q^T$ is a sum of $\lambda_1 q_1 q_1^T + \cdots + \lambda_r q_n q_n^T$ of rank one matrices

With $S = A^T A$ this will lead to the singular values of A

$A = U\Sigma V^T$ is a sum $\sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$ of rank one matrices

Singular values σ_1 to σ_r in Σ. Singular vectors in U and V

$A^T A$ is square, symmetric, nonnegative definite

- Square
- Symmetric : $(A^T A)^T =\ ?$

Moore-Penrose Pseudoinverse

$$\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{y}$$

- If the equation has:
  - Exactly one solution: this is the same as the inverse.
  - No solution: this gives us the solution with the smallest error $||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2$.
  - Many solutions: this gives us the solution with the smallest norm of $\boldsymbol{x}$.

Computing the Pseudoinverse

- The SVD allows the computation of the pseudoinverse:

$$A^{+} = V D^{+} U^{T}$$

Take reciprocal of non-zero entries

Trace

$$Tr(A) = \sum_i A_{i,i}.$$

$$Tr(ABC) = Tr(CAB) = Tr(BCA)$$

Manifolds

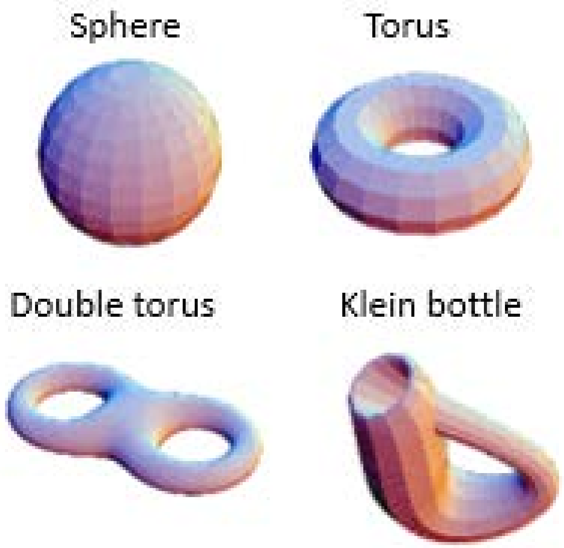
- Earlier we learned that hyperplanes generalize the concept of planes in high-dimensional spaces
  - Similarly, manifolds can be informally imagined as generalization of the concept of surfaces in high-dimensional spaces
- To begin with an intuitive explanation, the surface of the Earth is an example of a two-dimensional manifold embedded in a three-dimensional space
  - This is true because the Earth looks locally flat, so on a small scale it is like a 2-D plane
  - However, if we keep walking on the Earth in one direction, we will eventually end up back where we started
    - This means that Earth is not really flat, it only looks locally like a Euclidean plane, but at large scales it folds up on itself, and has a different global structure than a flat plane

Manifolds

- An n-dimensional manifold is defined as a topological space with the property that each point has a neighborhood that is homeomorphic to the Euclidean space of dimension n
- This means that a manifold locally resembles Euclidean space near each point
- Informally, a Euclidean space is locally smooth, it does not have holes, edges, or other sudden changes, and it does not have intersecting neighborhoods
- Although the manifolds can have very complex structure on a large scale, resemblance of the Euclidean space on a small scale allows to apply standard math concepts

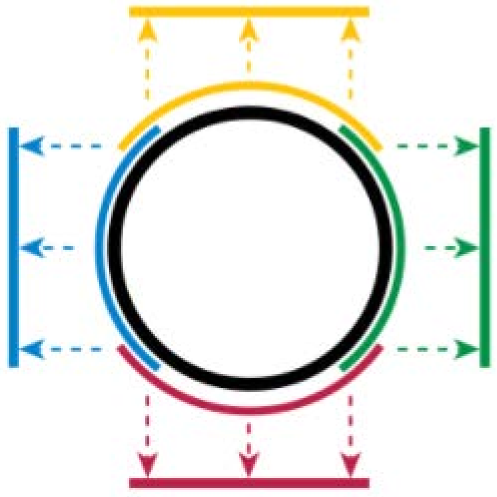
Examples of 2-dimensional manifolds

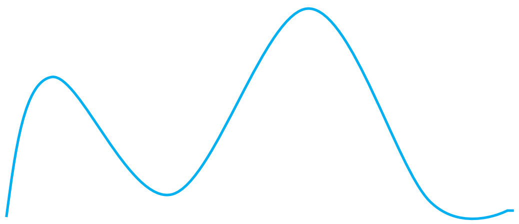
- The surfaces in the figure have been conveniently cut up into little rectangles that were glued together
- Those small rectangles locally look like flat Euclidean planes



Examples of one-dimensional manifolds

- :a circle is a l-D manifold embedded in 2-D, where each arc of the circle locally resembles a line segment
- Lower figures: other examples of 1-D manifolds
- Note that a number 8 figure is not a manifold because it has an intersecting point (it is not Euclidean locally)
- It is hypothesized that in the real-world, high-dimensional data (such as images) lie on low-dimensional manifolds embedded in the high-dimensional space
  - E.g., in ML, let's assume we have a training set of images with size $224 \times 224 \times 3$ pixels
  - Learning an arbitrary function in such high-dimensional space would be intractable
  - Despite that, all images of the same class ("cats" for example) might lie on a low-dimensional manifold
  - This allows function learning and image classification





Manifolds

- Example:
  - The data points have 3 dimensions (left figure), i.e., the input space of the data is 3-dimensional
  - The data points lie on a 2-dimensional manifold, shown in the right figure
  - Most ML algorithms extract lower-dimensional data features that enable to distinguish between various classes of high-dimensional input data
    - The low-dimensional representations of the input data are called embeddings