CM22009

Introduction to Linear Algebra

Dr. Georgios Exarchakis

Announcements

Coursework deadline on Friday!

Math

Probability

- Probability Space
- Random Variables
- Conditional Prob/Independence
- Probability Densities/Distributions
- Expectation / (Co)Variance
- Bayes Theorem
- Central Limit Theorem
- Law of Large Numbers

Linear Algebra

- Vector Spaces
- Matrix Operations
- Linear Independence
- Orthogonality
- Eigenvalues/Eigenvectors

Multiple Dimensions

Machine Learning is a successful approach at dealing with statistics of high dimensional observations

A **vector** is an ordered collection of numbers. An n-dimensional vector of real numbers is represented as $\mathbf{v}$ or $\vec{v} \in \mathbb{R}^n$:

e.g. $\mathbf{v} = \vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$

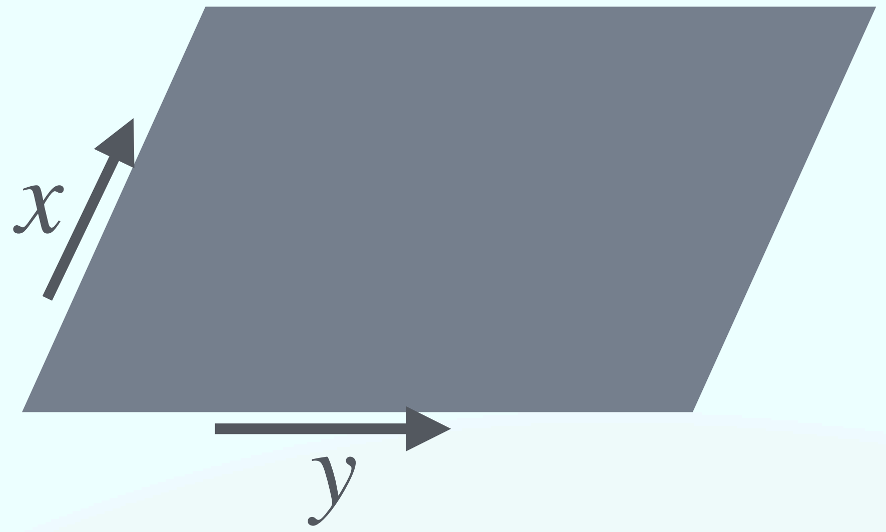
It is commonly used to store the **features** of a datapoint

A **matrix** is a collection of numbers in a two dimensional layout, with rows or columns, that can be viewed as vectors, often represented by an uppercase bold letter $\mathbf{A} \in \mathbb{R}^{m \times n}$

$$\mathbf{A} = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{m,n} \\ \vdots & & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix}$$

Examples

Spatial coordinates $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$



Student marks $\mathbf{v}_s$(or $\mathbf{v}^{(s)}$) $= \begin{bmatrix} \mathrm{c}_1 \\ \mathrm{c}_2 \\ \vdots \\ \mathrm{c}_n \end{bmatrix}$

$$\mathbf{V} = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_N^T \end{bmatrix} =$$

Marks	Module 1	Module 2	...	Module n
Student 1	c1	c2		cn
Student 2	c1	c2		cn
Student 3	c1	c2		cn
...				

Multiple Dimensions

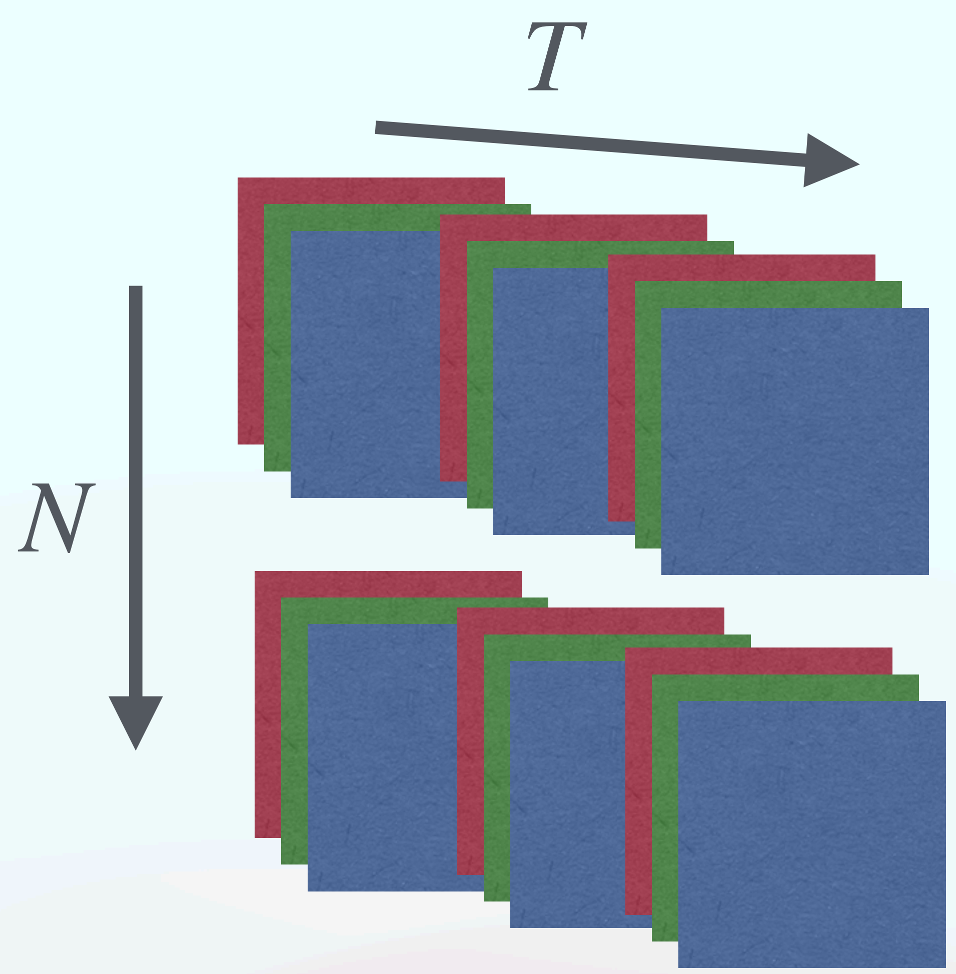
Machine Learning is a successful approach at dealing with statistics of high dimensional observations

A **tensor** is an ordered collection of numbers with more dimensions than a vector or a matrix, e.g $\mathbf{A} \in \mathbb{R}^{m \times n \times \ldots \times q}$.

An **example** of a tensor that appears frequently in machine learning is a data tensor that holds an ensemble of images or videos:

$\mathbf{I} \in \mathbb{R}^{N \times C \times H \times W}$

$\mathbf{V} \in \mathbb{R}^{N \times T \times C \times H \times W}$



Where N: the number of Images/Videos, T the number of frames, C the number of channels (red/green/blue), H the vertical resolution, and W the horizontal resolutions

Often software packages treat vectors and matrices as 1-dimensional and 2-dimensional tensors respectively.

Matrix Transpose

- It is often useful to rearrange the position of rows and columns in a matrix.
- Definition:

The transpose, $\mathbf{A}^T \in \mathbb{R}^{m \times n}$, of a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ is formed by reflecting it around the principal diagonal

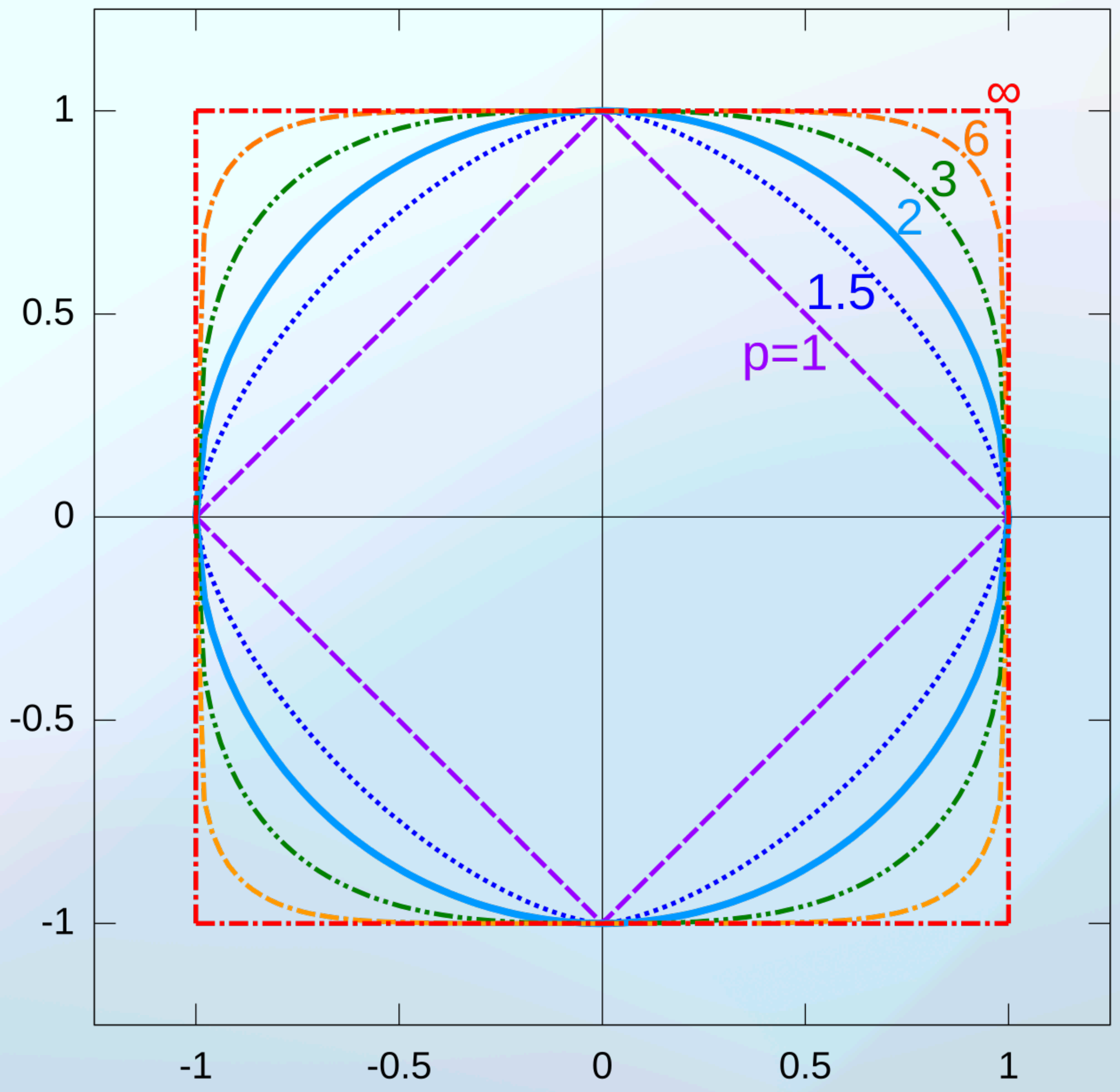
$$\mathbf{A} = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{m,n} \\ \vdots & & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix} \quad \mathbf{A}^T = \begin{bmatrix} A_{1,1} & A_{2,1} & \cdots & A_{m,1} \\ A_{1,2} & A_{2,2} & \cdots & A_{m,n} \\ \vdots & & \ddots & \vdots \\ A_{1,n} & A_{2,n} & \cdots & A_{m,n} \end{bmatrix}$$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \mathbf{v}^T = [v_1 \quad v_2 \quad \cdots \quad v_n]$$

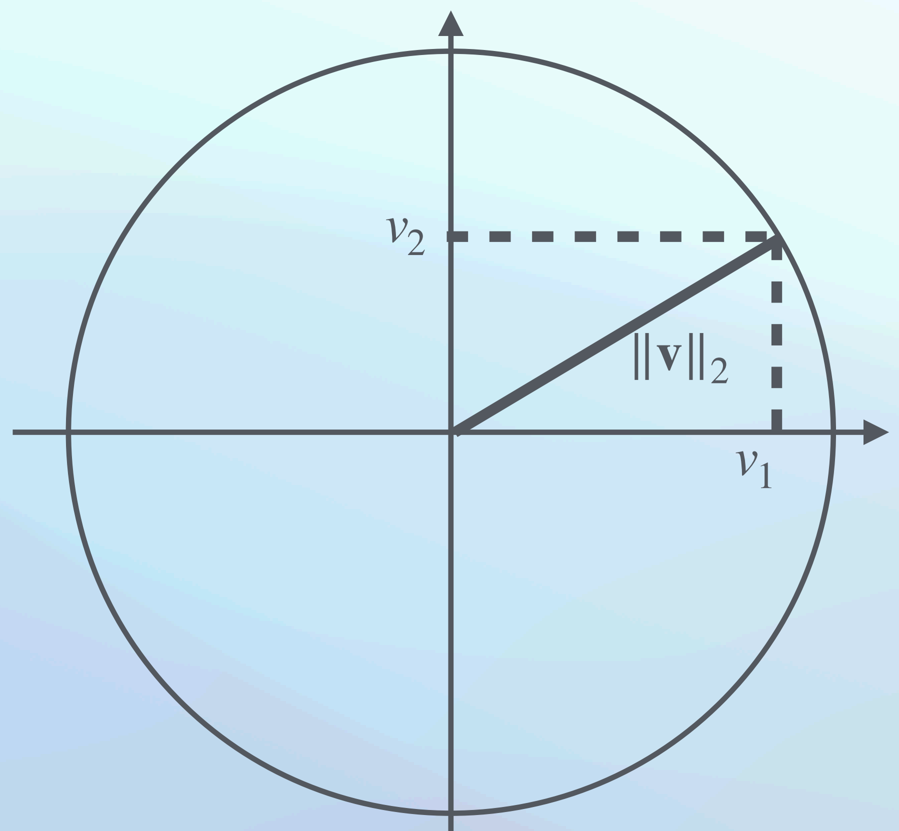
Vector Norms

For a vector $\mathbf{v}$ the ℓ_p norm is defined as :

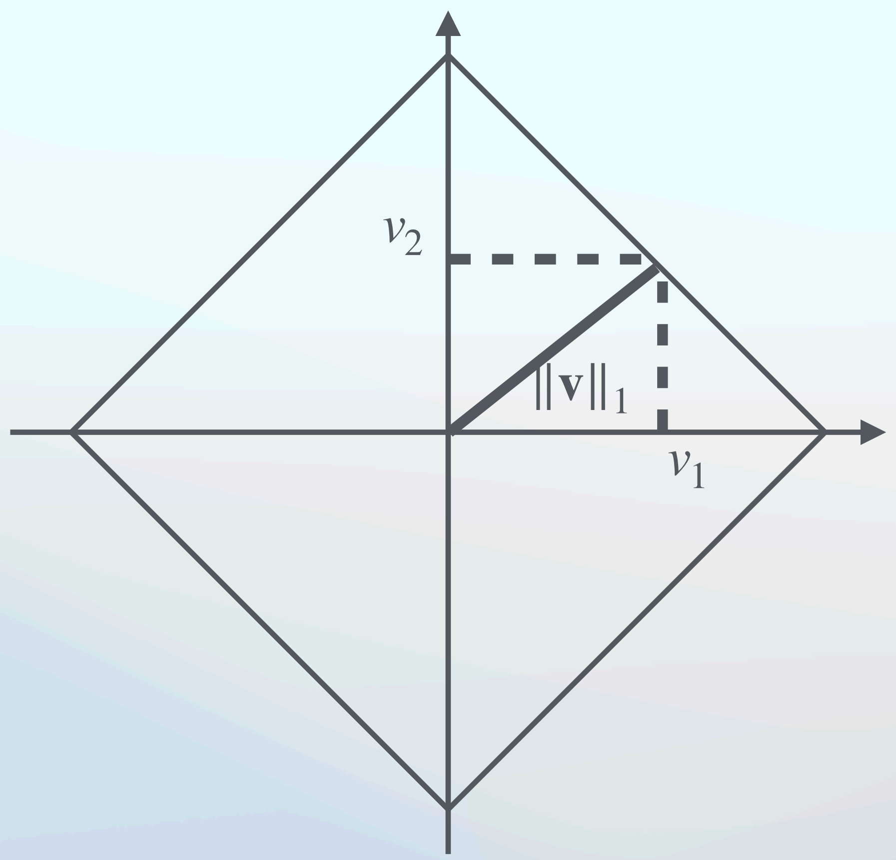
$$\|\mathbf{v}\|_p = \left(\sum_{d=1}^{D} |v_d|^p \right)^{1/p}$$



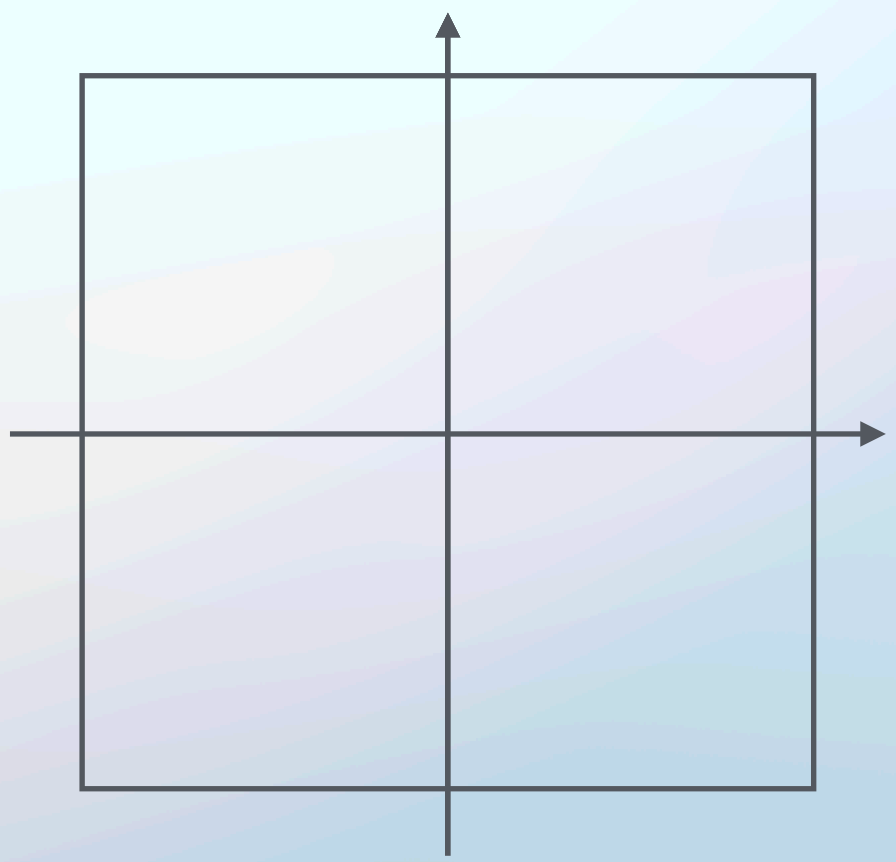
For $D = 2$



$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2}$$



$$\|\mathbf{v}\|_1 = |v_1| + |v_2|$$

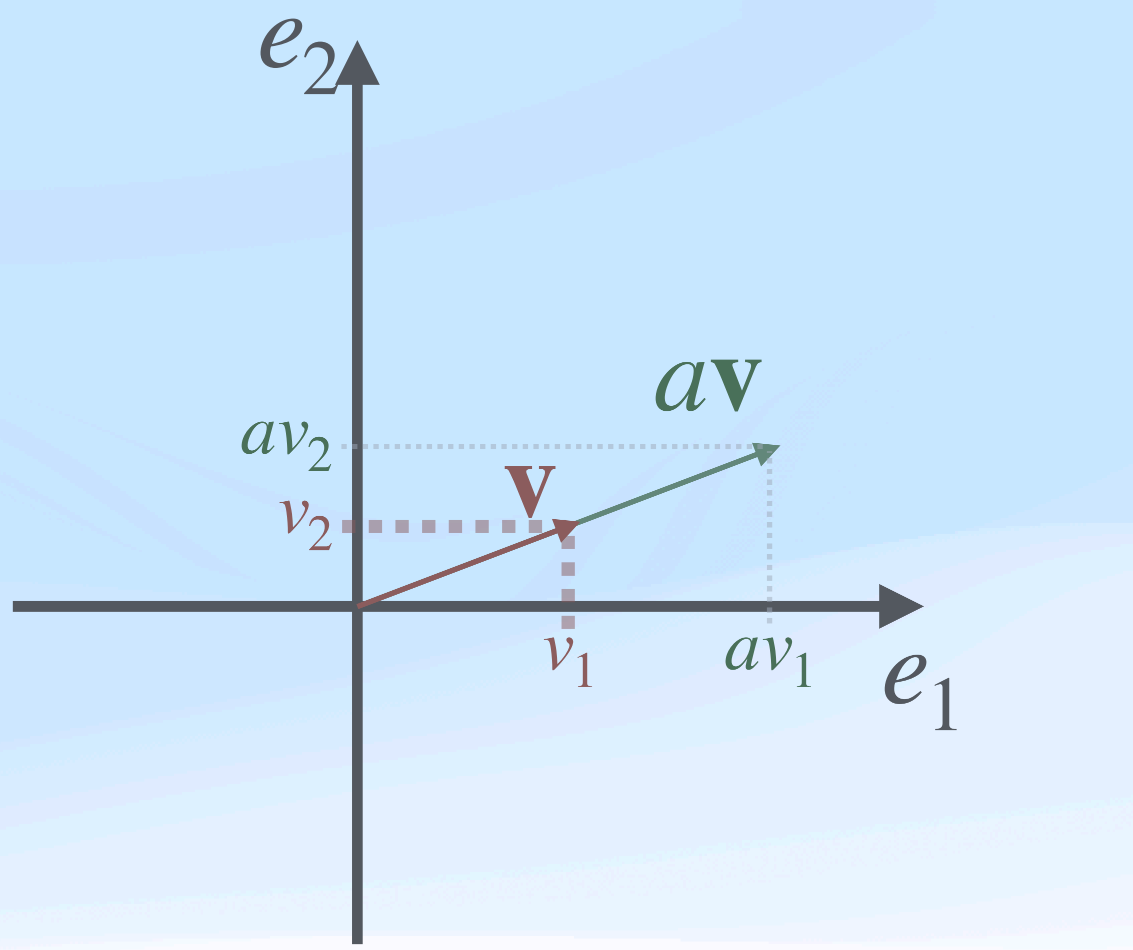


$$\|\mathbf{v}\|_\infty = \max\{v_1, v_2\}$$

Vector Operations

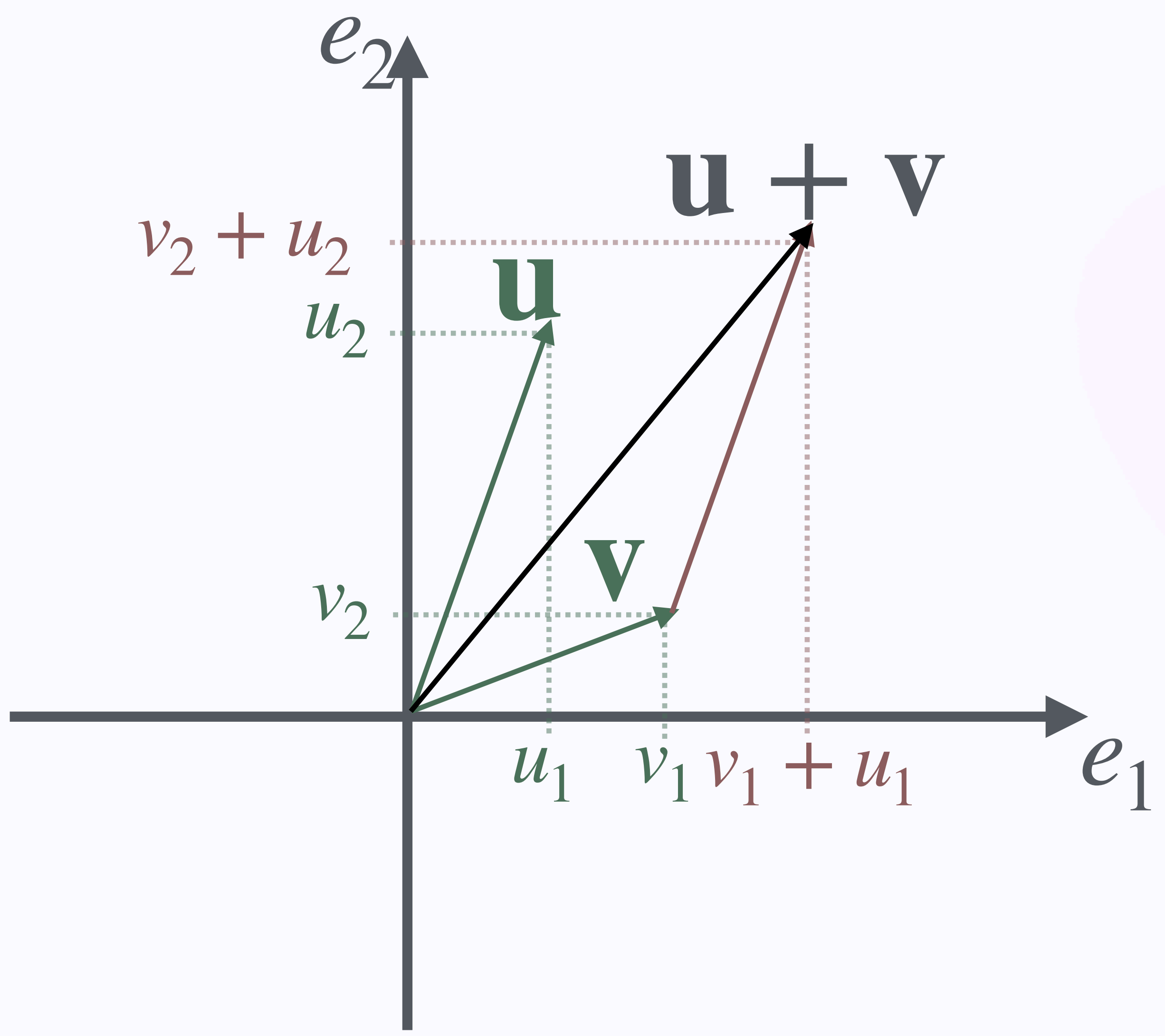
- Multiplication of a vector, $\mathbf{v} \in \mathbb{R}^n$, with a scalar, $a \in \mathbb{R}$:

$$a\mathbf{v} = \begin{bmatrix} av_1 \\ av_2 \\ \vdots \\ av_n \end{bmatrix}$$



- Addition of two vectors, $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{u} \in \mathbb{R}^n$, :
  (they must have the same dimensions)

$$\mathbf{v} + \mathbf{u} = \begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ \vdots \\ v_n + u_n \end{bmatrix}$$



Basis Vectors

Vectors as linear combinations of vectors

Linear dependence:

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is said to be linearly dependent if there is a set of scalar values $\{a_1, a_2, \ldots, a_n\}$, not all 0, such that:

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n = \mathbf{0}$$

Linear independence:

If a set of vectors is not linear dependent they are linearly independent.

Span of vectors:

For a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, the span of S is the set of all vectors that can be written as linear combinations of the vectors in S.

$$\text{span}(S) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n\}$$

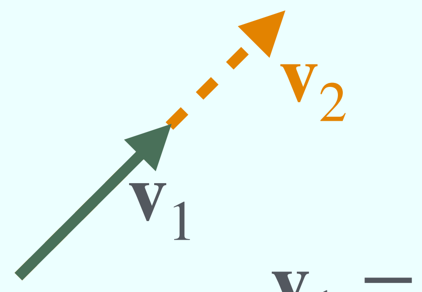
Basis:

A basis B of a vector space, e.g. $\mathbb{R}^n$, is a subset of that space, e.g. $S \subset \mathbb{R}^n$, that spans the vector space, e.g. $\text{span}(S) = \mathbb{R}^n$.

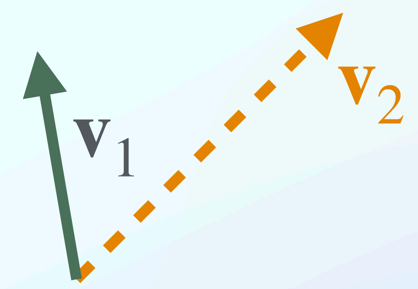
Example

Dependent | Independent

2D:



$\mathbf{v}_1 = 0.5\mathbf{v}_2$



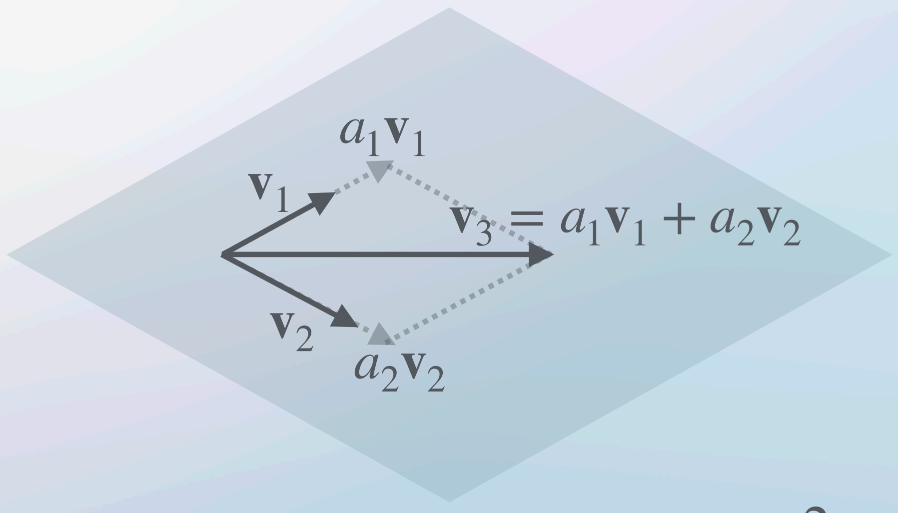
So for $a_1 = -2$ and $a_2 = 1$

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 = -2\mathbf{v}_1 + \mathbf{v}_2$$
$$= -2 \cdot 0.5\mathbf{v}_2 + \mathbf{v}_2 = 0$$

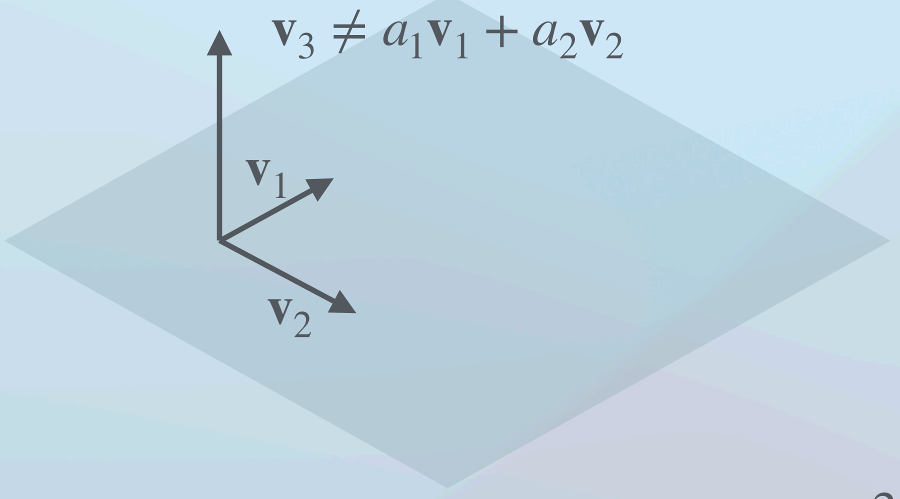
Unless $a_1 = a_2 = 0$

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 \neq 0$$

3D:



$\text{span}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}) = \mathbb{R}^2$



$\text{span}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}) = \mathbb{R}^3$

Basis Vectors

Vectors as linear combinations of vectors

Linear dependence:

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is said to be linearly dependent if there is a set of scalar values $\{a_1, a_2, \ldots, a_n\}$, not all 0, such that:

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n = \mathbf{0}$$

Linear independence:

If a set of vectors is not linear dependent they are linearly independent.

Span of vectors:

For a set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, the span of S is the set of all vectors that can be written as linear combinations of the vectors in S.

$$\text{span}(S) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n\}$$

Basis:

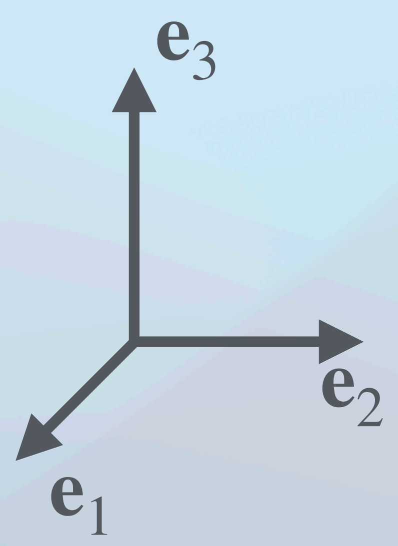
A basis B of a vector space, e.g. $\mathbb{R}^n$, is a subset of that space, e.g. $S \subset \mathbb{R}^n$, that spans the vector space, e.g. $\text{span}(S) = \mathbb{R}^n$.

Example

The **Standard (Canonical) Basis** $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, where $\mathbf{e}_i$ is a vector with the i-th dimension equal to 1 and all others equal to 0.

For example the standard basis for $\mathbb{R}^3$ is:

$$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix} \right\}$$



Dot product

For two vectors, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we define the dot product as:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i$$

The dot product can also be written as:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta$$

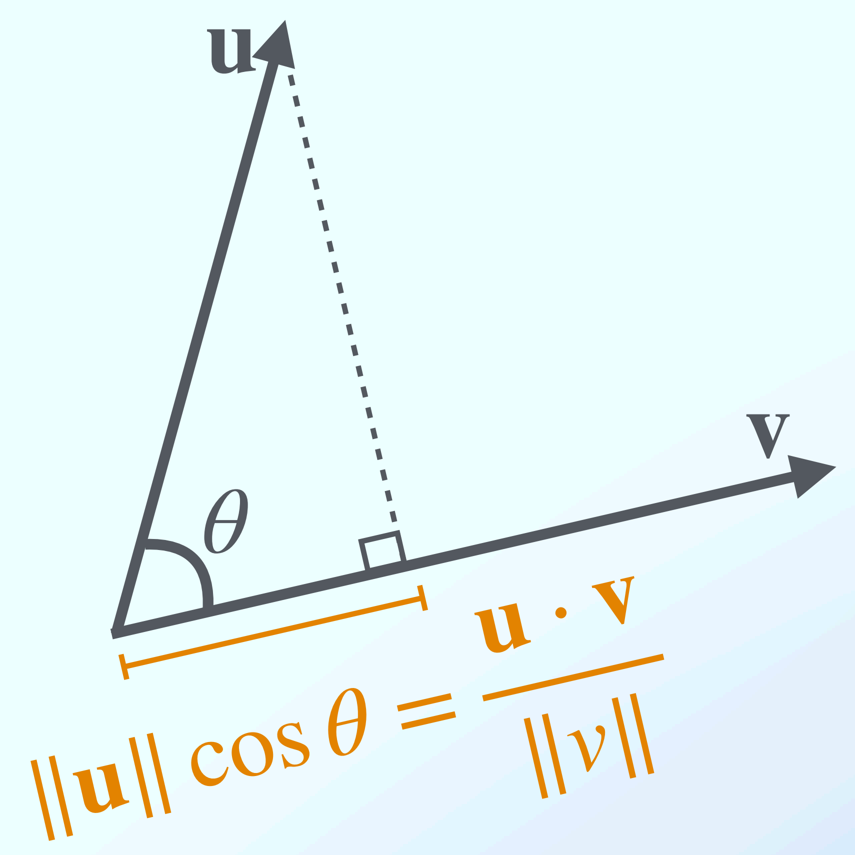
Length of a vector:

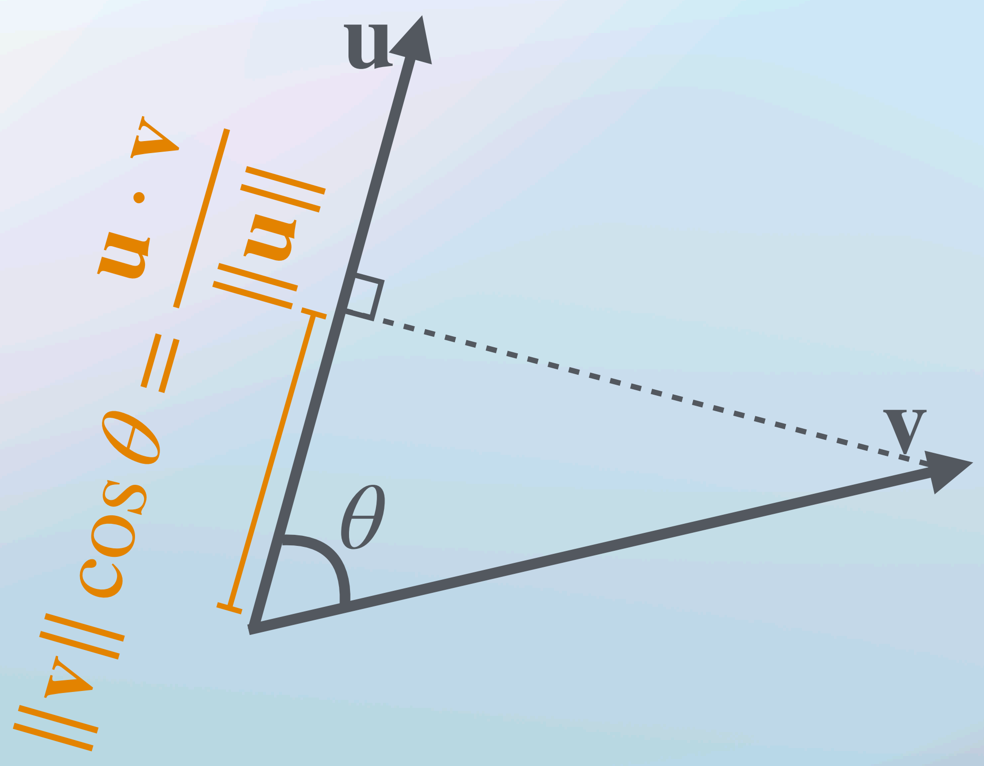
$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{\sum_{i=1}^{n} v_i^2} = \sqrt{\|\mathbf{v}\|^2 \cancel{\cos 0}}$$

Angle between vectors:

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

Example



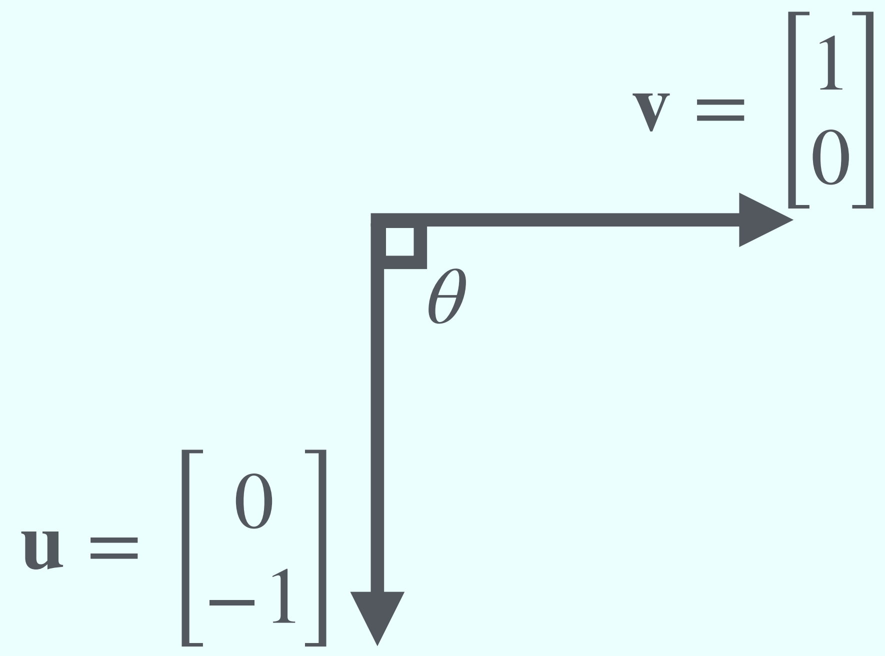


Orthogonal vectors

Two non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n \times n}$ are called orthogonal if their inner product is 0, i.e. $\mathbf{u} \cdot \mathbf{v} = 0$.

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta$$

Example



$$\mathbf{u} \cdot \mathbf{v} = 1 \cdot 0 + 0 \cdot (-1) = 0$$

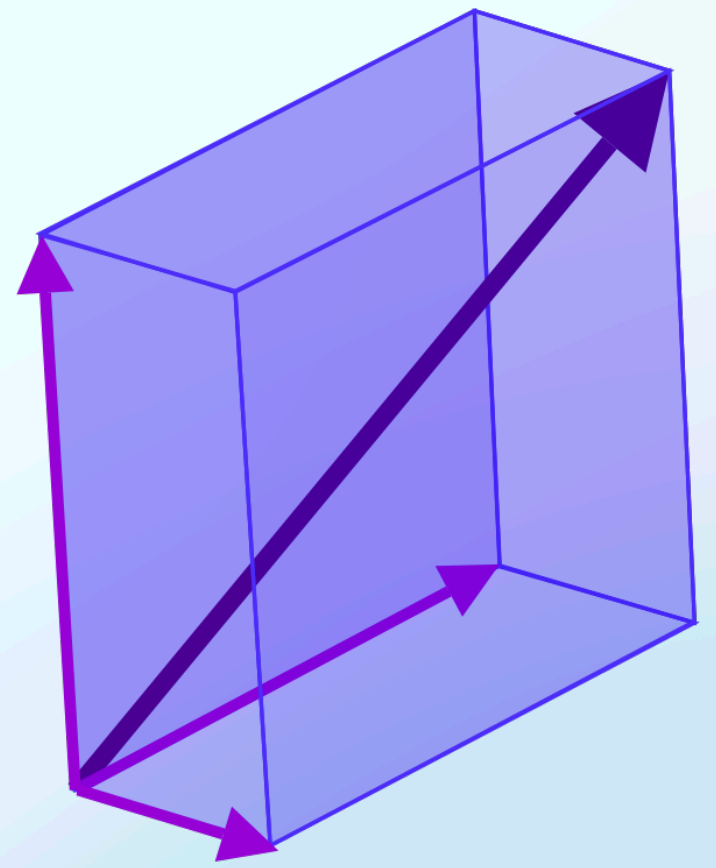
$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta = 1 \cdot 1 \cdot 0$$

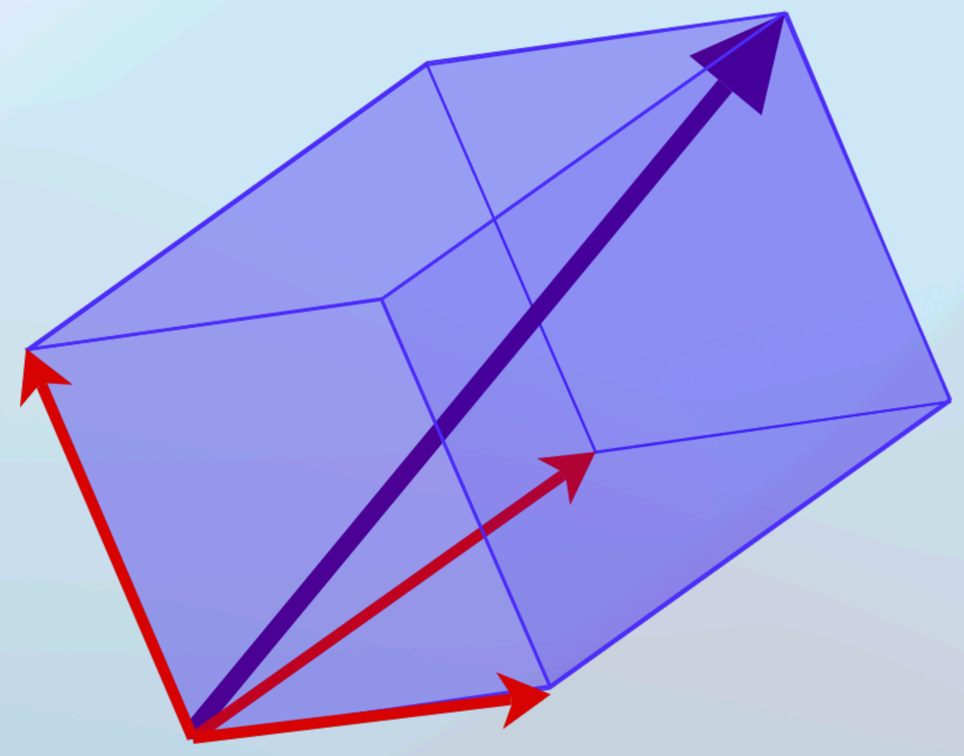
Matrix Rank

The matrix **rank** is the number of linearly independent columns of a matrix.

An orthogonal matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, has a rank equal to n, since a set of orthogonal vectors (the columns) are linearly independent.

For a non-square matrix, $\mathbf{B} \in \mathbb{R}^{m \times n}$, the matrix rank is at most as high as the minimum between its rows and columns, $rank(\mathbf{B}) \leq \min(m, n)$





Matrix-Vector Product

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a vector $\mathbf{v} \in \mathbb{R}^{n}$, we can define a product $\mathbf{u} = \mathbf{A}\mathbf{v} \in \mathbb{R}^{m}$ as:

$$u_i = \sum_{j=1}^{n} A_{i,j} v_j$$

<u>Interesting Matrices:</u>

Identity:

$$I = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_n] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Scaling:

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & s_n \end{bmatrix}$$

Example

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 2 \cdot 6 \\ 3 \cdot 5 + 4 \cdot 6 \end{bmatrix} = \begin{bmatrix} 17 \\ 49 \end{bmatrix}$$

Identity:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 0 \cdot 6 \\ 0 \cdot 5 + 1 \cdot 6 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

Scaling:

$$\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \cdot 5 + 0 \cdot 6 \\ 0 \cdot 5 + 0.5 \cdot 6 \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \end{bmatrix}$$

Matrix-Vector Product

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a vector $\mathbf{v} \in \mathbb{R}^{n}$, we can define a product $\mathbf{u} = \mathbf{A}\mathbf{v} \in \mathbb{R}^{m}$ as:

$$u_i = \sum_{j=1}^{n} A_{i,j} v_j$$

<u>Interesting Matrices:</u>

Rotation:

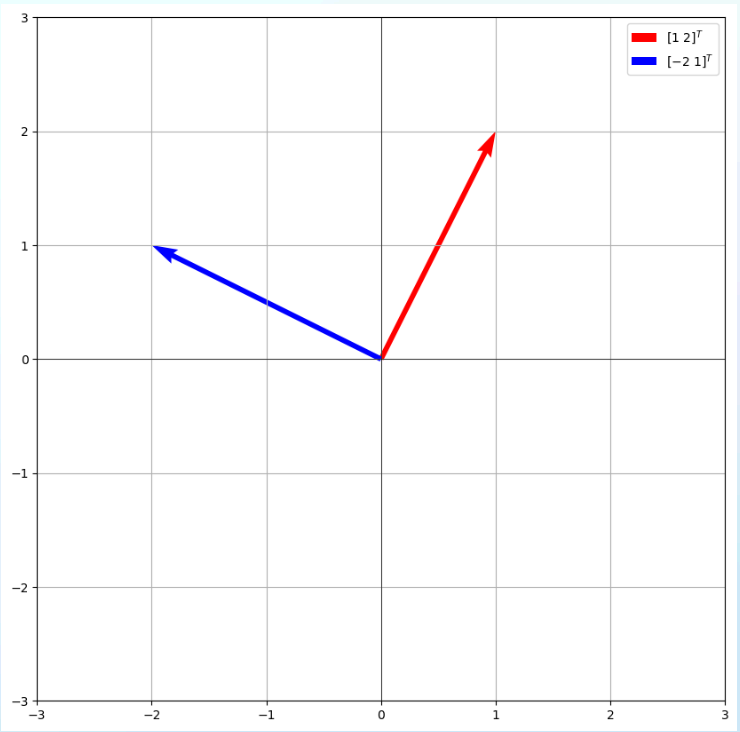
$$R = \begin{bmatrix} \cos\theta & -sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \cos\theta & \cdots & -sin\theta \\ \vdots & & \ddots & \vdots \\ 0 & \sin\theta & \cdots & \cos\theta \end{bmatrix}$$

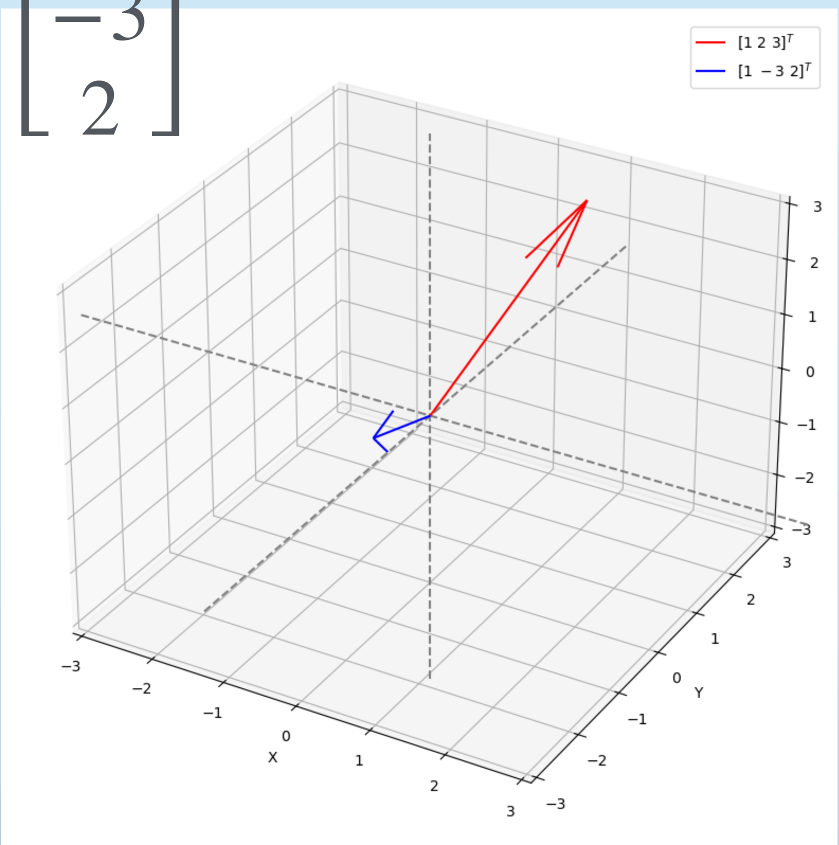
Example

Rotation by $\dfrac{\pi}{2}$

$$\begin{bmatrix} \cos\frac{\pi}{2} & -sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$



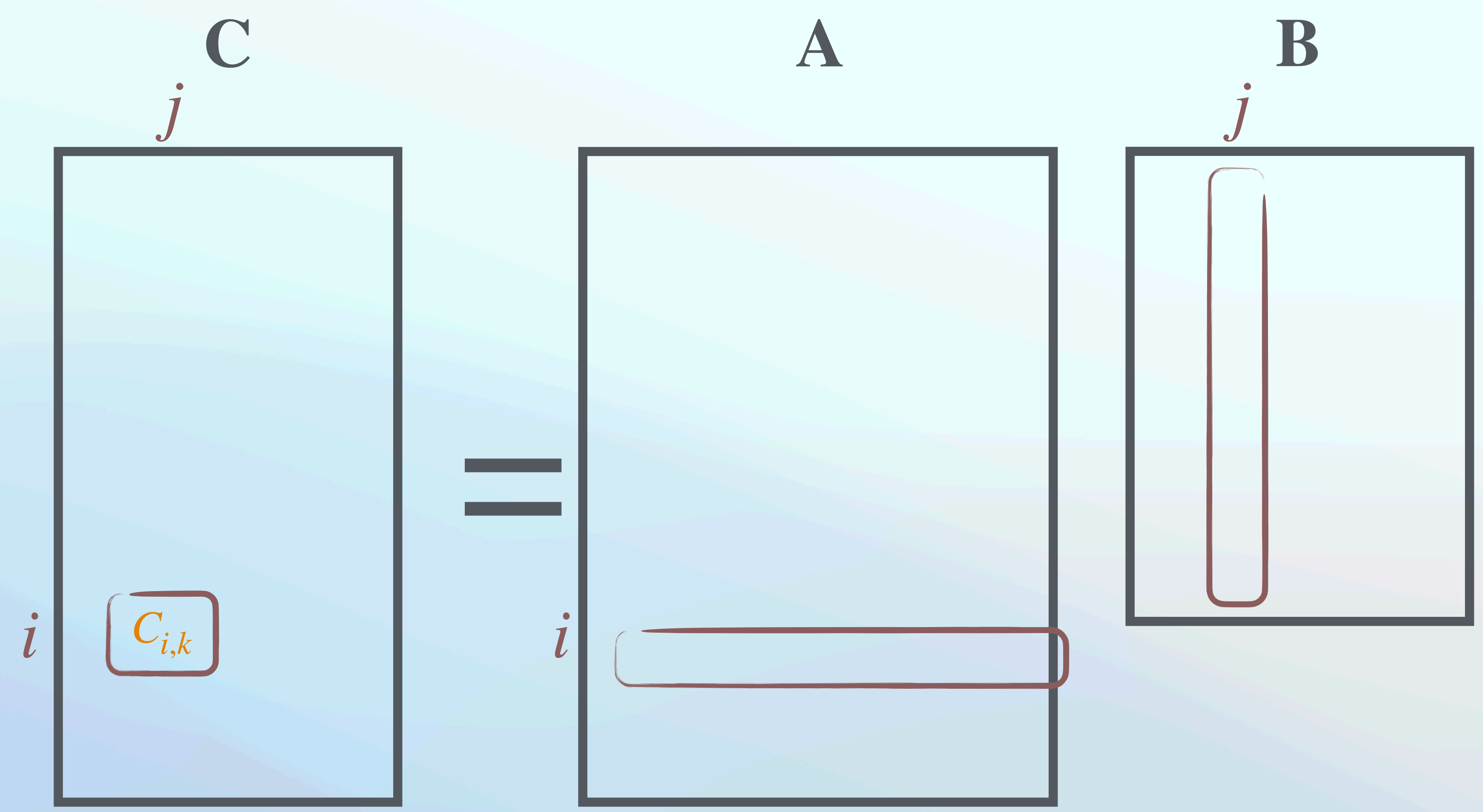
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\frac{\pi}{2} & -sin\frac{\pi}{2} \\ 0 & \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}$$



Matrix - Matrix Multiplication

A matrix $\mathbf{C} \in \mathbb{R}^{m \times k}$ is the product, $\mathbf{C} = \mathbf{AB}$, of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a matrix $\mathbf{B} \in \mathbb{R}^{n \times k}$ with:

$$C_{i,k} = \sum_{j=1}^{n} A_{i,j} B_{j,k}$$



Example

Let A be a matrix of student marks in different units. Some unit coordinators like to issue marks on a scale of 10, some in a scale of 20, and some in a scale of 100. What would you do to set them all to a scale of 100?

$$\overset{\mathbf{A}}{\begin{bmatrix} 6 & 20 & 86 \\ 7 & 13 & 96 \\ 4 & 11 & 76 \\ 10 & 3 & 65 \\ 9 & 17 & 54 \end{bmatrix}} \overset{\mathbf{B}}{\begin{bmatrix} 10 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}} = \overset{\mathbf{C}}{\begin{bmatrix} 60 & 80 & 86 \\ 70 & 65 & 96 \\ 40 & 55 & 76 \\ 100 & 15 & 65 \\ 90 & 85 & 54 \end{bmatrix}}$$

Change of Basis

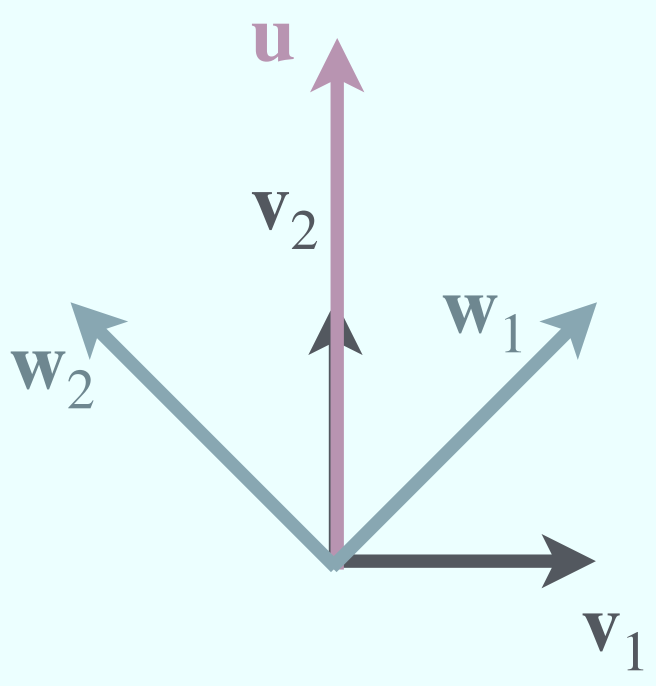
Let $\mathbf{B}_{\text{old}} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathbf{B}_{\text{new}} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be two bases of $\mathbb{R}^n$ (for simplicity assume that each vector in $\mathbf{B}_{\text{old}}$ and $\mathbf{B}_{\text{new}}$ have length 1). We can represent the new basis using the old one:

$$\mathbf{w}_j = \sum_{j,i} A_{j,i}\mathbf{v}_i$$

$$\mathbf{B}_{\text{old}} = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n] = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,n} \\ \vdots & & \ddots & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,n} \end{bmatrix}$$

$$\mathbf{B}_{\text{new}} = [\mathbf{w}_1 \quad \cdots \quad \mathbf{w}_n] = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,n} \\ \vdots & & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & w_{n,n} \end{bmatrix}$$

$$\mathbf{B}_{\text{new}} = \mathbf{A}\mathbf{B}_{\text{old}}$$



Example

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{w}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\mathbf{B}_{\text{old}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{B}_{\text{new}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \qquad \mathbf{B}_{\text{new}} = \mathbf{A}\mathbf{B}_{\text{old}}$$

Expressed in $\mathbf{B}_{\text{new}}$

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Expressed in $\mathbf{B}_{\text{old}}$

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

Invertible Matrix

A square matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, is called invertible if there is a matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ such that $\mathbf{AB} = I = \mathbf{BA}$, where I is the identity matrix.

The inverse of matrix $\mathbf{A}$ is often denoted as $\mathbf{A}^{-1}$.

Properties:

- $(\mathbf{A}^{-1})^{-1} = A$
- $(k\mathbf{A}^{-1}) = k^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}$
- For two invertible matrices $\mathbf{A}$ and $\mathbf{B}$ in $\mathbb{R}^{n \times n}$:
  - $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Example

$$\mathbf{A}^{-1} = \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} -2 \cdot 1 + 1 \cdot 3 & -2 \cdot 2 + 1 \cdot 4 \\ 1.5 \cdot 1 - 0.5 \cdot 3 & 1.5 \cdot 2 - 0.5 \cdot 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

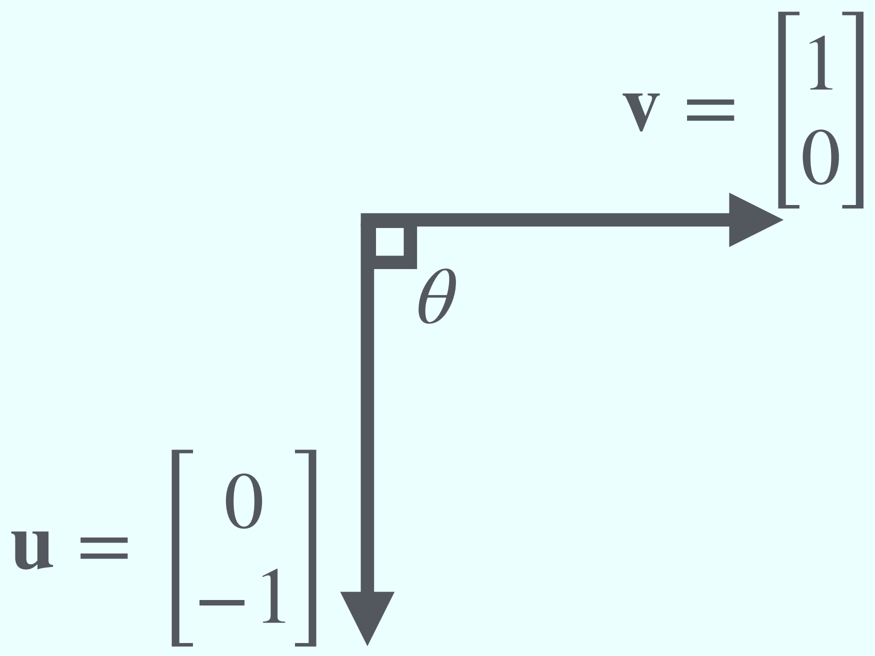
Orthogonal Vectors/Orthogonal Matrix

A real square matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, is called orthogonal (or orthonormal) if all its columns are orthogonal unit-length vectors.

In that case the inverse, I.e. $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = I$.

We also write $\mathbf{A}^T = \mathbf{A}^{-1}$

Example



$$\mathbf{u} \cdot \mathbf{v} = 1 \cdot 0 + 0 \cdot (-1) = 0$$

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta = 1 \cdot 1 \cdot 0$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \mathbf{A}^T = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 \cdot 0 + 1 \cdot 1 & 0 \cdot (-1) + 1 \cdot 0 \\ -1 \cdot 0 + 0 \cdot 1 & (-1) \cdot (-1) + 0 \cdot 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Symmetric Matrices

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric if and only if:

$$\mathbf{A} = \mathbf{A}^T \quad \text{or} \quad A_{i,j} = A_{j,i} \quad \text{for all } i, j$$

Properties:

If $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ are two symmetric matrices:

$\mathbf{A} + \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$ are also symmetric

Addition/Subtraction of Matrices is an elementwise operation, i.e. $\mathbf{A} - \mathbf{B} = \begin{bmatrix} A_{1,1} - B_{1,1} & \dots & A1,n - B_{1,n} \\ \vdots & \ddots & \vdots \\ A_{n,1} - B_{n,1} & \dots & An,n - B_{n,n} \end{bmatrix}$

$\mathbf{A}^n$ is symmetric for any n

$\mathbf{A}^{-1}$ is symmetric if and only if $\mathbf{A}$ is also symmetric

Example

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 4 \\ 2 & 2 & 1 & 3 \\ 1 & 1 & 3 & 1 \\ 4 & 3 & 1 & 2 \end{bmatrix} = \mathbf{A}^T \qquad \mathbf{B} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 3 \\ 1 & 1 & 2 & 1 \\ 1 & 3 & 1 & 2 \end{bmatrix} = \mathbf{B}^T$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 2 & 4 & 2 & 5 \\ 4 & 3 & 2 & 6 \\ 2 & 2 & 5 & 2 \\ 5 & 6 & 2 & 4 \end{bmatrix} \qquad \mathbf{A} - \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \begin{bmatrix} 22 & 19 & 10 & 19 \\ 19 & 18 & 10 & 21 \\ 10 & 10 & 12 & 12 \\ 19 & 21 & 12 & 30 \end{bmatrix}$$

Eigenvalues and Eigenvectors

Matrix decomposition

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ can be decomposed into a product of matrices like $A = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$, where $\mathbf{V} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, and $\boldsymbol{\Lambda} \in \mathbb{R}^{n \times n}$ is a diagonal matrix

Eigenvalues/Eigenvectors:

A vector $\mathbf{v} \in \mathbb{R}^n$ is called an eigenvector if:

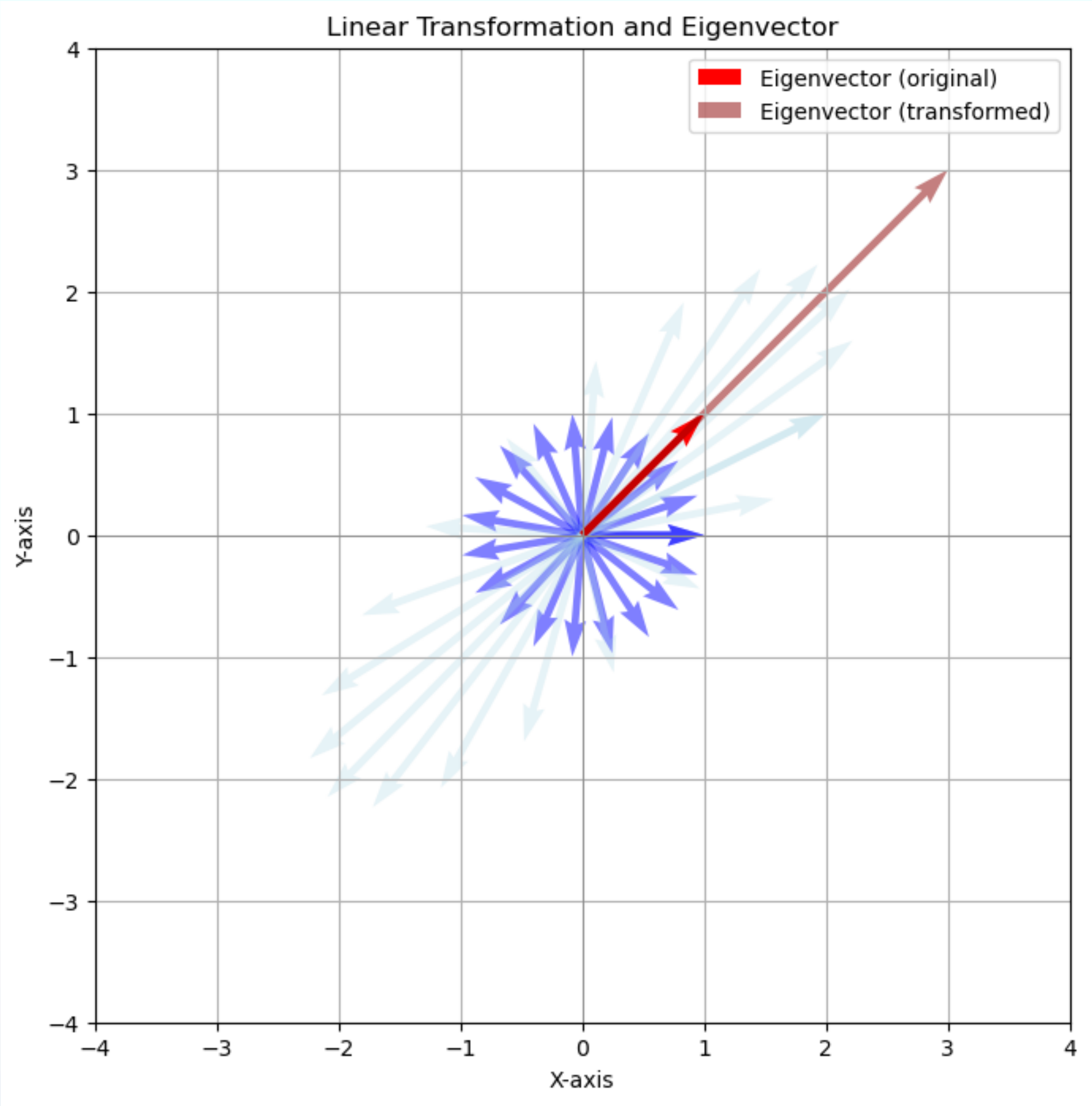
$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$

where λ is the corresponding eigenvalue of $\mathbf{v}$.

There are n different eigenvectors, we have:

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n] \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n]^T$$

Example



Covariance Matrix

For a random vector $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n]^\top$, the covariance matrix $\mathbf{\Sigma}$ is defined as

$$\Sigma = \text{Cov}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top\right]$$

For a data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ with n observations and p variables, the sample covariance matrix $\mathbf{S}$ can be calculated as:

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top$$

where:

- $\mathbf{X}_i$ is the i-th observation (a vector),
- $\bar{\mathbf{X}}$ is the mean vector of all observations, calculated as $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i$.

$$\Sigma = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & \text{Cov}(X_n, X_n) \end{bmatrix}$$

where $\text{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$.

Covariance Matrix Example

- Let $\mathbf{X} \in \mathbb{R}^{4\times3}$ be a data matrix of 4 samples from 3 observations. The covariance matrix is calculated as:

$$\mathbf{X} = \begin{bmatrix} 1 & -2 & 3 \\ 2 & -3 & 2 \\ -3 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$$

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix} \frac{1+2-3+1}{4} & \frac{-2-3+1+1}{4} & \frac{3+2+1-1}{4} \end{bmatrix} = [0.25 \quad -0.75 \quad 1.25]$$

$$\hat{\mathbf{X}} = \begin{bmatrix} (\mathbf{x}_1 - \mathbb{E}[\mathbf{x}_1]) & (\mathbf{x}_2 - \mathbb{E}[\mathbf{x}_2]) & (\mathbf{x}_3 - \mathbb{E}[\mathbf{x}_3]) \end{bmatrix} = \begin{bmatrix} 0.75 & -2.75 & 4.25 \\ 1.75 & -3.75 & 3.25 \\ -3.25 & 0.25 & 2.25 \\ 0.75 & 0.25 & 0.25 \end{bmatrix}$$

$$Cov\,[X] = \frac{1}{4-1}\hat{\mathbf{X}}^T\hat{\mathbf{X}} = \begin{bmatrix} 4.9 & -3.1 & 0.6 \\ -3.1 & 4.25 & -2.75 \\ 0.6 & -2.75 & 2.9 \end{bmatrix}$$

Summary

- Vectors/Matrices/Tensors
- Vector Operations (geometric interpretation/feature interpretation)
- Matrix Vector Operations
- Matrix Multiplication
- Matrix Inverse
- Special Matrices (Identity, Scaling, Orthogonal, Symmetric)
- Covariance Matrix

Anonymous Feedback on Georgios' Lectures for CM22009