Linear Algebra - Best of

Sibylle Hess

Technische Universiteit
Eindhoven
University of Technology

Vector Spaces

Vector Spaces

A vector space over the real numbers is a set of vectors $\mathcal{V}$ with two operations $+$ and $\cdot$ such that the following properties hold:

- Addition: for v, w we have $\mathrm{v} + \mathrm{w} \in \mathcal{V}$. The set of vectors with the addition $(\mathcal{V}, +)$ is an abelian group.
- Scalar multiplication: for $\alpha \in \mathbb{R}$ and $\mathrm{v} \in \mathcal{V}$, we have $\alpha \mathrm{v} \in \mathcal{V}$ such that the following properties hold:
  - $\alpha(\beta \mathrm{v}) = (\alpha\beta)\mathrm{v}$ for $\alpha, \beta \in \mathbb{R}$ and $\mathrm{v} \in \mathcal{V}$
  - $1\mathrm{v} = \mathrm{v}$ for $\mathrm{v} \in \mathcal{V}$
- Distributivity: the following properties hold:
  - $(\alpha + \beta)\mathrm{v} = \alpha \mathrm{v} + \beta \mathrm{v}$ for $\alpha, \beta \in \mathbb{R}$ and $\mathrm{v} \in \mathcal{V}$
  - $\alpha(\mathrm{v} + \mathrm{w}) = \alpha \mathrm{v} + \alpha \mathrm{w}$ for $\alpha \in \mathbb{R}$ and $\mathrm{v}, \mathrm{w} \in \mathcal{V}$

What is Allowed in a Vector Space?

A vector space is a structure where you can do most operations you know from real numbers, but not all. Let $\alpha \in \mathbb{R}, \mathsf{v}, \mathsf{w} \in \mathcal{V}$.

The following operations are well-defined:

- $\mathsf{v}/\alpha = \frac{1}{\alpha}\mathsf{v}$ for $\alpha \neq 0$
- $\mathsf{v} - \mathsf{w}$

What you can not do:

- $\mathsf{v} \cdot \mathsf{w}$
- α/v

The Vector Space $\mathbb{R}^d$

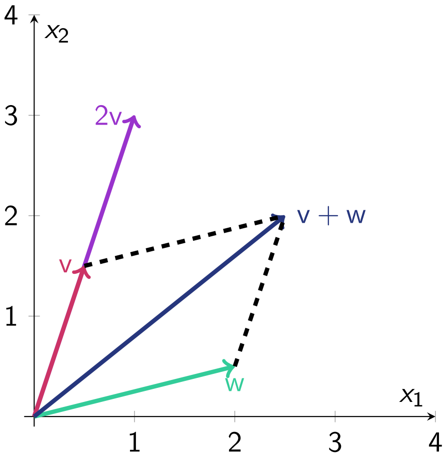
The elements of the vector space $\mathbb{R}^d$ are d-dimensional vectors

$$\mathrm{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix}, \quad v_i \in \mathbb{R} \text{ for } 1 \leq i \leq d.$$

For vectors, the addition between vectors and the scalar multiplication are defined for $\mathrm{v}, \mathrm{w} \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}$ as

$$\mathrm{v} + \mathrm{w} = \begin{pmatrix} v_1 + w_1 \\ \vdots \\ v_d + w_d \end{pmatrix}, \alpha \mathrm{v} = \begin{pmatrix} \alpha v_1 \\ \vdots \\ \alpha v_d \end{pmatrix}$$

Example: the Vector Space $\mathbb{R}^2$



$$v = \begin{pmatrix} 0.5 \\ 1.5 \end{pmatrix}$$

$$w = \begin{pmatrix} 2 \\ 0.5 \end{pmatrix}$$

$$v + w = \begin{pmatrix} 2.5 \\ 2 \end{pmatrix}$$

$$2v = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

Are there other important vector spaces next to $\mathbb{R}^d$?

Yes, the vector space of matrices $\mathbb{R}^{n \times d}$.

Why are matrices important?

Because data is represented as a matrix.

Data Representation by a Matrix

ID	F_1	F_2	F_3	...	F_d
1	5.1	3.5	1.4	...	0.2
2	6.4	3.5	4.5	...	1.2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	5.9	3.0	5.0	...	1.8

A data table of n observations of d features is represented by a $(n \times d)$ matrix.

Matrices and Their Notation

An $(n \times d)$ matrix concatenates n d-dimensional vectors column-wise ($A_{\cdot j}$ denotes the column-vector j of A)

$$A = \begin{pmatrix} | & & | \\ A_{\cdot 1} & \dots & A_{\cdot d} \\ | & & | \end{pmatrix} = \begin{pmatrix} A_{11} & \dots & A_{1d} \\ \vdots & & \vdots \\ A_{n1} & \dots & A_{nd} \end{pmatrix}$$

Simultaneously, we can see a matrix as concatenation of d row-vectors ($A_{i\cdot}$):

$$A = \begin{pmatrix} - & A_{1\cdot} & - \\ & \vdots & \\ - & A_{n\cdot} & - \end{pmatrix} = \begin{pmatrix} A_{11} & \dots & A_{1d} \\ \vdots & & \vdots \\ A_{n1} & \dots & A_{nd} \end{pmatrix}$$

The Vector Space $\mathbb{R}^{n \times d}$

The elements of the vector space $\mathbb{R}^{n \times d}$ are $(n \times d)$-dimensional matrices.

The addition between matrices and the scalar multiplication are defined for $A, B \in \mathbb{R}^{n \times d}$ and $\alpha \in \mathbb{R}$ as

$$A + B = \begin{pmatrix} A_{11} + B_{11} & \dots & A_{1d} + B_{1d} \\ \vdots & & \vdots \\ A_{n1} + B_{n1} & \dots & A_{nd} + B_{nd} \end{pmatrix}$$

$$\alpha A = \begin{pmatrix} \alpha A_{11} & \dots & \alpha A_{1d} \\ \vdots & & \vdots \\ \alpha A_{n1} & \dots & \alpha A_{nd} \end{pmatrix}$$

Matrix Operations:

The Transpose

The Transpose of a Matrix Swaps the Dimensionality

The transpose of a matrix changes row-vectors into column vectors and vice versa:

$$A = \begin{pmatrix} | & & | \\ A_{\cdot 1} & \dots & A_{\cdot d} \\ | & & | \end{pmatrix} = \begin{pmatrix} A_{11} & \dots & A_{1d} \\ \vdots & & \vdots \\ A_{n1} & \dots & A_{nd} \end{pmatrix} \in \mathbb{R}^{n\times d}$$

$$A^\top = \begin{pmatrix} - & A_{\cdot 1}^\top & - \\ & \vdots & \\ - & A_{\cdot d}^\top & - \end{pmatrix} = \begin{pmatrix} A_{11} & \dots & A_{n1} \\ \vdots & & \vdots \\ A_{1d} & \dots & A_{nd} \end{pmatrix} \in \mathbb{R}^{d\times n}$$

The Transpose of a Column Vector Makes it a Row Vector

The transpose of a d-dimensional vector has an interpretation as transpose of a $(d \times 1)$ matrix:

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix} \in \mathbb{R}^{d\times 1}$$

$$v^\top = \begin{pmatrix} v_1 & \dots & v_d \end{pmatrix} \in \mathbb{R}^{1\times d}$$

The Transpose of the Transpose Returns the Original Matrix

For any matrix $A \in \mathbb{R}^{n \times d}$ we have ${A^\top}^\top = A$

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad A^\top = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \qquad {A^\top}^\top = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

Symmetric Matrices are Invariant to Transposition

A symmetric matrix is a matrix $A \in \mathbb{R}^{n \times n}$ such that $A^\top = A$:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 7 \end{pmatrix} \qquad A^\top = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 7 \end{pmatrix}$$

Diagonal Matrices are Symmetric

A diagonal matrix is a symmetric matrix having only nonzero elements on the diagonal:

$$\operatorname{diag}(a_1, \ldots, a_n) = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ & & \ddots & \\ 0 & 0 & \ldots & a_n \end{pmatrix}$$

Okay, great, we can add, scale and transpose matrices/data. Isn't that kinda lame?

Yah, it gets interesting with the matrix product.

Inner and Outer Product of Vectors

The inner product of two vectors $v, w \in \mathbb{R}^d$ returns a scalar:

$$v^\top w = \begin{pmatrix} v_1 & \dots & v_d \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix} = \sum_{i=1}^{d} v_i w_i$$

The outer product of two vectors $v \in \mathbb{R}^d$ and $w \in \mathbb{R}^n$ returns a $(d \times n)$ matrix:

$$vw^\top = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix} \begin{pmatrix} w_1 & \dots & w_n \end{pmatrix} = \begin{pmatrix} v_1 w^\top \\ \vdots \\ v_d w^\top \end{pmatrix} = \begin{pmatrix} v_1 w_1 & \dots & v_1 w_n \\ \vdots & & \vdots \\ v_d w_1 & \dots & v_d w_n \end{pmatrix}$$

Matrix Multiplication

Given $A \in \mathbb{R}^{n \times r}$ and $B \in \mathbb{R}^{r \times d}$, the matrix product $C = AB \in \mathbb{R}^{n \times d}$ is defined as

$$C = \begin{pmatrix} A_{1\cdot}B_{\cdot 1} & \dots & A_{1\cdot}B_{\cdot d} \\ \vdots & & \vdots \\ A_{n\cdot}B_{\cdot 1} & \dots & A_{n\cdot}B_{\cdot d} \end{pmatrix} = \begin{pmatrix} - & A_{1\cdot} & - \\ & \vdots & \\ - & A_{n\cdot} & - \end{pmatrix} \begin{pmatrix} | & & | \\ B_{\cdot 1} & \dots & B_{\cdot d} \\ | & & | \end{pmatrix}$$

Every element C_{ji} is computed by the inner product of row j and column i (row-times-column)

$$C_{ji} = A_{j\cdot}B_{\cdot i} = \sum_{s=1}^{r} A_{js}B_{si}$$

Another View on Matrix Multiplication

Given $A \in \mathbb{R}^{n \times r}$ and $B \in \mathbb{R}^{r \times d}$, we can also state the product $C = AB$ in terms of the outer product:

$$C = \sum_{s=1}^{r} \begin{pmatrix} A_{1s}B_{s1} & \dots & A_{1s}B_{sd} \\ \vdots & & \vdots \\ A_{ns}B_{s1} & \dots & A_{ns}B_{sd} \end{pmatrix} = \begin{pmatrix} | & & | \\ A_{\cdot 1} & \dots & A_{\cdot r} \\ | & & | \end{pmatrix} \begin{pmatrix} - & B_{1\cdot} & - \\ & \vdots & \\ - & B_{r\cdot} & - \end{pmatrix}$$

The matrix product is the sum of outer products of corresponding column- and row-vectors (column-times-row):

$$C = \sum_{s=1}^{r} \begin{pmatrix} | \\ A_{\cdot s} \\ | \end{pmatrix} \begin{pmatrix} - & B_{s\cdot} & - \end{pmatrix} = \sum_{s=1}^{r} A_{\cdot s} B_{s\cdot}$$

Multiplying the Identity Matrix Doesn't Change Anything

The identity matrix I is a diagonal matrix having only ones on the diagonal:

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

Given $A \in \mathbb{R}^{n \times d}$, and I_n the $(n \times n)$ identity matrix and I_d the $(d \times d)$ identity matrix, then we have

$$I_n A = A = A I_d$$

The Transpose of a Matrix Product

We have for $A \in \mathbb{R}^{n\times r}$, $B \in \mathbb{R}^{r\times d}$ and $C = AB$

$$C^\top = \begin{pmatrix} A_{1\cdot}B_{\cdot 1} & \dots & A_{1\cdot}B_{\cdot d} \\ \vdots & & \vdots \\ A_{n\cdot}B_{\cdot 1} & \dots & A_{n\cdot}B_{\cdot d} \end{pmatrix}^\top = \begin{pmatrix} A_{1\cdot}B_{\cdot 1} & \dots & A_{n\cdot}B_{\cdot 1} \\ \vdots & & \vdots \\ A_{1\cdot}B_{\cdot d} & \dots & A_{n\cdot}B_{\cdot d} \end{pmatrix}$$

$$= \begin{pmatrix} B_{\cdot 1}^\top A_{1\cdot}^\top & \dots & B_{\cdot 1}^\top A_{n\cdot}^\top \\ \vdots & & \vdots \\ B_{\cdot d}^\top A_{1\cdot}^\top & \dots & B_{\cdot d}^\top A_{n\cdot}^\top \end{pmatrix} = B^\top A^\top$$

If we can multiply matrices, can we then also divide by them?

Just sometimes, if the matrix has an inverse.

Inverse Matrices

The inverse matrix to a matrix $A \in \mathbb{R}^{n\times n}$ is a matrix A^{-1} satisfying

$$AA^{-1} = A^{-1}A = I$$

Diagonal matrices with nonzero elements on the diagonal have an inverse:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} = I$$

Okay, but why is this now interesting?

Because matrix multiplication is computable fast, and almost every data operation can be written as a matrix operation.

Matrix Product Trivia

$A \in \mathbb{R}^{n\times r}$, $B \in \mathbb{R}^{m\times r}$, which product is well-defined?

a) BA b) $A^\top B$ c) $AB^\top$

$A \in \mathbb{R}^{n\times r}$, $B \in \mathbb{R}^{m\times r}$, what is $(AB^\top)^\top$?

a) $A^\top B$ b) $B^\top A^\top$ c) $BA^\top$

What is the matrix product computed by $C_{ji} = \sum_{s=1}^{r} A_{is} B_{js}$?

a) $C = AB^\top$ b) $C = B^\top A$ c) $C = BA^\top$

$A, B \in \mathbb{R}^{n\times n}$ have an inverse A^{-1}, B^{-1}, what is **not** equal to $AA^{-1}B$?

a) $A^{-1}BA$ b) B c) $BB^{-1}B$

Normed

Vector Spaces

Normed Vector Spaces

A normed vector space is a vector space $\mathcal{V}$ with a function $\|\cdot\| : \mathcal{V} \rightarrow \mathbb{R}_+$, called norm, satisfying the following properties for all $\mathsf{v}, \mathsf{w} \in \mathcal{V}$ and $\alpha \in \mathbb{R}$:

$$\|\mathsf{v} + \mathsf{w}\| \leq \|\mathsf{v}\| + \|\mathsf{w}\| \qquad \text{(triangle inequality)}$$

$$\|\alpha \mathsf{v}\| = |\alpha| \|\mathsf{v}\| \qquad \text{(homogeneity)}$$

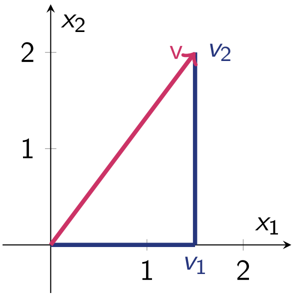
$$\|\mathsf{v}\| = 0 \Leftrightarrow \mathsf{v} = 0$$

The norm measures the length of a vector space

The Euclidean Space

The d-dimensional Euclidean space is the space of $\mathbb{R}^d$ with the Euclidean norm:

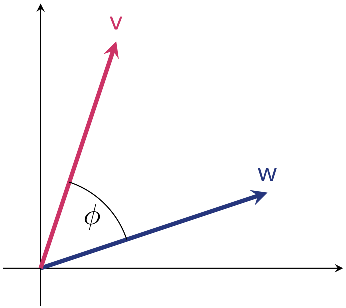
$$\|\mathrm{v}\|_2 = \|\mathrm{v}\| = \sqrt{\sum_{i=1}^{d} v_i^2}$$



The Euclidean norm computes the length of a vector by means of the Pythagorean theorem:

$$\|\mathrm{v}\|^2 = v_1^2 + v_2^2$$

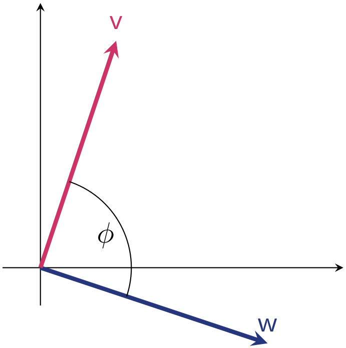
The Inner Product and the Euclidean Norm



The inner product is defined by the lengths of the vectors and the cosine of the angle between them.

$$\mathrm{v}^\top \mathrm{w} = \sum_{i=1}^{d} v_i w_i$$
$$= \cos \sphericalangle(\mathrm{v}, \mathrm{w}) \|\mathrm{v}\| \|\mathrm{w}\|$$

Orthogonal Vectors



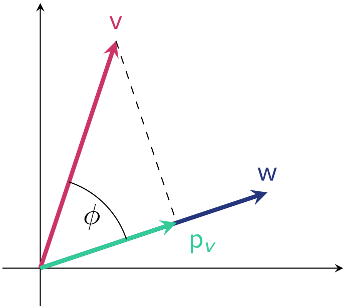
If two vectors are orthogonal, then $\cos \sphericalangle(\mathrm{v}, \mathrm{w}) = 0$ and the inner product is zero

$$\mathrm{v}^\top \mathrm{w} = \cos \sphericalangle(\mathrm{v}, \mathrm{w}) \|\mathrm{v}\| \|\mathrm{w}\| = 0$$

Two vectors are called orthonormal if they are orthogonal and have unit norm $\|\mathrm{v}\| = \|\mathrm{w}\| = 1$.

The Inner Product and Projections

The inner product of a vector v and a normalized vector $\frac{w}{\|w\|}$ computes the length of the projection p_v of v onto w:



$$\cos(\phi) = \frac{\|p_v\|}{\|v\|}$$

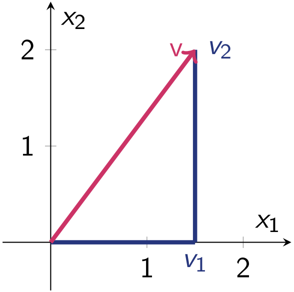
$$\Leftrightarrow \|p_v\| = \cos(\phi)\|v\| = v^\top \frac{w}{\|w\|}$$

$$\Rightarrow p_v = \frac{ww^\top}{\|w\|^2} v$$

The Manhattan Norm

The Manhattan norm is defined as:

$$\|\mathrm{v}\|_1 = |\mathrm{v}| = \sum_{i=1}^{d} |v_i|$$



The Manhattan norm computes the length of a vector coordinate-wise:

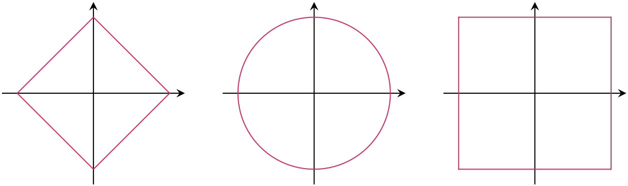
$$|\mathrm{v}| = |v_1| + |v_2|$$

L_p-norms

For $p \in [1, \infty]$, the function $\|\cdot\|_p$ is a norm, where

$$\|v\|_p = \left(\sum_{i=1}^{d} |v_i|^p \right)^{1/p}$$

The two-dimensional circles $\{v \in \mathbb{R}^2 | \|v\|_p = 1\}$ look as follows:



$p = 1$ $p = 2$ $p = \infty$

So, the norm measures the length of a vector. Can we also measure the length of a matrix?

Yes, matrix norms are the same but different.

Matrix Norms

We can extend the L_p vector normes to the element-wise L_p matrix norms:

$$\|A\|_p = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} |A_{ji}|^p \right)^{1/p}$$

Furthermore, we introduce the operator norm

$$\|A\|_{op} = \max_{\|\mathrm{v}\|=1} \|Av\|$$

Orthogonal Matrices

A matrix A with orthogonal columns satisfies

$$A^\top A = \operatorname{diag}(\|A_{\cdot 1}\|^2, \ldots, \|A_{\cdot d}\|^2)$$

A matrix A with orthonormal columns satisfies

$$A^\top A = \operatorname{diag}(1, \ldots, 1)$$

A square matrix $A \in \mathbb{R}^{n \times n}$ is called orthogonal if

$$A^\top A = AA^\top = I$$

Norms and Orthogonal Invariance

A vector norm $\|\cdot\|$ is called orthogonal invariant if for all $\mathrm{v} \in \mathbb{R}^n$ and orthogonal matrices $X \in \mathbb{R}^{n \times n}$ we have

$$\|X\mathrm{v}\| = \|\mathrm{v}\|$$

A matrix norm $\|\cdot\|$ is called orthogonal invariant if for all $V \in \mathbb{R}^{n \times d}$ and orthogonal matrices $X \in \mathbb{R}^{n \times n}$ we have

$$\|XV\| = \|V\|$$

Matrix Operations:

The Trace

The Trace of a Matrix

The trace sums the elements on the diagonal of a matrix. Let $A \in \mathbb{R}^{n\times n}$, then

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}$$

1. $\mathrm{tr}(cA + B) = c\,\mathrm{tr}(A) + \mathrm{tr}(B)$ (linearity)
2. $\mathrm{tr}(A^\top) = \mathrm{tr}(A)$
3. $\mathrm{tr}(ABCD) = \mathrm{tr}(BCDA) = \mathrm{tr}(CDAB) = tr(DABC)$ (cycling property)

The L_2-Norms are Induced by the Trace of the Product

For any vector $\mathrm{v} \in \mathbb{R}^d$ and matrix $A \in \mathbb{R}^{n \times d}$, we have

$$\|\mathrm{v}\|^2 = \mathrm{v}^\top \mathrm{v} = \mathrm{tr}(\mathrm{v}^\top \mathrm{v}) \qquad \|A\|^2 = \mathrm{tr}(A^\top A)$$

From this property derive the binomial formulas of vectors and matrices:

$$\|\mathrm{x} - \mathrm{y}\|^2 = (\mathrm{x} - \mathrm{y})^\top (\mathrm{x} - \mathrm{y}) = \|\mathrm{x}\|^2 - 2\langle \mathrm{x}, \mathrm{y} \rangle + \|\mathrm{y}\|^2$$
$$\|X - Y\|^2 = \mathrm{tr}((X - Y)^\top (X - Y)) = \|X\|^2 - 2\langle X, Y \rangle + \|Y\|^2$$

And now one super important cool thing:

The Singular Value Decomposition

Singular Value Decomposition

Theorem (SVD)

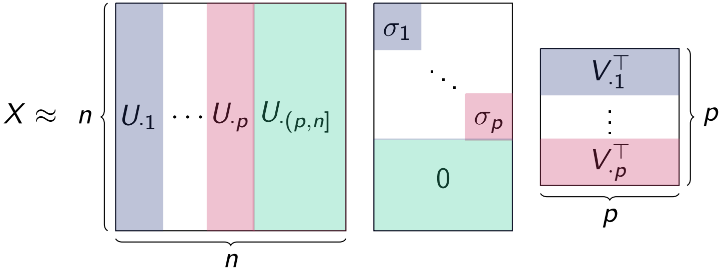
For every matrix $X \in \mathbb{R}^{n \times p}$ there exist orthogonal matrices $U \in \mathbb{R}^{n \times n}$, $V \in \mathbb{R}^{p \times p}$ and $\Sigma \in \mathbb{R}^{n \times p}$ such that

$$X = U \Sigma V^\top, \text{ where}$$

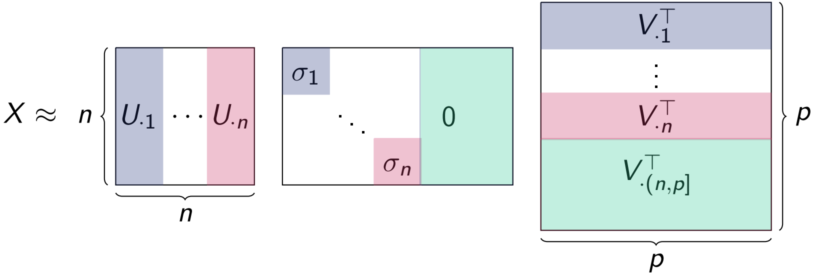
- $U^\top U = UU^\top = I_n, V^\top V = VV^\top = I_p$
- Σ *is a rectangular diagonal matrix,* $\Sigma_{11} \geq \ldots \geq \Sigma_{kk}$ *where* $k = \min\{n, p\}$

The column vectors $U_{\cdot s}$ and $V_{\cdot s}$ are called left and right singular vectors and the values $\sigma_i = \Sigma_{ii}$ are called singular values $(1 \leq i \leq l)$.

SVD Visualization for $n > p$



SVD Visualization for $p > n$



SVD Determines if a Matrix is Invertible

A $(n \times n)$ matrix $A = U\Sigma V^\top$ is invertible if all singular values are larger than zero. The inverse is given by

$$A^{-1} = V\Sigma^{-1}U^\top, \quad \text{where}$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \sigma_n \end{pmatrix}, \qquad \Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \frac{1}{\sigma_n} \end{pmatrix}$$

Since the matrices U and V of the SVD are orthogonal, we have:

$$AA^{-1} = U\Sigma V^\top V\Sigma^{-1}U^\top = U\Sigma\Sigma^{-1}U^\top = UU^\top = I$$
$$A^{-1}A = V\Sigma^{-1}U^\top U\Sigma V^\top = V\Sigma^{-1}\Sigma V^\top = VV^\top = I$$

Vector and Matrix Norm Trivia

$v, w \in \mathbb{R}^d$, $\alpha \in \mathbb{R}$, then $\|\alpha v + w\| \leq$

a) $\alpha\|v + w\|$ b) $|\alpha|\|v\| + \|w\|$ c) $\alpha\|v\| + \|w\|$

$A, B \in \mathbb{R}^{n \times r}$, $\alpha \in \mathbb{R}$, then $\|A\| \leq$

a) $\|A - B\| + \|B\|$ b) $\alpha\|\frac{1}{\alpha}A\|$ c) $\|A\|^2$

$A, B, C \in \mathbb{R}^{n \times n}$, what is equal to $\text{tr}(ABC)$?

a) $\text{tr}(ACB)$ b) $\text{tr}(A^\top C^\top B^\top)$ c) $\text{tr}(A)\,\text{tr}(BC)$

$A, B \in \mathbb{R}^{n \times n}$, A is orthogonal, what is **not** equal to $\text{tr}(ABA^\top)$?

a) $\text{tr}(A^\top BA)$ b) $\text{tr}(B)$ c) $\text{tr}(ABA)$