

# *Computational Intelligence Lab: Linear Algebra Recap*

*Cristian Perez Jensen*

*July 16, 2024*

Note that these are not the official lecture notes of the course, but only notes written by a student of the course. As such, there might be mistakes. The source code can be found at [github.com/cristianpjensen/eth-cs-notes](https://github.com/cristianpjensen/eth-cs-notes). If you find a mistake, please create an issue or open a pull request.

*Contents*

1	<i>Inner product and norms</i>	1
2	<i>Vector spaces</i>	2
3	<i>Matrices</i>	2
3.1	<i>Invertible matrices</i>	3
3.2	<i>Trace</i>	3
3.3	<i>Orthogonal projection</i>	4
3.4	<i>Special matrices</i>	4

## 1 Inner product and norms

**Definition 1** (Inner product). An inner product  $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$  is an operation defined on a vector space  $\mathcal{V}$  that satisfies the following properties  $\forall x, y, z, a, b \in \mathbb{R}$ ,

- Commutativity,

$$\langle x, y \rangle = \langle y, x \rangle;$$

- Linearity,

$$\langle x, ay + bz \rangle = a\langle x, y \rangle + b\langle x, z \rangle;$$

- Positive definiteness,

$$x \neq 0 \implies \langle x, x \rangle > 0$$

$$x = 0 \iff \langle x, x \rangle = 0;$$

- Bilinearity (follows from commutativity and linearity),

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$$

$$\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle.$$

**Corollary.**  $\langle x + y, x + y \rangle = \langle x, x \rangle + 2\langle x, y \rangle + \langle y, y \rangle.$

**Corollary.**  $\langle Ax, y \rangle = \langle x, A^\top y \rangle.$

The vector space  $\mathcal{V}$ , along with an inner product, defines an inner vector space. During this course, we will assume that we always work with real vectors in  $\mathbb{R}^n$ . An example of an inner product is the dot product,<sup>1</sup>

$$x \cdot y = x^\top y \in \mathbb{R}.$$

<sup>1</sup> Usually, this operation is what is meant by the inner product.

**Definition 2** (Norm). A norm  $\| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R}$  is a function that can be thought of as a way of measuring the distance from the origin. Norms satisfy the following properties,

- Positive definiteness,  $x \neq 0 \implies \|x\| > 0$ ;
- Triangle inequality,  $\|x + y\| \leq \|x\| + \|y\|$ ;
- Cauchy-Schwarz inequality,  $|\langle x, y \rangle| \leq \|x\| \|y\|.$

**Corollary.** For the Euclidean norm, the following holds,

$$\cos \theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|},$$

where  $\theta$  is the angle between  $\mathbf{x}$  and  $\mathbf{y}$ .

Each inner product defines a canonical norm  $\|\mathbf{x}\| \doteq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ . For example, the Euclidean norm is defined by the dot product,

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

The  $p$ -norm is a generalization of the Euclidean norm,

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}.$$

## 2 Vector spaces

The *vector space*  $\mathbb{R}^m$  consists of all column vectors with  $m$  elements. For a set of vectors  $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n \mid \mathbf{c}_i \in \mathbb{R}^m\}$ , we can define a subspace spanned by this set, denoted by  $\text{span}(\mathcal{C})$ . It is the set of all possible linear combinations of elements of  $\mathcal{C}$ . If a set of vectors that span a subspace are independent, they are called a *basis*,  $\mathbf{b}_1, \dots, \mathbf{b}_k \in \mathbb{R}^m$ . The number of basis vectors defines the *dimensionality* of the subspace.<sup>2</sup>

**Observation.** The following facts hold about subspaces,

- Every subspace contains the zero vector  $\mathbf{0}$ ;
- If  $\mathbf{x}$  and  $\mathbf{y}$  are in the subspace, then  $\mathbf{x} + \mathbf{y}$  is also in the subspace;
- If  $\mathbf{x}$  is in the subspace and  $a \in \mathbb{R}$ , then  $a\mathbf{x}$  is also in the subspace.

<sup>2</sup> We know that the amount of basis vectors must be smaller than the amount of vectors that span the subspace, which must be smaller than the dimensionality of the space,

$$k \leq m \leq n.$$

**Definition 3** (Orthogonal subspaces). Subspaces  $\mathcal{V}$  and  $\mathcal{W}$  are orthogonal when  $\mathbf{v}^\top \mathbf{w} = 0$  for all  $\mathbf{v} \in \mathcal{V}, \mathbf{w} \in \mathcal{W}$ .

## 3 Matrices

The rank  $r$  of a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is the dimensionality of its *column space*. It is bounded by

$$r \leq \min\{m, n\}.$$

The matrix is full-rank if  $r = \min\{m, n\}$ .

A matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  defines 4 fundamental subspaces,

- Column space  $\subseteq \mathbb{R}^m$  ( $r$  dimensional),  $\{\mathbf{b} \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}$ ;
- Null space  $\subseteq \mathbb{R}^n$  ( $n - r$  dimensional),  $\{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{0}\}$ ;
- Row space  $\subseteq \mathbb{R}^n$  ( $r$  dimensional),  $\{\mathbf{b} \mid \mathbf{A}^\top \mathbf{x} = \mathbf{b}\}$ ;

- Left null space  $\subseteq \mathbb{R}^m$  ( $m - r$  dimensional),  $\{x \mid A^\top x = 0\}$ .

The row space  $\text{row}(A)$  is the *orthogonal complement* of the null space  $\text{null}(A)$ , thus  $\text{row}(A) + \text{null}(A) = \mathbb{R}^n$ . Similarly,  $\text{col}(A) + \text{null}(A^\top) = \mathbb{R}^m$ .

### 3.1 Invertible matrices

A matrix is only invertible if it is a square full-rank matrix, i.e.,  $r = m = n$ .

**Properties.** Let  $A \in \mathbb{R}^{n \times n}$  be a full-rank matrix and  $k \in \mathbb{R}$ , then

$$\begin{aligned} A^{-1}A &= AA^{-1} = I \\ (kA)^{-1} &= \frac{1}{k}A^{-1} \\ \det(A^{-1}) &= \frac{1}{\det(A)} \\ (AB)^{-1} &= B^{-1}A^{-1}. \end{aligned}$$

The Moore-Penrose inverse (pseudo-inverse) is a generalization of the inverse. It is the solution to the general least squares problem  $\min_x \|Ax - b\|_2$ . For full-rank matrices, the left pseudo-inverse can be computed by

$$A^+ = (A^\top A)^{-1}A^\top.$$

The right pseudo-inverse can be computed as

$$A^+ = A^\top(AA^\top)^{-1}.$$

### 3.2 Trace

**Definition 4** (Trace). The trace of a square matrix  $A \in \mathbb{R}^{n \times n}$  is the sum of its diagonal,

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}.$$

**Properties.** Let  $A, B, C \in \mathbb{R}^{n \times n}$ ,  $x, y \in \mathbb{R}^n$ ,  $c, d \in \mathbb{R}$ , then

$$\begin{aligned} \text{tr}(cA + dB) &= c \cdot \text{tr}(A) + d \cdot \text{tr}(B) \\ \text{tr}(A) &= \text{tr}(A^\top) \\ \text{tr}(AB) &= \text{tr}(BA) \\ \text{tr}(ABC) &= \text{tr}(CAB) = \text{tr}(BCA) \\ \text{tr}(A^\top B) &= \text{tr}(B^\top A) = \text{tr}(AB^\top) = \text{tr}(BA^\top) \\ x^\top y &= \text{tr}(x^\top y) = \text{tr}(xy^\top). \end{aligned}$$

Furthermore, the trace of a matrix is equal to the sum of the eigenvalues of the matrix.



**Figure 1.** Illustration of the 4 spaces defined by a matrix  $A$ . It shows the perpendicular spaces. Furthermore, it shows that  $Ax_r = b$  for some  $x_r \in \text{col}(A)$ . Also, if you add a vector from the null space to the row vector, it still maps to the same  $b$ ,  $A(x_r + x_n) = Ax_r + Ax_n = Ax_r = b$ .

The left pseudo-inverse can be computed if  $r = n$  and has the property  $A^+A = I_{n \times n}$ .

The right pseudo-inverse can be computed if  $r = m$  and has the property  $AA^+ = I_{m \times m}$ .

Linearity.

Cyclic property.

$xy^\top$  is a rank-1 matrix.

### 3.3 Orthogonal projection

The projection  $a_1$  of a vector  $a$  on another vector  $b$  can be computed as

$$a_1 = \frac{b^\top a}{b^\top b} b.$$

The rejection can then be computed as

$$a_2 = a - a_1.$$

We can also project a vector  $a$  onto the column space of a matrix  $B \in \mathbb{R}^{m \times n}$ , denoted  $a_B$ . This can be computed by a matrix multiplication,

$$P = B(B^\top B)^{-1}B^\top,$$

which is equivalent to the left Moore-Penrose inverse.

**Definition 5** (Projection matrix). A square matrix  $P \in \mathbb{R}^{n \times n}$  is called a projection matrix if it is idempotent, i.e.,  $PP = P$ .

**Properties.** Let  $P$  be an orthogonal projection matrix, then  $P = P^\top$ , i.e.,  $P$  is symmetric. Furthermore, the eigenvalues of a projection matrix are all ones and zeros, because of idempotency.

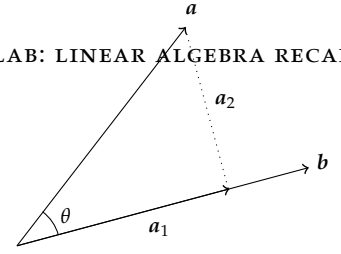
### 3.4 Special matrices

**Definition 6** (Orthogonal matrix). An orthogonal matrix is an invertible matrix whose columns  $q_1, \dots, q_n$  are all orthogonal to each other and of unit length, i.e.,

$$\begin{aligned} q_i^\top q_j &= 0 \quad \forall i \neq j \in [n] \\ q_i^\top q_i &= 1 \quad \forall i \in [n]. \end{aligned}$$

**Properties.** Let  $Q \in \mathbb{R}^{n \times n}$  be an orthogonal matrix,

$$\begin{aligned} Q^\top &= Q^{-1} \\ \langle x, y \rangle &= \langle Qx, Qy \rangle. \end{aligned}$$



**Figure 2.** Projection of  $a$  on  $b$ , denoted  $a_1$ , and the rejection of  $a$  from  $b$ , denoted  $a_2$ .