

# Subspaces

## Computational Intelligence, Lecture 2

by Sergei Savin

Spring 2024

# LEAST SQUARES AT A GLANCE (1)

Consider the following problem: find  $\mathbf{x}$  that minimizes  $\|\mathbf{Ax} - \mathbf{y}\|_2$ . This is the *least squares problem*.

- The value  $\mathbf{e} = \mathbf{Ax} - \mathbf{y}$  is called residual.
- $\|\cdot\|_2$  is a second norm, where  $\|\mathbf{e}\|_2 = \sqrt{\mathbf{e}^\top \mathbf{e}}$
- Least squares problem is about finding *least residual solution*.

## LEAST SQUARES AT A GLANCE (2)

Minimum of  $\|\mathbf{e}\|_2 = \|\mathbf{Ax} - \mathbf{y}\|_2$  is achieved at the same  $\mathbf{x}$  as minimum of  $(\mathbf{Ax} - \mathbf{y})^\top (\mathbf{Ax} - \mathbf{y})$ . With that in mind, let us find its extremum:

$$\frac{d}{d\mathbf{x}} \left( (\mathbf{Ax} - \mathbf{y})^\top (\mathbf{Ax} - \mathbf{y}) \right) = 0 \quad (1)$$

$$2\mathbf{A}^\top (\mathbf{Ax} - \mathbf{y}) = 0 \quad (2)$$

$$\mathbf{A}^\top \mathbf{Ax} = \mathbf{A}^\top \mathbf{y} \quad (3)$$

$$\mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} \quad (4)$$

Thus we can define a *pseudoinverse*:

$$\mathbf{A}^+ = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \quad (5)$$

Thus the least residual solution to  $\mathbf{Ax} = \mathbf{y}$  is written as:

$$\mathbf{x} = \mathbf{A}^+ \mathbf{y} \tag{6}$$

We already showed that it is the least-residual solution, later we will prove that it is also the *smallest norm* solution (out of all solutions with the same residual).

Let matrix  $\mathbf{M}$  be orthonormal (not necessarily square), meaning  $\mathbf{M}^\top \mathbf{M} = \mathbf{I}$ . Its pseudoinverse can be simplified:

$$\mathbf{M}^+ = (\mathbf{M}^\top \mathbf{M})^{-1} \mathbf{M}^\top = \mathbf{M}^\top \quad (7)$$

Then least squares solution to the equation  $\mathbf{M}\mathbf{x} = \mathbf{y}$  can be found as:

$$\mathbf{x}_{LS} = \mathbf{M}^\top \mathbf{y} \quad (8)$$

If  $\mathbf{M}$  is orthonormal and square, then  $\mathbf{M}^\top = \mathbf{M}^{-1}$ .

Given an equation  $\mathbf{Ax} = \mathbf{y}$  and least squares solution  $\mathbf{x}_{LS} = \mathbf{A}^+\mathbf{y}$ , let us compute the residual  $\mathbf{e} = \mathbf{y} - \mathbf{Ax}_{LS}$ . We substitute the solution:

$$\mathbf{e} = \mathbf{y} - \mathbf{AA}^+\mathbf{y} \tag{9}$$

We observe that:

- The residual can be found as  $\mathbf{e} = (\mathbf{I} - \mathbf{AA}^+)\mathbf{y}$ .
- The closest  $\mathbf{Ax}$  can get to  $\mathbf{y}$  is  $\mathbf{y}^* = \mathbf{AA}^+\mathbf{y}$ .
- Later we will find that  $\mathbf{AA}^+\mathbf{y}$  is a *projection* of  $\mathbf{y}$  onto a *column space* of  $\mathbf{A}$ .

# FOUR FUNDAMENTAL SUBSPACES

One of the key ideas in Linear Algebra is that every linear operator has four fundamental subspaces:

- Null space
- Row space
- Column space
- Left null space

Our goal is to understand them. The usefulness of this concept is enormous.

# NULL SPACE

## Definition

Consider the following task: find all solutions to the system of equations  $\mathbf{A}\mathbf{x} = \mathbf{0}$ .

It can be re-formulated as follows: find all elements of the *null space* of  $\mathbf{A}$ .

### Definition 1

*Null space* of  $\mathbf{A}$  is the set of all vectors  $\mathbf{x}$  that  $\mathbf{A}$  maps to  $\mathbf{0}$

We will denote null space as  $\text{null}(\mathbf{A})$ . Null space of an operator is sometimes called *kernel* and denoted as  $\text{ker}(\mathbf{A})$ .



# NULL SPACE

## Calculation

Now we can find all solutions to the system of equations  $\mathbf{Ax} = \mathbf{0}$  by using functions that generate an *orthonormal basis* in the null space of  $\mathbf{A}$ . In MATLAB it is function `null`, in Python/Scipy - `null_space`:

- `N = null(A).`

- `N = scipy.linalg.null_space(A).`

# NULL SPACE PROJECTION

## Local coordinates

Let  $\mathbf{N}$  be the orthonormal basis in the null space of matrix  $\mathbf{A}$ . Then, if a vector  $\mathbf{x}$  lies in the null space of  $\mathbf{A}$ , it can be represented as:

$$\mathbf{x} = \mathbf{N}\mathbf{z} \tag{10}$$

where  $\mathbf{z}$  are coordinates of  $\mathbf{x}$  in the basis  $\mathbf{N}$ .

However, there are vectors which not only are not lying in the null space of  $\mathbf{A}$ , but the closest vector to them in the null space is the zero vector.

# CLOSEST ELEMENT FROM A LINEAR SUBSPACE

$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ . Its null space has orthonormal basis  $\mathbf{N} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ .

■  $\begin{bmatrix} -2 \\ 0 \end{bmatrix} = -2\mathbf{N}$ ,  $\begin{bmatrix} 10 \\ 0 \end{bmatrix} = 10\mathbf{N}$ , - both are in the null space.

■ for  $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$  the closest vector in the null space is  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ .

■ for  $\mathbf{y} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$  the closest vector in the null space is  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ .

# ORTHOGONALITY, DEFINITION (1)

## Definition

Any two vectors,  $\mathbf{x}$  and  $\mathbf{y}$ , whose dot product is zero are said to be *orthogonal* to each other.

## Definition

Vector  $\mathbf{y}$ , whose dot product with any  $\mathbf{x} \in \mathcal{L}$  is orthogonal to the subspace  $\mathcal{L}$

## Definition (equivalent, see Appendix A)

If for a vector  $\mathbf{y}$ , the closest vector to it from a linear subspace  $\mathcal{L}$  is zero vector,  $\mathbf{y}$  is called orthogonal to the subspace  $\mathcal{L}$ .

## ORTHOGONALITY, DEFINITION (2)

### Definition

The space of all vectors  $\mathbf{y}$ , orthogonal to a linear subspace  $\mathcal{L}$  is called *orthogonal complement* of  $\mathcal{L}$  and is denoted as  $\mathcal{L}^\perp$ .

### Definition (equivalent)

The space of all vectors  $\mathbf{y}$ , such that  $\text{dot}(\mathbf{y}, \mathbf{x}) = 0, \forall \mathbf{x} \in \mathcal{L}$  is called *orthogonal complement* of  $\mathcal{L}$ .

Therefore  $\mathbf{x} \in \mathcal{L}$  and  $\mathbf{y} \in \mathcal{L}^\perp$  implies  $\text{dot}(\mathbf{y}, \mathbf{x}) = 0$ .

# PROJECTION

## Part 1

Let  $\mathbf{L}$  be an orthonormal basis in a linear subspace  $\mathcal{L}$ . Take vector  $\mathbf{a} = \mathbf{x} + \mathbf{y}$ , where  $\mathbf{x}$  lies in the subspace  $\mathcal{L}$ , and  $\mathbf{y}$  lies in the subspace  $\mathcal{L}^\perp$ .

### Definition

We call such vector  $\mathbf{x}$  an *orthogonal projection* of  $\mathbf{a}$  onto subspace  $\mathcal{L}$ , and such vector  $\mathbf{y}$  an orthogonal projection of  $\mathbf{a}$  onto subspace  $\mathcal{L}^\perp$ .

Orthogonal projection maps a vector to the element in the subspace closest to that vector. Orthogonal projection of  $\mathbf{a}$  onto  $\mathcal{L}$  can be found as:

$$\mathbf{x} = \mathbf{L}\mathbf{L}^+ \mathbf{a} \quad (11)$$

Since  $\mathbf{L}$  is orthonormal, this is the same as  $\mathbf{x} = \mathbf{L}\mathbf{L}^\top \mathbf{a}$

Since  $\mathbf{a} = \mathbf{x} + \mathbf{y}$ , and  $\mathbf{x} = \mathbf{L}\mathbf{L}^+\mathbf{a}$ , we can write:

$$\mathbf{a} = \mathbf{L}\mathbf{L}^+\mathbf{a} + \mathbf{y} \quad (12)$$

from which it follows that the projection of  $\mathbf{a}$  onto  $\mathcal{L}^\perp$  can be found as:

$$\mathbf{y} = (\mathbf{I} - \mathbf{L}\mathbf{L}^+)\mathbf{a} \quad (13)$$

where  $\mathbf{I}$  is an identity matrix. Since  $\mathbf{L}$  is orthonormal, this is the same as  $\mathbf{y} = (\mathbf{I} - \mathbf{L}\mathbf{L}^\top)\mathbf{a}$

# ROW SPACE

## Definition

### Definition

Let  $\mathcal{N}$  be null space of  $\mathbf{A}$ . Then orthogonal complement  $\mathcal{N}^\perp$  is called *row space* of  $\mathbf{A}$ .

Row space of  $\mathbf{A}$  is the space of all smallest-norm solutions of  $\mathbf{A}\mathbf{x} = \mathbf{y}$ , for  $\forall \mathbf{y}$ . We will denote row space as  $\text{row}(\mathbf{A})$ .



# VECTORS IN NULL AND ROW SPACES

Given vector  $\mathbf{x}$ , matrix  $\mathbf{A}$  and its null space basis  $\mathbf{N}$ , we check if  $\mathbf{x}$  is in the null space of  $\mathbf{A}$ . The simplest way is to check if  $\mathbf{Ax} = 0$ . But sometimes we may want to avoid computing  $\mathbf{Ax}$ , for example if the number of elements of  $\mathbf{A}$  is much larger than the number of elements of  $\mathbf{N}$ .

If  $\mathbf{x}$  is in the null space of  $\mathbf{A}$ , it will have zero projection onto the row space of  $\mathbf{A}$ . This gives us the condition we can check:

$$(\mathbf{I} - \mathbf{NN}^\top)\mathbf{x} = 0 \quad (14)$$

By the same logic, condition for being in the row space is as follows:

$$\mathbf{NN}^\top \mathbf{x} = 0 \quad (15)$$

Consider the following task: given a matrix  $\mathbf{A}$  find all linear combinations of its columns:  $\mathcal{C} = \{\mathbf{y} : \mathbf{y} = \mathbf{A}\mathbf{x}, \forall \mathbf{x}\}$ .

It can be re-formulated as follows: find all elements of the *column space* of  $\mathbf{A}$ .

## Definition - column space

*Column space* of  $\mathbf{A}$  is the set of all outputs of the matrix  $\mathbf{A}$ , for all possible inputs.

We will denote column space as  $\text{col}(\mathbf{A})$ . It is often called an *image* of  $\mathbf{A}$ .

The problem of finding orthonormal basis in the column space of a matrix is often called *orthonormalization* of that matrix. Hence in MATLAB and Python/Scipy the function that does it is called `orth`:

- `C = orth(A).`

- `C = scipy.linalg.orth(A).`

Let  $\mathbf{A}$  be a square matrix, a map from  $\mathbb{X} = \mathbb{R}^n$  to  $\mathbb{Y} = \mathbb{R}^n$ . Notice that if it has a non-trivial null space, it follows that multiple unique inputs are being mapped by it to the same output:

$$\begin{aligned}\mathbf{y} &= \mathbf{A}\mathbf{x}_r = \mathbf{A}(\mathbf{x}_r + \mathbf{x}_n), \\ \mathbf{x}_r &\in \text{row}(\mathbf{A}) \\ \forall \mathbf{x}_n &\in \text{null}(\mathbf{A})\end{aligned}\tag{16}$$

In fact, if null space of  $\mathbf{A}$  has  $k$  dimensions, it implies that an  $k$ -dimensional subspace of  $\mathbb{X}$  is mapped to a single element of  $\mathbb{Y}$ .

It follows that in this case the dimensionality of the column space could not exceed  $n - k$ .

Given vector  $\mathbf{y}$  and matrix  $\mathbf{A}$ , let us find  $\mathbf{y}_c$  - projection of  $\mathbf{y}$  onto the column space of  $\mathbf{A}$ .

Since  $\mathbf{y}_c \in \text{col}(\mathbf{A})$ , we can find such  $\mathbf{x}$  that  $\mathbf{Ax} = \mathbf{y}_c$ ; So, the problems is to minimize the residual  $e = \|\mathbf{y}_c - \mathbf{y}\|$  or equivalently  $e = \|\mathbf{Ax} - \mathbf{y}\|$ , which is least squares problem:  $\mathbf{x} = \mathbf{A}^+\mathbf{y}$ . So:

$$\mathbf{y}_c = \mathbf{AA}^+\mathbf{y} \in \text{col}(\mathbf{A}) \quad (17)$$

Remember that computing the pseudoinverse is based on SVD decomposition, same as finding a basis in the null space or the column space, so in terms of computational expense, all projections we discussed are similar.

Similarly we can define a projector onto the row space. Given vector  $\mathbf{x}$  and matrix  $\mathbf{A}$ , let us find projector of  $\mathbf{x}$  onto the row space of  $\mathbf{A}$ :

$$\mathbf{x}_r = \mathbf{A}^+ \mathbf{A} \mathbf{x} \in \text{row}(\mathbf{A}) \quad (18)$$

You can think of this in the following terms: first we find output  $\mathbf{A} \mathbf{x}$ , then we find the smallest norm vector that produces this same output; this vector has to 1) have the same row space projection (because output is the same), 2) has zero null space projection. Hence it is the row space projector of  $\mathbf{x}$ .

Notice that we implicitly used the fact that columns of  $\mathbf{A}^+$  lie in the row space of  $\mathbf{A}$ . We will prove this fact later. Additionally, we will prove that row space of  $\mathbf{A}$  is equivalent to the column space of  $\mathbf{A}^\top$ .

The subspace, orthogonal to the column space is called *left null space*.

## Definition

Space of all vectors  $\mathbf{y}$  orthogonal to the columns of  $\mathbf{A}$  is called *left null space*:  $\mathbf{y}^\top \mathbf{A} = 0$

You can think of left null space as a space of vectors that not only cannot be produced (as an output) by the operator  $\mathbf{A}$ , but the closest vector to them that can be produced is the zero vector.

Notice that  $\mathbf{y}^\top \mathbf{A} = 0$  implies  $\mathbf{A}^\top \mathbf{y} = 0$ , meaning that left null space of  $\mathbf{A}$  is equivalent to the null space of  $\mathbf{A}^\top$ .

If we want to project vector  $\mathbf{y}$  onto the left null space of  $\mathbf{A}$ , we project it onto the column space, and subtract the result from  $\mathbf{y}$ :

$$\mathbf{y}_l = (\mathbf{I} - \mathbf{A}\mathbf{A}^+) \mathbf{y} \in \text{left null}(\mathbf{A}) \quad (19)$$

If  $\mathbf{C}$  is an orthonormal basis in the column space of  $\mathbf{A}$ , the projection can be found the following way:

$$\mathbf{y}_l = (\mathbf{I} - \mathbf{C}\mathbf{C}^\top) \mathbf{y} \in \text{left null}(\mathbf{A}) \quad (20)$$



# SINGULAR VALUE DECOMPOSITION

Given  $\mathbf{A} \in \mathbb{R}^{n,m}$  we can find its Singular Value Decomposition (SVD):

$$\mathbf{A} = [\mathbf{C} \quad \mathbf{L}] \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}^\top \\ \mathbf{N}^\top \end{bmatrix} \quad (21)$$

$$\mathbf{A} = \mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top \quad (22)$$

where  $\mathbf{C}$ ,  $\mathbf{L}$ ,  $\mathbf{R}$  and  $\mathbf{N}$  are column, left null, row and null space bases (orthonormal),  $\mathbf{\Sigma}$  is the diagonal matrix of singular values. The singular values are positive and are sorted in the decreasing order.

Rank of the matrix is computed as the size of  $\mathbf{\Sigma}$ . Note that numeric tolerance applies when deciding if the singular value is non-zero.

Let us find SVD decomposition of a  $\mathbf{A}^\top$ :

$$\mathbf{A}^\top = [\mathbf{C}_t \quad \mathbf{L}_t] \begin{bmatrix} \boldsymbol{\Sigma}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}_t^\top \\ \mathbf{N}_t^\top \end{bmatrix} \quad (23)$$

Let us transpose it (remembering that transpose of a diagonal matrix the original matrix  $\boldsymbol{\Sigma}_t^\top = \boldsymbol{\Sigma}_t$ ):

$$\mathbf{A} = [\mathbf{R}_t \quad \mathbf{N}_t] \begin{bmatrix} \boldsymbol{\Sigma}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C}_t^\top \\ \mathbf{L}_t^\top \end{bmatrix} \quad (24)$$

Thus we can see that the row space of the original matrix  $\mathbf{A}$  is the column space of the transpose  $\mathbf{A}^\top$ . And the left null space of the original matrix  $\mathbf{A}$  is the null space of the transpose  $\mathbf{A}^\top$ .

Let us compute least squares - minimum of  $e = \|\mathbf{Ax} - \mathbf{y}\|_2$ . We find extremum:

$$\frac{d}{d\mathbf{x}} \left( (\mathbf{Ax} - \mathbf{y})^\top (\mathbf{Ax} - \mathbf{y}) \right) = 0 \quad (25)$$

$$2\mathbf{A}^\top (\mathbf{Ax} - \mathbf{y}) = 0 \quad (26)$$

$$\mathbf{A}^\top \mathbf{Ax} = \mathbf{A}^\top \mathbf{y} \quad (27)$$

We find SVD decomposition of  $\mathbf{A} = \mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top$ :

$$\mathbf{R}\mathbf{\Sigma}\mathbf{C}^\top \mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top \mathbf{x} = \mathbf{R}\mathbf{\Sigma}\mathbf{C}^\top \mathbf{y} \quad (28)$$

$$\mathbf{R}\mathbf{\Sigma}\mathbf{\Sigma}\mathbf{R}^\top \mathbf{x} = \mathbf{R}\mathbf{\Sigma}\mathbf{C}^\top \mathbf{y} \quad (29)$$

Since both sides lie in the column space of  $\mathbf{R}$ , we can multiply by  $\mathbf{R}^\top$ :

$$\mathbf{R}^\top \mathbf{R} \boldsymbol{\Sigma} \boldsymbol{\Sigma} \mathbf{R}^\top \mathbf{x} = \mathbf{R}^\top \mathbf{R} \boldsymbol{\Sigma} \mathbf{C}^\top \mathbf{y} \quad (30)$$

$$\boldsymbol{\Sigma} \boldsymbol{\Sigma} \mathbf{R}^\top \mathbf{x} = \boldsymbol{\Sigma} \mathbf{C}^\top \mathbf{y} \quad (31)$$

$$\mathbf{R}^\top \mathbf{x} = \boldsymbol{\Sigma}^{-1} \mathbf{C}^\top \mathbf{y} \quad (32)$$

Since  $\mathbf{R}$  and  $\mathbf{N}$  are orthogonal compliments, we can represent  $\mathbf{x}$  as its decomposition:  $\mathbf{x} = \mathbf{N}\mathbf{z} + \mathbf{R}\boldsymbol{\zeta}$ :

$$\mathbf{R}^\top \mathbf{N}\mathbf{z} + \mathbf{R}^\top \mathbf{R}\boldsymbol{\zeta} = \boldsymbol{\Sigma}^{-1} \mathbf{C}^\top \mathbf{y} \quad (33)$$

$$\boldsymbol{\zeta} = \boldsymbol{\Sigma}^{-1} \mathbf{C}^\top \mathbf{y} \quad (34)$$

With that we can compute  $\mathbf{x}$ :

$$\mathbf{x} = \mathbf{N}\mathbf{z} + \mathbf{R}\boldsymbol{\Sigma}^{-1} \mathbf{C}^\top \mathbf{y} \quad (35)$$

Expression  $\mathbf{x} = \mathbf{N}\mathbf{z} + \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top\mathbf{y}$  gives us all least-residual solutions.

Since  $\mathbf{N}\mathbf{z}$  is orthogonal to  $\mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top\mathbf{y}$ , we conclude that least-norm solution is given as:

$$\mathbf{x} = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top\mathbf{y} \quad (36)$$

With that we can define pseudoinverse matrix  $\mathbf{A}^+$  as:

$$\mathbf{A}^+ = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top \quad (37)$$

Note that this proves that  $\mathbf{A}^+$  lies in the row space of  $\mathbf{A}$ .

Let us prove that  $\mathbf{A}\mathbf{A}^+$  is equivalent to  $\mathbf{C}\mathbf{C}^\top$ :

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top\mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top \quad (38)$$

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{C}^\top \quad (39)$$

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{C}^\top \quad (40)$$

Let us prove that  $\mathbf{A}^+\mathbf{A}$  is equivalent to  $\mathbf{R}\mathbf{R}^\top$ :

$$\mathbf{A}^+\mathbf{A} = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top\mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top \quad (41)$$

$$\mathbf{A}^+\mathbf{A} = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{R}^\top \quad (42)$$

$$\mathbf{A}^+\mathbf{A} = \mathbf{R}\mathbf{R}^\top \quad (43)$$

Let us denote  $\mathbf{P} = \mathbf{A}\mathbf{A}^+$ . Let's prove that  $\mathbf{P}\mathbf{P} = \mathbf{P}$ :

$$\mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top\mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top\mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top\mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top \quad (44)$$

$$\mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{C}^\top \quad (45)$$

$$\mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{C}^\top \quad (46)$$

$$\mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}\mathbf{A}^+ \quad (47)$$

The same is true for  $\mathbf{P} = \mathbf{A}^+\mathbf{A}$ : we can prove that  $\mathbf{A}^+\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}^+\mathbf{A}$ .



Let's prove that  $\mathbf{P}^\top = \mathbf{P}$ .

Again, we use the fact that  $\mathbf{P} = \mathbf{C}\mathbf{C}^\top$ .

$$\mathbf{P}^\top = (\mathbf{C}\mathbf{C}^\top)^\top = \mathbf{C}\mathbf{C}^\top = \mathbf{P}. \quad \square \quad (48)$$

Let's prove that  $\mathbf{P}^+ = \mathbf{P}$ .

First, we find a basis in the linear space where  $\mathbf{P}$  projects onto:  $\mathbf{C} = \text{col}(\mathbf{P})$ , therefore  $\mathbf{P} = \mathbf{C}\mathbf{C}^\top = \mathbf{C}\mathbf{I}\mathbf{C}^\top$ , which is an SVD decomposition of  $\mathbf{P}$ . But we know how to find a pseudoinverse of a linear operator, given its SVD decomposition:

$$\mathbf{P} = \mathbf{C}\mathbf{I}\mathbf{C}^\top, \quad (49)$$

$$\mathbf{P}^+ = \mathbf{C}\mathbf{I}^{-1}\mathbf{C}^\top, \quad (50)$$

$$\mathbf{P}^+ = \mathbf{C}\mathbf{C}^\top, \quad (51)$$

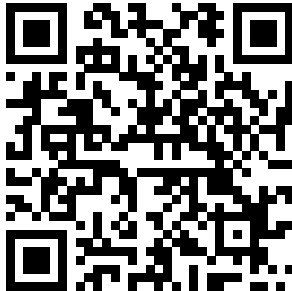
$$\mathbf{P}^+ = \mathbf{P}. \quad \square \quad (52)$$

- Minimum Norm Solutions, Math 484: Nonlinear Programming, Mikhail Lavrov
- Orthogonality, Math 484: Nonlinear Programming, Mikhail Lavrov
- Data Driven Science & Engineering. Machine Learning, Dynamical Systems, and Control, Steven L. Brunton, J. Nathan Kutz, chapter Singular Value Decomposition (SVD)

- Matrix  $\mathbf{M}$  is orthonormal and square, prove that  $\mathbf{M}^\top = \mathbf{M}^{-1}$ .
- Find minimum of  $\|\mathbf{Ax} - \mathbf{y}\|_2$  when columns of  $\mathbf{A}$  are not linearly independent.
- Given an equation  $\mathbf{Ax} = \mathbf{y}$  with a square matrix  $\mathbf{A}$ , prove that: either that equation has an exact solution for any  $\mathbf{y}$  or a related homogeneous equation  $\mathbf{Ax} = 0$  has a non-trivial solution.

Lecture slides are available via Github, links are on Moodle:

[github.com/SergeiSa/Computational-Intelligence-2024](https://github.com/SergeiSa/Computational-Intelligence-2024)



We have two definitions of orthogonality of a vector and a subspace:

- 1 Vector  $\mathbf{y}$ , whose dot product with any  $\mathbf{x} \in \mathcal{L}$  is orthogonal to the subspace  $\mathcal{L}$
- 2 If for a vector  $\mathbf{y}$ , the closest vector to it from a linear subspace  $\mathcal{L}$  is zero vector,  $\mathbf{y}$  is called orthogonal to the subspace  $\mathcal{L}$ .

Let us prove their equivalence. First we show that 1) implies 2). Let  $\mathbf{L}$  be orthonormal basis in  $\mathcal{L}$ . To find the closest element  $\mathbf{y}^*$  of  $\mathcal{L}$  to  $\mathbf{y}$ , we need to solve the least squares problem  $\mathbf{L}\mathbf{z} = \mathbf{y}$ , and multiply the solution by  $\mathbf{L}$ :

$$\mathbf{z}_{LS} = \mathbf{L}^\top \mathbf{y} = \mathbf{0} \quad (53)$$

$$\mathbf{y}^* = \mathbf{L}\mathbf{z}_{LS} = \mathbf{L}\mathbf{L}^\top \mathbf{y} = \mathbf{0} \quad (54)$$

Second, let us prove that 2) implies 1). Given that  $\mathbf{y}^* = \mathbf{L}\mathbf{z}_{LS} = \mathbf{L}\mathbf{L}^\top \mathbf{y} = \mathbf{0}$  we need to prove that  $\mathbf{L}^\top \mathbf{y} = \mathbf{0}$ . We start by multiplying the last equation by  $\mathbf{L}^\top$ :

$$\mathbf{L}\mathbf{L}^\top \mathbf{y} = \mathbf{0} \tag{55}$$

$$\mathbf{L}^\top \mathbf{L}\mathbf{L}^\top \mathbf{y} = \mathbf{0} \tag{56}$$

$$\mathbf{L}^\top \mathbf{y} = \mathbf{0} \quad \text{since } \mathbf{L}^\top \mathbf{L} = \mathbf{I}. \quad \square \tag{57}$$