

Fundamental subspaces

Fundamentals of Robotics, Lecture 5

by Sergei Savin

Fall 2022

LEAST SQUARES AT A GLANCE (1)

Consider the following problem: find the smallest 2-norm \mathbf{x} for which the equality $\mathbf{Ax} = \mathbf{y}$ has least residual. This is the *least squares problem*.

- Residual is defined as $\mathbf{e} = \mathbf{Ax} - \mathbf{y}$
- The idea that we need to find the smallest 2-norm \mathbf{x} and, at the same time, least residual \mathbf{e} seems to call for compromise. However, we will find that there exist a linear subspace of solutions with the exact same residual, and on that subspace we can find \mathbf{x} with the smallest norm.

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:

$$2\mathbf{A}^\top (\mathbf{Ax} - \mathbf{y}) = 0 \quad (1)$$

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

$$\mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} \quad (3)$$

Thus we can define a *pseudoinverse*:

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

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

$$\mathbf{x} = \mathbf{A}^+ \mathbf{y} \quad (5)$$

We proved that this is the least-residual solution, we will prove that the solution is smallest norm (out of all solutions with the same residual) in the following lectures.

You are given an equation $\mathbf{Ax} = \mathbf{y}$. You want to find an \mathbf{x} that gets the value of \mathbf{y} as close as possible to \mathbf{y}^* in terms of the 2-norm.

We know that $\mathbf{x} = \mathbf{A}^+\mathbf{y}^*$ gives us the least residual solution. Multiplying it by \mathbf{A} we get:

$$\mathbf{y} = \mathbf{AA}^+\mathbf{y}^* \tag{6}$$

This is the value of \mathbf{y} that we can achieve, which is closest to \mathbf{y}^* .

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 $\mathcal{N}(\mathbf{A})$. In the literature, it is often denoted as $\ker(\mathbf{A})$ or $\text{null}(\mathbf{A})$.

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).`

That is it! Space of solutions of $\mathbf{Ax} = \mathbf{0}$ is the span of the columns of \mathbf{N} , and all solutions \mathbf{x}^* can be represented as $\mathbf{x}^* = \mathbf{Nz}$; for any \mathbf{z} we get a unique solution, and for any solution - a unique \mathbf{z} .

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:

$$\mathbf{x} = \mathbf{N}\mathbf{z} \tag{7}$$

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

Let $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. 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{x} , whose dot product with any $\mathbf{y} \in \mathcal{L}$ is orthogonal to the subspace \mathcal{L}

Definition (equivalent)

If for a vector \mathbf{x} , the closest vector to it from a linear subspace \mathcal{L} is zero vector, \mathbf{x} 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} a *projection* of \mathbf{a} onto subspace \mathcal{L} , and such vector \mathbf{y} a projection of \mathbf{a} onto subspace \mathcal{L}^\perp .

Projection is the closest element in the subspace to a given vector. Projection of \mathbf{a} onto \mathcal{L} can be found as:

$$\mathbf{x} = \mathbf{L}\mathbf{L}^+ \mathbf{a} \tag{8}$$

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 $\mathbf{x} = \mathbf{L}\mathbf{L}^\top \mathbf{a}$, we can write:

$$\mathbf{a} = \mathbf{L}\mathbf{L}^\top \mathbf{a} + \mathbf{y} \quad (9)$$

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

$$\mathbf{y} = (\mathbf{I} - \mathbf{L}\mathbf{L}^\top) \mathbf{a} \quad (10)$$

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 subspace \mathcal{N}^\perp is called *row space* of \mathbf{A} .

Definition

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 \mathcal{R} .

VECTORS IN NULL SPACE, ROW SPACE

Given vector \mathbf{x} , matrix \mathbf{A} and its nulls space basis \mathbf{N} , and we check if \mathbf{x} is in the null space of \mathbf{A} . The simplest way is to check if $\mathbf{A}\mathbf{x} = 0$. But sometimes we may want to avoid computing $\mathbf{A}\mathbf{x}$, 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:

$$(\mathbf{I} - \mathbf{N}\mathbf{N}^\top)\mathbf{x} = 0 \quad (11)$$

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

$$\mathbf{N}\mathbf{N}^\top \mathbf{x} = 0 \quad (12)$$

Consider the following task: find all vectors \mathbf{y} such that $\mathbf{y} = \mathbf{A}\mathbf{x}$.

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 $\mathcal{C}(\mathbf{A})$. In the literature, 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).`

That is how one finds all the outputs of the matrix \mathbf{A} : as $\{\mathbf{Cz} : \forall \mathbf{z}\}$.

Notice that $\{\mathbf{Ax} : \forall \mathbf{x}\}$ might contain repeated entries if \mathbf{A} has a non-trivial null space.

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:

$$\begin{aligned}\mathbf{y} &= \mathbf{A}\mathbf{x}_r = \mathbf{A}(\mathbf{x}_r + \mathbf{x}_n), \\ \mathbf{x}_r &\in \mathcal{R}(\mathbf{A}) \\ \forall \mathbf{x}_n &\in \mathcal{N}(\mathbf{A})\end{aligned}\tag{13}$$

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 projector of \mathbf{y} onto the column space of \mathbf{A} .

This is done in the same manner as we did with the null space:

$$\mathbf{y}_c = \mathbf{A}\mathbf{A}^+\mathbf{y} \in \mathcal{C}(\mathbf{A}) \quad (14)$$

This feels nice, as it only requires the matrix itself, no need for the orthonormal basis as before. However, computing the pseudoinverse is based on SVD decomposition, same as finding a basis in the null space or the column space.

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} :

$$\mathbf{x}_r = \mathbf{A}^+ \mathbf{A} \mathbf{x} \in \mathcal{R}(\mathbf{A}) \quad (15)$$

You can think of this in the following terms: first we find what output \mathbf{x} makes, then we find the smallest norm vector that produces this same output, and this vector has to 1) have the same row space projection (because output is the same), 2) has to lie in the row space, 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 by operator \mathbf{A} , but the closest vector to them that can be produced is the zero vector.

We will later prove 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 do it as:

$$\mathbf{y}_l = (\mathbf{I} - \mathbf{A}\mathbf{A}^+) \mathbf{y} \in \mathcal{C}^\perp(\mathbf{A}) \quad (16)$$

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

$$\mathbf{y}_l = (\mathbf{I} - \mathbf{C}\mathbf{C}^\top) \mathbf{y} \in \mathcal{C}^\perp(\mathbf{A}) \quad (17)$$

SINGULAR VALUE DECOMPOSITION

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

$$\mathbf{A} = [\mathbf{C} \quad \mathbf{L}] \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}^\top \\ \mathbf{N}^\top \end{bmatrix} \quad (18)$$

$$\mathbf{A} = \mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top \quad (19)$$

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 softer in the decreasing order.

Rank of the matrix is computed as the size of Σ . Note that numeric tolerance applies when deciding if the singular value is non-zero.

Pseudoinverse \mathbf{A}^+ is defined as:

$$\mathbf{A}^+ = [\mathbf{R} \quad \mathbf{N}] \begin{bmatrix} \Sigma^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C}^\top \\ \mathbf{L}^\top \end{bmatrix} \quad (20)$$

$$\mathbf{A}^+ = \mathbf{R}\Sigma^{-1}\mathbf{C}^\top \quad (21)$$

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

Let's find SVD decomposition of a \mathbf{A}^\top :

$$\mathbf{A}^\top = [\mathbf{C}_t \quad \mathbf{L}_t] \begin{bmatrix} \boldsymbol{\Sigma}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}_t^\top \\ \mathbf{N}_t^\top \end{bmatrix} \quad (22)$$

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

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

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

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top\mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top \quad (24)$$

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{C}^\top \quad (25)$$

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{C}^\top \quad (26)$$

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

$$\mathbf{A}^+\mathbf{A} = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top\mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top \quad (27)$$

$$\mathbf{A}^+\mathbf{A} = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{R}^\top \quad (28)$$

$$\mathbf{A}^+\mathbf{A} = \mathbf{R}\mathbf{R}^\top \quad (29)$$

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

$$\mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{C}\Sigma\mathbf{R}^\top\mathbf{R}\Sigma^{-1}\mathbf{C}^\top\mathbf{C}\Sigma\mathbf{R}^\top\mathbf{R}\Sigma^{-1}\mathbf{C}^\top \quad (30)$$

$$\mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{C}\Sigma\Sigma^{-1}\Sigma\Sigma^{-1}\mathbf{C}^\top \quad (31)$$

$$\mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{C}^\top = \mathbf{A}\mathbf{A}^+ \quad (32)$$

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

THANK YOU!

Lecture slides are available via Moodle.

You can help improve these slides at:

github.com/SergeiSa/Fundamentals-of-robotics-2022

Check Moodle for additional links, videos, textbook suggestions.

