Subspaces

Computational Intelligence, Lecture 2

by Sergei Savin

Spring 2025

Given two vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}$ their *dot product* is:

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + ... + x_n y_n = \mathbf{x}^\top \mathbf{y} \tag{1}$$

A *2-norm* (also called Euclidean norm) of a vector is defined as:

$$||\mathbf{x}||_2 = \sqrt{\mathbf{x}^\top \mathbf{x}} = \sqrt{x_1 x_1 + x_2 x_2 + ... + x_n x_n} \tag{2}$$

In this course we will often have to find minimum of a square root of a function. We can make the following helpful observation:

Square of a positive-definite function

If a function $f(\mathbf{x}) \geq 0$ and $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x}$, then $(f(\mathbf{x}^*))^2 \leq (f(\mathbf{x}))^2$. Illustration on the next slide.

So, instead of finding minimum of the fucntion $f(\mathbf{x})$ we can find minimum of the function $f^2(\mathbf{x})$; both minimums will correspond to the same value of the argument $\mathbf{x}^*$.

So, if our function takes the form $f(x) = \sqrt{g(x)}$, instead of minimizing it, we can minimize $g(x)$ directly.

SQUARE OF POSITIVE X

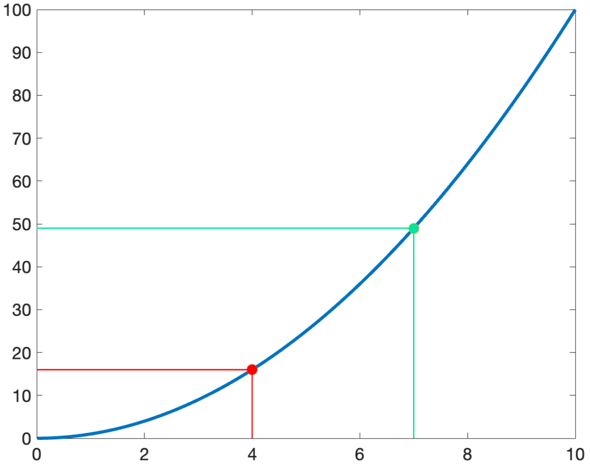

Figure 1: Graph of $f(x \geq 0) = x^2$; Since the function is monotonic, larger argument implies larger output.

Least squares at a glance (1)

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

- The value $\mathbf{e} = \mathbf{A}\mathbf{x} - \mathbf{y}$ is called residual.
- Least squares problem is about finding *least residual solution.*

Note that $f(\mathbf{x}) = ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2 = \sqrt{(\mathbf{A}\mathbf{x} - \mathbf{y})^\top(\mathbf{A}\mathbf{x} - \mathbf{y})}$; as we showed earlier, we can minimize the function $g(\mathbf{x}) = (\mathbf{A}\mathbf{x} - \mathbf{y})^\top(\mathbf{A}\mathbf{x} - \mathbf{y})$ to find the same optimal value of $\mathbf{x}$.

Least squares at a glance (2)

We find extremum of $g(\mathbf{x}) = (\mathbf{A}\mathbf{x} - \mathbf{y})^\top(\mathbf{A}\mathbf{x} - \mathbf{y})$:

$$\frac{d}{d\mathbf{x}}\left((\mathbf{A}\mathbf{x} - \mathbf{y})^\top(\mathbf{A}\mathbf{x} - \mathbf{y})\right) = 0 \tag{3}$$

$$(\mathbf{A}^\top(\mathbf{A}\mathbf{x} - \mathbf{y}))^\top + (\mathbf{A}\mathbf{x} - \mathbf{y})^\top\mathbf{A} = 0 \tag{4}$$

$$2\mathbf{A}^\top(\mathbf{A}\mathbf{x} - \mathbf{y}) = 0 \tag{5}$$

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

$$\mathbf{x} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{y} \tag{7}$$

Thus we can define a *pseudoinverse*:

$$\mathbf{A}^+ = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top \tag{8}$$

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

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{y} \tag{9}$$

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).

Pseudoinverse - orthonormal matrix

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

$$\mathbf{M}^+ = (\mathbf{M}^\top \mathbf{M})^{-1} \mathbf{M}^\top = \mathbf{M}^\top \tag{10}$$

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

$$\mathbf{x}_{LS} = \mathbf{M}^\top \mathbf{y} \tag{11}$$

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

Computing the residual

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:

$$\mathbf{e} = \mathbf{y} - \mathbf{AA}^{+}\mathbf{y} \tag{12}$$

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 $\mathbf{A}$.

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.

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 $\ker(\mathbf{A})$.

We can find all solutions of 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 we can use the function `null`, in Python/Scipy - `null_space`:

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

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:

$$\mathbf{x} = \mathbf{N}\mathbf{z} \tag{13}$$

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.

$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. 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 zero 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 $\mathrm{dot}(\mathbf{y}, \mathbf{x}) = 0, \forall \mathbf{x} \in \mathcal{L}$ is called *orthogonal complement* of $\mathcal{L}$.

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

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:

$$\mathbf{x} = \mathbf{L}\mathbf{L}^{+}\mathbf{a} \tag{14}$$

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:

$$\mathbf{a} = \mathbf{L}\mathbf{L}^{+}\mathbf{a} + \mathbf{y} \tag{15}$$

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

$$\mathbf{y} = (\mathbf{I} - \mathbf{L}\mathbf{L}^{+})\mathbf{a} \tag{16}$$

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

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 $\mathbf{A}\mathbf{x} = 0$. 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:

$$(\mathbf{I} - \mathbf{N}\mathbf{N}^\top)\mathbf{x} = 0 \tag{17}$$

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

$$\mathbf{N}\mathbf{N}^\top\mathbf{x} = 0 \tag{18}$$

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 $\mathbf{A}$.

Column space basis

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:

$$\begin{aligned} \mathbf{y} = \mathbf{A}\mathbf{x}_r = \mathbf{A}(\mathbf{x}_r + \mathbf{x}_n), \\ \mathbf{x}_r \in \mathrm{row}(\mathbf{A}) \\ \forall \mathbf{x}_n \in \mathrm{null}(\mathbf{A}) \end{aligned} \tag{19}$$

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 $n - k$.

Projector onto column space

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 $\mathbf{A}\mathbf{x} = \mathbf{y}_c$; so, the problems is to minimize the residual $e = ||\mathbf{y}_c - \mathbf{y}||$ or equivalently $e = ||\mathbf{A}\mathbf{x} - \mathbf{y}||$, which is least squares problem: $\mathbf{x} = \mathbf{A}^{+}\mathbf{y}$. So:

$$\mathbf{y}_c = \mathbf{A}\mathbf{A}^{+}\mathbf{y} \in \text{col}(\mathbf{A}) \tag{20}$$

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 $\mathbf{A}$:

$$\mathbf{x}_r = \mathbf{A}^{+}\mathbf{A}\mathbf{x} \in \mathrm{row}(\mathbf{A}) \tag{21}$$

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 1) has 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 $\mathbf{A}^{\top}$.

Left null space

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 $\mathbf{y}$:

$$\mathbf{y}_l = (\mathbf{I} - \mathbf{A}\mathbf{A}^+)\mathbf{y} \in \text{left null}(\mathbf{A}) \tag{22}$$

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

$$\mathbf{y}_l = (\mathbf{I} - \mathbf{C}\mathbf{C}^\top)\mathbf{y} \in \text{left null}(\mathbf{A}) \tag{23}$$

Further reading

- 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)

Exercise

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

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

github.com/SergeiSa/Computational-Intelligence-2025

Appendix A

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 $\mathbf{L}$:

$$\mathbf{z}_{LS} = \mathbf{L}^\top \mathbf{y} = \mathbf{0} \tag{24}$$

$$\mathbf{y}^* = \mathbf{L}\mathbf{z}_{LS} = \mathbf{L}\mathbf{L}^\top \mathbf{y} = \mathbf{0} \tag{25}$$

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 $\mathbf{L}^\top$:

$$\mathbf{L}\mathbf{L}^\top \mathbf{y} = \mathbf{0} \tag{26}$$

$$\mathbf{L}^\top \mathbf{L}\mathbf{L}^\top \mathbf{y} = \mathbf{0} \tag{27}$$

$$\mathbf{L}^\top \mathbf{y} = \mathbf{0} \qquad \text{since } \mathbf{L}^\top \mathbf{L} = \mathbf{I}. \qquad \square \tag{28}$$