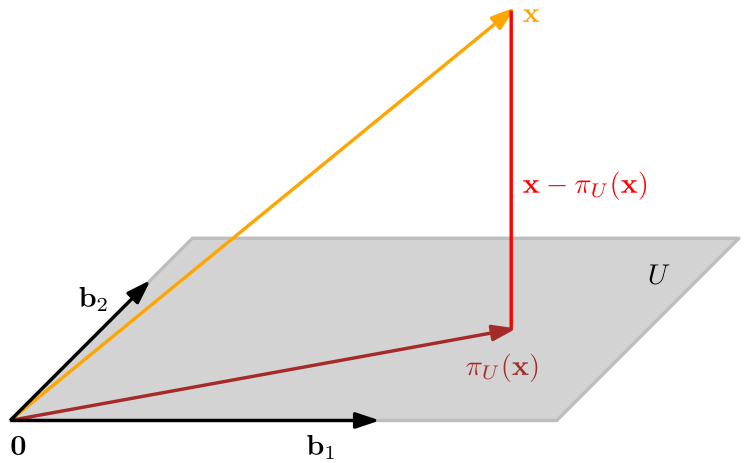
Figure 1: Projection onto a two-dimensional subspace U with basis $\boldsymbol{b}_1, \boldsymbol{b}_2$. The projection $\pi_U(\boldsymbol{x})$ of $\boldsymbol{x} \in \mathbb{R}^3$ onto U can be expressed as a linear combination of $\boldsymbol{b}_1, \boldsymbol{b}_2$ and the displacement vector $\boldsymbol{x} - \pi_U(\boldsymbol{x})$ is orthogonal to both $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$.



Orthogonal Projections onto Higher-Dimensional Subspaces

In the following, we will look at orthogonal projections of vectors $\boldsymbol{x} \in \mathbb{R}^n$ onto higher-dimensional subspaces $U \subseteq \mathbb{R}^n$ with $\dim(U) = m \geq 1$. An illustration is given in Figure 1.

Assume that $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m)$ is an ordered basis of U. Projections $\pi_U(\boldsymbol{x})$ onto U are elements of U. Therefore, they can be represented as linear combinations of the basis vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ of U, such that $\pi_U(\boldsymbol{x}) = \sum_{i=1}^m \lambda_i \boldsymbol{b}_i$.

If U is given by a set of spanning vectors, which are not a basis, make sure you determine a basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ before proceeding.

As in the 1D case, we follow a three-step procedure to find the projection $\boldsymbol{p} = \pi_U(\boldsymbol{x})$ and the projection matrix $\boldsymbol{P}_\pi$:

1. Find the coordinates $\lambda_1, \ldots, \lambda_m$ of the projection (with respect to the basis of U), such that the linear combination

$$\pi_U(\boldsymbol{x}) = \boldsymbol{p} = \sum_{i=1}^{m} \lambda_i \boldsymbol{b}_i = \boldsymbol{B}\boldsymbol{\lambda}\,, \tag{1}$$

$$\boldsymbol{B} = (\boldsymbol{b}_1 | \ldots | \boldsymbol{b}_m) \in \mathbb{R}^{n \times m},\ \boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]^\top \in \mathbb{R}^m\,, \tag{2}$$

The basis vectors form the columns of $\boldsymbol{B} \in \mathbb{R}^{n \times m}$, where $\boldsymbol{B} = (\boldsymbol{b}_1 | \ldots | \boldsymbol{b}_m)$.

is closest to $\boldsymbol{x} \in \mathbb{R}^n$. As in the 1D case, "closest" means "minimum distance", which implies that the vector connecting $\boldsymbol{p} \in U$ and $\boldsymbol{x} \in \mathbb{R}^n$ must be orthogonal to all basis vectors of U. Therefore, we obtain m simultaneous conditions (assuming the dot product as the inner product)

$$\langle \boldsymbol{b}_1, \boldsymbol{x} - \boldsymbol{p} \rangle = \boldsymbol{b}_1^\top (\boldsymbol{x} - \boldsymbol{p}) = 0 \tag{3}$$

$$\vdots \tag{4}$$

$$\langle \boldsymbol{b}_m, \boldsymbol{x} - \boldsymbol{p} \rangle = \boldsymbol{b}_m^\top (\boldsymbol{x} - \boldsymbol{p}) = 0 \tag{5}$$

which, with $\boldsymbol{p} = \boldsymbol{B}\boldsymbol{\lambda}$, can be written as

$$\boldsymbol{b}_1^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = 0 \tag{6}$$

$$\vdots \tag{7}$$

$$\boldsymbol{b}_m^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = 0 \tag{8}$$

such that we obtain a homogeneous linear equation system

$$\begin{bmatrix} \boldsymbol{b}_1^\top \\ \vdots \\ \boldsymbol{b}_m^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda} \end{bmatrix} = \boldsymbol{0} \iff \boldsymbol{B}^\top (\boldsymbol{x} - \boldsymbol{B}\boldsymbol{\lambda}) = \boldsymbol{0} \tag{9}$$

$$\iff \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{\lambda} = \boldsymbol{B}^\top \boldsymbol{x}\,. \tag{10}$$

The last expression is called *normal equation*. Since $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ are a basis of U and, therefore, linearly independent, $\boldsymbol{B}^\top \boldsymbol{B} \in \mathbb{R}^{m \times m}$ is regular and can be inverted. This allows us to solve for the optimal coefficients/ coordinates

normal equation

$$\boldsymbol{\lambda} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \boldsymbol{x}\,. \tag{11}$$

The matrix $(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top$ is also called the *pseudo-inverse* of $\boldsymbol{B}$, which can be computed for non-square matrices $\boldsymbol{B}$. It only requires that $\boldsymbol{B}^\top \boldsymbol{B}$ is positive definite, which is the case if $\boldsymbol{B}$ is full rank.[1]

pseudo-inverse

2. Find the projection $\pi_U(\boldsymbol{x}) = \boldsymbol{p} \in U$. We already established that $\boldsymbol{p} = \boldsymbol{B}\boldsymbol{\lambda}$. Therefore, with (11)

$$\boldsymbol{p} = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \boldsymbol{x}\,. \tag{12}$$

3. Find the projection matrix $\boldsymbol{P}_\pi$. From (12) we can immediately see that the projection matrix that solves $\boldsymbol{P}_\pi \boldsymbol{x} = \boldsymbol{p}$ must be

$$\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top\,. \tag{13}$$

Remark. Comparing the solutions for projecting onto a one-dimensional subspace and the general case, we see that the general case includes the 1D case as a special case: If $\dim(U) = 1$ then $\boldsymbol{B}^\top \boldsymbol{B} \in \mathbb{R}$ is a scalar and we can rewrite the projection matrix in (13) $\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top$ as $\boldsymbol{P}_\pi = \frac{\boldsymbol{B}\boldsymbol{B}^\top}{\boldsymbol{B}^\top \boldsymbol{B}}$, which is exactly the projection matrix for the one-dimensional case.

[1] In practical applications (e.g., linear regression), we often add a "jitter term" $\epsilon \boldsymbol{I}$ to $\boldsymbol{B}^\top \boldsymbol{B}$ to guarantee increase numerical stability and positive definiteness. This "ridge" can be rigorously derived using Bayesian inference.

Remark. The projections $\pi_U(\boldsymbol{x})$ are still vectors in $\mathbb{R}^n$ although they lie in an m-dimensional subspace $U \subseteq \mathbb{R}^n$. However, to represent a projected vector we only need the m coordinates $\lambda_1, \ldots, \lambda_m$ with respect to the basis vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ of U.

Remark. In vector spaces with general inner products, we have to pay attention when computing angles and distances, which are defined by means of the inner product.

Projections allow us to look at situations where we have a linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ without a solution. Recall that this means that $\boldsymbol{b}$ does not lie in the span of $\boldsymbol{A}$, i.e., the vector $\boldsymbol{b}$ does not lie in the subspace spanned by the columns of $\boldsymbol{A}$. Given that the linear equation cannot be solved exactly, we can find an *approximate solution*. The idea is to find the vector in the subspace spanned by the columns of $\boldsymbol{A}$ that is closest to $\boldsymbol{b}$, i.e., we compute the orthogonal projection of $\boldsymbol{b}$ onto the subspace spanned by the columns of $\boldsymbol{A}$. This problem arises often in practice, and the solution is called the *least squares solution* (assuming the dot product as the inner product) of an overdetermined system.

We can find approximate solutions to unsolvable linear equation systems using projections.

least squares solution