

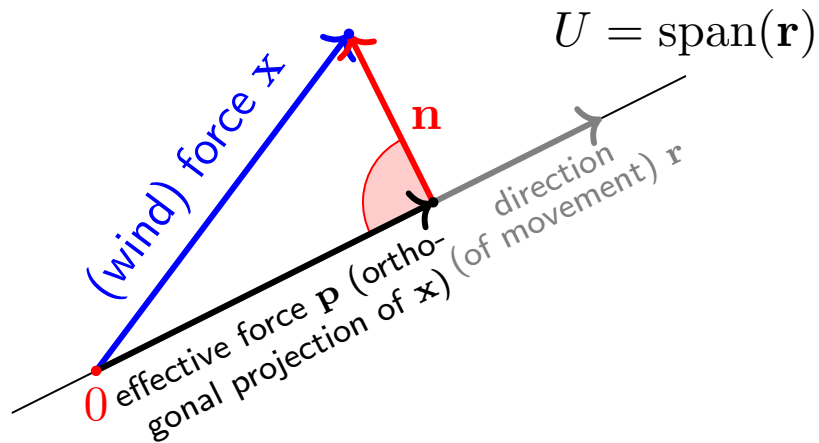
Mathematics 1 - Linear Algebra

Lecture 11 – §5.4, Orthogonal projection

Sabine Le Borne



Orthogonal projection onto a line



Given: direction (of movement) \mathbf{r} ($\neq \mathbf{o}$),
force (i.e., wind) \mathbf{x} .

Desired: (orthogonal) decomposition

$$\mathbf{x} = \mathbf{p} + \mathbf{n}$$

into effective force \mathbf{p} and
orthogonal (ineffective) force \mathbf{n} .

Definition 5.7 (Orthogonal projection onto a line in \mathbb{R}^n)

Let $\mathbf{x}, \mathbf{r} \in \mathbb{R}^n \setminus \{\mathbf{o}\}$ be given. In a decomposition

$$\mathbf{x} = \mathbf{p} + \mathbf{n}$$

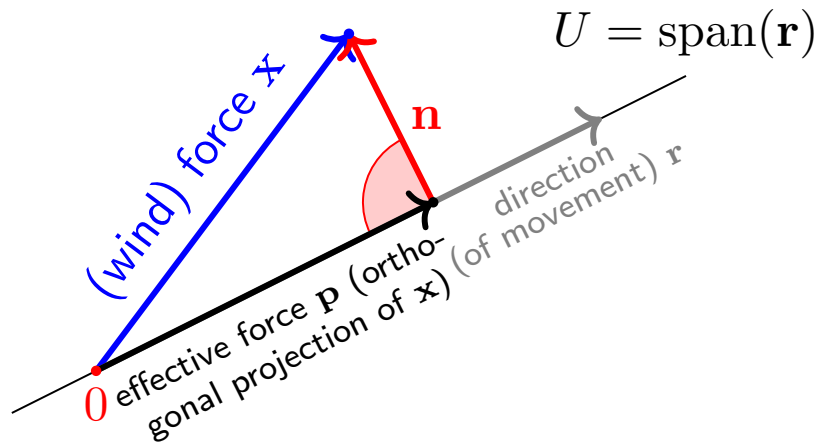
of $\mathbf{x} \in \mathbb{R}^n$ into two summands

$$\mathbf{p} \in U := \text{span}(\mathbf{r}) \quad \text{and} \quad \mathbf{n} \perp \mathbf{r},$$

the vector \mathbf{p} is called orthogonal projection of \mathbf{x} onto U .

The vector \mathbf{n} is called normal component of \mathbf{x} wrt. U .

Orthogonal projection onto a line



Given: direction (of movement) \mathbf{r} ,
force (i.e., wind) \mathbf{x} .

Desired: (orthogonal) decomposition

$$\mathbf{x} = \mathbf{p} + \mathbf{n}$$

into effective force \mathbf{p} and
orthogonal (ineffective) force \mathbf{n} .

Computation of \mathbf{p} and \mathbf{n} :

Use $\mathbf{p} = \lambda \mathbf{r}$ and form the scalar product with \mathbf{r} on both sides of the equation $\mathbf{x} = \mathbf{p} + \mathbf{n}$:

$$\langle \mathbf{x}, \mathbf{r} \rangle = \langle \underbrace{\lambda \mathbf{r}}_{\mathbf{p}} + \mathbf{n}, \mathbf{r} \rangle = \langle \lambda \mathbf{r}, \mathbf{r} \rangle + \underbrace{\langle \mathbf{n}, \mathbf{r} \rangle}_0 = \lambda \langle \mathbf{r}, \mathbf{r} \rangle \quad \text{and hence} \quad \lambda = \frac{\langle \mathbf{x}, \mathbf{r} \rangle}{\langle \mathbf{r}, \mathbf{r} \rangle}.$$

It follows that

$$\mathbf{p} = \lambda \mathbf{r} = \frac{\langle \mathbf{x}, \mathbf{r} \rangle}{\langle \mathbf{r}, \mathbf{r} \rangle} \mathbf{r}, \quad \mathbf{n} = \mathbf{x} - \mathbf{p}.$$

$$\text{Check } \langle \mathbf{p}, \mathbf{n} \rangle = \langle \lambda \mathbf{r}, \mathbf{x} - \lambda \mathbf{r} \rangle = \lambda \langle \mathbf{r}, \mathbf{x} \rangle - \lambda^2 \langle \mathbf{r}, \mathbf{r} \rangle = \frac{\langle \mathbf{x}, \mathbf{r} \rangle \langle \mathbf{r}, \mathbf{x} \rangle}{\langle \mathbf{r}, \mathbf{r} \rangle} - \frac{\langle \mathbf{x}, \mathbf{r} \rangle^2}{\langle \mathbf{r}, \mathbf{r} \rangle^2} \langle \mathbf{r}, \mathbf{r} \rangle = 0.$$

Orthogonal projection onto a line

Theorem 5.8 (orth. projection & normal component wrt. a line)

Let $\mathbf{x}, \mathbf{r} \in \mathbb{R}^n$ with $\mathbf{r} \neq \mathbf{0}$. For the orthogonal projection \mathbf{p} of \mathbf{x} onto $U = \text{span}(\mathbf{r})$ and the respective normal component \mathbf{n} of \mathbf{x} wrt. U there hold

$$\mathbf{p} = \frac{\langle \mathbf{x}, \mathbf{r} \rangle}{\langle \mathbf{r}, \mathbf{r} \rangle} \mathbf{r} \quad \text{and} \quad \mathbf{n} = \mathbf{x} - \mathbf{p} = \mathbf{x} - \frac{\langle \mathbf{x}, \mathbf{r} \rangle}{\langle \mathbf{r}, \mathbf{r} \rangle} \mathbf{r}.$$

Proof. See the previous slide. □

Example

Compute the orthogonal projection \mathbf{p} of $\mathbf{x} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$ onto the subspace $U = \text{span} \left(\begin{pmatrix} -1 \\ 1 \end{pmatrix} \right)$ and the respective normal component \mathbf{n} .

$$\mathbf{p} = \frac{\langle \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \rangle}{\langle \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \rangle} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \frac{2}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \mathbf{n} = \mathbf{x} - \mathbf{p} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}.$$

Orthogonal projection onto a subspace

So far: projection of a vector \mathbf{x} onto a 1-dimensional subspace of \mathbb{R}^n (line).

Goal: projection of a vector \mathbf{x} onto an arbitrary subspace of \mathbb{R}^n .

For this, we need so-called *orthogonal subspaces*.

Definition 5.11 (orthogonal subspace M^\perp)

Let M be a non-empty subset of \mathbb{R}^n . Then the orthogonal subspace M^\perp wrt. M is defined by

$$M^\perp := \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{m} \rangle = 0 \text{ for all } \mathbf{m} \in M\}.$$

Instead of $\mathbf{x} \in M^\perp$, we also write $\mathbf{x} \perp M$ (since $\mathbf{x} \perp \mathbf{m}$ for all $\mathbf{m} \in M$).

Examples for orthogonal subspaces

set M	subspace M^\perp
$\{\mathbf{e}_1\}, \{3\mathbf{e}_1\}, \{\mathbf{e}_1, 3\mathbf{e}_1\}$ or $\text{span}(\mathbf{e}_1)$ in \mathbb{R}^2	$\text{span}(\mathbf{e}_2) \subset \mathbb{R}^2$
$\{\mathbf{e}_1, \mathbf{e}_2\}$ or $\text{span}(\mathbf{e}_1, \mathbf{e}_2)$ in \mathbb{R}^3	$\text{span}(\mathbf{e}_3) \subset \mathbb{R}^3$
$\{\mathbf{n}\} \subset \mathbb{R}^3 \setminus \{\mathbf{o}\}$	plane in \mathbb{R}^3 through \mathbf{o} with normal vector \mathbf{n}
$\{\mathbf{o}\} \subset \mathbb{R}^n$	\mathbb{R}^n

Orthogonal projection onto a subspace

Theorem 5.13

For all subsets $M \neq \emptyset$ of \mathbb{R}^n there holds

- a) $M^\perp = (\text{span}(M))^\perp$.
- b) $(M^\perp)^\perp = \text{span}(M)$.

Proof. Exercise.

Orthogonal projection onto a subspace

Theorem 5.14 (properties of M^\perp)

- a) For every subset $M \neq \emptyset$ of \mathbb{R}^n , M^\perp is a subspace \mathbb{R}^n .
- b) For subspaces U of \mathbb{R}^n there holds $U \cap U^\perp = \{\mathbf{o}\}$.
- c) For subspaces U of \mathbb{R}^n there holds $\dim(U^\perp) = \dim(\mathbb{R}^n) - \dim(U)$.

Sketch of a proof.

- a) $\mathbf{o} \in M^\perp$, hence M^\perp is non-empty.
 $\mathbf{x}, \mathbf{x}' \in M^\perp, \alpha \in \mathbb{R} \implies$ for all $\mathbf{m} \in M$ there holds $\langle \mathbf{x} + \mathbf{x}', \mathbf{m} \rangle = \langle \mathbf{x}, \mathbf{m} \rangle + \langle \mathbf{x}', \mathbf{m} \rangle = 0$
and $\langle \alpha \mathbf{x}, \mathbf{m} \rangle = \alpha \langle \mathbf{x}, \mathbf{m} \rangle = \alpha 0 = 0$
 $\implies \mathbf{x} + \mathbf{x}' \in M^\perp, \alpha \mathbf{x} \in M^\perp$.
- b) $\mathbf{x} \in U \cap U^\perp \implies \mathbf{x} \perp \mathbf{x} \implies \langle \mathbf{x}, \mathbf{x} \rangle = 0 \implies \mathbf{x} = \mathbf{o}$.
- c) The subsequent Theorem 5.17 guarantees that every $\mathbf{x} \in \mathbb{R}^n$ can be written uniquely as $\mathbf{x} = \mathbf{p} + \mathbf{n}$ with $\mathbf{p} \in U$ and $\mathbf{n} \in U^\perp$.
Hence one obtains a basis of \mathbb{R}^n by forming the union of a basis of U and a basis of U^\perp .
Counting the basis elements yields $\dim(\mathbb{R}^n) = \dim(U) + \dim(U^\perp)$.

□

Orthogonal projection onto a subspace

Theorem 5.15 (orthogonality to all basis elements is sufficient)

Let U be a subspace of \mathbb{R}^n and let $\mathcal{B} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ be a basis of U . Then there holds for all $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x} \perp U \iff \mathbf{x} \perp \mathcal{B}.$$

In other words: A vector \mathbf{x} is orthogonal to all vectors in U if and only if it is orthogonal to each basis vector of U .

Proof.

“ \Rightarrow ”: $\mathbf{x} \perp \mathbf{u} \quad \forall \mathbf{u} \in U \implies \mathbf{x} \perp \mathbf{u}_i \quad \forall \mathbf{u}_i \in \mathcal{B} \subset U$.

“ \Leftarrow ”: Let $\mathbf{x} \perp \mathcal{B}$, i.e., $\langle \mathbf{x}, \mathbf{u}_i \rangle = 0$ for $i = 1, \dots, k$.

Let $\mathbf{u} = \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k \in U$ be arbitrary. Then there holds

$$\begin{aligned} \langle \mathbf{x}, \mathbf{u} \rangle &= \langle \mathbf{x}, \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k \rangle \\ &= \alpha_1 \langle \mathbf{x}, \mathbf{u}_1 \rangle + \dots + \alpha_k \langle \mathbf{x}, \mathbf{u}_k \rangle \\ &= \alpha_1 0 + \dots + \alpha_k 0 = 0. \end{aligned}$$



True or false?

1. Let

$$\mathbf{w} = \begin{pmatrix} 0 \\ -2 \\ -1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}.$$

- a) The orthogonal projection of \mathbf{w} onto $\text{span}(\mathbf{z})$ is $\begin{pmatrix} -1 \\ -2 \\ -2 \end{pmatrix}$.
- b) There holds $\mathbf{w} \perp \mathbf{x}$.
- c) There holds $\{\mathbf{w}\}^\perp = \{\mathbf{x}\}$.
- d) There holds $\{\mathbf{w}\}^\perp = \text{span}(\mathbf{x})$.
- e) There holds $\{\mathbf{w}\}^\perp = \text{span}(\mathbf{x}, \mathbf{y})$.
- f) There holds $\{\mathbf{w}\}^\perp = \text{span}(\mathbf{x}, \mathbf{y}, \mathbf{z})$.
- g) There holds $(\{\mathbf{w}\}^\perp)^\perp = \{\mathbf{w}\}$.

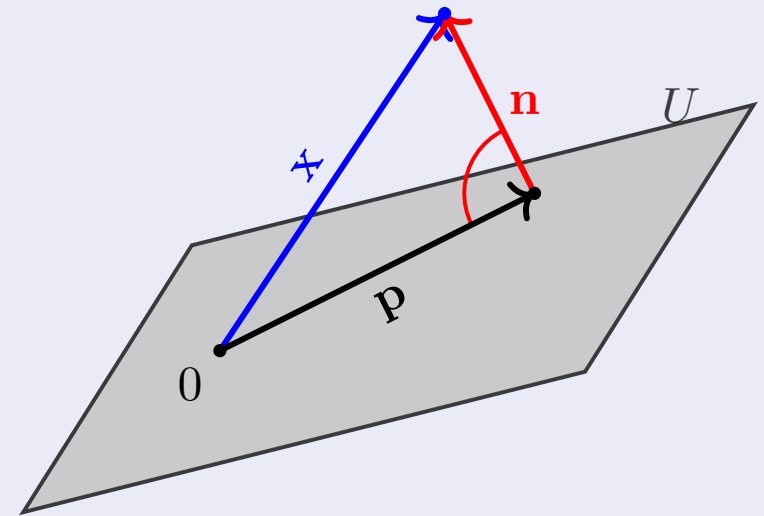
Orthogonal projection onto a subspace

Definition 5.16 (orthogonal projection onto a subspace U)

Let U be a subspace of \mathbb{R}^n and let $\mathbf{x} \in \mathbb{R}^n$.
We are looking for two vectors

$$\mathbf{p} \in U \text{ and } \mathbf{n} \perp U \text{ with } \mathbf{x} = \mathbf{p} + \mathbf{n}.$$

In other words: We write \mathbf{x} as the sum of two vectors
 $\mathbf{p} \in U$ and $\mathbf{n} \in U^\perp$.



The vector \mathbf{p} is called orthogonal projection of \mathbf{x} onto U ,
the vector \mathbf{n} is called normal component of \mathbf{x} wrt. U .

For the orthogonal projection \mathbf{p} of \mathbf{x} onto U we also write $\mathbf{x}_{\downarrow U}$.

The orthogonal decomposition $\mathbf{x} = \mathbf{p} + \mathbf{n}$ from above then looks as follows:

$$\mathbf{x} = \mathbf{x}_{\downarrow U} + \mathbf{x}_{\downarrow U^\perp}, \quad \text{i.e.,} \quad \mathbf{p} = \mathbf{x}_{\downarrow U} \quad \text{and} \quad \mathbf{n} = \mathbf{x}_{\downarrow U^\perp}.$$

Orthogonal projection onto a subspace

Computation of the orthogonal projection $\mathbf{p} = \mathbf{x}_{\downarrow U}$
(analogous to the projection onto a one-dimensional subspace):

Let $\mathcal{B} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ be a basis of U .

We want to compute coefficients $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ of

$$\mathbf{p} = \mathbf{x}_{\downarrow U} = \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k \in U.$$

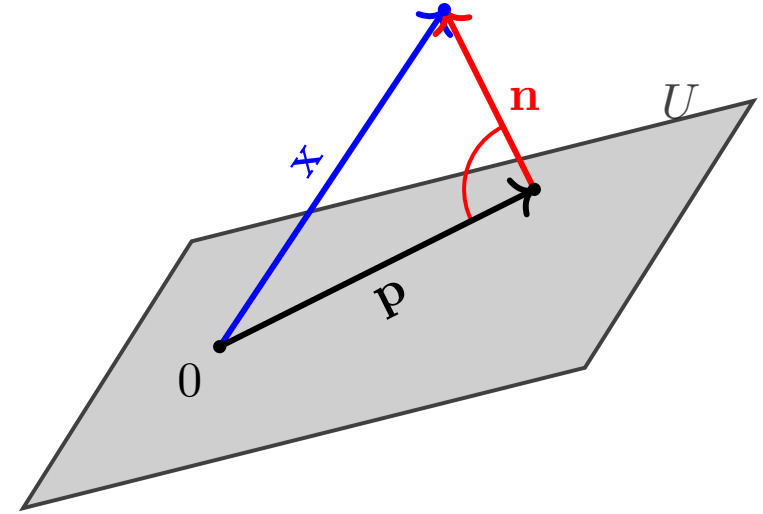
Let $\mathbf{n} = \mathbf{x}_{\downarrow U^\perp} \perp U$ such that

$$\mathbf{x} = \mathbf{p} + \mathbf{n}.$$

We now form the scalar product on both sides with the k basis vectors \mathbf{u}_i of U :

$$\begin{aligned} \langle \mathbf{x}, \mathbf{u}_i \rangle &= \langle \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k + \mathbf{n}, \mathbf{u}_i \rangle \\ &= \alpha_1 \langle \mathbf{u}_1, \mathbf{u}_i \rangle + \dots + \alpha_k \langle \mathbf{u}_k, \mathbf{u}_i \rangle + \underbrace{\langle \mathbf{n}, \mathbf{u}_i \rangle}_0 \quad \text{for all } i = 1, \dots, k. \end{aligned}$$

We hence have a system of k equations for the k unknowns $\alpha_1, \dots, \alpha_k$. The following Theorem 5.17 asserts that there (always) is a unique solution.



Orthogonal projection onto a subspace

Theorem 5.17 (orthogonal projection $\mathbf{x}_{\downarrow U}$ onto a subspace U)

Let $\mathbf{x} \in \mathbb{R}^n$ and let U be a subspace of \mathbb{R}^n with basis $\mathcal{B} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$. Then there holds

$$\mathbf{x}_{\downarrow U} = \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k$$

where $(\alpha_1, \dots, \alpha_k)^\top$ is the solution of the LES

$$\begin{pmatrix} \langle \mathbf{u}_1, \mathbf{u}_1 \rangle & \langle \mathbf{u}_2, \mathbf{u}_1 \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \langle \mathbf{u}_2, \mathbf{u}_2 \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{u}_1, \mathbf{u}_k \rangle & \langle \mathbf{u}_2, \mathbf{u}_k \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_k \rangle \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix} = \begin{pmatrix} \langle \mathbf{x}, \mathbf{u}_1 \rangle \\ \langle \mathbf{x}, \mathbf{u}_2 \rangle \\ \vdots \\ \langle \mathbf{x}, \mathbf{u}_k \rangle \end{pmatrix}.$$

The above $k \times k$ matrix is also called *Gram matrix* $G(\mathcal{B})$. It is invertible. Hence $\alpha_1, \dots, \alpha_k$ and thus also $\mathbf{x}_{\downarrow U}$ are uniquely determined.

The normal component $\mathbf{n} = \mathbf{x}_{\downarrow U}^\perp$ follows as $\mathbf{n} = \mathbf{x} - \mathbf{x}_{\downarrow U}$.

Orthogonal projection onto a subspace

Proof. We have already derived the system of equations for the α_i . We still have to show that the Gram matrix

$$G(\mathcal{B}) =: \mathbf{G} = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{u}_1 \rangle & \langle \mathbf{u}_2, \mathbf{u}_1 \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \langle \mathbf{u}_2, \mathbf{u}_2 \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{u}_1, \mathbf{u}_k \rangle & \langle \mathbf{u}_2, \mathbf{u}_k \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_k \rangle \end{pmatrix}$$

is invertible. Since \mathbf{G} is square, thanks to Theorem 3.35 we only have to show that there holds $\text{Ker}(\mathbf{G}) = \{\mathbf{o}\}$. Hence let $(\beta_1, \dots, \beta_k)^\top \in \text{Ker}(\mathbf{G})$. Then there holds

$$\begin{aligned} 0 &= (\mathbf{G}\beta)_i = \beta_1 \langle \mathbf{u}_1, \mathbf{u}_i \rangle + \dots + \beta_k \langle \mathbf{u}_k, \mathbf{u}_i \rangle \\ &= \langle \beta_1 \mathbf{u}_1 + \dots + \beta_k \mathbf{u}_k, \mathbf{u}_i \rangle \quad \text{for all } i = 1, \dots, k, \\ \implies \mathbf{u} &:= \beta_1 \mathbf{u}_1 + \dots + \beta_k \mathbf{u}_k \perp \mathbf{u}_i \quad \text{for all } i = 1, \dots, k, \\ \xRightarrow{\text{Theorem 5.15}} \mathbf{u} &\in U^\perp \implies \mathbf{u} \in U^\perp \cap U \stackrel{\text{Theorem 5.14b}}{=} \{\mathbf{o}\}, \text{ hence } \mathbf{u} = \mathbf{o}. \end{aligned}$$

Since $(\mathbf{u}_1, \dots, \mathbf{u}_k)$ is a basis and hence linearly independent, $\mathbf{u} = \mathbf{o}$ implies that $\beta_1 = \dots = \beta_k = 0$. □

Orthogonal projection onto a subspace

Projections are extremely useful: The projection $\mathbf{x}_{\downarrow U}$ of an element \mathbf{x} onto a subspace U is the best approximation of \mathbf{x} among all elements in U .

Theorem 5.18 (approximation theorem)

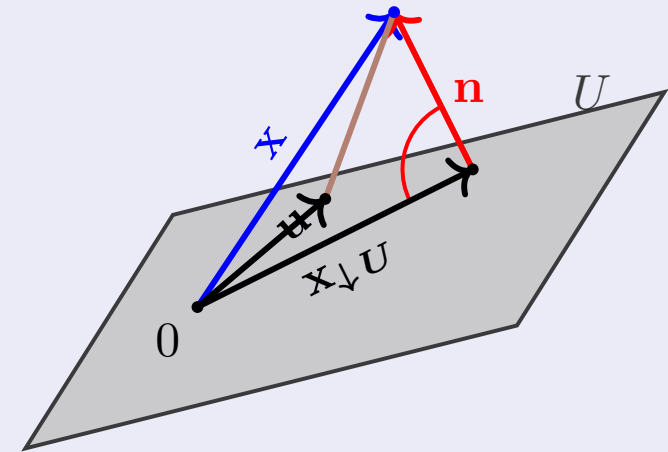
Let $\mathbf{x} \in \mathbb{R}^n$ and let U be a subspace of \mathbb{R}^n .

The orthogonal projection $\mathbf{x}_{\downarrow U}$ minimizes the distance of \mathbf{x} to points of U :

$$\underbrace{\|\mathbf{x} - \mathbf{x}_{\downarrow U}\|}_{\mathbf{n}} = \min_{\mathbf{u} \in U} \|\mathbf{x} - \mathbf{u}\| = \text{dist}(\mathbf{x}, U)$$

In other words:

No other vector of U is as close to \mathbf{x} as $\mathbf{x}_{\downarrow U}$.



Proof. For all $\mathbf{u} \in U$ there holds

$$\|\mathbf{x} - \mathbf{u}\|^2 = \underbrace{\|\mathbf{x} - \mathbf{x}_{\downarrow U}\|}_{\mathbf{n}}^2 + \underbrace{\|\mathbf{x}_{\downarrow U} - \mathbf{u}\|}_{=: \mathbf{v}}^2 = \langle \mathbf{n} + \mathbf{v}, \mathbf{n} + \mathbf{v} \rangle = \underbrace{\langle \mathbf{n}, \mathbf{n} \rangle}_{\|\mathbf{n}\|^2} + 2 \underbrace{\langle \mathbf{n}, \mathbf{v} \rangle}_0 + \underbrace{\langle \mathbf{v}, \mathbf{v} \rangle}_{\geq 0} \geq \|\mathbf{n}\|^2$$

such that $\|\mathbf{x} - \mathbf{u}\| \geq \|\mathbf{n}\| = \|\mathbf{x} - \mathbf{x}_{\downarrow U}\|$. Equality holds if and only if $\mathbf{v} = \mathbf{0}$, i.e., $\mathbf{u} = \mathbf{x}_{\downarrow U}$. \square

Orthogonal projection onto a subspace

Different view on the Approximation Theorem

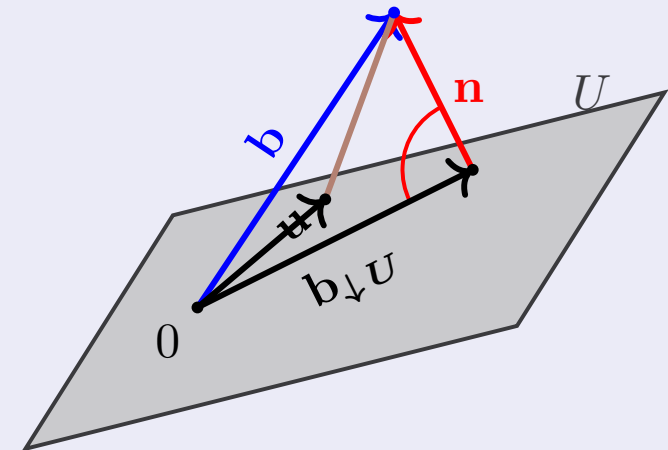
Theorem 5.18 (approximation theorem)

Let $\mathbf{b} \in \mathbb{R}^n$ and let U be a subspace of \mathbb{R}^n with basis $(\mathbf{a}_1, \dots, \mathbf{a}_k)$. The orthogonal projection $\mathbf{b}_{\downarrow U}$ minimizes the distance of \mathbf{b} to points of U :

$$\begin{aligned}\|\mathbf{b} - \mathbf{b}_{\downarrow U}\| &= \min_{\mathbf{u} \in U} \|\mathbf{b} - \mathbf{u}\| \\ &= \min_{\mathbf{x} \in \mathbb{R}^k} \|\mathbf{b} - (x_1 \mathbf{a}_1 + \dots + x_k \mathbf{a}_k)\| \\ &= \min_{\mathbf{x} \in \mathbb{R}^k} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|\end{aligned}$$

with matrix $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_k) \in \mathbb{R}^{n \times k}$.

The problem $\min_{\mathbf{x} \in \mathbb{R}^k} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$ is called a linear least squares problem. The vector $\mathbf{x} \in \mathbb{R}^k$ that solves the problem can be obtained via computation of $\mathbf{b}_{\downarrow U}$ using Theorem 5.17.



Orthogonal projection onto a subspace

Theorem 5.18 (approximation theorem, linear least squares problem)

Let $\mathbf{b} \in \mathbb{R}^n$ and let U be a subspace of \mathbb{R}^n with basis $(\mathbf{a}_1, \dots, \mathbf{a}_k)$. Define $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_k) \in \mathbb{R}^{n \times k}$. Let $\mathbf{b}_{\downarrow U}$ be the orthogonal projection of \mathbf{b} onto U . Then

$$\|\mathbf{b} - \mathbf{b}_{\downarrow U}\| = \min_{\mathbf{u} \in U} \|\mathbf{b} - \mathbf{u}\| = \min_{\mathbf{x} \in \mathbb{R}^k} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|.$$

The vector $\mathbf{x} \in \mathbb{R}^k$ that solves the problem is obtained using Theorem 5.17:

$$\underbrace{\begin{pmatrix} \langle \mathbf{a}_1, \mathbf{a}_1 \rangle & \langle \mathbf{a}_2, \mathbf{a}_1 \rangle & \dots & \langle \mathbf{a}_k, \mathbf{a}_1 \rangle \\ \langle \mathbf{a}_1, \mathbf{a}_2 \rangle & \langle \mathbf{a}_2, \mathbf{a}_2 \rangle & \dots & \langle \mathbf{a}_k, \mathbf{a}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{a}_1, \mathbf{a}_k \rangle & \langle \mathbf{a}_2, \mathbf{a}_k \rangle & \dots & \langle \mathbf{a}_k, \mathbf{a}_k \rangle \end{pmatrix}}_{\mathbf{A}^T \mathbf{A}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} \langle \mathbf{b}, \mathbf{a}_1 \rangle \\ \langle \mathbf{b}, \mathbf{a}_2 \rangle \\ \vdots \\ \langle \mathbf{b}, \mathbf{a}_k \rangle \end{pmatrix}}_{\mathbf{A}^T \mathbf{b}}$$

The system of equations $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$ is called normal equations. In particular, there holds

$$\mathbf{b}_{\downarrow U} = \mathbf{A} \mathbf{x} = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}.$$

Example for Approximation Theorem/linear least squares problem

$$\text{Let } \mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \text{ and}$$

$$U = \text{span}(\mathbf{a}_1, \mathbf{a}_2) = \text{Ran}(\mathbf{A}).$$

Compute the orthogonal projection $\mathbf{b}_{\downarrow U}$. Solve the linear least squares problem $\min_{\mathbf{x} \in \mathbb{R}^k} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$.

Gram matrix: $\mathbf{A}^T \mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}$

right hand side: $\mathbf{A}^T \mathbf{b} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

normal equations:

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b} \implies \left(\begin{array}{cc|c} 2 & 2 & 0 \\ 2 & 3 & 1 \end{array} \right) \implies \left(\begin{array}{cc|c} 2 & 2 & 0 \\ 0 & 1 & 1 \end{array} \right) \implies \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

orthogonal projection:

$$\mathbf{b}_{\downarrow U} = \mathbf{A}\mathbf{x} = -\mathbf{a}_1 + \mathbf{a}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{dist}(\mathbf{b}, U) = \min_{\mathbf{x} \in \mathbb{R}^k} \|\mathbf{b} - \mathbf{A}\mathbf{x}\| = \|\mathbf{b} - \mathbf{b}_{\downarrow U}\| = \sqrt{2}.$$

True or false?

1. Let

$$\mathbf{x} = \begin{pmatrix} 0 \\ -2 \\ -1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad U = \text{span}(\mathbf{x}, \mathbf{y}), \quad V = \text{span}(\mathbf{y}, \mathbf{z}).$$

a) There holds $\mathbf{x}_{\downarrow U} = \mathbf{x}$.

b) There holds $\mathbf{x}_{\downarrow V} = -\frac{3}{2} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.

c) There holds $\mathbf{x}_{\downarrow V^\perp} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

d) The Gram matrix $G(\mathbf{x}, \mathbf{y})$ is a diagonal matrix.