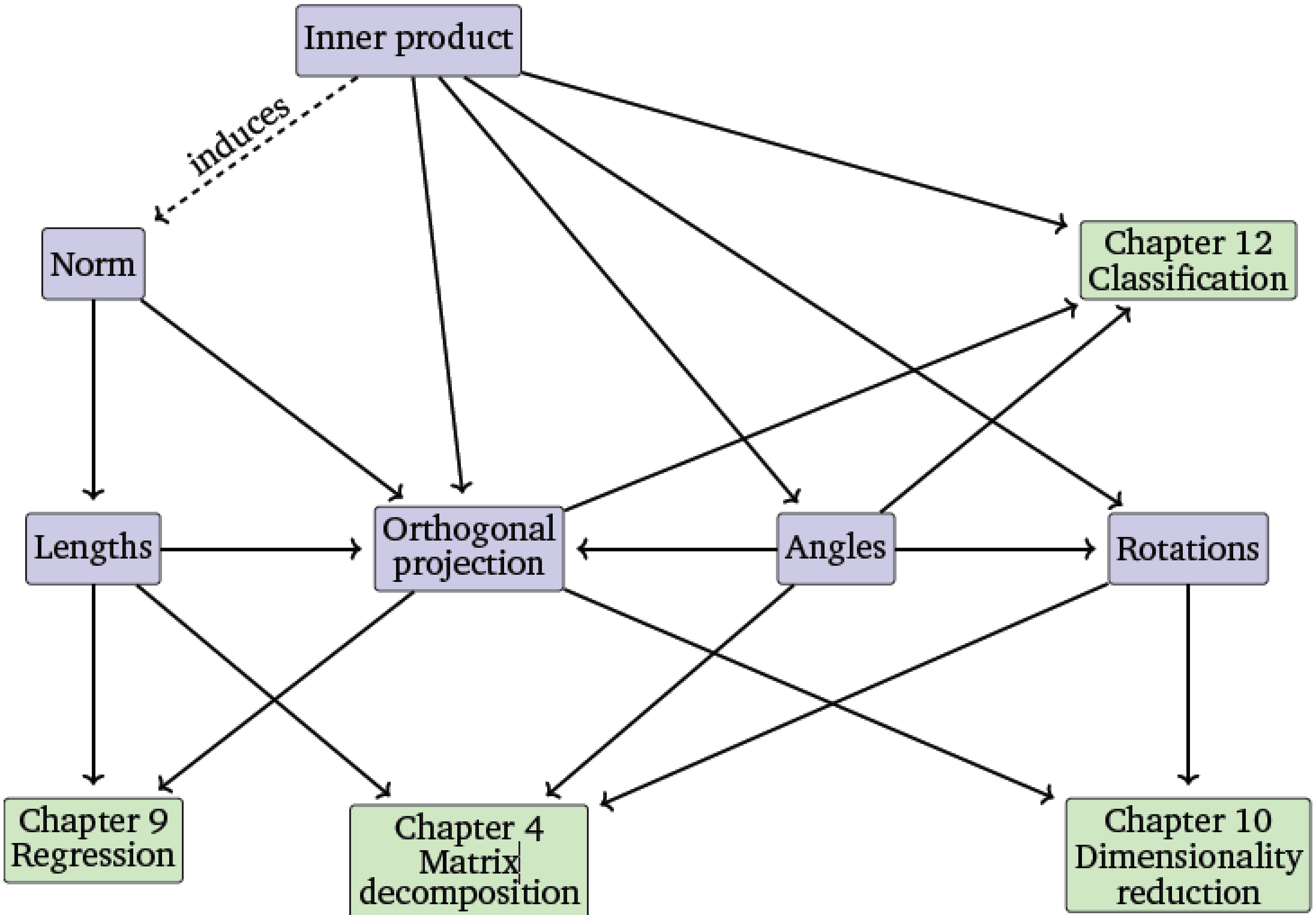
Analytic Geometry

- In particular, we will look at geometric vectors and compute their lengths and distances or angles between two vectors.
- To be able to do this, we equip the vector space with an inner product that induces the geometry of the vector space.
- Inner products and their corresponding norms and metrics capture the intuitive notions of similarity and distances, which we use to develop the support vector machine.

- The concepts of lengths and angles between vectors to discuss orthogonal projections, which will play a central role when we discuss principal component analysis.
- Regression via maximum likelihood estimation


Inner product
induces
Norm
Chapter 12
Classification
Lengths
Orthogonal
projection
Angles
Rotations
Chapter 9
Regression
Chapter 4
Matrix
decomposition
Chapter 10
Dimensionality
reduction

Norm

$$\| \cdot \| : V \rightarrow \mathbb{R} ,$$
$$x \mapsto \|x\| ,$$

- Definition 3.1 (Norm). A *norm on a vector space V is a function.*

which assigns each vector x its *length $||x|| \in R$, such that for all $\lambda \in R$* and x; y $\in$ V the following hold:

- *Absolutely homogeneous: $|| \lambda x|| = |\lambda| \; ||\lambda x||$*
- *Triangle inequality: $||x + y|| <= ||x|| + ||y||$*
- *Positive definite: $||x|| > =0$ and $||x|| = 0 \Leftrightarrow x = 0$*
- **Example 3.1 (Manhattan Norm)**

The *Manhattan norm on R^n is defined for $x \in R^n$ as Manhattan norm*

$$\|x\|_1 := \sum_{i=1}^{n} |x_i| ,$$

- where |.| is the absolute value. The left panel shows all vectors $x \in R^2$ with $||x||_1 = 1$. The Manhattan norm is also called l_1 norm.

Example 3.2 (Euclidean Norm)
The *Euclidean norm* of $x \in \mathbb{R}^n$ is defined as

$$\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^\top x} \tag{3.4}$$

and computes the *Euclidean distance* of x from the origin. The right panel of Figure 3.3 shows all vectors $x \in \mathbb{R}^2$ with $\|x\|_2 = 1$. The Euclidean norm is also called *ℓ_2 norm*.

- Inner products allow for the introduction of intuitive geometrical concepts, such as the length of a vector and the angle or distance between two vectors.
- A major purpose of inner products is to determine whether vectors are orthogonal to each other.

Example 3.5 (Lengths of Vectors Using Inner Products)
In geometry, we are often interested in lengths of vectors. We can now use an inner product to compute them using (3.16). Let us take $\boldsymbol{x} = [1, 1]^\top \in \mathbb{R}^2$. If we use the dot product as the inner product, with (3.16) we obtain

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} = \sqrt{1^2 + 1^2} = \sqrt{2} \tag{3.18}$$

as the length of $\boldsymbol{x}$. Let us now choose a different inner product:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \boldsymbol{y} = x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) + x_2 y_2 \,. \tag{3.19}$$

If we compute the norm of a vector, then this inner product returns smaller values than the dot product if x_1 and x_2 have the same sign (and $x_1 x_2 > 0$); otherwise, it returns greater values than the dot product. With this inner product, we obtain

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = x_1^2 - x_1 x_2 + x_2^2 = 1 - 1 + 1 = 1 \implies \|\boldsymbol{x}\| = \sqrt{1} = 1 \,, \tag{3.20}$$

such that $\boldsymbol{x}$ is "shorter" with this inner product than with the dot product.

Definition 3.6 (Distance and Metric). Consider an inner product space $(V, \langle\cdot,\cdot\rangle)$. Then

$$d(\boldsymbol{x}, \boldsymbol{y}) := \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle} \tag{3.21}$$

is called the *distance* between $\boldsymbol{x}$ and $\boldsymbol{y}$ for $\boldsymbol{x}, \boldsymbol{y} \in V$. If we use the dot product as the inner product, then the distance is called *Euclidean distance*.

The mapping

$$d : V \times V \to \mathbb{R} \tag{3.22}$$

$$(\boldsymbol{x}, \boldsymbol{y}) \mapsto d(\boldsymbol{x}, \boldsymbol{y}) \tag{3.23}$$

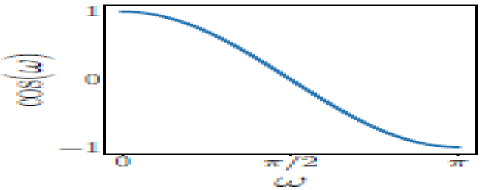
is called a *metric*.

Angles and Orthogonality

- In addition to enabling the definition of lengths of vectors, as well as the distance between two vectors, inner products also capture the geometry of a vector space by defining the angle ω between two vectors.
- We use the Cauchy-Schwarz inequality (3.17) to define angles ω in inner product spaces between two vectors x, y, and this notion coincides with our intuition in R^2 and R^3. Assume that x != 0, y != 0. Then

$$-1 \leqslant \frac{\langle x, y \rangle}{\|x\| \, \|y\|} \leqslant 1 \, .$$

- When restricted to [0,π] then f(ω) = cos(ω) returns a unique number in the interval [-1, 1].



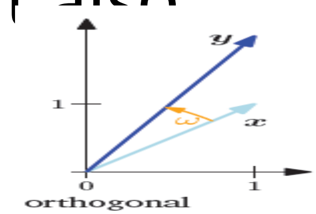
- Therefore, there exists a unique ω ϵ [0, π], illustrated in Figure, with $\cos \omega = \frac{\langle x, y \rangle}{\|x\| \, \|y\|}$.
- The number ω is the *angle between the vectors x and y.*
- Intuitively, the angle between two vectors tells us how similar their orientations are.
- For example, using the dot product, the angle between x and y = 4x, i.e., y is a scaled version of x, is 0: Their orientation is the same.

Example 3.6 (Angle between Vectors)
Let us compute the angle between $\boldsymbol{x} = [1, 1]^\top \in \mathbb{R}^2$ and $\boldsymbol{y} = [1, 2]^\top \in \mathbb{R}^2$; see Figure 3.5, where we use the dot product as the inner product. Then we get

$$\cos\omega = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle}} = \frac{\boldsymbol{x}^\top \boldsymbol{y}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{x} \boldsymbol{y}^\top \boldsymbol{y}}} = \frac{3}{\sqrt{10}}, \tag{3.26}$$

and the angle between the two vectors is $\arccos(\frac{3}{\sqrt{10}}) \approx 0.32\,\mathrm{rad}$, which corresponds to about $18°$.

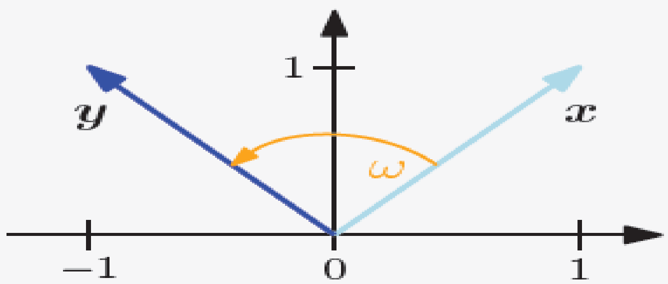
- A key feature of the inner product is that it also allows us to characterize vectors that are orthogonal.



- **Definition 3.7 (Orthogonality). Two vectors x and y are *orthogonal if and*** only if <x, y> = 0, and we write x perpendicular to y. If additionally ||x|| = 1 = ||y||, i.e., the vectors are unit vectors, then x and y are *orthonormal.*

- *Remark. Orthogonality is the generalization of the concept of perpendicularity* to bilinear forms that do not have to be the dot product.
- In our context, geometrically, we can think of orthogonal vectors as having a right angle with respect to a specific inner product.

Example 3.7 (Orthogonal Vectors)



Consider two vectors $\boldsymbol{x} = [1, 1]^\top, \boldsymbol{y} = [-1, 1]^\top \in \mathbb{R}^2$; see Figure 3.6. We are interested in determining the angle ω between them using two different inner products. Using the dot product as the inner product yields an angle ω between $\boldsymbol{x}$ and $\boldsymbol{y}$ of $90°$, such that $\boldsymbol{x} \perp \boldsymbol{y}$. However, if we choose the inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{y}, \tag{3.27}$$

we get that the angle ω between $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$\cos \omega = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} = -\frac{1}{3} \implies \omega \approx 1.91 \,\mathrm{rad} \approx 109.5°, \tag{3.28}$$

and $\boldsymbol{x}$ and $\boldsymbol{y}$ are not orthogonal. Therefore, vectors that are orthogonal with respect to one inner product do not have to be orthogonal with respect to a different inner product.

Definition 3.8 (Orthogonal Matrix)**.** A square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is an *orthogonal matrix* if and only if its columns are orthonormal so that

$$\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I} = \boldsymbol{A}^\top\boldsymbol{A}\,, \tag{3.29}$$

which implies that

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top\,, \tag{3.30}$$

i.e., the inverse is obtained by simply transposing the matrix.

Transformations by orthogonal matrices are special because the length of a vector $\boldsymbol{x}$ is not changed when transforming it using an orthogonal matrix $\boldsymbol{A}$. For the dot product, we obtain

$$\|\boldsymbol{A}\boldsymbol{x}\|^2 = (\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{x} = \|\boldsymbol{x}\|^2\,. \tag{3.31}$$

Moreover, the angle between any two vectors $\boldsymbol{x}, \boldsymbol{y}$, as measured by their inner product, is also unchanged when transforming both of them using an orthogonal matrix $\boldsymbol{A}$. Assuming the dot product as the inner product, the angle of the images $\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{A}\boldsymbol{y}$ is given as

$$\cos\omega = \frac{(\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{A}\boldsymbol{y})}{\|\boldsymbol{A}\boldsymbol{x}\|\,\|\boldsymbol{A}\boldsymbol{y}\|} = \frac{\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{y}}{\sqrt{\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}\boldsymbol{y}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{y}}} = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}\,, \tag{3.32}$$

Orthonormal Basis

- We characterized properties of basis vectors and found that in an n-dimensional vector space, we need n basis vectors, i.e., n vectors that are linearly independent.
- The special case where the basis vectors are orthogonal to each other and where the length of each basis vector is 1.

Definition 3.9 (Orthonormal Basis). Consider an n-dimensional vector space V and a basis $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ of V. If

$$\langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle = 0 \quad \text{for } i \neq j \tag{3.33}$$

$$\langle \boldsymbol{b}_i, \boldsymbol{b}_i \rangle = 1 \tag{3.34}$$

for all $i, j = 1, \ldots, n$ then the basis is called an *orthonormal basis* (*ONB*). If only (3.33) is satisfied, then the basis is called an *orthogonal basis*. Note that (3.34) implies that every basis vector has length/norm 1.

Example 3.8 (Orthonormal Basis)

The canonical/standard basis for a Euclidean vector space $\mathbb{R}^n$ is an orthonormal basis, where the inner product is the dot product of vectors.

In $\mathbb{R}^2$, the vectors

$$\boldsymbol{b}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{b}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{3.35}$$

form an orthonormal basis since $\boldsymbol{b}_1^\top \boldsymbol{b}_2 = 0$ and $\|\boldsymbol{b}_1\| = 1 = \|\boldsymbol{b}_2\|$.

Ortohogonal Complement

- Having defined orthogonality, we will now look at vector spaces that are orthogonal to each other.

Consider a D-dimensional vector space V and an M-dimensional subspace $U \subseteq V$. Then its *orthogonal complement* $U^{\perp}$ is a $(D-M)$-dimensional subspace of V and contains all vectors in V that are orthogonal to every vector in U. Furthermore, $U \cap U^{\perp} = \{\mathbf{0}\}$ so that any vector $x \in V$ can be

orthogonal complement

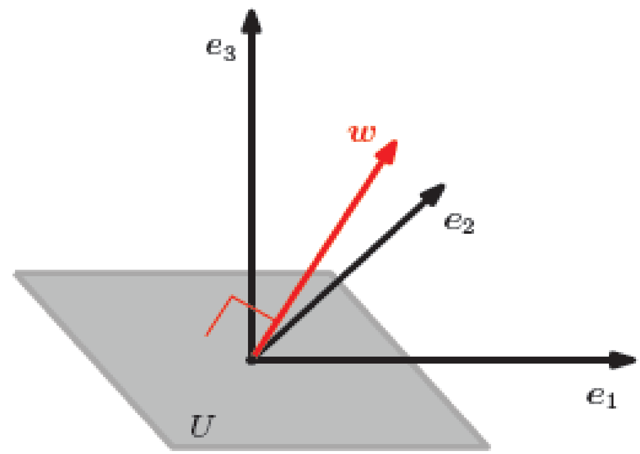


Figure 3.7 A plane U in a three-dimensional vector space can be described by its normal vector, which spans its orthogonal complement $U^\perp$.

uniquely decomposed into

$$x = \sum_{m=1}^{M} \lambda_m b_m + \sum_{j=1}^{D-M} \psi_j b_j^\perp, \quad \lambda_m,\ \psi_j \in \mathbb{R}, \tag{3.36}$$

where $(b_1, \dots, b_M)$ is a basis of U and $(b_1^\perp, \dots, b_{D-M}^\perp)$ is a basis of $U^\perp$.

Therefore, the orthogonal complement can also be used to describe a plane U (two-dimensional subspace) in a three-dimensional vector space. More specifically, the vector w with $\|w\| = 1$, which is orthogonal to the plane U, is the basis vector of $U^\perp$. Figure 3.7 illustrates this setting. All vectors that are orthogonal to w must (by construction) lie in the plane U. The vector w is called the *normal vector* of U.

normal vector

Generally, orthogonal complements can be used to describe hyperplanes in n-dimensional vector and affine spaces.

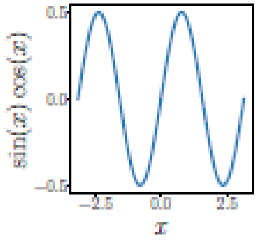
Inner product of Function

- The inner products we discussed so far were defined for vectors with a finite number of entries. We can think of a vector $x \in R^n$ as a function with n function values.
- The concept of an inner product can be generalized to vectors with an infinite number of entries (countably infinite) and also continuous-valued functions (uncountably infinite).
- Then the sum over individual components of vectors turns into an integral.
- An inner product of two functions u : R->R and v : R->R can be defined as the definite integral.

$$\langle u, v \rangle := \int_a^b u(x)v(x)dx$$

Example 3.9 (Inner Product of Functions)
If we choose $u = \sin(x)$ and $v = \cos(x)$, the integrand $f(x) = u(x)v(x)$ of (3.37), is shown in Figure 3.8. We see that this function is odd, i.e., $f(-x) = -f(x)$. Therefore, the integral with limits $a = -\pi, b = \pi$ of this product evaluates to 0. Therefore, $\sin$ and $\cos$ are orthogonal functions.

Figure 3.8 $f(x) = \sin(x)\cos(x)$.



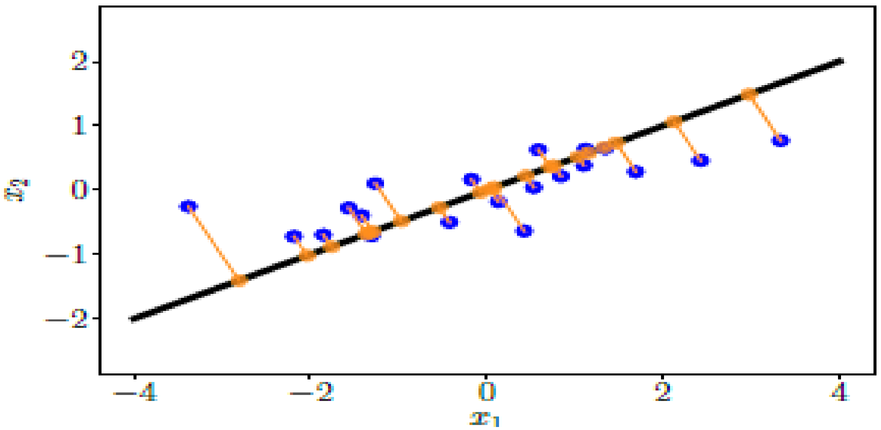
Remark. It also holds that the collection of functions

$$\{1, \cos(x), \cos(2x), \cos(3x), \dots\} \tag{3.38}$$

Orthogonal Projection

- Projections are an important class of linear transformations (besides rotations and reflections) and play an important role in graphics, coding theory, statistics and machine learning.
- In machine learning, we often deal with data that is high-dimensional. High-dimensional data is often hard to analyze or visualize.
- High-dimensional data quite often possesses the property that only a few dimensions contain most information, and most other dimensions are not essential to describe key properties of the data.
- When we compress or visualize high-dimensional data, we will lose information.
- To minimize this compression loss, we ideally find the most informative dimensions in the data.
- More specifically, we can project the original high-dimensional data onto a lower-dimensional feature space and work in this lower-dimensional space to learn more about the dataset and extract relevant patterns.

Figure 3.9 Orthogonal projection (orange dots) of a two-dimensional dataset (blue dots) onto a one-dimensional subspace (straight line).



- orthogonal projections, which we will use in Chapter 10 for linear dimensionality reduction and in Chapter 12 for classification.
- Even linear regression, which we discuss in Chapter 9, can be interpreted using orthogonal projections.
- For a given lower-dimensional subspace, orthogonal projections of high-dimensional data retain as much information as possible and minimize the difference/ error between the original data and the corresponding projection.

projection

Definition 3.10 (Projection). Let V be a vector space and $U \subseteq V$ a subspace of V. A linear mapping $\pi : V \to U$ is called a *projection* if $\pi^2 = \pi \circ \pi = \pi$.

projection matrix

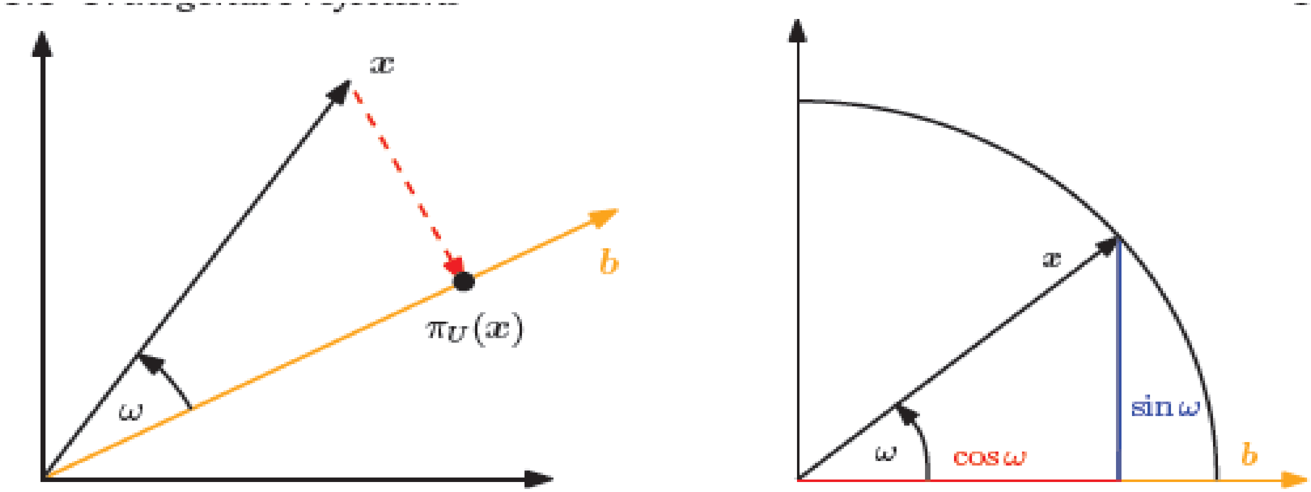
Since linear mappings can be expressed by transformation matrices (see Section 2.7), the preceding definition applies equally to a special kind of transformation matrices, the *projection matrices* P_π, which exhibit the property that $P_\pi^2 = P_\pi$.

line

In the following, we will derive orthogonal projections of vectors in the inner product space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ onto subspaces. We will start with one-dimensional subspaces, which are also called *lines*. If not mentioned otherwise, we assume the dot product $\langle x, y \rangle = x^\top y$ as the inner product.

Projection onto One-Dimensional Subspaces (Lines)

Assume we are given a line (one-dimensional subspace) through the origin with basis vector $b \in \mathbb{R}^n$. The line is a one-dimensional subspace $U \subseteq \mathbb{R}^n$ spanned by b. When we project $x \in \mathbb{R}^n$ onto U, we seek the vector $\pi_U(x) \in U$ that is closest to x. Using geometric arguments, let



(a) Projection of $x \in \mathbb{R}^2$ onto a subspace U with basis vector b.

(b) Projection of a two-dimensional vector x with $\|x\| = 1$ onto a one-dimensional subspace spanned by b.

Figure 3.10 Examples of projections onto one-dimensional subspaces.

us characterize some properties of the projection $\pi_U(x)$ (Figure 3.10(a) serves as an illustration):

- The projection $\pi_U(x)$ is closest to x, where "closest" implies that the distance $\|x - \pi_U(x)\|$ is minimal. It follows that the segment $\pi_U(x) - x$ from $\pi_U(x)$ to x is orthogonal to U, and therefore the basis vector b of U. The orthogonality condition yields $\langle \pi_U(x) - x, b\rangle = 0$ since angles between vectors are defined via the inner product.
- The projection $\pi_U(x)$ of x onto U must be an element of U and, therefore, a multiple of the basis vector b that spans U. Hence, $\pi_U(x) = \lambda b$, for some $\lambda \in \mathbb{R}$.

λ is then the coordinate of $\pi_U(x)$ with respect to b.