Definition (Transformation Matrix)

For

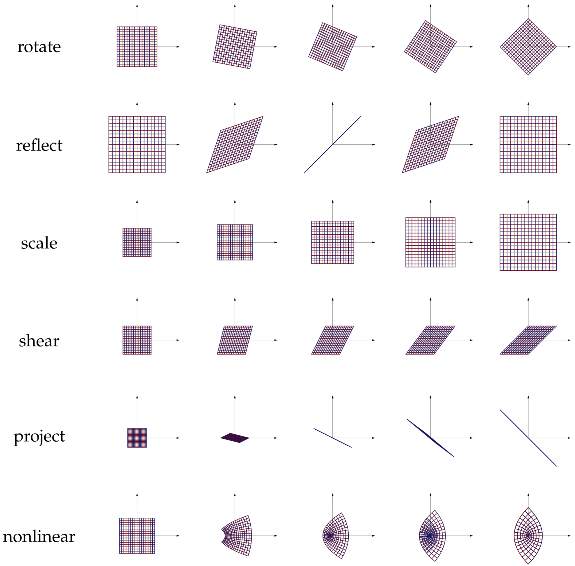
- vector spaces V and W with corresponding (ordered) bases $\mathrm{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $\mathrm{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_m)$.
- a *linear mapping* $\Phi : \mathrm{V} \to \mathrm{W}$ and $j = \{1, \ldots, n\}$, $\Phi(\mathbf{b}_j) = \alpha_{1j}\mathbf{c}_1 + \cdots + \alpha_{mj}\mathbf{c}_m = \sum_{i=1}^{m} \alpha_{ij}\mathbf{c}_i$ is the unique representation of $\Phi(\mathbf{b}_j)$ with respect to C.

$m \times n-$matrix $\mathbf{A}_\Phi$, whose elements are given by

$$A_\Phi(i, j) = \alpha_{ij},$$

called *transformation matrix* of Φ (w.r.t. the ordered bases of B of V and C of W.)

Examples of Transformation of Vectors



Theorem (Basis Change)

For a linear mapping $\Phi : \mathrm{V} \to \mathrm{W}$*, ordered bases*

$$\mathrm{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_\mathrm{n}), \quad \widetilde{\mathrm{B}} = (\widetilde{\mathbf{b}}_1, \ldots, \widetilde{\mathbf{b}}_\mathrm{n})$$

of V *and*

$$\mathrm{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_\mathrm{m}), \quad \widetilde{\mathrm{C}} = (\widetilde{\mathbf{c}}_1, \ldots, \widetilde{\mathbf{c}}_\mathrm{m})$$

of W, transformation matrices w.r.t. to the preceding ordered bases are given as:

$$\widetilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1}\mathbf{A}_\Phi \mathbf{S},$$

where $\mathbf{S} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$ *maps* $\widetilde{\mathrm{B}}$ *onto* B*, and* $\mathbf{T} \in \mathbb{R}^{\mathrm{m}\times\mathrm{m}}$ *maps* $\widetilde{\mathrm{C}}$ *onto* C*.*

Definition (Equivalence)

Two matrices $\mathbf{A}$, $\widetilde{\mathbf{A}} \in \mathcal{R}^{m \times n}$ are *equivalent* if there exist regular matrices $\mathbf{S} \in \mathbb{R}^{n \times n}$ and $\mathbf{T} \in \mathbb{R}^{m \times m}$, s.t.

$$\widetilde{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{S}.$$

Definition (Similarity)

Two matrices $\mathbf{A}$, $\widetilde{\mathbf{A}} \in \mathbb{R}^{n \times n}$ are *similar* if there exists a regular matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ with

$$\widetilde{\mathbf{A}} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}.$$

Example 4. For a linear mapping $\Phi : \mathbb{R}^3 \to \mathbb{R}^4$ and the given transformation matrix in bases B, C, find the same matrix w.r.t bases $\widetilde{\mathrm{B}}, \widetilde{\mathrm{C}}$

$$\mathbf{A}_\Phi = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}$$

$$\mathrm{B} = (\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}), \quad \mathrm{C} = (\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix})$$

$$\widetilde{\mathrm{B}} = (\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}), \quad \widetilde{\mathrm{C}} = (\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix})$$

Example 4.

$$\implies \quad \mathbf{S} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\widetilde{\mathbf{A}}_\Phi = \mathbf{T}^{-1}\mathbf{A}_\Phi\mathbf{S} = \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 3 & 2 & 1 \\ 0 & 4 & 2 \\ 10 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} -4 & -4 & -2 \\ 6 & 0 & 0 \\ 4 & 8 & 4 \\ 1 & 6 & 3 \end{bmatrix}.$$

Definition (Image & Kernel)

For $\Phi : \mathrm{V} \to \mathrm{W}$, the *kernel/null space* is:

$$\ker(\Phi) := \Phi^{-1}(\mathbf{0}_{\mathrm{W}}) = \{\mathbf{v} \in \mathrm{V} : \Phi(v) = \mathbf{0}_{\mathrm{W}}\}$$

and the *image/range* is:

$$\mathrm{Im}(\Phi) := \Phi(\mathrm{V}) = \{\mathbf{w} \in \mathrm{W} \mid \exists \mathbf{v} \in \mathrm{V} : \Phi(\mathbf{v}) = \mathbf{w}\}.$$

V and W $\longrightarrow$ *domain* and *codomain* of Φ

Graphical interpretation of “Image” and “Kernel”

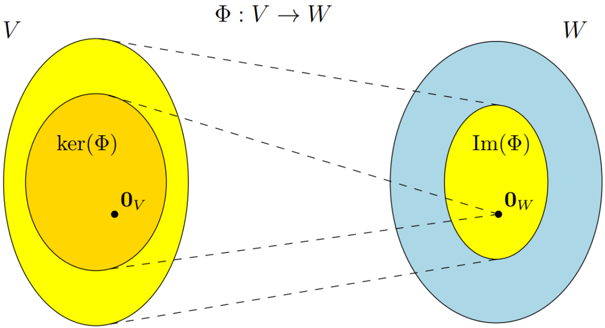


Figure: Kernel and image of a linear mapping $\Phi : \mathrm{V} \to \mathrm{W}$.

Theorem (Rank-Nullity / Fundamental Theorem of Linear Mappings)

For vector spaces V, W *and a linear mapping* $\Phi : \mathrm{V} \to \mathrm{W}$*:*

$$\dim(\ker(\Phi)) + \dim(\mathrm{Im}(\Phi)) = \dim(\mathrm{V})$$

- If $\dim(\mathrm{Im}(\Phi)) < \dim(\mathrm{V})$
  - $\ker(\Phi)$ is non-trivial $\longrightarrow$ kernel contains more than $\mathbf{0}_{\mathrm{V}}$ and $\dim(\ker(\Phi)) \geq 1$
  - If $\mathbf{A}_{\Phi}$ is the transformation matrix of Φ: $\mathbf{A}_{\Phi}\mathbf{x} = \mathbf{0}$ has infinite number of solutions
- If $\dim(\mathrm{V}) = \dim(\mathrm{W})$, these three are equivalent:
  - Φ is injective
  - Φ is surjective
  - Φ is bijective

Definition (Norm)

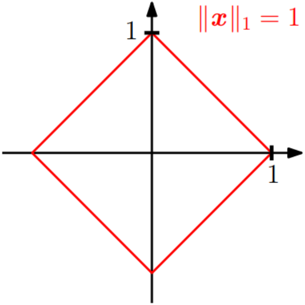
A *norm* on a vector space V is a function

$$\begin{aligned} \|\cdot\| : \mathrm{V} &\to \mathbb{R}, \\ \mathbf{x} &\mapsto \|\mathbf{x}\|, \end{aligned}$$

which assigns each vector $\mathbf{x}$ its *length* $\|\mathbf{x}\| \in \mathbb{R}$, such that:

- $\|\lambda \mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|$
- $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
- $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0 \iff \mathbf{x} = 0$

Example 5. Manhattan & Euclidean distance



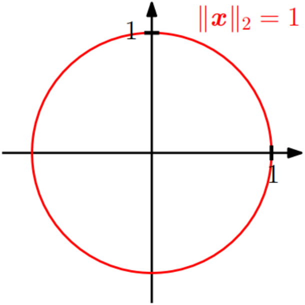


Figure: The red lines indicate the set of vectors with norm 1. Left: Manhattan norm (ℓ_1); Right: Euclidean norm (ℓ_2)

Dot product

A particular type of *Inner Product* is *dot product*:

$$\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} \mathrm{x_i y_i}$$

Definition (bilinear mapping)

A mapping Ω is a *bilinear mapping* with two arguments and is linear in each argument:

$$\Omega(\lambda \mathbf{x} + \psi \mathbf{y}, \mathbf{z}) = \lambda \Omega(\mathbf{x}, \mathbf{z}) + \psi \Omega(\mathbf{y}, \mathbf{z})$$
$$\Omega(\mathbf{x}, \lambda \mathbf{y} + \psi \mathbf{z}) = \lambda \Omega(\mathbf{x}, \mathbf{y}) + \psi \Omega(\mathbf{x}, \mathbf{z})$$

Definition (Inner Product)

For a vector space V and a bilinear mapping $\Omega : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$,

- Ω is *symmetric* if $\Omega(\mathbf{x}, \mathbf{y}) = \Omega(\mathbf{y}, \mathbf{x})$
- Ω is *positive definite* if

$$\forall \mathbf{x} \in \mathrm{V} \backslash \{\mathbf{0}\} : \Omega(\mathbf{x}, \mathbf{x}) > 0, \quad \Omega(\mathbf{0}, \mathbf{0}) = 0$$

- *Inner product* $\langle \mathbf{x}, \mathbf{y} \rangle$: A positive, symmetric bilinear mapping $\Omega : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$
- *Inner product space*: The pair $(\mathrm{V}, \langle \cdot, \cdot \rangle)$
  - For dot product, $(\mathrm{V}, \langle \cdot, \cdot \rangle)$ is a *Euclidean vector space*

Definition (Symmetric, Positive Definite Matrix)

For each symmetric matrix $\mathbf{A}$, if we have

$$\forall \mathbf{x} \in V \setminus \{\mathbf{0}\} : \mathbf{x}^\top \mathbf{A} \mathbf{x} > 0 \qquad (*)$$

$\longrightarrow$ *symmetric, positive definite*

if only $\geq$ holds $\longrightarrow$ *symmetric, positive semidefinite*

Example 6. Symmetric, Positive Definite Matrices

$$\mathbf{A}_1 = \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 9 & 6 \\ 6 & 3 \end{bmatrix}$$

$$\begin{aligned} \mathbf{x}^\top \mathbf{A}_1 \mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 9x_1^2 + 12x_1x_2 + 5x_2^2 = (3x_1 + 2x_2)^2 + x_2^2 > 0 \end{aligned}$$

$\Longrightarrow$ $\mathbf{A}_1$ is positive definite

However, $\mathbf{A}_2$ is symmetric but not positive definite (why?)

Theorem

For a real-valued, finite-dimensional vector space V*,* $\langle\cdot,\cdot\rangle : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$ *is an inner product if and only if there exists a symmetric, positive definite matrix* $\mathbf{A} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$ *with*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \hat{\mathbf{x}}^\top \mathbf{A} \hat{\mathbf{y}}$$

Thus, if $\mathbf{A} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$ is symmetric and positive definite:

- The null space (kernel) of $\mathbf{A}$ consists only of $\mathbf{0} \Longrightarrow \mathbf{Ax} \neq \mathbf{0}$ if $\mathbf{x} \neq \mathbf{0}$
- Diagonal elements a_{ii} of $\mathbf{A}$ are positive

Definition (Distance and Metric)

For an inner product space $(\mathrm{V}, \langle \cdot, \cdot \rangle)$,

$$\mathrm{d}(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}$$

is the *distance* between **x** and **y**

The mapping

$$\begin{aligned} d : \mathrm{V} \times \mathrm{V} &\to \mathbb{R} \\ (\mathbf{x}, \mathbf{y}) &\mapsto \mathrm{d}(\mathbf{x}, \mathbf{y}) \end{aligned}$$

is a *metric*

A metric d satisfies:

- d is positive definite
- d is symmetric
- $\mathrm{d}(\mathbf{x}, \mathbf{z}) \leq \mathrm{d}(\mathbf{x}, \mathbf{y}) + \mathrm{d}(\mathbf{y}, \mathbf{z})$

Angle between vectors
Example 7.

The angle between $\mathbf{x} = \begin{bmatrix}1 & 1\end{bmatrix}^\top \in \mathbb{R}^2$ and $\mathbf{y} = \begin{bmatrix}1 & 2\end{bmatrix}^\top \in \mathbb{R}^2$?

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

$$\Longrightarrow \arccos\left(\frac{3}{\sqrt{10}}\right) \approx 0.32\,\mathrm{rad}$$

Definition (Orthogonality)

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* ($\mathbf{x} \perp \mathbf{y}$) if and only if

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0$$

If additionally $\|\mathbf{x}\| = 1 = \|\mathbf{y}\|$, i.e., the vectors are unit vectors, then $\mathbf{x}$ and $\mathbf{y}$ are *orthonormal*

Definition (Orthogonal Matrix)

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is an *orthogonal matrix* if and only if its columns are orthonormal ($\mathbf{A}\mathbf{A}^\top = \mathbf{I} = \mathbf{A}^\top \mathbf{A}$), implying that

$$\mathbf{A}^{-1} = \mathbf{A}^\top$$

Transformations by Orthogonal Matrices

- Length of a vector $\mathbf{x}$ is not changed

$$\|\mathbf{Ax}\|^2 = (\mathbf{Ax})^\top(\mathbf{Ax}) = \mathbf{x}^\top\mathbf{A}^\top\mathbf{Ax} = \mathbf{x}^\top\mathbf{Ix} = \mathbf{x}^\top\mathbf{x} = \|\mathbf{x}\|^2$$

- The angle between any two vectors $\mathbf{x}, \mathbf{y}$ is unchanged

$$\cos\omega = \frac{(\mathbf{Ax})^\top(\mathbf{Ay})}{\|\mathbf{Ax}\|\,\|\mathbf{Ay}\|} = \frac{\mathbf{x}^\top\mathbf{A}^\top\mathbf{Ay}}{\sqrt{\mathbf{x}^\top\mathbf{A}^\top\mathbf{Axy}^\top\mathbf{A}^\top\mathbf{Ay}}} = \frac{\mathbf{x}^\top\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}$$

Definition (Orthonormal Basis)

For an *n*-dimensional vector space V and a basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of V, the basis is an *orthonormal basis* if

$$\langle \mathbf{b}_\mathrm{i}, \mathbf{b}_\mathrm{j} \rangle = 0 \quad \text{for } \mathrm{i} \neq \mathrm{j}$$
$$\langle \mathbf{b}_\mathrm{i}, \mathbf{b}_\mathrm{i} \rangle = 1$$

Example 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

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

form an orthonormal basis. (Why?)

Orthogonal Complement

- Generally used to describe hyperplanes in n-dimensional vector and affine spaces (important in *linear dimensionality reduction*
- For a D-dimensional vector space V and an M-dimensional subspace $U \subseteq V$, its *orthogonal complement* $U^{\perp}$ is a $(D - M)$-dimensional subspace of V
- Contains all vectors in V that are orthogonal to every vector in U
- Since $U \cap U^{\perp} = \{\mathbf{0}\}$, any vector $\mathbf{x} \in V$ can be uniquely decomposed into

$$\mathbf{x} = \sum_{m=1}^{M} \lambda_m \mathbf{b}_m + \sum_{j}^{D-M} \psi_j \mathbf{b}_j^{\perp}$$

- The vector ω with $\|\omega\| = 1$, which is orthogonal to a 2D subspace U, is the basis vector of $U^{\perp}$ (see next slide)
- The vector ω is the *normal vector* of U

Mathematical Foundations: Linear Algebra

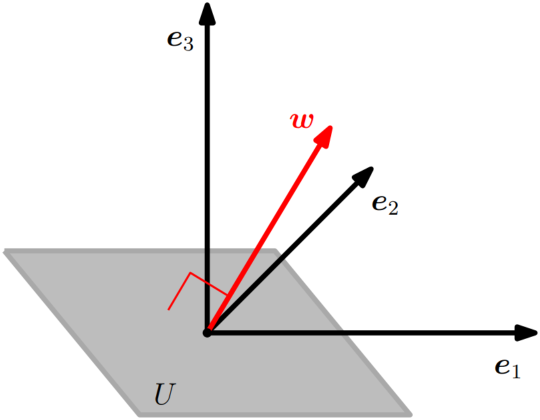


Figure: A plane U in a three-dimensional vector space can be described by its normal vector, which spans its orthogonal complement $\mathrm{U}^{\perp}$

Proejections

- An important class of linear transformations
- Used to represent the original high-dimensional data onto a lower-dimensional feature space
- A fundamental mathematical tool in *data compression* tasks
- To retain as much information as possible is to minimize the difference/error between the original high-dimensional data and the projected lower-dimensional subspace (illustrated in the next slide)

An example of an **Orthogonal Projection**

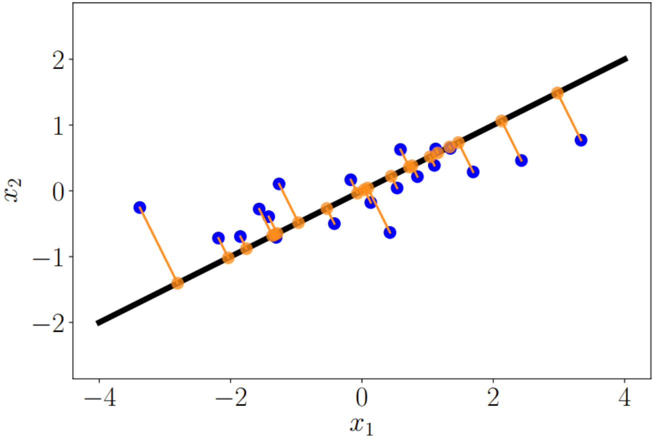


Figure: Orthogonal projection (orange dots) of a two-dimensional dataset (blue dots) onto a one-dimensional subspace (straight line)

Definition (Projection)

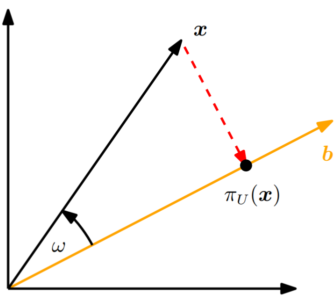
For a vector space V and $\mathrm{U} \subseteq \mathrm{V}$ a subspace of V, a linear mapping $\pi : \mathrm{V} \rightarrow \mathrm{U}$ is called a *projection* if

$$\pi^2 = \pi \circ \pi = \pi$$

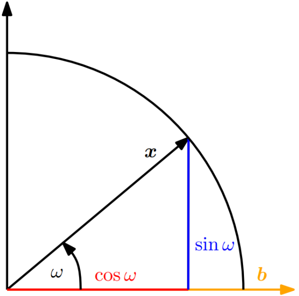
The preceding definition applies to a special kind of transformation matrices, the *projection matrices* $\mathbf{P}_\pi$, which exhibits the property that $\mathbf{P}_\pi^2 = \mathbf{P}_\pi$.

Projection onto 1D Subspaces (Lines)

By projecting $\mathbf{x} \in \mathbb{R}^n$ onto U, the vector $\pi_{\mathrm{U}}(\mathbf{x}) \in \mathrm{U}$ that is closest to $\mathbf{x}$ is sought.



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



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

Figure: Examples of projections onto 1D subspaces

Projection onto 1D Subspaces (Lines)

1. Finding the coordinate λ using $\langle \pi_{\mathrm{U}}(\mathbf{x}) - x, \pi_{\mathrm{U}}(\mathbf{x}) \rangle = 0$(note that $\pi_{\mathrm{U}}(\mathbf{x}) = \lambda \mathbf{b}$)

$$\lambda = \frac{\mathbf{b}^\top \mathbf{x}}{\mathbf{b}^\top \mathbf{b}} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2}$$

2. Finding the projection point $\pi_{\mathrm{U}}(\mathbf{x}) \in \mathrm{U}$

$$\pi_{\mathrm{U}}(\mathbf{x}) = \lambda \mathbf{b} = \frac{\langle \mathbf{x}, \mathbf{b} \rangle}{\|\mathbf{b}\|^2} \mathbf{b} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} \mathbf{b},$$

$$\|\pi_{\mathrm{U}}(\mathbf{x})\| = |\cos \omega| \, \|\mathbf{x}\|$$

3. Finding the projection matrix $\mathbf{P}_\pi$

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\|\mathbf{b}\|^2}$$

Example 9. Projection onto a Line

Find the projection matrix onto the line through the origin spanned by $\mathbf{b} = \begin{bmatrix} 1 & 2 & 2 \end{bmatrix}^\top$.

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\mathbf{b}^\top\mathbf{b}} = \frac{1}{9}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}$$

$$\pi_{\mathrm{U}}(\mathbf{x}) = \mathbf{P}_\pi\mathbf{x} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 \\ 10 \\ 10 \end{bmatrix} \in \mathrm{span}[\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}]$$

We can further show that $\pi_{\mathrm{U}}(\mathbf{x})$ is an *eigenvector* of $\mathbf{P}_\pi$, and the corresponding *eigenvaue* is 1

Projection onto a 2D subspace

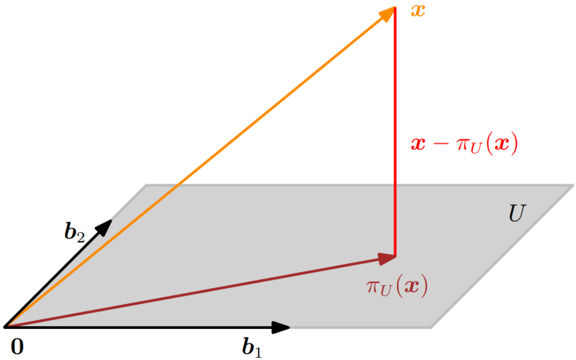


Figure: Projection onto a 2D subspace U with basis $\mathbf{b}_1$ and $\mathbf{b}_2$. The projection $\pi_{\mathrm{U}}(\mathbf{x})$ of $\mathbf{x} \in \mathbb{R}^3$ onto U can be expressed as a linear combination of $\mathbf{b}_1$, $\mathbf{b}_2$ and the displacement vector $\mathbf{x} - \pi_{\mathrm{U}}(\mathbf{x}$ is orthogonal to both $\mathbf{b}_1$ and $\mathbf{b}_2$

Projection onto General Subspaces

1. Finding the coordinates $\lambda_1, \ldots, \lambda_m$

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

$$\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_m] \in \mathbb{R}^{n\times m}, \quad \lambda = [\lambda_1, \ldots, \lambda_m]^\top \in \mathbb{R}^m,$$

$$\begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} \left[\mathbf{x} - \mathbf{B}\lambda\right] = \mathbf{0} \iff \mathbf{B}^\top(\mathbf{x} - \mathbf{B}\lambda) = \mathbf{0}$$

$$\iff \mathbf{B}^\top\mathbf{B}\lambda = \mathbf{B}^\top\mathbf{x} \ (\textit{normal equation}),$$

$$\lambda = \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{x}$$

Projection onto General Subspaces (contd.)

3. Finding the projection $\pi_{\mathrm{U}}(\mathbf{x}) \in \mathrm{U}$

$$\pi_{\mathrm{U}}(\mathbf{x}) = \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{x}$$

4. Finding the projection matrix $\mathbf{P}_\pi$

$$\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top$$

Example 10. Projection onto a 2D Subspace

For a subspace $\mathrm{U} = \mathrm{span}[\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\2\end{bmatrix}] \subseteq \mathbb{R}^3$ and $\mathbf{x} = \begin{bmatrix}6\\0\\0\end{bmatrix} \in \mathbb{R}^3$, find the projection matrix.

$$\Longrightarrow \mathbf{B} = \begin{bmatrix}1 & 0\\1 & 1\\1 & 2\end{bmatrix}$$

To find λ, we solve the normal equation $\mathbf{B}^\top \mathbf{B}\lambda = \mathbf{B}^\top \mathbf{x}$:

$$\begin{bmatrix}3 & 3\\3 & 5\end{bmatrix}\begin{bmatrix}\lambda_1 & \lambda_2\end{bmatrix} = \begin{bmatrix}6\\0\end{bmatrix} \iff \lambda = \begin{bmatrix}5\\-3\end{bmatrix}$$

Example 10. Projection onto a 2D Subspace (contd.)

$$\pi_{\mathrm{U}}(\mathbf{x}) = \mathbf{B}\lambda = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

$$\|\mathbf{x} - \pi_{\mathrm{U}}(\mathbf{x})\| = \left\| \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}^\top \right\| = \sqrt{6}$$

The projection matrix (for any $\mathbf{x} \in \mathbb{R}^3$) is

$$\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

Projections as least-squares solutions

For a linear system $\mathbf{Ax} = \mathbf{b}$, we can have approximate solutions if the systems cannot be solved exactly

- Find the vector in the subspace spanned by the columns of $\mathbf{A}$ that is closest to $\mathbf{b}$
- Computing the orthogonal projection of $\mathbf{b}$ onto the subspace spanned by the columns of $\mathbf{A}$
- This *least-squares solution* is one possible approach to derive *PCA*

Gram-Schmidt Orthogonalization

Transforming any basis $(\mathbf{b}_1, \ldots, \mathbf{b}_\mathrm{n})$ for an n-dimensional vector space V into an orthogonal/orthonormal basis $(\mathbf{u}_1, \ldots, \mathbf{u}_\mathrm{n})$ of V.

$$\mathbf{u}_1 := \mathbf{b}_1$$
$$\mathbf{u}_\mathrm{k} := \mathbf{b}_\mathrm{k} - \pi_{\mathrm{span}[\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}]}(\mathbf{b}_\mathrm{k}), \quad \mathrm{k} = 2, \ldots, \mathrm{n}.$$

- The kth basis vector $\mathbf{b}_k$ is projected onto the subspace spanned by the first $k-1$ constructed orthogonal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$
- This projection is then subtracted from $\mathbf{b}_k$ and yields a vector $\mathbf{u}_k$ that is orthogonal to the $(k-1)$-dimensional subspace spanned by $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$
- Repeating this for all n basis vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n \Longrightarrow$ an orthogonal basis $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ of V
- Normalizing the $\mathbf{u}_k \Longrightarrow$ an orthonormal basis where $\|\mathbf{u}_k\| = 1$

Example 11. Gram-Schmidt Orthogonalization

For a basis $(\mathbf{b}_1, \mathbf{b}_2)$ of $\mathbb{R}^2$,

$$\mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

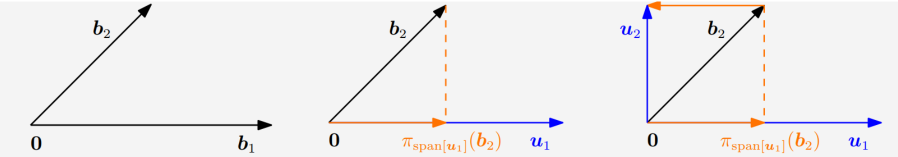
construct an orthogonal basis $(\mathbf{u}_1, \mathbf{u}_2)$ of $\mathbb{R}^2$.

$$\mathbf{u}_1 := \mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$

$$\mathbf{u}_2 := \mathbf{b}_2 - \pi_{\mathrm{span}[\mathbf{u}_1]}(\mathbf{b}_2) = \mathbf{b}_2 - \frac{\mathbf{u}_1 \mathbf{u}_1^\top}{\|\mathbf{u}_1\|^2} \mathbf{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

This procedure is illustrated in the next slide.

Example 11. Gram-Schmidt Orthogonalization (contd.)



(a) Original non-orthogonal basis vectors $\boldsymbol{b}_1, \boldsymbol{b}_2$.

(b) First new basis vector $\boldsymbol{u}_1 = \boldsymbol{b}_1$ and projection of $\boldsymbol{b}_2$ onto the subspace spanned by $\boldsymbol{u}_1$.

(c) Orthogonal basis vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2 = \boldsymbol{b}_2 - \pi_{\mathrm{span}[\boldsymbol{u}_1]}(\boldsymbol{b}_2)$.

Figure: Gram-Schmidt Orthogonalization. (a) non-orthogonal basis $(\mathbf{b}_1, \mathbf{b}_2)$; (b) first constructed basis vector $\mathbf{u}_1$ and orthogonal projection of $\mathbf{b}_2$ onto $\mathrm{span}[\mathbf{u}_1]$; (c) orthogonal basis of $(\mathbf{u}_1, \mathbf{u}_2)$

Rotation

An automorphism of a Euclidean vector space that rotates a plane by angle θ about a fixed point, like the origin

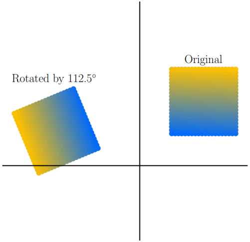


Figure: A rotation rotates objects in a plane about the origin. If the rotation angle is positive, we rotate ccw.

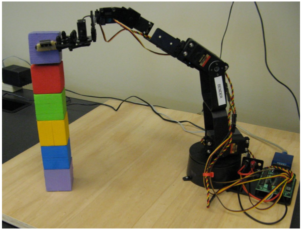
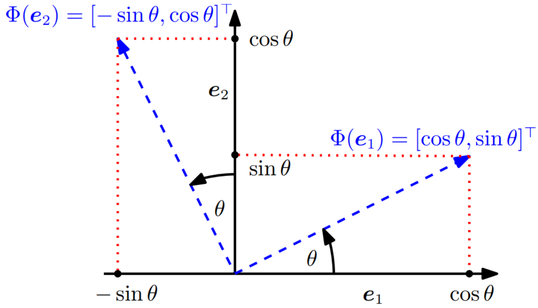


Figure: The robotic arm needs to rotate its joints in order to pick up objects or to place them correctly.

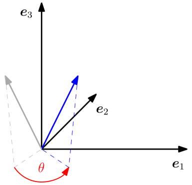
Rotations in $\mathbb{R}^2$



$$\Phi(\mathbf{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad \Phi(\mathbf{e}_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

$$\mathbf{R}(\theta) = \begin{bmatrix} \Phi(\mathbf{e}_1) & \Phi(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Rotations in $\mathbb{R}^3$



$$\mathbf{R}_1(\theta) = \begin{bmatrix} \Phi(\mathbf{e}_1) & \Phi(\mathbf{e}_2) & \Phi(\mathbf{e}_3) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix},$$

$$\mathbf{R}_2(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix},$$

$$\mathbf{R}_3(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Rotations in n Dimensions

Definition (Givens Rotation)

For an n-dimensional Euclidean vector space V and $\Phi : \mathrm{V} \to \mathrm{V}$ an automorphism with transformation matrix

$$\mathbf{R}_{\mathrm{ij}}(\theta) := \begin{bmatrix} \mathbf{I}_{\mathrm{i}-1} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cos\theta & \mathbf{0} & -\sin\theta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{\mathrm{j}-\mathrm{i}-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sin\theta & \mathbf{0} & \cos\theta & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{I}_{\mathrm{n}-\mathrm{j}} \end{bmatrix} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}},$$

with $1 \leq \mathrm{i} \leq \mathrm{j} \leq \mathrm{n}$ and $\theta \in \mathbb{R}$, the $\mathbf{R}_{\mathrm{ij}}(\theta)$ is a *Givens rotation*.

$\mathbf{R}_{\mathrm{ij}}(\theta)$ is $\mathbf{I}_{\mathrm{n}}$ with

$$r_{ii} = \cos\theta, \quad r_{ij} = -\sin\theta, \quad r_{ji} = \sin\theta, \quad r_{jj} = \cos\theta.$$

Properties of Rotations

- Preserving distances: $\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{R}_\theta(\mathbf{x}) - \mathbf{R}_\theta(\mathbf{y})\|$
- Preserving angles: the angle between $\mathbf{R}_\theta(\mathbf{x})$ and $\mathbf{R}_\theta(\mathbf{y})$ equals the angle between $\mathbf{x}$ and $\mathbf{y}$
- Not commutative in three (or more) dimensions $\longrightarrow$ order is important
- Commutative only in two dimensions: $\mathbf{R}(\phi)\mathbf{R}(\theta) = \mathbf{R}(\theta)\mathbf{R}(\phi)$

Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong.
Mathematics for machine learning.
Cambridge University Press, Cambridge and New York NY, 2020.