 

*Exercises*

**7.1.8 Problem.** Determine which of the following statements are true. If a statement is false, then provide a counterexample.

*i.* The norm of any vector is a complex number.
*ii.* The only vector of norm 0 is the zero vector.
*iii.* In any inner product space, there exists only finitely many vectors having norm 1.
*iv.* Orthonormal vectors always satisfy non-trivial linear relations.

**7.1.9 Problem.** Let $n$ be a nonnegative integer. For any vectors $\boldsymbol{v}$ in $\mathbb{C}^n$, consider the norm defined by $\|\boldsymbol{v}\|_1 := |v_1| + |v_2| + \cdots + |v_n| \in \mathbb{R}$.

*i.* Show that this norm satisfies the following four properties:

$$\begin{array}{rll} \text{(homogeneity)} & \|c\boldsymbol{v}\|_1 = |c|\,\|\boldsymbol{v}\|_1 & \text{for all } c \text{ in } \mathbb{C} \text{ and all } \boldsymbol{v} \text{ in } \mathbb{C}^n. \\ \text{(nonnegativity)} & \|\boldsymbol{v}\|_1 \geqslant 0 & \text{for all } \boldsymbol{v} \text{ in } \mathbb{C}^n. \\ \text{(positivity)} & \|\boldsymbol{v}\|_1 = 0 \ \text{ if and only if } \boldsymbol{v} = \boldsymbol{0}. & \\ \text{(subadditivity)} & \|\boldsymbol{v}+\boldsymbol{w}\|_1 \leqslant \|\boldsymbol{v}\|_1 + \|\boldsymbol{w}\|_1 & \text{for all } \boldsymbol{v} \text{ and } \boldsymbol{w} \text{ in } \mathbb{C}^n. \end{array}$$

*ii.* Whenever $n \geqslant 2$, prove that this norm does not satisfy the parallelogram identity.

**7.1.10 Problem.** For any two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ in a complex inner product space , prove that $\|\boldsymbol{v}+\boldsymbol{w}\|\;\|\boldsymbol{v}-\boldsymbol{w}\| \leqslant \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2$. When does equality hold?

**7.1.11 Problem.** For all vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ in a complex inner product space, prove that $\|\boldsymbol{u}-\boldsymbol{v}\|\,\|\boldsymbol{w}\| \leqslant \|\boldsymbol{v}-\boldsymbol{w}\|\,\|\boldsymbol{u}\| + \|\boldsymbol{w}-\boldsymbol{u}\|\,\|\boldsymbol{v}\|$.

## 7.2 Orthonormalization

HOW DO WE CONSTRUCT AN ORTHONORMAL BASIS? There is an effective process for producing an orthonormal set from any linearly independent set of vectors in an inner product space.

**7.2.0 Algorithm** (Orthonormalization).

Erhard Schmidt published this process in 1907, indicating that Jorgen Gram had essentially the same idea in 1883. However, Pierre-Simon Laplace also described this process in 1816.

input: a list $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)$ of linearly independent vectors in an inner product space.
output: an orthonormal list $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m)$ of vectors such that $\operatorname{Span}(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k) = \operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ for all $1 \leqslant k \leqslant m$.

For $k$ from 1 to $m$ do — *loop over input list*
  Set $\boldsymbol{w}_k := \boldsymbol{v}_k - \langle \boldsymbol{v}_k, \boldsymbol{u}_1 \rangle\, \boldsymbol{u}_1 - \langle \boldsymbol{v}_k, \boldsymbol{u}_2 \rangle\, \boldsymbol{u}_2 - \cdots - \langle \boldsymbol{v}_k, \boldsymbol{u}_{k-1} \rangle\, \boldsymbol{u}_{k-1}$; — *create orthogonal vectors*
  Set $\boldsymbol{u}_k := \frac{1}{\|\boldsymbol{w}_k\|} \boldsymbol{w}_k$; — *normalize the vectors*
Return the list $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m)$.

Before explaining why this algorithm produces the expected output, we illustrate it with an example.

**7.2.1 Problem.** Consider the real vector space $\mathbb{R}^4$ equipped with the standard inner product. Find an orthonormal basis for $\operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ where $\boldsymbol{v}_1 := \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^\mathsf{T}$, $\boldsymbol{v}_2 := \begin{bmatrix}0 & 1 & 1 & 1\end{bmatrix}^\mathsf{T}$, and $\boldsymbol{v}_3 := \begin{bmatrix}0 & 0 & 1 & 1\end{bmatrix}^\mathsf{T}$.

*Solution.* Applying the orthonormalization algorithm [7.2.0] gives

$$\begin{aligned}
k = 1: \quad & \boldsymbol{w}_1 = \boldsymbol{v}_1 = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} && \Rightarrow \boldsymbol{u}_1 = \tfrac{1}{2}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} \\
k = 2: \quad & \boldsymbol{w}_2 = \boldsymbol{v}_2 - \langle \boldsymbol{v}_2, \boldsymbol{u}_1 \rangle \, \boldsymbol{u}_1 \\
& \quad = \begin{bmatrix}0 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} - \tfrac{3}{4}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} = \tfrac{1}{4}\begin{bmatrix}-3 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} && \Rightarrow \boldsymbol{u}_2 = \tfrac{1}{2\sqrt{3}}\begin{bmatrix}-3 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} \\
k = 3: \quad & \boldsymbol{w}_3 = \boldsymbol{v}_3 - \langle \boldsymbol{v}_3, \boldsymbol{u}_1 \rangle \, \boldsymbol{u}_1 - \langle \boldsymbol{v}_3, \boldsymbol{u}_2 \rangle \, \boldsymbol{u}_2 \\
& \quad = \begin{bmatrix}0 & 0 & 1 & 1\end{bmatrix}^\mathsf{T} - \tfrac{2}{4}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} - \tfrac{2}{12}\begin{bmatrix}-3 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} \\
& \quad = \tfrac{1}{3}\begin{bmatrix}0 & -2 & 1 & 1\end{bmatrix}^\mathsf{T} && \Rightarrow \boldsymbol{u}_3 = \tfrac{1}{\sqrt{6}}\begin{bmatrix}0 & -2 & 1 & 1\end{bmatrix}^\mathsf{T},
\end{aligned}$$

so the list $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ is an orthonormal basis for $\operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. □

*Correctness of Algorithm 7.2.0.* We demonstrate, by induction on the number $m$ of input vectors, that the output does have the desired properties. If $m = 0$, then the algorithm returns the empty set, which is the unique basis for the zero linear space. Suppose that the number $m$ is positive. When $k = m - 1$ in the loop, the induction hypothesis establishes that $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_{m-1})$ is an orthonormal list and $\operatorname{Span}(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_{m-1}) = \operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{m-1})$. For each index $j$ satisfying $1 \leqslant j < m$, the linearity of inner products [7.0.0] and orthonormality [7.1.3] give

$$\begin{aligned}
\langle \boldsymbol{w}_m, \boldsymbol{u}_j \rangle &= \langle \boldsymbol{v}_m - \langle \boldsymbol{v}_m, \boldsymbol{u}_1 \rangle \, \boldsymbol{u}_1 - \langle \boldsymbol{v}_m, \boldsymbol{u}_2 \rangle \, \boldsymbol{u}_2 - \cdots - \langle \boldsymbol{v}_m, \boldsymbol{u}_{m-1} \rangle \, \boldsymbol{u}_{m-1}, \boldsymbol{u}_j \rangle \\
&= \langle \boldsymbol{v}_m, \boldsymbol{u}_j \rangle - \langle \boldsymbol{v}_m, \boldsymbol{u}_1 \rangle \, \langle \boldsymbol{u}_1, \boldsymbol{u}_j \rangle - \langle \boldsymbol{v}_m, \boldsymbol{u}_2 \rangle \, \langle \boldsymbol{u}_k, \boldsymbol{u}_2 \rangle - \cdots - \langle \boldsymbol{v}_m, \boldsymbol{u}_{m-1} \rangle \, \langle \boldsymbol{u}_{m-1}, \boldsymbol{u}_j \rangle \\
&= \langle \boldsymbol{v}_m, \boldsymbol{u}_j \rangle - \langle \boldsymbol{v}_m, \boldsymbol{u}_j \rangle = 0 \, .
\end{aligned}$$

Hence, the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_{m-1}, \boldsymbol{w}_m$ are pairwise orthogonal. Since the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$ are linearly independent, we deduce that the vector $\boldsymbol{v}_m$ is not in $\operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{m-1}) = \operatorname{Span}(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_{m-1})$, so $\boldsymbol{w}_m \neq \boldsymbol{0}$. Thus, the vector $\boldsymbol{u}_m = \frac{1}{\|\boldsymbol{w}_m\|} \boldsymbol{w}_m$ is well-defined and has unit length. It follows that $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$ is an orthonormal list and, in particular, linearly independent [7.1.6]. The defining equations for the vectors $\boldsymbol{u}_m$ and $\boldsymbol{w}_m$ imply that the vector $\boldsymbol{v}_m$ lies in $\operatorname{Span}(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$ and $\operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m) \subseteq \operatorname{Span}(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$. Since $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ and $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$ are linearly independent, the linear subspaces they span have the same dimension, so we conclude that $\operatorname{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m) = \operatorname{Span}(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m)$. □

This algorithm has two immediate consequences.

**7.2.2 Corollary.** *Let $V$ be a finite-dimensional inner product space.*

i. *The vector space $V$ has an orthonormal basis*
ii. *Every orthonormal list in $V$ can be extended to an orthonormal basis.*

 

*Proof.*

*i.* Because $V$ is a finite-dimensional vector space, it has a basis $v_1, v_2, \dots, v_n$ [2.2.1] where $n := \dim(V)$ is a nonnegative integer. Applying the orthonormalization algorithm to these vectors produces an orthonormal list $(u_1, u_2, \dots, u_n)$ such that

$$\operatorname{Span}(u_1, u_2, \dots, u_n) = \operatorname{Span}(v_1, v_2, \dots, v_n) = V\,.$$

An orthonormal set of vectors in linearly independent [7.1.6], so we conclude that $u_1, u_2, \dots, u_n$ is a basis for $V$.

*ii.* Suppose the list $(u_1, u_2, \dots, u_m)$ of vectors in $V$ is orthonormal. Since the vectors $u_1, u_2, \dots, u_n$ are linearly independent [7.1.6], they can be extended to a basis [2.2.1]. In other words, there exists vectors $v_{m+1}, v_{m+2}, \dots, v_n$ in $V$ such that the list

$$(u_1, u_2, \dots, u_m, v_{m+1}, v_{m+2}, \dots, v_n)$$

forms a basis for $V$. Applying the orthonormalization algorithm to this list produces an orthonormal basis $V$. Moreover, the algorithm does not change the first $m$ vectors because they are already orthonormal. More explicitly, for all $1 \leqslant k \leqslant m$, we have

$$w_k = u_k - \langle u_k, u_1\rangle\, u_1 - \langle u_k, u_2\rangle\, u_2 - \cdots - \langle u_k, u_{k-1}\rangle\, u_{k-1} = u_k\,. \quad \square$$

Since the empty set is vacuously an orthonormal list, part *ii* implies part *i*.

As a second illustration of the orthonormalization algorithm, we construct some orthogonal polynomials.

**7.2.3 Problem.** Consider the $\mathbb{R}$-vector space $\mathbb{R}[t]_{\leqslant 2}$ equipped with the inner product $\langle f, g\rangle := \int_{-1}^{1} f(x)\, g(x)\, dx$ where $f$ and $g$ in $\mathbb{R}[t]_{\leqslant 2}$. Convert the monomial basis $(1, t, t^2)$ into an orthonormal basis.

*Solution.* The orthonormalization algorithm gives

$$\begin{aligned}
k = 1: \quad & w_1 = 1\,, \\
& \|w_1\|^2 = \int_{-1}^{1} 1\, dx = 2\,, && \Rightarrow\; u_1 = \tfrac{1}{\|w_1\|} w_1 = \tfrac{1}{\sqrt{2}}\,; \\
k = 2: \quad & w_2 = t - \langle t, u_1\rangle\, u_1 = t - \tfrac{1}{2}\int_{-1}^{1} x\, dx = t - \tfrac{1}{2}\Big[\tfrac{1}{2}x^2\Big]_{-1}^{1} = t\,, \\
& \|w_2\|^2 = \int_{-1}^{1} x^2\, dx = \Big[\tfrac{1}{3}x^3\Big]_{-1}^{1} = \tfrac{2}{3}\,, && \Rightarrow\; u_2 = \tfrac{1}{\|w_2\|} w_2 = \tfrac{\sqrt{3}}{\sqrt{2}}\, t\,; \\
k = 3: \quad & w_3 = t^2 - \langle t^2, u_1\rangle\, u_1 - \langle t^2, u_2\rangle\, u_2 = t^2 - \tfrac{1}{2}\int_{-1}^{1} x^2\, dx - \tfrac{3}{2}\, t \int_{-1}^{1} x^3\, dx \\
& \quad = t^2 - \tfrac{1}{2}\Big[\tfrac{1}{3}\, x^3\Big]_{-1}^{1} - \tfrac{3}{2}\Big[\tfrac{1}{4}\, x^4\Big]_{-1}^{1} t = t^3 - \tfrac{1}{3}\,, \\
& \|w_3\|^2 = \int_{-1}^{1} (x^2 - \tfrac{1}{3})^2\, dx = \tfrac{1}{9}\int_{-1}^{1} 9\, x^4 - 6\, x^2 + 1\, dx \\
& \quad = \tfrac{1}{9}\Big[\tfrac{9}{5}\, x^5 - 2\, x^3 + x\Big]_{-1}^{1} = \tfrac{2}{45}(9 - 10 + 5) = \tfrac{8}{45}\,, && \Rightarrow\; u_3 = \tfrac{1}{\|w_3\|} w_3 = \tfrac{\sqrt{5}}{2\sqrt{2}}(3\, t^2 - 1)\,.
\end{aligned}$$

Therefore, the polynomials $\frac{1}{\sqrt{2}}, \frac{\sqrt{3}}{\sqrt{2}}\, t, \frac{\sqrt{5}}{2\sqrt{2}}(3\, t^2 - 1)$ form an orthonormal basis of the space $\mathbb{R}[t]_{\leqslant 2}$ of polynomials of degree at most 2. $\square$

This orthonormal basis already appeared in Problem 7.1.4.

*Exercises*

**7.2.4 Problem.** Determine which of the following statements are true. If a statement is false, then provide a counterexample.

*i.* Every linearly independent set of vectors in an inner product space is orthonormal.
*ii.* Every orthogonal list can be extended to an basis of pairwise orthogonal vectors.
*iii.* By rescaling the vectors, any orthogonal list can be converted into an orthonormal list.
*iv.* The output of the orthonormalization algorithm depends on the order of the vectors in the input list.

**7.2.5 Problem.** Set $n \coloneqq 2$. Consider the $\mathbb{R}$-vector space $\mathbb{R}[t]_{\leqslant n}$ with the inner product defined, for all polynomials $f$ and $g$ in $\mathbb{R}[t]_{\leqslant n}$, by

$$\langle f, g \rangle \coloneqq \int_0^\infty f(x)\, g(x)\, e^{-x}\, dx\,.$$

*i.* Transform the monomial basis $(1, t, t^2, \dots, t^n)$ of $\mathbb{R}[t]_{\leqslant n}$ into a orthonormal basis $\big(L_0(t), L_1(t), \dots, L_n(t)\big)$ by applying the orthonormalization algorithm.
*ii.* For each integer $j$ satisfying $0 \leqslant j \leqslant n$, consider the linear operator $D_j \colon \mathbb{R}[t]_{\leqslant n} \to \mathbb{R}[t]_{\leqslant n}$ defined, for all polynomials $f$ in $\mathbb{R}[t]_{\leqslant n}$, by $D_j[f] \coloneqq t\, f''(t) + (1-t)\, f'(t) + j\, f(t)$. For all $0 \leqslant j \leqslant n$, show that $\operatorname{Span}(L_j) = \operatorname{Ker}(D_j)$.

# 8
# Orthogonal Projections



Orthogonal projections produce to a data-fitting technique. The best fit in the least-squares sense minimizes the sum of squared residuals—the difference between an observed value and the fitted value provided by a model. This chapter develops this idea.

## 8.0 Projections

HOW DO WE UNDERSTAND GENERAL ORTHOGONAL PROJECTIONS? Although we have discussed orthogonal projections onto a line, we gain new insights by generalizing these ideas to linear operators. We first identify a special type of linear operator.

**8.0.0 Definition.** A ***projection*** is a linear operator $P$ such that $P^2 = P$.

For any scalar $c$ in $\mathbb{C}$, the matrix $\left[\begin{smallmatrix} 1 & c \\ 0 & 0 \end{smallmatrix}\right]$ defines a projection on $\mathbb{C}^2$ via left multiplication. Indeed, we have

$$\begin{bmatrix} 1 & c \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & c \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1+0 & c+0 \\ 0+0 & 0+0 \end{bmatrix} = \begin{bmatrix} 1 & c \\ 0 & 0 \end{bmatrix}.$$

**8.0.1 Lemma** (Properties of projections)**.** *Let $V$ be a vector space and let $P \colon V \to V$ be a projection.*

i. *For any vector $\boldsymbol{w}$ in the image* $\operatorname{Im}(P)$*, we have* $P[\boldsymbol{w}] = \boldsymbol{w}$.
ii. *We have* $\operatorname{Im}(P) \cap \operatorname{Ker}(P) = \{\boldsymbol{0}\}$.
iii. *For any vector $\boldsymbol{v}$ in $V$, there exists unique vectors $\boldsymbol{w}$ in* $\operatorname{Im}(P)$ *and $\boldsymbol{z}$ in* $\operatorname{Ker}(P)$ *such that* $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{z}$, $P[\boldsymbol{v}] = \boldsymbol{w}$, *and* $(\operatorname{id}_V - P)[\boldsymbol{v}] = \boldsymbol{z}$.

*Proof.*

*i.* For any vector $\boldsymbol{w}$ in $\operatorname{Im}(P)$, there is a vector $\boldsymbol{v}$ such that $\boldsymbol{w} = P[\boldsymbol{v}]$. Since $P^2 = P$, we have $P[\boldsymbol{w}] = P\big[P[\boldsymbol{v}]\big] = P^2[\boldsymbol{v}] = P[\boldsymbol{v}] = \boldsymbol{w}$. Thus, the restriction of $P$ to $\operatorname{Im}(P)$ is the identity operator.

*ii.* Suppose that the vector $\boldsymbol{w}$ lies in the intersection $\operatorname{Im}(P) \cap \operatorname{Ker}(P)$. Part *i* shows that $P[\boldsymbol{w}] = \boldsymbol{w}$. Since $\boldsymbol{w}$ is in $\operatorname{Ker}(P)$, we also have $P[\boldsymbol{w}] = \boldsymbol{0}$. Therefore, we have $\boldsymbol{w} = \boldsymbol{0}$ and $\operatorname{Im}(P) \cap \operatorname{Ker}(P) = \{\boldsymbol{0}\}$.

*iii.* We first prove existence. For any vector $\boldsymbol{v}$ in $V$, consider the vectors $\boldsymbol{w} := P[\boldsymbol{v}]$ and $\boldsymbol{z} := \boldsymbol{v} - \boldsymbol{w}$. It follows that $\boldsymbol{w}$ lies in $\operatorname{Im}(P)$ and $\boldsymbol{w} + \boldsymbol{z} = \boldsymbol{w} + (\boldsymbol{v} - \boldsymbol{w}) = \boldsymbol{v}$. Using part *i*, linearity gives $P[\boldsymbol{z}] = P[\boldsymbol{v} - \boldsymbol{w}] = P[\boldsymbol{v}] - P[\boldsymbol{w}] = \boldsymbol{w} - \boldsymbol{w} = \boldsymbol{0}$, so $\boldsymbol{z}$ lies in $\operatorname{Ker}(P)$. Hence, the required expression is $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{z}$.

To prove uniqueness, suppose that we have $\boldsymbol{w} + \boldsymbol{z} = \widetilde{\boldsymbol{w}} + \widetilde{\boldsymbol{z}}$ where $\boldsymbol{w}, \widetilde{\boldsymbol{w}} \in \operatorname{Im}(P)$ and $\boldsymbol{z}, \widetilde{\boldsymbol{z}} \in \operatorname{Ker}(P)$. Since both $\operatorname{Im}(P)$ and $\operatorname{Ker}(P)$ are linear subspaces [3.1.2], it follows that $\boldsymbol{w} - \widetilde{\boldsymbol{w}} \in \operatorname{Im}(P)$ and $\widetilde{\boldsymbol{z}} - \boldsymbol{z} \in \operatorname{Ker}(P)$, so the vector $\boldsymbol{w} - \widetilde{\boldsymbol{w}} = \widetilde{\boldsymbol{z}} - \boldsymbol{z}$ lies in the intersection $\operatorname{Im}(P) \cap \operatorname{Ker}(P)$. Part *ii* implies that $\boldsymbol{w} - \widetilde{\boldsymbol{w}} = \boldsymbol{0}$ and

 

$\widetilde{z} - z = \mathbf{0}$ or $w = \widetilde{w}$ and $\widetilde{z} = z$. Thus, an expression $v = w + z$, where $w \in \operatorname{Im}(P)$ and $z \in \operatorname{Ker}(P)$, is unique. □

An inner product distinguishes a special class of projections.

**8.0.2 Definition.** Let $V$ be an inner product space. A projection $P\colon V \to V$ is ***orthogonal*** if the linear subspaces $\operatorname{Im}(P)$ and $\operatorname{Ker}(P)$ are orthogonal.

Saying that $\operatorname{Im}(P)$ and $\operatorname{Ker}(P)$ are orthogonal means that, for any vector $w$ in $\operatorname{Im}(P)$ and any vector $z$ in $\operatorname{Ker}(P)$, we have $\langle w, z \rangle = 0$.

**8.0.3 Lemma** (Characterization of orthogonal projections)**.** *Let $V$ be an inner product space. A projection $P\colon V \to V$ is orthogonal if and only if, for all vectors $u$ and $v$ in $V$, we have $\langle u, P[v] \rangle = \langle P[u], v \rangle$.*

*Proof.*
⇒: Suppose that the linear operator $P$ is an orthogonal projection. The properties of projections show that there exist unique vectors $x$ and $w$ in $\operatorname{Im}(P)$ and unique vectors $y$ and $z$ in $\operatorname{Ker}(P)$ such that $u = x + y$ and $v = w + z$. The orthogonality of the linear operator $P$ implies that $\langle y, w \rangle = 0$ and $\langle x, z \rangle = 0$. It follows that $\langle u, P[v] \rangle = \langle x, w \rangle = \langle P[u], v \rangle$ because

$$\langle u, P[v] \rangle = \langle x + y, P[w + z] \rangle = \langle x + y, P[w] + P[z] \rangle = \langle x + y, w + \mathbf{0} \rangle = \langle x, w \rangle + \langle y, w \rangle = \langle x, w \rangle \,,$$
$$\langle P[u], v \rangle = \langle P[x + y], w + z \rangle = \langle P[x] + P[y], w] + z \rangle = \langle x + \mathbf{0}, w + z \rangle = \langle x, w \rangle + \langle x, z \rangle = \langle x, w \rangle \,.$$

⇐: For all vectors $u$ and $v$ in $V$, suppose that $\langle u, P[v] \rangle = \langle P[u], v \rangle$. For any vector $w$ in $\operatorname{Im}(P)$ and any vector $z$ in $\operatorname{Ker}(P)$, there exists a vector $v$ such that $w = P[v]$ and $P[z] = \mathbf{0}$. The properties of an inner product [7.0.9] show that

$$\langle w, z \rangle = \langle P[v], z \rangle = \langle v, P[z] \rangle = \langle v, \mathbf{0} \rangle = \mathbf{0} \,,$$

so the linear subspaces $\operatorname{Im}(P)$ and $\operatorname{Ker}(P)$ are orthogonal. □

Orthogonal projections have an elegant description.

**8.0.4 Proposition** (Projection formula)**.** *Let $U$ be a finite-dimensional linear subspace in an inner product space $V$. For any orthonormal basis $u_1, u_2, \dots, u_m$ of the linear subspace $U$, the unique orthogonal projection $P\colon V \to V$ satisfying $\operatorname{Im}(P) = U$ is defined, for all vectors $v$ in $V$, by $P[v] \coloneqq \langle v, u_1 \rangle\, u_1 + \langle v, u_2 \rangle\, u_2 + \cdots + \langle v, u_m \rangle\, u_m$.*

*Proof.* For any vector $v$ in $V$, the properties of projections show that there exists unique vectors $w$ in $\operatorname{Im}(P)$ and $z$ in $\operatorname{Ker}(P)$ such that $v = w + z$ and $P[v] = P[w + z] = P[w] + P[z] = w + \mathbf{0} = w$. Orthonormal coordinates [7.1.7] relative to the basis $u_1, u_2, \cdots, u_m$ establish that $P[v] = w = \langle w, u_1 \rangle\, u_1 + \langle w, u_2 \rangle\, u_2 + \cdots + \langle w, u_m \rangle\, u_m$. Since the vectors $u_1, u_2, \dots, u_m$ are linearly independent, we see that

$$\begin{aligned}
\mathbf{0} &= \big( \langle v, u_1 \rangle\, u_1 + \langle v, u_2 \rangle\, u_2 + \cdots + \langle v, u_m \rangle\, u_m \big) - P[v] \\
&= \langle v - w, u_1 \rangle\, u_1 + \langle v - w, u_2 \rangle\, u_2 + \cdots + \langle v - w, u_m \rangle\, u_m \\
&= \langle z, u_1 \rangle\, u_1 + \langle z, u_2 \rangle\, u_2 + \cdots + \langle z, u_m \rangle\, u_m
\end{aligned}$$

if and only if, for all $1 \leqslant k \leqslant m$, we have $\langle \boldsymbol{z}, \boldsymbol{u}_k \rangle = 0$. Thus, to have $P[\boldsymbol{v}] = \langle \boldsymbol{v}, \boldsymbol{u}_1 \rangle \boldsymbol{u}_1 + \langle \boldsymbol{v}, \boldsymbol{u}_2 \rangle \boldsymbol{u}_2 + \cdots + \langle \boldsymbol{v}, \boldsymbol{u}_m \rangle \boldsymbol{u}_m$ for any vector $\boldsymbol{v}$ in $V$, it both necessary and sufficient that each vector $\boldsymbol{z}$ in $\operatorname{Ker}(P)$ be orthogonal to the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m$. Equivalently, the linear subspace $\operatorname{Ker}(P)$ must be orthogonal to linear subspace $U$. For any vector $\boldsymbol{v}$ in $V$, setting $P[\boldsymbol{v}] = \langle \boldsymbol{v}, \boldsymbol{u}_1 \rangle \boldsymbol{u}_1 + \langle \boldsymbol{v}, \boldsymbol{u}_2 \rangle \boldsymbol{u}_2 + \cdots + \langle \boldsymbol{v}, \boldsymbol{u}_m \rangle \boldsymbol{u}_m$ implies that $P[\boldsymbol{u}_k] = \boldsymbol{u}_k$ for all $1 \leqslant k \leqslant m$, so $\operatorname{Im}(P) = U$. □

*Exercises*

**8.0.5 Problem.** Determine which of the following statements are true. If a statement is false, then provide a counterexample.

*i.* Every linear operator is a projection.

*ii.* A linear operator $T$ that satisfies $\operatorname{Im}(T) \cap \operatorname{Ker}(T) = \{\boldsymbol{0}\}$ must be a projection.

*iii.* Given any linear subspace $W$ in an inner product space $V$, there is a unique projection $P$ on $V$ such that $\operatorname{Im}(P) = W$.

*iv.* Given any linear subspace $W$ in an inner product space $V$, there is a unique orthogonal projection $P$ on $V$ such that $\operatorname{Im}(P) = W$.

**8.0.6 Problem.** Determine all of the eigenvalues of a projection and describe the corresponding eigenspaces.

## 8.1 Approximate Solutions

WHAT IS THE BEST APPROXIMATE SOLUTION TO A LINEAR SYSTEM? For an inconsistent linear system, it is fruitful to find an approximate solution. The next theorem demonstrates that orthogonal projections play a pivotal role in locating optimal approximations.

**8.1.0 Theorem** (Orthogonal projections minimize norms)**.** *Let $V$ be an inner product space and let $P \colon V \to V$ be an orthogonal projection. For any vector $\boldsymbol{v}$ in $V$ and any vector $\boldsymbol{u}$ in $\operatorname{Im}(P)$, we have $\|\boldsymbol{v} - P[\boldsymbol{v}]\| \leqslant \|\boldsymbol{v} - \boldsymbol{u}\|$. Furthermore, we have $\|\boldsymbol{v} - P[\boldsymbol{v}]\| = \|\boldsymbol{v} - \boldsymbol{u}\|$ if and only if $\boldsymbol{u} = P[\boldsymbol{v}]$.*

*Proof.* For any vector $\boldsymbol{v}$ in $V$, the properties of projections [8.0.1] show that there exist unique vectors $\boldsymbol{w}$ in $\operatorname{Im}(P)$ and $\boldsymbol{z}$ in $\operatorname{Ker}(P)$ such that $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{z}$ and $P[\boldsymbol{v}] = P[\boldsymbol{w}] + P[\boldsymbol{z}] = \boldsymbol{w} + \boldsymbol{0} = \boldsymbol{w}$. The difference $\boldsymbol{w} - \boldsymbol{u}$ is also the image $\operatorname{Im}(P)$, because both $\boldsymbol{u}$ and $\boldsymbol{w}$ lie in $\operatorname{Im}(P)$ and $\operatorname{Im}(P)$ is a linear subspace [3.1.2]. The orthogonality [8.0.2] of the projection $P$ asserts that the linear subspaces $\operatorname{Im}(P)$ and $\operatorname{Ker}(P)$ are orthogonal, so $\langle \boldsymbol{z}, \boldsymbol{w} - \boldsymbol{u} \rangle = 0$. Hence, the nonnegativity of inner products [7.0.0] and the Pythagorean theorem [7.1.2] yield

$$\|\boldsymbol{v} - P[\boldsymbol{v}]\|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2 = \|\boldsymbol{z}\|^2 \leqslant \|\boldsymbol{z}\|^2 + \|\boldsymbol{w} - \boldsymbol{u}\|^2 = \|\boldsymbol{z} + \boldsymbol{w} - \boldsymbol{u}\|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2 ,$$

Taking square roots, we obtain $\|\boldsymbol{v} - P[\boldsymbol{v}]\| \leqslant \|\boldsymbol{v} - \boldsymbol{u}\|$. Equality holds if and only if $0 = \|\boldsymbol{w} - \boldsymbol{u}\| = \|P[\boldsymbol{v}] - \boldsymbol{u}\|$ which, by the properties of norms [7.1.1], is equivalent to $P[\boldsymbol{v}] = \boldsymbol{u}$. □

**8.1.1 Remark.** Let $T\colon V \to V$ be linear operator on an inner product space $V$. For any vector $\boldsymbol{b}$ in $V$, consider the equation $T[\boldsymbol{x}] = \boldsymbol{b}$. When the vector $\boldsymbol{b}$ lies in the image $\operatorname{Im}(T)$, this equation has a solution: there exists $\boldsymbol{v} \in V$ such that $T[\boldsymbol{v}] = \boldsymbol{b}$. When the vector $\boldsymbol{b}$ does not lie in the image $\operatorname{Im}(T)$, the best approximate solution is a vector $\boldsymbol{v}$ in $V$ minimizing $\|\boldsymbol{b} - T[\boldsymbol{v}]\|$. Since orthogonal projections minimize norms, the optimal approximation is a vector $\boldsymbol{v}$ in $V$ such that the vector $T[\boldsymbol{v}]$ equals the orthogonal projection of $\boldsymbol{b}$ onto the image of $T$.

This method of least squares is a standard approach used to approximate the solution of overdetermined systems.

**8.1.2 Problem.** Let $V$ be the $\mathbb{R}$-vector space of continuous functions over the interval $[-1,1] \subset \mathbb{R}$ with $\langle f, g\rangle := \int_{-1}^{1} f(s)\,g(s)\; ds$. Find the quadratic polynomial $g(x)$ that is the best approximation to the function $f(x) = e^x$ over the interval $[-1,1]$.

The best approximation is not the quadratic Taylor polynomial for $e^x$.

*Solution.* Problem 7.2.3 establishes that $\frac{1}{\sqrt{2}}, \frac{\sqrt{3}}{\sqrt{2}}x, \frac{\sqrt{5}}{2\sqrt{2}}(3x^2-1)$ form an orthonormal basis for the linear subspace $W$ of the inner product space $V$ consisting of polynomials of degree at most 2. Since the best approximation is given by the orthogonal projection onto $W$, projection formula [8.0.4] gives

$$\begin{aligned}
&\tfrac{1}{2}\,\langle f, 1\rangle\, 1 + \tfrac{3}{2}\,\langle f, x\rangle\, x + \tfrac{5}{8}\,\langle f, 3x^2-1\rangle\,(3x^2-1)\\
&= \tfrac{1}{2}\left(\int_{-1}^{1} e^s\; ds\right) 1 + \tfrac{3}{2}\left(\int_{-1}^{1} s\,e^s\; ds\right) x + \tfrac{5}{8}\left(\int_{-1}^{1} (3\,s^2-1)\,e^s\; ds\right)(3\,x^2-1)\\
&= \tfrac{1}{2}\left(\big[e^s\big]_{-1}^{1}\right) 1 + \tfrac{3}{2}\left(\big[(s-1)\,e^s\big]_{-1}^{1}\right) x + \tfrac{5}{8}\left(\big[(3\,s^2-6\,s+5)\,e^s\big]_{-1}^{1}\right)(3x^2-1)\\
&= \tfrac{1}{2}(e - e^{-1}) + 3\,e^{-1}\,x + \tfrac{5}{8}(2\,e - 14\,e^{-1})(3x^2-1)\\
&= \big(\tfrac{1}{4}(7\,e - 33\,e^{-1})\big)1 + (3\,e^{-1})x + \big(\tfrac{15}{4}(e - 7\,e^{-1})\big)x^2\,. \qquad \square
\end{aligned}$$

**8.1.3 Problem.** Consider the inner product space of trigonometric polynomials having degree at most $n$ with $\langle f, g\rangle := \frac{1}{\pi}\int_{-\pi}^{\pi} f(s)\,g(s)\; ds$. Prove that $\big(\frac{1}{\sqrt{2}}, \cos(x), \sin(x), \cos(2x), \sin(2x), \ldots, \cos(nx), \sin(nx)\big)$ is an orthonormal basis for this vector space.

Fourier series are the least-squares approximations of periodic functions in terms of (typically infinite) sums of sines and cosines.

*Solution.* For all nonnegative integer $j$ and all positive integers $k$ such that $j \neq k$, using integration by parts twice gives

$$\begin{aligned}
\int_{-\pi}^{\pi} \cos(js)\cos(ks)\; ds &= \left[\tfrac{1}{k}\cos(js)\sin(ks)\right]_{s=-\pi}^{s=\pi} + \tfrac{j}{k}\int_{-\pi}^{\pi} \sin(js)\sin(ks)\; ds\\
&= \left[-\tfrac{j}{k^2}\sin(js)\cos(ks)\right]_{s=-\pi}^{s=\pi} + \tfrac{j^2}{k^2}\int_{-\pi}^{\pi} \cos(js)\cos(ks)\; ds\,,
\end{aligned}$$

It follows that $0 = \big((k^2-j^2)/(k^2)\big)\int_{-\pi}^{\pi}\cos(js)\cos(ks)\; ds$, so we deduce that $\langle \cos(jx), \cos(kx)\rangle = 0$. Similar calculations demonstrate that $\langle \cos(jx), \sin(kx)\rangle = 0$ and $\langle \sin(jx), \sin(kx)\rangle = 0$. As $1 = \cos(0x)$, this establishes the desired orthogonality.

It remains to see that the functions are appropriately normalized. Observe that $\langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{1}{2}\; ds = \frac{1}{2\pi}\big[s\big]_{s=-\pi}^{s=\pi} = 1$. For a positive integer $k$, integration by parts gives

$$\int_{-\pi}^{\pi}\cos^2(ks)\; ds = \left[\tfrac{1}{k}\cos(ks)\sin(ks)\right]_{s=-\pi}^{s=\pi} + \int_{-\pi}^{\pi}\sin^2(ks)\; ds = \int_{-\pi}^{\pi} 1 - \cos^2(ks)\; ds\,.$$

It follows that $2\int_{-\pi}^{\pi} \cos^2(ks)\ ds = \int_{-\pi}^{\pi} 1\ ds = \big[s\big]_{s=-\pi}^{s=\pi} = 2\pi$, so we see that $\langle \cos(kx), \cos(kx) \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2(ks)\ ds = 1$. A similar computation gives $\langle \sin(kx), \sin(kx) \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2(ks)\ ds = 1$. □

The ***sawtooth wave function*** $\operatorname{saw}(x)$ is piecewise linear: $\operatorname{saw}(x) = \frac{x}{\pi}$ for all $-\pi < x \leqslant \pi$; $\operatorname{saw}(x + 2j\pi) = \operatorname{saw}(x)$ for all $x \in \mathbb{R}$ and all $j \in \mathbb{Z}$.

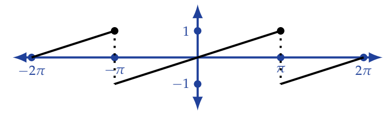


Figure 8.0: Graph of sawtooth wave

**8.1.4 Problem.** Find the trigonometric polynomial of degree at most $n$ that best approximates the function $\operatorname{saw}(x) = 2\big(\frac{x}{2\pi} - \big\lfloor \frac{1}{2} + \frac{x}{2\pi} \big\rfloor\big)$.

*Solution.* Problem 8.1.3 provides the relevant orthonormal basis. The best approximation is given by the orthogonal projection, which by the projection formula [8.0.4] is

$$\langle \operatorname{saw}(x), 1 \rangle \tfrac{1}{2} + \sum_{j=1}^{n} \langle \operatorname{saw}(x), \cos(jx) \rangle \cos(jx) + \sum_{k=1}^{n} \langle \operatorname{saw}(x), \sin(kx) \rangle \sin(kx)\,.$$

Since $\langle \operatorname{saw}(x), \cos(jx) \rangle = \frac{1}{\pi^2}\int_{-\pi}^{\pi} s\cos(js)\ ds = 0$ and

$$\begin{aligned} \langle \operatorname{saw}(x), \sin(kx) \rangle &= \tfrac{1}{\pi^2}\int_{-\pi}^{\pi} s\sin(ks)\ ds \\ &= \Big[-\tfrac{1}{k\pi^2} s\cos(ks)\Big]_{s=-\pi}^{s=\pi} + \tfrac{1}{k\pi^2}\int_{-\pi}^{\pi} \cos(ks)\ ds \\ &= -\frac{2\cos(k\pi)}{k\pi} + \tfrac{1}{k^2\pi^2}\big[\sin(ks)\big]_{s=-\pi}^{s=\pi} = \frac{2(-1)^{k+1}}{k\pi}\,, \end{aligned}$$

the best approximation of $\operatorname{saw}(x)$ is $\frac{2}{\pi}\sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} \sin(kx)$. □

**8.1.5 Remark.** One can prove that, as $n \to \infty$, this approximation converges pointwise to $\operatorname{saw}(x)$.

*Exercises*

**8.1.6 Problem.** Determine which of the following statements are true. If a statement is false, then provide a counterexample.

*i.* Any projection onto a linear subspace in an inner product space minimizes norms.

*ii.* When the vector $\boldsymbol{v}$ is a solution to the linear equation $T[\boldsymbol{x}] = \boldsymbol{b}$, the orthogonal projection of $T[\boldsymbol{v}]$ onto the image of $T$ equals $\boldsymbol{b}$.

*iii.* Taylor polynomials always provide the best approximation to a function.

*iv.* The defining basis for the space of trigonometric polynomials is an orthonormal basis.

**8.1.7 Problem.** Fix a nonnegative integer $n$ and consider the $n$-element set $\mathcal{X} := \big\{\frac{2\pi\ell}{n} \in \mathbb{R} \;\big|\; 0 \leqslant \ell \leqslant n-1\big\}$. Let $V := \mathbb{C}^{\mathcal{X}}$ the complex inner product space, consisting of all functions from the finite set $\mathcal{X}$ of real numbers to $\mathbb{C}$ with the inner product

$$\langle f, g \rangle := \sum_{x \in \mathcal{X}} f(x)\,\overline{g(x)} = \sum_{\ell=0}^{n-1} f\big(\tfrac{2\pi\ell}{n}\big)\,\overline{g\big(\tfrac{2\pi\ell}{n}\big)}\,.$$