CHAPTER 3

Euclidean Space

We define the geometric concepts of length, distance, angle and perpendicularity for $\mathbb{R}^n$. This gives $\mathbb{R}^n$ the structure of an *Euclidean Space*.

3.1 Inner Product

We write a point $\mathbf{u} \in \mathbb{R}^n$ as a column vector, i.e. $1 \times n$ matrix.

Definition 3.1. The **inner product** of $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, i.e.

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

is given by

$$\mathbf{u} \cdot \mathbf{v} := \mathbf{u}^T \mathbf{v} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^n u_i v_i \in \mathbb{R}$$

Note. To avoid confusion, I will omit the dot for scalar multiplication: I use $c\mathbf{u}$ instead of $c \cdot \mathbf{u}$.

Some easily checked properties:

Theorem 3.2. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n, c \in \mathbb{R}$

(1) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

(2) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$

(3) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

(4) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ iff $\mathbf{u} = \mathbf{0}$.

More generally:

Definition 3.3. Any vector space V over $\mathbb{R}$ equipped with an inner product $V \times V \longrightarrow \mathbb{R}$ satisfying Theorem 3.2 is called an **inner product space**. When $V = \mathbb{R}^n$ it is called an **Euclidean space**.

Example 3.1 (Optional)**.** An example of inner product space that is *infinite dimensional*: Let $C[a,b]$ be the vector space of real-valued continuous function defined on a closed interval $[a,b] \subset \mathbb{R}$. Then for $f, g \in C[a,b]$,

$$f \cdot g := \int_a^b f(t)g(t)dt$$

gives an inner product on $C[a,b]$.

Remark. All the Definitions and Theorems below applies to inner product spaces.

Remark. When $K = \mathbb{C}$, the inner product involves the complex conjugate

$$\mathbf{u} \cdot \mathbf{v} := \mathbf{u}^* \mathbf{v} := \begin{pmatrix} \overline{u_1} & \cdots & \overline{u_n} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^n \overline{u_i} v_i \in \mathbb{C}$$

so that the last property (4) can hold. Also the third property have to be replaced by

(3*) $(\bar{c}\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

Properties 3.4. If $\mathbf{A} = (a_{ij})$ is an $m \times n$ matrix, then the matrix entries are given by

$$a_{ij} = \mathbf{e}'_i \cdot \mathbf{A}\mathbf{e}_j$$

where $\{\mathbf{e}_j\}$ is the standard basis for $\mathbb{R}^n$ and $\{\mathbf{e}'_i\}$ is the standard basis for $\mathbb{R}^m$.

Definition 3.5. The **norm** (or **length**) of $\mathbf{v}$ is the nonnegative scalar

$$\|\mathbf{v}\| := \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2+\cdots+v_n^2} \in \mathbb{R}_{\geq 0}$$

For $c \in \mathbb{R}$, we have $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$.

Definition 3.6. The vector $\mathbf{u}$ with unit length, i.e. $\|\mathbf{u}\| = 1$ is called a **unit vector**.

Given $\mathbf{v} \neq 0$, $\frac{1}{\|\mathbf{v}\|}\mathbf{v}$ has unit length and is called the **normalization** of $\mathbf{v}$

Example 3.2. $\mathbf{v} = \begin{pmatrix} 1 \\ -2 \\ 2 \\ 0 \end{pmatrix} \in \mathbb{R}^4$ has norm $\|\mathbf{v}\| = \sqrt{1^2+(-2)^2+2^2+0^2} = \sqrt{9} = 3$.

$$\frac{1}{3}\mathbf{v} = \begin{pmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{pmatrix}$$

is a unit vector.

Definition 3.7. The distance between $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is defined by

$$dist(\mathbf{u},\mathbf{v}) := \|\mathbf{u}-\mathbf{v}\|$$

Theorem 3.8 (Law of cosine)**.** The angle θ between $\mathbf{u}$ and $\mathbf{v}$ can be calculated by

$$\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$

When $\theta = 90°$, we have

Definition 3.9. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are **orthogonal** (or **perpendicular**) to each other if $\mathbf{u} \cdot \mathbf{v} = 0$.

Properties 3.10 (Pythagorean Theorem)**.** If $\mathbf{u} \cdot \mathbf{v} = 0$,

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Properties 3.11 (Cauchy-Schwarz Inequality)**.** For all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$|\mathbf{u} \cdot \mathbf{v}| \le \|\mathbf{u}\|\|\mathbf{v}\|$$

Properties 3.12 (Triangle Inequality)**.** For all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$$

Example 3.3. $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ are orthogonal to each other in $\mathbb{R}^2$.

Example 3.4. $\mathbf{0}$ is orthogonal to every vector in $\mathbb{R}^n$.

3.2 Orthogonal Basis

Definition 3.13. Let $S = \{\mathbf{u}_1, ..., \mathbf{u}_p\} \in \mathbb{R}^n$.

- S is called an **orthogonal set** if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for all $i \neq j$.
- If in addition S is a basis of $W \subset \mathbb{R}^n$, it is called an **orthogonal basis** for W.
- If in addition all vectors in S has unit norm, it is called an **orthonormal basis** for W.

Example 3.5. The standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \cdots \mathbf{e}_n\}$ for $\mathbb{R}^n$ is an orthonormal basis.

Example 3.6. The set $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$ is an orthogonal basis for $\mathbb{R}^2$.

Its rescaled version, the set $\left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\}$ is an orthonormal basis for $\mathbb{R}^2$.

Theorem 3.14. Let $\mathcal{B} = \{\mathbf{u}_1, ..., \mathbf{u}_p\}$ be an orthogonal basis for a subspace $W \subset \mathbb{R}^n$. Then for $\mathbf{x} \in W$ we can solve for the coordinates with respect to $\mathcal{B}$ explicitly as

$$[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

where

$$c_i = \frac{\mathbf{x} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}, \qquad i = 1, ..., p$$

3.3 Orthogonal Projection

Definition 3.15. Let $W \subset \mathbb{R}^n$ be a subspace. The **orthogonal complement** of W is the set

$$W^{\perp} := \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \text{ for every } \mathbf{w} \in W\}$$

Properties 3.16. We have the following properties:

- $W^{\perp}$ is a subspace of $\mathbb{R}^n$.
- If $L = W^{\perp}$, then $W = L^{\perp}$.
- $\mathbf{x} \in W^{\perp}$ iff $\mathbf{x}$ is orthogonal to every vector in a spanning set ot W.

Theorem 3.17. Let $A \in M_{m\times n}(\mathbb{R})$. Then

$$(\text{Row}A)^{\perp} = \text{Nul}A, \qquad (\text{Col}A)^{\perp} = \text{Nul}A^T$$

Definition 3.18. The orthogonal projection of $\mathbf{b}$ onto $\mathbf{u}$ is given by

$$\text{Proj}_{\mathbf{u}}(\mathbf{b}) := (\mathbf{b} \cdot \mathbf{e})\mathbf{e} = \frac{\mathbf{b} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$$

where $\mathbf{e} := \frac{\mathbf{u}}{\|\mathbf{u}\|}$ is the normalization of $\mathbf{u}$.

Theorem 3.19 (Orthogonal Decomposition Theorem)**.** Let $W \subset \mathbb{R}^n$ a subspace. Then each $\mathbf{x} \in \mathbb{R}^n$ can be written *uniquely* in the form

$$\mathbf{x} = \widehat{\mathbf{x}} + \mathbf{z}$$

where $\widehat{\mathbf{x}} \in W$ and $\mathbf{z} \in W^\perp$. Therefore we have

$$\mathbb{R}^n = W \oplus W^\perp$$

We sometimes write $\mathrm{Proj}_W(\mathbf{x}) := \widehat{\mathbf{x}}$. Note that $\mathrm{Proj}_W \in L(\mathbb{R}^n, \mathbb{R}^n)$ with

$$\mathrm{Im}(\mathrm{Proj}_W) = W, \qquad \mathrm{Ker}(\mathrm{Proj}_W) = W^\perp$$

Proof. Explicitly, if $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthogonal basis of W, then

$$\begin{aligned}\widehat{\mathbf{x}} &= \mathrm{Proj}_{\mathbf{u}_1}(\mathbf{x}) + \cdots + \mathrm{Proj}_{\mathbf{u}_p}(\mathbf{x}) \\ &= \frac{\mathbf{x} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{x} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p\end{aligned}$$

and $\mathbf{z} = \mathbf{x} - \widehat{\mathbf{x}}$. ☐

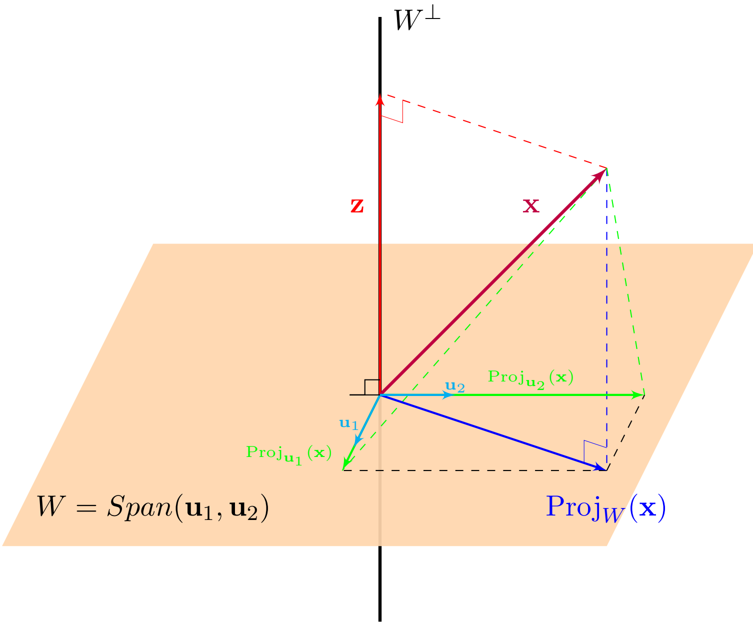

Figure 3.1: Orthogonal Projection

Remark. In particular, the uniqueness statement says that the orthogonal decomposition, i.e. the formula for $\widehat{\mathbf{x}}$, does not depend on the basis used for W in the proof.

Properties 3.20. If $\mathbf{x} \in W$, then $\text{Proj}_W(\mathbf{x}) = \mathbf{x}$.

By using orthonormal basis, we can represent Proj_W as a matrix:

Theorem 3.21. If $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthonormal basis for $W \subset \mathbb{R}^n$, then

$$\text{Proj}_W(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots (\mathbf{x} \cdot \mathbf{u}_p)\mathbf{u}_p$$

Equivalently, if $U = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{pmatrix}$ is an $n \times p$ matrix, then

$$\text{Proj}_W(\mathbf{x}) = UU^T\mathbf{x}$$

The matrix $P := UU^T$ is an $n \times n$ matrix which is called an **orthogonal projection matrix**.

Definition 3.22. A **projection matrix** is an $n \times n$ matrix such that

$$P^2 = P$$

It is an **orthogonal projection** matrix if in addition

$$P^T = P$$

Example 3.7. If $W = Span(\mathbf{v}_1, \mathbf{v}_2)$ where $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for W: $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

The normalization

$$\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}$$

is then an orthonormal basis for W. We have

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}$$

and therefore

$$\text{Proj}_W = UU^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} \frac{5}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{5}{6} \end{pmatrix}$$

Theorem 3.23 (Best Approximation Theorem)**.** Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$. Then

$$\|\mathbf{x} - \mathrm{Proj}_W \mathbf{x}\| \leq \|\mathbf{x} - \mathbf{v}\|, \qquad \text{for any } \mathbf{v} \in W$$

i.e. $\mathrm{Proj}_W \mathbf{x} \in W$ is the closest point in W to $\mathbf{x}$.

3.4 Orthogonal Matrix

Definition 3.24. A linear transformation $T \in L(V, W)$ between inner product spaces is called an **isometry** if it preserves the inner product:

$$(T\mathbf{u}) \cdot (T\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$$

for any vector $\mathbf{u}, \mathbf{v} \in V$.

Theorem 3.25. If $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ is a linear isometry which is represented by an $m \times n$ matrix $\mathbf{U}$, then for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

- $\mathbf{U}$ has orthonormal columns
- $\mathbf{U}^T\mathbf{U} = \mathbf{Id}_{n \times n}$
- $\|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|$ (i.e. it preserves length)
- $(\mathbf{U}\mathbf{x}) \cdot (\mathbf{U}\mathbf{y}) = 0$ iff $\mathbf{x} \cdot \mathbf{y} = 0$ (i.e. it preserves right angle)

Definition 3.26. If $n = m$, the square matrix $\mathbf{U}$ corresponding to a linear isometry is called an **orthogonal matrix**. It is invertible with

$$\mathbf{U}^{-1} = \mathbf{U}^T$$

The set of $n \times n$ orthogonal matrices is denoted by $O(n)$.

Properties 3.27. Orthogonal matrices satisfy the following *"group properties"*:

- $\mathbf{Id}_{n\times n} \in O(n)$.
- If $\mathbf{U} \in O(n)$, then $\mathbf{U}^{-1} \in O(n)$.
- If $\mathbf{U}, \mathbf{V} \in O(n)$, then $\mathbf{UV} \in O(n)$.

Example 3.8. In $\mathbb{R}^2$ and $\mathbb{R}^3$, an orthogonal matrix corresponds to combinations of rotations and mirror reflections.

In $\mathbb{R}^2$, all orthogonal matrix is of the form Rotations by angle θ counterclockwise

$$\mathbf{U} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Mirror reflections along the line with slope $\tan\frac{\theta}{2}$ passing through the origin

$$\mathbf{U} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

Example 3.9. The change-of-coordinate matrix $P^{\mathcal{B}}_{\mathcal{B}'}$ between orthonormal bases $\mathcal{B}$ and $\mathcal{B}'$ is an orthogonal matrix.

Non-Example 3.10. Projection Proj_W is in general not an orthogonal matrix: It does not preserve lengths.

3.5 Gram-Schmidt Process

Gram-Schmidt Process gives a simple algorithm to compute an orhtogonal basis from an arbitrary basis.

Theorem 3.28 (Gram-Schmidt Process)**.** Let $\{\mathbf{x}_1, ..., \mathbf{x}_p\}$ be a basis for a subspace $W \subset \mathbb{R}^n$. Define

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{x}_1 \\
\mathbf{v}_2 &:= \mathbf{x}_2 - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_2) \\
\mathbf{v}_3 &:= \mathbf{x}_3 - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_3) - \mathrm{Proj}_{\mathbf{v}_2}(\mathbf{x}_3) \\
&\vdots \\
\mathbf{v}_p &:= \mathbf{x}_p - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_p) - \mathrm{Proj}_{\mathbf{v}_2}(\mathbf{x}_3) - \cdots - \mathrm{Proj}_{\mathbf{v}_{p-1}}(\mathbf{x}_p)
\end{aligned}$$

where

$$\mathrm{Proj}_{\mathbf{v}}(\mathbf{x}) := \frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}$$

is the orthogonal projection (see Definition 3.18).

Then $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ is an orthogonal basis for W. Furthermore,

$$Span(\mathbf{v}_1, ..., \mathbf{v}_k) = Span(\mathbf{x}_1, ..., \mathbf{x}_k), \qquad \text{for all } 1 \leq k \leq p$$

To obtain an orthonormal basis, just normalize the vectors:

$$\mathbf{v}_i \rightsquigarrow \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$$

The Gram-Schmidt Process implies the following factorization, which is very important in computation algorithms, best linear approximations, and eigenvalues decomposition.

Theorem 3.29 (QR Decomposition)**.** If $\mathbf{A}$ is an $m \times n$ matrix with linearly independent columns, then

$$\mathbf{A} = \mathbf{QR}$$

where

- $\mathbf{Q}$ is a $m \times n$ matrix with orhtonormal columns forming a basis for Col$\mathbf{A}$,
- $\mathbf{R}$ is an $n \times n$ upper triangular invertible matrix with positive entries on the diagonal.

Example 3.11. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Let $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$ be the columns of $\mathbf{A}$.

Part 1: Gram-Schmidt Process.

Apply the formula, we obtain

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathrm{Proj}_{\mathbf{v}_1}\mathbf{x}_2 = \begin{pmatrix}0\\1\\1\\1\end{pmatrix} - \frac{3}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \frac{1}{4}\begin{pmatrix}-3\\1\\1\\1\end{pmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathrm{Proj}_{\mathbf{v}_1}\mathbf{x}_3 - \mathrm{Proj}_{\mathbf{v}_2}\mathbf{x}_3 = \begin{pmatrix}0\\0\\1\\1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{1}{6}\begin{pmatrix}-3\\1\\1\\1\end{pmatrix} = \frac{1}{3}\begin{pmatrix}0\\-2\\1\\1\end{pmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for $\mathrm{Col}\mathbf{A}$.

Part 2: Orthonormal basis and Q.

The vectors have lengths

$$\|\mathbf{v}_1\| = 2, \|\mathbf{v}_2\| = \frac{\sqrt{12}}{4}, \|\mathbf{v}_3\| = \frac{\sqrt{6}}{3},$$

hence the corresponding orthonormal basis is

$$\mathbf{v}_1' = \begin{pmatrix}1/2\\1/2\\1/2\\1/2\end{pmatrix}, \mathbf{v}_2' = \begin{pmatrix}-3/\sqrt{12}\\1/\sqrt{12}\\1/\sqrt{12}\\1/\sqrt{12}\end{pmatrix}, \mathbf{v}_3' = \begin{pmatrix}0\\-2/\sqrt{6}\\1/\sqrt{6}\\1/\sqrt{6}\end{pmatrix}$$

Then $\mathbf{Q}$ is formed by $\{\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_3'\}$, i.e.

$$\mathbf{Q} = \begin{pmatrix}1/2 & -3/\sqrt{12} & 0\\1/2 & 1/\sqrt{12} & -2/\sqrt{6}\\1/2 & 1/\sqrt{12} & 1/\sqrt{6}\\1/2 & 1/\sqrt{12} & 1/\sqrt{6}\end{pmatrix}$$

Part 3: The triangular matrix R.

To find $\mathbf{R}$, recall that $\mathbf{Q}$ is a linear isometry, hence

$$\mathbf{Q}^T\mathbf{A} = \mathbf{Q}^T(\mathbf{QR}) = \mathbf{Id}\cdot\mathbf{R} = \mathbf{R}$$

Therefore

$$\begin{aligned}\mathbf{R} &= \mathbf{Q}^T\mathbf{A}\\ &= \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}\\ &= \begin{pmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{pmatrix}\end{aligned}$$

Note that the diagonal of $\mathbf{R}$ is the same as the length of $\|\mathbf{v}_1\|, \|\mathbf{v}_2\|, \|\mathbf{v}_3\|$ used for the normalization in Part 2.

3.6 Least Square Approximation

If $\mathbf{A}$ is an $m \times n$ matrix with $m > n$, then the system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

is *overdetermined* and we may not have a solution.

If $\mathbf{A}$ has linear independent columns, we can use the QR decomposition to find a best approximation $\widehat{\mathbf{x}}$ to the solution of $\mathbf{A}\widehat{\mathbf{x}} = \mathbf{b}$ such that $\|\mathbf{A}\widehat{\mathbf{x}} - \mathbf{b}\|$ is the smallest.

By the Best Approximation Theorem (Theorem 3.23), the closest point $\mathbf{A}\mathbf{x} \in \mathrm{Col}\mathbf{A}$ to $\mathbf{b} \in \mathbb{R}^m$ should be $\mathrm{Proj}_{\mathrm{Col}\mathbf{A}}\mathbf{b}$. But $\mathrm{Col}\mathbf{A}$ has orthonormal basis given by columns of Q, so $\mathrm{Proj}_{\mathrm{Col}\mathbf{A}} = QQ^T$. Hence $\mathbf{A}\widehat{\mathbf{x}} = QQ^T\mathbf{b}$. Using $\mathbf{A} = QR$ we obtain:

Theorem 3.30 (Least Square Approximation)**.** If $\mathbf{A}$ is an $m \times n$ matrix with $m > n$ and has linear independent columns, such that we have the QR decomposition $\mathbf{A} = \mathbf{QR}$, then

$$\widehat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b} \in \mathbb{R}^n$$

is the vector such that for every $\mathbf{x} \in \mathbb{R}^n$,

$$\|\mathbf{A}\widehat{\mathbf{x}} - \mathbf{b}\| \leq \|\mathbf{A}\mathbf{x} - \mathbf{b}\|$$

Note. It is easier to solve for $\widehat{\mathbf{x}}$ by

$$R\widehat{\mathbf{x}} = \mathbf{Q}^T\mathbf{b}$$

instead of finding $\mathbf{R}^{-1}$, because $\mathbf{R}$ is upper triangular, so that we can use backward substitution.

Example 3.12. Continue our example, if $\mathbf{b} = \begin{pmatrix} 1\\2\\3\\4 \end{pmatrix}$, then the closest point $\widehat{\mathbf{x}}$ such that $\|\mathbf{Ax}-\mathbf{b}\|$ is smallest is given by

$$\mathbf{R}\widehat{\mathbf{x}} = \mathbf{Q}^T\mathbf{b}$$

$$\begin{pmatrix} 2 & 3/2 & 1\\ 0 & 3/\sqrt{12} & 2/\sqrt{12}\\ 0 & 0 & 2/\sqrt{6} \end{pmatrix} \begin{pmatrix} x_1\\x_2\\x_3 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2\\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12}\\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 1\\2\\3\\4 \end{pmatrix}$$

$$= \begin{pmatrix} 5\\ \sqrt{3}\\ 3/\sqrt{6} \end{pmatrix}$$

This is a very simple system of linear equations, and we can solve for $\widehat{\mathbf{x}}$ to get

$$\widehat{\mathbf{x}} = \begin{pmatrix} x_1\\x_2\\x_3 \end{pmatrix} = \begin{pmatrix} 2/3\\1\\3/2 \end{pmatrix}$$

Therefore

$$\mathbf{A}\widehat{\mathbf{x}} = \begin{pmatrix} 1&0&0\\1&1&0\\1&1&1\\1&1&1 \end{pmatrix} \begin{pmatrix} 2/3\\1\\3/2 \end{pmatrix} = \begin{pmatrix} 2/3\\5/3\\19/6\\19/6 \end{pmatrix}$$

is the closest approximation in Col$\mathbf{A}$ to $\begin{pmatrix} 1\\2\\3\\4 \end{pmatrix}$.

Note. It is called *"least square"* because $\|\mathbf{u}\|$ is computed by summing the squares of the coordinates. We want to find the smallest $\|\mathbf{Ax}-\mathbf{b}\|$. This is very useful in regression problems in statistics, where we want to fit the data onto a linear model as closely as possible.