Chapter 2

Orthogonality

Notation $\mathbb{F}, V$

- $\mathbb{F}$ *denotes* $\mathbb{R}$ *or* $\mathbb{C}$.
- V *denotes a vector space over* $\mathbb{F}$.

Learning objectives for this chapter:

- *Cauchy-Schwartz Inequality*
- *Gram-Schmidt Procedure*
- *Linear Functional on inner product spaces*
- *Calculating minimum distance to a subspace*

2.1 The inner product

The notion of inner product generalizes the notion of dot product of vectors in $\mathbb{R}^n$.

Definition 2.1.0.7. *Let* V *be a vector space. A function* $\langle , \rangle : V \times V \to \mathbb{R}$, *usually denoted* $\beta(x, y) = \langle x, y\rangle$, *is called an* ***inner product*** *on* V *if it positive, symmetric, and bilinear. That is, if*

i. $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geq 0$, *and* $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ *only for* $x = 0$ *(positivity)*

ii. $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle \overline{\boldsymbol{y}, \boldsymbol{x}}\rangle$ *(symmetry)*

iii. $\langle r\boldsymbol{u}, \boldsymbol{v}\rangle = r\langle \boldsymbol{u}, \boldsymbol{v}\rangle$ *for all scalars* r *and* $\langle v_1 + v_2, v\rangle = \langle v_1, v\rangle + \langle v_2, v\rangle$ *for all* $v_1, v_2, v, u \in V$. *(bilinear, that is linear (in both factors))*

iv. $\langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{z}\rangle + \langle \boldsymbol{y}, \boldsymbol{z}\rangle$ *(distributive law)*

v. ***positive*** *that is* $\langle \boldsymbol{v}, \boldsymbol{v}\rangle \geq 0$.

vi. ***non-degenerate*** *that is if* $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$ *for every* $v \in V$ *then* $u = 0$.

The vector space $\mathbb{R}^n$ *is an inner product space with respect to the usual dot product :*

$$\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \rangle = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i.$$

2.2 Inner product spaces

An inner product space is a vector space V with an inner product $\langle , \rangle$, $(V, \langle , \rangle)$. if V is a vector space over $\mathbb{R}$ (real vector space) it is called a Euclidean space and V is a vector space over $\mathbb{C}$, we call it is Unitary space.

Example 2.2.0.7. . $V = \mathbb{R}^n$.

2.2.1 Cauchy-Schwartz and triangular inequalities

Theorem 2.2.1.1. *(Cauchy-Schewartz inequality)*
Let V be an inner product space. Then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}||.||\mathbf{v}||$$

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

Proof. Let us consider the real case first. If $y = 0$, the statement is trivial, so we can assume that $y \neq 0$. By the properties of an inner product, for all scalar t

$$0 \leq ||\mathbf{x} - t\mathbf{y}||^2 = (\mathbf{x} - t\mathbf{y}, \mathbf{x} - t\mathbf{y}) = ||\mathbf{x}||^2 - 2t(\mathbf{x}, \mathbf{y}) + t^2||\mathbf{y}||^2.$$

In particular, this inequality should hold for

$$t = \frac{(\mathbf{x}, \mathbf{y})}{||\mathbf{y}||^2},$$

and for this point the inequality becomes

$$0 \leq ||\mathbf{x}||^2 - 2\frac{(\mathbf{x}, \mathbf{y})^2}{||\mathbf{y}||^2} + \frac{(\mathbf{x}, \mathbf{y})^2}{||\mathbf{y}||^2} = ||\mathbf{x}||^2 - \frac{(\mathbf{x}, \mathbf{y})^2}{||\mathbf{y}||^2},$$

which is exactly the inequality we need to prove. There are several possible ways to treat the complex case. One is to replace $\mathbf{x}$ by $\alpha\mathbf{x}$, where α is a complex constant, $\mid \alpha \mid= 1$ such that $(\alpha\mathbf{x}, \mathbf{y})$ is real, and then repeat the proof for the real case. The other possibility is again to consider

$$0 \leq ||\mathbf{x} - t\mathbf{y}||^2 = (\mathbf{x} - t\mathbf{y}, \mathbf{x} - t\mathbf{y}) = (\mathbf{x}, \mathbf{x} - t\mathbf{y}) - t(\mathbf{y}, \mathbf{x} - t\mathbf{y}) = ||\mathbf{x}||^- t(\mathbf{y}, \mathbf{x}) - \bar{t}(\mathbf{x} - \mathbf{y}) + |t|^2||\mathbf{y}||^2.$$

Substituting $t = \frac{(\mathbf{x},\mathbf{y})}{||\mathbf{y}||^2} = \frac{(\mathbf{y},\mathbf{x})}{||\mathbf{y}||^2}$ into this inequality, we get

$$0 \leq ||\mathbf{x}||^2 - \frac{|(\mathbf{x}, \mathbf{y})|^2}{||\mathbf{y}||^2}$$

which is the inequality we need. Note, that the above paragraph is in fact a complete formal proof of the theorem. The reasoning before that was only to explain why do we need to pick this particular value of t.
Equivalently, consider the function

$$y = y(t) := \langle \mathbf{u} + \mathbf{tv}, \mathbf{u} + \mathbf{tv} \rangle,\ t \in \mathbb{R}.$$

Then $y(t) \geq 0$ by the third property of inner product. Note that $y(t)$ is a quadratic function of t. In fact ,

$$y(t) = \langle \mathbf{u}, \mathbf{u} + \mathbf{tv} \rangle + \langle \mathbf{tv}, \mathbf{u} + \mathbf{tv} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle t + \langle \mathbf{v}, \mathbf{v} \rangle t^2.$$

Thus the quadratic equation

$$\langle \mathbf{u}, \mathbf{u}\rangle + 2\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle t^2 = 0$$

has at most one solution as $y(t) \geq 0$. This implies that its discriminant must be less than or equal to zero, i.e.,

$$(2\langle \mathbf{u}, \mathbf{v}\rangle)^2 - 4\langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle \leq 0.$$

□

Theorem 2.2.1.2. *(Triangle Inequality) Let V be an inner product space. Then*

$$||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$$

for all $\mathbf{x}, \mathbf{y} \in V$.

Proof.

$$\begin{aligned} ||\mathbf{x} + \mathbf{y}||^2 = (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) =& ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + (\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{x}) \\ &\leq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2|(\mathbf{x}, \mathbf{y})| \\ &\leq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2||\mathbf{x}||.||\mathbf{y}|| = (||\mathbf{x}|| + ||\mathbf{y}||)^2. \end{aligned}$$

The following polarization identities allow one to reconstruct the inner product from the norm: □

Lemma 2.2.1.1. *(Polarization identities). For $\mathbf{x}, \mathbf{y} \in V$*

$$(\mathbf{x}, \mathbf{y}) = \frac{1}{4}(||\mathbf{x} + \mathbf{y}||^2 - ||\mathbf{x} - \mathbf{y}||^2)$$

if V is a real inner product space, and

$$(\mathbf{x}, \mathbf{y}) = \frac{1}{4}(\sum_{\alpha=\pm 1, \pm i} \alpha||\mathbf{x} + \alpha\mathbf{y}||^2)$$

if V is a complex space.

Lemma 2.2.1.2. *(Parallelogram Identity). For any vectors $\mathbf{u}, \mathbf{v}$*

$$||\mathbf{u} + \mathbf{v}||^2 - ||\mathbf{u} - \mathbf{v}||^2 = 2(||\mathbf{u}||^2 + ||\mathbf{v}||^2).$$

In 2-dimensional space this lemma relates sides of a parallelogram with its diagonals, which explains the name. It is a well-known fact from planar geometry.

Normed Vector Spaces

Definition 2.2.1.1. *Let V be a real or complex vector space. Suppose to each $\mathbf{v} \in V$ there is assigned real number, denoted by $||\mathbf{v}||$. This function $||.||$ is called a norm on V if it satisfies the following axioms:*

- $||\mathbf{v}|| \geq 0$; *and* $||\mathbf{v}|| = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$.
- $||k\mathbf{v}|| = |k|||\mathbf{v}||$.
- $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$.

Then the vector space with a norm $(V, ||.||)$ is called ***normed vector space****.*

Norms on $\mathbb{R}^n$ and $\mathbb{C}^n$

Norm of Vector

$$||\mathbf{u}|| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

If $||\mathbf{u}|| = 1$ or, equivalently, if $\langle \mathbf{u}, \mathbf{u} \rangle = 1$, then $\mathbf{u}$ is called a unit vector and is said to be normalized. Every non-zero vector v in V can be multiplied by the reciprocal of its length to obtain the unit vector

$$\hat{\mathbf{v}} = (\frac{1}{||\mathbf{v}||})\mathbf{v}$$

which is a positive multiple of $\mathbf{v}$. This process is called normalizing $\mathbf{v}$. We have proved before that the norm $||\mathbf{v}||$ satisfies the following properties:

1. *. Homogeneity: $||\alpha\mathbf{v}|| = |\alpha|||\mathbf{v}||$ for all vectors $\mathbf{v}$ and all scalars α.*
2. *Triangle inequality: $||k\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}||| + ||\mathbf{v}||$.*
3. *Non-negativity: $||\mathbf{v}|| \geq 0$ for all vectors $\mathbf{v}$.*
4. *Non-degeneracy: $||\mathbf{v}|| = 0$ if and only if $\mathbf{v} = 0$.*

Suppose in a vector space V we assigned to each vector $\mathbf{v}$ a number $||\mathbf{v}||$ such that above properties 1-4 are satisfied. Then we say that the function

$$\mathbf{v} \mapsto ||\mathbf{v}||$$

*is a norm. A vector space V equipped with a norm is called a **normed space**. The following define three important norms on $\mathbb{R}^n$ and $\mathbb{C}^n$:*

$$|| < a_1, a_2, ..., a_n > ||_\infty = max\{|a_i|, \ \forall a_i \text{ coordinates of the vector}\}$$

$$|| < a_1, a_2, ..., a_n > ||_1 = |a_1| + ...|a_n|$$

$$|| < a_1, a_2, ..., a_n > ||_2 = \sqrt{|a_1|^2 + ... + |a_n|^2}.$$

Example 2.2.1.1. *Any inner product space is a normed space, because the norm $||\mathbf{v}|| = \sqrt{(\mathbf{v}, \mathbf{v}}$ satisfies the above properties 1-4. However, there are many other normed spaces.*

Example 2.2.1.2. *Consider vectors $< 1, -5, 3 >$ and $< 4, 2, -3 >$ in $\mathbb{R}^3$. Find*

a. *the infinity norm of the vectors.*

b. *the one-norm.*

c. *the two-norm.*

d. *d_∞, d_1, and d_2.*

A linear space with a norm such as:

$$\|x\|_p = \left(\sum_{i=1}^{\infty} |\xi_i|^p\right)^{\frac{1}{p}}$$ *$x = \{\xi_i\} \in \ell^p, p \neq 2$, is a normed space but not an inner product space, because this norm does not satisfy the parallelogram equality required of a norm to have an inner product associated with it.*

However, inner product spaces have a naturally defined norm based upon the inner product of the space itself that does satisfy the parallelogram equality:

$\|x\| = \sqrt{\langle x, x\rangle}$.

This is well defined by the nonnegativity axiom of the definition of inner product space. The norm is thought of as the length of the vector x. Directly from the axioms, we can prove the following:

Definition 2.2.1.2. *(Matrix Norms) Matrix norms are functions $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ that satisfy the same properties as vector norms. Let $A \in \mathbb{R}^{m\times n}$. Here are a few examples of matrix norms:*

- *The Frobenius norm:* $\| A \|_F = \sqrt{Tr(A^T A)} = \sqrt{\sum_{i,j} A_{i,j}^2}$
- *The sum-absolute-value norm:* $\| A \|_s\, av = \sum_{i,j} |X_{i,j}|$
- *The max-absolute-value norm:* $\| A \|_m\, av = max_{i,j}|A_{i,j}|$

Definition 2.2.1.3. *(Operator norm). An operator (or induced) matrix is a norm $\| \cdot \|_{a,b}$: $\mathbb{R}^{m\times n} \to \mathbb{R}$ defined as*

$$\| A \|_{a,b} = max_x \| Ax \|_a$$

$$s.t. \| x \|_b \leq 1,$$

where $\| \cdot \|_a$ is a vector norm on $\mathbb{R}^m$ and $\| \cdot \|_b$ is a vector norm on $\mathbb{R}^n$.

Notation:

When the same vector norm is used in both spaces, we write

$$\| A \|_c = max \| Ax \|_c$$

$$s.t. \| x \|_c \leq 1.$$

Example 2.2.1.3. *:*

- $\| A \|_2 = \sqrt{\lambda_m ax(A^T A)}$, *where $\lambda_m ax$ denotes the largest eigenvalue.*
- $\| A \|_1 = max_j \sum_i |A_{ij}|$, *i.e., the maximum column sum.*
- $\| A \|_\infty = max_j \sum_j |A_{ij}|$, *i.e., the maximum row sum.*

Notice that not all matrix norms are induced norms. An example is the frobenius norm given above as $\| I \|_* = 1$ for any induced norm, but $\| I \|_F = \sqrt{n}$.

Lemma 2.2.1.3. *Every induced norm is submultiplicative, i.e., $\| AB \| \leq \| A \| \| B \|$.*

Proof: We first show that $\| Ax \| \leq \| A \| \| x \|$.

Suppose that this is not the case, then

$$\| Ax \| > \| A \| \| x \|$$

$$\Rightarrow \frac{1}{\| x \|} \| Ax \| > \| A \|$$

$$\Rightarrow \| A.\frac{x}{\| x \|} \| > \| A \|$$

$\frac{x}{\|x\|}$ *is a unit vector.*

This contradicts the definition of $\| A \|$. *Now we need to prove the claim.*

$$\| AB \| = \max_{\|x\|\leq 1} \| ABx \| \leq \max_{\|x\|\leq 1} \| A \| \| Bx \| = \| A \| \max_{\|x\|\leq 1} \| Bx \| = \| A \| \| B \| .$$

Theorem 2.2.1.3. *A norm in a normed space is obtained from some inner product if and only if it satisfies the Parallelogram Identity*

$$||\mathbf{u}+\mathbf{v}||^2 + ||\mathbf{u}-\mathbf{v}||^2 = 2(||\mathbf{u}||^2 + ||\mathbf{v}||^2),\ \forall \mathbf{u},\ \mathbf{v} \in V.$$

Exercise 2.2.1.4. *1. Let* $D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 2 \end{pmatrix}$. *Consider the inner product defined on* $\mathbb{R}^3$ *by* $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T D \mathbf{v},\ \mathbf{u} = \begin{pmatrix} -2 \\ 3 \\ -7 \end{pmatrix},\ \mathbf{v} = \begin{pmatrix} -4 \\ 8 \\ 9 \end{pmatrix}$. *Verify that both inequalities hold.*

2.3 Orthogonal and Orthonormal sets

Definition 2.3.0.4. *:*
Let V *be an inner product space.*

- *We say that* $\mathbf{v},\ \mathbf{w}$ *are* ***orthogonal*** *if* $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.
- *A set* β *of vectors is orthogonal if every pair of vectors within it are orthogonal.*
- *A set* A *of vectors is orthonormal if it is orthogonal and every vector in* A *has unit norm.*

Thus the set $\beta = \{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}\}$ *is orthonormal if*

$$\langle v_i, v_j \rangle = \begin{cases} 0 & if\ i \neq j \\ 1 & if\ i = j \end{cases} .$$

An orthonormal basis for an inner product space V *is a basis which is itself an orthonormal set.*

Lemma 2.3.0.4. *: Let* V *be an inner product space. Suppose* $v \in V$ *is orthogonal to each of the vectors* $\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n} \in V$. *Then* v *is orthogonal to any linear combination of vectors* $\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}$.

Proof. Since $\langle \mathbf{v}, \mathbf{v_i} \rangle = 0$ for all $i \in \{1, 2, \cdots, n\}, if \alpha_1, \cdots, \alpha_n \in F$, we have

$$\langle \mathbf{v}, \sum_{i=1}^{n} \alpha_i \mathbf{v_i} \rangle = \sum_{i=1}^{n} \overline{\alpha_i} \langle \mathbf{v}, \mathbf{v_i} \rangle = 0.$$

□

Theorem 2.3.0.5. *(Generalized Pythagorean Theorem). Let* V *be an inner product space, and* $\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n} \in V$ *be mutually orthogonal vectors. Then*

$$\| \sum_{i=1}^{n} \mathbf{v_i} \|^2 = \sum_{i=1}^{n} \| \mathbf{v_i} \|^2$$

.

Proof. For simplicity, we assume $k = 2$. If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then

$$\begin{aligned} ||\mathbf{u}+\mathbf{v}||^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= ||\mathbf{u}||^2 + ||\mathbf{v}||^2. \end{aligned}$$

□

Example 2.3.0.4. *The three vectors*

$$\mathbf{u} = [1,2,1]^T, \ \mathbf{v} = [2,1,-4]^T, \ \mathbf{w} = [3,-2,1]^T$$

are mutually orthogonal. Express the vector $\mathbf{a} = [7,1,9]^T$ *as a linear combination of* $\mathbf{u}, \ \mathbf{v}, \ \mathbf{w}$. *Set*

$$x_1\mathbf{u} + x_2\mathbf{v} + x_3\mathbf{w} = \mathbf{a}.$$

There are two ways to find $x_1, \ x_2, \ x_3$.
Method 1: Solving the linear system by performing row operations to its augmented matrix

$$[\mathbf{v_1}, \ \mathbf{v_2}, \ \mathbf{v_3}|\mathbf{a}].$$

we obtain $x_1 = 3, x_2 = -1, x_3 = 2$. *So*

$$\mathbf{v} = 3\mathbf{v_1} - \mathbf{v_2} + 2\mathbf{v_3}.$$

Method 2: Since $\mathbf{v_i} \perp \mathbf{v_j}$ *for* $i \neq j$, *we have*

$$\langle \mathbf{a}, \mathbf{v_i} \rangle = \langle x_1\mathbf{v_1} + x_2\mathbf{v_2} + x_3\mathbf{v_3}, \mathbf{v_i} \rangle = x_i \langle \mathbf{v_i}, \mathbf{v_i} \rangle,$$

where $i = 1,2,3$. *Then*

$$x_i = \frac{\langle \mathbf{v}, \mathbf{v_i} \rangle}{\langle \mathbf{v_i}, \mathbf{v_i} \rangle}, \ i = 1,2,3$$

Then we have

$$\begin{aligned} x_1 &= \frac{7+2+9}{1+4+1} = \frac{18}{6} = 3, \\ x_2 &= \frac{14+1-36}{4+1+16} = \frac{-21}{21} = 1, \\ x_3 &= \frac{21-2+9}{9+4+1} = \frac{28}{14} = 2. \end{aligned}$$

Corollary 2.3.0.1. *Let* V *be an inner product space, and* $v_1, v_2, \cdots, v_n \in V$ *be orthogonal to each other. That is* $\langle v_i, v_j \rangle = 0$ *for all* $i, j \in \{1, 2, \cdots, n\}$ *with* $i \neq j, \alpha_1, \cdots, \alpha_n \in F$. *Then*

$$\| \sum_{i=1}^{n} \alpha_i v_i \|^2 = \sum_{i=1}^{n} |\alpha_i|^2 \| v_i \|^2 .$$

Corollary 2.3.0.2. *Let* V *be an inner product space and* $v_1, v_2, \cdots, v_n \in V$ *be an orthonormal set of vectors. Then*

$$\| \sum_{i=1}^{n} \alpha_i v_i \|^2 = \sum_{i=1}^{n} |\alpha_i|^2.$$

Corollary 2.3.0.3. *Any set of orthonormal vectors is linearly independent.*

2.3.1 Orthogonal complements

Definition 2.3.1.1. *Let V be an inner product space. If U is a subspace of V, the orthogonal complement to U is*

$$U^{\perp} = \{\mathbf{v} \in V | \langle \mathbf{v},\ \mathbf{u} \rangle = 0 \ for\ all\ \mathbf{u} \in U\}.$$

If $V = W \oplus U$ is a direct sum of subspaces with $W \subset U^{\perp}$, we write $b = b_W \perp b_U$ and say b is the the (internal) orthogonal sum of b_W and b_U. The subspace $V^{\perp}$ is called the radical of b and denoted by $rad(b)$. The form b is non-degenerate if and only if $rad(b) = 0$.

Lemma 2.3.1.1. *Let V be an inner product space and U be a subspace of V. Then*

(i) $U^{\perp}$ is a subspace of V, and

(ii) $U \cap U^{\perp} = \{0\}$.

Example 2.3.1.1. *Consider a line $L = \{x, 0, 0) |\ x \in \mathbb{R}\}$ and plane $\Pi = \{(0, y, z)\ y, z \in \mathbb{R}\}$ in $\mathbb{R}^3$. Then $L^{\perp} = \Pi$ and $\Pi^{\perp} = L$.*

Theorem 2.3.1.1. *Let V be a finite-dimensional inner product space and U be a subspace of V. Then*

$$V = U \oplus U^{\perp}.$$

We can consider the projection map $P_U : V \to V$ onto U associated to the decomposition $V = U \oplus U^{\perp}$. This is given by

$$P_U(\mathbf{v}) = u$$

where $\mathbf{v} = \mathbf{u} + \mathbf{w}$ is the unique decomposition of $\mathbf{v}$ with $\mathbf{u} \in U$ and $\mathbf{w} \in U^{\perp}$.

Theorem 2.3.1.2. *Let V be a finite-dimensional inner product space and U be a subspace of V. Let $P_U : V \to V$ be the projection map onto U associated to the direct sum decomposition $V = U \oplus U^{\perp}$. If $\mathbf{v} \in V$, then $P_U(\mathbf{v})$ is the vector in U closest to $\mathbf{v}$.*

Proposition 2.3.1.1. *Let S be a nonempty subset of an inner product space V. Then the orthogonal complement $S^{\perp}$ is a subspace of V.*

Proof. To show that $S^{\perp}$ is a subspace. We need to show that $S^{\perp}$ is closed under addition and scalar multiplication. Let $\mathbf{u},\ \mathbf{v} \in S^{\perp}$ and $c \in \mathbb{R}$. Since $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ and $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{w} \in S$, then

$$\mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle = 0,$$

$$\langle c\mathbf{u}, \mathbf{w} \rangle = c \langle \mathbf{u}, \mathbf{w} \rangle = 0$$

for all $w \in S$. So $\mathbf{u} + \mathbf{v}, c\mathbf{u}S^{\perp}$. Hence $S^{\perp}$ is a subspace of $\mathbb{R}^n$. ☐

Proposition 2.3.1.2. *Let S be a subset of an inner product space V. Then every vector of $S^{\perp}$ is orthogonal to every vector of $Span(S)$, i.e.,*

$$\langle \mathbf{u}, \mathbf{v} \rangle \ = 0, \forall \mathbf{u} \in Span(S),\ \mathbf{v} \in S^{\perp}.$$

Proof. For any $\mathbf{u} \in Span(S)$, the vector $\mathbf{u}$ must be a linear combination of some vectors in S, say,

$$\mathbf{u} = a_1\mathbf{u_1} + a_2\mathbf{u_2} + ... + a_k\mathbf{u_k}.$$

Then for any $\mathbf{v} \in S^{\perp}$,

$$\langle \mathbf{u}, \mathbf{v} \rangle = a_1 \langle \mathbf{u_1}, \mathbf{v} \rangle + ... + a_n \langle \mathbf{u_n}, \mathbf{v} \rangle = 0.$$

☐

Theorem 2.3.1.3. *(Orthogonal sets are linearly independent)*

If $S = \{v_1, v_2, \cdots, v_n\}$ is an orthogonal set of nonzero vectors in an inner product space V, then S is linealry independent.

Proof:

S is an orthogonal set of nonzero vectors, i.e., $\langle v_i, v_j\rangle = 0, for i \neq j, and \langle v_i, v_\rangle > 0$
For

$$\sum_{i=1}^{n} c_i \boldsymbol{v}_i = \boldsymbol{0} \text{ (If there is only the trivial solution for } c_i's \text{ are} 0, S \text{ is linearly independent)}$$

$$\Rightarrow \langle c_i \sum_{i=1}^{n} \boldsymbol{v}_i, \boldsymbol{v}_i\rangle = \langle 0, \boldsymbol{v}_i = \boldsymbol{0}, \forall i$$

$$\Rightarrow c_1\langle \boldsymbol{v}_1, \boldsymbol{v}_i\rangle + c_2\langle \boldsymbol{v}_2, \boldsymbol{v}_i\rangle + \cdots + c_n\langle \boldsymbol{v}_n, \boldsymbol{v}_i\rangle = c_i\langle \boldsymbol{v}_i, \boldsymbol{v}_i\rangle = \boldsymbol{0} \text{ why?}$$

$$\because \langle \boldsymbol{v}_i, \boldsymbol{v}_i\rangle \neq 0 \Rightarrow c_i = 0 \forall i$$

$\therefore S$ *is linearly independent*

Theorem 2.3.1.4. *(Coordinate relative to an orthonormal basis)*

If $B = \{v_1, v_2, \cdots, v_n\}$ is an orthonormal basis for an inner product space V, then the unique coordinate representation of vector $\boldsymbol{w}$ with respect to B is

$$\boldsymbol{w} = \sum_{i=1}^{n} \langle \boldsymbol{w}, \boldsymbol{v}_i\rangle \boldsymbol{v}_i$$

The above theorem tells us that it is easy to derive the coordinate representation of a vector relative to an orthonormal basis, which is another advantage of using orthonormal bases.

Proof

$B = \{v_1, v_2, \cdots, v_n\}$ is an orthonormal basis for V, $\boldsymbol{w} = \sum_{i=1}^{n} k_i \boldsymbol{v}_i \in V$ since $\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$

$$\boldsymbol{w}, \boldsymbol{v}_i\rangle = \langle \sum_{i=1}^{n} k_i \boldsymbol{v}_i, \boldsymbol{v}_i\rangle$$

$$= k_1\langle \boldsymbol{v}_1, \boldsymbol{v}_i\rangle + k_2\langle \boldsymbol{v}_2, \boldsymbol{v}_i\rangle + \cdots + k_n\langle \boldsymbol{v}_n, \boldsymbol{v}_i\rangle$$

$$= k_i \quad for i = 1 \text{ to } n$$

$$\Rightarrow \boldsymbol{w} = \langle \boldsymbol{w}, \boldsymbol{v}_1\rangle \boldsymbol{v}_1 + \langle \boldsymbol{w}, \boldsymbol{v}_2\rangle \boldsymbol{v}_2 + \cdots + \langle \boldsymbol{w}, \boldsymbol{v}_n\rangle \boldsymbol{v}_n$$

Note If $B = \{v_1, v_2, \cdots, v_n\}$ is an orthonormal basis for V and $\boldsymbol{w} \in V$, then the corresponding coordinate matrix of $\boldsymbol{w}$ relative to B is $[\boldsymbol{W}]_B = \begin{pmatrix} \langle \boldsymbol{w}, \boldsymbol{v}_1\rangle \\ \langle \boldsymbol{w}, \boldsymbol{v}_2\rangle \\ \vdots \\ \langle \boldsymbol{w}, \boldsymbol{v}_n\rangle \end{pmatrix}$

Exercise 2.3.1.5. *:*

1. *Find all values of the scalar k for which the two vectors are orthogonal.*

   a. $\mathbf{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} k+1 \\ k-1 \end{bmatrix}$ *b.* $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} k^2 \\ k \\ -3 \end{bmatrix}$

2. *Describe all vectors that are orthogonal to*

   a. $\mathbf{u} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ *b.* $\mathbf{u} \begin{bmatrix} a \\ b \end{bmatrix}$

2.3.2 The gram Schmidt orthogonalization/orthonormalization process

Projections and Orthogonal Projection

Projections :
Let V be an inner product space, $\mathbf{w}$ is a given nonzero vector in V, $\mathbf{v}$ any vector. The projection of $\mathbf{v}$ along $\mathbf{w}$ is denoted and given by

$$proj_{\mathbf{v}}^{\mathbf{w}} = (\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{w}, \mathbf{w} \rangle}).\mathbf{w}.$$

Orthogonal projection *Let V be an inner product space. Let $\mathbf{v}$ be a nonzero vector of V . We want to decompose an arbitrary vector y into the form*

$$\mathbf{y} = \alpha\mathbf{v} + \mathbf{z}, \ \textit{where } \mathbf{z} \in \mathbf{v}^{\perp}.$$

Since $\mathbf{z} \perp \mathbf{v}$

$$\langle \mathbf{v}, \mathbf{y} \rangle = \langle \alpha\mathbf{v}, \mathbf{v} \rangle.$$

This implies that

$$\alpha = \frac{\langle \mathbf{v}, \mathbf{y} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

We define the vector

$$Proj_{\mathbf{v}}(\mathbf{y}) = \frac{\langle \mathbf{v}, \mathbf{y} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v},$$

called the orthogonal projection of $\mathbf{y}$ along $\mathbf{v}$. The linear transformation $Proj_{\mathbf{u}} : V \to V$ is called the orthogonal projection of V onto the direction $\mathbf{v}$.

Proposition 2.3.2.1. *Let $\mathbf{v}$ be a nonzero vector of the Euclidean n-space $\mathbb{R}^n$. Then the orthogonal projection $Proj_u : \mathbb{R}^n \to \mathbb{R}^n$ is given by*

$$Proj_{\mathbf{v}}(\mathbf{y}) = \left(\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}.\mathbf{v}}\right)\mathbf{y};$$

$Proj_{\mathbf{v}^{\perp}} : \mathbb{R}^n \to \mathbb{R}^n$ *is given by*

$$Proj_{\mathbf{v}}(\mathbf{y}) = \left(I - \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}.\mathbf{v}}\right)\mathbf{y}.$$

Write the vector $\mathbf{v}$ as $\mathbf{v} = [a_, a_2, ..., a_n]^T$. The for any scalar c,

$$c\mathbf{v} = \begin{bmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{bmatrix} = \begin{bmatrix} a_1c \\ a_2c \\ \vdots \\ a_nc \end{bmatrix} = \mathbf{v}[c],$$

where $[c]$ is the 1×1 matrix with the only entry c. Note that

$$[\mathbf{v}, \mathbf{y}] = \mathbf{v}^T\mathbf{y}.$$

Then the orthogonal projection $Proj_{\mathbf{v}}$ can be written as

$$\begin{aligned} Proj_{\mathbf{v}}(\mathbf{y}) &= \left(\tfrac{1}{\mathbf{v}.\mathbf{v}}\right)(\mathbf{v}.\mathbf{y})\mathbf{v} \\ &= \left(\tfrac{1}{\mathbf{v}.\mathbf{v}}\right)\mathbf{v}[\mathbf{v}.\mathbf{y}] \\ &= \left(\tfrac{1}{\mathbf{v}.\mathbf{v}}\right)\mathbf{v}\mathbf{v}^T\mathbf{y}. \end{aligned}$$

This means that the standard matrix of $Proj_{\mathbf{v}}$ is

$$\left(\tfrac{1}{\mathbf{v}.\mathbf{v}}\right)\mathbf{v}\mathbf{v}^T.$$

Indeed, $\mathbf{v}$ is an $n \times 1$ matrix and $\mathbf{v}^T$ is a $1 \times n$ matrix, the product $\mathbf{v}\mathbf{v}^T$ is an $n \times n$ matrix. The orthogonal projection $Proj_{\mathbf{v}^\perp} : \mathbb{R}^n \to \mathbb{R}^n$ is given by

$$Proj_{\mathbf{v}^\perp}(\mathbf{y}) = y - Proj_{\mathbf{v}}(\mathbf{y}) = \left(I - \tfrac{1}{\mathbf{v}.\mathbf{v}}\mathbf{v}\mathbf{v}^T\right)\mathbf{y}.$$

This means that the standard matrix of $Proj_{\mathbf{v}^\perp}$ is

$$\left(I - \tfrac{1}{\mathbf{v}.\mathbf{v}}\mathbf{v}\mathbf{v}^T\right).$$

Example 2.3.2.1. *Find the linear mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$ that is a the orthogonal projection of $\mathbb{R}^3$ onto the plane $x_1 + x_2 + x_3 = 0$.*

Solution

To find the orthogonal projection of $\mathbb{R}^3$ onto the subspace $\mathbf{v}^\perp$, where $\mathbf{v} = [1, 1, 1]^T$, we find the following orthogonal projection

$$\begin{aligned} Proj_{\mathbf{v}}(\mathbf{y}) = \left(\tfrac{\mathbf{v}.\mathbf{y}}{\mathbf{v}.\mathbf{v}}\right)\mathbf{v} &= \frac{y_1 + y_2 + y_3}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix}2 & -1 & -1\\-1 & 2 & -1\\-1 & -1 & 2\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}. \end{aligned}$$

Let W be a subspace of V, and let $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ be an orthogonal basis of W. We want to decompose an arbitrary vector $\mathbf{y} \in V$ into the form

$$\mathbf{y} = \mathbf{w} + \mathbf{z}$$

with $\mathbf{w} \in W$, $\mathbf{z} \in W^\perp$. Then there exist scalars $\alpha_1, ..., \alpha_k$ such that

$$\widehat{\mathbf{y}} = \alpha_1\mathbf{v_1} + ... + \alpha_k\mathbf{v_k}.$$

Since $\mathbf{z} \perp \mathbf{v_1}, \mathbf{z} \perp \mathbf{v_2}, ..., \mathbf{z} \perp \mathbf{v_k}$, we have

$$\langle \mathbf{v_i}, \mathbf{y}\rangle = \langle \mathbf{v_i}, \alpha_1\mathbf{v_1} + ... + \alpha_k\mathbf{v_k}\rangle = \alpha_i\langle \mathbf{v_i}, \mathbf{v_i}\rangle.$$

Then

$$\alpha_i = \frac{\langle \mathbf{v_i}, \mathbf{y}\rangle}{\langle \mathbf{v_i}, \mathbf{v_i}\rangle},\ 1 \le i \le k.$$

Thus we define

$$Proj_W(\mathbf{y}) = \sum_{i=1}^{k} \frac{\langle \mathbf{v_i}, \mathbf{y} \rangle}{\langle \mathbf{v_i}, \mathbf{v_i} \rangle} \mathbf{v_i},$$

called the orthogonal projection of $\mathbf{v}$ *along* W*. The linear transformation*

$$Proj_W : V \to V$$

is called the orthogonal projection of V *onto* W*.*

Definition 2.3.2.1. *An ordered basis which is orthonormal is called* ***orthonormal basis***

Theorem 2.3.2.1. *Suppose* $S = \{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_k}\}$ *is an orthogonal subset of* V *such that* $\mathbf{v_i} \neq 0$*. For* $y \in span(S)$*,*

$$y = \sum_{i=1}^{k} \left(\frac{\langle \mathbf{y}, \mathbf{v_i} \rangle}{||\mathbf{v_i}||^2}\right).\mathbf{v_i}.$$

Example 2.3.2.2. *Find the orthogonal projection*

$$Proj_W : \mathbb{R}^3 \to \mathbb{R}^3,$$

where W *is the plane*

$$x_1 + x_2 + x_3 = 0.$$

Solution:
By inspection, the following two vectors

$$\mathbf{v_1} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \text{ and } \mathbf{v_2} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

form an orthogonal basis of W*. Then*

$$\begin{aligned} Proj_W(\mathbf{y}) &= \left(\frac{\mathbf{v_1}.\mathbf{y}}{\mathbf{v_1}.\mathbf{v_1}}\right)\mathbf{v_1} + \left(\frac{\mathbf{v_2}.\mathbf{y}}{\mathbf{v_2}.\mathbf{v_2}}\right)\mathbf{v_2} \\ &= \frac{y_1 - y_2}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{y_1 + y_2 - 2y_3}{6} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \\ &= \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}. \end{aligned}$$

Example 2.3.2.3. *Find the matrix of the orthogonal projection*

$$Proj_W : \mathbb{R}^3 \to \mathbb{R}^3,$$

where

$$W = Span\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}.$$

Solution:
The following two vectors

$$\mathbf{u_1} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \ \mathbf{u_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ 0 \end{bmatrix}$$

form form an orthonormal basis of W. *Then the standard matrix of* $Proj_W$ *is the product*

$$\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix},$$

which results the matrix

$$\begin{bmatrix} \frac{5}{6} & \frac{-1}{6} & \frac{1}{3} \\ \frac{-1}{6} & \frac{5}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Alternatively, the matrix can be found by computing the orthogonal projection:

$$\begin{aligned} Proj_w(\mathbf{y}) &= \frac{y_1 + y_2 + y_3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{y_1 - y_2}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 5y_1 - y_2 + 2y_3 \\ -y_1 + 5y_2 + 2y_3 \\ 2y_1 + 2y_2 + 2y_3 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 5 & -1 & 2 \\ -1 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}. \end{aligned}$$

Gram-Schmidt process

Problem: *Given a (finite-dimensional) inner product space* V , *how do we find an orthonormal basis?*

Definition 2.3.2.2. *Gram–Schmidt Process: is the process of forming an orthogonal sequence* $\{\mathbf{v_k}\}$ *from a linearly independent sequence* $\{u_k\}$ *of members of an inner-product space.*

Theorem 2.3.2.2. ***(Gram-Schmidt algorithm)*** *Suppose that* V *is a finite-dimensional inner product space with basis* $S = \{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}\}$. *The following procedure constructs an an orthogonal and orthonormal basis* $S' = \{\mathbf{u_1}, \ \mathbf{u_2}, ..., \ \mathbf{u_k}\}$, $S'' = \{\mathbf{e_1}, \mathbf{e_2}, \cdots, \mathbf{e_n}\}$ *for* V *respectively.*

Step 1 *Set* $\mathbf{u_1} := \mathbf{v_1}$, $\mathbf{e_1} = (\frac{1}{||\mathbf{u_1}||}).\mathbf{u_1}$.

Step k *For* $k \geq 2$, $\mathbf{u_2} = \mathbf{v_2} - (\frac{\langle \mathbf{v_2}, \mathbf{u_1} \rangle}{||\mathbf{v_2}||^2}).\mathbf{u_1}$, $e_2 = (\frac{1}{||\mathbf{u_2}||}).\mathbf{u_2}$.

$$u_k = v_k - \sum_{j=1}^{k-1} \left(\frac{\langle \mathbf{v_k}, \mathbf{u_j} \rangle}{||\mathbf{u_j}||^2} \right).u_j, \ e_k = (\frac{1}{||\mathbf{u_k}||}).\mathbf{u_k}.$$

Then S' *is orthogonal and span(*S' *= span(s),* S'' *is orthonormal.*

Example 2.3.2.4. *By using gram-schmidt process of constructing basis of vector space, find the orthogonal and orthonormal basis of the vector space, with relative basis*

$$\{< 1, 2, 1 >, < 2, 1, -4 >, < 3, -2, 1 >\}.$$

Example 2.3.2.5. *Let V_0 be a subspace of dimension $\mathbf{K}$ in $\mathbb{R}^n$. Let $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}$ be a basis for V_0.*

i. *Find an orthogonal basis for V_0.*

ii. *Extend it to an orthogonal basis for $\mathbb{R}^n$.*

Approach 1. *Extend $\mathbf{v_1}, ..., \mathbf{v_k}$ to a basis $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n}$ for $\mathbb{R}^n$. Then apply the Gram-Schmidt process to the extended basis. We shall obtain an orthogonal basis $\mathbf{u_1}, ..., \mathbf{u_n}$ for $\mathbb{R}^n$.*
By construction, $Span(\mathbf{u_1}, ..., \mathbf{u_k}) = Span(\mathbf{v_1}, ..., \mathbf{v_k}) = V_0$.
It follows that $\mathbf{u_1}, ..., \mathbf{u_n}$ is a basis for V_0. Clearly, it is orthogonal.
Approach 2. *First apply the Gram-Schmidt process to $\mathbf{v_1}, ..., \mathbf{v_k}$ and obtain an orthogonal basis $\mathbf{u_1}, ..., \mathbf{u_k}$ for V_0.*
Secondly, find a basis $\mathbf{y_1}, ..., \mathbf{y_m}$ for the orthogonal complement $V_0^{\perp}$ and apply the Gram-Schmidt process to it obtaining an orthogonal basis $\mathbf{x_1}, ..., \mathbf{x_m}$ for $V_0^{\perp}$. Then $\mathbf{u_1}, ..., \mathbf{u_k}, \mathbf{x_1}, ..., \mathbf{x_m}$ is an orthogonal basis for $\mathbb{R}^n$.

Example 2.3.2.6. *Let π be the plane in $\mathbb{R}^3$ spanned by vectors $\mathbf{x_1} = (1, 2, 2)$ and $\mathbf{x_2} = (-1, 0, 2)$.*

(i) *Find an orthonormal basis for π.*

(ii) *Extend it to an orthonormal basis for $\mathbb{R}^3$.*

$\mathbf{x_1}, \mathbf{x_2}$ is a basis for the plane π. We can extend it to a basis for $\mathbb{R}^3$ by adding one vector from the standard basis. For instance, vectors $\mathbf{x_1}, \mathbf{x_2}$, and $\mathbf{x_3} = (0, 0, 1)$ form a basis for $\mathbb{R}^3$ because

$$\begin{vmatrix} 1 & 2 & 2 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ -1 & 0 \end{vmatrix} = 2 \neq 0.$$

Using the Gram-Schmidt process, we orthogonalize the basis $\mathbf{x_1} = (1, 2, 2), \mathbf{x_2} = (-1, 0, 2), \mathbf{x_3} = (0, 0, 1)$:

$$\begin{aligned}
\mathbf{v_1} = \mathbf{x_1} &= (1, 2, 2), \\
\mathbf{v_2} = \mathbf{x_2} - \frac{\langle \mathbf{x_2}, \mathbf{v_1} \rangle}{\langle \mathbf{v_1}, \mathbf{v_1} \rangle}\mathbf{v_1} &= (-1, 0, 2) - \frac{3}{9}(1, 2, 2) = (-4/3, -2/3, 4/3), \\
\mathbf{v_3} = \mathbf{x_3} - \frac{\langle \mathbf{x_3}, \mathbf{v_1} \rangle}{\langle \mathbf{v_1}, \mathbf{v_1} \rangle}\mathbf{v_1} - \frac{\langle \mathbf{x_3}, \mathbf{v_2} \rangle}{\langle \mathbf{v_2}, \mathbf{v_2} \rangle}\mathbf{v_2} &= (0, 0, 1) - \frac{2}{9}(1, 2, 2) - \frac{4/3}{4}(-4/3, -2/3, 4/3) = (2/9, -2/9, 1/9).
\end{aligned}$$

Now $\mathbf{v_1} = (1, 2, 2), \mathbf{v_2} = (-4/3, -2/3, 4/3), \mathbf{v_3} = (2/9, -2/9, 1/9)$ is an orthogonal basis for $\mathbb{R}^3$ while $\mathbf{v_1}, \mathbf{v_2}$ is an orthogonal basis for π. It remains to normalize these vectors.

$$\begin{aligned}
\langle \mathbf{v_1}, \mathbf{v_1} \rangle &= 9 \Rightarrow ||\mathbf{v_1}|| = 3, \\
\langle \mathbf{v_2}, \mathbf{v_2} \rangle &= 4 \Rightarrow ||\mathbf{v_2}|| = 2, \\
\langle \mathbf{v_3}, \mathbf{v_3} \rangle &= 1/9 \Rightarrow ||\mathbf{v_3}|| = 1/3.
\end{aligned}$$

$$\begin{aligned}
\mathbf{w_1} &= \mathbf{v_1}/||\mathbf{v_1}|| = (1/3, 2/3, 2/3) = \frac{1}{3}(1, 2, 2), \\
\mathbf{w_2} &= \mathbf{v_2}/||\mathbf{v_2}|| = (-2/3, -1/3, 2/3) = \frac{1}{3}(-2, -1, 2), \\
\mathbf{w_3} &= \mathbf{v_3}/||\mathbf{v_3}|| = (2/3, -2/3, 1/3) = \frac{1}{3}(2, -2, 1).
\end{aligned}$$

$\mathbf{w_1},\ \mathbf{w_2}$ *is an orthonormal basis for* π.
$\mathbf{w_1},\ \mathbf{w_2},\ \mathbf{w_3}$ *is an orthonormal basis for* $\mathbb{R}^3$.

Example 2.3.2.7. *Find the distance from the point* $y = (0,0,0,1)$ *to the subspace* $V \subset \mathbb{R}^4$ *spanned by vectors* $\mathbf{x_1} = (1,-1,1,-1),\ \mathbf{x_2} = (1,1,3,-1),$ *and* $\mathbf{x_3} = (-3,7,1,3)$.
Solution:
Let us apply the Gram-Schmidt process to vectors $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{y}$. *We should obtain an orthogonal system* $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}$. *The desired distance will be* $||\mathbf{v_4}||$.

$$\mathbf{x_1} = (1,-1,1,-1), \mathbf{x_2} = (1,1,3,-1), \mathbf{x_3} = (-3,7,1,3), \mathbf{y} = (0,0,0,1).$$

$$\begin{aligned}
\mathbf{v_1} &= \mathbf{x_1} = (1,-1,1,-1),\\
\mathbf{v_2} &= \mathbf{x_2} - \frac{\langle \mathbf{x_2}, \mathbf{v_1}\rangle}{\langle \mathbf{v_1}, \mathbf{v_1}\rangle}\mathbf{v_1}\\
&= (1,1,3,-1) - \frac{4}{4}(1,-1,1,-1) = (0,2,2,0),\\
\mathbf{v_3} &= \mathbf{x_3} - \frac{\langle \mathbf{x_3}, \mathbf{v_1}\rangle}{\langle \mathbf{v_1}, \mathbf{v_1}\rangle}\mathbf{v_1} - \frac{\langle \mathbf{x_3}, \mathbf{v_2}\rangle}{\langle \mathbf{v_2}, \mathbf{v_2}\rangle}\mathbf{v_2}\\
&= (-3,7,1,3) - \frac{-12}{4}(1,-1,1,-1) - \frac{16}{8}(0,2,2,0) = (0,0,0,0).
\end{aligned}$$

The Gram-Schmidt process can be used to check linear independence of vectors!.
The vector $\mathbf{x_3}$ *is a linear combination of* $\mathbf{x_1}$ *and* $\mathbf{x_2}$. V *is a plane, not a 3-dimensional subspace.*
We should orthogonalize vectors $\mathbf{x_1}, \mathbf{x_2}, \mathbf{y}$.

$$\begin{aligned}
\hat{\mathbf{v}}_\mathbf{3} &= \mathbf{y} - \frac{\langle \mathbf{y}, \mathbf{v_1}\rangle}{\langle \mathbf{v_1}, \mathbf{v_1}\rangle}\mathbf{v_1} - \frac{\langle \mathbf{y}, \mathbf{v_2}\rangle}{\langle \mathbf{v_2}, \mathbf{v_2}}\mathbf{v_2}\\
&= (0,0,0,1) - \frac{-1}{4}(1,-1,1,-1) - \frac{0}{8}(0,2,2,0)\\
&= (\frac{1}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4}).\\
||\hat{\mathbf{v}}_\mathbf{3}|| &= ||\left(\tfrac{1}{4}, -\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{3}{4}\right)|| = \frac{1}{4}||(1,-1,1,3)||\\
&= \frac{\sqrt{12}}{4} = \frac{\sqrt{3}}{2}.
\end{aligned}$$

Example 2.3.2.8. *Find the distance from the point* $\mathbf{z} = (0,0,1,0)$ *to the plane* π *that passes through the point* $\mathbf{x_0} = (1,0,0,0)$ *and is parallel to the vectors* $\mathbf{v_1} = (1,-1,1,-1)$ *and* $\mathbf{v_2} = (0,2,2,0)$.
The plane π *is not a subspace of* $\mathbb{R}^4$ *as it does not pass through the origin. Let* $\pi_0 = Span(\mathbf{v_1}, \mathbf{v_2})$. *Then*

$$\pi = \pi_0 + \mathbf{x_0}.$$

Hence the distance from the point $\mathbf{z}$ *to the plane* π *is the same as the distance from the point* $\mathbf{z} - \mathbf{x_0}$ *to the plane* π_0.
We shall apply the Gram-Schmidt process to vectors $\mathbf{v_1}, \mathbf{v_2}, \mathbf{z} - \mathbf{x_0}$.

This will yield an orthogonal system $\mathbf{w_1},\ \mathbf{w_2}, \mathbf{w_3}$. *The desired distance will be* $||\mathbf{w_3}||$.

$$\mathbf{v_1} = (1,-1,1,-1),\ \mathbf{v_2} = (0,2,2,0),\ \mathbf{z}-\mathbf{x_0} = (-1,0,1,0).$$
$$\mathbf{w_1} = \mathbf{v_1} = (1,-1,1,-1),$$
$$\mathbf{w_2} = \mathbf{v_2} - \frac{\langle \mathbf{v_2}, \mathbf{w_1}\rangle}{\langle \mathbf{w_1}, \mathbf{w_1}\rangle}\mathbf{w_1} = \mathbf{v_2} = (0,2,2,0) \ \textit{as}\ \mathbf{v_2} \perp \mathbf{v_1}.$$
$$\mathbf{w_3} = (\mathbf{z}-\mathbf{x_0}) - \frac{\langle \mathbf{z}-\mathbf{x_0}, \mathbf{w_1}\rangle}{\langle \mathbf{w_2}, \mathbf{w_1}\rangle}\mathbf{w_1} - \frac{\langle \mathbf{z}-\mathbf{x_0}, \mathbf{w_2}\rangle}{\langle \mathbf{w_2}, \mathbf{w_2}\rangle}\mathbf{w_2} = (-1,0,1,0) - \frac{0}{4}(1,-1,1,-1) - \frac{2}{8}(0,2,2,0)$$
$$= (-1, -\frac{1}{2}, \frac{1}{2}, 0).$$
$$||\mathbf{w_3}|| = \left|\Big(-1,-1/2,1/2,0\Big)\right| = \frac{1}{2}|(-2,-1,1,0)| = \frac{\sqrt{6}}{2} = \sqrt{\frac{3}{2}}.$$

Exercise 2.3.2.3. 1. *Let* W *be the subspace of* $\mathbb{R}^4$ *spanned by* $(1,0,-1,2)$ *and* $(2,1,0,-1)$ *and take the inner product to be the standard dot product.*

   (a) *Find the vector in* W *which is closest to the vector* $(4,3,2,1)$.

   (b) *Find a basis for* $W^{\perp}$ *and the dimension of* $W^{\perp}$.

   (c) *Find the unique* $\mathbf{w_1},\ \mathbf{w_2} \in W^{\perp}$ *such that* $(4,3,2,1) = \mathbf{w_1} + \mathbf{w_2}$.

2. *Let* $\mathbf{x_1} = (1,-1,1,-1),\ \mathbf{x_2} = (1,0,0,1),\ \mathbf{x_3} = (1,1,3,3)$ *in* $\mathbb{R}^4$ *and let* $S = Span(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3})$.

   (a) *If the Gram-Schmidt procedure is applied to the vectors* $\mathbf{x_1},\ \mathbf{x_2},\ \mathbf{x_3}$ *in that order, find the resulting orthonormal basis for* S.

   (b) *Find a basis for* $S^{\perp}$.

   (c) *Find an orthonormal basis for* $\mathbb{R}^4$ *which extends the orthonormal basis for* S *found in part (a).*

2.4 The dual space

Definition 2.4.0.3. *A mapping* $f: V \to K$ *is called linear functional iff* $f(a\boldsymbol{u}b + \boldsymbol{v}) = af(\boldsymbol{u}) + bf(\boldsymbol{v}), \forall \boldsymbol{u}, \boldsymbol{v} \in V$ *and* $a, b \in K$.*Let* V *be a vector space over the field* K.

Definition 2.4.0.4. *: - The dual space* $V^* := \{L(V,K)$ *of* V *set of all linear maps from* V *to* $K\}$. *The elements of* V^* *are linear functional on* V, *and the* ***annihilator*** *is defined by* $V^0 = \{l \in V^* : l(\boldsymbol{u}) = 0,\ \textit{for all}\ \mathbf{u} \in V\}$

Notation:

Let $\varphi \in V^*$ *and* $\boldsymbol{v} \in V^*$ *we shall use the notation* $< \varphi, \boldsymbol{v} >= \varphi(\boldsymbol{v})$

Note:

1. $< \varphi + \varphi_2, \mathbf{v} >= (\varphi + \varphi_2)(\mathbf{v}) = \varphi(\mathbf{v}) + \varphi_2(\mathbf{v}) =< \varphi, \mathbf{v} > + < \varphi_2, \mathbf{v} >$

2. $< \varphi, \mathbf{v_1} + \mathbf{v_2} >= \varphi(\mathbf{v_1} + \mathbf{v_2}) = \varphi(\mathbf{v_1}) + \varphi(\mathbf{v_2}) =< \varphi, \mathbf{v_1} > + < \varphi, \mathbf{v_2} >$

3. $< \lambda\varphi, \mathbf{v} >= (\lambda\varphi)(\mathbf{v}) = \lambda.\varphi(\mathbf{v}) = \lambda < \varphi, \mathbf{v} >$

4. $<\varphi, \lambda \mathbf{v}>=\lambda<\varphi, \mathbf{v}>$

Example 2.4.0.9. *Let* $V=\mathbb{R}^n$ *and* $\varphi:\mathbb{R}^n\to\mathbb{R}$*. The projection on the* i^{th} *component is defined by* $\varphi(\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n})=\mathbf{v_i}$*. Show that* φ *is a linear functional.*
Solution:

Lemma 2.4.0.1. *: - Assume that* V *and* W *are vector spaces over the field* K, $\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}\}$ *be basis of* V *and* $\{\mathbf{w_1}, \mathbf{w_2}, \cdots, \mathbf{w_n}\}$ *be an arbitrary element of* W*. Then there exist a unique linear map* $T:V\to W$ *defined by* $T(\mathbf{v_i})=\mathbf{w_i}$*, for* $i=1,2,\cdots,n$*.*
Proof

Note: Let V *be a finite dimensional vector space over a field* K*. Then* $dimV^*=dimV$*.*

Definition 2.4.0.5. *:- The basis* $\{\mathbf{v_1}^*, \mathbf{v_2}^*, \cdots, \mathbf{v_n}^*\}$ *of* V^* *called the dual basis of* $\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}\}$*.*

Theorem 2.4.0.4. *:- Let* V *be a finite dimensional inner product space,* $\varphi\in V^*$*. Then there exists a unique* $\mathbf{v}\in V$ *such that* $\varphi=L\mathbf{v}$*.*

Corollary 2.4.0.1. *: - Let* V *be a finite dimensional Euclidean space, then the map* $\varphi:V\to V^*$ *given by* $\varphi(\mathbf{v})=L\mathbf{v}$ *is an isomorphism.*

Note: If V *be a unitary space in the above corollary,* φ *fails to be linear.*

The Duality Principle *:*
For each ordered basis

$$\{\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_n}\}$$

of of a finite dimensional vector space V*, there exists a corresponding basis*

$$\{\omega_1, \omega_2, ..., \omega_n\}$$

for V^**, and vice versa such that*

$$\langle \omega_i, \mathbf{e_j}\rangle=\delta_{ij}.$$

The validation of the duality principle consists of the actual three-step construction of the basis dual to the given basis, which we denote by

$$B=\{\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_n}\}\subset V \text{ (basis for V).}$$

Steps of constructing dual basis

Step I For all vectors $\mathbf{x}$ *and* $\mathbf{y}$ *one has the following unique expansions:*

$$\begin{aligned}\mathbf{x}&=\alpha_1\mathbf{e_1}+...+\alpha_n\mathbf{e_n}\\ \mathbf{y}&=\beta_1\mathbf{e_1}+...+\beta_n\mathbf{e_n}\\ \mathbf{x}+\mathbf{y}&=(\alpha_1+\beta_1)\mathbf{e_1}+...+(\alpha_n+\beta_n)\mathbf{e_n}\\ \mathbf{cx}&=c\alpha_1\mathbf{e_1}+...+c\alpha_n\mathbf{e_n}\end{aligned}$$

Note that

$$\begin{aligned}\alpha_1 &\text{ is uniquely determined by } \mathbf{x}\\ \beta_1 &\text{ is uniquely determined by } \mathbf{y}\\ \alpha_1+\beta_1 &\text{ is uniquely determined by } \mathbf{x}+\mathbf{y}\\ c\alpha_1 &\text{ is uniquely determined by } \mathbf{cx}\end{aligned}$$

Step II These four relations determine a linear function, call it ω_1. Its defining properties are

$$\omega_1(\mathbf{x}) = \alpha_1, \quad \omega_1(\mathbf{y}) = \beta_1$$
$$\omega_1(\mathbf{x}+\mathbf{y}) = \alpha_1 + \beta_1, \quad \omega_1(c\mathbf{x}) = c\alpha_1$$

which imply

$$\omega_1(\mathbf{x}+\mathbf{y}) = \omega_1(\mathbf{x}) + \omega_1(\mathbf{y}), \quad \omega_1(c\mathbf{x}) = c\omega_1(\mathbf{x})$$

In particular, from the equation in step I, one has

$$\omega_1(\mathbf{e_1}) = 1, \quad \omega_1(\mathbf{e_2}) = 0, \quad \cdots, \quad \omega_1(\mathbf{e_n}) = 0.$$

*We conclude that ω_1 is a linear function, indeed. The function ω^1 is called the **first coordinate function**.*

Step III Similarly the j^{th} coordinate function, is defined by

$$\omega_j = \alpha_j, \ \ for \ j = 1, 2, ..., n.$$

By applying ω_j to the i^{th} basis vector $\mathbf{e_i}$, and using the equation in step I one obtains

$$\omega_j(\mathbf{e_i}) \equiv \langle \omega_j, \mathbf{e_i} \rangle = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}$$

or in terms of the Kronecker delta,

$$\langle \omega_j, \mathbf{e_i} \rangle = \delta_{ij}.$$

*This is called a **duality relation** or **duality principle**. The choice of a different vector basis would have resulted in a correspondingly different set of coordinate functions, but would have again resulted in a duality relation.*
Being elements in V^*, do these coordinate functions form a basis for V^*?
The answer to this important question is answered in the affirmative by the following:

Theorem 2.4.0.5. *(Dual Basis)*
If $B = \{\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_n}\}$ is a basis of finite dimensional inner product space V, then $B^* = \{\mathbf{e_1}^*, \mathbf{e_2}^*, ..., \mathbf{e_n}^*\}$ (the set of linear functions $B^* = \{\omega_j\}_{j=1}^n$ which satisfies the duality relation $\langle \omega_j, \mathbf{e_i} \rangle = \delta_{ij}$) is a basis for V^*.

Proof. **Spanning property**
Let $f \in V^*$ be some linear function on V. Evaluate $f(\mathbf{x})$ and use

$$\mathbf{x} = \sum_i \alpha_i \mathbf{e_i}.$$

Thus

$$f(\mathbf{x}) = f(\sum_i \alpha_i \mathbf{e_i}) = \sum_i f(\mathbf{e_i})\alpha_i, \ \alpha^i \text{ is the } i^{th} \ coordinate \ of \ \mathbf{x}, \ i.e. \ \alpha^i = \omega_i(\mathbf{x});$$
$$\sum_i f(\mathbf{e_i})\omega^i(\mathbf{x}), \ \forall \mathbf{x} \in V.$$

This holds for all $\mathbf{x} \in V$. Consequently,

$$f = \sum_i f(\mathbf{e_i})\omega^i.$$

which is an expansion of f in terms of the elements of B^*, which means that B^* is a spanning set for V^* indeed.

Linear independence property

Let us consider the equation

$$c_1\omega_1 + c_2\omega_2 + ... + c_n\omega_n = 0$$

where 0 is the function with constant value zero on V, then

$$\langle \sum_j c_i e_j^*, v \rangle = 0$$

for every $\mathbf{v} \in V$.

In particular, if $\mathbf{v} = \mathbf{e_i}$, we get

$$\mathbf{0} = \langle l, e_i \rangle = \sum_j c_j \langle e_j^*, e_i \rangle = \sum_j c_j \sigma_{ji} = c_i$$

for $1 \leq i \leq n$, By evaluating both sides on the i^{th} basis vector $\mathbf{e_i}$ and using the second step in the above process one obtains $c_i = 0$ for $i = 1, 2, ..., n$.

Consequently, B^* does have the linear independence property. Together with its spanning property, this validates the claim made in the Theorem that B^* is a basis for V^*. ☐

Example 2.4.0.10. *(Column space∗ =Row space)*

GIVEN:

Let

$$B = \{\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{e_2} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{e_3} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\}$$

be a basis for the column space $V = \mathbb{R}^3$.

a. Identify V^**, the space dual to* V.

b. Find the basis

$$B^* = \{\omega_1, \omega_2, \omega_3\} \text{ dual to } B.$$

***Solution**:*

a. The space dual to V *consists of the row space*

$$V^* = \{\sigma = [a \ b \ c] : \ a, b, c \in \mathbb{R}\}.$$

Indeed, for any $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$,

$$\langle \sigma, \mathbf{x} \rangle = \sigma \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = [a \ b \ c] \begin{bmatrix} x \\ y \\ z \end{bmatrix} = ax + by + cz.$$

$$\langle \omega_1, \mathbf{e_1} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1 \Rightarrow a = 1$$

$$\langle \omega_1, \mathbf{e_2} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0 \Rightarrow a + b = 0 \Rightarrow b = -a = -1$$

$$\langle \omega_1, \mathbf{e_3} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0 \Rightarrow c = -a - b = -1 + 1 = 0$$

$$\Rightarrow \omega_1 = [1\ \ -1\ 0],$$

$$\langle \omega_2, \mathbf{e_1} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 0 \Rightarrow a = 0$$

$$\langle \omega_2, \mathbf{e_2} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 1 \Rightarrow b = 1$$

$$\langle \omega_2, \mathbf{e_3} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0 \Rightarrow c = -a - b = 0 - 1 = -1$$

$$\Rightarrow \omega_2 = [0\ 1\ \ -1],$$

$$\langle \omega_3, \mathbf{e_1} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 0 \Rightarrow a = 0$$

$$\langle \omega_3, \mathbf{e_2} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0 \Rightarrow b = 0$$

$$\langle \omega_3, \mathbf{e_3} \rangle = [a\ b\ c] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 1 \Rightarrow c = 1$$

$$\Rightarrow \omega_3 = [0\ 0\ 1].$$

Thus the basis of duals for V^*, the space dual to $V = \mathbb{R}^3$ is

$$B^* = \{\omega_i\}_{i=1}^3 = \{[1\ \ -1\ 0], [0\ 1\ \ -1], [0\ 0\ 1]\}.$$

Example 2.4.0.11. *Let $\mathcal{B} = \{f_1, f_2, f_3\}$ denote the dual basis for V^* the standard ordered basis for V be $\mathcal{B}' = \{\mathbf{e_1} = (1,0,0),\ \mathbf{e_2} = (0,1,0),\ \mathbf{e_3} = (0,0,1)\}$.*
***Solution**:*
To find an explicit formula for f_1 we need to consider the equations

$$\begin{aligned} 1 &= f_1((1,0,1)) = f_1(\mathbf{e_1} + \mathbf{e_3}) = f_1(\mathbf{e_1}) + f_1(\mathbf{e_3}) = f_1(\mathbf{e_1}), \\ 0 &= f_1((1,2,1)) = f_1(\mathbf{e_1}) + 2f_1(\mathbf{e_2}) + f_1(\mathbf{e_3}) = 1 + 2f_1(\mathbf{e_2}), \\ 0 &= f_1((0,0,1)) = f_1(\mathbf{e_3}) \end{aligned}$$

implying

$$f_1(\mathbf{e_1}) = 1,\ f1(\mathbf{e_2}) = \frac{1}{2},\ f_1(\mathbf{e_3}) = 0$$

implying

$$f_1(x, y, z) = x - \frac{1}{2}y.$$

To find an explicit formula for f_2 *we need to consider the equations*

$$\begin{aligned} 0 = f_2((1,0,1)) = f_2(\mathbf{e_1} + \mathbf{e_3}) = f_2(\mathbf{e_1}) + f_2(\mathbf{e_3}) = f_2(\mathbf{e_1}), \\ 1 = f_2((1,2,1)) = f_2(\mathbf{e_1}) + 2f_2(\mathbf{e_2}) + f_2(\mathbf{e_3}) = 2f_2(\mathbf{e_2}), \\ 0 = f_2((0,0,1)) = f_2(\mathbf{e_3}) \end{aligned}$$

implying

$$f_2(\mathbf{e_1}) = 0,\ f_2(\mathbf{e_2}) = \frac{1}{2},\ f_2(\mathbf{e_3}) = 0$$

implying

$$f_2(x, y, z) = \frac{1}{2}y.$$

Finally, to find an explicit formula for f_3 *we need to consider the equations*

$$\begin{aligned} 0 = f_3((1,0,1)) = f_3(\mathbf{e_1} + \mathbf{e_3}) = f_3(\mathbf{e_1}) + f_3(\mathbf{e_3}) = f_3(\mathbf{e_1}) + 1 \\ 0 = f_3((1,2,1)) = f_3(\mathbf{e_1}) + 2f_3(\mathbf{e_2}) + f_3(\mathbf{e_3}) = -1 + 2f_3(\mathbf{e_2}) + 1 = 2f_3(\mathbf{e_2}), \\ 1 = f_3((0,0,1)) = f_3(\mathbf{e_3}) \end{aligned}$$

implying

$$f_3(\mathbf{e_1}) = -1,\ f_3(\mathbf{e_2}) = 0,\ f_3(\mathbf{e_3}) = 1$$

implying

$$f_3(x, y, z) = -x + z.$$

Exercise 2.4.0.6. *: Identify* V^**, the space dual to* V *and find the basis*

$$B^* = \{\omega_1, \omega_2, \omega_3\} \text{ dual to } B.$$

given the basis for the column space $V = \mathbb{R}^3$.

a.

$$B = \{\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},\ \mathbf{e_2} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},\ \mathbf{e_3} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}\}$$

b.

$$B = \{\mathbf{u} = (1,0,1),\ \mathbf{v} = (1,-1,0), \mathbf{w} = (2,0,-1)\}.$$

2.4.1 Adjoint of linear operator

Definition 2.4.1.1. *: - Let $< V, <, >>$ be an n-dimensional Euclidean space and $T : V \to V$ a linear operator. The adjoint T^* of T is the linear operator $S : V \to V$ such that $< T(u), v >=< u, S(v) >$, for all $u, v \in V$.*
Here if A is an $m \times n$ matrix which rpesents the linear operator T, then the adjoint of A is denoted and given by

$$A^* = (A^*)_{ij} = \overline{A_{ji}}$$

Example 2.4.1.1. *: Find $T^*(x, y, z)$ of the linear operator $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(x, y, z) = (x + 2y, 3x - 4z, y)$, for all $x, y, z \in \mathbb{R}$.*
Solution:
T have usual basis of $\mathbb{R}^3$.

$$T(1,0,0) = (1,3,0) = 1.(10,0) + 3(0,1,0) + 0.(0,0,1)T(0,1,0) \quad = (2,0,1) = 2.(1,0,0) + 0.(0,1,0) + 1.(0,0,1$$

$M_{ST} = \begin{pmatrix} 1 & 3 & 0 \\ 2 & 0 & 1 \\ 0 & -4 & 0 \end{pmatrix}^t = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 0 & -4 \\ 0 & 1 & 0 \end{pmatrix}$. *Therefore,* $[T^*]_B = A^* = (A^{-1})^t = (\overline{A})^t = \begin{pmatrix} 1 & 3 & 0 \\ 2 & 0 & 1 \\ 0 & -4 & 0 \end{pmatrix}$.
Thus $T^*(x, y, z) = (x + 3y, 2x + z, -4y)$.

Theorem 2.4.1.1.

- $\langle x, y + z\rangle = \langle x, yi + hx, z\rangle$.
- $\langle x, cy\rangle = \bar{c}\langle x, y\rangle$.
- $\langle x, 0\rangle = \langle 0, x\rangle = 0$.
- $\langle x, x\rangle = 0$ *if and only if* $x = 0$.
- *If $\langle x, y\rangle = \langle x, z\rangle$ for every $x \in V$, then $y = z$.*

Theorem 2.4.1.2. *: - Let V be a finite dimensional inner product space, $\lambda \in K$,T and S be linear operators on V. Then*

1. $(T + S)^* = T^* + S^*$
2. $(\lambda T)^* = \overline{\lambda} T^*$
3. $(TOS)^* = S^* OT^*$
4. $(T^*)^* = T$
5. $(I_n)^* = I_n$
6. $(0)^* = 0$

2.4.2 Self-adjoint linear operators

Preliminary Concept

Conjugate of matrix

Definition 2.4.2.1. *: - Let A be a matrix over a field $\mathbb{C}$, conjugate of A is denoted and defined as $\overline{A}$ is a matrix obtained from A by replacing each entry by its conjugate.*

Example 2.4.2.1. *:-* $A = \begin{pmatrix} 1+3i & 4i \\ 2 & 5-6i \end{pmatrix}$ *then* $\overline{A} = \begin{pmatrix} 1-3i & -4i \\ 2 & 5+6i \end{pmatrix}$

Theorem 2.4.2.1. *: - Let A, B be matrices over $\mathbb{C}$, $\lambda \in \mathbb{C}$ then*
1. $\overline{\overline{A}} = A$ *2.* $\overline{\lambda A} = \overline{\lambda}.\overline{A}$ *3.* $\overline{A+B} = \overline{A} + \overline{B}$
4. $\overline{AB} = (\overline{A})(\overline{B})$ *5.* $(\overline{A})^t = \overline{(A)^t}$

Hermitian and Skew-Harmitian matrices

Definition 2.4.2.2. *: - Let A be a square matrix over $\mathbb{C}$. A is called Hermitian if $A^* = A$ and Skew-Harmitian $A^* = -A$*

Note:
1. The diagonal elements of a hermitian matrix are real numbers.
2. The diagonal elements of a skew-hermitian matrices either zero or pure imaginaries.

orthogonal matrix

Definition 2.4.2.3. *: - Let A be a square matrix over $\mathbb{R}$. A is called orthogonal if $A^t = A^{-1}$*

Example 2.4.2.2. *: Show that the matrix* $A = \begin{pmatrix} 2/3 & -1/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \\ -1/3 & 2/3 & 2/3 \end{pmatrix}$ *is an orthogonal matrix and the columns A form an orthonormal set.*

Solution:

Theorem 2.4.2.2. *: - Let A, B be orthogonal matrices of the same size, then A^t, A^{-1} and AB are orthogonal $det(A) = \pm 1$*

Unitary Matrices

Definition 2.4.2.4. *: - Let A be a square matrix over $\mathbb{C}$. A is called unitary if and only if $A^* = A^{-1} = (A^{-1})^t$.*

Theorem 2.4.2.3. *: - Suppose A and B are unitary matrices of the same size, then A^t, A^{-1} and AB are also unitary.*

Self-Adjoint Linear operators

Definition 2.4.2.5. *: - An operator $T \in V$ is called self-adjoint if $T^* = T$.*

Note:-
If V is a Euclidean space and T is self adjoint linear operator on V, then T is called symmetric.
If V is a unitary space and T is self adjoint linear operator on V, then T is called hermitian.

Example 2.4.2.3. *:- The linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x, y) = (y, x)$for all $x, y \in \mathbb{R}$ is self-adjoint.*

Example 2.4.2.4. *: - let $V = \mathbb{R}^2$, with an inner product $<,>$ defined by $A = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ w.r.t a fixed basis $E = \{(1,0),(0,1)\}$. Let $T : V \to V$ a linear operator, defined by $T(E) = EB$,where $B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$. Show that T is self adjoint operator or not. If so, find the matrix of its adjoint operator w.r.t E.*

Solution:

If T were self-adjoint, we would have $< T(Ex), Ey >=< Ex, T(Ey) >; \forall Ex, Ey \in V$. From which $x^t(B^tA)y = x^t(AB)y, \quad$ for all $x, y \in M_{n,1}(R)$(minor).

Therefore, we should have $B^tA = AB$, but: $B^tA = \begin{pmatrix} 0 & 1 \\ 3 & -1 \end{pmatrix} \neq \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix}$. Hence T is not self-adjoint.

The adjoint of T is defined by: $S(E) = EC$ with $C = A^{-1}B^tA = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$.

Lemma 2.4.2.1. *:- Let T be a linear operator on a finite dimensional inner product space V, then $< T(v), w >= 0$, for every $v, w \in V$.*

Lemma 2.4.2.2. *: - Let T be a self adjoint linear operator on a finite dimensional inner product space V, if and only if $< T(v), v >= 0$, for all $v \in V$.*

Lemma 2.4.2.3. *: - Let T be a linear operator on a finite dimensional unitary space V, then $< T(v), v >= 0, \forall v \in V$.*

Theorem 2.4.2.4. *: - T be a self adjoint linear operator on V if and only if each eigenvalue of T is real and admits an orthonormal eigenbasis with real eigenvalues*

Proof. i. Let $v \in V$ be a non zero vector such that $Tv = \lambda v$. Then

$$\lambda < v, v >=< Tv, v >=< v, T^*v >= \overline{\lambda} < v, v > .$$

Thus, $\lambda = \overline{\lambda}$, which means λ is real.

ii. For $F = \mathbb{C}$: we already know that M_T is upper-triangular is some orthonormal basis.

Method II Suppose that $T(\mathbf{v}) = \lambda \mathbf{v}$ for $\mathbf{v} \leq 0$. Because a self-adjoint operator is normal, then if $\mathbf{v}$ is an eigenvector of T then $\mathbf{v}$ is also an eigenvector of T^*. Thus

$$\lambda \mathbf{v} = T(\mathbf{v}) = T^*(\mathbf{v}) = \overline{\lambda}\mathbf{v}.$$

□

Theorem 2.4.2.5. *Let T be a linear operator on a finite dimensional inner product space V, $B = \{v_1, v_2, \cdots, v_n\}$ be an orthonormal basis of V, $A = \left(a_{ij}\right) n \times n$, $[T]_B =$ the matrix representation of T with respect to B, then $a_{ij} =< T(v_j), v_i >$.*

Proof:

Corollary 2.4.2.1. *: - Let T be a linear operator on finite dimensional inner product space V, $B = \{v_1, v_2, \cdots, v_n\}$ be an orthonormal basis of V, $A = [T]_B$ =the matrix representation of T with respect to B, then $[T^*]_B = A^*$.*

2.5 Isometry

Definition 2.5.0.6. *:- An isometry is an operator such that* $< u, v >=< Tu, Tv > \forall u, v \in V$. *That is, isometries are operators that preserve the inner product.*

Example 2.5.0.5. *For the Euclidean n-space* $\mathbb{R}^n$ *with the dot product, rotations and reflections are isometries.*

Proposition 2.5.0.1. *:- T is an isometry iff it preserves length (norm)i.e.* $\| T(v) \|=\| v \|$, *for every* $v \in V$.

Proof.

$$\begin{aligned} ||T(\mathbf{u}+\mathbf{v}||^2 = \langle T(\mathbf{u}+\mathbf{v}), T(\mathbf{u}+\mathbf{v})\rangle &= \langle T(\mathbf{u}), T(\mathbf{u})\rangle + \langle T(\mathbf{v}), T(\mathbf{v})\rangle\langle +2T(\mathbf{u}), T(\mathbf{v})\rangle \\ &= ||T(\mathbf{u})||^2 + ||T(\mathbf{v})||^2 + 2\langle T(\mathbf{u}), T(\mathbf{v})\rangle, \\ ||\mathbf{u}+\mathbf{v}||^2 = \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle + 2\langle \mathbf{u}, \mathbf{v}\rangle \\ &= ||\mathbf{u}||^2 + ||\mathbf{v}||^2 + 2\langle \mathbf{u}, \mathbf{v}\rangle. \end{aligned}$$

It is clear that the length preserving is equivalent to the inner product preserving. ☐

Characterization of isometries *Let* $V = W$ *be finite dimensional inner product space.* $T \in L(V)$ *is an isometry iff* $T^*T = I = TT^*$.
An $n \times n$ *matrix* Q *is called orthogonal if* $QQ^T = I$, *i.e.,*

$$Q^{-1} = Q^T.$$

Theorem 2.5.0.6. *- Let V be a finite dimensional inner product space, T be a linear operator on V. Then the following statements are equivalent.*

a. $T^* = T^{-1}$

b. *T preserves inner product i.e.* $< T(v), T(w) >=< v, w >, \forall v, w \in V$.

c. *T preserves length (norm) i.e.* $\| T(v) \|=\| v \|$, *for every* $v \in V$.

Definition 2.5.0.7. *: - Let T be a linear operator on a finite dimensional inner product space V.*
An isometry on a Euclidean space is called orthogonal operator.
An isometry on a unitary space is called unitary operator
Let T be an isometry and $\{v_1, v_2, \cdots, v_n\}$ *be an orthonormal basis of V. Then*
$\| T(v_i) \|=\| v_i \|= 1$
$< T(v_i), T(v_j) >=< v_i, v_j >= \delta_{ij}$ *hence* $\{T(v_1), T(v_2), \cdots, T(v_n)\}$ *is also an orthonormal basis for V. Because an isometry measures the inner product it measures orthogonality and other matrix notions.*

Theorem 2.5.0.7. *: - Let T be a linear operator on a finite dimensional inner product space V. Let A be the matrix representation of T relative to the orthonormal basis, then T is an isometry if and only if* $A^* = A^{-1}$.
Two inner product spaces V and W over $\mathbb{F}$ *are isometric, if we can find an isometry* $L : V \to W$, *i.e. an isomorphism such that*

$$\langle L(\mathbf{x}), L(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle.$$

2.6 Normal operators and the spectral theorem

Definition 2.6.0.8. *:- Let T be a linear operator on a finite dimensional inner product space V. T is called a normal operator if and only if $ToT^* = T^*oT$.*

Example 2.6.0.6. *: - show that the operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (y, -x), \forall x, y \in \mathbb{R}$ is normal operator.*

Proof. $T(x, y) = (y, -x), \forall x, y \in \mathbb{R}$ then $T^*(x, y) = (-y, x)$, by the definition of T^*.
Then, $TOT^*(x, y) = T(-yx) = (x, y)$ and $T^*T(x, y) = T^*(y, -x) = (x, y)$.
So,$T^*T = TT^*$,
Hence, T is normal . □

Exercise 2.6.0.8. *show that every diagonal matrix is normal.*

Lemma 2.6.0.4. *:- An operator T on V is normal iff $\| T(v) \| = \| T^*(v) \|, \forall v \in V$.*

Proof. $\| T(v) \| = \| T^*(v) \|$
$\Leftrightarrow < T(v), T(v) > = < T^*(v), T^*(v) >$
$\Leftrightarrow < v, T^*T(v) > = < v, TT^*(v) >$
$\Leftrightarrow < (T^*T - TT^*)v, v > = 0$
$\Leftrightarrow T^*T = TT^*$
That is, Since $T^*T - TT^*$ is self adjoint $\forall v$ iff T is normal. □

Lemma 2.6.0.5. *: - Let T be normal operator be a finite dimensional inner product space on V. Then for any $\lambda \in K, T - \lambda I$ is also normal operator.*

Theorem 2.6.0.9. *: - Let T be normal operator be a finite dimensional inner product space on V. Then $v \in V$ such that $v \neq 0$ is an eigenvector of T with eigenvallue λ iff v is the eigenvector of T^* with eigenvalue $\overline{\lambda}$.*

Proof. $T(v) = \lambda v \leftrightarrow T(v) - \lambda v = 0 \Leftrightarrow (T - \lambda I)(v) = 0 \Leftrightarrow$ since T is normal $T - \lambda I$ is also normal.
Now, $(T - \lambda I)^* = T^* - \overline{\lambda} I$.
Hence $(T^* - \overline{\lambda} I)(v) = 0$
$\Leftrightarrow T^*(v) - \overline{\lambda} I(v) = 0$
$\Leftrightarrow T^*(v) = \overline{\lambda} v$.
$\Leftrightarrow$ Since $v \neq 0, \overline{\lambda}$ is an eigenvalue of T^*. □

Definition 2.6.0.9. *:- Let V be a vector space over K, T be a linear operator on V, W be a subspace of V. we say W is T invariant if for each $w \in W$ the vector $T(w) \in W$.*

Lemma 2.6.0.6. *: - Let T be a linear operator an inner product space on V, W be a T invariant subspace of V. Then W orthogonal $W^\perp$ is T^* invariant.*

Theorem 2.6.0.10. *: - (Spectral Theorem) Let T be a self-adjoint operator on a finite dimensional Euclidean space or a normal operator on a finite dimensional unitary space V, $\lambda_1, \lambda_2, \cdots, \lambda_n$ be distinct eigenvalues of T, T_j orthogonal projection of V on w_j. Then*

i. $w_j \perp w_i, i \neq j$

ii. $V = w_1 \bigoplus w_2 \bigoplus, \cdots, \bigoplus w_n$

iii. $T = \sum_{i=1}^{n}(\lambda_i T_i)$

(Equivalently, any real symmetric matrix A can be diagonalized by an orthogonal matrix. More specifically, there exists an orthonormal basis $\mathcal{B} = \{\mathbf{u_1}, ..., \mathbf{u_n}\}$ of $\mathbb{R}^n$ such that

$$A\mathbf{u_i} = \lambda_i \mathbf{u_i},\ 1 \le i \le n,$$

$$Q^{-1}AQ = Q^T AQ = Diag[\lambda_1, ...\lambda_n],$$

where $Q = [\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_n}]$; and spectral decomposition

$$A = \lambda_1 \mathbf{u_1}\mathbf{u_1}^T + ... + \lambda_n \mathbf{u_n}\mathbf{u_n}^T.$$

Proof. Let $T : \mathbb{R}^n \to \mathbb{R}^N$ has an orthonormal basis $\mathcal{B} = \{\mathbf{u_1}, ..., \mathbf{u_n}\}$ of eigenvectors of T,

$$T(\mathbf{u_i}) = \lambda_i \mathbf{u_i},\ 1 \le i \le n;$$

and $Q = [\mathbf{u_1}, ..., \mathbf{u_n}]$. Then

$$Q^{-1}AQ = Q^T AQ = Diag[\lambda_1, ...\lambda_n] = D.$$

Alternatively,

$$\begin{aligned} A &= QDQ^{-1} = QDQ^T \\ &= [\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_n}] \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u_1}^T \\ \mathbf{u_2}^T \\ \vdots \\ \mathbf{u_n}^T \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 \mathbf{u_1} + \lambda_2 \mathbf{u_2} + ... + \lambda_n \mathbf{u_n} \end{bmatrix} \begin{bmatrix} \mathbf{u_1}^T \\ \mathbf{u_2}^T \\ \vdots \\ \mathbf{u_n}^T \end{bmatrix} \\ &= \lambda_1 \mathbf{u_1}\mathbf{u_1}^T + \lambda_2 \mathbf{u_2}\mathbf{u_2}^T + ... + \lambda_n \mathbf{u_n}\mathbf{u_n}^T. \end{aligned}$$

□

Note. It is clear that if a real square matrix A is orthogonally diagonalizable, then A is symmetric.

Example 2.6.0.7. *Is the matrix $A = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ 3 & 3 & 1 \end{bmatrix}$ orthogonally diagonalizzable?*

Solution:
The characteristic polynomial of A is

$$\Delta(t) = (t+2)^2(t-7).$$

There are eigenvalues $\lambda_1 = -2$ and $\lambda_2 = 7$.
For $\lambda_1 = -2$, there are two independent eigenvectors

$$\mathbf{v_1} = [-1, 1, 0]^T,\ \mathbf{v_2} = [-1, 0, 1]^T.$$

Set $\mathbf{w_1} = \mathbf{v_1}$,

$$\mathbf{w_2} = \mathbf{v_2} - \frac{\mathbf{v_2}.\mathbf{w_1}}{\mathbf{w_1}.\mathbf{w_1}}\mathbf{w_1} = [\frac{-1}{2}, \frac{1}{2}, 1]^T.$$

Then

$$\mathbf{u_1} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \ \mathbf{u_2} = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{2}} \end{bmatrix}$$

form an orthonormal basis of E_{λ_1}.
For $\lambda_2 = 7$, *there is one independent eigenvector*

$$\mathbf{v_3} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The orthonormal basis of E_{λ_2} *is*

$$\mathbf{u_3} \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

Then the orthogonal matrix

$$Q = \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

diagonalizes the symmetric matrix A.

2.7 Factorization of a matrix

In linear algebra, a matrix decomposition or matrix factorization is a factorization of a matrix into a product of matrices. There are many different matrix decompositions. Among those

Preliminary concepts

Definition 2.7.0.10.

- *A is positive definite if A is symmetric and*

$$x^T Ax > 0 \ \forall x \geq 0$$

That is definite if it's symmetric and all its pivots are positive. the k^{th} *pivot of a matrix is*

$$d_k = \frac{det(A_k)}{det(A_{k_l})}$$

where A_k *is the upper left* $k \times k$ *submatrix. All the pivots will be positive if and only if* $det(A_k) > 0$ *for all* $1 \leq k \leq n$. *So, if all upper left* $k \times k$ *determinants of a symmetric matrix are positive, and all of its eigenvalues are positive the matrix is positive definite.*

- *A is positive semi-definite if A is symmetric and*

$$x^T Ax \geq 0 \ \forall \ x.$$

and if all of its eigenvalues are non-negative

Note: if A is symmetric of order n,then

$$x^TAx = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j = \sum_{i=1} a_{ii}x_i^2 + 2\sum_{i>j} a_{ij}x_ix_j.$$

Example 2.7.0.8. *Determine whether the following matrices are positive, semipositive of neither.*

a. $A = \begin{pmatrix} 9 & 6 \\ 6 & 5 \end{pmatrix}$ *b.* $B = \begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}$ *c.* $C = \begin{pmatrix} 9 & 6 \\ 6 & 3 \end{pmatrix}$

Solution

A is positive since

$$x^TAx = 9x_1^2 + 12x_1x_2 + 5x_2^2 = (3x_1 + 2x_2)^2 + x_2^2$$

B is semi-positive definite but not positive definite: since

$$x^TAx = 9x_1^2 + 12x_1x_2 + 4x_2^2 = (3x_1 + 2x_2)^2.$$

C is not positive semi-definite:

$$x^TAx = 9x_1^2 + 12x_1x_2 + 3x_2^2 = (3x_1 + 2x_2)^2 - x_2^2.$$

2.7.1 LU decomposition

key point *: An LU decomposition of a matrix A is the product of a lower triangular matrix and an upper triangular matrix that is equal to A.*

Definition 2.7.1.1. *A $m \times n$ matrix is said to have a LU-decomposition if there exists matrices L and U with the following properties:*

(i) L is a $m \times n$ lower triangular matrix with all diagonal entries being 1.

(ii) U is a $m \times n$ matrix in some echelon form.

(iii) $A = LU$.

Suppose we have the system of equations

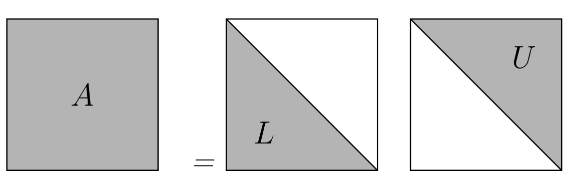
$$AX = B.$$

The motivation for an LU decomposition is based on the observation that systems of equations involving triangular coefficient matrices are easier to deal with. Indeed, the whole point of Gaussian elimination is to replace the coefficient matrix with one that is triangular. The LU decomposition is another approach designed to exploit triangular systems. We suppose that we can write

$$A = LU$$

where L is a lower triangular matrix and U is an upper triangular matrix. Our aim is to find L and U and once we have done so we have found an LU decomposition of A.

It turns out that we need only consider lower triangular matrices L that have $1'$s down the diagonal.



Types of LU decomposition :

*The principle difference between **Doolittle's** and **Crout's** LU decomposition method is the calculation sequence these methods follow. Both the methods exhibit similarity in terms of inner product accumulation.*

*In **Doolittle's method**, calculations are sequenced to compute one row of L followed by the corresponding row of U until A is exhausted. Below is the computational sequence and algorithm for Doolittle's LU decomposition.*

For each

$$i = 1, 2, ..., n-1:$$

$$u_{ik} = a_{ik} - \sum_{j=1}^{i}(L_{ij}u_{jk})$$

For $k = i, i+1, ..., n-1$ *produces the* k^{th} *row of* U.

$$L_{ik} = \frac{(a_{ik} - \sum_{j=1}^{i}(L_{ij}u_{jk}))}{u_{kk}}$$

For $i = k+1, k+2, ..., n-1$ *and* $L_{ii} = 1$ *produces the* k^{th} *column of* L.

Crout's** LU **Factorization

An equivalent LU decomposition of $A = LU$ *may be obtained by assuming that L is lower triangular and U is unit upper triangular. This factorization scheme is referred to as Crout's method. The defining equations for Crout's method are*

$$L_{ij} = a_{ij} - \sum_{k=1}^{i-1} L_{ik}u_{kj}, \text{ where } i \geq j$$

and

$$u_{ij} = \frac{a_{ij} - \sum_{k=1}^{i-1} L_{ik}u_{kj}}{L_{ii}}, \text{ where } i < j.$$

Example 2.7.1.1.

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{pmatrix} = LU,$$

where

$$L = \begin{pmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{pmatrix} \text{ and } U = \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ 0 & U_{22} & U_{23} \\ 0 & 0 & U_{33} \end{pmatrix}.$$

Then

$$LU = \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ L_{21}U_{11} & L_{21}U_{12} + U_{22} & L_{21}U_{13} + U_{23} \\ L_{31}U_{11} & L_{31}U_{12} + L_{32}U_{12} + L_{32}U_{22} & L_{31}U_{13} + L_{32}U_{23} + U_{33} \end{pmatrix} = A.$$

Now we use this to find the entries in L and U. Fortunately this is not nearly as hard as it might at first seem. We begin by running along the top row to see that

$U_{11} = 1,\ U_{12} = 2,\ U_{13} = 4$.

Now consider the second row $L_{21}U_{11} = 3 \therefore L_{21} \times 1 = 3 \Rightarrow L_{21} = 3$,

$L_{21}U_{12} + U_{22} = 8 \Rightarrow 3 * 2 + U_{22} = 8 \Rightarrow U_{22} = 2$,

$L_{21}U_{13} + U_{23} = 14 \Rightarrow 3 * 4 + U_{23} = 14 \Rightarrow U_{23} = 2$.

Notice how, at each step, the equation being considered has only one unknown in it, and other quantities that we have already found. This pattern continues on the last row

$$L_{31}U_{11} = 2 \Rightarrow L_{31} * 1 = 2 \Rightarrow L_{31} = 2,$$
$$L_{31}U_{12} + L_{32}U_{22} = 6 \Rightarrow 2 * 2 + L_{32} * 2 = 6 \Rightarrow L_{32} = 1,$$
$$L_{31}U_{13} + L_{32}U_{23} + U_{33} = 13 \Rightarrow (2 * 4) + (1 * 2) + U_{33} = 13 \Rightarrow U_{33} = 3.$$

We have shown that

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix} . \begin{pmatrix} 1 & 2 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix},$$

and this is an LU decomposition of A.

Exercise 2.7.1.1. *Find an LU decomposition of*

a. $\begin{pmatrix} 3 & 1 \\ -6 & -4 \end{pmatrix}$ *b.* $\begin{pmatrix} 3 & 1 & 6 \\ -6 & 0 & -16 \\ 0 & 8 & -17 \end{pmatrix}$

Theorem 2.7.1.2. *Let A be a $m \times n$ matrix and $E_1, E_2, ..., E_k$ be elementary matrices such that*

$$U = E_k E_{k-1} ... E_1 A$$

is in row echelon form . If none of the $E_i^{'}s$ corresponds to the operation of row interchange, then

$$C = E_k ... E_1$$

is a lower triangular invertible matrix. Further $L = C^{-1}$ is also a lower triangular matrix with $A = LU$.

Proof:
We observe that if an elementary row operation does not involve a row-interchange , then to reduce A to row echelon form, each row operation involved is either multiplying a row by a nonzero scalar, or adding a row to some row below it.
Thus, all the elementary matrices $E_1, E_2, ..., E_k$ such that

$$U = E_K ... E_1 A$$

is in row echelon form , are lower triangular and invertible.
Hence, $C = E_k E_{k-1} ... E_1$ is also lower triangular and invertible. Finally

$$L = C^{-1}$$

is also lower triangular and

$$A = LU.$$

Do matrices always have an LU decomposition? *:*
No. Sometimes it is impossible to write a matrix in the form "lower triangular " × " upper triangular".
Why not?
An invertible matrix A has an LU decomposition provided that all its leading sub-matrices have non-zero determinants. The k^{th} leading submatrix of A is denoted A_k and is the $k \times k$ matrix found by looking only at the top k rows and leftmost k columns.

Example 2.7.1.2. *Determine whether the matrix* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 1 & 3 & 4 \end{pmatrix}$ *has LU decomposition or not.*

Solution:
The second leading submatrix has determinant equal to

$$\left| \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \right| = (1 \times 4) - (2 \times 2) = 0.$$

0 *is not positive. Due to this the matrix A doesn't have an LU decomposition.*

Example 2.7.1.3. *Let*

$$\begin{pmatrix} 2 & -1 & 3 \\ 4 & 2 & 1 \\ -6 & -1 & 2 \end{pmatrix}$$

Then by applying row operations $R_2 \leftrightarrow R_2 - 2R_1$, $R_3 \leftrightarrow R_3 + 3R_1$, $R_3 \leftrightarrow R_3 + R_2$ *on A, we obtain*

$$A \sim \begin{pmatrix} 2 & -1 & 3 \\ 0 & 4 & -5 \\ 0 & -4 & 11 \end{pmatrix} \sim \begin{pmatrix} 2 & -1 & 3 \\ 0 & 4 & -5 \\ 0 & 0 & 6 \end{pmatrix} = U.$$

The corresponding elementary matrices and their inverses are

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; E_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; E_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \; E_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

Thus,

$$L = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & -1 & 1 \end{pmatrix}.$$

We observe that

$$A = \begin{pmatrix} 2 & -1 & 3 \\ 4 & 2 & 1 \\ -6 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 3 \\ 0 & 4 & -5 \\ 0 & 0 & 6 \end{pmatrix}.$$

Exercise 2.7.1.3. *Determine whether the following matrices has an LU decomposition or not.*

a. $\begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix}$ *b.* $\begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix}$ *c.* $\begin{pmatrix} 1 & -3 & 7 \\ -2 & 6 & 1 \\ 0 & 3 & -2 \end{pmatrix}$ *d.* $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 1 & 3 & 4 \end{pmatrix}$

e. $\begin{pmatrix} 1 & -3 & 7 \\ -2 & 6 & 1 \\ 0 & 3 & -2 \end{pmatrix}$

Once a matrix A has been decomposed into lower and upper triangular parts it is possible to obtain the solution to $AX = B$ in a direct way. The procedure can be summarized as follows

- *Given A, find L and U so that $A = LU$. Hence $LUX = B$.*
- *Let $Y = UX$ so that $LY = B$. Solve this triangular system for Y .*
- *Finally solve the triangular system $UX = Y$ for X.*

The benefit of this approach is that we only ever need to solve triangular systems. The cost is that we have to solve two of them.

Example 2.7.1.4. *Find the solution of $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ of the system $\begin{pmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 13 \\ 4 \end{pmatrix}$.*

Solution:

The first step is to calculate the LU decomposition of the coefficient matrix on the left-hand side. In this case that job has already been done since this is the matrix we considered earlier. We found that

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix}, \; U = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}.$$

The next step is to solve $LY = B$ for the vector $Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$. That is we consider

$$LY = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \begin{pmatrix} 3 \\ 13 \\ 4 \end{pmatrix} = B$$

*which can be solved using **forward substitution**. From the top equation we see that $y_1 = 3$. The middle equation states that $3y_1 + y_2 = 13$ and hence $y_2 = 4$. Finally the bottom line says that $2y_1 + y_2 + y_3 = 4$ from which we see that $y_3 = -6$.*

Now that we have found Y we finish the procedure by solving $UX = Y$ for X. That is we solve

$$UX = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ -6 \end{pmatrix} = Y$$

*by using **back substitution**. Starting with the bottom equation we see that $3x_3 = -6$ so clearly $x_3 = -2$. The middle equation implies that $2x_2 + 2x_3 = 4$ and it follows that $x_2 = 4$. The top equation states that $x_1 + 2x_2 + 4x_3 = 3$ and consequently $x_1 = 3$.*

Therefore we have found that the solution to the system of simultaneous equations

$$\begin{pmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 13 \\ 4 \end{pmatrix}$$

is $X = \begin{pmatrix} 3 \\ 4 \\ -2 \end{pmatrix}$.

Exercise 2.7.1.4. *Solve the system using LU decomposition*

$$\begin{pmatrix} 3 & 1 & 6 \\ -6 & 0 & -16 \\ 0 & 8 & -17 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 17 \end{pmatrix}.$$

2.7.2 Cholesky

*The **Cholesky decomposition** or Cholesky factorization is a decomposition of a **Hermitian, positive-definite** matrix into the product of a **lower triangular** matrix and its **conjugate transpose**.*
The Cholesky decomposition is roughly twice as efficient as the LU decomposition for solving systems of linear equations.
***Every symmetric, positive definite** matrix A can be decomposed into a product of a unique lower triangular matrix L and its transpose:*

$$A = LL^T.$$

The following formulas are obtained by solving above lower triangular matrix and its transpose. These are the basis of Cholesky Decomposition Algorithm :
Let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = LL^T = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{pmatrix}.$$

$$\Rightarrow \begin{cases} l_{11}^2 = a_{11} \Rightarrow l_{11} = \sqrt{a_{11}},\ l_{11}l_{21} = a_{12} \Rightarrow l_{12} = \frac{a_{12}}{l_{ll}},\ l_{11}l_{31} = a_{13} \Rightarrow l_{31} = \frac{a_{13}}{l_{11}} \\ l_{21}^2 + l_{22}^2 = a_{22} \Rightarrow l_{22} = \sqrt{a_{22} - l_{21}^2},\ l_{21}l_{31} + l_{22}l_{32} = a_{23} \Rightarrow l_{32} = \frac{a_{23} - l_{21}l_{31}}{l_{22}}, \\ l_{31}^2 + l_{32}^2 + l_{33}^2 = a_{33} \Rightarrow l_{33} = \sqrt{a_{33} - (l_{31}^2 + l_{32}^2}. \end{cases}$$

Generally, if A is an $n \times n$ symmetric positive definite matrix, it can be decomposed into $L * L^T$ by the following algorithm:

$$L_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{i-1} L_{jk)^2}}$$

$$L_{ij} = \frac{1}{L_{jj}}(a_{jj} - \sum_{k=1}^{i-1} L_{ik}L_{jk}).$$

Example 2.7.2.1. *Find the Cholesky decomposition of*

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

***solution**:*
From the beginning, we observe A is symmetric and diagonally dominant matrix which positive definite. Hence we can use Cholesky decomposition to decompose.

Let $L = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix}$ *be it's corresponding lower triangular matrix and from the above algorithm we have,*

$$\begin{aligned}
l_{11} &= \sqrt{a_{11}} = \sqrt{2}, \\
l_{21} &= \frac{a_{12}}{l_{11}} \Rightarrow l_{21} = \frac{-\sqrt{2}}{2}, \\
l_{22} &= \sqrt{a_{22} - l_{21}^2} \Rightarrow l_{22} = \sqrt{2 - \left(\tfrac{-\sqrt{2}}{2}\right)^2} = \sqrt{\frac{3}{2}}, \\
l_{31} &= \frac{a_{13}}{l_{11}} \Rightarrow l_{31} = \frac{0}{\sqrt{2}} = 0, \\
l_{32} &= \frac{a_{23} - l_{21}l_{31}}{l_{22}} \Rightarrow l_{32} = \frac{-1 + \frac{\sqrt{2}}{2}.0}{\sqrt{\frac{3}{2}}} = -\sqrt{\frac{2}{3}}, \\
l_{33} &= \sqrt{a_{33} - (l_{31}^2 + l_{32}^2)} \Rightarrow l_{33} = \sqrt{2 - 0 - \frac{2}{3}} = \sqrt{\frac{4}{3}}.
\end{aligned}$$

Thus,

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ \frac{-\sqrt{2}}{2} & \sqrt{\frac{3}{2}} & 0 \\ 0 & -\sqrt{\frac{2}{3}} & \sqrt{\frac{4}{3}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & \frac{-\sqrt{2}}{2} & 0 \\ 0 & \sqrt{\frac{3}{2}} & -\sqrt{\frac{2}{3}} \\ 0 & 0 & \sqrt{\frac{4}{3}} \end{pmatrix}.$$

2.7.3 The QR-Decomposition

The Gram-Schmidt formulas can be organized as matrix multiplication $A = QR$, *where* $\vec{a_1}, ..., \vec{a_n}$ *are the independent columns of* A, *and* Q *has columns equal to the Gram-Schmidt orthonormal vectors* $\vec{u_1}, ..., \vec{u_n}$, *which are the unitized Gram-Schmidt vectors.*
Square matrix
Any real square matrix A *may be decomposed as*

$$A = QR,$$

where Q *is an orthogonal matrix (its columns are orthogonal unit vectors meaning* $Q^TQ = QQ^T = I$ *and* R *is an upper triangular matrix (also called* ***right/upper triangular matrix****).*
If A *is invertible, then the factorization is unique if we require the diagonal elements of* R *to be positive.*
If instead A *is a complex square matrix, then there is a decomposition* $A = QR$ *where* Q *is a unitary matrix (so*

$$Q^*Q = QQ^* = I, \; Q^* = (\overline{Q})^T \text{ is the adjoint of } Q.$$

If A *has* n *linearly independent columns, then the first* n *columns of* Q *form an orthonormal basis for the column space of* A. *More generally, the first* k *columns of* Q *form an orthonormal basis for the span of the first* k *columns of* A *for any* $1 \leq k \leq n$. *The fact that any column* k *of* A *only depends on the first* k *columns of* Q *is responsible for the triangular form of* R.
More generally, we can factor a complex $m \times n$ *matrix* A, *with* $m \geq n$, *as the product of an* $m \times m$ *unitary matrix* Q *and an* $m \times n$ *upper triangular matrix* R. *As the bottom* $(m - n)$ *rows of an*

$m \times n$ upper triangular matrix consist entirely of zeroes, it is often useful to partition R, or both R and Q:

$$A = QR = Q\begin{bmatrix} R_1 \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}\begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1,$$

where R_1 is an $n \times n$ upper triangular matrix, 0 is an $(m-n) \times n$ zero matrix, Q_1 is $m \times n$, Q_2 is $m \times (m-n)$, and Q_1 and Q_2 both have orthogonal columns.
Golub and Van Loan (1996) call $Q_1 R_1$ the thin QR factorization of A; Trefethen and Bau call this the reduced QR factorization.
If A is of full rank n and we require that the diagonal elements of R_1 are positive then R_1 and Q_1 are unique, but in general Q_2 is not. R_1 is then equal to the upper triangular factor of the Cholesky decomposition of A^*A ($= A^T A$ if A is real).

Computing the QR decomposition *There are several methods for actually computing the QR decomposition, such as by means of the Gram–Schmidt process, Householder transformations, or Givens rotations. Each has a number of advantages and disadvantages.*

Using the Gram–Schmidt process *Consider the Gram–Schmidt process applied to the columns of the full column rank matrix*

$$A = [\mathbf{a}_1 \cdots \mathbf{a}_n],$$

with inner product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \mathbf{w}$$

(or

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^* \mathbf{w}$$

for the complex case).
By recalling projection of vectors:

$$\mathbf{u_1} = \mathbf{a_1}, \ \mathbf{e_1} = \frac{\mathbf{u_1}}{||\mathbf{u_1}||},$$
$$\mathbf{u_2} = \mathbf{a_2} - proj_{\mathbf{u_1}}^{\mathbf{a_2}} = \mathbf{a_2} - \frac{\langle \mathbf{a_2}, \mathbf{u_1} \rangle}{||\mathbf{u_1}||^2}\mathbf{u_1}, \ \mathbf{e_2} = \frac{\mathbf{u_2}}{||\mathbf{u_2}||},$$
$$\vdots$$
$$\mathbf{u_n} = \mathbf{a_n} - \sum_{j=1}^{n-1} proj_{\mathbf{u_j}}^{\mathbf{a_n}} = \mathbf{a_n} - \sum_{j=1}^{n-1} \frac{\langle \mathbf{a_n}, \mathbf{u_j} \rangle}{||\mathbf{u_j}||^2}\mathbf{u_j}, \ \mathbf{e_j} = \frac{\mathbf{u_j}}{||\mathbf{u_j}||}.$$

We can now express the $a_i^{'}s$ over our newly computed orthonormal basis:

$$\begin{aligned} \mathbf{a}_1 &= \langle \mathbf{e}_1, \mathbf{a}_1 \rangle \mathbf{e}_1 \\ \mathbf{a}_2 &= \langle \mathbf{e}_1, \mathbf{a}_2 \rangle \mathbf{e}_1 + \langle \mathbf{e}_2, \mathbf{a}_2 \rangle \mathbf{e}_2 \\ \mathbf{a}_3 &= \langle \mathbf{e}_1, \mathbf{a}_3 \rangle \mathbf{e}_1 + \langle \mathbf{e_2}, \mathbf{a_3} \rangle \mathbf{e}_2 + \langle \mathbf{e}_3, \mathbf{a}_3 \rangle \mathbf{e}_3 \\ &\vdots \\ \mathbf{a}_k &= \sum_{j=1}^{k} \langle \mathbf{e_j}, \mathbf{a_k} \rangle \mathbf{e_j} \end{aligned}$$

where

$$\langle \mathbf{e_i}, \mathbf{a_i} \rangle = \|\mathbf{u_i}\|.$$

This can be written in matrix form:

$$A = QR$$

where:

$$Q = \begin{pmatrix} \mathbf{e_1} & \mathbf{e_2} & \dots & \mathbf{e_n} \end{pmatrix}, \; R = \begin{pmatrix} \langle \mathbf{e_1}, \mathbf{a_1} \rangle & \langle \mathbf{e_1}, \mathbf{a_2} \rangle & \langle \mathbf{e_1}, \mathbf{a_3} \rangle & \dots \\ 0 & \langle \mathbf{e_2}, \mathbf{a_2} \rangle & \langle \mathbf{e_2}, \mathbf{a_3} \rangle & \dots \\ 0 & 0 & \langle \mathbf{e_3}, \mathbf{a_3} \rangle & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \mathbf{e_1}^T\mathbf{a_1} & \mathbf{e_1}^T\mathbf{a_2} & \dots & \mathbf{e_1}^T\mathbf{a_n} \\ 0 & \mathbf{e_2}^T\mathbf{a_2} & \dots & \mathbf{e_2}^T\mathbf{a_n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{e_n}^T\mathbf{a_n} \end{pmatrix}$$

Example 2.7.3.1. *Consider the decomposition of* $A = \begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix}$.
Recall that an orthonormal matrix Q has the property

$$Q^T Q = I.$$

Then, we can calculate Q by means of Gram–Schmidt as follows:

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix} = \begin{pmatrix} 12 & -69 & -58/5 \\ 6 & 158 & 6/5 \\ -4 & 30 & -33 \end{pmatrix};$$

$$Q = \begin{pmatrix} \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} & \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} & \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \end{pmatrix} = \begin{pmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{pmatrix}.$$

Thus, we have

$$Q^T A = Q^T Q\, R = R;$$

$$R = Q^T A = \begin{pmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{pmatrix}.$$

Advantages and disadvantages *The Gram-Schmidt process is inherently numerically unstable. While the application of the projections has an appealing geometric analogy to orthogonalization, the orthogonalization itself is prone to numerical error. A significant advantage however is the ease of implementation, which makes this a useful algorithm to use for prototyping if a pre-built linear algebra library is unavailable.*

Example 2.7.3.2. *Compute the QR factorization of the matrix*

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

.

Solution: *Consider the matrix with the vectors* $\mathbf{a_1} = (1,1,0)^T$, $\mathbf{a_2}^T = (1,0,1)$, $\mathbf{a_3}^T = (0,1,1)$.

Performing the Gram-Schmidt procedure, we obtain:

$$\mathbf{u_1} = a_{11} = (1,1,0)^T,\ \mathbf{e_1} = \frac{\mathbf{u_1}}{||\mathbf{u_1}||} = \frac{1}{\sqrt{2}}(1,1,0),$$

$$\mathbf{u_2} = \mathbf{a_2} - \frac{\langle \mathbf{a_2}, \mathbf{u_1}\rangle}{||\mathbf{u_1}||^2}\mathbf{u_1} = (1,0,1) - \frac{1}{\sqrt{2}}\left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0)\right) = \left(\tfrac{1}{2}, -\tfrac{1}{2}, 1\right),$$

$$\mathbf{e_2} = \frac{\mathbf{u_2}}{||\mathbf{u_2}||} = \left(\tfrac{1}{\sqrt{6}}, -\tfrac{1}{\sqrt{6}}, \tfrac{2}{\sqrt{6}}\right),$$

$$\mathbf{u_3} = \mathbf{a_3} - (\mathbf{a_3}, \mathbf{e_1})\mathbf{e_1} - (\mathbf{a_3}, \mathbf{e_2})\mathbf{e_2} = (0,1,1) - \frac{1}{\sqrt{2}}\left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0\right) - \frac{1}{\sqrt{6}}\left(\tfrac{1}{\sqrt{6}}, -\tfrac{1}{\sqrt{6}}, \tfrac{2}{\sqrt{6}}\right) = \left(-\tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}\right),$$

$$\mathbf{e_3} = \frac{\mathbf{u_3}}{||\mathbf{u_3}||} = \left(-\tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}\right).$$

Thus,

$$Q = \left[\mathbf{e_1}|\mathbf{e_2}|\mathbf{e_3}\right] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$R = \begin{bmatrix} \mathbf{a_1}.\mathbf{e_1} & \mathbf{a_2}.\mathbf{e_1} & \mathbf{a_3}.\mathbf{e_1} \\ 0 & \mathbf{a_2}.\mathbf{e_2} & \mathbf{a_3}.\mathbf{e_2} \\ 0 & 0 & \mathbf{a_3}.\mathbf{e_3} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{3}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{2}{\sqrt{3}} \end{bmatrix}.$$

Exercise 2.7.3.1. *Compute the QR factorization for the matrix*

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

(Ans.)

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{pmatrix} \text{ and } R = \begin{pmatrix} \sqrt{2} & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \frac{\sqrt{6}}{2} \end{pmatrix}.$$

2.7.4 Singular value decomposition

The decomposition has been used as a data compression algorithm. A geometric interpretation will be given in the next subsection.
If M is an $m \times n$ matrix, then we may write A as a product of three factors:

$$M = U\Sigma V^*, \tag{2.1}$$

where U is an orthogonal/unitary $m \times m$ matrix, V is an orthogonal/unitary $n \times n$ matrix, V^* is the conjugate transpose of V, and Σ is an $m \times n$ matrix, then M^TM is a real symmetric matrix whose eigen pairs $(\lambda, \mathbf{v})$ satisfy

$$\lambda = \frac{||M\mathbf{v}||^2}{||\mathbf{v}||^2} \geq 0. \tag{2.2}$$

If the real symmetric matrix M^TM the eigenvalues

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_> 0 = \lambda_{r+1} = ... = \lambda_n.$$

The numbers

$$\sigma_k = \sqrt{\lambda_k},\ 1 \le k \le n,$$

*are called **singular values** of the matrix M. The ordering of the singular values is always with decreasing magnitude. The diagonal matrix, Σ, is uniquely determined by M (though not the matrices U and V if M is not square, see below).*
In the special, yet common case when M is an $m \times m$ real square matrix with positive determinant: U, V^*, and Σ are real $m \times m$ matrices as well. Σ can be regarded as a scaling matrix, and U, V^* can be viewed as rotation matrices.
*Thus, the expression $U\Sigma V^*$ can be intuitively interpreted as a composition of three geometrical transformations: a **rotation or reflection**, a **scaling**, and another **rotation or reflection**.*

Theorem 2.7.4.1. *(The singular value decomposition (svd))*
Let M be a real $m \times n$ matrix, $(\lambda_1, \mathbf{v_1}), ..., (\lambda_n, \mathbf{v_n})$ be a set of orthonormal eigenpairs for M^TM such that $\sigma_k = \sqrt{\lambda_k},\ 1 \le k \le r$ defines the positive singular values of A and $\lambda_k = 0$ for $r < k \le n$. Complete $\mathbf{u_1} = \frac{1}{\sigma_1}M\mathbf{v_1},\ ...,\ \mathbf{u_r} = \frac{1}{\sigma_r}M\mathbf{v_r}$ to an orthonormal basis $\{\mathbf{u_k}_{k=1}^m\}$ for $\mathbb{R}^m$. Then the columns of U and V are orthonormal and

$$\underbrace{\mathbf{M}}_{W\times D} = \underbrace{\mathbf{U}}_{W\times W} \times \underbrace{\mathbf{\Sigma}}_{W\times D} \times \underbrace{\mathbf{V}^T}_{D\times D} = \begin{pmatrix} \sigma_1 & & & \\ & \cdot & 0 & \\ & 0 & \cdot \ \ \sigma_r & \\ & & & 0 \end{pmatrix} = U\Sigma V^T$$

$$= \underbrace{\begin{bmatrix}\vec{u_1} & \mathbf{u}_2 & \dots & \vec{u_r}\end{bmatrix}}_{\text{Col}\,M}\ \underbrace{\begin{matrix}\vec{u_{r+1}} & \dots & \vec{u_m}\end{matrix}}_{\text{Nul}\,M^T} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \dots & & & & & & \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & & & & & & \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \vec{v_1}^T \\ \vec{v_2}^T \\ \dots \\ \vec{v_r}^T \\ \vec{v_{r+1}}^T \\ \dots \\ \vec{v_n}^T \end{bmatrix} \begin{matrix} \left.\right\} \text{Row}\,M \\ \left.\right\} \text{Nul}\,M \end{matrix}$$

$$= \sum_{i=1}^{r} \sigma_i \mathbf{u_i}\mathbf{v_i^T} = M\sum_{i=1}^{r} \mathbf{v_i}\mathbf{v_i^T}.$$

Example 2.7.4.1. *Find the singular value decomposition of*

$$A = \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix}$$

The eigenvalues of

$$A^TA = \begin{pmatrix} 5 & 3 \\ 8 & 3 \end{pmatrix}$$

are 2, 8 and their corresponding unit eigenvectors are

$$\vec{v_1} = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix},\ and\ \vec{v_2} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

respectively. Hence the singular values are $\sigma_1 = \sqrt{\lambda_1} = \sqrt{2}$, and $\sigma_2 = \sqrt{\lambda_2} = \sqrt{8} = 2\sqrt{2}$.
We have

$$A\mathbf{v_1} = \sigma_1\mathbf{u_1}\mathbf{v_1}^T\mathbf{v_1} = \sigma_1\mathbf{u_1} = \begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix},\ so\ \mathbf{u_1} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$A\mathbf{v_2} = \sigma_2 \mathbf{u_2} \mathbf{v_2}^T \mathbf{v_2} = \sigma_2 \mathbf{u_2} = \begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}, \ \textit{so}\ \mathbf{u_2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The SVD of A is therefore

$$A = U \sum V^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Exercise 2.7.4.2. *:*

(1) *Show that if the matrix A is square and symmetric then the singular values of A are the absolute values of the eigenvalues of A and that the singular vectors are eigenvectors of A.*

(2) *Show that $\|A\|_2 = \sigma_1$. If A is square and A^{-1} exists, then $\|A^{-1}\|_2 = \sigma_n^{-1}$. Show that the condition number of A (for the 2-norm) is σ_1/σ_n.*

(3) *"Regularization" is a way to compromise between accurate solution of a least squares problem and the size of the solution. Regularization is used to improve the conditioning of ill conditioned least squares problems. The simplest regularization strategy is to replace the least squares problem*

$$\min_x \|Ax - b\|_2$$

with the "regularized" problem

$$\min_x \|Ax - b\|_2^2 + \epsilon \|x\|_2^2$$

Give an algorithm to solve the regularized least squares problem, (2), using the SVD of A. Your algorithm will involve $1/(\sigma_i + \epsilon)$. Does this make dimensional sense?

(4) *Use the SVD to show that the non zero eigenvalues of A^*A are identical to the non zero eigenvalues of AA^*. Note that these matrices have different dimensions so they have a different number of eigenvalues.*

(5) *The "polar decomposition" of a square matrix, A, is a decomposition $A = RQ$ where R is symmetric and positive semi-definite and Q is orthogonal. This is a generalization of the polar decomposition of a complex number $z = re^{i\theta}$ into a product of a non negative real number and a complex number of magnitude 1. The polar decomposition of matrices is used in elasticity theory. Show how to find the polar decomposition of a matrix from its SVD.*