Unit 5
Inner Product Spaces and Orthogonal Vectors

5.1 Introduction

In previous units we have looked at general vector spaces defined by a very small number of defining axioms, and have looked at many specific examples of such vector spaces. However, in our usual view of the space in which we live we always think of it having certain properties such as distance and angles that do not occur in many vector spaces. In this Unit we investigate vector spaces that have one additional property, an inner product, that allows us to define distance, angle and other properties and we call these *inner product spaces*. We have already studied some special cases of these in the Euclidean Vector Spaces,, R^n, for $n = 1,\ 2,\ \cdots$, but in this unit we show how the same results apply to other vector spaces. In particular the Euclidean space aspect is covered in the subsection ”Geometry of Linear Transformations Between R^2, R^3 and R” of Unit 3, Section 3: Linear Transformations from R^n to R^m.

The approach in this section follows a very important and powerful mathematical approach in which a very general concept, such as inner product spaces, is developed from only a very few basic defining properties, or axioms. The properties developed for the general concept apply to a wide variety of specific and often very different-looking manifestations. That is, any theorem or property that is proved using the defining axioms of an inner product space, will then apply to every specific example or manifestation of inner product space.

In particular, theorems like that of Pythagoras (sum of squares of the two sides adjacent to a right angle in a triangle is equal to the square of the length of the hypotenuse) are shown to apply in all inner product spaces. A method for creating an orthogonal basis (any two basis vectors are orthogonal to each other), called the Gram-Schmidt process, is developed for all inner product spaces. Orthonormal bases (orthogonal bases of unit vectors) were previously shown to be important in using eigenvectors/eigenvalue to compute powers of a matrix (see Unit 4, Section 5: Diagonalizing a Matrix). A method called *least squares approximation* is shown to apply to all inner product spaces. Least squares approximation has many important applications in Euclidean vector spaces, and in finding best approximations of data sets, such as linear regression.

The topics are:

- Basic definitions and properties of inner product spaces
- Constructing an orthonormal basis, using the Gram-Schmidt process, and applications

- Least squares approximation and applications

5.2 Learning objectives

Upon completion of this unit you should be able to:

- write down the defining axioms of an inner product space;
- define and give properties satisfied by basic concepts of an inner product space, such as angle, orthogonality, length/distance, norm, orthogonal complement;
- write down and prove a number of basic results that are true in inner product spaces, such as the triangle inequality, Cauchy-Schwarz inequality, Pythagoras' Theorem, the parallelogram theorem;
- describe and analyse a number of specific examples of inner product spaces;
- describe the Gram-Schmidt process for creating an orthogonal, or orthonormal, basis from any other basis of an inner product space;
- apply the Gram-Schmidt process to find an orthogonal/orthonormal basis from any given basis of a specific inner product spaces;
- show how an orthogonal basis allows properties of vectors to be easily calculated including coordinates, norms, inner products, distances, orthogonal projections;
- show how the Gram-Schmidt process is equivalent to finding a $QR-$ decomposition of a certain matrix;
- calculate the $QR-$ decomposition of a matrix;
- explain the basic concept of "best approximation" of a vector in terms of projections in an inner product space;
- explain how the "best approximation" theory is applied to find an approximate solution, called the "least squares solution", of a system of linear equations $A\mathbf{x} = \mathbf{b}$ that has no exact solution;
- show how the least squares solution of the linear system $A\mathbf{x} = \mathbf{b}$ is given by $\mathbf{x} = \left(A^T A\right)^{-1} A^T\mathbf{b}$, provided the columns of A are linearly independent; and
- derive least squares solutions for a variety of practical problems.

5.3 Assigned readings

- Section 5.5, read sections 6.1 and 6.2 in your textbook.
- Section 5.6, read sections 6.3, 6.5 and 6.6 in your textbook.
- Section 5.7, read section 6.4 in your textbook.

5.4 Unit activities

1. Read each section in the unit and carefully work through each illustrative example. Make sure you understand each concept, process, or Theorem and how it is used. Add all key points to your personal course summary sheets.

   Work through **on your own** all examples and exercises throughout the unit to check your understanding of each concept.

2. Read through the corresponding sections in your textbook and work through the sample problems and exercises.

3. If you have difficulty with a problem after giving it a serious attempt, check the discussion topic for this unit to see if others are having similar problems. The link to the discussion area is found in the left hand menu of your course. If you are still having difficulty with the problem then ask your instructor for assistance.

4. After completing the unit, review the learning objectives again. Make sure that you are familiar with each objective and understand the meaning and scope of the objective.

5. Review the course content in preparation for the final examination.

6. Complete the online course evaluation.

5.5 Basic definitions and properties of inner product spaces

An inner product space is any vector space that has, in addition, a special kind of function, called the inner product. The inner product computes a real number from any two vectors, in a way similar to the previously-encountered scalar product in Euclidean spaces R^n. The inner product function has linearity properties, is commutative, and the inner product of a non-zero vector with itself is a positive number. These properties are exactly the properties satisfied by the scalar product (dot product) of vectors in a Euclidean vector space, and so Euclidean spaces R^2, R^3 and more generally R^n, with the scalar product, are already inner product spaces. In the Euclidean spaces our intuitive idea of distance between two points/vectors $\mathbf{u}, \mathbf{v}$ is given by the scalar product $\sqrt{(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})} = \|\mathbf{u}-\mathbf{v}\|$, and other concepts such as perpendicular vectors and the angle between vectors are defined in terms of the scalar product. Analogously, an inner product can be used to define a distances, perpendiculartiy, and angles in any inner product space as shown in this section. Many theorems of Euclidean spaces, depending on these concepts, hold true in inner product spaces, such as the well known Pythagoras' Theorem for right angle triangles. A number of examples of inner product spaces are given in this section. The following sections develop more extensive applications for **orthogonal bases** and **least squares approximations**.

5.5.1 Definition of inner product space

Definition.
In a vector space V an **inner product** is a function, written as $\langle \mathbf{u}, \mathbf{v} \rangle$, for any two vectors $\mathbf{u}, \mathbf{v} \in V$, satisfying the following five axioms (properties):

(1) $\langle \mathbf{u}, \mathbf{v} \rangle$ is a real number for every $\mathbf{u}, \mathbf{v} \in V$ (real number axiom).

(2) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ for every $\mathbf{u}, \mathbf{v} \in V$ (symmetry or commutative axiom).

(3) $\langle \mathbf{u}+\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ for every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ (additive linearity axiom for the first variable).

(4) $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$ for every $\mathbf{u}, \mathbf{v} \in V$ and every $k \in R$ (homogenity or scalar linearity axiom for the first variable).

(5) $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ for every $\mathbf{v} \neq 0 \in V$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if $\mathbf{v} = 0$ (positivity axiom).

Note: The symmetry axiom (2) shows that the linearity of axioms (3) and (4) also applies to the second variable. That is, for every $\mathbf{w}, \mathbf{u}, \mathbf{v} \in V$, and every $k \in R$:

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

Note: Combining properties (3) and (4) shows that it is also true that it is fully linear on the first variable:

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

Using the note above it also follows that a similar result holds for the second variable:

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

Note: Setting $k = 1$ and $l = -1$ in the result immediately above shows that it is also true that:

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

Definition.
A real vector space that has an inner product is called an **inner product space**.

Example 5.5.1.
This example has two parts:

(a) In R^2, for any two vectors $\mathbf{u} = (u_1,\ u_2)$ and $\mathbf{v} = (v_1,\ v_2)$ we have previously defined the scalar product $\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2$, which is a real number. Show that $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$ is an inner product, called the **Euclidean inner product**.

(b) Similarly in R^n, for any $n \geq 1$ show that the scalar product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$ is an inner product.

Solution. It is left as an exercise for the reader to show that the scalar product satisfies all five axioms of the definition. If you have difficulty with this then please consult your instructor or textbook.

An inner product space can have more than one choice of inner product, as the next example shows for Euclidean vector spaces. This means that there are other ways to define "distance" in Euclidean spaces that are different than our normal definition of distance.

Example 5.5.2.
Show that each of the following $\langle \mathbf{u}, \mathbf{v} \rangle$ is an inner product (the first three are called **weighted Euclidean inner products**):

(a) In R^2, for any two vectors $\mathbf{u} = (u_1,\ u_2)$ and $\mathbf{v} = (v_1,\ v_2)$ define $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + 2u_2v_2$.

(b) In R^2, define $\langle \mathbf{u}, \mathbf{v} \rangle = au_1v_1 + bu_2v_2$, where $a, b \in R$ are any two positive numbers.

(c) In R^n, for a fixed $n \geq 1$ define $\langle \mathbf{u}, \mathbf{v} \rangle = a_1u_1v_1 + a_2u_2v_2 + \cdots + a_nu_nv_n$, where $a_1,\ a_2,\ \cdots\ a_n$ are any non-negative real numbers.

(d) In R^2, for any two vectors $\mathbf{u} = (u_1,\ u_2)$ and $\mathbf{v} = (v_1,\ v_2)$ define $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 - u_1v_2 - u_2v_1 + 2u_2v_2$.

Solution. We show here the proofs for parts (b) and (d). The proofs for parts (a) and (c) are similar and are left as an exercise for the reader.

(b) Suppose $\mathbf{u} = (u_1,\ u_2)\,,\ \mathbf{v} = (v_1,\ v_2)\,,\ \mathbf{w} = (w_1,\ w_2)$ are any three vectors in R^2.
Proof of axiom (1):
$au_1v_1 + bu_2v_2$ is clearly a real number since it consists of products and sums of real numbers a, b, u_1, u_2, v_1, v_2.
Proof of axiom (2):
Using the definition $\langle \mathbf{u}, \mathbf{v} \rangle = au_1v_1 + bu_2v_2$ and $\langle \mathbf{v}, \mathbf{u} \rangle = av_1u_1 + bv_2u_2$ but $au_1v_1 + bu_2v_2 = av_1u_1 + bv_2u_2$ since real number multiplication is commutative. Hence, $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.
Proof of axiom (3):

$$\begin{aligned}\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= \langle (u_1 + v_1,\ u_2 + v_2)\,, (w_1,\ w_2) \rangle \\ &= a\,(u_1 + v_1)\,w_1 + b\,(u_2 + v_2)\,w_2 \\ &= (au_1w_1 + bu_2w_2) + (av_1w_1 + bv_2w_2) \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle\end{aligned}$$

Proof of axiom (4):

$$\begin{aligned}\langle k\mathbf{u},\mathbf{v}\rangle &= \langle (ku_1,\ ku_2),(v_1,\ v_2)\rangle \\ &= aku_1v_1 + bku_2v_2 \\ &= k(au_1v_1 + bu_2v_2) \\ &= k\langle \mathbf{u},\mathbf{v}\rangle\end{aligned}$$

Proof of axiom (5):

$$\langle \mathbf{v},\mathbf{v}\rangle = av_1^2 + bv_2^2$$

and $av_1^2 + bv_2^2 \geq 0$ since $a > 0,\ ,b > 0,\ v_1^2 \geq 0,\ v_2^2 \geq 0$. That is, this inner product, being composed of products and and sums of non-negative numbers, is also non-negative and so satisfies:

$$\langle \mathbf{v},\mathbf{v}\rangle = av_1^2 + bv_2^2 \geq 0$$

Furthermore, the only way that $\langle \mathbf{v},\mathbf{v}\rangle = 0$ is if $v_1 = v_2 = 0$ when $\mathbf{v} = (v_1,\ v_2)$ is the zero vector.

(d) Proof of axiom (1):
$\langle \mathbf{u},\mathbf{v}\rangle = 2u_1v_1 - u_1v_2 - u_2v_1 + 2u_2v_2$ is clearly a real number.
Proofs of axioms (2), (3), (4):
These proofs are left as an exercise for the reader and should present no problems.
Proof of axiom (5):
$\langle \mathbf{v},\mathbf{v}\rangle = 2v_1^2 - 2v_1v_2 + 2v_2^2 = (v_1 - v_2)^2 + v_1^2 + v_2^2$. Hence:

$$\langle \mathbf{v},\mathbf{v}\rangle = (v_1 - v_2)^2 + v_1^2 + v_2^2 \geq 0$$

and $\langle \mathbf{v},\mathbf{v}\rangle = 0$ only if $v_1 = v_2 = 0$ in which case $\mathbf{v} = \mathbf{0}$.

Example 5.5.3.
In R^2 let any two vectors be given by $\mathbf{u} = (u_1,\ u_2)$ and $\mathbf{v} = (v_1,\ v_2)$. Show that each of the following **is not** an inner product in R^2 :

(a) $\langle \mathbf{u},\mathbf{v}\rangle = 2u_1v_1 - 3u_2v_2$

(b) $\langle \mathbf{u},\mathbf{v}\rangle = u_1v_1$

(c) $\langle \mathbf{u},\mathbf{v}\rangle = \sqrt{u_1v_1 + u_2v_2}$

(d) $\langle \mathbf{u},\mathbf{v}\rangle = u_1^2v_1^2 + u_2^2v_2^2$

Note: In order to show that a function $\langle \mathbf{u},\mathbf{v}\rangle$ is not an inner product it is only necessary to find two specific vectors $\mathbf{u},\mathbf{v}$f or which one of the five axioms fails to hold.

Solution. It is left for the reader to show that each one fails to satisfy at least one axiom of the definition as follows:

Show that (a) does not satisfy the positivity axiom 5 (for example, when $\mathbf{u} = (0,\ 1),\ \mathbf{v} = (1,1)$). In addition it does not satisfy axiom (2).

Show that (b) does not satisfy axiom 5 (for example, if $\mathbf{u} = (0,\ 1),\ \mathbf{v} = (1,1)$ - but for a different reason than in part (a).

Show that (c) does not satisfy axiom (1) because it is not even defined as a real number for some choices of vectors $\mathbf{u},\mathbf{v}$. In addition, even when it is defined axioms 3 and 4 are not satisfied.

Show that (d) does not satisfy axiom (3). In additon it does not satisfy axiom (4).

Example 5.5.4.
In R^n show that any $n \times n$ real non-singular matrix A generates an inner product defined in terms of the Euclidean inner product by:

$$\langle \mathbf{u},\ \mathbf{v} \rangle = A\mathbf{u} \cdot A\mathbf{v}$$

Note: This can be re-written in the standard way as a matrix product:

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

Solution. Proving each axiom:
Axiom 1 (it is a real number) is clearly satisfied.
Axiom 2: $\langle \mathbf{u},\ \mathbf{v} \rangle = \mathbf{u}^T A^T A\mathbf{v}$ and $\langle \mathbf{v},\ \mathbf{u} \rangle = \mathbf{v}^T A^T A\mathbf{u}$ appear to be different, but are in fact the same. This is because $\mathbf{u}^T A^T A\mathbf{v}$ is a real number and so transposing it does not change its value. Hence, transposing, using the usual matrix formula that the individual parts of the product are transposed in reverse order:

$$\begin{aligned} \mathbf{v}^T A^T A\mathbf{u} &= \left(\mathbf{v}^T A^T A\mathbf{u}\right)^T \\ &= \mathbf{u}^T A^T \left(A^T\right)^T \left(\mathbf{v}^T\right)^T \\ &= \mathbf{u}^T A^T A\ \mathbf{v} \end{aligned}$$

Axiom 3: For any $\mathbf{u},\ \mathbf{v},\ \mathbf{w} \in R^n$:

$$\begin{aligned} \langle \mathbf{u}+\mathbf{v}, \mathbf{w} \rangle &= (\mathbf{u}+\mathbf{v})^T A^T A\ \mathbf{w} \\ &= \left(\mathbf{u}^T + \mathbf{v}^T\right) A^T A\ \mathbf{w} \\ &= \mathbf{u}^T A^T A\mathbf{w} + \mathbf{v}^T A^T A\ \mathbf{w} \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

Axiom 4: For any $\mathbf{u},\ \mathbf{v} \in R^n$ and $k \in R$:

$$\begin{aligned} \langle k\mathbf{u}, \mathbf{v} \rangle &= (k\mathbf{u})^T A^T A\ \mathbf{v} \\ &= k\mathbf{u}^T A^T A\ \mathbf{v} \\ &= k \langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

Axiom 5: For any $\mathbf{v} \in R^n$

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v} \rangle &= \mathbf{v}^T A^T A\mathbf{v} \\ &= (A\mathbf{v})^T A\mathbf{v} \\ &= \|A\mathbf{v}\|^2 \text{ (the usual Euclidean norm)} \end{aligned}$$

Since A is non-singular it follows that $A\mathbf{v} = 0$ only when $\mathbf{v} = \mathbf{0}$, and so $\|A\mathbf{v}\| > 0$. when $\mathbf{v} \neq \mathbf{0}$. Hence:

$$\begin{aligned} &\langle \mathbf{v}, \mathbf{v} \rangle > 0 \text{ when } \mathbf{v} \neq \mathbf{0} \\ &\langle \mathbf{v}, \mathbf{v} \rangle = 0 \text{ when } \mathbf{v} = \mathbf{0} \end{aligned}$$

Theorem 5.1. *An inner product exists in every finite dimensional vector space V. That is, every finite dimensional vector space can be made into an inner product space.*

Proof. Suppose V has dimension n and has a basis $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \cdots, \mathbf{b}_n\}$. For any two vectors $\mathbf{u},\ \mathbf{v} \in V$ suppose that the unique linear combinations of the basis vectors are:

$$\mathbf{u} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2 + k_3\mathbf{b}_3 + \cdots k_n\mathbf{b}_n$$
$$\mathbf{v} = l_1\mathbf{b}_1 + l_2\mathbf{b}_2 + l_3\mathbf{b}_3 + \cdots l_n\mathbf{b}_n$$

Define the inner product as the scalar product (Euclidean inner product) of the coordinates of the two vectors:

$$\langle \mathbf{u}, \mathbf{v} \rangle = k_1 l_1 + k_2 l_2 + k_3 l_3 + \cdots + k_n l_n$$

This is clearliy a real number, so Axiom 1 holds. It is easy to show that $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$, $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle k\mathbf{u}, \mathbf{v} \rangle = k \langle \mathbf{u}, \mathbf{v} \rangle$, so Axioms (2), (3), (4) hold (verify these for yourself). For Axiom 5, $\langle \mathbf{u}, \mathbf{u} \rangle = (k_1)^2 + (k_2)^2 + (k_3)^2 + \cdots + (k_n)^2 \geq 0$ and clearly $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ only if $k_1 = k_2 = k_3 = \cdots = k_n = 0$, which means $\mathbf{u} = \mathbf{0}$.

□

Theorem 5.2. *If V is an inner product space, and S is a subspace of V then S is also an inner product space with the same inner product function as V.*

Proof. This is left as an exercise for the reader. Convince yourself that all five axioms for the inner product of V will also be true in the subspace S. □

Example 5.5.5.
In R^3 find a formula for the inner product induced, as in Theorem 5.1, by the basis $B = \{(1,0,0),\ (1,2,0),\ (1,1,1)\}$.

Solution. Put the basis vectors as the columns of the matrix P. To express a vector $\mathbf{v} = (x,\ y,\ z)$, with respect to the standard basis, in terms the basis B we need to find a vector multiplying P on the right that gives the vector $\mathbf{v}$. That is, using column matrices for vectors, the required coordinates, X, Y, Z, for the basis B satisfy:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \overset{P}{\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \implies \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \overset{P^{-1}}{\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}^{-1}} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x - \frac{y}{2} - \frac{z}{2} \\ \frac{y}{2} - \frac{z}{2} \\ z \end{bmatrix}$$

Hence, using this formula for the coordinates, the induced inner product for two vectors $\mathbf{u} = (u_1, u_2, u_3),\ \mathbf{v} = (v_1, v_2, v_3)$ is:

$$\begin{aligned} \langle \mathbf{u},\ \mathbf{v} \rangle &= \left(u_1 - \frac{u_2}{2} - \frac{u_3}{2},\ \frac{u_2}{2} - \frac{u_3}{2},\ u_3\right) \cdot \left(v_1 - \frac{v_2}{2} - \frac{v_3}{2},\ \frac{v_2}{2} - \frac{v_3}{2},\ v_3\right) \\ &= \left(u_1 - \frac{u_2}{2} - \frac{u_3}{2}\right)\left(v_1 - \frac{v_2}{2} - \frac{v_3}{2}\right) + \left(\frac{u_2}{2} - \frac{u_3}{2}\right)\left(\frac{v_2}{2} - \frac{v_3}{2}\right) + u_3 v_3 \\ \langle \mathbf{u},\ \mathbf{v} \rangle &= u_1 v_1 - \frac{1}{2}u_1 v_2 - \frac{1}{2}u_2 v_1 - \frac{1}{2}u_1 v_3 + \frac{1}{2}u_2 v_2 - \frac{1}{2}u_3 v_1 + \frac{3}{2}u_3 v_3 \end{aligned}$$

5.5.2 Norm, length, distance, angle, and projections

Definition.
In an inner product space with inner product $\langle \mathbf{u}, \mathbf{v} \rangle$, define:

1. the *norm* or *length* of a vector $\mathbf{v}$ as:
$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

2. the *distance* $d(\mathbf{u},\ \mathbf{v})$ between the two point/vectors $\mathbf{u},\ \mathbf{v}$ as the length of the vector $\mathbf{u}-\mathbf{v}$, namely:
$$d(\mathbf{u},\ \mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| = \sqrt{\langle \mathbf{u}-\mathbf{v},\ \mathbf{u}-\mathbf{v} \rangle}$$

3. the *angle* θ between two non-zero vectors as the angle $0 \leq \theta \leq \pi$ satisfying:
$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|}$$

Note: The angle definition is analogous to the formula found for angles in a Euclidean space defined in terms of the scalar product. That is, we previously saw the scalar product formula, $\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$, where θ is the angle between the Euclidean vectors $\mathbf{u},\ \mathbf{v}$. Thus gives the analogous Euclidean space formula:
$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}$$
This is the same as the inner product formula above because $\mathbf{u}\cdot\mathbf{v}$ is an inner product (see Example 5.5.1).
Note: For any angle θ the cosine function satisfies $-1 \leq \cos\theta \leq 1$. Hence, the above formula for angles in an inner product space can only make sense if $-1 \leq \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1$ for every pair of vectors $\mathbf{u}, \mathbf{v}$ in an inner product space. This result is true and it is known as the Cauchy-Schwarz inequality, described next in Theorem 5.3.

Theorem 5.3. *The **Cauchy-Scwarz** inequality. For any two vectors $\mathbf{u}, \mathbf{v}$ in an inner product space the inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ satisfies:*
$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$$

Note: The inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ is a positive or negative real number so $|\langle \mathbf{u}, \mathbf{v} \rangle|$ means the absolute value of $\langle \mathbf{u}, \mathbf{v} \rangle$, whereas $\mathbf{u}, \mathbf{v}$ are vectors and so $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$, $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ are the norms or lengths of the vectors.

Proof. The proof is short but non-intuitive, and so is not given here. The proof may be found in most textbooks. □

Note: Since $|\langle \mathbf{u}, \mathbf{v} \rangle| = \langle \mathbf{u}, \mathbf{v} \rangle$ or $|\langle \mathbf{u}, \mathbf{v} \rangle| = -\langle \mathbf{u}, \mathbf{v} \rangle$ (whichever is positve), then the theorem can be restated:
$$\pm\langle \mathbf{u}, \mathbf{v} \rangle \leq \|\mathbf{u}\|\,\|\mathbf{v}\| \Longrightarrow \begin{cases} \langle \mathbf{u}, \mathbf{v} \rangle \leq \|\mathbf{u}\|\,\|\mathbf{v}\| \\ -\langle \mathbf{u}, \mathbf{v} \rangle \leq \|\mathbf{u}\|\,\|\mathbf{v}\| \end{cases} \Longrightarrow \begin{cases} \langle \mathbf{u}, \mathbf{v} \rangle \leq \|\mathbf{u}\|\,\|\mathbf{v}\| \\ \langle \mathbf{u}, \mathbf{v} \rangle \geq -\|\mathbf{u}\|\,\|\mathbf{v}\| \end{cases}$$
since multiplying an inequality by a negative number reverses its direction. Hence, the theorem is equivalent to:
$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \leq 1$$
This justifies the definition of angle θ between vectors given by $\cos\theta = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}$ (since $\cos\theta$ assumes every value between -1 and 1 for just one value of θ with $0 \leq \theta \leq \pi$, often written in terms of the inverse cosine formula, $\theta = \arccos\left(\frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}\right)$).

Definition.
Two non-zero vectors $\mathbf{u}, \mathbf{v}$ in an inner product space are said to be **perpendicular** if the angle θ between them is $\theta = \frac{\pi}{2}$ radians (90 degrees), which is equivalent to $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ (since for angles between 0 and π, $\cos\theta = 0 \Longleftrightarrow \theta = \frac{\pi}{2}$ and therefore $\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta = 0 \Longleftrightarrow \theta = \frac{\pi}{2}$).

Theorem 5.4. *The norm or length of a vector in an inner product space satisfies the normal properties that we expect of length and distance. That is, for any two vectors* $\mathbf{u}, \mathbf{v}$ *of an inner product space:*

(a) $\|\mathbf{v}\| \geq 0$ *and* $\|\mathbf{v}\| = 0$ *only if* $\mathbf{v} = \mathbf{0}$ *(the zero vector).*

(b) $\|k\mathbf{v}\| = |k| \, \|\mathbf{v}\|$ *for any* $k \in R$. *- multiplying a vector by a scalar changes the length of the vector by the positive value of that scalar.*

(c) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ *- sometimes called the* ***triangle inequality****. That is, the sum of two vectors cannot be longer than the lengths of the two individual vectors.*

Proof. The following outlines the proof:

(a) See the exercise set.

(b) For any $k \in R$:

$$\begin{aligned} \|k\mathbf{v}\| &= \sqrt{\langle k\mathbf{v},\ k\mathbf{v}\rangle} = \sqrt{k\,\langle \mathbf{v},\ k\mathbf{v}\rangle} \text{ by Axiom 4} \\ &= \sqrt{k^2\,\langle \mathbf{v},\ \mathbf{v}\rangle} \text{ by Axioms 2, 3, 4 (see Note after Axioms)} \\ &= |k|\,\sqrt{\langle \mathbf{v},\ \mathbf{v}\rangle} \text{ since } \sqrt{k^2} = |k| \text{ for any } k \in R \end{aligned}$$

(c) Starting with the square of the left side and using the axioms of inner products:

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

Using the Cauchy-Schwarz inequality formula: $\langle \mathbf{u}, \mathbf{v}\rangle \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$ this becomes:

$$\begin{aligned} &\|\mathbf{u} + \mathbf{v}\|^2 \leq \|\mathbf{u}\|^2 + 2\,\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 \implies \\ &\|\mathbf{u} + \mathbf{v}\|^2 \leq (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \end{aligned}$$

Taking square roots of both sides gives the result.

□

Example 5.5.6.
In R^3 let $\mathbf{u} = (1,\ 0,\ 0)$ and $\mathbf{v} = (2,\ -1,\ 3)$. Define the inner product by $\langle \mathbf{u},\ \mathbf{v}\rangle = A\mathbf{u} \cdot A\mathbf{v} = \mathbf{u}^T A^T A\mathbf{v}$ where A is the matrix:

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & -1 \\ 2 & 0 & 2 \end{bmatrix}$$

(a) Compute $\langle \mathbf{u},\ \mathbf{v}\rangle$.

(b) Compute the norms $\|\mathbf{u}\|, \|\mathbf{v}\|$.

(c) Find all vectors $\mathbf{w}$ perpendicular to $\mathbf{u}$.

(d) Find the equation satisfied by all vectors $\mathbf{w} = (x,\ y,\ z)$ with $\|\mathbf{w}\| = 1$.

Solution. The following outlines the solution:

(a) Using the scalar product form:

$$\langle \mathbf{u},\ \mathbf{v}\rangle = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & -1 \\ 2 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & -1 \\ 2 & 0 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

$$\langle \mathbf{u},\ \mathbf{v}\rangle = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ -6 \\ 10 \end{bmatrix} = 28$$

(b)

$$\|\mathbf{u}\|^2 = \langle \mathbf{u},\ \mathbf{u}\rangle = A\mathbf{u}\cdot A\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = 5 \implies \|\mathbf{u}\| = \sqrt{5}$$

Show for yourself that $\|\mathbf{v}\| = \sqrt{200} = 10\sqrt{2}$.

(c) If $\mathbf{w} = (x,\ y,\ z)$ then it is perpendicular to $\mathbf{u} = (1,\ 0,\ 0)$ if $\langle \mathbf{u},\ \mathbf{w}\rangle = A\mathbf{u}\cdot A\mathbf{w} = 0$. That is:

$$0 = A\mathbf{u}\cdot A\mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & -1 \\ 2 & 0 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} x+2z \\ 3y-z \\ 2x+2z \end{bmatrix} = 5x + 6z$$

Hence, $\mathbf{w} = (x,\ y,\ z)$ is perpendicular to $\mathbf{u}$ exactly when $5x + 6z = 0$.
Note: The equation $5x + 6z = 0$ defines a set of vectors $\mathbf{w}$ that is a plane through the origin of R^3 (that is, the vectors $\mathbf{w}$ go from the origin to points on the plane $5x + 6z = 0$).

(d) Since $\|\mathbf{w}\| > 0$ when $\mathbf{w} \neq 0$ it follows that $\|\mathbf{w}\| = 1$ if, and only if, $\|\mathbf{w}\|^2 = 1$. Hence, the vectors satisfy:

$$1 = \|\mathbf{w}\|^2 = A\mathbf{w}\cdot A\mathbf{w} = \begin{bmatrix} x+2z \\ 3y-z \\ 2x+2z \end{bmatrix} \cdot \begin{bmatrix} x+2z \\ 3y-z \\ 2x+2z \end{bmatrix} = 5x^2 + 9y^2 + 9z^2 + 12xz - 6yz$$

Hence, $\|\mathbf{w}\| = 1$ exactly when $5x^2 + 9y^2 + 9z^2 + 12xz - 6yz = 1$.
Note: In a Euclidean coordinate system this is the equation of an ellipsoid with centre at the origin.

Example 5.5.7.
Using the weighted Euclidean inner product on R^2 given by:

$$\langle \mathbf{u},\ \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$$

Let $\mathbf{u} = (3,\ -2)\,,\ \mathbf{v} = (1,\ 4)\,.$

(a) Find the norms $\mathbf{u},\ \mathbf{v}$.

(b) Find the inner product $\langle \mathbf{u},\ \mathbf{v}\rangle$.

(c) Find the distance between the vectors/points $\mathbf{u},\ \mathbf{v}$.

(d) Find the angle between the vectors $\mathbf{u},\ \mathbf{v}$.

(e) Find the set of all vectors perpendicular to $\mathbf{u}$.

Solution. The following outlines the solution:

(a) $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u},\ \mathbf{u}\rangle =}\sqrt{2u_1^2 + 3u_2^2} = \sqrt{30}$. Similarly $\|\mathbf{v}\| = \sqrt{50}$

(b) $\langle \mathbf{u},\ \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2 = -18$

(c) $\|\mathbf{u}-\mathbf{v}\| = \sqrt{\langle \mathbf{u}-\mathbf{v},\ \mathbf{u}-\mathbf{v}\rangle} = \sqrt{\langle (2,-6)\,,\ (2,-6)\rangle} = \sqrt{126}$ and this is the distance between the vectors/points $\mathbf{u},\ \mathbf{v}$.

(d) The angle θ satisfies $\cos\theta = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{-18}{\sqrt{30}\sqrt{50}} = -\frac{9}{5\sqrt{15}}$. Using a calculator, the approximate value of the angle θ is:

$$\theta = \arccos\left(-\frac{9}{5\sqrt{15}}\right) \simeq 2.054\,2 \text{radians}$$

In degrees the approximate value is:

$$2.054\,2 \times \frac{180}{\pi} \simeq 117.70 \text{ degrees}$$

Note: This angle is different from the angle calculated using the standard scalar product, which is $\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{\mathbf{u}\cdot\mathbf{v}}{\sqrt{\mathbf{u}\cdot\mathbf{u}}\sqrt{\mathbf{v}\cdot\mathbf{v}}} = -\frac{5}{\sqrt{13}\sqrt{17}}$, giving $\theta \simeq 1.913\,8$ radians or about 109.65 degrees.

(e) The vector $\mathbf{w} = (w_1, w_2)$ is perpendicular to $\mathbf{u}$ if:

$$0 = \langle \mathbf{u},\ \mathbf{w}\rangle = 2u_1w_1 + 3u_2w_2 = 6w_1 - 6w_2$$

The set of vectors perpendicular to $\mathbf{u}$ therefore satisfies $6w_1 - 6w_2 = 0,$ or simply $w_1 = w_2$. In a Euclidean coordinate system this is a line through the origin at 45 degrees to the axes. That is every vector from the origin along this line is perpendicular to $\mathbf{u}$.

Theorem 5.5. *This theorem is in two parts:*

(a) ***Pythagoras' Theorem**. For any two perpendicular vectors $\mathbf{u}, \mathbf{v}$ in an inner product space both of the following hold:*

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

*(b) The **cosine law**. If θ is the angle between two vectors $\mathbf{u}, \mathbf{v}$ in an inner product space, then:*

$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$$

Note: To see the connection with the usual theorem of Pythagoras and the cosine law in R^2 think of $\mathbf{u}, \mathbf{v}$ as being two vectors in R^2, starting from the origin with angle θ between them. The vector joining the end of $\mathbf{v}$ to the end of $\mathbf{u}$ (the hypotenuse of the triangle) is $\mathbf{u}-\mathbf{v}$. Hence, part (b) of the theorem in R^2 states that the square of the length of the hypotenues is the sum of the squares on the other two sides of the triangle minus $2\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$. This last term is zero when the vectors are perpendicular, thus giving Pythagoras' Theorem. See Figure 5.1.

Proof. By the definition of norm and the axioms of the inner product definition:it follows that:

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

Using the definition of angle, $\langle \mathbf{u},\ \mathbf{v}\rangle = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$, between the vectors, this becomes the part (b) result:

$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$$

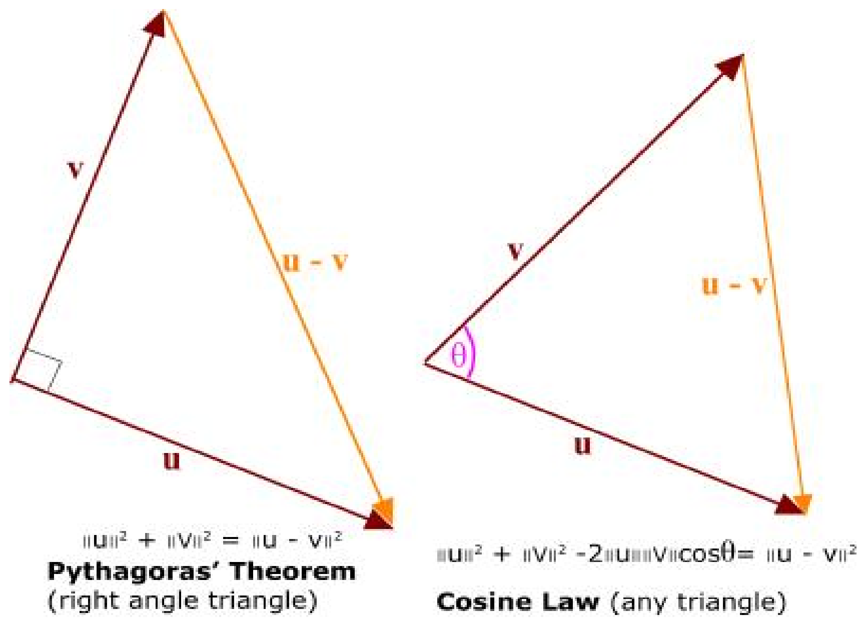

Figure 5.1: Pythagoras' Theorem

If the vectors $\mathbf{u}, \mathbf{v}$ are perpendicular then $\cos\theta = 0$, and part (a) of the theorem follows:

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

Replacing "$\mathbf{v}$" by "$-\mathbf{v}$" gives the alternate form of the part (a).

□

Theorem 5.6. *The **parallelogram theorem**. Given two vectors* $\mathbf{u},\ \mathbf{v}$*:*

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

Note: This can be interpreted as follows. The sum of the squares of the lengths of the four sides of a parallelogram is equal to the sum of the squares of the lengths of the diagonals, as in Figure 5.2.

Proof. Try this for yourself. Use the norm property $\|\mathbf{w}\|^2 = \langle \mathbf{w}, \mathbf{w}\rangle$ applied to $\|\mathbf{u}-\mathbf{v}\|^2$ and $\|\mathbf{u}+\mathbf{v}\|^2$, together with the axioms satisfied by the inner product.

□

Example 5.5.8.
In M_{22} (all 2 by 2 matrices) prove:

(a) The following is an inner product:

$$\langle A, B\rangle = \left\langle \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \right\rangle = a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{21} + a_{22}b_{22}$$

That is, multiply the matrix entries in the corresponding positions and add them together.

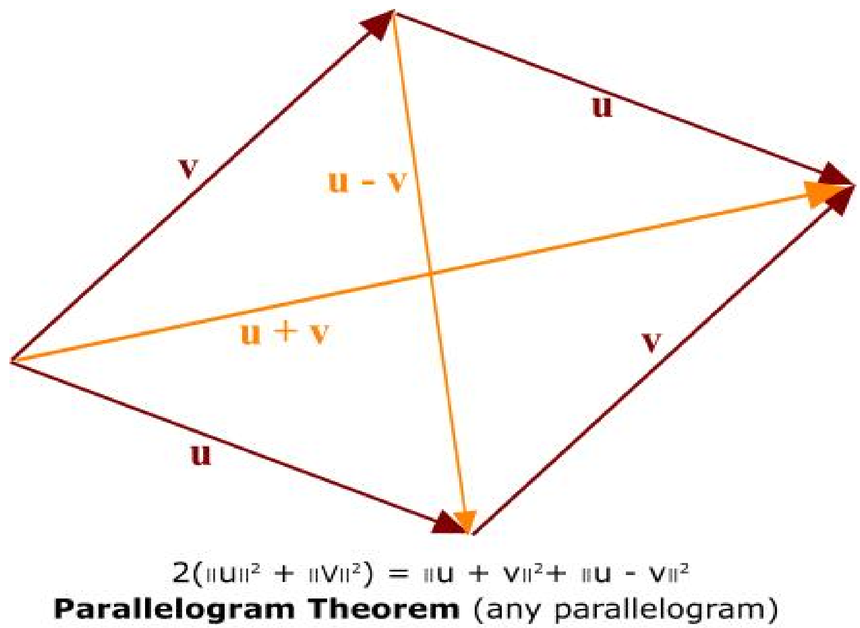

Figure 5.2: Parallelogram Theorem

(b) Prove that the two vectors/matrices are orthogonal with respect to the inner product:

$$C = \left[\begin{array}{cc} 1 & 3 \\ 2 & -4 \end{array}\right],\ D = \left[\begin{array}{cc} 0 & 2 \\ 3 & 3 \end{array}\right]$$

(c) Find the hypotenuse H of the triangle for which two of the sides are the orthogonal vectors/matrices $C,\ D$, and confirm that these satisfy Pythagoras' Theorem.

(d) Find the angles between C and H and between D and H. Do the angles of this triangle add up to 180 degrees?

Solution. The following outlines the solution:

(a) We could prove that each of the five axioms hold for this formula. However, note that the matrix shape plays no role in the formula for $\langle A, B\rangle$. That is, if we re-write the matrix entries as vectors:

$$A \longrightarrow (a_{11},\ a_{12},\ a_{21},\ a_{22}),\quad B \longrightarrow (b_{11},\ b_{12},\ b_{21},\ b_{22})$$

then the formula is exactly the same as the Euclidean inner product on R^4, and so it must be an inner product in M_{22}.

(b) $\langle C,\ D\rangle = 1\times 0 + 3\times 2 + 2\times 3 + (-4)\times 3 = 0$. Hence, the matrices/vectors are orthogonal.

(c) The hypotenuse vector/matrix is given by:

$$H = C - D = \left[\begin{array}{cc} 1 & 3 \\ 2 & -4 \end{array}\right] - \left[\begin{array}{cc} 0 & 2 \\ 3 & 3 \end{array}\right] = \left[\begin{array}{cc} 1 & 1 \\ -1 & -7 \end{array}\right]$$

Computing the norms (squaring all of the entries and adding them):

$$\|H\|^2 = \langle H, H\rangle = 52,\ \ \|C\|^2 = \langle C, C\rangle = 30,\ \ \|D\|^2 = \langle D, D\rangle = 22$$

and we have:

$$\|C\|^2 + \|D\|^2 = 30 + 22 = 52 = \|H\|^2$$

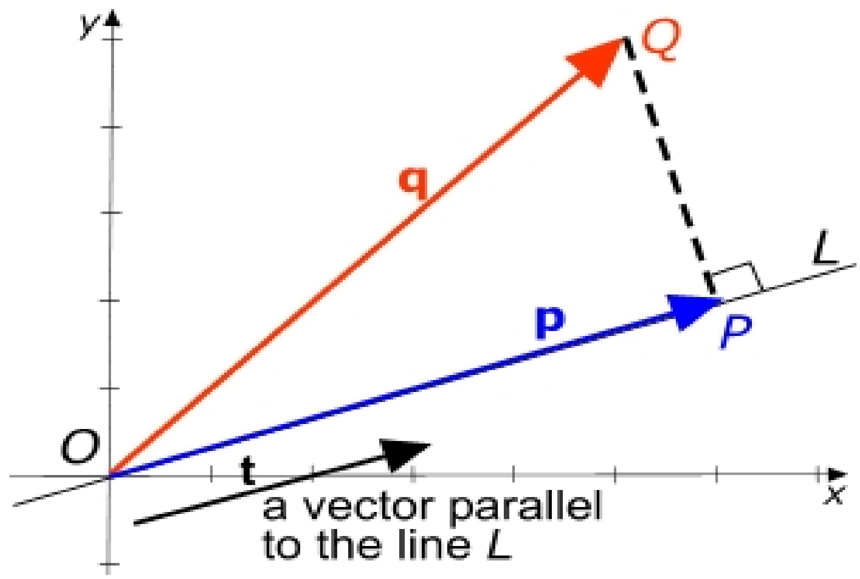

Figure 5.3: Projection onto a Line

(d) The angle θ between C and H is given by:

$$\cos\theta = \frac{\langle C,\ H\rangle}{\|C\|\,\|H\|} = \frac{32}{\sqrt{30}\sqrt{52}} \Longrightarrow \theta \simeq 0.626\,32 \text{ radians, } \simeq 35.885\text{degrees}$$

The angle ϕ between D and H is given by:

$$\cos\phi = \frac{\langle D,\ H\rangle}{\|D\|\,\|H\|} = \frac{-22}{\sqrt{22}\sqrt{52}} \Longrightarrow \phi \simeq 2.279\,0 \text{ radians, } \simeq 130.58 \text{ degrees}$$

Since the third angle in the triangle is $\frac{\pi}{2}$ radians or 90 degrees, the angles of this triangle clearly do not add up to 180 degrees. That theorem only works with the standard Euclidean norm.

Definition.
In an inner product space, the *projection* (also called *orthogonal projection*) of a vector $\mathbf{q}$ onto a vector $\mathbf{t}$ is a vector $\mathbf{p}$ parallel to $\mathbf{t}$ such that $\mathbf{p}-\mathbf{q}$ is orthogonal to $\mathbf{t}$, as in Figure 5.3.

Note: This is exactly analogous to the previously-defined projection in an Euclidean vector space (see Unit3, Section 3: Linear Transformations from R^n to R^m) where the projection is given in terms of the scalar product by $\mathbf{p} = \frac{\mathbf{q}\cdot\mathbf{t}}{\|\mathbf{t}\|^2}\mathbf{t} = \frac{\mathbf{q}\cdot\mathbf{t}}{\mathbf{t}\cdot\mathbf{t}}\mathbf{t}$.

Definition.
In an inner product space, the **reflection** of a vector $\mathbf{q}$ in the line formed by a vector $\mathbf{t}$ is a vector $\mathbf{r}$ such that $\mathbf{p}-\mathbf{r}$ is orthogonal to $\mathbf{t}$ and such that the mean, $\frac{1}{2}(\mathbf{p}+\mathbf{r})$ is the projection of $\mathbf{q}$ ono $\mathbf{t}$, as in Figure 5.4.

Note: This is exactly analogous to the previously-defined reflection in an Euclidean vector space (see Unit3, Section 3: Linear Transformations from R^n to R^m) where the reflection is given in terms of the scalar product by $\mathbf{r} = \frac{2(\mathbf{t}\cdot\mathbf{q})}{\|\mathbf{t}\|^2}\mathbf{t} - \mathbf{q}$.

Theorem 5.7. *If $\mathbf{q}, \mathbf{t}$ are any two vectors in an inner product space V then:*

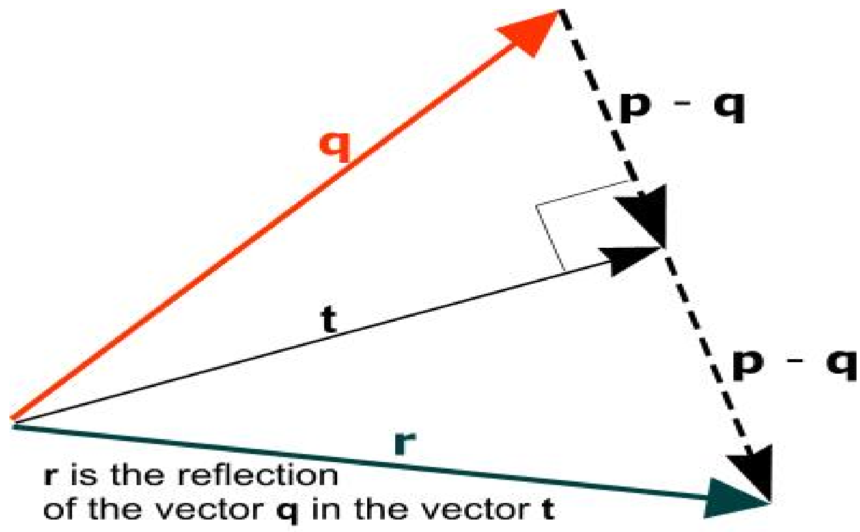

Figure 5.4: Reflection in a Line

(a) *There exists a unique vector* $\mathbf{p}$ *that is the projection of the vector* $\mathbf{q}$ *onto a vector* $\mathbf{t}$*, given by the formula:*

$$\mathbf{p} = \frac{\langle \mathbf{q}, \mathbf{t} \rangle}{\|\mathbf{t}\|^2}\mathbf{t} = \frac{\langle \mathbf{q}, \mathbf{t} \rangle}{\langle \mathbf{t}, \mathbf{t} \rangle}\mathbf{t}$$

(b) *There exists a unique vector* $\mathbf{r}$ *that is the reflection of* $\mathbf{q}$ *in* $\mathbf{t}$*, given by the formula:*

$$\mathbf{p} = 2\frac{\langle \mathbf{q}, \mathbf{t} \rangle}{\|\mathbf{t}\|^2}\mathbf{t} - \mathbf{q} = 2\frac{\langle \mathbf{q}, \mathbf{t} \rangle}{\langle \mathbf{t}, \mathbf{t} \rangle}\mathbf{t} - \mathbf{q}$$

Proof. The proofs are exactly the same as for Euclidean spaces, except that the scalar product is replaced by the inner product. Try it for yourself, and look at the proofs in Unit 3 if you have difficulty.

□

Example 5.5.9.
Using the M_{22} inner product and matrix D of Example 5.5.8, find the projection of the matrix C onto the matrix D and the projection of the matrix E onto D where:

$$E = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

Proof. From Theorem 5.7 the matrix M that is the projection of C onto D is given by:

$$M = \frac{\langle C, D \rangle}{\langle D, D \rangle}D = \frac{0}{22}\begin{bmatrix} 0 & 2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and so the projection is the zero vector, because the two matrices are orthogonal to each other.
From Theorem 5.7 the matrix M that is the projection of E onto D is given by:

$$M = \frac{\langle E, D \rangle}{\langle D, D \rangle}D = \frac{24}{22}\begin{bmatrix} 0 & 2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 0 & \frac{24}{11} \\ \frac{36}{11} & \frac{36}{11} \end{bmatrix}$$

□

Example 5.5.10.
Requires calculus. In P_4 find the reflection of the polynomial $f(x) = x$ in the polynomial $g(x) = 1 + x + x^2 + x^3$, using the inner product $\int_{-1}^{1} f(x)\, g(x)\ dx$.

Solution. The reflection polynomial $h(x)$ is given by the formula:

$$\begin{aligned} h(x) &= 2\frac{\langle f, g\rangle}{\langle g, g\rangle} g(x) - f(x) \\ &= 2\frac{\int_{-1}^{1} f(x)\, g(x)\ dx}{\int_{-1}^{1} f(x)\, g(x)\ dx} g(x) - g(x) \end{aligned}$$

Section 5.5 exercise set

Check your understanding by answering the following questions.

1. In R^2 prove that $\langle \mathbf{u}, \mathbf{v}\rangle$ is an inner product (show it satisfies the five axioms):

$$\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$$

for any two vectors $\mathbf{u} = (u_1,\ u_2)$ and $\mathbf{v} = (v_1,\ v_2)$.

2. In R^2 let any two vectors be given by $\mathbf{u} = (u_1,\ u_2)$ and $\mathbf{v} = (v_1,\ v_2)$. Show that each of the following **is not** an inner product in R^2. Recall that you need only produce one example where the result fails in order to disprove something.

   (a) $\langle \mathbf{u}, \mathbf{v}\rangle = -2u_1v_1 + 3u_2v_2$

   (b) $\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 + u_2$

   (c) $\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 + u_2v_2 + 1$

   (d) $\langle \mathbf{u}, \mathbf{v}\rangle = |u_1v_1 + u_2v_2|$ (absolute value)

   (e) $\langle \mathbf{u}, \mathbf{v}\rangle = \frac{u_1}{v_1} + \frac{u_2}{v_2}$

3. Using the inner product on R^2 given by $\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$:

   (a) Find the lengths of the vectors $(1,\ 0), (2,\ -1)$.

   (b) Find the inner product of the two vectors above.

   (c) Find the angle between the two vectors.

4. Define an inner product on R^3 by:

$$\langle (u_1,\ u_2,\ u_3),\ (v_1,\ v_2,\ v_3)\rangle = 2u_1v_1 + 3u_2v_2 + u_3v_3$$

   (a) Find the length of the vector $(3,\ 2,\ 1)$.

   (b) Find the inner product of the vector $(3,\ 2,\ 1)$ with the vector $(0,\ 1,\ -2)$.

   (c) Find all vectors perpendicular to the vector $(3,\ 2,\ 1)$.

   (d) Find the distance from $(3,\ 2,\ 1)$ to $(0,\ 1,\ -2)$.

5. Define an function on P_3 for any two polynomials $p(x) = a_0 + a_1x + a_2x^2,\ q(x) = b_0 + b_1x + b_2x^2$ by:

$$\langle p,\ q\rangle = a_0b_0 + a_1b_1 + a_2b_2$$

   (a) Prove that it is an inner product.

   (b) Find the length of the vector $p(x) = 2 - 3x + 4x^2$.

(c) Find all vectors perpendicular to the vector $p(x) = 1$.

(d) Find all vectors perpendicular to the vector $p(x) = 1 + 2x - x^2$.

6. **Note: Requires calculus.** Repeat the previous problem, parts (a), (b), (c), but with the inner product defined for any two polynomials in $p,\ q\ \in P_3$ by:

$$\langle p,\ q\rangle = \int_{-1}^{1} p(x)\, q(x)\ dx$$

7. Using the matrix-based inner product for R^2 defined by $\langle \mathbf{u},\ \mathbf{v}\rangle = \mathbf{u}^T A^T A\mathbf{v}$ (see Examples 5.5.4 and 5.5.5), where:

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$$

(a) Find the length of the vector $(1,\ 1)$.

(b) Find the inner product of the vector $(1,\ 1)$ with the vector $(0,\ 1)$.

(c) Find the distance from $(1,\ 1)$ to $(0,\ 1)$.

(d) Find the angle between the vectors $(1,\ 1)$ and $(0,\ 1)$.

8. In R^2 define a function $\langle \mathbf{u},\ \mathbf{v}\rangle = \mathbf{u}^T A^T A\mathbf{v}$ where:

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

Is it an inner product for R^2? Justify your answer.

9. In R^2 with inner product $\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$, find the equation satisfied by all vectors with $\|\mathbf{u}\| = 1$.

10. Find a formula for the inner product in P_3 given in Theorem 5.1, when the basis of P_3 is $\left\{1,\ x,\ 1 + x^2\right\}$.

11. In M_{22}, for any two matrices $A = [a_{ij}]\,, B = [b_{ij}]$ define:

$$\langle A,\ B\rangle = a_{11}b_{11} + 2a_{12}b_{12} + 3a_{21}b_{21} + 4a_{22}b_{22}$$

That is, multipy elements in corresponding positions in each matrix and form a weighted sum.

(a) Show this is an inner product.

(b) Compute the norm of the matrix/vector given in question 7.

(c) Compute the inner product of the two matrices/vectors:

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ D = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}$$

(d) Compute the distance between the two matrices/vectors above.

(e) Compute the angle between the two matrices/vectors above.

12. Use the Cauchy-Schwarz inequality to prove:

(a) For any two vectors $\mathbf{u}, \mathbf{v} \in R^2$

$$|u_1v_1 + u_2v_2| \leq \left(u_1^2 + u_2^2\right)^{\frac{1}{2}} \left(v_1^2 + v_2^2\right)^{\frac{1}{2}}$$

(b) For any two vectors $\mathbf{u}, \mathbf{v} \in R^n$

$$|u_1v_1 + u_2v_2 + \ \cdots \ + u_nv_n| \le \left(u_1^2 + u_2^2 + \ \cdots \ + u_n^2\right)^{\frac{1}{2}} \left(v_1^2 + v_2^2 + \ \cdots \ + v_n^2\right)^{\frac{1}{2}}$$

(c) **Requires calculus.** For any two functions $f, \ g$ continuous on $[0, 1]$:

$$\left[\int_0^1 f(x)\, g(x) \ dx\right]^2 \le \left[\int_0^1 f(x) \ dx\right] \left[\int_0^1 g(x) \ dx\right]$$

(d) **Requires calculus.** For any two functions $f, \ g$ continuous on $[0, 1]$:

$$\left[\int_0^1 [f(x) + g(x)]^2 \ dx\right]^{\frac{1}{2}} \le \left[\int_0^1 [f(x)]^2 \ dx\right]^{\frac{1}{2}} + \left[\int_0^1 [g(x)]^2 \ dx\right]^{\frac{1}{2}}$$

13. Let V be any inner product space. Prove that the inner product function can be expressed in terms of norms of the vectors by:

$$\langle \mathbf{u}, \ \mathbf{v} \rangle = \frac{1}{4} \|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4} \|\mathbf{u} - \mathbf{v}\|^2$$

14. Show that **Pythagoras' Theorem** can be written in the following form and prove this result. For all non-zero vectors $\mathbf{p}, \ \mathbf{q}, \ \mathbf{r}$ satisfying $(\mathbf{p} - \mathbf{r}) \cdot (\mathbf{r} - \mathbf{q}) = 0$, it is true that:

$$\|\mathbf{p} - \mathbf{r}\|^2 + \|\mathbf{r} - \mathbf{q}\|^2 = \|\mathbf{p} - \mathbf{q}\|^2$$

15. In any inner product space prove that if $\|\mathbf{v}\| = 0$ then $\mathbf{v} = \mathbf{0}$ (the zero vector).

16. If V is an inner product space then prove:

(a) If $\mathbf{u} \in V$, is a fixed vector then prove that the set $S = \{\mathbf{v} \in V \mid \langle \mathbf{u}, \mathbf{v} \rangle = 0\}$, of all vectors perpendicular to $\mathbf{u}$, is a subspace of V.

(b) If $\mathbf{w}$ is perpendicular to each of the vectors in $T = \{\mathbf{v}_1, \mathbf{v}_2, \ \cdots \ \mathbf{v}_k\}$ then $\mathbf{w}$ is perpendicular to every vector in the linear span of T (that is, all linear combinations of the vectors v_j).

(c) The set S of all vectors $\mathbf{w} \in V$ perpendicular to T is a subspace of V. **Note:** Sometime S is designated by the symbol $T^{\perp}$.

(d) If the set T in part (b) is a basis of V then $\mathbf{w}$ must be the zero vector (that is, only the zero vector is perpendicular to every vector of a basis).

17. In an inner product space V prove that any three vectors satisfy:

$$\|\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3\| \le \|\mathbf{v}_1\| + \|\mathbf{v}_2\| + \|\mathbf{v}_3\|$$

Solutions

1. $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2$ is clearly a real number so Axiom 1 is satisfied, and clearly $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle = 2u_1v_1 + 3u_2v_2$ so Axiom 2 is satisfied.
Axiom 3 follows from the linearity of real number multiplication/addition:

$$\begin{aligned}\langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle &= 2(u_1 + w_1)\, v_1 + 3(u_2 + w_2)\, v_2 \\ &= 2u_1v_1 + 3u_2v_2 + 2w_1v_1 + 3w_2v_2 \\ &= \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle\end{aligned}$$

Axiom 4 follows in a similar way:

$$\begin{aligned}\langle k\mathbf{u},\mathbf{v}\rangle &= 2ku_1 + k3u_2 \\ &= k\left(2u_1 + 3u_2\right) \\ &= k\left\langle \mathbf{u},\mathbf{v}\right\rangle\end{aligned}$$

Axiom 5 follows easily:

$$\langle \mathbf{u},\mathbf{u}\rangle = 2u_1^2 + 3u_2^2 \geq 0$$

and clearly $\langle \mathbf{u},\mathbf{u}\rangle = 0$ only if $u_1 = u_2 = 0.$

2. **Note:** In each case one Axiom is shown to fail, but it is noted that other Axioms fail as well. Try yourself to find examples of failure for these other Axioms.

   (a) $\langle \mathbf{u},\mathbf{v}\rangle = -2u_1v_1 + 3u_2v_2$ satisfies the first four axioms but Axiom 5 fails if, for example, $\mathbf{u} = (1,0)$ when $\langle \mathbf{u},\mathbf{u}\rangle = -2$ is negative.

   (b) $\langle \mathbf{u},\mathbf{v}\rangle = u_1v_1 + u_2$ satisfies Axioms 1, but none of the other axioms. For example, Axiom 2 fails when $\mathbf{u} = (1,1)\,,\ \mathbf{v} = (1,0)$ because $\langle \mathbf{u},\mathbf{v}\rangle = 2 \neq \langle \mathbf{v},\mathbf{u}\rangle = 1.$

   (c) $\langle \mathbf{u},\mathbf{v}\rangle = u_1v_1 + u_2v_2 + 1$ satisfies Axioms 1, 2, 5 but not Axioms 3 and 4. For example, to disprove Axiom 3, if $\mathbf{u} = (1,0)\,,\ \mathbf{v} = (1,1)\,,\ \mathbf{w} = (1,0)$ then:

   $$\langle \mathbf{u}+\mathbf{w},\ \mathbf{v}\rangle = 3, \quad \langle \mathbf{u},\ \mathbf{v}\rangle + \langle \mathbf{w},\ \mathbf{v}\rangle = 2 + 2 = 4$$

   so $\langle \mathbf{u}+\mathbf{w},\ \mathbf{v}\rangle \neq \langle \mathbf{u},\ \mathbf{v}\rangle + \langle \mathbf{w},\ \mathbf{v}\rangle$.

   (d) $\langle \mathbf{u},\mathbf{v}\rangle = |u_1v_1 + u_2v_2|$ satisfies Axioms 1, 2, 5 but not Axioms 3 and 4. For example, to disprove Axiom 4, if $\mathbf{u} = (1,0)\,,\ \mathbf{v} = (1,0)\,,\ k = -2$ then:

   $$\langle \mathbf{u},\ \mathbf{v}\rangle = 1,\ \ \langle k\mathbf{u},\ \mathbf{v}\rangle = 2,\ k\left\langle \mathbf{u},\ \mathbf{v}\right\rangle = -2$$

   so $\langle k\mathbf{u},\ \mathbf{v}\rangle \neq k\left\langle \mathbf{u},\ \mathbf{v}\right\rangle$.

   (e) $\langle \mathbf{u},\mathbf{v}\rangle = \frac{u_1}{v_1} + \frac{u_2}{v_2}$ does not satisfy any of the axioms. For example, Axiom 1 does not hold if $\mathbf{u} = (1,0)$ because $\langle \mathbf{u},\mathbf{v}\rangle$ is not defined (cannot divide by zero).

3. If $\mathbf{u} = (1,0)\,, \mathbf{v} = (2,\ -1)\,,\ \ \langle \mathbf{u},\mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$, then:

   (a) $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u},\mathbf{u}\rangle} = \sqrt{2u_1^2 + 3u_2^2} = \sqrt{2}$. Similarly $\|\mathbf{v}\| = \sqrt{11}$.

   (b) $\langle \mathbf{u},\ \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2 = 4$

   (c) The angle θ between 0 and π is given by $\cos\theta = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{4}{\sqrt{22}}$, and it is approximately (using a calculator): $\theta = \arccos\left(\frac{4}{\sqrt{22}}\right) \simeq 0.549\,47$ radians (or $31.\,482$ degrees). **Note:** This is not the usual angle between these two vectors given by the Euclidean norm, which is about $0.463\,65$ radians or $26.\,565$ degrees -check this for yourself.

4. If $\langle (u_1,\ u_2,\ u_3)\,,\ (v_1,\ v_2,\ v_3)\rangle = 2u_1v_1 + 3u_2v_2 + u_3v_3$ and $\mathbf{u} = (3,\ 2,\ 1)\,, \mathbf{v} = (0,\ 1,\ -2)$ then:

   (a) $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u},\mathbf{u}\rangle} = \sqrt{2u_1^2 + 3u_2^2 + u_3^2} = \sqrt{31}$

   (b) $\langle \mathbf{u},\mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2 + u_3v_3 = 4$

   (c) The vector $\mathbf{w} = (w_1,\ w_2,\ w_3)$ is perpendicular to $\mathbf{u}$ if:

   $$\begin{gathered}2u_1w_1 + 3u_2w_2 + u_3w_3 = 0 \Longrightarrow \\ 6w_1 + 6w_2 + w_3 = 0\end{gathered}$$

   This is the equation of a plane through the origin. That is, all vectors from the origin to a point on this plane are perpendicular to $\mathbf{u}$.

(d) $\|\mathbf{u}-\mathbf{v}\| = \sqrt{\langle \mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle} = \sqrt{(3,\ 1,\ 3)\,,(3,\ 1,\ 3)} = \sqrt{30}$

5. If $f(x) = a_0 + a_1 x + a_2 x^2$, $g(x) = b_0 + b_1 x + b_2 x^2$ and $\langle f,\ g\rangle = a_0 b_0 + a_1 b_1 + a_2 b_2$ then:

(a) Axioms 1 and 2 clearly hold since $a_0 b_0 + a_1 b_1 + a_2 b_2$ is a real number, and the real number multiplications are all commutative. Axiom 3 holds because if $h(x) = c_0 + c_1 x + c_2 x^2$ then:

$$\begin{aligned}\langle f+h,\ g\rangle &= (a_0+c_0)\,b_0 + (a_1+c_1)\,b_1 + (a_2+c_2)\,b_2 \\ &= (a_0 b_0 + a_1 b_1 + a_2 b_2) + (c_0 b_0 + c_1 b_1 + c_2 b_2) \\ &= \langle f,\ g\rangle + \langle h,\ g\rangle\end{aligned}$$

Axiom 4 holds because:

$$\langle f,\ f\rangle = a_0^2 + a_1^2 + a_2^2 \geq 0$$

and $\langle f,\ f\rangle = 0$ only if $a_0 = a_1 = a_2 = 0$ (that is, when $f(x)$ is the zero function).

(b) $\|p(x)\| = \sqrt{\langle p,\ p\rangle} = \sqrt{p_0^2 + p_1^2 + p_2^2} = \sqrt{29}$

(c) $q(x) = q_0 + q_1 x + q_2 x^2$ is perpendicular to $p(x) = p_0 + p_1 x + p_2 x^2 = 1$ if:

$$\begin{aligned}q_0 p_0 + q_1 p_1 + q_2 p_2 = 0 \Longrightarrow& \\ q_0 = 0 \text{ (since } p_0 = 1, p_1 = p_2 = 0)&\end{aligned}$$

Hence, the vectors/polynomials perpendicular to $p(x) = 1$ are all polynomials of the form $q(x) = q_1 x + q_2 x^2$ for any real values q_1, q_2.

(d) Similar to the previous part $q(x)$ must satisfy:

$$q_0 p_0 + q_1 p_1 + q_2 p_2 = 0$$

where $p_0 = 1,\ p_1 = 2,\ p_2 = -1$. That is:

$$q_0 + 2q_1 - q_2 = 0$$

That is, replacing q_2 by $(q_0 + 2q_1)$, all vectors/polynomials perpendicular to $p(x) = 1 + 2x - x^2$ are $q(x) = q_0 + q_1 x + (q_0 + 2q_1)\,x^2$, for any real values q_0, q_1.

6. If $f(x) = a_0 + a_1 x + a_2 x^2$, $g(x) = b_0 + b_1 x + b_2 x^2$ and $\langle f,\ g\rangle = \int_{-1}^{1} p(x)\,q(x)\ dx$ then:

(a) Axiom 1 holds because the integral always exists (when the functions are continuous) and is a real number.
Axiom 2 holds because $\langle f,\ g\rangle = \langle g,\ f\rangle = \int_{-1}^{1} f(x)\,g(x)\ dx$.
Axiom 3 holds because if $h(x) = c_0 + c_1 x + c_2 x^2$ then:

$$\begin{aligned}\langle f\ +h,\ g\rangle &= \int_{-1}^{1} (f(x) + h(x))\,g(x)\ dx \\ &= \int_{-1}^{1} f(x)\,g(x) + h(x)\,g(x)\ dx \\ &= \int_{-1}^{1} f(x)\,g(x)\ dx + \int_{-1}^{1} h(x)\,g(x)\ dx \\ &= \langle f,\ g\rangle + \langle h,\ g\rangle\end{aligned}$$

Axiom 4 holds because if $k \in R$:

$$\begin{aligned}\langle kf,\ g\rangle &= \int_{-1}^{1} kf(x)\,g(x)\ dx \\ &= k\int_{-1}^{1} f(x)\,g(x)\ dx \\ &= k\,\langle f,\ g\rangle\end{aligned}$$

Axiom 5 holds because:

$$\langle f,\ f\rangle = \int_{-1}^{1} [f(x)]^2\ dx \geq 0$$

since the integral of a non-negative function over any interval is a non-negative value. A theorem of calculus shows that the only way this integral can be equal to zero when f is continous on the interval $[-1,\ 1]$ is when $f(x)$ is equal to zero for every x in the interval.
Note: This proof also works for the vector space of functions for which the integrals exist, such as the vector space of all functions continuous on $[-1,\ 1]$. The proof also works if different limits of integration are used to define the inner product formula.

(b) $\|p(x)\| = \sqrt{\langle p,\ p\rangle} = \sqrt{\int_{-1}^{1} [p(x)]^2\ dx}$ and this is given by:

$$\begin{aligned}\sqrt{\int_{-1}^{1} [p(x)]^2\ dx} &= \sqrt{\int_{-1}^{1} (2-3x+4x^2)^2\ dx} \\ &= \sqrt{\int_{-1}^{1} (16x^4-24x^3+25x^2-12x+4)\ dx} \\ &= \sqrt{\frac{16}{5}x^5-6x^4+\frac{25}{3}x^3-6x^2+4x\Big|_{-1}^{1}} \\ &= \sqrt{\frac{466}{15}}\end{aligned}$$

(c) $q(x) = q_0+q_1x+q_2x^2$ is perpendicular to $p(x) = p_0+p_1x+p_2x^2 = 1$ if:

$$\begin{aligned}0 &= \langle p,\ q\rangle = \int_{-1}^{1} p(x)\,q(x)\ dx \\ 0 &= \int_{-1}^{1} \left(q_0+q_1x+q_2x^2\right)\ dx \\ 0 &= q_0x+q_1\frac{x^2}{2}+q_2\frac{x^3}{3}\Big|_{-1}^{1} \\ 0 &= 2q_0+\frac{2}{3}q_2 \\ 0 &= q_0+\frac{q_2}{3}\end{aligned}$$

Hence, substituting $q_0 = -\frac{q_2}{3}$, the polynomials perpendicular to $p(x) = 1$ are $q(x) = -\frac{q_2}{3}+q_1x+q_2x^2$.

7. Let $\mathbf{u} = (1,1)\,,\ \mathbf{v} = (0,\ 1)\,$. First compute:

$$A^TA = \begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 10 \end{bmatrix}$$

(a) $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u},\mathbf{u}\rangle} = \sqrt{\mathbf{u}^TA^TA\mathbf{u}}$ and this is given by:

$$\mathbf{u}^TA^TA\mathbf{u} = \begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} 4 & 6 \\ 6 & 10 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 26$$

Hence, $\|\mathbf{u}\| = \sqrt{26}$. Similarly, $\|\mathbf{v}\| = \sqrt{10}$.

(b) This is given by:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T A^T A \mathbf{v} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 6 \\ 6 & 10 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 16$$

(c) $\|\mathbf{u} - \mathbf{v}\| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle}$ and $\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle$ is given by:

$$\begin{aligned} \langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle &= (\mathbf{u} - \mathbf{v})^T A^T A (\mathbf{u} - \mathbf{v}) \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 6 \\ 6 & 10 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= 4 \end{aligned}$$

Hence, the distance from $(1,\ 1)$ to $(0,\ 1)$ is $\sqrt{4} = 2$.

(d) The angle θ between $\mathbf{u}$ and $\mathbf{v}$ is given by:

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|} = \frac{16}{\sqrt{26}\sqrt{10}} = \frac{8}{\sqrt{65}}$$

The approximate angle (using a calculator) is $\theta = \arccos \frac{8}{\sqrt{65}} \simeq 0.124\,35$ radians or about $7.124\,7$ degrees. **Note:** This has no relationship to the angle calculated using the Euclidean inner product, which is $\frac{\pi}{4}$ or 45 degrees.

8. The function $\langle \mathbf{u},\ \mathbf{v} \rangle = \mathbf{u}^T A^T A \mathbf{v}$ is defined for all vectors $\mathbf{u},\ \mathbf{v} \in R^2$ and is a real number (since the sizes of the four parts of the product are compatible: $1 \times 2,\ 2 \times 3,\ 3 \times 2,\ 2 \times 1$, and the result has size 1×1). Hence, Axiom 1 is satisfied.

Axiom 2 is satisfied because $\langle \mathbf{v},\ \mathbf{u} \rangle = \mathbf{v}^T A^T A \mathbf{u} = \left(\mathbf{v}^T A^T A \mathbf{u}\right)^T$ (since the transpose of a number is the same number). Hence:

$$\langle \mathbf{v},\ \mathbf{u} \rangle = \left(\mathbf{v}^T A^T A \mathbf{u}\right)^T = \mathbf{u}^T A^T A \ \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle$$

Axiom 3 is satisfied since:

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

Axiom 4 is satisfed since for $k \in R$:

$$\begin{aligned} \langle k\mathbf{u}, \mathbf{v} \rangle &= (k\mathbf{u})^T A^T A \mathbf{v} \\ &= k \left(\mathbf{u}^T A^T A \mathbf{v}\right) \\ &= k \langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

Axiom 5 is satisfied since:

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v} \rangle &= \mathbf{v}^T A^T A \mathbf{v} \\ &= (A\mathbf{v})^T \\ &= \|A\mathbf{v}\|^2 \ \text{(standard Euclidean norm)} \end{aligned}$$

and so $\langle \mathbf{v}, \mathbf{v} \rangle = \|A\mathbf{v}\|^2 \geq 0$. Furthermore, $A\mathbf{v} = \mathbf{0}$ only if:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} v_1 \\ v_2 \\ v_1 + v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which shows $\mathbf{v} = \mathbf{0}$.
Hence, all five axioms are satisfied and so the function is an inner product.

9. $\|\mathbf{u}\| = 1$ if, and only if, $\|\mathbf{u}\|^2 = 1$, which is $2u_1^2 + 3u_2^2 = 1$. In the standard Euclidean axis system, this is an ellipse.

10. Any vector/polynomial $p(x) = p_0 + p_1 x + p_2 x^2$ can be expressed in terms of the basis by:

$$p(x) = (p_0 - p_2) \times 1 + p_1 \times x + p_2 \times \left(1 + x^2\right)$$

Hence, by Theorem 5.1 the inner product defined as a dot product by this basis is:

$$\begin{aligned}\langle p, q\rangle &= (p_0 - p_2,\ p_1,\ p_2) \cdot (q_0 - q_2,\ q_1, q_2) \\ &= (p_0 - p_2)(q_0 - q_2) + p_1 q_1 + p_2 q_2 \\ \langle p, q\rangle &= p_0 q_0 - p_0 q_2 + p_1 q_1 - p_2 q_0 + 2 p_2 q_2\end{aligned}$$

Hence, the inner product is:

$$\langle p, q\rangle = p_0 q_0 - p_0 q_2 + p_1 q_1 - p_2 q_0 + 2 p_2 q_2$$

11.

(a) We could prove that each of the five axioms hold for this formula. However, we can avoid this by using the method of Example 5.5.8. Note that the matrix shape plays no role in the formula for $\langle A, B\rangle$. The formula is exactly the same as a weighted (weights 1, 2, 3, and 4) Euclidean inner product on R^4 where the matrix entries are re-written as vectors:

$$A \longrightarrow (a_{11},\ a_{12},\ a_{21},\ a_{22}),\ \ B \longrightarrow (b_{11},\ b_{12},\ b_{21},\ b_{22})$$

Since a weighted Euclidean inner product is always an inner product, then the matrix formula $\langle A, B\rangle$ is also an inner product.

(b) $\left\|\begin{bmatrix} 2 & 3 \\ 0 & 1\end{bmatrix}\right\| = 1 \times 2^2 + 2 \times 3^2 + 3 \times 0 + 4 \times 1^2 = 26$

(c) $\langle C,\ D\rangle = 1 \times 0 + 2 \times 0 \times 3 + 3 \times 0 \times 1 + 4 \times 1 \times 0 = 0$

(d) $\langle C - D,\ C - D\rangle = \left\langle \begin{bmatrix} 1 & -3 \\ -1 & 1\end{bmatrix},\ \begin{bmatrix} 1 & -3 \\ -1 & 1\end{bmatrix}\right\rangle = 1 + 2(-3)^2 + 3(-1)^2 + 4 = 26$. Hence:

$$\|C - D\| = \sqrt{\langle C - D,\ C - D\rangle} = \sqrt{26}$$

(e) If θ is the angle between C and D then:

$$\cos\theta = \frac{\langle C, D\rangle}{\|C\|\,\|D\|} = 0 \text{ since } \langle C, D\rangle = 0$$

Hence, $\theta = \frac{\pi}{2}$ (90 degrees). The matrices are perpendicular to each other.

12. The Cauchy-Schwarz result (Theorem 5.3) is: $|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$.

(a) If $\mathbf{u} = (u_1,\ u_2),\ \mathbf{v} = (v_1,\ v_2)$ with the inner product is the standard scalar or dot product so that $\langle \mathbf{u}, \mathbf{v}\rangle = u_1 v_1 + u_2 v_2$ and $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2},\ \ \|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2}$. Applying these in the Cauchy-Schwarz formula gives the required result:

$$|u_1 v_1 + u_2 v_2| \le \sqrt{u_1^2 + u_2^2}\sqrt{v_1^2 + v_2^2}$$

(b) The proof is very similar to part (a). Try this for yourself.

(c) The function $\langle f,\ g\rangle = \int_0^1 f(x)\, g(x)\ dx$ is an inner product for the vector space of functions continuous on $[0,1]$ (see the proof of question 6(a). With this inner product: $\|f\| = \sqrt{\int_0^1 [f(x)]^2\ dx},\ \ \|g\| = \sqrt{\int_0^1 [g(x)]^2\ dx}$, and the Cauchy-Schwarz result, $|\langle f,g\rangle| \le \|f\|\,\|g\|$, becomes:

$$\left|\int_0^1 f(x)\, g(x)\ dx\right| \le \left(\int_0^1 [f(x)]^2\ dx\right)^{\frac{1}{2}} \left(\int_0^1 [g(x)]^2\ dx\right)^{\frac{1}{2}}$$

Both sides are positive, so squaring both sides gives the required result:

$$\left[\int_0^1 f(x)\, g(x)\ dx\right]^2 \le \left[\int_0^1 [f(x)]^2\ dx\right] \left[\int_0^1 [g(x)]^2\ dx\right]$$

(d) This result is simply the triangle inequality (see Theorem 5.4, part (c)):

$$\|f+g\| \le \|f\| + \|g\|$$

13. By the definition of norm in terms of the inner product and the axioms of the inner product:

$$\begin{aligned}\frac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2 &= \frac{1}{4}\langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\ \rangle - \frac{1}{4}\langle \mathbf{u}-\mathbf{v},\ \mathbf{u}-\mathbf{v}\rangle \\ &= \frac{1}{4}[\langle \mathbf{u},\mathbf{u}\ \rangle + 2\langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{v}\ \rangle] - \frac{1}{4}[\langle \mathbf{u},\mathbf{u}\ \rangle - 2\langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{v}\ \rangle] \\ &= \langle \mathbf{u},\mathbf{v}\rangle\end{aligned}$$

14. One version of Pythagoras' Theorem (Theorem 5.5) states that $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u}+\mathbf{v}\|^2$ when $\langle \mathbf{u},\mathbf{v}\rangle = 0$. Replace $\mathbf{u}$ by $\mathbf{p}-\mathbf{r}$ and $\mathbf{v}$ by $\mathbf{r}-\mathbf{q}$ so that $\mathbf{u}+\mathbf{v}$ is replaced by $\mathbf{p}-\mathbf{q}$ and $\langle \mathbf{u},\mathbf{v}\rangle = 0$ is replaced by $\langle \mathbf{p}-\mathbf{r}, \mathbf{r}-\mathbf{q}\rangle = 0$, thus giving the required result:

$$\|\mathbf{p}-\mathbf{r}\|^2 + \|\mathbf{r}-\mathbf{q}\|^2 = \|\mathbf{p}-\mathbf{q}\|^2$$

15. Proof of the "if" part: If $\mathbf{v} = \mathbf{0}$ then $\langle \mathbf{v},\mathbf{v}\rangle = 0$ by Axiom 5. Hence, $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v},\mathbf{v}\rangle} = 0.$
Proof of the "only if" part: If $\|\mathbf{v}\| = 0$ then $\|\mathbf{v}\|^2 = \langle \mathbf{v},\mathbf{v}\rangle = 0$. Axiom 5 states that $\langle \mathbf{v},\mathbf{v}\rangle = 0$ only when $\mathbf{v} = \mathbf{0}$ and so the result follows.

16.

(a) The set S of vectors in V perpendicular to a particular vector $\mathbf{v}$ is a subset of a vector space V. To prove S is a vector space it is only necessary to prove it is closed under addition and scalar multiplication (see Theorem 2.3 of Unit 2, Section 3: Subspaces of a Vector Space). Suppose $k \in R$ and $\mathbf{v},\ \mathbf{w}$ are any two vectors in S, so $\langle \mathbf{v},\mathbf{u}\rangle = \langle \mathbf{w},\mathbf{u}\rangle = 0$. Using the axioms of the inner product:

$$\begin{aligned}\langle \mathbf{v}+\mathbf{w},\mathbf{u}\rangle &= \langle \mathbf{v},\mathbf{u}\rangle + \langle \mathbf{w},\mathbf{u}\rangle = 0+0 = 0 \\ \langle k\mathbf{v},\mathbf{u}\rangle &= k\langle \mathbf{v},\mathbf{u}\rangle = k\times 0 = 0\end{aligned}$$

Hence, S is closed under addition and scalar multiplication and so it is a vector space (a subspace of V). **Note:** S must also be an inner product space since the inner product of V is also an inner product for S.

(b) If $\mathbf{w}$ satisfies $\langle \mathbf{v}_1,\ \mathbf{w}\rangle = 0,\ \langle \mathbf{v}_2,\ \mathbf{w}\rangle = 0,\ \cdots, \langle \mathbf{v}_k,\ \mathbf{w}\rangle = 0$ and then by the axioms of inner products, for any scalars $r_1, r_2,\ \cdots, r_k$:

$$\begin{aligned}\langle r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \ \cdots\ + r_k\mathbf{v}_k,\ \mathbf{w}\rangle &= \langle r_1\mathbf{v}_1,\ \mathbf{w}\rangle + \langle r_2\mathbf{v}_2,\ \mathbf{w}\rangle + \ \cdots\ + \langle r_k\mathbf{v}_k,\ \mathbf{w}\rangle \\ &= r_1\langle \mathbf{v}_1,\ \mathbf{w}\rangle + r_2\langle \mathbf{v}_2,\ \mathbf{w}\rangle + \ \cdots\ + r_k\langle \mathbf{v}_k,\ \mathbf{w}\rangle \\ &= 0\end{aligned}$$

Hence, $\mathbf{w}$ is perpendicular to the linear span of T.

(c) As in part (a), it is only necessary to prove that S is closed under addition and scalar multiplication. The proof is very similar to the proof of part (a) and is not given here.

(d) The vector $\mathbf{w}$ is perpendicular to all of the basis vectors, and $\mathbf{w}$ is also a linear combination of the basis vectors. However $\mathbf{w}$ is orthogonal to all linear combinations of the basis vectors, by part (b), and so $\mathbf{w}$ is perpendicular to itself:

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

By Axiom 5 for linear products, it follows that $\mathbf{w} = \mathbf{0}$.

17. The triangle inequality for any two vectors $\mathbf{u},\ \mathbf{v} \in V$, states that $\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$. Writing $\mathbf{u} = \mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{v} = \mathbf{v}_3$ changes this to:

$$\|(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3\| \le \|\mathbf{v}_1 + \mathbf{v}_2\| + \|\mathbf{v}_3\|$$

Applying the triangle inequality a second time to $\mathbf{v}_1, \mathbf{v}_2$ shows that $\|\mathbf{v}_1 + \mathbf{v}_2\| \le \|\mathbf{v}_1\| + \|\mathbf{v}_2\|$ and so the above inequality becomes the required result:

$$\|\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3\| \le \|\mathbf{v}_1\| + \|\mathbf{v}_2\| + \|\mathbf{v}_3\|$$

5.6 Orthogonal bases, the Gram-Schmidt process and $QR-$ factorization

We have previously seen that a basis of a vector space can be used to develop most processes and properties of interest. Usually it does not matter which particular basis is used, but sometimes a special basis is easier to use than other bases. In particular, if the vector space is an Inner Product Space then it is often very advantageous to work with an **orthogonal basis** (each basis vector is perpendicular to every other basis vector). Orthogonal bases are required in some applications, such as the method for diagonalizing a matrix using eigenvalues and eigenvectors (see Unit 4 Eigenvalues, Eigenvectors, and Diagonalization of Matrices, Diagonalizing a Matrix).

In this section an algorithm , the **Gram-Schmidt Process**, is described for converting any non-orthogonal basis of an Inner Product Space into an orthogonal basis. If the vectors of the original basis are the columns of a matrix A then the Gram-Schmidt process is shown to be equivalent to finding a $QR-$ factorization, $A = QR$, where Q is an orthogonal matrix, and R is upper triangular. The $QR-$ factorization is used in the $QR-$ algorithm, one of the most successful numerical methods for finding the eigenvalues a matrix (see Unit 4 Eigenvalues, Eigenvectors and Diagonalization of Matrices, Methods for finding Eigenvalues and Eigenvectors).

5.6.1 The Gram-Schmidt process

Definition.
A set of vectors $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3, \cdots,\ \mathbf{v}_n\}$ of an Inner Product space is said to be **orthogonal** if the vectors are mutually orthogonal, meaning:

$$\langle \mathbf{v}_i,\ \mathbf{v}_j \rangle = 0 \text{ for } i \ne j = 1, 2, 3, \cdots, n$$

The set $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3, \cdots,\ \mathbf{v}_n\}$ is said to be **orthonormal** if it is orthogonal and all vectors are unit vectors. That is, for $i, j = 1, 2, 3, \cdots, n$:

$$\langle \mathbf{v}_i,\ \mathbf{v}_i \rangle = 1 \text{ and } \langle \mathbf{v}_i,\ \mathbf{v}_j \rangle = 0 \text{ if } i \ne j$$

Note: A vector $\mathbf{v}$ is a unit vector if its length is 1, meaning $\|\mathbf{v}\| = 1$. Since $\|\mathbf{v}\|^2 = \langle \mathbf{v},\ \mathbf{v} \rangle$, and so $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v},\ \mathbf{v} \rangle}$, it follows that $\mathbf{v}$ is also a unit vector if $\langle \mathbf{v},\ \mathbf{v} \rangle = 1$.
Note: An orthogonal set of vectors can always be converted to an orthonormal set by simply converting each vector to a unit vector by dividing it by its length (change each vector $\mathbf{v}$ to the vector $\frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{\sqrt{\langle \mathbf{v},\ \mathbf{v} \rangle}}\mathbf{v}$).

The Gram-Schmidt Process, or algorithm, uses a known basis of an Inner Product space to construct an orthogonal basis. The next two examples show how this process works in simple cases, and Theorem 5.8 gives the general process. The process makes extensive use of the formula for the orthogonal projection of one vector, $\mathbf{u}$, onto another vector $\mathbf{v}$ derived in the previous section.

$$\text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\langle \mathbf{u},\ \mathbf{v} \rangle}{\|\mathbf{v}\|^2}\mathbf{v} = \frac{\langle \mathbf{u},\ \mathbf{v} \rangle}{\langle \mathbf{v},\ \mathbf{v} \rangle}\mathbf{v} \text{ - orthogonal projection of } \mathbf{u} \text{ onto } \mathbf{v}$$

The formula for the projection of $\mathbf{u}$ onto vector $\mathbf{v}$ in Euclidean spaces was developed in Unit 3, Linear Transformations from R^n to R^m in the subsection: Projection Operators $R^2 \to R$ and is the same formula with the scalar product being the inner product:

$$\text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v} = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}\mathbf{v}$$

Example 5.6.1.
Given the basis $\{\mathbf{u}_1,\ \ \mathbf{u}_2\} = \{(1,\ 1)\,,\ (1,\ 0)\}$ of R^2, construct an orthogonal basis $\{\mathbf{v}_1,\ \mathbf{v}_2\}$.

Solution. Note that the inner product here is the normal scalar product, and the existing basis is not orthogonal, since

$$\mathbf{u}_1 \cdot\ \mathbf{u}_2 = (1,\ 1) \cdot (1,\ 0) = 1 \neq 0$$

First we will compute an orthogonal basis.

Step 1. Choose arbitrarily $\mathbf{v}_1$ as one of the basis vectors, say:

$$\mathbf{v}_1 = \mathbf{u}_1 = (1,\ 1)$$

Step 2. Choose $\mathbf{v}_2$ as the second original basis vector, $\mathbf{u}_2$, minus the projection of $\mathbf{u}_2$ onto $\mathbf{v}_1$:

$$\begin{aligned} \mathbf{v}_2 &= \mathbf{u}_2 - \frac{(\mathbf{u}_2 \cdot\ \mathbf{v}_1)}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\ &= (1,\ 0) - \frac{(1,\ 0) \cdot (1,\ 1)}{2}(1,\ 1) \\ &= (1,\ 0) - \frac{1}{2}(1,\ 1) \\ \mathbf{v}_2 &= \left(\frac{1}{2},\ -\frac{1}{2}\right) \end{aligned}$$

Multiplying $\mathbf{v}_2$ by 2 to simplify it without changing the orthogonality gives the orthogonal basis $\{\mathbf{v}_1,\ \mathbf{v}_2\} = \{(1,\ 1)\,,\ (1,\ -1)\}$. **Note:** Normalizing the vectors gives the orthonormal basis:

$$\{\mathbf{v}_1,\ \mathbf{v}_2\} = \left\{\left(\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right),\ \left(\frac{1}{\sqrt{2}},\ -\frac{1}{\sqrt{2}}\right)\right\}$$

Note: Check for yourself that these are orthogonal.
Note: The process used here is followed in more general cases. Start with one of the original basis vectors, then modify the second one by subtracting its projection on the first vector. In the next example, with three vectors, the third original basis vector is modified by subtracting its projection on the first two vectors of the orthogonal basis.

Example 5.6.2.
Use this to construct an orthonormal basis $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3\}$.

Solution. Note that the inner product here is the normal scalar product, and the existing basis is clearly not orthogonal. It is easiest to construct an orthogonal basis first, then normalize (make into unit vectors) the basis afterwards. Start as in the previous example:

Step 1. Choose arbitrarily $\mathbf{v}_1$ as one of the basis vectors, say:

$$\mathbf{v}_1 = \mathbf{u}_1 = (1,\ 1,\ 1)$$

Step 2. You might notice that $\mathbf{u}_2$ is already orthogonal to $\mathbf{v}_1$ and so we can choose $\mathbf{v}_2 = \mathbf{u}_2 = (1,\ 0\ -1)$. If you did not notice this then the solution process gives the same result as follows. Choose $\mathbf{v}_2$ as the second original basis vector, $\mathbf{u}_2$, minus the projection of $\mathbf{u}_2$ onto $\mathbf{v}_1$:

$$\begin{aligned}
\mathbf{v}_2 &= \mathbf{u}_2 - \frac{(\mathbf{u}_2 \cdot\ \mathbf{v}_1)}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\
&= (1,\ 0,\ -1) - \frac{(1,\ 0,\ -1)\cdot(1,\ 1,\ 1)}{(1,\ 1,\ 1)\cdot(1,\ 1,\ 1)}(1,\ 1,\ 1) \\
&= (1,\ 0\ -1) - \frac{0}{3}(1,\ 1,\ 1) \\
\mathbf{v}_2 &= (1,\ 0\ -1)
\end{aligned}$$

Step 3. In order to find $\mathbf{v}_3$ apply the method of Step 2 to $\mathbf{u}_3$, but this time subtract off the projections of $\mathbf{u}_3$ onto both $\mathbf{v}_1$ and $\mathbf{v}_2$.

$$\begin{aligned}
\mathbf{v}_3 &= \mathbf{u}_3 - \frac{(\mathbf{u}_3 \cdot\ \mathbf{v}_1)}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{(\mathbf{u}_3 \cdot\ \mathbf{v}_2)}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 \\
&= (2,\ -3,\ 4) - \frac{(2,\ -3,\ 4)\cdot(1,\ 1,\ 1)}{(1,\ 1,\ 1)\cdot(1,\ 1,\ 1)}(1,\ 1,\ 1) - \frac{(2,\ -3,\ 4)\cdot(1,\ 0,\ -1)}{(1,\ 0,\ -1)\cdot(1,\ 0,\ -1)}(1,\ 0,\ -1) \\
&= (2,\ -3,\ 4) - \frac{3}{3}(1,\ 1,\ 1) - \frac{-2}{2}(1,\ 0,\ -1) \\
\mathbf{v}_3 &= (2,\ -4,\ 2)
\end{aligned}$$

Hence, the orthogonal basis is $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3\} = \{(1,\ 1,\ 1),\ (1,\ 0\ -1),\ (2,-4,\ 2)\}$.
Note: Check for yourself that this set is orthogonal. Notice that $\mathbf{v}_3 = (1,\ -2,\ 1)$ can be simplified, by dividing by 2, to give $\mathbf{v}_3 = (1,\ -2,\ 1)$, and the set is still orthogonal.
Note: Normalizing the orthogonal basis gives an orthonormal basis:

$$\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3\} = \left\{\left(\frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}}\right),\ \left(\frac{1}{\sqrt{2}},\ 0\ -\frac{1}{\sqrt{2}}\right),\ \left(\frac{1}{\sqrt{6}},-\frac{2}{\sqrt{6}},\ \frac{1}{\sqrt{6}}\right)\right\}$$

The general method for producing an orthogonal basis is given next in Theorem 5.8. It is a simple extension of the process in Examples 5.6.1 and 5.6.2 above, except that a dot product like $\mathbf{u}_2 \cdot\ \mathbf{v}_1$ is replaced by the inner product $\langle \mathbf{u}_2,\ \mathbf{v}_1\rangle$.

Theorem 5.8. ***The Gram-Schmidt Process*** *If $\{\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3, \cdots,\ \mathbf{u}_m\}$ is a set of linearly independent vectors spanning a subspace S of an inner-product space V then an* ***orthogonal*** *set of vectors, $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3, \cdots,\ \mathbf{v}_m\}$, spanning the same subspace S is produced by the following process (algorithm):*

Step 1. Set $\mathbf{v}_1 = \mathbf{u}_1$

Step 2. Set $\mathbf{v}_2 = \mathbf{u}_2 - \dfrac{\langle \mathbf{u}_2,\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$

Step 3. Set $\mathbf{v}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3,\ \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3,\ \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2$

Step 4. Set $\mathbf{v}_4 = \mathbf{u}_4 - \frac{\langle \mathbf{u}_4,\ \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_4,\ \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \frac{\langle \mathbf{u}_4,\ \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3$

and continuing this pattern until step m *is reached:*

Step m*. Set* $\mathbf{v}_m = \mathbf{u}_m - \frac{\langle \mathbf{u}_m,\ \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_m,\ \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \frac{\langle \mathbf{u}_m,\ \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3 - \cdots\cdots\ - \frac{\langle \mathbf{u}_m,\ \mathbf{v}_{m-1} \rangle}{\|\mathbf{v}_{m-1}\|^2}\mathbf{v}_{m-1}$

Note: *In the equations above we can replace* $\|\mathbf{v}_j\|^2$ *by* $\langle \mathbf{v}_j,\ \mathbf{v}_j \rangle$ *for each* $j = 1, 2,\ \cdots, m$*, and an* ***orthonormal basis*** *is obtained by changing each* $\mathbf{v}_j$ *into the unit vector* $\frac{1}{\|\mathbf{v}j\|}\mathbf{v}_j = \frac{1}{\sqrt{\langle \mathbf{v}_j,\ \mathbf{v}_j \rangle}}\mathbf{v}_j$*.*

Note: *If* $\{\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3, \cdots,\ \mathbf{u}_m\}$ *is a basis of* V *then* $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3, \cdots,\ \mathbf{v}_m\}$ *will be an orthogonal basis of* V*.*

Proof. Consult your textbook or other source for a proof of this result. The proof is conceptually simple but rather ”messy” and confusing.

□

The Gram-Schmidt Process can be written in a slightly different and, in some ways, simpler form in order to directly produce an orthonormal basis, as in the next theorem.

Theorem 5.9. *If* $\{\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3, \cdots,\ \mathbf{u}_m\}$ *is a set of linearly independent vectors spanning a subspace* S *of an inner-product space* V *then an* ***orthonormal*** *set of vectors,* $\{\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3, \cdots,\ \mathbf{w}_m\}$*, spanning the same subspace* S *is produced by the following process (algorithm):*

Step 1. Set $\mathbf{v}_1 = \mathbf{u}_1$ *and define* $\mathbf{w}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1$*. That is, normalize the vector* $\mathbf{v}_1$ *so it has length one by dividing by* $\|\mathbf{v}_1\| = \sqrt{\langle \mathbf{v}_1,\ \mathbf{v}_1 \rangle}$*.*

Step 2. Set $\mathbf{v}_2 = \mathbf{u}_2 - \langle \mathbf{u}_2,\ \mathbf{w}_1 \rangle \mathbf{w}_1$ *and define* $\mathbf{w}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2$*.*

Step 3. Set $\mathbf{v}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3,\ \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{u}_3,\ \mathbf{w}_2 \rangle \mathbf{w}_2$ *and define* $\mathbf{w}_3 = \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3$*.*

Step 4. Set $\mathbf{v}_4 = \mathbf{u}_4 - \langle \mathbf{u}_4,\ \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{u}_4,\ \mathbf{w}_2 \rangle \mathbf{w}_2 - \langle \mathbf{u}_4,\ \mathbf{w}_3 \rangle \mathbf{w}_3$ *and define* $\mathbf{w}_4 = \frac{1}{\|\mathbf{v}_4\|}\mathbf{v}_4$*,*

and continuing this pattern until step n *is reached.*

Step m*. Set* $\mathbf{v}_m = \mathbf{u}_m - \langle \mathbf{u}_m,\ \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{u}_m,\ \mathbf{w}_2 \rangle \mathbf{w}_2 - \langle \mathbf{u}_m,\ \mathbf{w}_3 \rangle \mathbf{w}_3 - \cdots\cdots\ - \langle \mathbf{u}_w,\ \mathbf{w}_{m-1} \rangle \mathbf{w}_{m-1}$*, and define* $\mathbf{w}_m = \frac{1}{\|\mathbf{v}_m\|}\mathbf{v}_m$*.*

Proof. In step 2 of Theorem 5.8 show that the formula there is the same as the one used here by verifying that:

$$\begin{aligned}\mathbf{v}_2 &= \mathbf{u}_2 - \frac{\langle \mathbf{u}_2,\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\ &= \mathbf{u}_2 - \left\langle \mathbf{u}_2,\ \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}\right\rangle \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} \\ &= \mathbf{u}_2 - \langle \mathbf{u}_2,\ \mathbf{w}_1\rangle \mathbf{w}_1\end{aligned}$$

Show that the other formulae are also the same by following the same method. □

Example 5.6.3.
The set of polynomials $\{g_1(x),\ g_2(x),\ g_3(x),\ g_4(x)\} = \{1,\ 1+x,\ 1-2x^2,\ x+x^3\}$ is a basis of P_3, the vector space of all polynomials of degree 3 or less. P_3 is an inner product space with the inner product computed as the scalar product of the coefficients of the polynomials (see previous section of this unit for details).

$$\langle a_0 + a_1x + a_2x^2 + a_3x^3,\ b_0 + b_1x + b_2x^2 + b_3x^3\rangle = a_0b_0 + a_1b_1 + a_2b_2 + a_3b_3$$

Use the Gram-Schmidt process to find an orthogonal basis $\{f_1,\ f_2,\ f_3,\ f_4\}$ for P_3.

Solution. Step 1: Define $f_1(x) = g_1(x) = 1$

Step 2: Define $f_2(x) = g_2(x) - \frac{\langle g_2,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) = 1 + x - \frac{\langle 1+x,\ 1\rangle}{\langle 1,\ 1\rangle} \times 1 = 1 + x - 1 = x$

Step 3: Define:

$$\begin{aligned}f_3(x) &= g_3(x) - \frac{\langle g_3,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_3,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) \\ &= 1 - 2x^2 - \frac{\langle 1-2x^2,\ 1\rangle}{\langle 1,\ 1\rangle} - \frac{\langle 1-2x^2,\ x\rangle}{\langle x,\ x\rangle} x \\ &= 1 - 2x^2 - 1 - \frac{0}{1}x \\ f_3(x) &= -2x^2\end{aligned}$$

Step 4: Define:

$$\begin{aligned}f_4(x) &= g_4(x) - \frac{\langle g_4,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_4,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) - \frac{\langle g_4,\ f_3\rangle}{\langle f_3,\ f_3\rangle} f_3(x) \\ &= x + x^3 - \frac{\langle x+x^3,\ 1\rangle}{\langle 1,\ 1\rangle} - \frac{\langle x+x^3,\ x\rangle}{\langle x,\ x\rangle} x - \frac{\langle x+x^3,\ -2x^2\rangle}{\langle -2x^2,\ -2x^2\rangle}\left(-2x^2\right) \\ &= x + x^3 - \frac{0}{1} - \frac{1}{1}x - \frac{0}{4}\left(-2x^2\right) \\ f_4(x) &= x^3\end{aligned}$$

Hence, the orthogonal basis is: $\{1,\ x,\ -2x^2,\ x^3\}$.

Note: We can divide the third polynomial by -2 without changing the orthogonality, thus giving the standard basis of P_3 :

$$\{1,\ x,\ x^2,\ x^3\}$$

In fact we could have done this at step 3 when we found $f_3(x) = -2x^2$, thus simplifying step 4 slightly. Satisfy yourself that this basis is orthogonal.

Example 5.6.4.
Requires calculus. The set of functions $f\ (x)$ that are continuous on an interval $[a,b]$, written $C[a,b]$ is a an infinite dimensional vector space and is an inner product space with the inner product defined as the integral of the product of the two functions between a and b :

$$\langle f,g\rangle = \int_a^b f\ (x)\, g(x)\ dx$$

This is also an inner product for the subspaces of $C[a,b]$ of polynomials $P_n,\ n=1,2,3,\cdots$. Given $a=-1, b=1$ and the four functions: $g_1(x)=1,\ g_2(x)=x,\ g_3(x)=x^2,\ g_4(x)=x^3$ then find four orthogonal functions, $f_1,\ f_2,\ f_3, f_4,$ that span the same subspace.

Solution. The following is the solution:

Step 1: Define $f_1(x)=g_1(x)=1$

Step 2: Define:

$$\begin{aligned} f_2(x) &= g_2(x) - \frac{\langle g_2,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) \\ &= x - \frac{\int_{-1}^{1} x\ dx}{\int_{-1}^{1}\ dx}\times 1 \\ &= x - \frac{0}{2} \\ f_2(x) &= x \end{aligned}$$

Step 3: Define:

$$\begin{aligned} f_3(x) &= g_3(x) - \frac{\langle g_3,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_3,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) \\ &= x^2 - \frac{\int_{-1}^{1} x^2\ dx}{\int_{-1}^{1}\ dx}\times 1 - \frac{\int_{-1}^{1} x^2\times x\ dx}{\int_{-1}^{1} x^2\ dx}\ x \\ &= x^2 - \frac{\left(\frac{2}{3}\right)}{2} - \frac{0}{\left(\frac{2}{3}\right)}\ x \\ f_3(x) &= x^2 - \frac{1}{3} \end{aligned}$$

Step 4: Define:

$$\begin{aligned} f_4(x) &= g_4(x) - \frac{\langle g_4,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_4,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) - \frac{\langle g_4,\ f_3\rangle}{\langle f_3,\ f_3\rangle} f_3(x) \\ &= x^3 - \frac{\int_{-1}^{1} x^3\ dx}{\int_{-1}^{1}\ dx}\times 1 - \frac{\int_{-1}^{1} x^3\times x\ dx}{\int_{-1}^{1} x^2\ dx}\ x - \frac{\int_{-1}^{1} x^3\left(x^2-\frac{1}{3}\right)\ dx}{\int_{-1}^{1}\left(x^2-\frac{1}{3}\right)\left(x^2-\frac{1}{3}\right)\ dx}\ \left(x^2-\frac{1}{3}\right) \\ &= x^3 - \frac{0}{2} - \frac{\left(\frac{2}{5}\right)}{\left(\frac{2}{3}\right)}\ x - \frac{0}{\left(\frac{8}{45}\right)}\ \left(x^2-\frac{1}{3}\right) \\ f_4(x) &= x^3 - \frac{3}{5}\ x \end{aligned}$$

Note: The orthogonal polynomials $f_1(x) = 1,\ f_2(x) = x,\ f_3(x) = x^2 - \frac{1}{3},\ f_4(x) = x^2 - \ \frac{3}{5}\, x$ are multiples of the first four **Legendre Polynomials**, which are:

$$p_0(x) = 1,\ p_1(x) = x,\ p_2(x) = \frac{1}{2}\left(x^2 - 1\right),\ p_3(x) = \frac{1}{2}\left(5x^3 - 3x\right)$$

These multiples are chosen so that $p_k(1) = 1$, so they are an orthogonal set but are not normalized with respect to the inner product and so do not form an orthonormal set. Legendre Polynomials are important in some Engineering applications and more details can be found on the web at http://en.wikipedia.org/wiki/Legendre_polynomials

5.6.2 Uses of orthogonal and orthonormal bases

When vectors are expressed as linear combinations of orthogonal basis vectors, many operations are much easier to carry out, as is shown in the following theorems. If the basis is orthonormal, then it becomes even easier, and vectors with respect to this orthonormal basis interact very like vectors in Euclidean spaces, R^n. The first theorem notes the fairly obvious fact that orthogonal sets of vectors must be linearly independent.

Theorem 5.10. *If $S = \{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3, \cdots,\ \mathbf{v}_n\}$ is an orthogonal set of (non-zero) vectors in an inner product space then the vectors are linearly independent.*
Note: *Hence, if S is produced by the Gram-Schmidt process applied to a basis of an inner product space V, then S must also be a basis of V. Hence, every inner product space has an orthogonal basis.*

Proof. If the vectors are linearly dependent then a non-zero linear combination of the vectors gives the zero vector:

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + k_3\mathbf{v}_3 + \cdots\ + k_n\mathbf{v}_n = \mathbf{0}$$

Take the inner product with $\mathbf{v}_1$, using the axioms:

$$\begin{aligned}\langle (k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + k_3\mathbf{v}_3 + \cdots\ + k_n\mathbf{v}_n),\ \ \mathbf{v}_1\rangle &= \langle \mathbf{0},\ \mathbf{v}_1\rangle \\ k_1\langle \mathbf{v}_1,\ \mathbf{v}_1\rangle + k_2\langle \mathbf{v}_2,\ \mathbf{v}_1\rangle + k_3\langle \mathbf{v}_3,\ \mathbf{v}_1\rangle + \cdots\ + k_n\langle \mathbf{v}_n,\ \mathbf{v}_1\rangle &= 0 \\ k_1\langle \mathbf{v}_1,\ \mathbf{v}_1\rangle &= 0\end{aligned}$$

This shows that $k_1 = 0$, since $\langle \mathbf{v}_1,\ \mathbf{v}_1\rangle = \|\mathbf{v}\|^2 \neq 0$. Repeating the above process, taking inner products with $\mathbf{v}_2,\ \mathbf{v}_3, \cdots,\ \mathbf{v}_n$, shows that all of the coefficients are zero:

$$k_1 = k_2 = k_3 = \cdots\ = k_n = 0$$

Hence, there is no linear combination of the vectors giving the zero vector except the trivial one with all coefficients equal to zero, and so the vectors are linearly independent.

□

Theorem 5.11. *If $B = \{\mathbf{b}_1,\ \mathbf{b}_2,\ \mathbf{b}_3, \cdots,\ \mathbf{b}_n\}$ is an **orthonormal** basis of an inner product space and vectors $\mathbf{u}, \mathbf{v}$ have coordinates with respect to this basis:*

$$(\mathbf{u})_B = (u_1,\ u_2,\ u_3, \cdots,\ u_n) \text{ and } (\mathbf{v})_B = (v_1,\ v_2,\ v_3, \cdots,\ v_n)$$

(that is: $\mathbf{u} = u_1\mathbf{b}_1 +\ u_2\mathbf{b}_2 +\ u_3\mathbf{b}_3 + \cdots +\ u_n\mathbf{b}_n$ and similarly for $\mathbf{v}$), then:

(a) $\langle \mathbf{u},\ \mathbf{v}\rangle = u_1v_1 + u_2v_2 + u_3v_3 + \cdots + u_nv_n$

(b) $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + u_3^2 + \cdots + u_n^2}$

(c) $d(\mathbf{u},\ \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2 + \cdots + (u_n - v_n)^2}$

Note: *These quantities do not depend on the values of the basis vectors, but only on the coordinates relative to the basis, and are exactly the same formulae as for Euclidean vectors in R^n with respect to the standard basis of R^n.*

Proof. The following is the proof:

(a) When $n = 2$**:** Using the linearity of the inner product:

$$\begin{aligned}\langle \mathbf{u},\ \mathbf{v}\rangle &= \langle u_1\mathbf{b}_1 +\ u_2\mathbf{b}_2\ ,\ \ v_1\mathbf{b}_1 +\ v_2\mathbf{b}_2\rangle \\ &= u_1v_1\,\langle \mathbf{b}_1,\ \mathbf{b}_1\rangle + u_1\ v_2\,\langle \mathbf{b}_1,\ \mathbf{b}_2\rangle + u_2v_1\,\langle \mathbf{b}_2,\ \mathbf{b}_1\rangle +\ u_2v_2\,\langle \mathbf{b}_2,\ \mathbf{b}_2\rangle\end{aligned}$$

Using the orthonormality $\langle \mathbf{b}_1,\ \mathbf{b}_1\rangle = \langle \mathbf{b}_2,\ \mathbf{b}_2\rangle = 1$, and $\langle \mathbf{b}_1,\ \mathbf{b}_2\rangle = \langle \mathbf{b}_2,\ \mathbf{b}_1\rangle = 0$, this simplifies to:

$$\langle \mathbf{u},\ \mathbf{v}\rangle = u_1v_1 +\ u_2v_2$$

Note: Try yourself the same proof for $n = 3$, and attempt to generalize your proof for all values of n.
Note: Try yourself to prove (b) and (c) for the cases $n = 2,\ 3$ and for all n.

□

Theorem 5.12. *If $B = \{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3, \cdots,\ \mathbf{v}_n\}$ is an* ***orthogonal*** *basis of an inner product space and a vector $\mathbf{u}$ has coordinates $(\mathbf{u})_S = (u_1,\ u_2,\ u_3, \cdots,\ u_n)$ with respect to this basis then:*

$$u_1 = \frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2},\ u_2 = \frac{\langle \mathbf{u},\ \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}, \cdots, u_n = \frac{\langle \mathbf{u},\ \mathbf{v}_n\rangle}{\|\mathbf{v}_n\|^2},\ \textit{so}\ \mathbf{u} = \frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\ \mathbf{v}_1 +\ \frac{\langle \mathbf{u},\ \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\ \mathbf{v}_2 +\ \cdots\ + \frac{\langle \mathbf{u},\ \mathbf{v}_n\rangle}{\|\mathbf{v}_n\|^2}\ \mathbf{v}_n$$

If the basis is also ***orthonormal*** *then:*

$$u_1 = \langle \mathbf{u},\ \mathbf{v}_1\rangle\,,\ u_2 = \langle \mathbf{u},\ \mathbf{v}_2\rangle\,,\ \cdots, u_n = \langle \mathbf{u},\ \mathbf{v}_n\rangle\ \textit{so that}\ \mathbf{u} = \langle \mathbf{u},\ \mathbf{v}_1\rangle\ \mathbf{v}_1 +\ \langle \mathbf{u},\ \mathbf{v}_2\rangle\ \mathbf{v}_2 +\ \cdots\ + \langle \mathbf{u},\ \mathbf{v}_n\rangle\ \mathbf{v}_n$$

Note: Recall that $\frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\ \mathbf{v}_1$ is the formula for the perpendicular projection of the vector $\mathbf{u}$ onto the vector $\mathbf{v}_1$. Hence, the coordinates with respect to an orthogonal basis are given by the perpendicular projections onto the basis vectors. Recall also that for any vector $\mathbf{v}$, $\|\mathbf{v}\|^2 = \langle \mathbf{v},\ \mathbf{v}\rangle$.

Proof. **Proof for** $n = 2$: Suppose that $\mathbf{u} = u_1\mathbf{v}_1 + u_2\mathbf{v}_2$ then taking the inner product with $\mathbf{v}_1$ and using the linearity axioms:

$$\begin{aligned}\langle \mathbf{u},\ \mathbf{v}_1\rangle &= \langle u_1\mathbf{v}_1 + u_2\mathbf{v}_2,\ \mathbf{v}_1\rangle \\ &= u_1\,\langle \mathbf{v}_1,\ \mathbf{v}_1\rangle + u_2\,\langle \mathbf{v}_2,\ \mathbf{v}_1\rangle \\ \langle \mathbf{u},\ \mathbf{v}_1\rangle &= u_1\,\langle \mathbf{v}_1,\ \mathbf{v}_1\rangle\end{aligned}$$

Hence, solving for u_1 :

$$u_1 = \frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\langle \mathbf{v}_1,\ \mathbf{v}_1\rangle} = \frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}$$

Taking the inner product of $\mathbf{u}$ with $\mathbf{v}_2$ gives in the same way the formula for u_2 :

$$u_2 = \frac{\langle \mathbf{u},\ \mathbf{v}_2\rangle}{\langle \mathbf{v}_2,\ \mathbf{v}_2\rangle} = \frac{\langle \mathbf{u},\ \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}$$

Note: Try yourself the proof for $n = 3$, and think about the generalization to the proof for all n.

□

Generalizing the previous theorem to projections onto a subspace of an inner product space gives the following result.

Theorem 5.13. *Suppose that W is a $r-$ dimensional subspace with orthogonal basis $B = \{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r\}$ of an inner product space V and $\mathbf{u} \in V$, then:*

(a) *The perpendicular projection of $\mathbf{u}$ onto W is given by:*

$$\text{proj}_W\ \mathbf{u} = \frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\ \mathbf{v}_1 +\ \frac{\langle \mathbf{u},\ \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\ \mathbf{v}_2 +\ \cdots\ + \frac{\langle \mathbf{u},\ v_r\rangle}{\|\mathbf{v}_r\|^2}\ \mathbf{v}_r$$

or if $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r\}$ is orthonormal then:

$$\text{proj}_W\ \mathbf{u} = \langle \mathbf{u},\ \mathbf{v}_1\rangle\ \mathbf{v}_1 +\ \langle \mathbf{u},\ \mathbf{v}_2\rangle\ \mathbf{v}_2 +\ \cdots\ + \langle \mathbf{u},\ v_r\rangle\ \mathbf{v}_r$$

(b) *The basis B can be extended to an orthogonal basis of V :*

$$\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r,\ \mathbf{v}_{r+1},\ \cdots,\mathbf{v}_n\}$$

and the additional basis vectors $\mathbf{v}_{r+1},\ \cdots,\mathbf{v}_n$ are a basis of $W^\perp$ (the subspace of all vectors in W that are perpendicular to every vector in W).

(c) *Every vector $\mathbf{u} \in V$ can be expressed in exactly one way:*

$$\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2 \text{ where } \mathbf{u}_1 \in W \text{ and } \mathbf{u}_2 \in W^\perp$$

Note: Part (a) states that the projection of $\mathbf{u}$ onto W is simply the sum of the projections onto each individual vector of B (the orthogonal basis of W).
Note: The formula in part (a) for the projection of $\mathbf{u}$ onto the space S spanned by the orthogonal set $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r\}$ is exactly the same as the formula from Theorem 5.12. for expressing $\mathbf{u}$ as a linear combination of the $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r\}$. That is, the formula gives that linear combination if it exists (if $\mathbf{u}$ is in the span of S) and otherwise gives the projection of the vector $\mathbf{u}$ onto S.

Proof. The following is the proof:

(a) If $\mathbf{w} = \text{proj}_W\ \mathbf{u} = k_1\mathbf{v}_1 +\ k_2\mathbf{v}_2 +\ \cdots\ + k_r\mathbf{v}_r$ then use the fact that $\mathbf{w} - \mathbf{u}$ is orthogonal to every vector in W, first taking the inner product with $\mathbf{v}_1$:

$$0 = \langle \mathbf{w} - \mathbf{u},\ \mathbf{v}_1\rangle$$
$$0 = \langle k_1\mathbf{v}_1 +\ k_2\mathbf{v}_2 +\ \cdots\ + k_r\mathbf{v}_r,\ \mathbf{v}_1\rangle - \langle \mathbf{u},\ \mathbf{v}_1\rangle$$

Using the linearity of the inner product and orthonormality of B:

$$0 = k_1\langle \mathbf{v}_1,\ \mathbf{v}_1\rangle - \langle \mathbf{u},\ \mathbf{v}_1\rangle \Longrightarrow$$
$$k_1 = \frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\langle \mathbf{v}_1,\ \mathbf{v}_1\rangle} = \frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}$$

The formulae for other k_i values can be established using the same method, computing $0 = \langle \mathbf{w} - \mathbf{u},\ \mathbf{v}_j\rangle$ for $j = 2,\ 3,\ \cdots,r$.

(b) (outline only) The basis $B = \{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r\}$ can be extended to a basis of V by first simply adding any vector $\mathbf{w}_{r+1}$ not in W, then repeating this by adding another vector not in the span of $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r,\ \mathbf{w}_{r+1}\}$ and so on until a basis of V is created. Apply the Gram-Schmidt process to this basis, starting with $\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r$ (i.e., leave them unchanged) to produce an orthogonal basis $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r,\ \mathbf{v}_{r+1},\ \cdots,\mathbf{v}_n\}$ of V. By orthogonality all of the basis vectors $\mathbf{v}_{r+1},\ \cdots,\mathbf{v}_n$ are orthogonal to every vector $\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r$. Linearity of the inner product shows that every vector of the subspace $W^\perp$ spanned by $\mathbf{v}_{r+1},\ \cdots,\mathbf{v}_n$ is perpendicular to every vector of the subspace W spanned by $\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r$. Furthermore, any vector $\mathbf{w}$ orthogonal to W can easily be shown to belong to $W^\perp$.

(c) This immediately follows from the proof of (b). That is, express $\mathbf{u}$ as a linear (unique) combination of the basis vectors $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r,\ \mathbf{v}_{r+1},\ \cdots,\mathbf{v}_n\}$. The part involving $\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_r$ will be $\mathbf{u}_1$ and the part involving $\mathbf{v}_{r+1},\ \cdots,\mathbf{v}_n$ will be $\mathbf{u}_2$.

□

Example 5.6.5.
This example has four parts:

(a) Write the vector $\mathbf{u} = (1, 2,\ 3)$ of R^3 as a linear combination of the orthogonal basis vectors:

$$\mathbf{v}_1 = (1, 0, 0)\,,\ \mathbf{v}_2 = (0,\ 1,\ 1)\,,\ \mathbf{v}_3 = (0,\ 1,\ -1)$$

Note: Check for yourself that this is an orthogonal set of vectors.

(b) Use Theorem 5.11 to compute $\|\mathbf{u}\|$ and compare it with the value by the standard calculation.

(c) Find the projection of $\mathbf{u}$ onto the subspace S of R^3 spanned by $\mathbf{v}_2$ and $\mathbf{v}_3$.

(d) Write $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ where $\mathbf{u}_1 \in S$ and $\mathbf{u}_2 \in S^{\perp}$.

Solution. The following is the solution:

(a) We could solve for x, y, z the system of equations formed by equating the components:

$$(1, 2,\ 3) = x\,(1, 0, 0) + y\,(0,\ 1,\ 1) + z\,(0,\ 1,\ -1)$$

However, Theorem 5.12 gives us an easier way to do this when the basis is orthogonal, namely:

$$\begin{aligned}
\mathbf{u} &= \frac{\langle \mathbf{u},\ \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\ \mathbf{v}_1 + \frac{\langle \mathbf{u},\ \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\ \mathbf{v}_2 + \frac{\langle \mathbf{u},\ \mathbf{v}_n\rangle}{\|\mathbf{v}_n\|^2}\ \mathbf{v}_n \\
&= \frac{(1,2,3)\cdot(1,0,0)}{(1,0,0)\cdot(1,0,0)}\,(1,0,0) + \frac{(1,2,3)\cdot(0,1,1)}{(0,1,1)\cdot(0,1,1)}\,(0,1,1) + \frac{(1,2,3)\cdot(0,1,-1)}{(0,1,-1)\cdot(0,1,-1)}\,(0,1,-1) \\
\mathbf{u} &= (1,0,0) + \frac{5}{2}\,(0,1,1) - \frac{1}{2}\,(0,1,-1)
\end{aligned}$$

(b) Converting the basis vectors to unit vectors, the expression for $\mathbf{u}$ becomes:

$$\mathbf{u} = 1\,(1,0,0) + \frac{5}{2}\sqrt{2}\left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) - \frac{1}{2}\sqrt{2}\left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$

By Theorem 5.12 the norm is:

$$\|\mathbf{u}\| = \sqrt{1^2 + \left(\frac{5\sqrt{2}}{2}\right)^2 + \left(\frac{\sqrt{2}}{2}\right)^2} = \sqrt{14}$$

In comparison the direct calculation of the norm is:

$$\|(1, 2,\ 3)\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$$

(c) The projection of $\mathbf{u}$ onto S is, according to Theorem 5.13:

$$\frac{5}{2}\,(0,1,1) - \frac{1}{2}\,(0,1,-1)$$

(that part of the linear combination for $\mathbf{u}$ from part (a) that involves the basis vectors of S).

(d) From Theorem 5.13, $\mathbf{u}_2 = \frac{5}{2}(0,1,1) - \frac{1}{2}(0,1,-1)$ (from part (c) above) and $\mathbf{u}_1 = (1,0,0)$ - the remaining part of the linear combination of orthogonal basis vectors given in part (a).

Example 5.6.6.
M_{22} is an inner product space with inner product $\langle U,\ V\rangle = \text{trace}\big(U^T V\big)$. Recall that the trace is the sum of the diagonal entries (it is the same as multiplying the entries in the same position in U and V and adding those four products).

(a) Show that the four matrices A_1, A_2, A_3, A_4 are mutually orthogonal.

(b) Show that $\{A_1, A_2, A_3, A_4\}$ is a basis of M_{22}.

(c) Express the matrix B as a linear combination of A_1, A_2, A_3, A_4.

$$A_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},\ A_2 = \begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix},\ A_3 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},\ A_4 = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}$$
$$B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Solution. The following is the solution:

(a) We must show that the trace is zero for each of the six products (there are 12 products but the others are transposes of these and so have the same trace):

$$A_1^T A_2,\ A_1^T A_3,\ A_1^T A_4,\ A_2^T A_3,\ A_2^T A_4,\ A_3^T A_4$$

The first one is:

$$\overset{A_1^T}{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}} \overset{A_2}{\begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix}} = \begin{bmatrix} 1 & -1 \\ 2 & -1 \end{bmatrix} \text{ with trace } 1 + (-1) = 0$$

An alternate, perhaps simpler, way of computing this is to multiply the corresponding entries in each matrix and form the sum:

$$1 \times 2 + 1 \times (-1) + 1 \times (-1) + 0 \times 0 = 0$$

Verify that the other five inner products are also zero. Hence, $\{A_1, A_2, A_3, A_4\}$ is an orthogonal set.

(b) The set $\{A_1, A_2, A_3, A_4\}$ is linearly independent by Theorem 5.10. Hence, it must be a basis of M_{22} since the dimension of M_{22} is four.

(c) By Theorem 5.12, omitting details of the computations of the inner products:

$$B = \frac{\langle B,\ A_1\rangle}{\langle A_1,\ A_1\rangle} A_1 + \frac{\langle B,\ A_2\rangle}{\langle A_2,\ A_2\rangle} A_2 + \frac{\langle B,\ A_3\rangle}{\langle A_3,\ A_3\rangle} A_3 + \frac{\langle B,\ A_4\rangle}{\langle A_4,\ A_4\rangle} A_4$$
$$B = \frac{6}{3} A_1 + \frac{(-3)}{6} A_2 + \frac{(-1)}{2} A_3 + \frac{12}{9} A_4$$

Note: Check for yourself that the result is correct by computing the matrices on the right hand side as follows:

$$B \overset{?}{=} 2\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + \frac{4}{3}\begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}$$

Example 5.6.7.
Requires calculus. Compute the projection of the polynomial $g(x) = x + x^3$ onto the subspace S of P_4, spanned by the first three Legendre polynomials:

$$f_1(x) = 1,\ f_2(x) = x,\ f_3(x) = x^2 - \frac{1}{3}$$

with the inner product defined by:

$$\langle f, g\rangle = \int_{-1}^{1} f\ (x)\, g(x)\ dx$$

Solution. It was shown in Example 5.6.4 that the set $\{f_1,\ \ f_2,\ f_3\}$ is orthogonal and spans the same subspace as $\left\{1,\ x,\ x^2\right\}$. Consequently, it is tempting to suppose that the projection of $g(x)$ on S is the polynomial x (that is, drop the x^3 term). However, we will confirm this, or otherwise, using Theorem 5.13.

According to Theorem 5.13, the projection is given by:

$$\begin{aligned}
\text{proj}_S\ g &= \frac{\langle g,\ f_1\rangle}{\langle f_1,\ f_1\rangle}\ f_1(x) + \ \frac{\langle g,\ f_2\rangle}{\langle f_2,\ f_2\rangle}\ f_2(x) + \ \frac{\langle g,\ f_3\rangle}{\langle f_3,\ f_3\rangle} f_3(x) \\
&= \frac{\int_{-1}^{1} g(x)\ \ f_1(x)\, dx}{\int_{-1}^{1} (f_1(x))^2\ \ dx}\ f_1(x) + \ \frac{\int_{-1}^{1} g(x)\ \ f_2(x)\, dx}{\int_{-1}^{1} (f_2(x))^2\ \ dx}\ f_2(x) + \ \frac{\int_{-1}^{1} g(x)\ \ f_3(x)\, dx}{\int_{-1}^{1} (f_3(x))^2\ \ dx} f_3(x) \\
&= \frac{\int_{-1}^{1} \left(x + x^3\right) dx}{\int_{-1}^{1} 1\ dx}\ \times 1 + \ \frac{\int_{-1}^{1} \left(x + x^3\right) x dx}{\int_{-1}^{1} x^2\ dx}\ x + \ \frac{\int_{-1}^{1} \left(x + x^3\right)\ \ \left(x^2 - \frac{1}{3}\right) dx}{\int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)^2\ \ dx}\ \left(x^2 - \frac{1}{3}\right) \\
&= \frac{0}{2}\ +\ \frac{\left(\frac{16}{15}\right)}{\left(\frac{2}{3}\right)}\ x + \ \frac{0}{\left(\frac{8}{45}\right)}\ \left(x^2 - \frac{1}{3}\right) \\
\text{proj}_S\ g &= \frac{8}{5}\ x
\end{aligned}$$

Note: The intuitive argument above, that $\text{proj}_S\ g = x$, is clearly wrong. This is because the projection of the term x^3 of $g(x)$ onto S is not zero, but is in fact the polynomial $\frac{3}{5}x$ (check this for yourself). Since the other term in $g(x)$ is x, and this is already in S, it follows that the projection of g onto S is $\frac{3}{5}x + x = \frac{8}{5}\ x$ as was found above.

5.6.3 The $QR-$ factorization of a matrix

The Gram-Schmidt Process takes a linearly independent set of vectors $\{\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3, \cdots,\ \mathbf{u}_m\}$ and converts it into an orthogonal set $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3, \cdots,\ \mathbf{v}_m\}$ that spans the same vector space, that can be changed to an orthonormal basis $\{\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3, \cdots,\ \mathbf{w}_m\}$ by normalizing the vectors: $\mathbf{w}_j = \frac{1}{\|\mathbf{v}_j\|}\mathbf{v}_j$. This process can be expressed as a matrix factorization, $A = QR$, called the $QR-$ factorization or $QR-$ decomposition, of the matrix A. In this factorization the columns of A are the vectors $\mathbf{u}_j$, and the columns of Q are the normalized vectors $\mathbf{w}_j$. The matrix R is upper triangular (entries below the main diagonal are zero), and each non-zero row i, column j entry is equal to the inner product of the form $\langle \mathbf{w}_i,\ \mathbf{u}_j\rangle$. In fact the whole Gram-Schmidt orthogonalization process can be carried out very efficiently by working with matrices, rather than vectors.

The $QR-$ factorization is usually applied to Euclidean spaces R^n, and the components of the vectors $\mathbf{u}_i$ are written as columns of the matrix A. Similarly the $\mathbf{w}_i$ components are columns of Q. In this case, if the number of vectors $m = n$, the dimension of the vector space, then the matrix Q will be an $n \times n$ orthogonal matrix. However, the matrix form applies to any inner product space.

Examples 5.6.8 and 5.6.9 show the $QR-$ factorization in a simple two-vector case previously examined in example 5.6.1. The complete result is given in Theorem 5.14.

Example 5.6.8.
Construct the $QR-$ factorization for orthogonalizing any set of two linearly independent vectors $\{\mathbf{u}_1,\ \mathbf{u}_2\}$.

Solution. The set $\{\mathbf{u}_1,\ \mathbf{u}_2\}$ is used to create an orthonormal basis $\{\mathbf{w}_1,\ \mathbf{w}_2\}$. Refer to Theorem 5.9 to see the equations used in the Gram-Schmidt Process which are rewritten here with $\mathbf{u}_1,\ \mathbf{u}_2$ on the left, as follows:

$$\left\{\begin{array}{l}\mathbf{v}_1 = \mathbf{u}_1 \text{ and } \mathbf{w}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 \\ \mathbf{v}_2 = \mathbf{u}_2 - \langle \mathbf{u}_2,\ \mathbf{w}_1\rangle \mathbf{w}_1 \text{ and } \mathbf{w}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2\end{array}\right. \Longrightarrow \left\{\begin{array}{l}\mathbf{u}_1 = \mathbf{v}_1 \\ \mathbf{u}_2 = \mathbf{v}_2 + \langle \mathbf{u}_2,\ \mathbf{w}_1\rangle \mathbf{w}_1\end{array}\right.$$

$$\Longrightarrow \left.\begin{array}{l}\mathbf{u}_1 = \|\mathbf{v}_1\|\,\mathbf{w}_1 + 0 \times \mathbf{w}_2 \\ \mathbf{u}_2 = \langle \mathbf{u}_2,\ \mathbf{w}_1\rangle\ \mathbf{w}_1 + \|\mathbf{v}_2\|\,\mathbf{w}_2\end{array}\right\}$$

Writing the equations as matrix products, with vectors being columns, gives the $QR-$ factorization:

$$\left.\begin{array}{r}[\mathbf{u}_1] = [\mathbf{w}_1 \mid \mathbf{w}_2]\begin{bmatrix}\|\mathbf{v}_1\| \\ 0\end{bmatrix} \\ [\mathbf{u}_2] = [\mathbf{w}_1 \mid \mathbf{w}_2]\begin{bmatrix}\langle \mathbf{u}_2,\ \mathbf{w}_1\rangle \\ \|\mathbf{v}_2\|\end{bmatrix}\end{array}\right\} \Longrightarrow \overset{A}{[\mathbf{u}_1 \mid \mathbf{u}_2]} = \overset{Q}{[\mathbf{w}_1 \mid \mathbf{w}_2]}\overset{R}{\begin{bmatrix}\|\mathbf{v}_1\| & \langle \mathbf{u}_2,\ \mathbf{w}_1\rangle \\ 0 & \|\mathbf{v}_2\|\end{bmatrix}}$$

To give the matrix R a more consistent look, note that $\|\mathbf{v}_1\| = \langle \mathbf{w}_1,\ \mathbf{u}_1\rangle$, and $\|\mathbf{v}_2\| = \langle \mathbf{w}_2,\ \mathbf{u}_2\rangle$ because:

$$\begin{aligned}\langle \mathbf{u}_1,\ \mathbf{w}_1\rangle &= \langle (\|\mathbf{v}_1\|\,\mathbf{w}_1),\ \mathbf{w}_1\rangle = \|\mathbf{v}_1\|\,\langle \mathbf{w}_1,\ \mathbf{w}_1\rangle = \|\mathbf{v}_1\| \\ \langle \mathbf{u}_2,\ \mathbf{w}_2\rangle &= \langle \langle \mathbf{u}_2,\ \mathbf{w}_1\rangle\ \mathbf{w}_1 + \|\mathbf{v}_2\|\,\mathbf{w}_2,\ \mathbf{w}_2\rangle \\ &= \langle \mathbf{u}_2,\ \mathbf{w}_1\rangle\,\langle\ \mathbf{w}_1,\ \mathbf{w}_2\rangle + \|\mathbf{v}_2\|\,\langle \mathbf{w}_2,\ \mathbf{w}_2\rangle \\ &= \|\mathbf{v}_2\| \text{ since } \langle\ \mathbf{w}_1,\ \mathbf{w}_2\rangle = 0 \text{ and } \langle \mathbf{w}_2,\ \mathbf{w}_2\rangle = 1\end{aligned}$$

Hence, using the symmetry of the inner product, $\langle \mathbf{u}_i,\ \mathbf{w}_j\rangle = \langle \mathbf{w}_j,\ \mathbf{u}_i\rangle$, the $A = QR$ formula becomes:

$$\overset{A}{[\mathbf{u}_1 \mid \mathbf{u}_2]} = \overset{Q}{[\mathbf{w}_1 \mid \mathbf{w}_2]}\overset{R}{\begin{bmatrix}\langle \mathbf{w}_1,\ \mathbf{u}_1\rangle & \langle \mathbf{w}_1,\ \mathbf{u}_2\rangle \\ 0 & \langle \mathbf{w}_2,\ \mathbf{u}_2\rangle\end{bmatrix}}$$

Note: That is, the columns $\mathbf{u}_j$ of A are the original vectors, the columns $\mathbf{w}_j$ of Q are the normalized vectors produced by the Gram-Schmidt process, and the row i column j entry of R is $\langle \mathbf{w}_i,\ \mathbf{u}_j\rangle$ (same as $\langle \mathbf{u}_j,\ \mathbf{w}_i\rangle$).

Example 5.6.9.
Find the actual $QR-$ decomposition for Example 5.6.1, that starts with the two vectors in R^2:

$$\{\mathbf{u}_1,\ \ \mathbf{u}_2\} = \{(1,\ 1),\ (1,\ 0)\}$$

Solution. In Example 5.6.1 the Gram-Schmidt process was used to derive the following orthogonal set:

$$\{\mathbf{v}_1,\ \ \mathbf{v}_2\} = \left\{(1,\ 1),\ \left(\frac{1}{2},\ -\frac{1}{2}\right)\right\}$$

Hence, the orthonormal set is:

$$\{\mathbf{w}_1,\ \ \mathbf{w}_2\} = \left\{\left(\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right),\ \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)\right\}$$

The scalar product replaces the inner product in the formulae of the previous Example 5.6.8. Noting that $\mathbf{u}_1 \cdot\ \mathbf{w}_1 = \sqrt{2},\ \mathbf{u}_2 \cdot\ \mathbf{w}_2 = \frac{1}{\sqrt{2}}$, and $\mathbf{u}_2 \cdot\ \mathbf{w}_1 = \frac{1}{\sqrt{2}}$. Putting these values into the formula derived in Example 5.6.8 (with the vectors becoming columns of the matrices):

$$\overset{A}{[\mathbf{u}_1 \mid \mathbf{u}_2]} = \overset{Q}{[\mathbf{w}_1 \mid \mathbf{w}_2]}\overset{R}{\begin{bmatrix}\mathbf{u}_1 \cdot\ \mathbf{w}_1 & \mathbf{w}_1 \cdot \mathbf{u}_2 \\ 0 & \mathbf{w}_2 \cdot\ \mathbf{u}_2\end{bmatrix}}$$

$$\overset{A}{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}} \overset{R}{\begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}}$$

Note: Check for yourself that this decomposition is correct.

Theorem 5.14. *Let $\{\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3, \cdots,\ \mathbf{u}_m\}$ be a set of m linearly independent vectors in an inner product space V of dimension n. If $\{\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3, \cdots,\ \mathbf{w}_m\}$ is the orthonormal set produced by the Gram-Schmidt process (spanning the same subspace as the u_j), produced as in Theorem 5.9, then these satisfy the matrix equation:*

$$A = QR$$

where A has the vectors $\mathbf{u}_j$ as columns, Q is an orthogonal matrix with the vectors $\mathbf{w}_j$ as columns, and R is the non-singular upper triangular matrix given by:

$$A = [\mathbf{u}_1 \mid \mathbf{u}_2 \cdots \mid \mathbf{u}_m] = \overset{Q}{[\mathbf{w}_1 \mid \mathbf{w}_2 \cdots \mid \mathbf{w}_m]} \overset{R}{\begin{bmatrix} \langle \mathbf{w}_1,\ \mathbf{u}_1\rangle & \langle \mathbf{w}_1,\ \mathbf{u}_2\rangle & \langle \mathbf{w}_1,\ \mathbf{u}_3\rangle & \cdots & \langle \mathbf{w}_1,\ \mathbf{u}_m\rangle \\ 0 & \langle \mathbf{w}_2,\ \mathbf{u}_2\rangle & \langle \mathbf{w}_2,\ \mathbf{u}_3\rangle & \cdots & \langle \mathbf{w}_2,\ \mathbf{u}_m\rangle \\ 0 & 0 & \langle \mathbf{w}_3,\ \mathbf{u}_3\rangle & \cdots & \langle \mathbf{w}_3,\ \mathbf{u}_m\rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \langle \mathbf{w}_m,\ \mathbf{u}_m\rangle \end{bmatrix}}$$

Proof. Multiplying out the matrix product in $A = QR$ gives the same formulae as the Gram-Schmidt orthonormalization formulae of Theorem 5.9

However the details are complicated, but you may be able to prove it yourself using the results that $\mathbf{v}_j = \|\mathbf{v}_j\|\,\mathbf{w}_j$ and $\|\mathbf{v}_j\| = \langle \mathbf{u}_j,\ \mathbf{w}_j\rangle$ for each $j = 1, 2, \cdots, m$.

□

Example 5.6.10.
Given the basis $\{\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3\} = \{(1,\ 1,\ 1),\ (1,\ 0,\ -1),\ (2,\ -3,\ 4)\}$ of R^3, write out the associated $QR-$ factorization, using the Gram-Schmidt orthogonalization previously found in Example 5.6.2.

Solution. Example 5.6.2 used the Gram-Schmidt orthogonalization algorithm to find the orthogonal set of vectors:

$$\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3\} = \{(1,\ 1,\ 1),\ (1,\ 0\ -1),\ (2, -4,\ 2)\}$$

Normalizing these gives the orthonormal set:

$$\{\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3\} = \left\{ \left(\frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}}\right),\ \left(\frac{1}{\sqrt{2}},\ 0\ -\frac{1}{\sqrt{2}}\right),\ \left(\frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}},\ \frac{1}{\sqrt{6}}\right) \right\}$$

Using the scalar product as the inner product, the $QR-$ factorization is therefore:

$$\overset{A}{[\mathbf{u}_1 \mid \mathbf{u}_2 \cdots \mid \mathbf{u}_m]} = \overset{Q}{[\mathbf{w}_1 \mid \mathbf{w}_2 \cdots \mid \mathbf{w}_m]} \overset{R}{\begin{bmatrix} \mathbf{w}_1 \cdot \mathbf{u}_1 & \mathbf{w}_1 \cdot \mathbf{u}_2 & \mathbf{w}_1 \cdot \mathbf{u}_3 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \mathbf{w}_2 \cdot \mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}}$$

The inner products of R are easily calculated
as: $\langle \mathbf{w}_1 \cdot \mathbf{u}_1\rangle = \sqrt{3},\ \mathbf{w}_1 \cdot \mathbf{u}_2 = 0,\ \mathbf{w}_1 \cdot \mathbf{u}_3 = \sqrt{3},\ \mathbf{w}_2 \cdot \mathbf{u}_2 = \sqrt{2},\ \mathbf{w}_2 \cdot \mathbf{u}_3 = -\sqrt{2},\ \mathbf{w}_3 \cdot \mathbf{u}_3 = 2\sqrt{6}$, so the $QR-$ factorization is:

$$\overset{A}{\begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & -3 \\ 1 & -1 & 4 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}} \overset{R}{\begin{bmatrix} \sqrt{3} & 0 & \sqrt{3} \\ 0 & \sqrt{2} & -\sqrt{2} \\ 0 & 0 & 2\sqrt{6} \end{bmatrix}}$$

Note: Check for yourself that the matrix equation is correct, and that Q is orthogonal ($QQ^T = I$).

Example 5.6.11.
Find an orthonormal set of vectors spanning the same subspace of R^4 as the vectors;

$$\mathbf{u}_1 = (1,0,0,2)\,,\ \mathbf{u}_2 = (0,\ 1,\ 1,0)\,,\ \mathbf{u}_3 = (1,\ 1,0,-1)$$

carrying out the calculations within the matrices $A = QR$.

Solution. The following is the solution:
Step 1: By Theorem 5.9, the matrices have the following form, with the first column of Q being the normalized $\mathbf{u}_1$. The first row of R is then computed:

$$\overset{A}{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & -1 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} \frac{1}{\sqrt{5}} & * & * \\ 0 & * & * \\ 0 & * & * \\ \frac{2}{\sqrt{5}} & * & * \end{bmatrix}} \overset{R}{\begin{bmatrix} \mathbf{w}_1\cdot\mathbf{u}_1 & \mathbf{w}_1\cdot\mathbf{u}_2 & \mathbf{w}_1\cdot\mathbf{u}_3 \\ 0 & \mathbf{w}_2\cdot\mathbf{u}_2 & \mathbf{w}_2\cdot\mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3\cdot\mathbf{u}_3 \end{bmatrix}}$$
$$= \begin{bmatrix} \frac{1}{\sqrt{5}} & * & * \\ 0 & * & * \\ 0 & * & * \\ \frac{2}{\sqrt{5}} & * & * \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 & -\frac{1}{\sqrt{5}} \\ 0 & \mathbf{w}_2\cdot\mathbf{u}_2 & \mathbf{w}_2\cdot\mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3\cdot\mathbf{u}_3 \end{bmatrix}$$

Step 2: The second column of Q is equal to the vector
$\mathbf{v}_2 = \mathbf{u}_2 - \langle \mathbf{u}_2,\ \mathbf{w}_1\rangle\,\mathbf{w}_1 = (0,\ 1,\ 1,0) - 0\times\left(\frac{1}{\sqrt{5}},0,0,\frac{2}{\sqrt{5}}\right) = (0,\ 1,\ 1,0)$, which is then normalized to $\mathbf{w}_2 = \left(0,\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right)$. The second row of R is then computed:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & * \\ 0 & \frac{1}{\sqrt{2}} & * \\ 0 & \frac{1}{\sqrt{2}} & * \\ \frac{2}{\sqrt{5}} & 0 & * \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 - \frac{1}{\sqrt{5}} & \\ 0 & \mathbf{w}_2\cdot\mathbf{u}_2 & \mathbf{w}_2\cdot\mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3\cdot\mathbf{u}_3 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & * \\ 0 & \frac{1}{\sqrt{2}} & * \\ 0 & \frac{1}{\sqrt{2}} & * \\ \frac{2}{\sqrt{5}} & 0 & * \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 & -\frac{1}{\sqrt{5}} \\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \mathbf{w}_3\cdot\mathbf{u}_3 \end{bmatrix}$$

Step 3: The third column of Q is equal to the vector:

$$\begin{aligned} \mathbf{v}_3 &= \mathbf{u}_3 - \langle \mathbf{u}_3,\ \mathbf{w}_1\rangle\,\mathbf{w}_1 - \langle \mathbf{u}_3,\ \mathbf{w}_2\rangle\,\mathbf{w}_2 \\ &= (1,\ 1,\ 0,-1) + \frac{1}{\sqrt{5}}\left(\frac{1}{\sqrt{5}},0,0,\frac{2}{\sqrt{5}}\right) - \frac{1}{\sqrt{2}}\left(0,\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right) \\ &= 1\times\left(\frac{6}{5},\frac{1}{2},-\frac{1}{2},-\frac{3}{5}\right) \end{aligned}$$

which is then normalized to $\mathbf{w}_3 = \frac{\sqrt{10}}{\sqrt{23}}\left(\frac{6}{5},\frac{1}{2},-\frac{1}{2},-\frac{3}{5}\right)$. The bottom right entry of R is then $\mathbf{w}_3\cdot\mathbf{u}_3 = \frac{\sqrt{23}}{\sqrt{10}}$:

$$\overset{A}{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & -1 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & \frac{6\sqrt{10}}{5\sqrt{23}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{\sqrt{10}}{2\sqrt{23}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{\sqrt{10}}{2\sqrt{23}} \\ \frac{2}{\sqrt{5}} & 0 & -\frac{3\sqrt{10}}{5\sqrt{23}} \end{bmatrix}} \overset{R}{\begin{bmatrix} \sqrt{5} & 0 & -\frac{1}{\sqrt{5}} \\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \sqrt{\frac{23}{10}} \end{bmatrix}}$$

Note: Verify for yourself that the matrix equation is correct.
Note: Some authors make Q into an orthogonal matrix by adding appropriate extra columns (one is needed in this case), and by adding rows of zeros to the bottom of R to make the matrix multiplication compatible. The extra column of Q is found by choosing an arbitrary fourth vector $\mathbf{u}_4$ (it is very unlikely to be in the span of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$) and then continuing the Gram-Schmidt algorithm for one more step, but leaving out the extra column in the final matrix equation. This form of the factorization in this example is:

$$\overbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & -1 \end{bmatrix}}^{A} = \overbrace{\begin{bmatrix} \frac{1}{\sqrt{5}} & 0 & \frac{6\sqrt{10}}{5\sqrt{23}} & \frac{2}{\sqrt{23}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{\sqrt{10}}{2\sqrt{23}} & -\frac{3}{\sqrt{23}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{\sqrt{10}}{2\sqrt{23}} & \frac{3}{\sqrt{23}} \\ \frac{2}{\sqrt{5}} & 0 & -\frac{3\sqrt{10}}{5\sqrt{23}} & -\frac{1}{\sqrt{23}} \end{bmatrix}}^{Q} \overbrace{\begin{bmatrix} \sqrt{5} & 0 & -\frac{1}{\sqrt{5}} \\ 0 & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \sqrt{\frac{23}{10}} \\ 0 & 0 & 0 \end{bmatrix}}^{R}$$

For completeness we give a $QR-$ factorization for P_3, but in fact this matrix factorization is rarely used outside of Euclidean spaces.

Example 5.6.12.
Referring to Example 5.6.3, starting with the polynomials $\{g_1(x), \ g_2(x), \ g_3(x), \ g_4(x)\} = \left\{1, \ 1+x, \ 1-2x^2, \ x+x^3\right\}$ and inner product defined by:

$$\left\langle a_0 + a_1x + a_2x^2 + a_3x^3, \ b_0 + b_1x + b_2x^2 + b_3x^3 \right\rangle = a_0b_0 + a_1b_1 + a_2b_2 + a_3b_3$$

write the Gram-Schmidt orthogonalization as a $QR-$ factorization.

Solution. The orthogonal set found in Example 5.6.3 is:

$$\{f_1(x), \ f_2(x), \ f_3(x), \ f_4(x)\} = \left\{1, \ x, \ -2x^2, \ x^3\right\}$$

Normalizing these gives the orthonormal basis of P_3 :

$$\{h_1(x), \ h_2(x), \ h_3(x), \ h_4(x)\} = \left\{1, \ x, \ x^2, \ x^3\right\}$$

The $QR-$ factorization is therefore:

$$\begin{bmatrix} g_1(x) & g_2(x) & g_3(x) & g_4(x) \end{bmatrix} =$$

$$\begin{bmatrix} h_1(x) & h_2(x) & h_3(x) & h_4(x) \end{bmatrix} \begin{bmatrix} \langle h_1, \ g_1 \rangle & \langle h_1, \ g_2 \rangle & \langle h_1, \ g_3 \rangle & \langle h_1, \ g_4 \rangle \\ 0 & \langle h_2, \ g_2 \rangle & \langle h_2, \ g_3 \rangle & \langle h_2, \ g_4 \rangle \\ 0 & 0 & \langle h_3, \ g_3 \rangle & \langle h_3, \ g_4 \rangle \\ 0 & 0 & 0 & \langle h_4, \ g_4 \rangle \end{bmatrix}$$

and $\langle h_1, \ g_1 \rangle = 1, \ \langle h_1, \ g_2 \rangle = 1, \ \langle h_1, \ g_3 \rangle = 1 \ \langle h_1, \ g_3 \rangle = 0, \ \langle h_2, \ g_2 \rangle = 1, \ \langle h_2, \ g_3 \rangle = 0, \ \langle h_2, \ g_3 \rangle = 1,$ $\langle h_3, \ g_3 \rangle = -2, \ \langle h_3, \ g_4 \rangle = 0, \ \langle h_4, \ g_4 \rangle = 1$, so

$$\overbrace{\begin{bmatrix} 1 & 1+x & 1-2x^2 & x+x^3 \end{bmatrix}}^{A} = \overbrace{\begin{bmatrix} 1 & x & x^2 & x^3 \end{bmatrix}}^{Q} \overbrace{\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{R}$$

Note: Check for yourself that the matrix equality is correct.

Section 5.6 exercise set

Check your understanding by answering the following questions.

1. Let S be the subspace of R^3 spanned by the following set of two vectors:

$$\{(1,\ 2,\ -1),\ (3,\ 0,\ 3)\}$$

   (a) Check that the set of vectors is orthogonal.

   (b) Convert it to an orthonormal set.

   (c) Extend it to an orthonormal basis of R^3 (Hint: add to the two orthogonal vectors any other vector not in S - a random choice usually works - and apply one step of the Gram-Schmidt process).

   (d) Find the projection of the vector $(1, 2, 3)$ onto S (use Theorem 5.13).

2. For the following set of vectors in R^3 :

$$\{(1,\ 2,\ 0),\ (1,\ 1,\ 1),\ (3,\ -2,\ 1)\}$$

   (a) Use the Gram-Schmidt process of Theorem 5.8 to find an orthogonal basis of R^3.

   (b) Find an orthonormal basis of R^3.

   (c) Can you be certain that the original three vectors are linearly independent?

3. Apply the Gram-Schmidt orthogonalization process (Theorem 5.8) to the following set of vectors in R^4. What do you conclude about this set?

$$\{(1,\ 2,\ 0,\ -3),\ (1,\ 1,\ 1,\ 1),\ (1,\ 0,\ 2,\ 5)\}$$

4. For the following set of vectors in R^3 :

$$\{(1,\ 0,\ -1),\ (1,\ 1,\ 0),\ (1,\ 1,\ 1)\}$$

   (a) Use the Gram-Schmidt process of Theorem 5.8 to find an orthonormal basis of R^3.

   (b) Use the Gram-Schmidt process of Theorem 5.9 to find an orthonormal basis of R^3. Discuss which of the two methods in (a) and (b) is easiest.

   (c) Use Theorem 5.12 to find linear combination of the basis vectors, giving the two vectors $\mathbf{u}_1 = (1, 3, -2)\,, \mathbf{u}_2 = (4, 0, 1)$.

   (d) Show that Theorem 5.11 gives the correct values for $\mathbf{u}_1 \cdot \mathbf{u}_2$, for the norms of $\mathbf{u}_1$. and $\mathbf{u}_2$, and for the distance between $\mathbf{u}_1$ and $\mathbf{u}_2$.

5. The following set of polynomials is a basis of P_2:

$$\left\{1 - x^2,\ x + 2x^2,\ -2 + x + 3x^2\right\}$$

   (a) Find an orthogonal basis using the Gram-Schmidt orthogonalization process (Theorem 5.8) with the inner product of Example 5.6.3 (scalar product of the vectors of coefficients).

   (b) Find an orthogonal basis using the Gram-Schmidt orthogonalization process (Theorem 5.8) with the inner product:

$$\left\langle a_0 + a_1 x + a_2 x^2,\ b_0 + b_1 x + b_2 x^2 \right\rangle = a_0 b_0 + 2 a_1 b_1 + 3 a_2 b_2$$

   (c) **Requires calculus and more difficult.** Find an orthogonal basis using the Gram-Schmidt orthogonalization process (Theorem 5.8) with the inner product of Example 5.6.7 (the integral of the product from $x = -1$ to $x = 1$).

   (d) Find orthonormal basis with the inner product of part (a).

   (e) Find orthonormal basis with the inner product of part (b).

6. For the basis vectors found in the previous question

   (a) Express the polynomial $2+3x-4x^2$ as a linear combination of the orthogonal polynomials found in part (a) of the previous question (see Theorems 5.11and 5.12).

   (b) Express the polynomial $2+3x-4x^2$ as a linear combination of the orthogonal polynomials found in part (b)of the previous question.

   (c) Express the Gram-Schmidt process covered in part (a) and (d) of the previous question as a $QR-$ factorization.

7. Apply the Gram-Schmidt orthogonalization process (Theorem 5.8) with the inner product of Example 5.6.3 (scalar product of the vectors of coefficients).to the set of polynomials in P_3

$$\left\{1-x^2,\ x+2x^2,\ -2+x+4x^2\right\}$$

   What do you conclude?

8. Let S be the subspace of M_{22} spanned by the three matrices:

$$S=\left\{\begin{bmatrix}1&3\\-2&0\end{bmatrix},\begin{bmatrix}2&1\\2&1\end{bmatrix},\begin{bmatrix}0&3\\2&1\end{bmatrix}\right\}$$

   (a) Find an orthogonal basis of S using the Gram-Schmidt orthogonalization process (Theorem 5.8) with the inner product of Example 5.6.6 (scalar product of the $4-$vectors formed by the entries of the matrices).

   (b) Find an orthonormal basis of S.

   (c) Attempt to express the following matrix as a linear combination of the orthonormal basis matrices of S (see Theorems 5.11 and 5.12):

$$\begin{bmatrix}5&2\\0&1\end{bmatrix}$$

   (d) Attempt to express the following matrix as a linear combination of the orthonormal basis matrices of S :

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

9. **Requires calculus and is more difficult.** Starting with the following five basis polynomials of P_4 and the inner product $\langle f,\ g\rangle=\int_{-1}^{1}f(x)\,g(x)\ dx$:

$$\left\{1,\ x,\ x^2,\ x^3,\ x^4\right\}$$

   (a) Use the Gram-Schmidt orthogonalization process (Theorem 5.8) to find the fifth function of the orthogonal basis (the first four are given in Example 5.6.4).
   Note: These orthogonal basis polynomials are multiples of the first four **Legendre Polynomials** that are very useful in some Engineering applications.

   (b) Convert the basis from part (a) into an orthonormal basis of P_4.

10. Let W be the subspace of R^4 spanned by the two vectors:

$$\{(1,\ 0,\ 0,\ 1),\ (1,\ 2,\ 0,\ 1)\}$$

   (a) Find a basis of $W^{\perp}$.

   (b) If $\mathbf{u}=(2,3,0,4)$ then find two (unique) vectors $\mathbf{u}_1\in W,\ \mathbf{u}_2\in W^{\perp}$ such that $\mathbf{u}=\mathbf{u}_1+\mathbf{u}_2$.

   (c) Find the $QR-$ factorization equivalent to the Gram-Schmidt process in part (a) (that is, including all four basis vectors of W and $W^{\perp}$).

11. Find, if possible, the $QR-$ factorization of the following matrices:

$$\text{(a) } A=\begin{bmatrix}1&3\\0&4\end{bmatrix},\ \text{(b) } B=\begin{bmatrix}1&1&0\\0&1&1\\1&-1&1\end{bmatrix},\ \text{(c) } C=\begin{bmatrix}3&5\\4&0\\0&-2\end{bmatrix},\ \text{(d) } D=\begin{bmatrix}1&0&2\\0&1&3\\2&-1&1\end{bmatrix}$$

Solutions

1. For the vectors $(1,\ 2,\ -1)\,,\ (3,\ 0,\ 3):$

(a) They are orthogonal since $(1,\ 2,\ -1)\cdot(3,\ 0,\ 3)=1\times 3+2\times 0+(-1)\times 3=0.$

(b) Normalizing (dividing the vector by its norm/length) the vectors gives the orthonormal set:

$$\mathbf{w}_1=\left(\frac{1}{\sqrt{6}},\ \frac{2}{\sqrt{6}},\ -\frac{1}{\sqrt{6}}\right),\ \mathbf{w}_2=\left(\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}}\right)$$

(c) Choose, randomly, the vector $\mathbf{u}=(1,\ 0,\ 0)$ and apply the last step of the Gram-Schmidt algorithm (Theorem 5.9) to the vector $\{\mathbf{w}_1,\mathbf{w}_2,\mathbf{u}\}$ to form an orthonormal set $\{\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3\}$ using the intermediate vector $\mathbf{v}$:

$$\begin{aligned}\mathbf{v}&=\mathbf{u}-(\mathbf{w}_1\cdot\mathbf{u})\,\mathbf{w}_1-(\mathbf{w}_2\cdot\mathbf{u})\,\mathbf{w}_2\\&=(1,\ 0,\ 0)-\frac{1}{\sqrt{6}}\left(\frac{1}{\sqrt{6}},\ \frac{2}{\sqrt{6}},\ -\frac{1}{\sqrt{6}}\right)-\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}}\right)\\&=(1,\ 0,\ 0)-\left(\frac{1}{6},\ \frac{1}{3},\ -\frac{1}{6}\right)-\left(\frac{1}{2},\ 0,\ \frac{1}{2}\right)\\\mathbf{v}&=\left(\frac{1}{3},\ -\frac{1}{3},\ -\frac{1}{3}\right)\end{aligned}$$

Normalizing $\mathbf{v}$ gives the required result $\mathbf{w}_3=\left(\frac{1}{\sqrt{3}},\ -\frac{1}{\sqrt{3}},\ -\frac{1}{\sqrt{3}}\right)$ and the orthonormal basis:

$$\mathbf{w}_1=\left(\frac{1}{\sqrt{6}},\ \frac{2}{\sqrt{6}},\ -\frac{1}{\sqrt{6}}\right),\ \mathbf{w}_2=\left(\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}}\right),\ \mathbf{w}_3=\left(\frac{1}{\sqrt{3}},\ -\frac{1}{\sqrt{3}},\ -\frac{1}{\sqrt{3}}\right)$$

(d) The projection of the vector $\mathbf{u}=(1,2,3)$ onto the subspace S with orthonormal basis $\{\mathbf{w}_1,\ \mathbf{w}_2\}$ is given by (see Theorem 5.13):

$$\begin{aligned}\text{proj}_S\,\mathbf{u}&=\langle\mathbf{u},\ \mathbf{w}_1\rangle\ \mathbf{w}_1+\ \langle\mathbf{u},\ \mathbf{w}_2\rangle\ \mathbf{w}_2\\&=\left[(1,2,3)\cdot\left(\frac{1}{\sqrt{6}},\ \frac{2}{\sqrt{6}},\ -\frac{1}{\sqrt{6}}\right)\right]\mathbf{w}_1+\left[(1,2,3)\cdot\left(\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}}\right)\right]\mathbf{w}_2\\&=\frac{1}{3}\sqrt{6}\left(\frac{1}{\sqrt{6}},\ \frac{2}{\sqrt{6}},\ -\frac{1}{\sqrt{6}}\right)+2\sqrt{2}\left(\frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}}\right)\\&=\left(\frac{1}{3},\ \frac{2}{3},\ -\frac{1}{3}\right)+(2,\ 0,\ 2)\\\text{proj}_S\,\mathbf{u}&=\left(\frac{7}{3},\ \frac{2}{3},\ \frac{5}{3}\right)\end{aligned}$$

2. For the vectors:

$$\mathbf{u}_1=(1,\ 2,\ 0)\,,\ \mathbf{u}_2=(1,\ 1,\ 0)\,,\ \mathbf{u}_3=(3,\ -2,\ 1)$$

(a) From Theorem 5.8, steps 1, 2 and 3:

$$\mathbf{v}_1 = \mathbf{u}_1 = (1,\ 2,\ 0)$$

$$\begin{aligned}\mathbf{v}_2 &= \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\ &= (1,\ 1,\ 0) - \frac{3}{5}(1,\ 2,\ 0) \\ \mathbf{v}_2 &= \left(\frac{2}{5},\ -\frac{1}{5},\ 0\right)\end{aligned}$$

$$\begin{aligned}\mathbf{v}_3 &= \mathbf{u}_3 - \frac{\mathbf{u}_3 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{u}_3 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 \\ &= (3,\ -2,\ 1) + \frac{1}{5}(1,\ 2,\ 0) - 8\left(\frac{2}{5},\ -\frac{1}{5},\ 0\right) \\ \mathbf{v}_3 &= (0,\ 0,\ 1)\end{aligned}$$

Note: Check for yourself that the three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are mutually orthogonal by showing $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0, \mathbf{v}_1 \cdot \mathbf{v}_3 = 0, , \mathbf{v}_2 \cdot \mathbf{v}_3 = 0.$

(b) Normalizing the three vectors from part (a) gives the orthonormal basis:

$$\left\{\left(\frac{1}{\sqrt{5}},\ \frac{2}{\sqrt{5}},\ 0\right),\ \left(\frac{2}{\sqrt{5}},\ -\frac{1}{\sqrt{5}},\ 0\right),\ (0,\ 0,\ 1)\right\}$$

(c) The original vectors must be linearly independent because otherwise the Gram-Schmidt process would not be able to produce three mutually orthogonal (and therefor linearly independent by Theorem 5.10) vectors as linear combinations of the original vectors. For example, if $\mathbf{u}_3$ is linearly dependent on $\mathbf{u}_1$ and $\mathbf{u}_2$ then $\mathbf{v}_3$ is also linearly dependent on $\mathbf{u}_1$ and $\mathbf{u}_2$. This is because $\mathbf{v}_3$ is defined as a linear combination of $\mathbf{u}_3,\ \mathbf{v}_1,\ \mathbf{v}_2$, and each of these is itself a linear combination of, or is equal to, one of, $\mathbf{u}_1$ and $\mathbf{u}_2$. In that case $\mathbf{v}_3$ could not be orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$, since orthogonal vectors are always linearly independent (Theorem 5.10).
Note: If the Gram-Schmidt process is applied to three vectors $\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3\ \mathbf{u}_3$ for which $\mathbf{u}_3$ is linearly dependent on $\mathbf{u}_1$ and $\mathbf{u}_2$ then a zero vector will be produced: $\mathbf{v}_3 = (0,0,0)$. In fact whenever the Gram-Schmidt process produces a zero vector, it shows that the original vectors used up to that point are linearly dependent. An example of this is in the next question.

3. For the set of vectors:

$$\{\mathbf{u}_1 = (1,\ 2,\ 0,\ -3),\ \mathbf{u}_2 = (1,\ 1,\ 1,\ 1),\ \mathbf{u}_3 = (1,\ 0,\ 2,\ 5)\}$$

From Theorem 5.8, steps 1, 2 and 3:

$$\mathbf{v}_1 = \mathbf{u}_1 = (1,\ 2,\ 0,\ -3)$$

$$\begin{aligned}\mathbf{v}_2 &= \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\ &= (1,\ 1,\ 1,\ 1) - \frac{0}{5}(1,\ 2,\ 0,\ -3) \\ \mathbf{v}_2 &= (1,\ 1,\ 1,\ 1)\end{aligned}$$

That is $\mathbf{v}_2 = \mathbf{u}_2$ because $\mathbf{u}_2$ was already orthogonal to $\mathbf{v}_1$. Finally,

$$\begin{aligned}
\mathbf{v}_3 &= \mathbf{u}_3 - \frac{\mathbf{u}_3 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{u}_3 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 \\
&= (1,\ 0,\ 2,\ 5) + \frac{14}{14}(1,\ 2,\ 0, -3) - \frac{8}{4}(1,\ 1,\ 1,\ 1) \\
&= (1,\ 0,\ 2,\ 5) + (1,\ 2,\ 0, -3) - (2,\ 2,\ 2,\ 2) \\
\mathbf{v}_3 &= (0,\ 0,\ 0,\ 0)
\end{aligned}$$

Hence, $\mathbf{v}_3 = \mathbf{0}$, the zero vector, and this indicates that it is not possible to find a non-zero vector orthogonal to $\mathbf{v}_1$ and $\mathbf{v}_2$. This only happens when the original set of vectors is linearly dependent. **Note:** Check for yourself that the linear dependence is: $\mathbf{u}_3 = 2\mathbf{u}_2 - \mathbf{u}_1$.

4. For the vectors:

$$\mathbf{u}_1 = (1,\ 0,\ -1),\ \mathbf{u}_2 = (1,\ 1,\ 0),\ \mathbf{u}_3 = (1,\ 2,\ -1)$$

(a) From Theorem 5.8, steps 1, 2 and 3:

$$\mathbf{v}_1 = \mathbf{u}_1 = (1,\ 0,\ -1)$$

$$\begin{aligned}
\mathbf{v}_2 &= \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\
&= (1,\ 1,\ 0) - \frac{1}{2}(1,\ 0,\ -1) \\
\mathbf{v}_2 &= \left(\frac{1}{2},\ 1,\ \frac{1}{2}\right)
\end{aligned}$$

$$\begin{aligned}
\mathbf{v}_3 &= \mathbf{u}_3 - \frac{\mathbf{u}_3 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{u}_3 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 \\
&= (1,\ 2,\ -1) - \frac{2}{2}(1,\ 0,\ -1) - \frac{2}{\left(\frac{3}{2}\right)}\left(\frac{1}{2},\ 1,\ \frac{1}{2}\right) \\
\mathbf{v}_3 &= \left(-\frac{2}{3},\ \frac{2}{3},\ -\frac{2}{3}\right)
\end{aligned}$$

Hence, simplifying by multiplying two of the vectors by an appropriate constant, an orthogonal set is:

$$\left\{\mathbf{v}_1,\ 2\mathbf{v}_2,\ -\frac{3}{2}\mathbf{v}_3\right\} = \{(1,\ 0,\ -1),\ (1,\ 2,\ 1),\ (1,\ -1,\ 1)\}$$

Hence, normalizing the vectors gives the orthonormal set:

$$\begin{aligned}
\{\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3\} &= \left\{\left(\frac{1}{\sqrt{2}},\ 0,\ -\frac{1}{\sqrt{2}}\right),\ \left(\frac{1}{\sqrt{6}},\ \frac{2}{\sqrt{6}},\ \frac{1}{\sqrt{6}}\right),\ \left(\frac{1}{\sqrt{3}},\ -\frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}}\right)\right\} \\
\{\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3\} &= \left\{\frac{1}{\sqrt{2}}(1,\ 0, -1),\ \frac{1}{\sqrt{6}}(1,\ 2,\ 1),\ \frac{1}{\sqrt{3}}(1, -1,\ 1)\right\}
\end{aligned}$$

(b) Repeat the process, using Theorem 5.9, steps 1, 2 and 3, starting with the vectors $\mathbf{u}_1 = (1,\ 0,\ -1),\ \mathbf{u}_2 = (1,\ 1,\ 0),\ \mathbf{u}_3 = (1,\ 2,\ -1)$:

$$\mathbf{w}_1 = \frac{1}{\|\mathbf{u}_1\|}\mathbf{u}_1 = \left(\frac{1}{\sqrt{2}},\ 0,\ -\frac{1}{\sqrt{2}}\right)$$

$$
\begin{aligned}
\mathbf{v}_2 &= \mathbf{u}_2 - (\mathbf{u}_2 \cdot \mathbf{w}_1)\,\mathbf{w}_1 \\
&= (1,\ 1,\ 0) - \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}},\ 0,\ \ -\frac{1}{\sqrt{2}}\right) = \left(\frac{1}{2},\ 1,\ \frac{1}{2}\right) \\
\mathbf{w}_2 &= \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \left(\frac{1}{\sqrt{6}},\ \frac{2}{\sqrt{6}},\ \frac{1}{\sqrt{6}}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{v}_3 &= \mathbf{u}_3 - (\mathbf{u}_3 \cdot \mathbf{w}_1)\,\mathbf{w}_1 - (\mathbf{u}_3 \cdot \mathbf{w}_2)\,\mathbf{w}_2 \\
&= (1,\ 2,\ \ -1) - \sqrt{2}\left(\frac{1}{\sqrt{2}},\ 0,\ \ -\frac{1}{\sqrt{2}}\right) - \frac{4}{\sqrt{6}}\left(\frac{1}{\sqrt{6}},\ \frac{2}{\sqrt{6}},\ \frac{1}{\sqrt{6}}\right) = \left(-\frac{2}{3},\ \frac{2}{3},\ \ -\frac{2}{3}\right) \\
\mathbf{w}_3 &= \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 = \left(= \frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}},\ \ -\frac{1}{\sqrt{3}}\right)
\end{aligned}
$$

Note: There is a slight difference from part (a) in that $\mathbf{w}_3$ is the negative of the vector found in part (a) but that just reflects the fact that a unit vector along a given line can have two different directions.

Part (b) has simpler formulae and fewer computations (for example, in part (a) division by $\|\mathbf{v}_1\|$ occurs at step (2), (3) and in the final normalization, but only occurs once in part (b). However, when using hand calculations the part (b) method produces more complicated calculations.

(c) With $\mathbf{u}_1 = (1, 3, -2)$, $\mathbf{u}_2 = (4, 0, 1)$ and using $\{\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3\}$ from part (a), by Theorem 5.12:

$$\mathbf{u}_1 = (\mathbf{u}_1 \cdot \mathbf{w}_1)\ \mathbf{w}_1 + (\mathbf{u}_1 \cdot \mathbf{w}_2)\ \mathbf{w}_2 + (\mathbf{u}_1 \cdot \mathbf{w}_3)\ \mathbf{w}_3$$

$$= \frac{1}{\sqrt{2}}\left((1,3,-2)\cdot(1,0,-1)\right)\ \mathbf{w}_1 + \frac{1}{\sqrt{6}}\left((1,3,-2)\cdot(1,2,1)\right)\ \mathbf{w}_2 + \frac{1}{\sqrt{3}}\left((1,3,-2)\cdot(1,-1,1)\right)\ \mathbf{w}_3$$

$$\mathbf{u}_1 = \frac{3}{\sqrt{2}}\mathbf{w}_1 + \frac{5}{\sqrt{6}}\mathbf{w}_2 + \left(-\frac{4}{\sqrt{3}}\right)\ \mathbf{w}_3$$

Similarly, show for yourself that:

$$\mathbf{u}_2 = (\mathbf{u}_2 \cdot \mathbf{w}_1)\ \mathbf{w}_1 + (\mathbf{u}_2 \cdot \mathbf{w}_2)\ \mathbf{w}_2 + (\mathbf{u}_2 \cdot \mathbf{w}_3)\ \mathbf{w}_3 = \frac{3}{\sqrt{2}}\mathbf{w}_1 + \frac{5}{\sqrt{6}}\mathbf{w}_2 + \frac{5}{\sqrt{3}}\mathbf{w}_3$$

(d) Using the coefficients of $\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3$ found in part (c) for $\mathbf{u}_1$ and $\mathbf{u}_2$ in the formulae of Theorem 5.11 gives:

$$
\begin{aligned}
\mathbf{u}_1 \cdot \mathbf{u}_2 &= \frac{3}{\sqrt{2}} \times \frac{3}{\sqrt{2}} + \frac{5}{\sqrt{6}} \times \frac{5}{\sqrt{6}} + \left(-\frac{4}{\sqrt{3}}\right) \times \frac{5}{\sqrt{3}} = 2 \\
\|\mathbf{u}_1\| &= \sqrt{\left(\frac{3}{\sqrt{2}}\right)^2 + \left(\frac{5}{\sqrt{6}}\right)^2 + \left(-\frac{4}{\sqrt{3}}\right)^2} = \sqrt{14} \\
\|\mathbf{u}_2\| &= \sqrt{\left(\frac{3}{\sqrt{2}}\right)^2 + \left(\frac{5}{\sqrt{6}}\right)^2 + \left(\frac{5}{\sqrt{3}}\right)^2} = \sqrt{17} \\
d(\mathbf{u}_1, \mathbf{u}_2) &= \sqrt{\left(\frac{3}{\sqrt{2}} - \frac{3}{\sqrt{2}}\right)^2 + \left(\frac{5}{\sqrt{6}} - \frac{5}{\sqrt{6}}\right)^2 + \left(-\frac{4}{\sqrt{3}} - \frac{5}{\sqrt{3}}\right)^2} = \sqrt{27}
\end{aligned}
$$

Computing $\mathbf{u}_1 \cdot \mathbf{u}_2$, $\|\mathbf{u}_1\|$, $\|\mathbf{u}_2\|$, $d(\mathbf{u}_1, \mathbf{u}_2)$ for $\mathbf{u}_1 = (1, 3, -2)$, $\mathbf{u}_2 = (4, 0, 1)$ in the usual way confirms that these values are correct. That is:

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = 1 \times 4 + 3 \times 0 + (-2) \times 1 = 2$$

$\|\mathbf{u}_1\| = \sqrt{1^2 + 3^2 + (-2)^2} = \sqrt{14}$ and $\|\mathbf{u}_2\| = \sqrt{4^2 + 0^2 + 1^2} = \sqrt{17}$

$$d(\mathbf{u}_1, \mathbf{u}_2) = \sqrt{(1-4)^2 + (3-0)^2 + (-2-1)^2} = \sqrt{27}$$

5.

$$g_1(x) = 1 - x^2,\ g_2(x) = x + 2x^2,\ g_3(x) = -2 + x + 3x^2$$

(a) From Theorem 5.8, steps 1, 2 and 3 (using $f_1,\ f_2,\ f_3$ as the orthogonal polynomials) with inner product:

$$\left\langle a_0 + a_1x + a_2x^2,\ b_0 + b_1x + b_2x^2 \right\rangle = a_0b_0 + a_1b_1 + a_2b_2$$

$$f_1(x) = g_1(x) = 1 - x^2$$

$$\begin{aligned} f_2(x) &= g_2(x) - \frac{\langle g_2,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) \\ &= x + 2x^2 - \frac{(-2)}{2}\left(1 - x^2\right) \\ f_2(x) &= 1 + x + x^2 \end{aligned}$$

$$\begin{aligned} f_3(x) &= g_3(x) - \frac{\langle g_3,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_3,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) \\ &= -2 + x + 3x^2 - \frac{(-5)}{2}\left(1 - x^2\right) - \frac{2}{3}\left(1 + x + x^2\right) \\ f_3(x) &= -\frac{1}{6} + \frac{1}{3}x - \frac{1}{6}x^2 \end{aligned}$$

For a simpler answer we multiply $f_3(x)$ by 6 (this does not change the orthogonality) so that:

$$f_3(x) = -1 + 2x - x^2$$

Hence, the orthogonal set is:

$$\{f_1(x),\ f_2(x),\ f_3(x)\} = \left\{1 - x^2,\ 1 + x + x^2,\ -1 + 2x - x^2\right\}$$

(b) With $g_1(x) = 1 - x^2,\ g_2(x) = x + 2x^2,\ g_3(x) = -2 + x + 3x^2$ and with inner product $\left\langle a_0 + a_1x + a_2x^2,\ b_0 + b_1x + b_2x^2 \right\rangle = a_0b_0 + 2a_1b_1 + 3a_2b_2$, Theorem 5.8, steps 1, 2 and 3 (using $f_1,\ f_2,\ f_3$ as the orthogonal polynomials) are:

$$f_1(x) = g_1(x) = 1 - x^2$$

$$\begin{aligned} f_2(x) &= g_2(x) - \frac{\langle g_2,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) \\ &= x + 2x^2 - \frac{(-6)}{4}\left(1 - x^2\right) \\ f_2(x) &= \frac{3}{2} + x + \frac{1}{2}x^2 \end{aligned}$$

For simpler calculations we multiply $f_2(x)$ by 2 (this does not change the orthogonality) so that:

$$f_2(x) = 3 + 2x + x^2$$

$$\begin{aligned} f_3(x) &= g_3(x) - \frac{\langle g_3,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_3,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) \\ &= -2 + x + 3x^2 - \frac{(-11)}{4}\left(1 - x^2\right) - \frac{7}{20}\left(3 + 2x + x^2\right) \\ f_3(x) &= -\frac{3}{10} + \frac{3}{10}x - \frac{1}{10}x^2 \end{aligned}$$

For a simpler answer we multiply $f_2(x)$ by 2 (this does not change the orthogonality) so that:

$$f_3(x) = -3 + 3x - x^2$$

Hence, the orthogonal set is:

$$\{f_1(x),\ f_2(x),\ f_3(x)\} = \left\{1 - x^2,\ 3 + 2x + x^2,\ -3 + 3x - x^2\right\}$$

Note: Check for yourself that these three polynomials are mutually orthogonal but only for the inner product $\left\langle a_0 + a_1x + a_2x^2,\ b_0 + b_1x + b_2x^2\right\rangle = a_0b_0 + 2a_1b_1 + 3a_2b_2$.

(c) With $g_1(x) = 1 - x^2,\ g_2(x) = x + 2x^2,\ g_3(x) = -2 + x + 3x^2$ and with inner product $\langle p(x),\ q(x)\rangle = \int_{-1}^{1} p\ (x)\, q(x)\ dx$, Theorem 5.8, steps 1, 2 and 3, using $f_1,\ f_2,\ f_3$ as the orthogonal polynomials, and omitting details of the integrations, are:

$$f_1(x) = g_1(x) = 1 - x^2$$

$$\begin{aligned} f_2(x) &= g_2(x) - \frac{\langle g_2,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) \\ &= x + 2x^2 - \frac{\int_{-1}^{1}\left(x + 2x^2\right)\left(1 - x^2\right)\ dx}{\int_{-1}^{1}\left(1 - x^2\right)^2\ dx}\left(1 - x^2\right) \\ &= x + 2x^2 - \frac{\left(\frac{8}{15}\right)}{\left(\frac{16}{15}\right)}\left(1 - x^2\right) \\ f_2(x) &= -\frac{1}{2} + x + \frac{5}{2}x^2 \end{aligned}$$

For simpler calculations we multiply $f_2(x)$ by 2 (this does not change the orthogonality) so that:

$$f_2(x) = -1 + 2x + 5x^2$$

$$\begin{aligned} f_3(x) &= g_3(x) - \frac{\langle g_3,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_3,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) \\ &= -2 + x + 3x^2 - \frac{\int_{-1}^{1}\left(-2 + x + 3x^2\right)\left(1 - x^2\right)\ dx}{\int_{-1}^{1}\left(1 - x^2\right)^2\ dx}\left(1 - x^2\right) \\ &\qquad - \frac{\int_{-1}^{1}\left(-2 + x + 3x^2\right)\left(-1 + 2x + 5x^2\right)\ dx}{\int_{-1}^{1}\left(-1 + 2x + 5x^2\right)^2\ dx}\left(-1 + 2x + 5x^2\right) \\ &= -2 + x + 3x^2 - \frac{\left(-\frac{28}{15}\right)}{\left(\frac{16}{15}\right)}\left(1 - x^2\right) - \frac{\left(\frac{8}{3}\right)}{8}\left(-1 + 2x + 5x^2\right) \\ f_3(x) &= \frac{1}{12} + \frac{1}{3}x - \frac{5}{12}x^2 \end{aligned}$$

For a simpler answer we multiply $f_3(x)$ by 12 (this does not change the orthogonality) so that:

$$f_3(x) = 1 + 4x - 5x^2$$

Hence, the orthogonal set is:

$$\{f_1(x),\ f_2(x),\ f_3(x)\} = \left\{1 - x^2,\ -1 + 2x + 5x^2,\ 1 + 4x - 5x^2\right\}$$

Note: Check for yourself that these three polynomials are mutually orthogonal but only for the inner product used in this part.

(d) The orthonormal basis in part (a) is (recall that $\|f\| = \sqrt{\langle f,\ f\rangle}$):

$$\begin{aligned}
&\left\{\frac{1}{\|f_1\|} f_1(x), \frac{1}{\|f_2\|}\ f_2(x),\ \frac{1}{\|f_3\|} f_3(x)\right\} \\
&= \left\{\frac{1}{\sqrt{2}}\left(1-x^2\right),\ \frac{1}{\sqrt{3}}\left(1+x+x^2\right),\ \frac{1}{\sqrt{6}}\left(-1+2x-x^2\right)\right\} \\
&= \left\{\left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}x^2\right),\ \left(\frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}}x + \frac{1}{\sqrt{3}}x^2\right),\ \left(-\frac{1}{\sqrt{6}} + \frac{2}{\sqrt{6}}x - \frac{1}{\sqrt{6}}x^2\right)\right\}
\end{aligned}$$

(e) The orthonormal basis in part (b) is:

$$\begin{aligned}
&\left\{\frac{1}{\|f_1\|} f_1(x), \frac{1}{\|f_2\|}\ f_2(x),\ \frac{1}{\|f_3\|} f_3(x)\right\} \\
&= \left\{\frac{1}{\sqrt{4}}\left(1-x^2\right),\ \frac{1}{\sqrt{20}}\left(3+2x+x^2\right),\ \frac{1}{\sqrt{30}}\left(-3+3x-x^2\right)\right\} \\
&= \left\{\left(\frac{1}{2} - \frac{1}{2}x^2\right),\ \left(\frac{3}{\sqrt{20}} + \frac{2}{\sqrt{20}}x + \frac{1}{\sqrt{20}}x^2\right),\ \left(-\frac{3}{\sqrt{30}} + \frac{3}{\sqrt{30}}x - \frac{1}{\sqrt{30}}x^2\right)\right\}
\end{aligned}$$

6. To illustrate Theorem 5.12 fully, the part (a) answer uses the orthogonal polynomials from the previous question - part (a) , whereas part (b) of this question uses the orthonormal polynomials from the previous question - part (e).

(a) Let $g(x) = 2 + 3x - 4x^2$ and take $f_1,\ f_2,\ f_3$ to be the orthogonal polynomials set from part (a) of the previous question:

$$\{f_1(x),\ f_2(x),\ f_3(x)\} = \left\{1-x^2,\ 1+x+x^2,\ -1+2x-x^2\right\}$$

From Theorem 5.12, with the inner product used in the previous question - part (a):

$$\begin{aligned}
g(x) &= \frac{\langle g,\ f_1\rangle}{\langle f_1,\ f_1\rangle}\ f_1(x) + \frac{\langle g,\ f_2\rangle}{\langle f_2,\ f_2\rangle}\ f_2(x) + \frac{\langle g,\ f_3\rangle}{\langle f_3,\ f_3\rangle}\ f_3(x) \\
g(x) &= \frac{6}{2}\left(1-x^2\right) + \frac{1}{3}\left(1+x+x^2\right) + \frac{8}{6}\left(-1+2x-x^2\right)
\end{aligned}$$

That is,

$$g(x) = 2 + 3x - 4x^2 = 3\left(1-x^2\right) + \frac{1}{3}\left(1+x+x^2\right) + \frac{4}{3}\left(-1+2x-x^2\right)$$

Note: Check for yourself that $g(x)$ does equal the given linear combination of $f_1,\ f_2,\ f_3$

(b) Let $g(x) = 2 + 3x - 4x^2$ and take $f_1,\ f_2,\ f_3$ to be the orthonormal polynomials set from part (e) of the previous question:

$$\begin{aligned}
&\{f_1(x),\ f_2(x),\ f_3(x)\} \\
&= \left\{\left(\frac{1}{2} - \frac{1}{2}x^2\right),\ \frac{6}{\sqrt{20}}\left(\frac{3}{\sqrt{20}} + \frac{2}{\sqrt{20}}x + \frac{1}{\sqrt{20}}x^2\right),\ \left(-\frac{3}{\sqrt{30}} + \frac{3}{\sqrt{30}}x - \frac{1}{\sqrt{30}}x^2\right)\right\}
\end{aligned}$$

Using the Theorem 5.12 formula for orthonormal sets and with the inner product used in the previous question - part (b):

$$\begin{aligned}
g(x) &= \langle g,\ f_1\rangle\ f_1(x) + \langle g,\ f_2\rangle\ f_2(x) + \langle g,\ f_3\rangle\ f_3(x) \\
g(x) &= 7\left(\frac{1}{2} - \frac{1}{2}x^2\right) + \frac{6}{\sqrt{20}}\left(\frac{3}{\sqrt{20}} + \frac{2}{\sqrt{20}}x + \frac{1}{\sqrt{20}}x^2\right) + \frac{24}{\sqrt{30}}\left(-\frac{3}{\sqrt{30}} + \frac{3}{\sqrt{30}}x - \frac{1}{\sqrt{30}}x^2\right)
\end{aligned}$$

Note: Check for yourself that $g(x)$ does equal the given linear combination of $f_1,\ f_2,\ f_3$.

(c) Recall that the starting polynomials in the previous question are $g_1(x) = 1 - x^2,\ g_2(x) = x + 2x^2,\ g_3(x) = -2 + x + 3x^2$, and the orthonormal set of polynomials found in part (d) of the previous question is:

$$\begin{aligned}&\{f_1(x),\ f_2(x),\ f_3(x)\}\\ &= \left\{\left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}x^2\right),\ \left(\frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}}x + \frac{1}{\sqrt{3}}x^2\right),\ \left(-\frac{1}{\sqrt{6}} + \frac{2}{\sqrt{6}}x - \frac{1}{\sqrt{6}}x^2\right)\right\}\end{aligned}$$

The $QR-$ factorization formula uses the inner product from part (a) of the previous question, and it is:

$$\begin{bmatrix} g_1(x) & g_2(x) & g_3(x) \end{bmatrix} = \begin{bmatrix} f_1(x) & f_2(x) & f_3(x) \end{bmatrix} \begin{bmatrix} \langle f_1, g_1\rangle & \langle f_1,\ g_2\rangle & \langle f_1,\ g_3\rangle \\ 0 & \langle f_2,\ g_2\rangle & \langle f_2,\ g_3\rangle \\ 0 & 0 & \langle f_3,\ g_3\rangle \end{bmatrix}$$

Hence, the factorization is:

$$\begin{aligned}&\begin{bmatrix} 1 - x^2 & x + 2x^2 & -2 + x + 3x^2 \end{bmatrix}\\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}x^2 & \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}}x + \frac{1}{\sqrt{3}}x^2 & -\frac{1}{\sqrt{6}} + \frac{2}{\sqrt{6}}x - \frac{1}{\sqrt{6}}x^2 \end{bmatrix} \begin{bmatrix} \sqrt{2} & -\sqrt{2} & -\frac{5}{\sqrt{2}} \\ 0 & \sqrt{3} & \frac{2}{\sqrt{3}} \\ 0 & 0 & \frac{1}{\sqrt{6}} \end{bmatrix}\end{aligned}$$

Note: It is a somewhat challenging exercise to show that the product of the two right-hand side matrices does give the left-hand side matrix.

7. With the original polynomials $g_1(x) = 1 - x^2,\ g_2(x) = x + 2x^2,\ g_3(x) = -2 + x + 4x^2$, from Theorem 5.8, steps 1, 2 and 3 (using $f_1,\ f_2,\ f_3$ as the orthogonal polynomials) with inner product:

$$\left\langle a_0 + a_1x + a_2x^2,\ b_0 + b_1x + b_2x^2\right\rangle = a_0b_0 + a_1b_1 + a_2b_2$$

$$f_1(x) = g_1(x) = 1 - x^2$$

$$\begin{aligned} f_2(x) &= g_2(x) - \frac{\langle g_2,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) \\ &= x + 2x^2 - \frac{(-2)}{2}\left(1 - x^2\right) \\ f_2(x) &= 1 + x + x^2 \end{aligned}$$

$$\begin{aligned} f_3(x) &= g_3(x) - \frac{\langle g_3,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_3,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) \\ &= -2 + x + 4x^2 - \frac{(-6)}{2}\left(1 - x^2\right) - \frac{3}{3}\left(1 + x + x^2\right) \\ f_3(x) &= 0 \end{aligned}$$

That is, the polynomial $f_3(x)$ is the zero polynomial, and this indicates that it is not possible to find a non-zero polynomial orthogonal to f_1 and f_2. This only happens when the original set of polynomials is linearly dependent. In this case the linear dependence is:

$$g_3(x) = g_2(x) - 2g_1(x)$$

8. For the matrices $A_1 = \begin{bmatrix} 1 & 1 \\ -2 & 0 \end{bmatrix}, A_2 = \begin{bmatrix} 2 & 0 \\ 4 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

(a) the Gram-Schmidt process of Theorem 5.8 gives the orthogonal matrices B_1, B_2, B_3 as follows:

$$B_1 = A_1 = \begin{bmatrix} 1 & 1 \\ -2 & 0 \end{bmatrix}$$

$$\begin{aligned} B_2 &= A_2 - \frac{\langle A_2,\ B_1\rangle}{\langle B_1,\ B_1\rangle} B_1 \\ &= \begin{bmatrix} 2 & 0 \\ 4 & 1 \end{bmatrix} - \frac{(-6)}{6} \begin{bmatrix} 1 & 1 \\ -2 & 0 \end{bmatrix} \\ B_2 &= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} B_3 &= A_3 - \frac{\langle A_3,\ B_1\rangle}{\langle B_1,\ B_1\rangle} B_1 - \frac{\langle A_3,\ B_2\rangle}{\langle B_2,\ B_2\rangle} B_2 \\ &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - \frac{2}{6} \begin{bmatrix} 1 & 1 \\ -2 & 0 \end{bmatrix} - \frac{5}{15} \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \\ B_3 &= \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} \\ 0 & \frac{2}{3} \end{bmatrix} \end{aligned}$$

For a simpler answer we multiply B_3 by 3 (this does not affect orthogonality), and so:

$$B_3 = \begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix}$$

Hence, the orthogonal set is:

$$\{B_1, B_2, B_3\} = \left\{ \begin{bmatrix} 1 & 1 \\ -2 & 0 \end{bmatrix}, \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix} \right\}$$

(b) Normalizing the orthogonal set:

$$\begin{aligned} \left\{ \frac{1}{\|B_1\|} B_1, \frac{1}{\|B_2\|} B_2, \frac{1}{\|B_3\|} B_3 \right\} &= \left\{ \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 1 \\ -2 & 0 \end{bmatrix}, \frac{1}{\sqrt{15}} \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} & 0 \end{bmatrix}, \begin{bmatrix} \frac{3}{\sqrt{15}} & \frac{1}{\sqrt{15}} \\ \frac{2}{\sqrt{15}} & \frac{1}{\sqrt{15}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \right\} \end{aligned}$$

(c) The matrix is:

$$C = \begin{bmatrix} 3 & 1 \\ 0 & 0 \end{bmatrix}$$

Using Theorem 5.12 and using the orthonormal basis found in part (b), the required linear combination is (if it exists):

$$\begin{aligned} C &= \langle C,\ B_1\rangle\ B_1 + \ \langle C,\ B_2\rangle\ B_2 + \ \langle C,\ B_3\rangle\ B_3 \\ &= \frac{4}{\sqrt{6}} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} & 0 \end{bmatrix} + \frac{10}{\sqrt{15}} \begin{bmatrix} \frac{3}{\sqrt{15}} & \frac{1}{\sqrt{15}} \\ \frac{2}{\sqrt{15}} & \frac{1}{\sqrt{15}} \end{bmatrix} + \frac{-2}{\sqrt{6}} \begin{bmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \end{aligned}$$

Checking this result by multiplying out the matrices and computing the sums shows that it is correct:

$$\begin{aligned} &= \begin{bmatrix} \frac{2}{3} & \frac{2}{3} \\ -\frac{4}{3} & 0 \end{bmatrix} + \begin{bmatrix} 2 & \frac{2}{3} \\ \frac{4}{3} & \frac{2}{3} \end{bmatrix} + \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ 0 & -\frac{2}{3} \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 \\ 0 & 0 \end{bmatrix} = C \end{aligned}$$

(d) The matrix is:

$$C = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$$

Using Theorem 5.12 and using the orthonormal basis found in part (b), the required linear combination is (if it exists):

$$\begin{aligned} C &= \langle C,\ B_1\rangle\ B_1 + \ \langle C,\ B_2\rangle\ B_2 + \ \langle C,\ B_3\rangle\ B_3 \\ &= \frac{3}{\sqrt{6}} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} & 0 \end{bmatrix} + \frac{3}{\sqrt{15}} \begin{bmatrix} \frac{3}{\sqrt{15}} & \frac{1}{\sqrt{15}} \\ \frac{2}{\sqrt{15}} & \frac{1}{\sqrt{15}} \end{bmatrix} + \frac{1}{\sqrt{6}} \begin{bmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \end{aligned}$$

Checking this result by multiplying out the matrices and computing the sums shows that it is correct:

$$\begin{aligned} &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} \frac{3}{5} & \frac{1}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix} + \begin{bmatrix} -\frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{3} \end{bmatrix} \\ &= \begin{bmatrix} \frac{14}{15} & \frac{13}{15} \\ -\frac{3}{5} & \frac{8}{15} \end{bmatrix} \neq C \end{aligned}$$

In this case the linear combination of the orthogonal matrices is not equal to C. This means C is not in the span of B_1, B_2, B_3. By Theorem 5.13, the linear combination is actually the projection of C onto the space S.

9. For the polynomials:

$$g_1(x) = 1,\ g_2(x) = x,\ g_3(x) = x^2,\ g_4(x) = x^3,\ g_5(x) = x^4$$

it was shown in Example 5.6.4 that the first four orthogonal polynomials produced by the Gram-Schmidt process (Theorem 5.8) are:

$$\{f_1(x),\ f_2(x),\ \ f_3(x),\ \ f_4(x)\} = \left\{1,\ x,\ x^2 - \frac{1}{3},\ x^3 - \frac{3}{5}x\right\}$$

(a) The fifth orthogonal polynomial is given by:

$$\begin{aligned} f_5(x) &= g_5(x) - \frac{\langle g_5,\ f_1\rangle}{\langle f_1,\ f_1\rangle} f_1(x) - \frac{\langle g_5,\ f_2\rangle}{\langle f_2,\ f_2\rangle} f_2(x) - \frac{\langle g_5,\ f_3\rangle}{\langle f_3,\ f_3\rangle} f_3(x) - \frac{\langle g_5,\ f_4\rangle}{\langle f_4,\ f_4\rangle} f_4(x) \\ &= x^4 - \frac{\int_{-1}^{1} x^4\ dx}{\int_{-1}^{1} 1^2\ dx} 1 - \frac{\int_{-1}^{1} x^5\ dx}{\int_{-1}^{1} x^2\ dx} x - \frac{\int_{-1}^{1} x^4 \left(x^2 - \frac{1}{3}\right)\ dx}{\int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)^2\ dx} \left(x^2 - \frac{1}{3}\right) \\ &\quad - \frac{\int_{-1}^{1} x^4 \left(x^3 - \frac{3}{5}x\right)\ dx}{\int_{-1}^{1} \left(x^3 - \frac{3}{5}x\right)^2\ dx} \left(x^3 - \frac{3}{5}x\right) \\ &= x^4 - \frac{\left(\frac{2}{5}\right)}{2} - \frac{0}{\left(\frac{2}{3}\right)} x - \frac{\left(\frac{16}{105}\right)}{\left(\frac{8}{45}\right)} \left(x^2 - \frac{1}{3}\right) - \frac{0}{\left(\frac{8}{175}\right)} \left(x^3 - \frac{3}{5}x\right) \\ &= x^4 - \frac{1}{5} - \frac{6}{7}\left(x^2 - \frac{1}{3}\right) \\ f_3(x) &= x^4 - \frac{6}{7}x^2 + \frac{3}{35} \end{aligned}$$

Hence, the orthogonal basis of P_4 is:

$$\{f_1(x),\ f_2(x),\ \ f_3(x),\ \ f_4(x),\ \ f_5(x)\} = \left\{1,\ x,\ x^2 - \frac{1}{3},\ x^3 - \frac{3}{5}x,\ x^4 - \frac{6}{7}x^2 + \frac{3}{35}\right\}$$

Note: The Legendre Polynomials are multiples of these polynomials, and the first five Legendre Polynomials are:

$$p_0(x) = 1,\ p_1(x) = x,\ p_2(x) = \frac{1}{2}\left(x^2 - 1\right),\ p_3(x) = \frac{1}{2}\left(5x^3 - 3x\right),\ p_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right)$$

These multiples are chosen so that $p_k(1) = 1$, so they are an orthogonal set but are not normalized with respect to the inner product and so do not form an orthonormal set. For more details see: http://en.wikipedia.org/wiki/Legendre_polynomials

(b) The corresponding orthonormal basis is:

$$\frac{1}{\|f_1\|} f_1(x),\ \frac{1}{\|f_2\|} f_2(x),\ \frac{1}{\|f_3\|} f_3(x),\ \frac{1}{\|f_4\|} f_4(x),\ \frac{1}{\|f_5\|} f_5(x)$$

$$= \frac{1}{\sqrt{\langle f_1,\ f_1\rangle}} f_1(x),\ \frac{1}{\sqrt{\langle f_2,\ f_2\rangle}} f_2(x),\ \frac{1}{\sqrt{\langle f_3,\ f_3\rangle}} f_3(x),\ \frac{1}{\sqrt{\langle f_4,\ f_4\rangle}} f_4(x),\ \frac{1}{\sqrt{\langle f_5,\ f_5\rangle}} f_5(x)$$

$$= \frac{1}{\sqrt{\int_{-1}^{1} 1\, dx}} \times 1,\ \frac{1}{\sqrt{\int_{-1}^{1} x^2\, dx}}\, x,\ \frac{1}{\sqrt{\int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)^2 dx}} \left(x^2 - \frac{1}{3}\right),$$

$$\frac{1}{\sqrt{\int_{-1}^{1} \left(x^3 - \frac{3}{5}x\right)^2 dx}} \left(x^3 - \frac{3}{5}x\right), \frac{1}{\sqrt{\int_{-1}^{1} \left(x^4 - \frac{6}{7}x^2 + \frac{3}{35}\right)^2 dx}} \left(x^4 - \frac{6}{7}x^2 + \frac{3}{35}\right)$$

$$= \frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{\frac{2}{3}}}x,\ \frac{1}{\left(\frac{2\sqrt{2}}{3\sqrt{5}}\right)} \left(x^2 - \frac{1}{3}\right),\ \frac{1}{\left(\frac{2\sqrt{2}}{5\sqrt{7}}\right)} \left(x^3 - \frac{3}{5}x\right),\ \frac{1}{\left(\frac{8\sqrt{2}}{105}\right)} \left(x^4 - \frac{6}{7}x^2 + \frac{3}{35}\right)$$

$$= \frac{1}{\sqrt{2}},\ \sqrt{\frac{3}{2}}x,\ \frac{3\sqrt{5}}{2\sqrt{2}} \left(x^2 - \frac{1}{3}\right),\ \frac{5\sqrt{7}}{2\sqrt{2}} \left(x^3 - \frac{3}{5}x\right),\ \frac{105}{\left(8\sqrt{2}\right)} \left(x^4 - \frac{6}{7}x^2 + \frac{3}{35}\right)$$

10. For the vectors:

$$\mathbf{u}_1 = (1,\ 0,\ 0,\ 1),\ \mathbf{u}_2 = (1,\ 2,\ 0,\ 1)$$

(a) Using the method of Theorem 5.9 to find an orthonormal basis of W :

$$\mathbf{v}_1 = \mathbf{u}_1 \text{ and } \mathbf{w}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \left(\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right)$$

$$\mathbf{v}_2 = \mathbf{u}_2 - (\mathbf{u}_2 \cdot \mathbf{w}_1)\,\mathbf{w}_1 = (1,\ 2,\ 0,\ 1) - \sqrt{2}\left(\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right) = (0,\ 2,\ 0,\ 0)$$

$$\mathbf{w}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = (0,\ 1,\ 0,\ 0)$$

Hence, an orthonormal basis of W is:

$$\{\mathbf{w}_1,\ \mathbf{w}_2\} = \left\{\left(\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right),\ (0,\ 1,\ 0,\ 0)\right\}$$

In order to find a basis of $W^{\perp}$ we add vectors, and use the Gram-Schmidt process to find an orthonormal basis of the whole space R^4 - the extra vectors are an orthonormal basis of $W^{\perp}$. That is, we first add a third vector $\mathbf{w}_3$ to the set $\mathbf{w}_1,\ \mathbf{w}_2$. The choice of $\mathbf{w}_3$ is arbitrary (as long as it is not in W) but by noticing that the third components of $\mathbf{w}_1,\ \mathbf{w}_2$ are both zero we can choose $\mathbf{w}_3 = (0,\ 0,\ 1,\ 0)$ so that it is already normalized and orthogonal to the first two vectors.
We add a fourth vector $\mathbf{u}_4$ to the set $\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3$ and arbitrarily choose it to be

$\mathbf{u}_4 = (1,\ 0,\ 0,\ 3)$. We use one step of the method from Theorem 5.9 to make the whole set orthonormal:

$$\mathbf{v}_4 = \mathbf{u}_4 - (\mathbf{u}_4 \cdot \mathbf{w}_1)\,\mathbf{w}_1 - (\mathbf{u}_4 \cdot \mathbf{w}_2)\,\mathbf{w}_2 - (\mathbf{u}_4 \cdot \mathbf{w}_3)\,\mathbf{w}_3$$

$$= (1,\ 0,\ 0,\ 3) - 2\sqrt{2}\left(\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right) -\ 0 \times (0,\ 1,\ 0,\ 0) - 0 \times (0,\ 0,\ 1,\ 0) = (-1,\ 0,\ 0,\ 1)$$

$$\mathbf{w}_4 = \frac{1}{\|\mathbf{v}_4\|}\mathbf{v}_4 = \left(-\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right)$$

Hence, we now have an orthonormal basis of R^4 :

$$\{\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3,\ \mathbf{w}_4\} = \left\{\left(\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right),\ (0,\ 1,\ 0,\ 0),\ (0,\ 0,\ 1,\ 0),\ \left(-\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right)\right\}$$

and by Theorem 5.13, the extra two vectors are a basis for $W^{\perp}$, namely:

$$\{\mathbf{w}_3,\ \mathbf{w}_4\} = \left\{(0,\ 0,\ 1,\ 0),\ \left(-\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right)\right\}$$

(b) By Theorem 5.12 we can express $\mathbf{u} = (2, 3, 0, 4)$ as a linear combination of $\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3,\ \mathbf{w}_4$ by the formula:

$$\mathbf{u} = (\mathbf{u} \cdot \mathbf{w}_1)\ \mathbf{w}_1 + (\mathbf{u} \cdot \mathbf{w}_2)\ \mathbf{w}_2 + (\mathbf{u} \cdot \mathbf{w}_3)\ \mathbf{w}_3 + (\mathbf{u} \cdot \mathbf{w}_4)\ \mathbf{w}_4$$
$$= 3\sqrt{2}\mathbf{w}_1 + 3\mathbf{w}_2 + 0 \times \mathbf{w}_3 + 2\sqrt{2}\mathbf{w}_4$$

Hence, by Theorem 5.13 $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ with $\mathbf{u}_1 \in W,\ \mathbf{u}_2 \in W^{\perp}$ and:

$$\mathbf{u}_1 = 3\sqrt{2}\mathbf{w}_1 + 3\mathbf{w}_2 = 3\sqrt{2}\left(\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right) + 3\,(0,\ 1,\ 0,\ 0)$$
$$= 3\sqrt{2}\left(\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right) + 3\,(0,\ 1,\ 0,\ 0) = (3,\ 3,\ 0,\ 3)$$
$$\mathbf{u}_2 = 0 \times \mathbf{w}_3 + 2\sqrt{2}\mathbf{w}_4 = 2\sqrt{2}\left(-\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right) = (-2,\ 0,\ 0,\ 2)$$

(c) The $QR-$ factorization formula is, with the vectors as columns, $\mathbf{u}_1 = (1,\ 0,\ 0,\ 1)$, $\mathbf{u}_2 = (1,\ 2,\ 0,\ 1)$, $\mathbf{w}_1 = \left(\frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}\right)$, $\mathbf{w}_2 = (0,\ 1,\ 0,\ 0)$:

$$\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \cdot \mathbf{u}_1 & \mathbf{w}_1 \cdot \mathbf{u}_2 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 2 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & 2 \end{bmatrix}$$

11. The Gram-Schmidt process is given in matrix form, as shown in Example 5.6.11 11. The columns of the original matrix are referred to as $\mathbf{u}_1, \mathbf{u}_2, \cdots$ and the columns of Q are referred to as $\mathbf{w}_1, \mathbf{w}_2, \cdots$.

(a) **Step 1:** No special work needed since the first column is already normalized:

$$\overset{A}{\begin{bmatrix} 1 & 3 \\ 0 & 4 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} 1 & * \\ 0 & * \end{bmatrix}} \overset{R}{\begin{bmatrix} 1 & \mathbf{w}_1 \cdot \mathbf{u}_2 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 \end{bmatrix}}$$

Step 2: The second column of Q is equal to the vector $\mathbf{v}_2 = \mathbf{u}_2 - (\mathbf{u}_2 \cdot \mathbf{w}_1)\,\mathbf{w}_1 = (3,\ 4) - 3\,(1,0) = (0,\ 4)$, which is then normalized to $\mathbf{w}_2 = (0,1)$. The second column of R has non-zero entries identical to $\mathbf{u}_2$:

$$\begin{bmatrix} 1 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 4 \end{bmatrix}$$

Note: This is rather a trivial case with $R = A$ and $Q = I$. That is because A is already upper triangular.

(b) **Step 1:** The first column of Q is the normalized $\mathbf{u}_1$, and the first row of R is computed:

$$\overset{B}{\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} \frac{1}{\sqrt{2}} & * & * \\ 0 & * & * \\ \frac{1}{\sqrt{2}} & * & * \end{bmatrix}} \overset{R}{\begin{bmatrix} \mathbf{w}_1 \cdot \mathbf{u}_1 & \mathbf{w}_1 \cdot \mathbf{u}_2 & \mathbf{w}_1 \cdot \mathbf{u}_3 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \mathbf{w}_2 \cdot \mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & * & * \\ 0 & * & * \\ \frac{1}{\sqrt{2}} & * & * \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \mathbf{w}_2 \cdot \mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}$$

Step 2: The second column of Q is equal to the vector $\mathbf{v}_2 = \mathbf{u}_2 - (\mathbf{w}_1 \cdot \mathbf{u}_2)\,\mathbf{w}_1 = (1,\ 1,\ -1) - 0 \times \mathbf{w}_1 = (1,\ 1,\ -1)$, which is then normalized to $\mathbf{w}_2 = \left(\frac{1}{\sqrt{3}},\ \frac{1}{\sqrt{3}},\ -\frac{1}{\sqrt{3}}\right)$. The second row of R is computed.

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & * \\ 0 & \frac{1}{\sqrt{3}} & * \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & * \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \mathbf{w}_2 \cdot \mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & * \\ 0 & \frac{1}{\sqrt{3}} & * \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & * \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{3} & \mathbf{0} \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}$$

Step 3: The third column of Q is equal to the vector $\mathbf{v}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3,\ \mathbf{w}_1 \rangle\,\mathbf{w}_1 - \langle \mathbf{u}_3,\ \mathbf{w}_2 \rangle\,\mathbf{w}_2$ $= (0,\ 1,\ 1) - \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) - 0 \times \mathbf{w}_2 = \left(-\frac{1}{2}, 1, \frac{1}{2}\right)$ normalized to $\mathbf{w}_3 = \left(-\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)$.

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{3} & \mathbf{0} \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}$$

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

The last row of R is computed, giving the $QR-$ decomposition of B :

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

(c) **Step 1:** The first column of Q is the normalized $\mathbf{u}_1$, and the first row of R is computed:

$$\overset{C}{\begin{bmatrix} 3 & 5 \\ 4 & 0 \\ 0 & -2 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} \frac{3}{5} & * \\ \frac{4}{5} & * \\ 0 & * \end{bmatrix}} \overset{R}{\begin{bmatrix} \mathbf{w}_1 \cdot \mathbf{u}_1 & \mathbf{w}_1 \cdot \mathbf{u}_2 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 \end{bmatrix}} = \begin{bmatrix} \frac{3}{5} & * \\ \frac{4}{5} & * \\ 0 & * \end{bmatrix} \begin{bmatrix} 5 & 3 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 \end{bmatrix}$$

Step 2: The second column of Q is equal to the vector $\mathbf{v}_2 = \mathbf{u}_2 - (\mathbf{w}_1 \cdot \mathbf{u}_2)\,\mathbf{w}_1 = (5,\ 0,\ -2) - 3\left(\frac{3}{5},\ \frac{4}{5},\ 0\right) = \left(\frac{16}{5},\ -\frac{12}{5},\ -2\right)$ which is then normalized to $\mathbf{w}_2 = \left(\frac{8}{5\sqrt{5}},\ -\frac{6}{5\sqrt{5}},\ -\frac{1}{\sqrt{5}}\right)$The second row of R is computed giving the $QR-$factorization:

$$\begin{bmatrix} 3 & 5 \\ 4 & 0 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{8}{5\sqrt{5}} \\ \frac{4}{5} & -\frac{6}{5\sqrt{5}} \\ 0 & -\frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 5 & 3 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 \end{bmatrix} = \overset{Q}{\begin{bmatrix} \frac{3}{5} & \frac{8}{5\sqrt{5}} \\ \frac{4}{5} & -\frac{6}{5\sqrt{5}} \\ 0 & -\frac{1}{\sqrt{5}} \end{bmatrix}} \overset{R}{\begin{bmatrix} 5 & 3 \\ 0 & 2\sqrt{5} \end{bmatrix}}$$

(d) **Step1**: By Theorem 5.9, the matrices have the following form, with the first column of Q being the normalized $\mathbf{u}_1$. The first row of R is then computed:

$$\overset{D}{\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 2 & -1 & 1 \end{bmatrix}} = \overset{Q}{\begin{bmatrix} \frac{1}{\sqrt{5}} & * & * \\ 0 & * & * \\ \frac{2}{\sqrt{5}} & * & * \end{bmatrix}} \overset{R}{\begin{bmatrix} \mathbf{w}_1 \cdot \mathbf{u}_1 & \mathbf{w}_1 \cdot \mathbf{u}_2 & \mathbf{w}_1 \cdot \mathbf{u}_3 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \mathbf{w}_2 \cdot \mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{5}} & * & * \\ 0 & * & * \\ \frac{2}{\sqrt{5}} & * & * \end{bmatrix} \begin{bmatrix} \sqrt{5} & -\frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \mathbf{w}_2 \cdot \mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}$$

Step 2: The second column of Q is equal to the vector $\mathbf{v}_2 = \mathbf{u}_2 - (\mathbf{u}_2 \cdot \mathbf{w}_1)\,\mathbf{w}_1 = (0,\ 1,\ -1) - \left(-\frac{2}{\sqrt{5}}\right)\left(\frac{1}{\sqrt{5}}, 0, \frac{2}{\sqrt{5}}\right) = \left(\frac{2}{5},\ 1,\ -\frac{1}{5}\right)$, which is then normalized to $\mathbf{w}_2 = \sqrt{\frac{5}{6}}\left(\frac{2}{5},\ 1,\ -\frac{1}{5}\right) = \left(\frac{\sqrt{6}}{3\sqrt{5}},\ \frac{\sqrt{5}}{\sqrt{6}},\ -\frac{1}{\sqrt{5}\sqrt{6}}\right)$. The second row of R is then computed:

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 2 & -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{\sqrt{6}}{3\sqrt{5}} & * \\ 0 & \frac{\sqrt{5}}{\sqrt{6}} & * \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}\sqrt{6}} & * \end{bmatrix} \begin{bmatrix} \sqrt{5} & -\frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \mathbf{w}_2 \cdot \mathbf{u}_3 \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{\sqrt{6}}{3\sqrt{5}} & * \\ 0 & \frac{\sqrt{5}}{\sqrt{6}} & * \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}\sqrt{6}} & * \end{bmatrix} \begin{bmatrix} \sqrt{5} & -\frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \\ 0 & \sqrt{\frac{6}{5}} & \frac{3\sqrt{6}}{\sqrt{5}} \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{bmatrix}$$

Step 3: The third column of Q is equal to the vector:

$$\begin{aligned} \mathbf{v}_3 &= \mathbf{u}_3 - (\mathbf{w}_1 \cdot \mathbf{u}_3)\,\mathbf{w}_1 - (\mathbf{w}_2 \cdot \mathbf{u}_3)\,\mathbf{w}_2 \\ &= (2,\ 3,\ 1) - \frac{4}{\sqrt{5}}\left(\frac{1}{\sqrt{5}}, 0, \frac{2}{\sqrt{5}}\right) - \frac{3\sqrt{6}}{\sqrt{5}}\left(\frac{\sqrt{6}}{3\sqrt{5}}, \frac{\sqrt{5}}{\sqrt{6}}, -\frac{1}{\sqrt{5}\sqrt{6}}\right) \\ \mathbf{v}_3 &= (0,\ 0,\ 0) \end{aligned}$$

The zero vector found for $\mathbf{v}_3$ indicates that it is not possible to find a third non-zero vector orthogonal to the first two columns of Q. This happens because the original matrix D has a linear dependence in its columns (the third column equals 3 times the second plus 2 times the first), and so the columns of D only span a 2-dimensional subspace of R^3.
Note: Some authors produce a $QR-$ factorization anyway, by finding an extra non-zero column that makes Q orthogonal and then putting the bottom row of R to be all zeros, giving:

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 2 & -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{\sqrt{6}}{3\sqrt{5}} & \frac{\sqrt{2}}{\sqrt{3}} \\ 0 & \frac{\sqrt{5}}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \sqrt{5} & -\frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \\ 0 & \sqrt{\frac{6}{5}} & \frac{3\sqrt{6}}{\sqrt{5}} \\ 0 & 0 & 0 \end{bmatrix}$$

5.7 Least squares approximations

When S is a subspace of an inner product space V any vector $\mathbf{b}$ not in S is shown to have a unique **orthogonal projection** vector in the space S, denoted $\text{proj}_S\ \mathbf{b}$, such that $\mathbf{b} - \text{proj}_S\ \mathbf{b}$ is orthogonal to every vector in S. The orthogonal projection vector solves an optimization problem because it is the nearest vector in S to the vector $\mathbf{b}$, when distance is defined in terms of the inner product of V. In this optimization context, $\text{proj}_S\ \mathbf{b}$ is often referred to as the **least squares approximation** of the vector $\mathbf{b}$ when V is a Euclidean vector space.

We have previously encountered some orthogonal projection formulae. In Euclidean spaces in Unit 3 Section 7: Linear Transformations from R^n to R^m, Theorem 3.14 gives the orthogonal projection onto a line through the origin, and Theorem 3.19 gives the orthogonal projection onto a plane through the origin. We also encountered the general inner product space formula for the orthogonal projection onto a line/vector through the origin in Theorem 5.7 of Unit 5 Section 5, Basic Properties of Inner Product Spaces. Finally we saw a formula for the orthogonal projection onto a subspace of an inner product space with a known orthogonal basis, in Theorem 5.13 of Unit 5, Section 6, Orthogonal Bases, the Gram-Schmidt Process and $QR-$ factorization. In this section we develop some formulae for orthogonal projections that work for any subspace of an inner product space and specific specializations of these formulae for Euclidean spaces. The orthogonal projection, or least squares approximation, has a wide variety of applications in Engineering, Science, and Mathematics.

Note: Some of the proofs of theorems are omitted here but can be found in your textbook. Vectors are shown interchangeably as column matrices and as comma - separated row vectors. The word "projection" is often used here to mean "orthogonal projection."

5.7.1 Properties of orthogonal projections in inner product spaces

Definition.
If S is a subspace of an inner product space V, and $\mathbf{b} \notin S$ is any vector in V then the **orthogonal projection** (or simply **projection**) of $\mathbf{b}$ in S is a vector $\mathbf{y} \in S$ such that $\mathbf{b} - \mathbf{y}$ is orthogonal to every vector in S. If $\mathbf{b}$ is in S we formally define the projection of $\mathbf{b}$ in S to be $\mathbf{b}$ (the same vector). The projection is denoted as $\text{proj}_S\ \mathbf{b}$.

Note: Think of going from $\mathbf{b}$ along a line that is perpendicular to every vector in S until the line meets S at the vector $\mathbf{y} = \text{proj}_S\ \mathbf{b}$. You can also think of it as being analogous to the Euclidean space projection that is easy to visualize, at least for R^2 and R^3. The picture for R^2 is shown in Figure 5.5, and a similar picture for R^3 can be found in Unit 3, Linear Transformations from R^n to R^m.

Theorem 5.15. *In any inner product space, the orthogonal projection* $\text{proj}_S\ \mathbf{b}$, *of a vector* $\mathbf{b}$ *into a finite dimensional subspace* S, *always exists and is unique.*

Proof. The subspace S must have a basis, and this basis can be converted to an orthogonal basis using the Gram-Schmidt algorithm of the previous section, Orthogonal Bases, the Gram-Schmidt Process, and $QR-$ factorization. Furthermore Theorem 5.13 of the previous section gives a formula for the orthogonal projection in terms of this orthogonal basis - hence, the projection always exists.

In order to show it is unique, suppose $\mathbf{y}_1$ and $\mathbf{y}_2$ are both projections of $\mathbf{b}$. Since $\mathbf{y}_1,\ \mathbf{y}_2 \in S$:

$$\left.\begin{array}{l}\langle \mathbf{b}-\mathbf{y}_1,\ \mathbf{y}_1\rangle = 0 \Longrightarrow \langle \mathbf{b},\ \mathbf{y}_1\rangle = \langle \mathbf{y}_1,\ \mathbf{y}_1\rangle \\ \langle \mathbf{b}-\mathbf{y}_1,\ \mathbf{y}_2\rangle = 0 \Longrightarrow \langle \mathbf{b},\ \mathbf{y}_1\rangle = \langle \mathbf{y}_1,\ \mathbf{y}_2\rangle \end{array}\right\} \text{ and } \left.\begin{array}{l}\langle \mathbf{b}-\mathbf{y}_2,\ \mathbf{y}_2\rangle = 0 \Longrightarrow \langle \mathbf{b},\ \mathbf{y}_2\rangle = \langle \mathbf{y}_2,\ \mathbf{y}_2\rangle \\ \langle \mathbf{b}-\mathbf{y}_2,\ \mathbf{y}_1\rangle = 0 \Longrightarrow \langle \mathbf{b},\ \mathbf{y}_2\rangle = \langle \mathbf{y}_2,\ \mathbf{y}_1\rangle \end{array}\right\}$$

Since $\langle \mathbf{y}_1,\ \mathbf{y}_2\rangle = \langle \mathbf{y}_2,\ \mathbf{y}_1\rangle$ (basis axiom that inner products are commutative) it follows that:

$$\langle \mathbf{y}_1,\ \mathbf{y}_1\rangle = \langle \mathbf{y}_2,\ \mathbf{y}_2\rangle = \langle \mathbf{y}_1,\ \mathbf{y}_2\rangle = \langle \mathbf{y}_2,\ \mathbf{y}_1\rangle$$

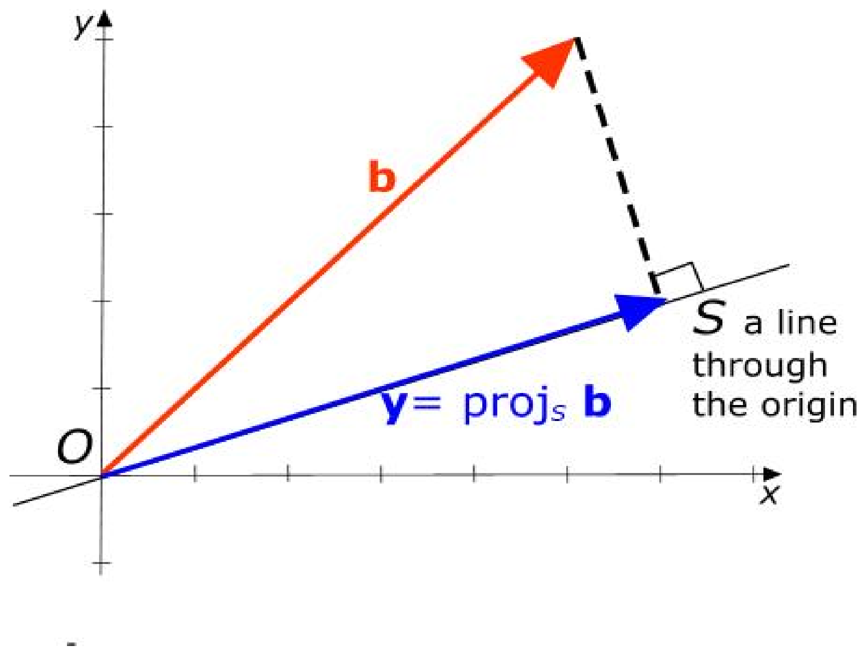

Figure 5.5: Projection onto a Line In R^2

Hence, the square of the distance from $\mathbf{y}_1$ to $\mathbf{y}_2$ is (using the linearity and commutativity axioms of the inner product):

$$\begin{aligned}[d(\mathbf{y}_1,\ \mathbf{y}_2)]^2 &= \|\mathbf{y}_1 - \mathbf{y}_2\|^2 = \langle \mathbf{y}_1 - \mathbf{y}_2,\ \mathbf{y}_1 - \mathbf{y}_2 \rangle \\ &= \langle \mathbf{y}_1,\ \mathbf{y}_1 \rangle + \langle \mathbf{y}_2,\ \mathbf{y}_2 \rangle - 2\langle \mathbf{y}_1,\ \mathbf{y}_2 \rangle \\ &= 0 \text{ (by the equations derived above)}\end{aligned}$$

Hence, $\|\mathbf{y}_1 - \mathbf{y}_2\| = 0$, but the fifth axiom of inner products (see the previous section, Basis Definitions and Properties of Inner Product Spaces) ensures this can only happen if $\mathbf{y}_1 - \mathbf{y}_2 = 0$. That is, $\mathbf{y}_1 = \mathbf{y}_2$ so the orthogonal projection is uniquely defined.

□

Theorem 5.16. *In any inner product space, the orthogonal projection* $\text{proj}_S\ \mathbf{b}$ *is the nearest vector in* S *to the vector* $\mathbf{b}$. *That is any other vector in* $\mathbf{x} \in S$ *satisfies:*

$$d(\mathbf{b},\ \mathbf{x}) > d(\mathbf{b},\ \text{proj}_S\ \mathbf{b}) \text{ or equivalently: } \|\mathbf{b} - \mathbf{x}\| > \|\mathbf{b} - \text{proj}_S\ \mathbf{b}\|$$

Proof. By the definition of projection, $\mathbf{b} - \text{proj}_S\ \mathbf{b}$ is perpendicular to any vector in S and, in particular, if $\mathbf{x} \neq \text{proj}_S\ \mathbf{b}$ is any other vector in S, then $\mathbf{b} - \text{proj}_S\ \mathbf{b}$ is perpendicular to $\mathbf{x} - \text{proj}_S\ \mathbf{b}$. Recall Theorem 5.5 (Pythagoras) of the previous section, Basic Definitions and Properties of Inner Product Spaces, that for two perpendicular vectors $\mathbf{u}, \mathbf{v}$:

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

Applying this with $\mathbf{u} = \mathbf{b} - \text{proj}_S\ \mathbf{b},\ \mathbf{v} = \mathbf{x} - \text{proj}_S\ \mathbf{b}$ to give the result:

$$\begin{aligned}\|(\mathbf{b} - \text{proj}_S\ \mathbf{b}) - (\mathbf{x} - \text{proj}_S\ \mathbf{b})\|^2 &= \|\mathbf{b} - \text{proj}_S\ \mathbf{b}\|^2 + \|\mathbf{x} - \text{proj}_S\ \mathbf{b}\|^2 \\ \|\mathbf{b} - \mathbf{x}\|^2 &= \|\mathbf{b} - \text{proj}_S\ \mathbf{b}\|^2 + \|\mathbf{x} - \text{proj}_S\ \mathbf{b}\|^2\end{aligned}$$

However, $\|\mathbf{x}-\mathrm{proj}_S\ \mathbf{b}\|^2>0$, since $\mathbf{x}-\mathrm{proj}_S\ \mathbf{b}\neq\mathbf{0}$, and so:

$$\|\mathbf{b}-\mathbf{x}\|^2>\|\mathbf{b}-\mathrm{proj}_S\ \mathbf{b}\|^2 \implies \|\mathbf{b}-\mathbf{x}\|>\|\mathbf{b}-\mathrm{proj}_S\ \mathbf{b}\|\text{, as asserted.}$$

□

Theorem 5.17. *Suppose V is an inner product space, and S is a 2-dimensional subspace with a basis $\{\mathbf{v}_1,\ \mathbf{v}_2\}$. If the orthogonal projection vector is given by the linear combination of these basis vectors as:*

$$\textit{proj}_S\ \mathbf{b}=k_1\mathbf{v}_1+k_2\mathbf{v}_2 \text{ where } k_1,\ k_2\in R$$

then $k_1,\ k_2$ are solutions of the linear equations with non-singular coefficient matrix:

$$\begin{bmatrix}\langle\mathbf{v}_1,\ \mathbf{v}_1\rangle & \langle\mathbf{v}_1,\ \mathbf{v}_2\rangle\\ \langle\mathbf{v}_2,\ \mathbf{v}_1\rangle & \langle\mathbf{v}_2,\ \mathbf{v}_2\rangle\end{bmatrix}\begin{bmatrix}k_1\\ k_2\end{bmatrix}=\begin{bmatrix}\langle\mathbf{v}_1,\ \mathbf{b}\rangle\\ \langle\mathbf{v}_2,\ \mathbf{b}\rangle\end{bmatrix}$$

Note: *The matrix is symmetric because $\langle\mathbf{v}_2,\ \mathbf{v}_1\rangle=\langle\mathbf{v}_1,\ \mathbf{v}_2\rangle$.*

Proof. If $\mathbf{b}-\mathrm{proj}_S\ \mathbf{b}$ is orthogonal to $\mathbf{v}_1$ and $\mathbf{v}_2$ then it is left as an exercise for the reader to show that $\mathbf{b}-\mathrm{proj}_S\ \mathbf{b}$ is orthogonal to all vectors in S. Hence, we have the two equations (using $\mathrm{proj}_S\ \mathbf{b}-\mathbf{b}$ instead of $\mathbf{b}-\mathrm{proj}_S\ \mathbf{b}$):

$$\left.\begin{array}{l}\langle\mathrm{proj}_S\ \mathbf{b}-\mathbf{b},\ \mathbf{v}_1\rangle=0\\ \langle\mathrm{proj}_S\ \mathbf{b}-\mathbf{b},\ \mathbf{v}_2\rangle=0\end{array}\right\}\implies\left.\begin{array}{l}\langle k_1\mathbf{v}_1-k_2\mathbf{v}_2-\mathbf{b},\ \mathbf{v}_1\rangle=0\\ \langle k_1\mathbf{v}_1-k_2\mathbf{v}_2-\mathbf{b},\ \mathbf{v}_2\rangle=0\end{array}\right\}\implies\left.\begin{array}{l}k_1\langle\mathbf{v}_1,\ \mathbf{v}_1\rangle+k_2\langle\mathbf{v}_2,\ \mathbf{v}_1\rangle=\langle\mathbf{b},\ \mathbf{v}_1\rangle\\ k_1\langle\mathbf{v}_1,\ \mathbf{v}_2\rangle+k_2\langle\mathbf{v}_2,\ \mathbf{v}_2\rangle=\langle\mathbf{b},\ \mathbf{v}_2\rangle\end{array}\right\}$$

The matrix form of the two equations is the required result (using the commutativity $\langle\mathbf{b},\ \mathbf{v}_1\rangle=\langle\mathbf{v}_1,\ \mathbf{b}\rangle$, $\langle\mathbf{b},\ \mathbf{v}_2\rangle=\langle\mathbf{v}_2,\ \mathbf{b}\rangle,\ \langle\mathbf{v}_2,\ \mathbf{v}_1\rangle=\langle\mathbf{v}_1,\ \mathbf{v}_2\rangle$.

The determinant of the coefficient matrix is (where θ is the angle between the two basis vectors):

$$\begin{aligned}\langle\mathbf{v}_1,\ \mathbf{v}_1\rangle\langle\mathbf{v}_2,\ \mathbf{v}_2\rangle-\langle\mathbf{v}_1,\ \mathbf{v}_2\rangle^2&=\|\mathbf{v}_1\|^2\|\mathbf{v}_2\|^2-\|\mathbf{v}_1\|^2\|\mathbf{v}_2\|^2\cos^2\theta\\ &=\|\mathbf{v}_1\|^2\|\mathbf{v}_2\|^2\left(1-\cos^2\theta\right)\\ &=\|\mathbf{v}_1\|^2\|\mathbf{v}_2\|^2\sin^2\theta\end{aligned}$$

This is zero only if one of $\|\mathbf{v}_1\|,\ \|\mathbf{v}_2\|$ is zero, in which case that vector is the zero vector, and so not a basis vector, or, $\sin\theta=0$ in which case $\theta=0$ and $\mathbf{v}_1,\ \mathbf{v}_2$ are parallel vectors and so cannot be a basis of S. Hence, the determinant cannot be zero when $\{\mathbf{v}_1,\ \mathbf{v}_2\}$ is a basis, and so the coefficient matrix is non-singular.

□

Theorem 5.18. *Suppose V is an inner product space, and S is a $k-$ dimensional subspace with a basis $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\ \mathbf{v}_k\}$. If the orthogonal projection vector is given by the linear combination of these basis vectors as:*

$$\textit{proj}_S\ \mathbf{b}=k_1\mathbf{v}_1+k_2\mathbf{v}_2+\cdots+k_k\mathbf{v}_k \text{ where } k_1,\ k_2,\cdots,\ k_k\in R$$

then $k_1,\ k_2\,_2,\cdots,\ k_k$ are solutions of the $k\times k$ system of linear equations with $k\times k$ non-singular coefficient matrix:

$$\begin{bmatrix}\langle\mathbf{v}_1,\ \mathbf{v}_1\rangle & \langle\mathbf{v}_1,\ \mathbf{v}_2\rangle & \cdots & \langle\mathbf{v}_1,\ \mathbf{v}_k\rangle\\ \langle\mathbf{v}_2,\ \mathbf{v}_1\rangle & \langle\mathbf{v}_2,\ \mathbf{v}_2\rangle & \cdots & \langle\mathbf{v}_2,\ \mathbf{v}_k\rangle\\ \vdots & \vdots & \ddots & \vdots\\ \langle\mathbf{v}_k,\ \mathbf{v}_1\rangle & \langle\mathbf{v}_k,\ \mathbf{v}_2\rangle & \cdots & \langle\mathbf{v}_k,\ \mathbf{v}_k\rangle\end{bmatrix}\begin{bmatrix}k_1\\ k_2\\ \vdots\\ k_k\end{bmatrix}=\begin{bmatrix}\langle\mathbf{v}_1,\ \mathbf{b}\rangle\\ \langle\mathbf{v}_2,\ \mathbf{b}\rangle\\ \vdots\\ \langle\mathbf{v}_k,\ \mathbf{b}\rangle\end{bmatrix}$$

Note: *The matrix is symmetric because $\langle\mathbf{v}_i,\ \mathbf{v}_j\rangle=\langle\mathbf{v}_j,\ \mathbf{v}_i\rangle$ for each i,j.*

Proof. The proof is not given but is similar to the proof of Theorem 5.17.

□

Example 5.7.1.
In P_3 find the projection of the polynomial $f(x) = 2 - 3x + 4x^2$ onto the subspace S with basis $\{b_1(x),\ b_2(x)\} = \{x,\ x^2\}$

(a) Using the inner product $\left\langle a_0 + a_1 x + a_2 x^2,\ b_0 + b_1 x + b_2 x^2 \right\rangle = a_0 b_0 + a_1 b_1 + a_2 b_2$

(b) **Requires calculus.** Using the inner product $\langle f,\ g\rangle = \int_{-1}^{1} f(x)\, g(x)\ dx$

Solution. The following is the solution:

(a) If the projection is the linear combination of the basis vectors $p(x) = k_1 x + k_2 x^2$ then by Theorem 5.17, k_1, k_2 satisfies:

$$\begin{bmatrix} \langle b_1,\ b_1\rangle & \langle b_1,\ b_2\rangle \\ \langle b_2,\ b_1\rangle & \langle b_2,\ b_2\rangle \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \langle b_1,\ f\rangle \\ \langle b_2,\ f\rangle \end{bmatrix}$$
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$$

with solution $k_1 = -3,\ k_2 = 4$. Hence, the projection of $f(x) = 2 - 3x + 4x^2$ onto the subspace with basis $\{x,\ x^2\}$ is the polynomial:

$$p(x) = -3x + 4x^2$$

Note: The projection just drops the constant term of the polynomial, which seems intuitively reasonable.

(b) If the projection is the linear combination of the basis vectors $p(x) = k_1 x + k_2 x^2$ then by Theorem 5.17, k_1, k_2 satisfies:

$$\begin{bmatrix} \langle b_1,\ b_1\rangle & \langle b_1,\ b_2\rangle \\ \langle b_2,\ b_1\rangle & \langle b_2,\ b_2\rangle \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \langle b_1,\ f\rangle \\ \langle b_2,\ f\rangle \end{bmatrix}$$
$$\begin{bmatrix} \int_{-1}^{1} x^2\ dx & \int_{-1}^{1} x^3\ dx \\ \int_{-1}^{1} x^3\ dx & \int_{-1}^{1} x^4\ dx \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \int_{-1}^{1} x\left(2 - 3x + 4x^2\right)\ dx \\ \int_{-1}^{1} x^2\left(2 - 3x + 4x^2\right)\ dx \end{bmatrix}$$

Evaluating the integrals (details not shown) the equations become:

$$\begin{bmatrix} \frac{2}{3} & 0 \\ 0 & \frac{2}{5} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} -2 \\ \frac{44}{15} \end{bmatrix}$$

with solution $k_1 = -3,\ k_2 = \frac{22}{3}$. Hence, the projection of $f(x) = 2 - 3x + 4x^2$ onto the subspace with basis $\{x,\ x^2\}$ is the polynomial:

$$p(x) = -3x + \frac{22}{3}x^2$$

Note: To check the accuracy of the result in part (b), compute the $f(x) - p(x) = 2 - \frac{10}{3}x^2$ (that is $f(x) - \operatorname{proj}_S\ f$), and show it is orthogonal to the basis polynomials, $b_1(x)$ and $b_2(x)$. That is, check that:

$$\int_{-1}^{1} x\left(2 - \frac{10}{3}x^2\right)\ dx = 0 \text{ and } \int_{-1}^{1} x^2\left(2 - \frac{10}{3}x^2\right)\ dx = 0$$

Example 5.7.2.
Find the orthogonal projection of the matrix A onto the subspace S of M_{22} with basis $\{B,\ C\}$, using the usual inner product (scalar product of the vectors formed from the four coefficients of the matrices):

$$A = \begin{bmatrix} 2 & 0 \\ -3 & 2 \end{bmatrix},\ B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ C = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

Solution. If the projection is the linear combination of the basis vectors $D = k_1B + k_2C$ then by Theorem 5.17, k_1, k_2 satisfies:

$$\begin{bmatrix} \langle B,\ B\rangle & \langle B,\ C\rangle \\ \langle C,\ B\rangle & \langle C,\ C\rangle \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \langle B,\ A\rangle \\ \langle C,\ A\rangle \end{bmatrix}$$
$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

with solution $k_1 = \frac{13}{5},\ k_2 = -\frac{6}{5}$. Hence, the projection is:

$$D = \frac{13}{5}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{6}{5}\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{13}{5} & -\frac{6}{5} \\ -\frac{6}{5} & \frac{7}{5} \end{bmatrix}$$

Note: To check the accuracy of the result in part (b), compute $A - D = \begin{bmatrix} -\frac{3}{5} & \frac{6}{5} \\ -\frac{9}{5} & \frac{3}{5} \end{bmatrix}$ (that is $A - \text{proj}_S\ A$), and show it is orthogonal to the basis matrices B, C. That is, check that:

$$\langle A - D, B\rangle = 0 \text{ and } \langle A - D, C\rangle = 0$$

5.7.2 Properties of orthogonal projections in Euclidean spaces R^n

Euclidean spaces, R^n, $n = 1, 2, 3, \cdots$ are inner product spaces. We assume the usual inner product $\langle \mathbf{u},\ \mathbf{v}\rangle = \mathbf{u} \cdot \mathbf{v}$ - the scalar product. The orthogonal projection in a Euclidean space R^n is often called the **Least Squares Approximation**. The reason for this name is that if $\mathbf{b}$ is a vector and $\mathbf{y} = \text{proj}_S\ \mathbf{b}$ is its projection on a subspace S then $\mathbf{y}$ is the nearest vector to $\mathbf{b}$ in S. Hence, if $\mathbf{e} = \mathbf{b} - \mathbf{y}$ is the "error" vector of the approximation of $\mathbf{b}$ by $\mathbf{y}$ then the length $\|\mathbf{e}\|$ is minimized. However:

$$\|\mathbf{e}\| = \sqrt{e_1^2 + e_2^2 + \cdots + e_n^2}$$

where $e_1, e_2, \cdots, e_n$ are the components of $\mathbf{e}$, and so the orthogonal projection minimizes the sum of the squares of the e_i - that is it finds the least value of the sum of squares.

The results in the previous section for general Inner Product Spaces, hold in Euclidean spaces but have some extra features. Theorem 5.17 becomes:

Theorem 5.19. *Suppose that in R^n, for some $n > 2$, S is a 2-dimensional subspace with a basis $\{\mathbf{v}_1,\ \mathbf{v}_2\}$. If the orthogonal projection of the vector $\mathbf{b}$ is given by the linear combination of these basis vectors as:*

$$\textit{proj}_S\ \mathbf{b} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 \textit{ where } k_1,\ k_2 \in R$$

then $k_1,\ k_2$ are solutions of the linear equations with symmetric non-singular coefficient matrix:

$$\begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{b} \\ \mathbf{v}_2 \cdot \mathbf{b} \end{bmatrix}$$

This can also be re-written in terms of products of matrices as:

$$A^TAX = A^TB$$

where:

$$\overbrace{\begin{bmatrix} \mathbf{v}_1^T \\ --- \\ \mathbf{v}_2^T \end{bmatrix}}^{A^T} \overbrace{\begin{bmatrix} \mathbf{v}_1 | & \mathbf{v}_2 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}}^{X} = \overbrace{\begin{bmatrix} \mathbf{v}_1^T \\ --- \\ \mathbf{v}_2^T \end{bmatrix}}^{A^T} \overbrace{[\mathbf{b}]}^{B}$$

That is, the matrix A has the column vectors $\mathbf{v}_1,\ \mathbf{v}_2$ as columns, the transpose matrix A^T has the vectors $\mathbf{v}_1,\ \mathbf{v}_2$ as rows, and B has the single vector $\mathbf{b}$ as its column.

Note: The matrix equations $A^TAX = A^TB$ are called the **normal equations** and are usually derived directly using matrix/vector methods in R^n without appealing to the general theorem of inner product spaces.

Proof. The first matrix equation containing the scalar products follows directly from Theorem 5.17. The proof that this is the same as the second matrix form follows directly by using block matrix multiplication of the rows $\mathbf{v}_1^T, \mathbf{v}_2^T$ of A^T by the columns $\mathbf{v}_1, \mathbf{v}_2, \mathbf{b}$ of A and B:

$$\begin{bmatrix} \mathbf{v}_1^T \\ --- \\ \mathbf{v}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 | & \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1^T \\ --- \\ \mathbf{v}_2^T \end{bmatrix} [\mathbf{b}] \Longrightarrow$$

$$\begin{bmatrix} \mathbf{v}_1^T\mathbf{v}_1 & \mathbf{v}_1^T\mathbf{v}_2 \\ \mathbf{v}_2^T\mathbf{v}_1 & \mathbf{v}_2^T\mathbf{v}_2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1^T\mathbf{b} \\ \mathbf{v}_2^T\mathbf{b} \end{bmatrix}$$

Since $\mathbf{v}_i^T\mathbf{v}_j = \mathbf{v}_i \cdot \mathbf{v}_j$ and $\mathbf{v}_i^T\mathbf{b} = \mathbf{v}_i \cdot \mathbf{b}$ this last equation is the same as first matrix equation of the theorem. □

Theorem 5.20. *Suppose that in R^n, for some $n > k$, S is a $k-$ dimensional subspace with a basis $\{\mathbf{v}_1,\ \mathbf{v}_2, \cdots, \mathbf{v}_k\}$. If the orthogonal projection of the vector $\mathbf{b}$ is given by the linear combination of these basis vectors as:*

$$\textit{proj}_S\ \mathbf{b} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_k\mathbf{v}_k \text{ where } k_1,\ k_2, \cdots, k_k \in R$$

then $k_1,\ k_2, \cdots, k_k$ are solutions of the linear equations with symmetric non-singular coefficient matrix:

$$\begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \cdots & \mathbf{v}_1 \cdot \mathbf{v}_k \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & \cdots & \mathbf{v}_2 \cdot \mathbf{v}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_k \cdot \mathbf{v}_1 & \mathbf{v}_k \cdot \mathbf{v}_2 & \cdots & \mathbf{v}_k \cdot \mathbf{v}_k \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_k \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{b} \\ \mathbf{v}_2 \cdot \mathbf{b} \\ \vdots \\ \mathbf{v}_k \cdot \mathbf{b} \end{bmatrix}$$

This can also be re-written in terms of products of matrices as:

$$A^TAX = A^TB$$

where:

$$\overbrace{\begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix}}^{A^T} \overbrace{\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_k \end{bmatrix}}^{X} = \overbrace{\begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix}}^{A^T} \overbrace{[\mathbf{b}]}^{B}$$

That is, the matrix A has the column vectors $\mathbf{v}_1,\ \mathbf{v}_2, \cdots, \mathbf{v}_k$ as columns, the transpose matrix A^T has the row vectors $\mathbf{v}_1^T,\ \mathbf{v}_2^T, \cdots, \mathbf{v}_k^T$ as rows, and B has the single vector $\mathbf{b}$ as its column.

Proof. The proof is not given here, but it would use methods analogous to the proof of Theorem 5.19. □

Theorem 5.21. *Suppose that in R^n, for some $n > k$, S is a $k-$ dimensional subspace with a basis $\{\mathbf{v}_1,\ \mathbf{v}_2, \cdots, \mathbf{v}_k\}$, and A has the $\mathbf{v}_i$ as columns:*

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}$$

The projection matrix P, such that for any $\mathbf{b} \in R^n$, $P\mathbf{b}$ is the projection of $\mathbf{b}$ onto S is given by:

$$P = A \left(A^T A\right)^{-1} A^T$$

Proof. From Theorem 5.20, the matrix $A^T A$ is non-singular, and the projection satisfies:

$$\text{proj}_S\ \mathbf{b} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_k\mathbf{v}_k = \overbrace{\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_k \end{bmatrix}}^{X} = AX$$

X is the solution of $A^T AX = A^T B$, where $B = [\mathbf{b}]$. This is given by the inverse matrix method as:

$$X = \left(A^T A\right)^{-1} A^T B$$

and so the projection is:

$$P = AX = A \left(A^T A\right)^{-1} A^T B$$

and so the projection matrix is $A \left(A^T A\right)^{-1} A^T$.

□

Example 5.7.3.
Find the orthogonal projection of the vector $\mathbf{b} = (1, 2, 3)$ onto the subspace S with basis $\{\mathbf{v}_1,\ \mathbf{v}_2\} = \{(1, 0, 2),\ (0, -1, 3)\}$.

Solution. If the projection is $\mathbf{p} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2$ then, using Theorem 5.19, the equations satisfied by k_1, k_2 are:

$$\begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{b} \\ \mathbf{v}_2 \cdot \mathbf{b} \end{bmatrix} \Longrightarrow \begin{bmatrix} 5 & 6 \\ 6 & 10 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

with solution $k_1 = 2,\ k_2 = -\frac{1}{2}$. Hence, the orthogonal projection is:

$$\mathbf{p} = 2\,(1, 0, 2) - \frac{1}{2}\,(0, -1, 3) = \left(2, \frac{1}{2}, \frac{5}{2}\right)$$

Note: To check the accuracy of this result, show that $\mathbf{b} - \mathbf{p} = \left(-1, \frac{3}{2}, \frac{1}{2}\right)$ is orthogonal to the basis vectors $\mathbf{v}_1,\ \mathbf{v}_2$. That is, check that:

$$\left(-1, \frac{3}{2}, \frac{1}{2}\right) \cdot (1, 0, 2) = 0 \text{ and } \left(-1, \frac{3}{2}, \frac{1}{2}\right) \cdot (0, -1, 3) = 0$$

Note: The alternative way of setting-up the equations, used in most textbooks, is as follows. Construct the matrices, A with the basis vectors as columns and B with $\mathbf{b}$ as its column:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 2 & 3 \end{bmatrix},\ B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Compute the normal equations $A^TAX = A^TB$ (these will be the same as the equations used above to solve this example):

$$\overset{A^T}{\begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 3 \end{bmatrix}} \overset{A}{\begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 2 & 3 \end{bmatrix}} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \overset{A^T}{\begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 3 \end{bmatrix}} \overset{B}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}$$

After multiplying the first two matrices on each side these become:

$$\begin{bmatrix} 5 & 6 \\ 6 & 10 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \end{bmatrix}$$

5.7.3 Applications of least squares in Euclidean spaces R^n

Approximate solutions of inconsistent systems of equations

In many applications the solution of a problem is given by the solution of a linear system of equations, but this system turns out to be inconsistent (it does not have a solution). This can happen when there is an excess of data that produces more equations than unknowns and inaccuracies in the data or model used ensure this system is inconsistent. In other applications the model deliberately creates an inconsistent system and requires a best possible approximate solution to the system (see Example 5.7.5 below).

There are many ways to find an approximate solution to an inconsistent system. The **Least Squares Approximation** of an inconsistent system $AX = B$ uses results established in Theorems 5.19, 5.20 and 5.21 that are, for convenience, re-written in equation terminology in Theorem 5.22.

Theorem 5.22. *Suppose a linear system is $AX = B$, where$A$ is an $m \times n$ matrix with $m \geq n,\ X$ is the $n \times 1$ column of unknowns, and B is the $m \times 1$ column vector of right hand values. The least squares approximate solution for X of the system is defined to be the orthogonal projection of B onto the space spanned by the columns of A, and it is given by the solution of the normal equations:*

$$A^TAX = A^TB$$

The least squares approximation minimizes the sum of the squares of the differences between the right hand side and left hand side of each of the equations (that is, each row of $AX = B$). If the columns of A are linearly independent then the coefficient matrix A^TA is non-singular and then the solution can be written:

$$X = \left(A^TA\right)^{-1} A^TB$$

The orthogonal projection of B onto the column space of A is given by:

$$AX = A\left(A^TA\right)^{-1} A^TB, \text{ and the projection matrix is: } A\left(A^TA\right)^{-1} A^T$$

Note: This formula holds even if the original system is consistent (has an exact solution). In that case the least squares approximation is the same as the exact solution of the system. When the columns of A are linearly dependent, the coefficient matrix A^TA is singular, and the normal equations have infinitely many solutions, but all of these will give the same value of AX.

Proof. This theorem is a variation of Theorems 5.19 and 5.20. To make the connection clearer, note that AX is simply a linear combination of the column vectors of A. That is if, in column form, $A = \begin{bmatrix} \mathbf{v}_1| & \mathbf{v}_2| & \cdots & |\,\mathbf{v}_n \end{bmatrix}$ and $X^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ then:

$$AX = \begin{bmatrix} \mathbf{v}_1| & \mathbf{v}_2| & \cdots & |\,\mathbf{v}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

Hence, the problem is to find the orthogonal projection of the right hand side B onto the subspace spanned by columns of A - the problem solved by Theorems 5.19 and 5.20.

Theorems 5.19 and 5.20 did not cover the case where the matrix A^TA is singular (because the columns of A are linearly dependent). However, in that case, if X is any one solution of the normal equations $A^TAX = A^TB$ the all other solutions are given by $X + Y$ where Y is any vector in the null space of A. In that case all solutions give the same value of for the projection AX because $A(X + Y) = AX + AY = AX$.

□

Example 5.7.4.
Find the least squares approximation, and find the error in the solution for the system of equations $AX = B$:

$$\begin{bmatrix} 1 & -3 \\ -2 & 4 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix}$$

Solution. Solving the first two equations gives $x_1 = -\frac{11}{2},\ x_2 = -\frac{5}{2}$, but these two values do not satisfy the third equation, so the system is inconsistent. Using the method of Theorem 5.19, the least square approximation X is the solution of the normal equations $A^TAX = A^TB$, and this is:

$$\begin{bmatrix} 1 & -2 & 1 \\ -3 & 4 & -4 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ -2 & 4 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 \\ -3 & 4 & -4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix}$$
$$\begin{bmatrix} 6 & -15 \\ -15 & 41 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ -26 \end{bmatrix}$$

with solution $x_1 = -\frac{48}{7}$ and $x_2 = -\frac{22}{7}$. The differences between the left and right sides gives the error vector:

$$\begin{bmatrix} 1 & -3 \\ -2 & 4 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} -\frac{48}{7} \\ -\frac{22}{7} \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{18}{7} \\ \frac{8}{7} \\ \frac{40}{7} \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{4}{7} \\ \frac{1}{7} \\ -\frac{2}{7} \end{bmatrix}$$

Hence, the errors in the three equations are $\frac{4}{7}, \frac{1}{7}, -\frac{2}{7}$.
Note: The method actually minimizes the sum of the squares of the errors. That is, in this example, $\left(\frac{4}{7}\right)^2 + \left(\frac{1}{7}\right)^2 + \left(-\frac{2}{7}\right)^2 = \frac{3}{7}$ is as small as it possibly can be.
Note: In many inconsistent systems of equations it would be better to minimize the sums of the absolute values of the errors (rather than the squares of these), but this is a more difficult problem to solve. In particular, when an inconsistent system has an equation constructed from data that is totally wrong, then the least squares method tends to exaggerate the affect of that invalid equation on the approximate solution (rather than minimizing it).

Linear regression

Example 5.7.5.
A large set of data in R^2: $(x_i,\ y_i)\,,\ i = 1, 2, \cdots, m$ is obtained in an experiment. The data should lie on a straight line but experimental errors, or inaccuracies in the model, have ensured that the data does not lie on a straight line. The experimenters want to find a straight line, $y = a + bx$, that best fits the data, and they decide to minimize the sums of the squares of the deviations in the $y-$ values. Show that the values for a, b are given by:

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{m \sum x^2 - \left(\sum x\right)^2},\ b = \frac{m \sum xy - \sum x \sum y}{m \sum x^2 - \left(\sum x\right)^2}$$

where the formulae use the following abbreviations for clarity: $\sum xy$ means $\sum_{i=1}^{m} x_i y_i$, $\sum x$ means $\sum_{i=1}^{m} x_i$, $\sum y$ means $\sum_{i=1}^{m} y_i$, $\sum x^2$means $\sum_{i=1}^{m} x_i^2$.

Solution. Set up each data point for the equation in the form $a + bx = y$, with unknowns a and b :

$$\left.\begin{array}{c} a + bx_1 = y_1 \\ a + bx_2 = y_2 \\ a + bx_3 = y_3 \\ \vdots \\ a + bx_m = y_m \end{array}\right\} \Longrightarrow \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

This is (almost certainly) an inconsistent system of the form $AX = B$, and so we can solve it by the method of Theorem 5.22, by solving the system $A^T AX = A^T B$:

$$\begin{bmatrix} 1 & 1 & 1 & \vdots & 1 \\ x_1 & x_2 & x_3 & \vdots & x_m \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \vdots & 1 \\ x_1 & x_2 & x_3 & \vdots & x_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

$$\begin{bmatrix} \sum_{i=1}^{m} 1 & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

Noting that $\sum_{i=1}^{m} 1 = m$, the solution given by the formula for the inverse of a 2×2 matrix is:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{m\sum x^2 - \left(\sum x\right)^2} \begin{bmatrix} \sum x^2 & -\sum x \\ -\sum x & m \end{bmatrix} \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{m\sum x^2 - \left(\sum x\right)^2} \begin{bmatrix} \sum y \sum x^2 - \sum x \sum xy \\ m\sum xy - \sum x \sum y \end{bmatrix} = \begin{bmatrix} \frac{\sum y \sum x^2 - \sum x \sum xy}{m\sum x^2 - \left(\sum x\right)^2} \\ \frac{m\sum xy - \sum x \sum y}{m\sum x^2 - \left(\sum x\right)^2} \end{bmatrix}$$

Note: The regression line is a basic result in statistics ,where it is often written in a simpler form using statistical constructs: $\overline{x} = \frac{\sum x}{m}$ (the mean of the x_i values), $\overline{y} = \frac{\sum y}{m}$. The formulae become (see the exercise set for more details):

$$b = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sum (x - \overline{x})^2} \text{ and } a = \overline{y} - b\overline{x}$$

Note: See the exercise set for numerical examples using this formula.

Least squares approximations for nonlinear functions

The least squares method uses linear methods but it can be used to find least squares data fits, using polynomials and exponential functions.

Example 5.7.6.
A given set of data in R^2: $(x_i,\ y_i)\,,\, i=1,2,\cdots,m$ is expected to fit a quadratic function but experimental errors, or inaccuracies in the model, have ensured that no quadratic fits it exactly. We want to find a quadratic, $y=a+bx+cx^2$, by minimizing the sums of the squares of the deviations in the $y-$ values. Show that the values of a,b,c are given by the solutions of the normal equations $A^TAX=A^TB$, where:

$$A=\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix},\quad X=\begin{bmatrix} a \\ b \\ c \end{bmatrix},\quad B=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

Solution. Set up each data point for the equation in the form $a+bx+cx^2=y$, with unknowns a,b,c :

$$\left.\begin{array}{c} a+bx_1+cx_1^2=y_1 \\ a+bx_2+cx_2^2=y_2 \\ a+bx_3+cx_3^2=y_3 \\ \vdots \\ a+bx_m+cx_m^2=y_m \end{array}\right\} \Longrightarrow \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix}=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

This is (almost certainly) an inconsistent system of the form $AX=B$ and so we can solve it by the method of Theorem 5.22 (and Example 5.7.5) by solving the system $A^TAX=A^TB$ as asserted.

Note: For numerical examples, see the exercise set.

Note: A method exactly analogous to this example can be used to find polynomials of any degree that approximate a data set. However, the problem becomes numerically unstable if the degree of the polynomial is large. The process generally works reasonably well with polynomials of degree 3 and 4.

Note: The method will also become unstable and may give meaningless results if the data really does not all lie reasonably close to a polynomial of the degree used in the approximation.

Example 5.7.7.
A given set of data in R^2: $(x_i,\ y_i)\,,\, i=1,2,\cdots,m$ is expected to fit an exponential function, $y=ae^{bx}$, but experimental errors, or inaccuracies in the model, have ensured that no such function fits it exactly. We want to find an exponential function $y=ae^{bx}$ by minimizing the sums of the squares of the deviations in the $y-$ values. Show that the values of a,b are given by the solutions of the normal equations $A^TAX=A^TB$, where:

$$A=\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix},\quad X=\begin{bmatrix} \ln a \\ b \end{bmatrix},\quad B=\begin{bmatrix} \ln y_1 \\ \ln y_2 \\ \ln y_3 \\ \vdots \\ \ln y_m \end{bmatrix}$$

Solution. Set up each data point for the equation in the form $ae^{bx}=y$, with unknowns a,b :

$$\left.\begin{array}{c} ae^{bx_1}=y_1 \\ ae^{bx_2}=y_2 \\ ae^{bx_3}=y_3 \\ \vdots \\ ae^{bx_m}=y_m \end{array}\right\} \underset{\text{both sides}}{\overset{\text{Take logs of}}{\Longrightarrow}} \left.\begin{array}{c} \ln a+bx_1=\ln y_1 \\ \ln a+bx_2=\ln y_2 \\ \ln a+bx_3=\ln y_3 \\ \vdots \\ \ln a+bx_m=\ln y_m \end{array}\right\} \Longrightarrow \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}\begin{bmatrix} \ln a \\ b \end{bmatrix}=\begin{bmatrix} \ln y_1 \\ \ln y_2 \\ \ln y_3 \\ \vdots \\ \ln y_m \end{bmatrix}$$

This is (almost certainly) an inconsistent system of the form $AX = B$, and so we can solve it by the method of Theorem 5.22 (and Examples 5.7.5, 5.7.6) by solving the system $A^T AX = A^T B$ as asserted.

Note: For numerical examples, see the exercise set.

Note: The solution does not actually solve the problem as stated, because it actually finds the least squares solution of the equations formed by taking logs of both sides. That is, the solution is not a least squares solution of the original equations $ae^{bx_i} = y_i$.

Section 5.7 exercise set

Check your understanding by answering the following questions.

1. Find the orthogonal projection in R^3 of the vector $(1,\ 2,\ 3)$ onto the two-dimensional subspace spanned by the vectors $\{(2, 0, -3)\,,\ (2, 1,\ 1)\}$, using the inner product:
   (a) The normal scalar product
   (a) The inner product: $\langle (a, b, c)\,, (e, f, g)\rangle = ae + bf + 2cg$

2. Find the orthogonal projection in P_4 of the polynomial $f(x) = x + x^4$ onto P_2 with basis $\{b_1(x)\,, b_2(x)\,, b_3(x)\} = \left\{1,\ x,\ x^2\right\}$ using the inner product:
   (a) $\left\langle a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4,\ \ b_0 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4 \right\rangle$
   $= a_0 b_0 + a_1 b_1 + a_2 b_2 + a_3 b_3 + a_4 b_4$
   (b) **Requires calculus:** $\langle p,\ q\rangle = \int_{-1}^{1} p(x)\, q(x)\ dx$

3. Find the projection of the matrix A onto the subspace of M_{22} spanned by the matrices B, C, D with the inner product defined in the usual way as the scalar product of the vectors formed from the four entries in each matrix or, equivalently, as the trace (sum of diagonal values) of the transpose of one matrix multiplied by the other matrix.

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix},\ B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},\ C = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},\ D = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

4. Find the projection matrix P for the projection in R^4 onto the subspace S spanned by the two vectors:

$$\{(1, 3, -2,\ 1)\,,\ (0,\ 1,\ 0,\ 1)\}$$

   That is, $P\mathbf{v}$ gives the orthogonal projection of $\mathbf{v}$ in S.

5. Find the least squares approximation, and find the errors in each equation, for the linear systems $AX = B$:
   (a) when:

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 2 \\ 0 & 1 \end{bmatrix},\ B = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

   (b) when:

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 2 \\ 0 & 1 \end{bmatrix},\ B = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

   and explain the meaning of the unusual set of errors.

(c) when:

$$A=\begin{bmatrix}2&0&1\\-1&2&0\\0&1&0\\0&0&1\end{bmatrix},\ B=\begin{bmatrix}2\\-1\\3\\2\end{bmatrix}$$

6. Find the regression line for the data:

$$\{(0,\ 1),\ (1,\ 1.3),\ (2,\ 1.9),\ (3,\ 2.6),\ (4,\ 3)\}$$

7. **More difficult and needs a calculator or computer.** Find a plane that fits the data:

$$(x_i,\ y_i,\ z_i)=\{(0,\ 1,\ 1),\ (1,\ 1.3,\ 1.9),\ (2,\ 1.9,\ 3.1),\ (3,\ 2.6,\ 4.2),\ (4,\ 3,\ 4.8)\}$$

That is, assume the plane is $z=a+bx+cy$, and do the least squares fit on the $z-$ values.

8. **More difficult.** Given the data:

$$\{(0,\ 2),\ (1,\ 0),\ (2,\ -1),\ (3,\ 0),\ (4,\ 1)\}$$

(a) Find a quadratic, $y=a+bx+cx^2$, that gives the best least squares fit to the data. Find the sum of squares of the errors.

(b) Find a cubic, $y=a+bx+cx^2+dx^3$, that gives the best least squares fit to the data. Find the sum of squares of the errors.

9. **More difficult.** Given the data:

$$\{(-1,\ 0.5),\ (0,\ 1),\ (0.5,\ 2),\ (1,\ 3),\ (1.5,\ 5)\}$$

(a) Find an exponential, $y=ae^{bx}$, that gives the best least squares fit to the data. Find all of the errors in the $y-$ values.

(b) Find a quadratic, $y=a+bx+cx^3$, that gives the best least squares fit to the data. Find the errors in the $y-$ values.

10. **Requires calculus and difficult.** Find the orthogonal projection of $f(x)=e^x$ onto the set of polynomials $\{g_1(x),\ g_2(x),\ g_3(x),\ g_4(x)\}=\{1,x,x^2,x^3\}$ in the (infinite dimensional) vector space of all continuous functions, with the inner product:

$$\langle p,\ q\rangle=\int_{-1}^{1}f(x)\,g(x)\,dx$$

11. **More difficult.** Show that the simpler forms of the formulae given in Example 5.7.5 are correct:

$$b=\frac{m\sum xy-\sum x\sum y}{m\sum x^2-\left(\sum x\right)^2}=\frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}$$

$$a=\frac{\sum y\sum x^2-\sum x\sum xy}{m\sum x^2-\left(\sum x\right)^2}=\overline{y}-b\overline{x}$$

Solutions

1. For the vector $\mathbf{b}=(1,\ 2,\ 3)$, and subspace S spanned by $\{\mathbf{v}_1,\ \mathbf{v}_2\}=\{(2,0,-3),\ (2,1,\ 1)\}$, the orthogonal projection of $\mathbf{b}$ onto S is given by $\mathbf{p}=k_1\mathbf{v}_1+k_2\mathbf{v}_2$, given by the solution of (Theorem 5.17):

$$\begin{bmatrix}\langle\mathbf{v}_1,\ \mathbf{v}_1\rangle&\langle\mathbf{v}_1,\ \mathbf{v}_2\rangle\\\langle\mathbf{v}_2,\ \mathbf{v}_1\rangle&\langle\mathbf{v}_2,\ \mathbf{v}_2\rangle\end{bmatrix}\begin{bmatrix}k_1\\k_2\end{bmatrix}=\begin{bmatrix}\langle\mathbf{v}_1,\ \mathbf{b}\rangle\\\langle\mathbf{v}_2,\ \mathbf{b}\rangle\end{bmatrix}$$

and this becomes:

(a) For the normal scalar product:

$$\begin{bmatrix} \mathbf{v}_1\cdot\mathbf{v}_1 & \mathbf{v}_1\cdot\mathbf{v}_2 \\ \mathbf{v}_2\cdot\mathbf{v}_1 & \mathbf{v}_2\cdot\mathbf{v}_2 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1\cdot\mathbf{b} \\ \mathbf{v}_2\cdot\mathbf{b} \end{bmatrix}$$

$$\begin{bmatrix} 13 & 1 \\ 1 & 6 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} -7 \\ 7 \end{bmatrix}$$

with solution $k_1 = -\frac{7}{11},\ k_2 = \frac{14}{11}$. Hence, the orthogonal projection is:

$$\mathbf{p} = \left(-\frac{7}{11}\right)(2,0,-3) + \frac{14}{11}(2,1,\ 1) = \left(\frac{14}{11},\ \frac{14}{11},\ \frac{35}{11}\right)$$

(b) For the inner product $\langle (a,b,c),(e,f,g)\rangle = ae + bf + 2cg$ this becomes:

$$\begin{bmatrix} 22 & -2 \\ -2 & 7 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} -16 \\ 10 \end{bmatrix}$$

with solution $k_1 = -\frac{46}{75},\ k_2 = \frac{94}{75}$. Hence, the orthogonal projection is:

$$\mathbf{p} = \left(-\frac{46}{75}\right)(2,0,-3) + \frac{94}{75}(2,1,\ 1) = \left(\frac{32}{25},\ \frac{94}{75},\ \frac{232}{75}\right)$$

2. Given $f(x) = x + x^4$ and basis $\{b_1(x), b_2(x), b_3(x)\} = \{1,\ x,\ x^2\}$ of P_2, the projection of f onto P_2 is given by $= p(x) = k_1 b_1(x) + k_2 b_2(x) + k_3 b_3(x)$ where k_1, k_2, k_3 are given by:

(a) With inner product $\langle a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4,\ \ b_0 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4\rangle$ $= a_0 b_0 + a_1 b_1 + a_2 b_2 + a_3 b_3 + a_4 b_4$, solve:

$$\begin{bmatrix} \langle b_1,\ b_1\rangle & \langle b_1,\ b_2\rangle & \langle b_1,\ b_3\rangle \\ \langle b_2,\ b_1\rangle & \langle b_2,\ b_2\rangle & \langle b_2,\ b_3\rangle \\ \langle b_3,\ b_1\rangle & \langle b_3,\ b_2\rangle & \langle b_3,\ b_3\rangle \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} \langle b_1,\ f\rangle \\ \langle b_2,\ f\rangle \\ \langle b_3,\ f\rangle \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Hence, the projection of $f(x) = x + x^4$ is $p(x) = x.$

(b) With inner product $\langle p,\ q\rangle = \int_{-1}^{1} p(x)\, q(x)\ dx$, solve:

$$\begin{bmatrix} \int_{-1}^{1} 1\ dx & \int_{-1}^{1} x\ dx & \int_{-1}^{1} x^2\ dx \\ \int_{-1}^{1} x\ dx & \int_{-1}^{1} x^2\ dx & \int_{-1}^{1} x^3\ dx \\ \int_{-1}^{1} x^2\ dx & \int_{-1}^{1} x^3\ dx & \int_{-1}^{1} x^4\ dx \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} \int_{-1}^{1} \left(x + x^4\right)\ dx \\ \int_{-1}^{1} x\left(x + x^4\right)\ dx \\ \int_{-1}^{1} x^2\left(x + x^4\right)\ dx \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{5} \end{bmatrix}\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} \\ \frac{2}{3} \\ \frac{2}{7} \end{bmatrix} \Longrightarrow \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} -\frac{3}{35} \\ 1 \\ \frac{6}{7} \end{bmatrix}$$

Hence, the projection of $f(x) = x + x^4$ is $p(x) = -\frac{3}{35} + x + \frac{6}{7}x^2$.

Note: To check your result, show that $\langle f - p,\ b_i\rangle = 0$ for each basis function b_i. Since $f(x) - p(x) = \frac{3}{35} - \frac{6}{7}x^2 + x^4$, we have to show (the correct results):

$$\int_{-1}^{1}\left(\frac{3}{35} - \frac{6}{7}x^2 + x^4\right)\ dx = 0,\quad \int_{-1}^{1}\left(\frac{3}{35}x - \frac{6}{7}x^3 + x^5\right)\ dx = 0$$

$$\int_{-1}^{1}\left(\frac{3}{35}x^2 - \frac{6}{7}x^4 + x^6\right)\ dx = 0$$

3. Given:

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix},\ B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\ C = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix},\ D = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

the projection is the linear combination of the basis vectors $D = k_1 B + k_2 C + k_3 D$. By Theorem 5.17, k_1, k_2, k_3 satisfy:

$$\begin{bmatrix} \langle B,\ B\rangle & \langle B,\ C\rangle & \langle B,\ D\rangle \\ \langle C,\ B\rangle & \langle C,\ C\rangle & \langle C,\ D\rangle \\ \langle D,\ B\rangle & \langle D,\ C\rangle & \langle D,\ D\rangle \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} \langle B,\ A\rangle \\ \langle C,\ A\rangle \\ \langle D,\ A\rangle \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \Longrightarrow \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

Hence, the required projection matrix $P = B + \frac{2}{3}C + \frac{2}{3}D$:

$$P = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & 1 \\ \frac{4}{3} & \frac{2}{3} \end{bmatrix}$$

Note: To check your result, show that $\langle (P-A),\ B\rangle = 0,\ \langle (P-A),\ C\rangle = 0,\ \langle (P-A),\ D\rangle = 0,$ or equivalently, the trace (sum of diagonal entries) is zero for each of $(P-A)^T B,\ (P-A)^T C,\ (P-A)^T D$.

4. If the basis vectors of S are the columns of the matrix A :

$$A = \begin{bmatrix} 1 & 0 \\ 3 & 1 \\ -2 & 0 \\ 1 & 1 \end{bmatrix}$$

then the projection matrix is given by Theorem 5.21 as: $P = A\left(A^T A\right)^{-1} A^T$:

$$P = \begin{bmatrix} 1 & 0 \\ 3 & 1 \\ -2 & 0 \\ 1 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 3 & -2 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 1 \\ -2 & 0 \\ 1 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 3 & -2 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 3 & 1 \\ -2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 15 & 4 \\ 4 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 3 & -2 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 3 & 1 \\ -2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{7} & -\frac{2}{7} \\ -\frac{2}{7} & \frac{15}{14} \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 3 & 1 \\ -2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{7} & \frac{1}{7} & -\frac{2}{7} & -\frac{1}{7} \\ -\frac{2}{7} & \frac{3}{14} & \frac{4}{7} & \frac{11}{14} \end{bmatrix}$$

$$P = \begin{bmatrix} \frac{1}{7} & \frac{1}{7} & -\frac{2}{7} & -\frac{1}{7} \\ \frac{1}{7} & \frac{9}{14} & -\frac{2}{7} & \frac{5}{14} \\ -\frac{2}{7} & -\frac{2}{7} & \frac{4}{7} & \frac{2}{7} \\ -\frac{1}{7} & \frac{5}{14} & \frac{2}{7} & \frac{9}{14} \end{bmatrix}$$

Note: To check this is correct, you can show that for any vector $\mathbf{v} \in R^4$, $(\mathbf{v} - P\mathbf{v}) \cdot (1, 3, -2,\ 1) = 0$ and $(\mathbf{v} - P\mathbf{v}) \cdot (0,\ 1,\ 0,\ 1) = 0$, but this is somewhat complicated.

5. The least squares projection of the system $AX = B$ is given, by Theorem 5.22, as the solution of the normal equations $A^T AX = A^T B$:

(a)

$$\begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -1 & 2 \\ 0 & 1 \end{bmatrix} X = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$
$$\begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix} X = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

with solution $X^T = \begin{bmatrix} \frac{9}{7} & \frac{5}{7} \end{bmatrix}$ (least square approximation for the original system). The errors for each equation are given by $AX - B$:

$$\begin{bmatrix} 2 & 0 \\ -1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{9}{7} \\ \frac{5}{7} \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{4}{7} \\ \frac{8}{7} \\ -\frac{16}{7} \end{bmatrix}$$

(b) This is the same matrix A as part (a) and so the left hand side matrix of the normal equations is the same:

$$\begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix} X = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

with solution $X^T = \begin{bmatrix} 1 & 1 \end{bmatrix}$ (least square approximation for the original system). The errors for each equation are given by $AX - B$:

$$\begin{bmatrix} 2 & 0 \\ -1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The errors are all zero. This means that the original system, even though it has more equations than variables, does have an exact solution, and it is the solution given by the least squares method.

(c)

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} X = \begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \\ 2 \end{bmatrix}$$
$$\begin{bmatrix} 5 & -2 & 2 \\ -2 & 5 & 0 \\ 2 & 0 & 2 \end{bmatrix} X = \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}$$

with solution $X^T = \begin{bmatrix} \frac{7}{11} & \frac{5}{11} & \frac{15}{11} \end{bmatrix}$ (least square approximation for the original system). The errors for each equation are given by $AX - B$:

$$\begin{bmatrix} 2 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{11} \\ \frac{5}{11} \\ \frac{15}{11} \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{7}{11} \\ \frac{14}{11} \\ -\frac{28}{11} \\ -\frac{7}{11} \end{bmatrix}$$

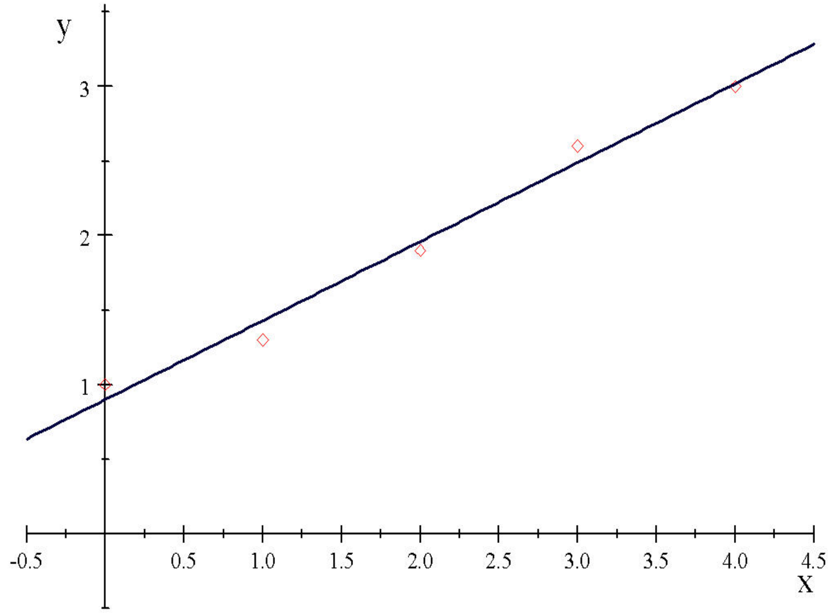

Figure 5.6: Least Squares Line

6. By Example 5.7.5, the regression line $y = a + bx$ for five data points is given by the least squares solution of $AX = B$:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 1.3 \\ 1.9 \\ 2.6 \\ 3 \end{bmatrix}$$

and the least squares solution is the solution of $A^T AX = A^T B$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1.3 \\ 1.9 \\ 2.6 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 9.8 \\ 24.9 \end{bmatrix}$$

with solution: $a = 0.9,\ b = 0.53$ and regression line $y = 0.9 + 0.53x$. To illustrate this solution, the graph of the data and regression line are shown in Figure 5.6.

Alternate solution: The solution can also be obtained using the formulae derived in Example 5.7.5, but these are complicated, and generally it is easier to develop the equations and solve the problem as shown above. The formulae for the least squares line $y = a + bx$ are:

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{m\sum x^2 - \left(\sum x\right)^2},\ b = \frac{m\sum xy - \sum x \sum y}{m\sum x^2 - \left(\sum x\right)^2}$$

Since $m = 5,\ \sum x = 10,\ \sum x^2 = 30,\ \sum y = 9.8,\ \sum xy = 24.9$ the formulae give the same values for a, b:

$$a = \frac{9.8 \times 30 - 10 \times 24.9}{5 \times 30 - (10)^2} = 0.9$$
$$b = \frac{5 \times 24.9 - 10 \times 9.8}{5 \times 30 - (10)^2} = 0.53$$

7. Extending the method of Example 5.7.5, the regression equation $z = a + bx + cy$: for five data points is given by the least squares solution of equations $AX = B$ formed by inserting the data in the form $a + bx + cy = z$

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \\ 1 & x_5 & y_5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1.3 \\ 1 & 2 & 1.9 \\ 1 & 3 & 2.6 \\ 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 1.9 \\ 3.1 \\ 4.2 \\ 4.8 \end{bmatrix}$$

and the least squares solution is the solution of $A^T AX = A^T B$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 1 & 1.3 & 1.9 & 2.6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1.3 \\ 1 & 2 & 1.9 \\ 1 & 3 & 2.6 \\ 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 1 & 1.3 & 1.9 & 2.6 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1.9 \\ 3.1 \\ 4.2 \\ 4.8 \end{bmatrix}$$
$$\begin{bmatrix} 5 & 10 & 9.8 \\ 10 & 30 & 24.9 \\ 9.8 & 24.9 & 22.06 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 15.0 \\ 39.9 \\ 34.68 \end{bmatrix}$$

with approximate solution: $a = 0.329,\ b = 0.583,\ c = 0.767$ and regression equation $z = 0.329 + 0.583x + 0.767y$.

8.

(a) By Example 5.7.6 the quadratic, $y = a + bx + cx^2$, that approximates the five data points $(x_i,\ y_i)$ in the least squares sense is given by the least squares approximation for the system $AX = B$:

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

with least squares approximation X given by the solution of $A^T AX = A^T B$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 4 & 9 & 16 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 4 & 9 & 16 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} 5 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 12 \end{bmatrix}$$

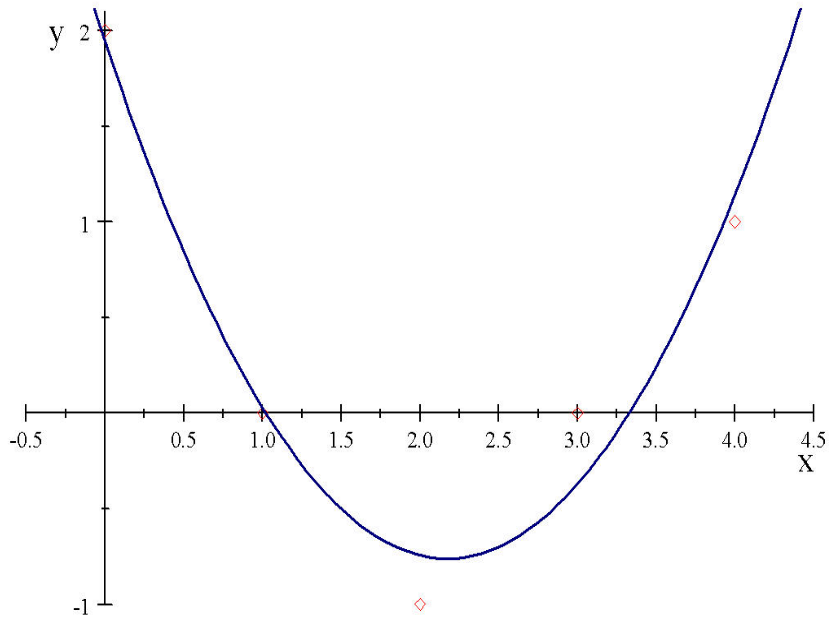

Figure 5.7: Least Squares Parabola

with solution $a = \frac{68}{35},\ b = -\frac{87}{35},\ c = \frac{4}{7}$ and approximating quadratic $y = a + bx + cx^2 = \frac{68}{35} - \frac{87}{35}x + \frac{4}{7}x^2$.
The sum of the squares of the errors is given by $\|AX - B\|^2$

$$\left\| \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} \frac{68}{35} \\ -\frac{87}{35} \\ \frac{4}{7} \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} \frac{68}{35} \\ \frac{1}{35} \\ -\frac{26}{35} \\ -\frac{13}{35} \\ \frac{8}{7} \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\|^2$$

$$= \left\| \begin{bmatrix} -\frac{2}{35} \\ \frac{1}{35} \\ \frac{9}{35} \\ -\frac{13}{35} \\ \frac{1}{7} \end{bmatrix} \right\|^2 = \frac{8}{35}$$

To illustrate this solution, the graph of the data and the approximating quadratic are shown in Figure 5.7.

(b) **Note:** Notice that the cubic has a significantly smaller error than the quadratic.

9.

(a) By Example 5.7.6 the exponential $y = ae^{bx}$ that approximates the five data points $(x_i,\ y_i)$ in the least squares sense is given by the least squares approximation for the system $AX = B$:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \end{bmatrix} \begin{bmatrix} \ln a \\ b \end{bmatrix} = \begin{bmatrix} \ln y_1 \\ \ln y_2 \\ \ln y_3 \\ \ln y_4 \\ \ln y_5 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & \frac{1}{2} \\ 1 & 1 \\ 1 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} \ln a \\ b \end{bmatrix} = \begin{bmatrix} \ln 0.5 \\ \ln 1 \\ \ln 2 \\ \ln 3 \\ \ln 5 \end{bmatrix} \simeq \begin{bmatrix} -0.693 \\ 0 \\ 0.693 \\ 1.098 \\ 1.609 \end{bmatrix}$$

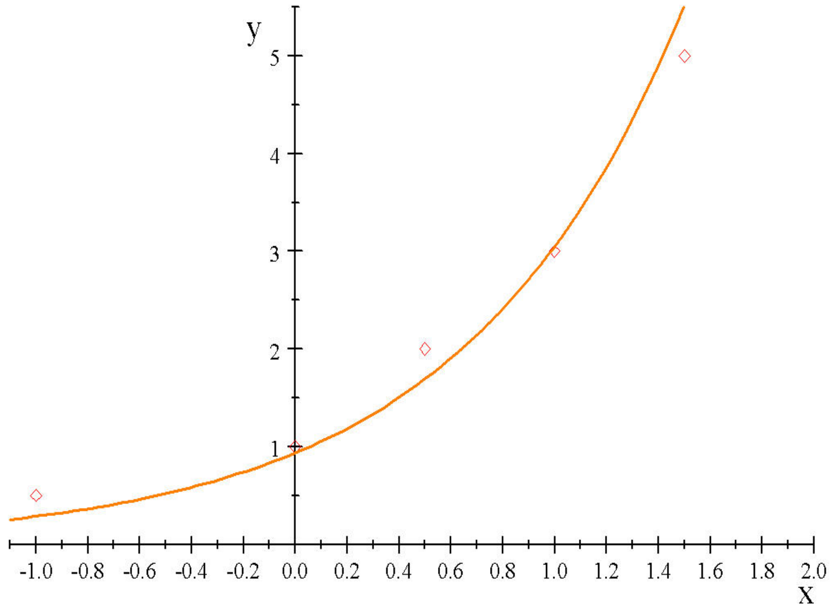

Figure 5.8: Least Squares Exponential

The least squares solution is given by the solution of $A^TAX = A^TB$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & \frac{1}{2} & 1 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & \frac{1}{2} \\ 1 & 1 \\ 1 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} \ln a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & \frac{1}{2} & 1 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} -0.693 \\ 0 \\ 0.693 \\ 1.098 \\ 1.609 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 2 \\ 2 & \frac{9}{2} \end{bmatrix} \begin{bmatrix} \ln a \\ b \end{bmatrix} = \begin{bmatrix} 2.707 \\ 4.551 \end{bmatrix}$$

with approximate solution $\ln a = 0.16646$, giving $a \simeq 1.181$ and $b = 0.937$. Hence, the approximating exponential is $y = 0.937e^{1.181x}$.
The errors in the $y-$ values are given by:

$$\begin{bmatrix} y_1 - 0.937e^{1.181x_1} \\ y_2 - 0.937e^{1.181x_2} \\ y_3 - 0.937e^{1.181x_3} \\ y_4 - 0.937e^{1.181x_4} \\ y_5 - 0.937e^{1.181x_5} \end{bmatrix} = \begin{bmatrix} 0.5 - 0.937e^{1.181\times(-1)} \\ 1 - 0.937e^{1.181\times 0} \\ 2 - 0.937e^{1.181\times 0.5} \\ 3 - 0.937e^{1.181\times 1} \\ 5 - 0.937e^{1.181\times 1.5} \end{bmatrix} \simeq \begin{bmatrix} 0.212 \\ 0.063 \\ 0.309 \\ -0.052 \\ -0.509 \end{bmatrix}$$

To illustrate this solution, the graph of the data and the approximating quadratic are shown in Figure 5.8.

(b) By Example 5.7.6 the quadratic $y = a + bx + cx^2$ that approximates the five data points $(x_i,\ y_i)$ in the least squares sense is given by the least squares approximation for the system $AX = B$:

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{4} \\ 1 & 1 & 1 \\ 1 & \frac{3}{2} & \frac{9}{4} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}$$

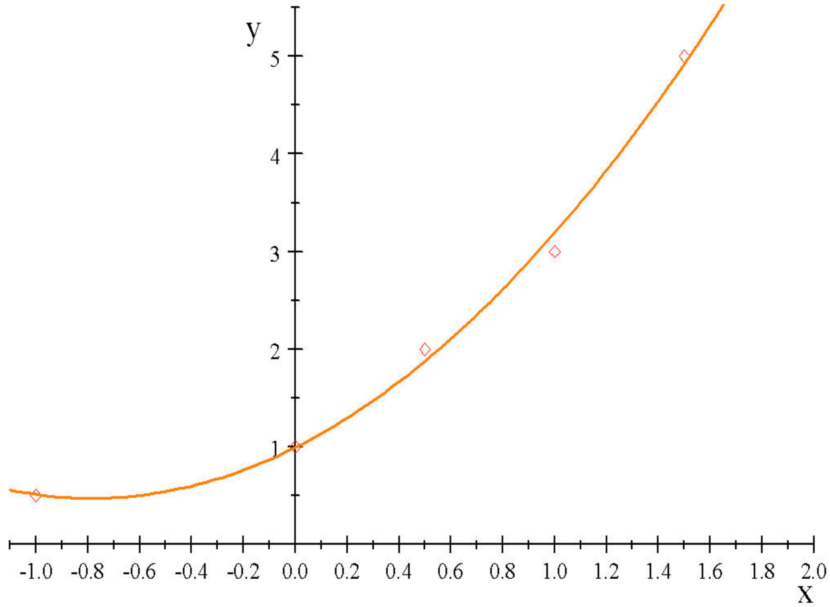

Figure 5.9: Least Squares Cubic

with least squares approximation X given by the solution of $A^TAX = A^TB$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & \frac{1}{2} & 1 & \frac{3}{2} \\ 1 & 0 & \frac{1}{4} & 1 & \frac{9}{4} \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{4} \\ 1 & 1 & 1 \\ 1 & \frac{3}{2} & \frac{9}{4} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & \frac{1}{2} & 1 & \frac{3}{2} \\ 1 & 0 & \frac{1}{4} & 1 & \frac{9}{4} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 2 & \frac{9}{2} \\ 2 & \frac{9}{2} & \frac{7}{2} \\ \frac{9}{2} & \frac{7}{2} & \frac{57}{8} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{23}{2} \\ 11 \\ \frac{61}{4} \end{bmatrix}$$

with approximate solution $a = 0.996,\ b = 1.337,\ c = 0.854$ and approximating quadratic $y = a + bx + cx^2 = 0.996 + 1.337x + 0.854x^2$.
The errors in the $y-$ values are:

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{4} \\ 1 & 1 & 1 \\ 1 & \frac{3}{2} & \frac{9}{4} \end{bmatrix} \begin{bmatrix} 0.996 \\ 1.337 \\ 0.854 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ 1 \\ 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 0.513 \\ 0.996 \\ 1.878 \\ 3.187 \\ 4.923 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ 1 \\ 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 0.013 \\ -0.004 \\ -0.122 \\ 0.187 \\ -0.077 \end{bmatrix}$$

To illustrate this solution, the graph of the data and the approximating quadratic are shown in Figure 5.9.

Note: The errors with a quadratic approximation are considerably smaller than for the exponential approximation, suggesting that the data is closer to a quadratic form.

10. By Theorem 5.18 the projection $p(x) = k_1 g_1(x) + k_2 g_2(x) + k_3 g_3(x) + k_4 g_4(x)$

$$\begin{bmatrix} \langle g_1, g_1\rangle & \langle g_1, g_2\rangle & \langle g_1, g_3\rangle & \langle g_1, g_4\rangle \\ \langle g_2, g_1\rangle & \langle g_2, g_2\rangle & \langle g_2, g_3\rangle & \langle g_2, g_4\rangle \\ \langle g_3, g_1\rangle & \langle g_3, g_2\rangle & \langle g_3, g_3\rangle & \langle g_3, g_4\rangle \\ \langle g_4, g_1\rangle & \langle g_4, g_2\rangle & \langle g_4, g_3\rangle & \langle g_4, g_4\rangle \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} = \begin{bmatrix} \langle g_1, f\rangle \\ \langle g_2, f\rangle \\ \langle g_3, f\rangle \\ \langle g4, f\rangle \end{bmatrix}$$

$$\begin{bmatrix} \int_{-1}^{1} 1dx & \int_{-1}^{1} xdx & \int_{-1}^{1} x^2dx & \int_{-1}^{1} x^3dx \\ \int_{-1}^{1} xdx & \int_{-1}^{1} x^2dx & \int_{-1}^{1} x^3dx & \int_{-1}^{1} x^4dx \\ \int_{-1}^{1} x^2dx & \int_{-1}^{1} x^3dx & \int_{-1}^{1} x^4dx & \int_{-1}^{1} x^5dx \\ \int_{-1}^{1} x^3dx & \int_{-1}^{1} x^4dx & \int_{-1}^{1} x^5dx & \int_{-1}^{1} x^6dx \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} = \begin{bmatrix} \int_{-1}^{1} e^x dx \\ \int_{-1}^{1} xe^x dx \\ \int_{-1}^{1} x^2e^x dx \\ \int_{-1}^{1} x^3e^x dx \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{2}{5} \\ \frac{2}{3} & 0 & \frac{2}{5} & 0 \\ 0 & \frac{2}{5} & 0 & \frac{2}{7} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} = \begin{bmatrix} 2\sinh 1 \\ 2e^{-1} \\ e - 5e^{-1} \\ 16e^{-1} - 2e \end{bmatrix}$$

with approximate solution: $k_1 = 0.996,\ k_2 = 0.998,\ k_3 = 0.537,\ k_4 = 0.176$ and approximating polynomial:

$$y = 0.996 + 0.998x + 0.537x^2 + 0.176x^3$$

Note: The first four terms of the Taylor Series approximation to e^x are:

$$\begin{aligned} e^x &\simeq 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 \\ &\simeq 1 + x + 0.5x^2 + 0.167\,x^3 \end{aligned}$$

These values are reasonably close to those found by the least squares method.

11. Using the formulae for the means, $\overline{x} = \frac{1}{m}\sum x^2, \overline{y} = \frac{1}{m}\sum y^2$ the simplified formula, $b = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}$, numerator and denomenator are computed separately:

$$\begin{aligned} \sum(x-\overline{x})(y-\overline{y}) &= \sum xy - \overline{x}\sum y - \overline{y}\sum x + \overline{xy}\sum 1 \\ &= \sum xy - \frac{1}{m}\sum x\sum y - \frac{1}{m}\sum y\sum x + \frac{1}{m}\sum x \times \frac{1}{m}\sum y \times m \\ \sum(x-\overline{x})(y-\overline{y}) &= \frac{1}{m}\left(m\sum xy - \sum y\sum x\right) \end{aligned}$$

$$\begin{aligned} \sum(x-\overline{x})^2 &= \sum x^2 - 2\overline{x}\sum x + \overline{x}^2\sum 1 \\ &= \sum x^2 - 2\frac{1}{m}\sum x \times \sum x + \left(\frac{1}{m}\sum x\right)^2 \times m \\ \sum(x-\overline{x})^2 &= \frac{1}{m}\left(m\sum x^2 - \left(\sum x\right)^2\right) \end{aligned}$$

Hence the result follows:

$$b = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2} = \frac{m\sum xy - \sum y\sum x}{m\sum x^2 - \left(\sum x\right)^2}$$ - the formula found in Exercise 5.7.5.

Starting with the simplified formula for a then:

$$\begin{aligned} \overline{y} - b\overline{x} &= \overline{y} - \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}\overline{x} \\ &= \frac{\overline{y}\sum(x-\overline{x})^2 - \overline{x}\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2} \end{aligned}$$

Computing the numerator separately:

$$= \overline{y}\sum x^2 - 2\overline{yx}\sum x + \overline{yx}^2\sum 1 - \overline{x}\sum xy + \overline{x}^2\sum y + \overline{xy}\sum x - \overline{x}^2\overline{y}\sum 1$$

Moving the fourth term of the numerator up to the second position:

$$\begin{aligned} &= \frac{1}{m}\sum y\sum x^2 - \frac{1}{m}\sum x\sum xy - 2m\overline{yx} \times \overline{x} + m\overline{yx}^2 + m\overline{x}^2\overline{y} + m\overline{xyx} - m\overline{x}^2\overline{y} \\ &= \frac{1}{m}\sum y\sum x^2 - \sum x\sum xy \end{aligned}$$

Putting this back with the denomenator and using the formula for the denomenator derived above:

$$a = \frac{\frac{1}{m}\sum y\sum x^2 - \sum x\sum xy}{\sum\left(x - \overline{x}\right)^2}$$

$$a = \frac{\sum y\sum x^2 - \sum x\sum xy}{m\sum x^2 - \left(\sum x\right)^2}$$ - the formula found in Exercise 5.7.5