
Week 13

June 5, 2026

Contents

1 Least-Squares Approximation

Often an exact solution to a problem in applied mathematics is difficult or impossible to obtain. However, it is usually just as useful to find an approximation to a solution. In particular, finding “linear approximations” is a powerful technique in applied mathematics. One basic case is the situation where a system of linear equations has no solution, and it is desirable to find a “best approximation” to a solution to the system.

We begin by defining the “best approximation” in a natural way, and showing that computing the best approximation reduces to solving a related system of linear equations called the *normal equations*. Next, we demonstrate a common application where a collection of data points is approximated by a line (or a curve). We conclude this section by showing that *QR*-factorization provides us with a more efficient way to solve the normal equations and compute the best approximation.

Best Approximate Solutions

Exploration 1 Let

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Consider the matrix equation $A\mathbf{x} = \mathbf{b}$. A quick examination of the last two rows should convince you that this equation has no solutions. In other words, $\mathbf{b}$ is not in the span of the columns of A .

If $\mathbf{z}$ were an exact solution to $A\mathbf{x} = \mathbf{b}$, then $\mathbf{b} - A\mathbf{z}$ would be $\mathbf{0}$. Since the equation does not have a solution, we will attempt to find the next best thing to a solution by finding $\mathbf{z}$ such that $\|\mathbf{b} - A\mathbf{z}\|$ is as small as possible. The quantity $\|\mathbf{b} - A\mathbf{z}\|$ is called the error.

The following GeoGebra interactive will help you understand the geometry behind finding $\mathbf{z}$.

RIGHT-CLICK and DRAG to rotate the image for a better view.

Access GeoGebra interactives through the online version of this text at

<https://ximera.osu.edu/linearalgebradv3/LinearAlgebraInteractiveIntro>.

Record your best guess for $\mathbf{z}$ – you will have a chance to check your answer in Example ??.

What did you discover about the geometry of minimizing $\|\mathbf{b} - A\mathbf{z}\|$? Select all that apply.

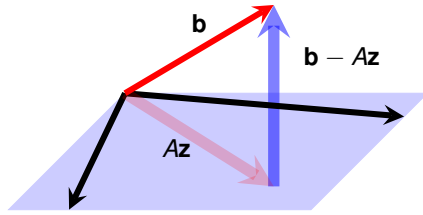
Select All Correct Answers:

1. $\mathbf{z}$ is orthogonal to the plane spanned by the columns of A .
2. $\|\mathbf{b} - A\mathbf{z}\|$ is orthogonal to $\text{col}(A)$.
3. $\mathbf{b} - A\mathbf{z}$ is orthogonal to $\text{col}(A)$.
4. $A\mathbf{z}$ is orthogonal to $\text{col}(A)$.
5. $A\mathbf{z}$ is an orthogonal projection of $\mathbf{b}$ onto $\text{col}(A)$.

Our geometric observations will help us develop a method for finding $\mathbf{z}$.

Suppose A is an $m \times n$ matrix, and $\mathbf{b}$ is a column vector in $\mathbb{R}^m$. Consider the matrix equation $A\mathbf{x} = \mathbf{b}$. If this equation does not have a solution, we can attempt to find a best approximation by finding $\mathbf{z}$ which minimizes the error, $\|\mathbf{b} - A\mathbf{z}\|$. The expression $\|\mathbf{b} - A\mathbf{z}\|$ is also sometimes called the *residual*. The error (or the residual) is given in terms of a vector norm. Recall that our definition of the norm involves the sum of squares of the vector components. When we minimize the norm, we minimize the sum of squares. This is why the method we are describing is often referred to as *least squares*. We will explore this idea further later in this section.

In the case when $\text{col}(A)$ is a subspace of $\mathbb{R}^3$, we can see geometrically that $\mathbf{z}$ is the best approximation if and only if $A\mathbf{z}$ is an orthogonal projection of $\mathbf{b}$ onto $\text{col}(A)$, and the error is the magnitude of $\mathbf{b} - A\mathbf{z}$, as shown below.



What we observed above, holds in general. We will use this fact to find $\mathbf{z}$.

Every vector in $\text{col}(A)$ can be written in the form $A\mathbf{x}$ for some $\mathbf{x}$ in $\mathbb{R}^n$. Our goal is to find $\mathbf{z}$ such that $A\mathbf{z}$ is the orthogonal projection of $\mathbf{b}$ onto $\text{col}(A)$. By Corollary ??, every vector $A\mathbf{x}$ in $\text{col}(A)$ is orthogonal to $\mathbf{b} - A\mathbf{z}$.

Therefore, we have

$$\begin{aligned} A^T(\mathbf{b} - A\mathbf{z}) &= \mathbf{0} \\ A^T\mathbf{b} - A^TA\mathbf{z} &= \mathbf{0} \\ A^TA\mathbf{z} &= A^T\mathbf{b} \end{aligned} \tag{1}$$

Since $\mathbf{b} - A\mathbf{z}$ is normal to the subspace $\text{col}(A)$, we call the system of linear equations in (??) the *normal equations for $\mathbf{z}$* . If A^TA is invertible, then we can write

$$\mathbf{z} = (A^TA)^{-1}A^T\mathbf{b} \tag{2}$$

We will return to the question of invertibility of A^TA in Theorem ??. For now, let's revisit the problem posed in Exploration ??.

Example 1. We now return to the matrix equation $A\mathbf{x} = \mathbf{b}$ of Exploration ?? to find $\mathbf{z}$ that best approximates a solution.

Explanation. Recall that

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

In this case, $(A^T A)^{-1}$ exists. Applying equation (??), we compute

$$\begin{aligned} \mathbf{z} &= \left(\begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 11 & 7 \\ 7 & 9 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 10 \end{bmatrix} = \begin{bmatrix} 0.18 & -0.14 \\ -0.14 & 0.22 \end{bmatrix} \begin{bmatrix} 10 \\ 10 \end{bmatrix} \\ &= \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix} \end{aligned}$$

Compare this answer to your guess in Exploration ???. If your guess was correct, nice job! If your guess was different, try setting $\mathbf{z}$ to the correct answer and use the GeoGebra interactive in Exploration ?? to examine the geometry of the problem.

We now come back to the question of when $A^T A$ is invertible.

Theorem 1. If columns of matrix A are linearly independent, then $A^T A$ is invertible.

Proof Let A be a matrix with linearly independent columns. We will show that $(A^T A) \mathbf{x} = \mathbf{0}$ has only the trivial solution. For $\mathbf{x}$, a solution of $A^T A \mathbf{x} = \mathbf{0}$, we have

$$\begin{aligned} \|\mathbf{Ax}\|^2 &= (\mathbf{Ax}) \cdot (\mathbf{Ax}) \\ &= (\mathbf{Ax})^T \mathbf{Ax} \\ &= \mathbf{x}^T A^T \mathbf{Ax} \\ &= \mathbf{x}^T \cdot \mathbf{0} = 0 \end{aligned}$$

Therefore $\mathbf{Ax} = \mathbf{0}$. By linear independence of the columns of A we conclude that $\mathbf{x} = \mathbf{0}$. ■

We summarize our findings in the following theorem.

Theorem 2. Let A be an $m \times n$ matrix, let $\mathbf{b}$ be a column vector in $\mathbb{R}^m$. Consider the matrix equation

$$\mathbf{Ax} = \mathbf{b}$$

- Any solution $\mathbf{z}$ to the normal equations

$$(A^T A) \mathbf{z} = A^T \mathbf{b}$$

is a best approximation to a solution to $\mathbf{Ax} = \mathbf{b}$ in the sense that $\|\mathbf{b} - \mathbf{Az}\|$ is minimized.

- If the columns of A are linearly independent, then $A^T A$ is invertible and $\mathbf{z}$ is given uniquely by

$$\mathbf{z} = (A^T A)^{-1} A^T \mathbf{b}$$

Example 2. The system of linear equations

$$\begin{aligned} 3x - y &= 4 \\ x + 2y &= 0 \\ 2x + y &= 1 \end{aligned}$$

has no solution. Find the vector $\mathbf{z}$ that best approximates a solution.

Explanation. We have

$$A = \begin{bmatrix} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}.$$

The normal equations are $(A^T A) \mathbf{z} = A^T \mathbf{b}$. We compute

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

$$\begin{bmatrix} 14 & 1 \\ 1 & 6 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 14 \\ -3 \end{bmatrix}$$

We observe that $A^T A$ is invertible. Multiplying on the left by $(A^T A)^{-1}$ yields

$$\mathbf{z} = \frac{1}{83} \begin{bmatrix} 87 \\ -56 \end{bmatrix}$$

With this vector $\mathbf{z}$, the left sides of the equations become

$$\begin{aligned} 3(87/83) - (-56/83) &\approx 3.82 \\ (87/83) + 2(-56/83) &\approx -0.30 \\ 2(87/83) + (-56/83) &\approx 1.42 \end{aligned}$$

This is as close as possible to a solution.

Example 3. The average number g of goals per game scored by a hockey player seems to be related linearly to two factors: the number x_1 of years of experience and the number x_2 of goals in the preceding 10 games. The data on the following page were collected on four players. Find the linear function $g = a_0 + a_1 x_1 + a_2 x_2$ that best fits the data.

g	x_1	x_2
0.8	5	3
0.8	3	4
0.6	1	5
0.4	2	1

Explanation. If the relationship is given by $g = a_0 + a_1 x_1 + a_2 x_2$, then the data can be described as follows:

$$\begin{bmatrix} 1 & 5 & 3 \\ 1 & 3 & 4 \\ 1 & 1 & 5 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix}$$

Using Theorem ??, we get

$$\mathbf{z} = \frac{1}{42} \underbrace{\begin{bmatrix} 119 & -17 & -19 \\ -17 & 5 & 1 \\ -19 & 1 & 5 \end{bmatrix}}_{(A^T A)^{-1}} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 3 & 1 & 2 \\ 3 & 4 & 5 & 1 \end{bmatrix}}_{A^T} \begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.14 \\ 0.09 \\ 0.08 \end{bmatrix}$$

Hence the best-fitting function is $g = 0.14 + 0.09x_1 + 0.08x_2$.

Application of Least Squares to Curve Fitting

In [Curve Fitting](#), we discussed how to fit a function to a set of data points, so that the graph of the function passes through each of the points. We also discussed why doing so is sometimes impossible (two points lie one above the other), and may not even be desirable (overfitting). In this section we will learn how to approximate a collection of data points with a line (or a curve) that fits the “trend” of the points. We will start with data that fit a linear pattern.

Exploration 2 Consider the points $(1, 1)$, $(2, 3)$ and $(4, 4)$. These points do not lie on a straight line, but they have a general upward linear trend. (Typically there would be many more points to consider, but we will limit our exploration to what we can do by hand.) Our goal is to find a line that fits these points as closely as possible.

We are looking for a function f of the form $f(x) = ax + b$ such that the following infeasible system is satisfied as closely as possible

$$\begin{aligned} a(1) + b &= 1 \\ a(2) + b &= 3 \\ a(4) + b &= 4 \end{aligned}$$

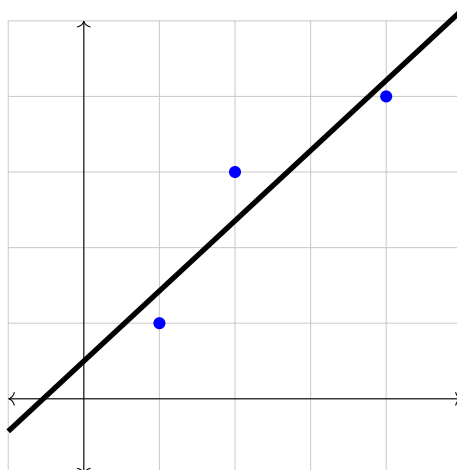
From the first part of this section we know how to find a best approximation. By Theorem ??, we have

$$\begin{aligned} \mathbf{z} = \begin{bmatrix} a \\ b \end{bmatrix} &= (A^T A)^{-1} A^T \mathbf{b} \\ &= \left(\begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 13/14 \\ 1/2 \end{bmatrix} \end{aligned}$$

According to our computations, the line that best fits the data is given by

$$f(x) = \frac{13}{14}x + \frac{1}{2}$$

Let's take a look.



We found this fit by minimizing $\|\mathbf{b} - \mathbf{Az}\|$. We will now investigate the meaning of this expression in relation to the line and the data points.

$$\mathbf{b} - \mathbf{Az} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} (13/14)(1) + 1/2 \\ (13/14)(2) + 1/2 \\ (13/14)(4) + 1/2 \end{bmatrix} \approx \begin{bmatrix} -0.43 \\ 0.64 \\ -0.21 \end{bmatrix} \quad (3)$$

Observe that each entry of $\mathbf{b} - \mathbf{Az}$ is the signed vertical distance between a particular point and the line.

Instead of computing the error, $\|\mathbf{b} - \mathbf{Az}\|$, we will compute $\|\mathbf{b} - \mathbf{Az}\|^2$ to avoid the square root.

$$\|\mathbf{b} - \mathbf{Az}\|^2 = (-0.43)^2 + 0.64^2 + (-0.21)^2 \approx 0.64 \quad (4)$$

Minimizing $\|\mathbf{b} - \mathbf{Az}\|$ also minimizes $\|\mathbf{b} - \mathbf{Az}\|^2$. Therefore, what we have minimized is the sum of squares of the vertical distances between the data points and the line. The following GeoGebra interactive will help you explore this idea.

Access GeoGebra interactives through the online version of this text at

<https://ximera.osu.edu/linearalgebradv3/LinearAlgebraInteractiveIntro>.

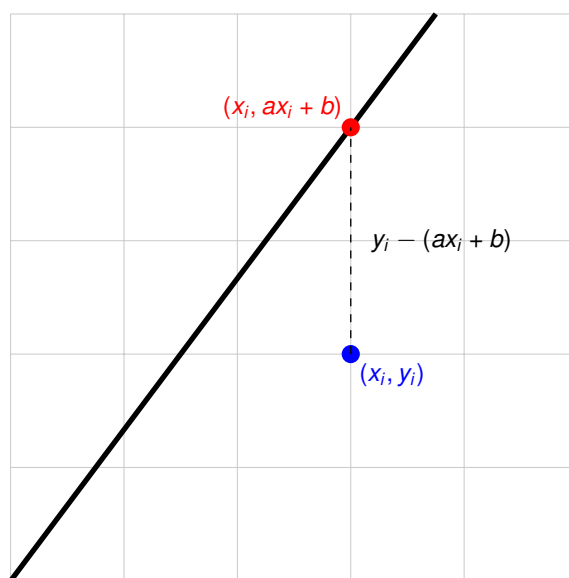
In Exploration ?? we discovered that $\|\mathbf{b} - \mathbf{Az}\|^2$ is the sum of squares of vertical distances between the given data points and the proposed line. By minimizing $\|\mathbf{b} - \mathbf{Az}\|$, we minimize the sum of squares of vertical distances. This observation holds in general. Given a collection of points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, finding a linear function of the form $f(x) = ax + b$ that best fits the points we would find a best solution to the system

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

by minimizing

$$\left\| \begin{bmatrix} y_1 - (ax_1 + b) \\ y_2 - (ax_2 + b) \\ \vdots \\ y_n - (ax_n + b) \end{bmatrix} \right\|^2 = (y_1 - (ax_1 + b))^2 + (y_2 - (ax_2 + b))^2 + \cdots + (y_n - (ax_n + b))^2$$

A geometric interpretation of $y_i - (ax_i + b)$ is shown below.



The line we obtain in this fashion is called a *line of best fit* or a *trendline*, and the method we used is referred to as the *method of least squares*.

We can apply the method of least squares to find best fitting non-linear functions.

Example 4. Find the least squares approximating quadratic polynomial of the form $f(x) = ax^2 + bx + c$ for the following points.

$$(-3, 3), (-1, 1), (0, 1), (1, 2), (3, 4)$$

Explanation. We are looking for an approximate solution to the system of equations

$$\begin{aligned} a(-3)^2 + b(-3) + c &= 3 \\ a(-1)^2 + b(-1) + c &= 1 \\ a(0)^2 + b(0) + c &= 1 \\ a(1)^2 + b(1) + c &= 2 \\ a(3)^2 + b(3) + c &= 4 \end{aligned}$$

This corresponds to the matrix equation

$$\begin{bmatrix} 9 & -3 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 9 & 3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$

Multiplying on the left by A^T gives us the normal equations.

$$A^T A \mathbf{z} = A^T \mathbf{b}$$

$$\begin{bmatrix} 164 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 5 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 66 \\ 11 \\ 4 \end{bmatrix}$$

It turns out that $A^T A$ is invertible, so it is easy to solve for $\mathbf{z}$. You can use technology to accomplish this.

We arrive at the solution

$$\mathbf{z} = \begin{bmatrix} 0.26 \\ 0.20 \\ 1.15 \end{bmatrix}$$

Therefore, the quadratic function of best fit is given by $f(x) = 0.26x^2 + 0.2x + 1.15$. You can see the graph and the points shown below.

For more information about doing least squares with Octave, please see [Octave for Chapter 9](#).

Example 5. Given the data points $(-1, 0)$, $(0, 1)$, and $(1, 4)$, find the least squares approximating function of the form $f(x) = ax + b2^x$.

Explanation. We are looking for an approximate solution to the system of equations

$$\begin{aligned} a(-1) + b(2^{-1}) &= 0 \\ a(0) + b(2^0) &= 1 \\ a(1) + b(2^1) &= 4 \end{aligned}$$

This corresponds to the matrix equation

$$\begin{bmatrix} -1 & 1/2 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$$

Using the normal equations, we obtain

$$A^T A \mathbf{z} = A^T \mathbf{b}$$

$$\frac{1}{4} \begin{bmatrix} 8 & 6 \\ 6 & 21 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

Solving for $\mathbf{z}$ yields

$$\mathbf{z} = \begin{bmatrix} 10/11 \\ 16/11 \end{bmatrix}$$

Therefore, the function of best fit (of the given form) is given by

$$f(x) = \frac{10}{11}x + \frac{16}{11}(2^x)$$

Practice Problems

Problem 3 Find the best approximation to a solution to the system of equations.

$$\begin{array}{rrcrcl} 3x & + & y & + & z & = & 6 \\ 2x & + & 3y & - & z & = & 1 \\ 2x & - & y & + & z & = & 0 \\ 3x & - & 3y & + & 3z & = & 8 \end{array}$$

Enter answers in fraction form below.

$$x = \dots\dots, \quad y = \dots\dots, \quad z = \dots\dots$$

Problem 4 Find a linear function of best fit for the given set of data points. Examine how well your line fits the points by typing the equation of the line into the Desmos window.

$$(2, 4), (4, 3), (7, 2), (8, 1)$$

Enter your answers in fraction form.

$$f(x) = \dots\dots x + \dots\dots$$

Problem 5 Find a linear function of best fit for the given set of data points. Examine how well your line fits the points by typing the equation of the line into the Desmos window.

$$(-2, 3), (-1, 1), (0, 0), (1, -2), (2, -4)$$

$$f(x) = \dots\dots x + \dots\dots$$

Problem 6 Use Octave to find the least squares approximating quadratic function for the following data points.

$$(-2, 1), (0, 0), (3, 2), (4, 3)$$

Round your answers to three decimal places.

$$f(x) = \dots\dots x^2 + \dots\dots x + \dots\dots$$

For more information about doing least squares with Octave, please see [Octave for Chapter 9](#).

Text Source

A portion of this section has been adapted from W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2021-A, Open Edition, pp. 308-319.

2 Abstract Vector Spaces

Earlier in the text we discussed $\mathbb{R}^n$ as a vector space and introduced the notion of a subspace of $\mathbb{R}^n$. In this section we will consider sets other than $\mathbb{R}^n$ that have two operations and satisfy the same properties as $\mathbb{R}^n$. Such sets, together with the operations of addition and scalar multiplication, will also be called vector spaces.

Properties of Vector Spaces

Recall that $\mathbb{R}^n$ is said to be a vector space because

$\mathbb{R}^n$ is closed under vector addition

$\mathbb{R}^n$ is closed under scalar multiplication

and satisfies the following properties:

1. Commutative Property of Addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. Associative Property of Addition: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. Existence of Additive Identity: $\mathbf{u} + \mathbf{0} = \mathbf{u}$
4. Existence of Additive Inverse: $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
5. Distributive Property over Vector Addition: $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
6. Distributive Property over Scalar Addition: $(k + p)\mathbf{u} = k\mathbf{u} + p\mathbf{u}$
7. Associative Property for Scalar Multiplication: $k(p\mathbf{u}) = (kp)\mathbf{u}$
8. Multiplication by 1: $1\mathbf{u} = \mathbf{u}$

In the next two examples we will explore two sets other than $\mathbb{R}^n$ endowed with addition and scalar multiplication and satisfying the same properties.

Example 6. Let $\mathbb{M}_{m,n}$ be the set of all $m \times n$ matrices. Matrix addition and scalar multiplication were defined in [Addition and Scalar Multiplication of Matrices](#).

Observe that the sum of two $m \times n$ matrices is also an $m \times n$ matrix. Likewise, a scalar multiple of an $m \times n$ matrix is an $m \times n$ matrix. Thus

$\mathbb{M}_{m,n}$ is closed under matrix addition

$\mathbb{M}_{m,n}$ is closed under scalar multiplication

In addition, Theorems ?? and ?? give us the following properties of matrix addition and scalar multiplication. Note that these properties are analogous to the eight vector properties above.

1. Commutative Property of Addition: $A + B = B + A$

2. Associative Property of Addition: $(A + B) + C = A + (B + C)$
3. Existence of Additive Identity: $A + O = A$ where O is the $m \times n$ zero matrix
4. Existence of Additive Inverse: $A + (-A) = O$
5. Distributive Property over Matrix Addition: $k(A + B) = kA + kB$
6. Distributive Property over Scalar Addition: $(k + p)A = kA + pA$
7. Associative Property for Scalar Multiplication: $k(pA) = (kp)A$
8. Multiplication by 1: $1A = A$

Example 7. Consider the set $\mathbb{L}$ of all linear functions. This set includes all polynomials of degree 1 and degree 0. We will use addition and scalar multiplication of polynomials as the two operations, and show that $\mathbb{L}$ is closed under those operations and satisfies eight properties analogous to those of vectors of $\mathbb{R}^n$.

Explanation. Elements of $\mathbb{L}$ are functions f given by

$$f(x) = mx + b$$

(Note that m and b can be equal to zero.)

Given f_1 and f_2 in $\mathbb{L}$, it is easy to verify that $f_1 + f_2$ is also in $\mathbb{L}$. This gives us closure under function addition.

For any scalar k , we have

$$kf(x) = k(mx + b) = (km)x + (kb)$$

Therefore kf is in $\mathbb{L}$, and $\mathbb{L}$ is closed under scalar multiplication.

We now proceed to formulate eight properties analogous to those of vectors of $\mathbb{R}^n$.

Let f_1 , f_2 and f_3 be elements of $\mathbb{L}$ given by $f_1(x) = m_1x + b_1$, $f_2(x) = m_2x + b_2$, and $f_3(x) = m_3x + b_3$. Let k and p be scalars.

1. Commutative Property of Addition: $f_1 + f_2 = f_2 + f_1$

This property holds because

$$\begin{aligned} f_1(x) + f_2(x) &= (m_1x + b_1) + (m_2x + b_2) \\ &= (m_2x + b_2) + (m_1x + b_1) \\ &= f_2(x) + f_1(x) \end{aligned}$$

2. Associative Property of Addition: $(f_1 + f_2) + f_3 = f_1 + (f_2 + f_3)$

This property is easy to verify.

3. Existence of Additive Identity: $f_1 + f_0 = f_1$

The additive identity f_0 is given by $f_0(x) = 0$. Note that f_0 is in $\mathbb{L}$.

4. Existence of Additive Inverse: $f_1 + (-f_1) = f_0$

The additive inverse of f_1 is a function $-f_1$ given by $-f_1(x) = -mx + (-b)$. Note that $-f_1$ is in $\mathbb{L}$.

5. Distributive Property over Vector Addition: $k(f_1 + f_2) = kf_1 + kf_2$

This property holds because

$$\begin{aligned} k(f_1(x) + f_2(x)) &= k((m_1x + b_1) + (m_2x + b_2)) \\ &= k(m_1x + b_1) + k(m_2x + b_2) \\ &= kf_1(x) + kf_2(x) \end{aligned}$$

6. Distributive Property over Scalar Addition: $(k + p)f_1 = kf_1 + pf_1$

This property holds because

$$\begin{aligned} (k + p)f_1(x) &= (k + p)(m_1x + b_1) \\ &= k(m_1x + b_1) + p(m_1x + b_1) \\ &= kf_1(x) + pf_1(x) \end{aligned}$$

7. Associative Property for Scalar Multiplication: $(k(pf_1)) = (kp)f_1$

This property holds because

$$\begin{aligned} k(p(f_1(x))) &= k(p(m_1x + b_1)) \\ &= k(pm_1x + pb_1) \\ &= kpm_1x + kpb_1 \\ &= (kp)m_1x + (kp)b_1 \\ &= (kp)(m_1x + b_1) \\ &= (kp)f_1(x) \end{aligned}$$

8. Multiplication by 1

$$1f_1 = f_1$$

Definition of a Vector Space

Examples ?? and ?? show us that there are many times in mathematics when we encounter a set with two operations (that we call addition and scalar multiplication) such that the set is closed under the two operations, and satisfies the same eight properties as $\mathbb{R}^n$. We will refer to such sets as *vector spaces*.

Definition 1. Let V be a nonempty set. Suppose that elements of V can be added together and multiplied by scalars. The set V , together with operations of addition and scalar multiplication, is called a *vector space* provided that

V is closed under addition

V is closed under scalar multiplication

and the following properties hold for $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in V and scalars k and p :

1. Commutative Property of Addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. Associative Property of Addition: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

3. Existence of Additive Identity: $\mathbf{u} + \mathbf{0} = \mathbf{u}$
4. Existence of Additive Inverse: $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
5. Distributive Property over Vector Addition: $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
6. Distributive Property over Scalar Addition: $(k + p)\mathbf{u} = k\mathbf{u} + p\mathbf{u}$
7. Associative Property for Scalar Multiplication: $k(p\mathbf{u}) = (kp)\mathbf{u}$
8. Multiplication by 1: $1\mathbf{u} = \mathbf{u}$

We will refer to elements of V as *vectors*.

When scalars k and p in the above definition are restricted to real numbers, as they are in this chapter, vector space V may be referred to as a *vector space over the real numbers*.

Example 8. $\mathbb{M}_{m,n}$ and $\mathbb{L}$ are vector spaces. (See Examples ?? and ??)

Sets of polynomials provide an important source of examples, so we review some basic facts. A *polynomial with real coefficients* in x is an expression

$$p(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

where $a_0, a_1, a_2, \dots, a_n$ are real numbers called the *coefficients* of the polynomial. If all the coefficients are zero, the polynomial is called the *zero polynomial* and is denoted simply as 0. If $p(x) \neq 0$, the highest power of x with a nonzero coefficient is called the *degree* of $p(x)$ denoted as $\deg(p(x))$. The coefficient itself is called the *leading coefficient* of $p(x)$. Hence $\deg(3 + 5x) = 1$, $\deg(1 + x + x^2) = 2$, and $\deg(4) = 0$. (The degree of the zero polynomial is not defined.)

Let $\mathbb{P}$ denote the set of all polynomials and suppose that

$$p(x) = a_0 + a_1x + a_2x^2 + \dots$$

$$q(x) = b_0 + b_1x + b_2x^2 + \dots$$

are two polynomials in $\mathbb{P}$ (possibly of different degrees). Then $p(x)$ and $q(x)$ are called *equal* (written $p(x) = q(x)$) if and only if all the corresponding coefficients are equal—that is, $a_0 = b_0$, $a_1 = b_1$, $a_2 = b_2$, and so on. In particular, $a_0 + a_1x + a_2x^2 + \dots = 0$ means that $a_0 = 0$, $a_1 = 0$, $a_2 = 0$, ...

The set $\mathbb{P}$ has an addition and scalar multiplication defined on it as follows: if $p(x)$ and $q(x)$ are as before and k is a real number,

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \dots$$

$$kp(x) = ka_0 + (ka_1)x + (ka_2)x^2 + \dots$$

Example 9. $\mathbb{P}$ is a vector space.

Explanation. It is easy to see that the sum of two polynomials is again a polynomial, and that a scalar multiple of a polynomial is a polynomial. Thus, $\mathbb{P}$ is closed under addition and scalar multiplication. The other eight vector space properties are easily verified, and we conclude that $\mathbb{P}$ is a vector space.

Example 10. Let Y be the set of all degree two polynomials in x . In other words,

$$Y = \{ax^2 + bx + c : a \neq 0\}$$

We claim that Y is not a vector space.

Explanation. Observe that Y is not closed under addition. To see this, let $y_1 = 2x^2 + 3x + 4$ and let $y_2 = -2x^2$. Then y_1 and y_2 are both elements of Y . However, $y_1 + y_2 = 3x + 4$ is not an element of Y , as it is only a degree one polynomial. We require the coefficient a of x^2 to be nonzero for a polynomial to be in Y , and this is not the case for $y_1 + y_2$.

As an exercise, check the remaining vector space properties one-by-one to see which properties hold and which do not.

Set Y in Example ?? is not a vector space, but if we make a slight modification, we can make it into a vector space.

Example 11. Let $\mathbb{P}^2$ be the set of polynomials of degree two or less. In other words,

$$\mathbb{P}^2 = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$$

Note that $\mathbb{P}^2$ contains the zero polynomial (let $a = b = c = 0$).

Unlike set Y in Example ??, $\mathbb{P}^2$ is closed under polynomial addition and scalar multiplication. It is easy to verify that all vector space properties hold, so $\mathbb{P}^2$ is a vector space.

Example 12. Let n be a natural number. Define $\mathbb{P}^n$ to be the set of polynomials of degree n or less than n , then by reasoning similar to Example ??, $\mathbb{P}^n$ is a vector space.

Subspaces

Definition 2. A nonempty subset U of a vector space V is called a *subspace* of V , provided that U is itself a vector space when given the same addition and scalar multiplication as V .

Example 13. In Example ?? we demonstrated that $\mathbb{P}^2$ is a vector space. From Example ?? we know that $\mathbb{P}$ is a vector space. But $\mathbb{P}^2$ is a subset of $\mathbb{P}$, and uses the same operations of polynomial addition and scalar multiplication. Therefore $\mathbb{P}^2$ is a subspace of $\mathbb{P}$.

Checking all ten properties to verify that a subset of a vector space is a subspace can be cumbersome. Fortunately we have the following theorem.

Theorem 3 (Subspace Test). Let U be a nonempty subset of a vector space V . If U is closed under the operations of addition and scalar multiplication of V , then U is a subspace of V .

Proof To prove that closure is a sufficient condition for U to be a subspace, we will need to show that closure under addition and scalar multiplication of V guarantees that the remaining eight properties are satisfied automatically.

Observe that Properties ??, ??, ??, ??, ?? and ?? hold for all elements of V . Thus, these properties will hold for all elements of U . We say that these properties are *inherited* from V .

To prove Property ?? we need to show that $\mathbf{0}$, which we know to be an element of V , is contained in U . Let $\mathbf{u}$ be an element of U (recall that U is nonempty). We will show that $0\mathbf{u} = \mathbf{0}$ in V . Then, by closure under scalar multiplication, we will be able to conclude that $0\mathbf{u} = \mathbf{0}$ must be in U .

$$0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u}$$

Adding the additive inverse of $0\mathbf{u}$ to both sides gives us

$$0\mathbf{u} + (-0\mathbf{u}) = (0\mathbf{u} + 0\mathbf{u}) + (-0\mathbf{u})$$

By Properties ?? and ??

$$\mathbf{0} = \mathbf{0}\mathbf{u} + (\mathbf{0}\mathbf{u} + (-\mathbf{0}\mathbf{u}))$$

By Properties ?? and ??

$$\mathbf{0} = \mathbf{0}\mathbf{u} + \mathbf{0} = \mathbf{0}\mathbf{u}$$

Because U is closed under scalar multiplication $\mathbf{0}\mathbf{u} = \mathbf{0}$ is in U .

We know that every element of U , being an element of V , has an additive inverse in V . We need to show that the additive inverse of every element of U is contained in U . Let $\mathbf{u}$ be any element of U . We will show that $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$. Then by closure, $(-1)\mathbf{u}$ will have to be contained in U . To show that $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$, we must show that $\mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$. We compute:

$$\mathbf{u} + (-1)\mathbf{u} = 1\mathbf{u} + (-1)\mathbf{u} = (1 + (-1))\mathbf{u} = \mathbf{0}\mathbf{u} = \mathbf{0}$$

Thus $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$. By closure, $(-1)\mathbf{u}$ is in U . ■

Example 14. Let A be a fixed matrix in $\mathbb{M}_{n,n}$. Show that the set C_A of all $n \times n$ matrices that commute with A under matrix multiplication is a subspace of $\mathbb{M}_{n,n}$.

Explanation. The set C_A consists of all $n \times n$ matrices X such that $AX = XA$. First, observe that C_A is not empty because I_n is an element. Now we need to show that C_A is closed under matrix addition and scalar multiplication.

Suppose that X_1 and X_2 lie in C_A . Then $AX_1 = X_1A$ and $AX_2 = X_2A$. Then

$$A(X_1 + X_2) = AX_1 + AX_2 = X_1A + X_2A = (X_1 + X_2)A$$

Therefore $(X_1 + X_2)$ commutes with A . Thus $(X_1 + X_2)$ is in C_A . We conclude that C_A is closed under matrix addition.

Now suppose X is in C_A . Let k be a scalar, then

$$A(kX) = k(AX) = k(XA) = (kX)A$$

Therefore (kX) commutes with A . We conclude that (kX) is in C_A , and C_A is closed under scalar multiplication. Hence C_A is a subspace of $\mathbb{M}_{n,n}$.

Suppose $p(x)$ is a polynomial and a is a number. Then the number $p(a)$ obtained by replacing x by a in the expression for $p(x)$ is called the *evaluation* of $p(x)$ at a . For example, if $p(x) = 5 - 6x + 2x^2$, then the evaluation of $p(x)$ at $a = 2$ is $p(2) = 5 - 12 + 8 = 1$. If $p(a) = 0$, the number a is called a *root* of $p(x)$.

Example 15. Consider the set U of all polynomials in $\mathbb{P}$ that have 3 as a root:

$$U = \{p(x) \in \mathbb{P} : p(3) = 0\}$$

Show that U is a subspace of $\mathbb{P}$.

Explanation. Observe that U is not empty because $r(x) = x - 3$ is an element of U . Suppose $p(x)$ and $q(x)$ lie in U . Then $p(3) = 0$ and $q(3) = 0$. We have $(p + q)(x) = p(x) + q(x)$ for all x , so $(p + q)(3) = p(3) + q(3) = 0 + 0 = 0$, and U is closed under addition. The verification that U is closed under scalar multiplication is similar.

Linear Combinations and Span

Definition 3. Let V be a vector space and let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be vectors in V . A vector $\mathbf{v}$ is said to be a *linear combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ if

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n$$

for some scalars $a_1, a_2, \dots, a_n$.

Definition 4. Let V be a vector space and let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ be vectors in V . The set S of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ is called the *span* of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$. We write

$$S = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p)$$

and we say that vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ *span* S . Any vector in S is said to be *in the span* of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$. The set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ is called a *spanning set* for S .

Example 16. Consider $p_1 = 1 + x + 4x^2$ and $p_2 = 1 + 5x + x^2$ in $\mathbb{P}^2$. Determine whether p_1 and p_2 lie in $\text{span}\{1 + 2x - x^2, 3 + 5x + 2x^2\}$.

Explanation. For p_1 , we want to determine if a and b exist such that

$$p_1 = a(1 + 2x - x^2) + b(3 + 5x + 2x^2)$$

Expanding the right hand side gives us:

$$a + 2ax - ax^2 + 3b + 5bx + 2bx^2$$

Combining like terms, we get:

$$(a + 3b) + (2a + 5b)x + (-a + 2b)x^2$$

Setting this equal to $p_1 = 1 + x + 4x^2$ and equating coefficients of powers of x gives us a system of equations

$$1 = a + 3b, \quad 1 = 2a + 5b, \quad \text{and} \quad 4 = -a + 2b$$

This system has the solution $a = -2$ and $b = 1$, so p_1 is indeed in $\text{span}\{1 + 2x - x^2, 3 + 5x + 2x^2\}$.

Turning to $p_2 = 1 + 5x + x^2$, we are looking for a and b such that

$$p_2 = a(1 + 2x - x^2) + b(3 + 5x + 2x^2)$$

Again equating coefficients of powers of x gives equations $1 = a + 3b$, $5 = 2a + 5b$, and $1 = -a + 2b$. But in this case there is no solution, so p_2 is not in $\text{span}\{1 + 2x - x^2, 3 + 5x + 2x^2\}$.

Theorem 4. Let V be a vector space. Let S be any subset of V . Then $U = \text{span}(S)$ is a subspace of V .

Proof See Practice Problem ??.

■

Practice Problems

Problem 7 Given the set of all points in $\mathbb{R}^2$, define the operations of addition and scalar multiplication as follows:

Addition: $(a, b) + (c, d) = (a + d, b + c)$

Scalar Multiplication: $k(a, b) = (ka, kb)$

Is addition, as defined above, commutative?

Multiple Choice:

1. Yes
2. No

Problem 8 Given the set of all points in $\mathbb{R}^2$, define the operations of addition and scalar multiplication as follows:

Addition: $(a, b) + (c, d) = (0, b + d)$

Scalar Multiplication: $k(a, b) = (ka, kb)$

Is addition, as defined above, commutative?

Multiple Choice:

1. Yes
2. No

Is there an additive identity?

Multiple Choice:

1. Yes
2. No

Problem 9 Given the set of all points in $\mathbb{R}^2$, define the operations of addition and scalar multiplication as follows:

Addition: $(a, b) + (c, d) = (a + c, b + d)$

Scalar Multiplication: $k(a, b) = (a, kb)$

Is the Distributive Property over Vector Addition satisfied?

Multiple Choice:

1. Yes
2. No

Is the Distributive Property over Scalar Addition satisfied?

Multiple Choice:

1. Yes
2. No

Problem 10 Let $\mathcal{F}$ be the set of all real-valued functions whose domain is all real numbers. Define addition and scalar multiplication as follows:

$$(f + g)(x) = f(x) + g(x) \quad (cf)(x) = cf(x)$$

Verify that $\mathcal{F}$ is a vector space.

Problem 11 Recall that for a fixed matrix A in $\mathbb{M}_{n,n}$, the set C_A of all $n \times n$ matrices that commute with A under matrix multiplication is a subspace of $\mathbb{M}_{n,n}$. (Example ??)

Describe all elements of C_I , where I is a 3×3 identity matrix.

Problem 12 Is the subset of all invertible $n \times n$ matrices a subspace of $\mathbb{M}_{n,n}$?

Multiple Choice:

1. Yes
2. No

Prove your claim.

Problem 13 Is the subset of all symmetric $n \times n$ matrices a subspace of $\mathbb{M}_{n,n}$?

Multiple Choice:

1. Yes
2. No

Prove your claim.

Problem 14 Let Z be a subset of $\mathbb{M}_{n,n}$ that consists of $n \times n$ matrices that commute with every matrix in $\mathbb{M}_{n,n}$ under matrix multiplication. In other words,

$$Z = \{B : BY = YB \text{ for all } Y \in \mathbb{M}_{n,n}\}$$

Is Z a subspace of $\mathbb{M}_{n,n}$?

Problem 15 List several elements of $\text{span} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right)$. Suggest a spanning set for $\mathbb{M}_{2,2}$.

Problem 16 Find $\text{span}(1, x, x^2, x^3)$.

Problem 17 Let V be a vector space. Let S be any subset of V . Prove that $U = \text{span}(S)$ is a subspace of V . (Theorem ??)

Text Source

The discussion on polynomials was adapted from Section 6.1 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 331-332

Example Source

Examples ?? and ?? were adapted from Examples 6.2.4 and 6.2.7 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 340-341

Exercise Source

Practice Problems ??-?? is adopted from Problems 9.1.1-9.1.4 of Ken Kuttler's *A First Course in Linear Algebra*. (CC-BY)

Ken Kuttler, *A First Course in Linear Algebra*, Lyryx 2017, Open Edition, p. 469.

3 Bases and Dimension of Abstract Vector Spaces

When working with $\mathbb{R}^n$ and subspaces of $\mathbb{R}^n$ we developed several fundamental ideas including *span*, *linear independence*, *bases* and *dimension*. We will find that these concepts generalize easily to abstract vector spaces and that analogous results hold in these new settings.

Linear Independence

Definition 5 (Linear Independence). Let V be a vector space. Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ be vectors of V . We say that the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ is *linearly independent* if the only solution to

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_p\mathbf{v}_p = \mathbf{0}$$

is the *trivial solution* $a_1 = a_2 = \dots = a_p = 0$.

If, in addition to the trivial solution, a *non-trivial solution* (not all $a_1, a_2, \dots, a_p$ are zero) exists, then we say that the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ is *linearly dependent*.

Example 17. Show that $P = \{1 + x, 3x + x^2, 2 + x - x^2\}$ is linearly independent in $\mathbb{P}^2$.

Explanation. Consider the linear combination equation

$$\begin{aligned} a(1 + x) + b(3x + x^2) + c(2 + x - x^2) &= 0 \\ a + ax + 3bx + bx^2 + 2c + cx - cx^2 &= 0 \\ (a + 2c) + (a + 3b + c)x + (b - c)x^2 &= 0 \end{aligned}$$

The constant term, as well as the coefficients in front of x and x^2 , must be equal to 0. This gives us the following system of equations.

$$\begin{array}{ccccccc} a & + & & & + & 2c & = & 0 \\ a & + & 3b & + & & c & = & 0 \\ & & b & - & c & & = & 0 \end{array}$$

The only solution is $a = b = c = 0$. We conclude that P is linearly independent in $\mathbb{P}^2$.

Bases and Dimension

Recall that our motivation for defining a basis of a subspace of $\mathbb{R}^n$ was to have a collection of vectors such that every vector of the subspace can be expressed as a unique linear combination of the vectors in that collection. Definition of a basis (??) generalizes to abstract vector spaces as follows.

Definition 6. Let V be a vector space. A set $\mathcal{B}$ of vectors of V is called a *basis* of V provided that

1. $\text{span}(\mathcal{B}) = V$
2. $\mathcal{B}$ is linearly independent.

Theorem 5. Let V be a vector space, and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a basis for V . Then every element $\mathbf{v}$ of V has a unique representation as linear combination of the elements of $\mathcal{B}$.

Proof By the definition of a basis, we know that $\mathbf{v}$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. Suppose there are two such representations. Then,

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n$$

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \dots + b_n\mathbf{v}_n$$

But then we have:

$$\begin{aligned} a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n &= b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \dots + b_n\mathbf{v}_n \\ a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n - (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \dots + b_n\mathbf{v}_n) &= \mathbf{0} \\ (a_1 - b_1)\mathbf{v}_1 + (a_2 - b_2)\mathbf{v}_2 + \dots + (a_n - b_n)\mathbf{v}_n &= \mathbf{0} \end{aligned}$$

Because $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are linearly independent, we have $a_i - b_i = 0$ for $1 \leq i \leq n$. Consequently $a_i = b_i$ for $1 \leq i \leq n$. ■

In [Bases and Dimension](#) we defined the dimension of a subspace of $\mathbb{R}^n$ to be the number of elements in a basis. (Definition ??) We will adopt this definition for abstract vector spaces. As before, to ensure that *dimension* is well-defined we need to establish that this definition is independent of our choice of a basis. The proof of the following theorem is identical to the proof of its counterpart in $\mathbb{R}^n$. (Theorem ??)

Theorem 6. Let V be a vector space. Suppose $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_s\}$ be two bases of V . Then $s = t$.

Now we can state the definition.

Definition 7. Let V be a subspace of a vector space W . The *dimension* of V is the number, m , of elements in any basis of V . We write

$$\dim(V) = m$$

In our discussions up to this point, we have always assumed that a basis is nonempty and hence that the dimension of the space is at least 1. However, the zero space $\{\mathbf{0}\}$ has *no* basis. To accommodate for this, we will say that the zero vector space $\{\mathbf{0}\}$ is defined to have dimension 0:

$$\dim\{\mathbf{0}\} = 0$$

Our insistence that $\dim\{\mathbf{0}\} = 0$ amounts to saying that the empty set of vectors is a basis of $\{\mathbf{0}\}$. Thus the statement that “the dimension of a vector space is the number of vectors in any basis” holds even for the zero space.

Example 18. Recall that the vector space $\mathbb{M}_{m,n}$ consists of all $m \times n$ matrices. (See Example ?? of [Abstract Vector Spaces](#)). Find a basis and the dimension of $\mathbb{M}_{m,n}$.

Explanation. Let $\mathcal{B}$ consist of $m \times n$ matrices with exactly one entry equal to 1 and all other entries equal to 0. It is clear that every $m \times n$ matrix can be written as a linear combination of elements of $\mathcal{B}$. It is also easy to see that the elements of $\mathcal{B}$ are linearly independent. Thus $\mathcal{B}$ is a basis for $\mathbb{M}_{m,n}$. The set $\mathcal{B}$ contains mn elements, so $\dim(\mathbb{M}_{m,n}) = mn$.

Example 19. Recall that $\mathbb{P}^n$ is the set of all polynomials of degree n or less. (See Example ?? of [Abstract Vector Spaces](#)) Show that $\dim(\mathbb{P}^n) = n + 1$ and that $\{1, x, x^2, \dots, x^n\}$ is a basis of $\mathbb{P}^n$.

Explanation. Each polynomial $p(x) = a_0 + a_1x + \dots + a_nx^n$ in $\mathbb{P}^n$ is clearly a linear combination of $1, x, \dots, x^n$, so $\mathbb{P}^n = \text{span}\{1, x, \dots, x^n\}$.

Suppose $a_0 \cdot 1 + a_1x + \dots + a_nx^n = 0$, then $a_0 = a_1 = \dots = a_n = 0$. So $\{1, x, \dots, x^n\}$ is linearly independent and hence is a basis containing $n + 1$ vectors. Thus, $\dim(\mathbb{P}^n) = n + 1$.

Example 20. Consider the subset

$$C_A = \{X \in \mathbb{M}_{2,2} : AX = XA\}$$

of $\mathbb{M}_{2,2}$.

It was shown in Example ?? of [Abstract Vector Spaces](#) that C_A is a subspace for any choice of the matrix A .

Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. Show that $\dim(C_A) = 2$ and find a basis of C_A .

Explanation. Suppose $X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is in C_A . Then

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} a+c & b+d \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a & a \\ c & c \end{bmatrix}$$

This gives us two relationships:

$$b + d = a \quad \text{and} \quad c = 0$$

We can now express a generic element X of C_A as

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} b+d & b \\ 0 & d \end{bmatrix} = \begin{bmatrix} b & b \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} = b \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + d \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

The set $\mathcal{B}$ is linearly independent. (See Practice Problem ??) Every element X of C_A can be written as a linear combination of elements of $\mathcal{B}$. Thus $C_A = \text{span}\mathcal{B}$. Therefore $\mathcal{B}$ is a basis of C_A , and $\dim(C_A) = 2$.

Example 21. In Practice Problem ?? of [Abstract Vector Spaces](#) you demonstrated that the set of all symmetric $n \times n$ matrices is a subspace of $\mathbb{M}_{n,n}$.

Let V be a subspace of $\mathbb{M}_{2,2}$ consisting of all 2×2 symmetric matrices. Find the dimension of V .

Explanation. A matrix A is symmetric if $A^T = A$. In other words, a matrix A is symmetric when entries directly across the main diagonal are equal, so each 2×2 symmetric matrix has the form

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

Hence the set $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$ spans V . The reader can verify that $\mathcal{B}$ is linearly independent.

Thus $\mathcal{B}$ is a basis of V , so $\dim(V) = 3$.

Finite-Dimensional Vector Spaces

Our definition of dimension of a vector space depends on the vector space having a basis. In this section we will establish that any vector space spanned by finitely many vectors has a basis.

Definition 8. A vector space is said to be *finite-dimensional* if it is spanned by finitely many vectors.

Given a finite-dimensional vector space V we will find a basis for V by starting with a linearly independent subset of V and expanding it to a basis. The following results are more general versions of Lemmas ?? and ??, and Theorem ?? of [Bases and Dimension](#). The proofs are identical and we will omit them.

Lemma 1. Let V be a vector space spanned by n vectors. If a linearly independent subset S of V contains m vectors, then $m \leq n$.

Lemma 2. Let V be a vector space. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a linearly independent subset of V . If $\mathbf{u}$ is not in $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$, then $\{\mathbf{u}, \mathbf{v}_1, \dots, \mathbf{v}_k\}$ is linearly independent.

Theorem 7. Let V be a finite-dimensional vector space. Any linearly independent subset of V can be expanded to a basis of V .

Coordinate Vectors

Recall that in the context of $\mathbb{R}^n$ (and subspaces of $\mathbb{R}^n$) the requirement that elements of a basis be linearly independent guarantees that every element of the vector space has a *unique* representation in terms of the elements of the basis. (See Theorem ?? of [Introduction to Bases](#)) We proved the same property for abstract vector spaces in Theorem ??.

Uniqueness of representation in terms of the elements of a basis allows us to associate every element of a vector space V with a unique *coordinate vector* with respect to a given basis. Coordinate vectors were first introduced in [Introduction to Bases](#). We now give a formal definition.

Definition 9. Let V be a vector space, and let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for V . If $\mathbf{v} = a_1\mathbf{v}_1 + \dots + a_n\mathbf{v}_n$, then the vector in $\mathbb{R}^n$ whose components are the coefficients $a_1, \dots, a_n$ is said to be the *coordinate vector for $\mathbf{v}$ with respect to $\mathcal{B}$* . We denote the coordinate vector by $[\mathbf{v}]_{\mathcal{B}}$ and write:

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Remark 1. The order of in which vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ appear in $\mathcal{B}$ of Definition ?? is important. Switching the order of these vectors would switch the order of the coordinate vector components. For this reason, we will often use the term *ordered basis* to describe $\mathcal{B}$ in the context of coordinate vectors.

Example 22. The coordinate vector for $p(x) = 4 - 3x^2 + 5x^3$ in $\mathbb{P}^4$ with respect to the ordered basis $\mathcal{B}_1 = \{1, x, x^2, x^3, x^4\}$ is

$$[p(x)]_{\mathcal{B}_1} = \begin{bmatrix} 4 \\ 0 \\ -3 \\ 5 \\ 0 \end{bmatrix}$$

Now let's change the order of the elements in $\mathcal{B}_1$. The coordinate vector for $p(x) = 4 - 3x^2 + 5x^3$ with respect to the ordered basis $\mathcal{B}_2 = \{x^4, x^3, x^2, x, 1\}$ is

$$[p(x)]_{\mathcal{B}_2} = \begin{bmatrix} 0 \\ 5 \\ -3 \\ 0 \\ 4 \end{bmatrix}$$

Example 23. Show that the set $\mathcal{B} = \{x, 1 + x, x + x^2\}$ is a basis for $\mathbb{P}^2$. Keep the order of elements in $\mathcal{B}$ and find the coordinate vector for $p(x) = 4 - x + 3x^2$ with respect to the ordered basis $\mathcal{B}$.

Explanation. We will begin by showing that the elements of $\mathcal{B}$ are linearly independent. Suppose

$$ax + b(1 + x) + c(x + x^2) = 0$$

then

$$b + (a + b + c)x + cx^2 = 0$$

This gives us the following system of equations:

$$\begin{array}{rcl} b & = & 0 \\ a + b + c & = & 0 \\ c & = & 0 \end{array}$$

The solution $a = b = c = 0$ is unique. We conclude that $\mathcal{B}$ is linearly independent.

Next, we need to show that $\mathcal{B}$ spans $\mathbb{P}^2$. To this end, we will consider a generic element $p(x) = \alpha + \beta x + \gamma x^2$ of $\mathbb{P}^2$ and attempt to express it as a linear combination of the elements of $\mathcal{B}$.

$$ax + b(1 + x) + c(x + x^2) = \alpha + \beta x + \gamma x^2$$

then

$$b + (a + b + c)x + cx^2 = \alpha + \beta x + \gamma x^2$$

Setting the coefficients of like terms equal to each other gives us

$$\begin{array}{rcl} b & = & \alpha \\ a + b + c & = & \beta \\ c & = & \gamma \end{array}$$

Solving this linear system of a , b and c gives us

$$a = \beta - \alpha - \gamma, \quad b = \alpha, \quad c = \gamma$$

(You should verify this.) This shows that every element of $\mathbb{P}^2$ can be written as a linear combination of elements of $\mathcal{B}$. Therefore $\mathcal{B}$ is a basis for $\mathbb{P}^2$.

To find the coordinate vector for $p(x) = 4 - x + 3x^2$ with respect to $\mathcal{B}$ we need to express $p(x)$ as a linear combination of the elements of $\mathcal{B}$. Fortunately, we have already done all the necessary work. For $p(x)$, $\alpha = 4$, $\beta = -1$ and $\gamma = 3$. This gives us the coefficients of the linear combination: $a = \beta - \alpha - \gamma = -8$, $b = \alpha = 4$, $c = \gamma = 3$. We now write $p(x)$ as a linear combination

$$p(x) = -8(x) + 4(1 + x) + 3(x + x^2)$$

The coordinate vector for $p(x)$ with respect to $\mathcal{B}$ is

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} -8 \\ 4 \\ 3 \end{bmatrix}$$

Example 24. Recall that the set V of all symmetric 2×2 matrices is a subspace of $\mathbb{M}_{2,2}$. In Example ?? we demonstrated that $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$ is a basis for V . Let $A = \begin{bmatrix} 2 & -3 \\ -3 & 1 \end{bmatrix}$. Observe that A is an element of V .

1. Find the coordinate vector with respect to the ordered basis $\mathcal{B}$ for A .
2. Let $\mathcal{B}' = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$ be another ordered basis for V . Find the coordinate vector for A with respect to $\mathcal{B}'$.

Explanation. We write A as a linear combination of the elements of $\mathcal{B}$.

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

Thus, the coordinate vector with respect to $\mathcal{B}$ is

$$[A]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$$

The coordinate vector with respect to $\mathcal{B}'$ is

$$[A]_{\mathcal{B}'} = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix}$$

Coordinate vectors will play a vital role in establishing one of the most fundamental results in linear algebra, that all n -dimensional vector spaces have the same structure as $\mathbb{R}^n$. In Example ?? of [Isomorphic Vector Spaces](#), for instance, we will show that $\mathbb{P}^2$ is essentially the same as $\mathbb{R}^3$.

Practice Problems

Problem 18 Consider the subset

$$C_A = \{X \in \mathbb{M}_{2,2} : AX = XA\}$$

of $\mathbb{M}_{2,2}$. (See Example ??)

Prove that set $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a linearly independent subset of C_A .

Problem 19 Show that the following subset of $\mathbb{P}^2$ is linearly independent, and is a basis of $\mathbb{P}^2$.

$$\{1 + x, 1 - x, x + x^2\}$$

Problem 20 Show that the following subset of $\mathbb{P}^2$ is linearly independent, and is a basis of $\mathbb{P}^2$.

$$\{x^2, x + 1, 1 - x - x^2\}$$

Problem 21 Show that the following subset of $\mathbb{M}_{2,2}$ is linearly independent, and is a basis of $\mathbb{M}_{2,2}$.

$$\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right\}$$

Problem 22 Find the coordinate vector for $p(x) = 6 - 2x + 4x^2$ in $\mathbb{P}^2$ with respect to

$$\mathcal{B} = \{1 + x, 1 - x, x + x^2\}$$

Answer:

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix}$$

Problem 23 Find the coordinate vector for $p(x) = 6 - 2x + 4x^2$ in $\mathbb{P}^2$ with respect to

$$\mathcal{B} = \{x^2, x + 1, 1 - x - x^2\}$$

Answer:

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix}$$

Problem 24 Find the coordinate vector for $A = \begin{bmatrix} 4 & -3 \\ 1 & 2 \end{bmatrix}$ in $\mathbb{M}_{2,2}$ with respect to the ordered basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right\}$$

Answer:

$$[A]_{\mathcal{B}} = \begin{bmatrix} \dots \\ \dots \\ \dots \\ \dots \end{bmatrix}$$

Text Source

The discussion of the zero space was adapted from Section 6.3 of Keith Nicholson's [Linear Algebra with Applications](#). (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 349

Example Source

Examples ?? and ?? were adapted from Examples 6.3.1 and 6.3.10 of Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 346, 350

Exercise Source

Practice Problems ??, ?? and ?? are Exercises 6.3(a)(b)(c) from Keith Nicholson's *Linear Algebra with Applications*. (CC-BY-NC-SA)

W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2018, Open Edition, p. 351