Introduction

Who This Book Is For

This book is for anyone who wants to understand linear algebra from scratch — no prerequisites beyond high school algebra. If you can solve $2x + 3 = 7$, you're ready.

Linear algebra is the mathematics of vectors, matrices, and the transformations they describe. It is the single most important branch of mathematics for modern technology: every neural network, every search engine, every computer game, every recommendation system runs on linear algebra.

Why a "Cookbook"?

Most textbooks explain concepts and leave you to figure out the mechanics. This book does the opposite: it shows you exactly how to compute, step by step, with many worked examples. Every chapter follows the same pattern:

1. **A concept introduction** that explains what we're doing and why
2. **Worked examples** — lots of them — with every step shown and the answer boxed
3. **Verification steps** so you can check your work
4. **Common mistakes** to avoid
5. **Exercises** to practice on your own

Think of it as a recipe book. You wouldn't learn to cook by reading about the chemistry of flavor — you'd follow recipes, see results, and build intuition through practice.

How This Book Is Organized

The book has five parts, each building on the last:

Part I: Vectors (Chapters 1–4) introduces vectors — the basic objects of linear algebra. You'll learn what they are, how to add and scale them, how to measure lengths and angles, and what it means for vectors to be independent.

Part II: Matrices (Chapters 5–10) introduces matrices — rectangular arrays of numbers that represent linear transformations. You'll learn how to multiply them, solve systems of equations, compute determinants and inverses, and understand what matrices do geometrically.

Part III: Deeper Structure (Chapters 11–13) reveals the hidden structure inside matrices. Eigenvalues tell you how a matrix stretches space. Orthogonality gives you perpendicular coordinates. The singular value decomposition is the ultimate factorization.

Part IV: Calculus for Linear Algebra (Chapters 14–16) introduces just enough calculus to understand neural networks. You'll learn derivatives, gradients, and optimization — the mathematical engine of machine learning.

Part V: The Mathematics of Neural Networks (Chapters 17–19) brings everything together. You'll see exactly how vectors, matrices, derivatives, and optimization combine to make a neural network learn. The final chapter on backpropagation is the grand synthesis of the entire book.

The Neural Network Thread

Every chapter includes a box explaining how that chapter's math appears in neural networks. These boxes build a continuous thread from Chapter 1 to Chapter 19:

- A neural network's input is a **vector** (Ch 1)
- A neuron computes a **linear combination** of its inputs (Ch 2)
- The dot product $\mathbf{w}^\top\mathbf{x}$ measures how much input matches weights (Ch 3)

- The **rank** of a weight matrix limits what a layer can learn (Ch 4)
- Weight matrices are literally **matrices** (Ch 5)
- The forward pass IS **matrix multiplication** (Ch 6)
- Training starts by **solving linear systems** (Ch 7)
- A near-zero **determinant** means a layer collapses information (Ch 8)
- The **transpose** (not the inverse) propagates errors backward (Ch 9)
- Each layer IS a **linear transformation** (Ch 10)
- **Eigenvalues** reveal which features carry the most information (Ch 11)
- **Orthogonal** initialization prevents exploding/vanishing gradients (Ch 12)
- **SVD** reveals effective rank and enables model compression (Ch 13)
- The **chain rule** IS backpropagation (Ch 14)
- The **gradient** tells you how to adjust every weight (Ch 15)
- **Gradient descent** IS how neural networks learn (Ch 16)
- A single **neuron** combines dot product + activation + loss (Ch 17)
- **Layers** chain matrix multiplications with nonlinearities (Ch 18)
- **Backpropagation** multiplies Jacobians via the chain rule (Ch 19)

By the end, you won't just know the math — you'll understand exactly why neural networks work.

Notation

We use standard mathematical notation throughout. A complete reference is in Appendix A, but here are the essentials:

Symbol	Meaning
x, y, z	Scalars (single numbers)
$\mathbf{v}, \mathbf{w}$	Vectors (bold lowercase)
A, B, W	Matrices (uppercase)
$\mathbb{R}^n$	The set of all n-dimensional real vectors
$\mathbf{v}^\top$	Transpose of $\mathbf{v}$

Symbol	Meaning
$\|\mathbf{v}\|$	Length (norm) of $\mathbf{v}$
$\mathbf{v} \cdot \mathbf{w}$	Dot product

Let's Begin

Open Chapter 1 and pick up a pencil. The best way to learn math is to do math.

What Is a Vector?

Linear algebra begins with vectors. Informally, a vector is an ordered list of real numbers — but this chapter gives that idea precise meaning. We define $\mathbb{R}^n$, establish notation, and develop the geometric and algebraic pictures that the rest of the book builds on.

Throughout, vectors are **column vectors**:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$$

📋 **Quick Reference: Vector Basics**

Definitions:

- $\mathbb{R}^n$ is the set of all column vectors with n real entries
- A **vector** $\mathbf{v} \in \mathbb{R}^n$ is an ordered n-tuple of real numbers, written as a column
- The **i-th component** of $\mathbf{v}$ is v_i
- The **zero vector** $\mathbf{0}$ has all components equal to zero

Notation:

$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$, and the row form is written $\mathbf{v}^\top = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}$

Equality: $\mathbf{u} = \mathbf{v}$ iff $u_i = v_i$ for all $i = 1, 2, \ldots, n$

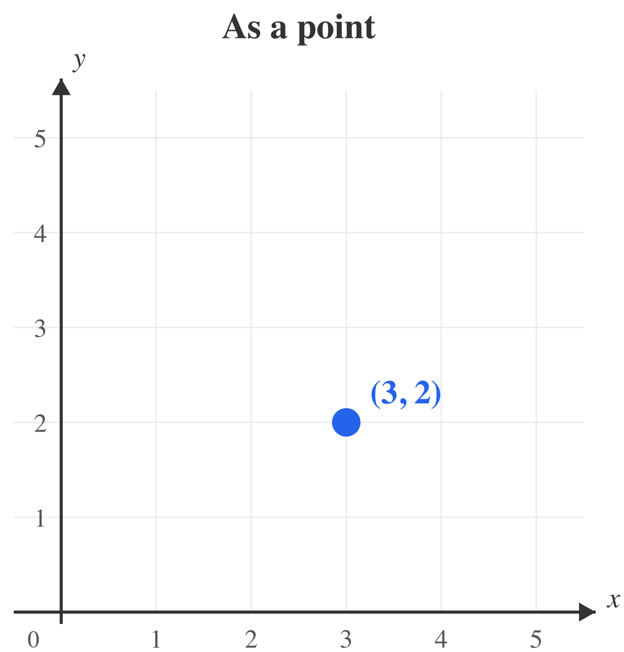
Space	Elements	Geometric picture
$\mathbb{R}^1$	Scalars on a number line	Points on a line
$\mathbb{R}^2$	2-component columns	Points/arrows in the plane
$\mathbb{R}^3$	3-component columns	Points/arrows in 3-space
$\mathbb{R}^n$	n-component columns	No visual picture for $n \geq 4$, but the algebra is identical

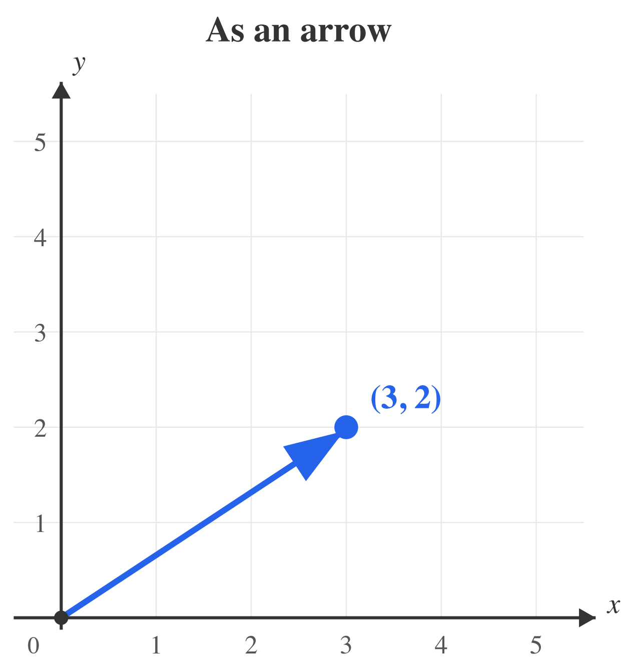
---

Geometric Interpretation

A vector $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2$ has two equivalent geometric readings:

1. **A point** in the plane at coordinates (v_1, v_2).
2. **A directed arrow** from the origin to that point, encoding both a direction and a magnitude.





Both pictures are used constantly. Data scientists tend to think “points in a cloud” — each data record is a point in some high-dimensional space. Physicists tend to think “arrows with magnitude and direction” — forces, velocities, and electric fields are all arrows. The algebra is the same either way, and that is one of the great strengths of linear algebra: a single formalism serves many disciplines.

Why two pictures? The point view emphasizes *location*: where is this data record relative to others? The arrow view emphasizes *change*: how far, and in what direction? In Chapter 2, when we add vectors, the arrow view makes the geometry transparent — we place arrows tip-to-tail. In Chapter 3, when we measure distance between data points, the point view is more natural.

Notice that the arrow always starts at the origin $(0, 0)$. This is the convention for a **free vector represented in standard position**. Later (in the displacement vector section below) we will see arrows that start at other points, but the default is always the origin. Every vector in $\mathbb{R}^n$ has a unique arrow in standard position, and every arrow from the origin corresponds to exactly one vector. This one-to-one correspondence is what lets us move freely between algebra and geometry.

In $\mathbb{R}^3$ the same duality holds: a vector is simultaneously a point in 3-space and an arrow from the origin. For $n \geq 4$ we can no longer draw pictures, but the algebraic definition — an ordered column of n real numbers — still works perfectly. The geometric intuition built in 2D and 3D guides our reasoning even when visualization fails.

Example 1: Vectors in R² and Their Geometry

Problem: Represent $\mathbf{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$ geometrically.

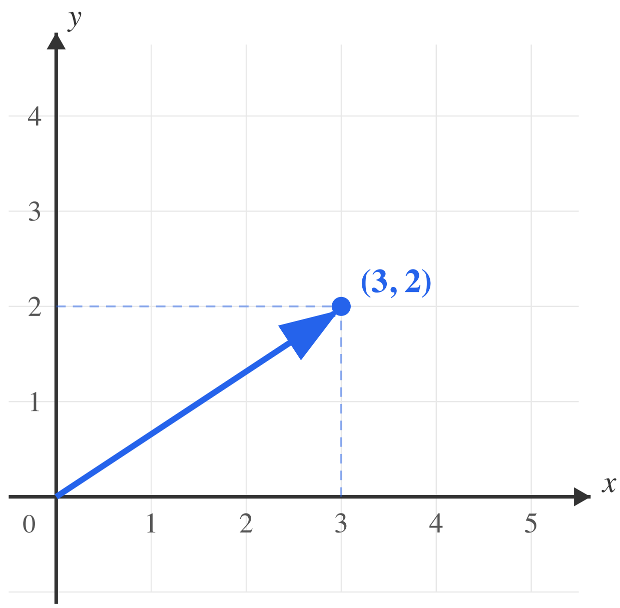
Step 1: Identify the components.

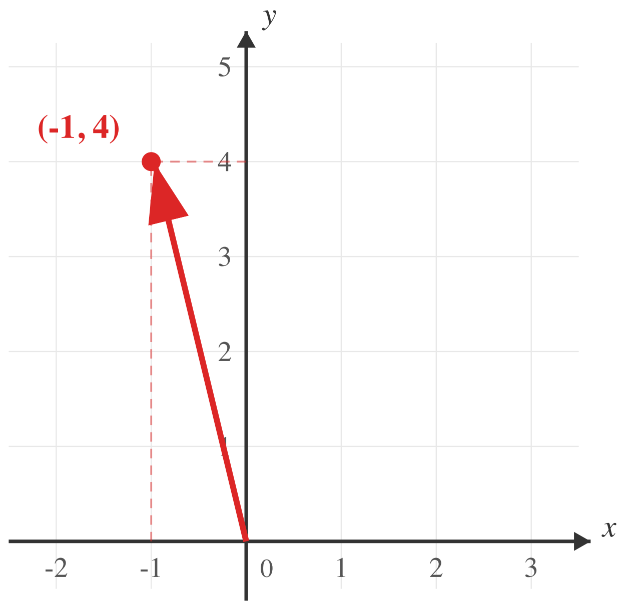
$$\mathbf{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}: \quad v_1 = 3,\ v_2 = 2$$

$$\mathbf{w} = \begin{pmatrix} -1 \\ 4 \end{pmatrix} : \quad w_1 = -1, \ w_2 = 4$$

Step 2: Plot as arrows from the origin.

For $\mathbf{v}$: move 3 right, 2 up. For $\mathbf{w}$: move 1 left, 4 up.





$$\mathbf{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \in \mathbb{R}^2, \quad \mathbf{w} = \begin{pmatrix} -1 \\ 4 \end{pmatrix} \in \mathbb{R}^2$$

Verify: Each vector has 2 components, so both live in $\mathbb{R}^2$. Negative components point in the negative axis direction. ✓

Example 2: A Vector in R³

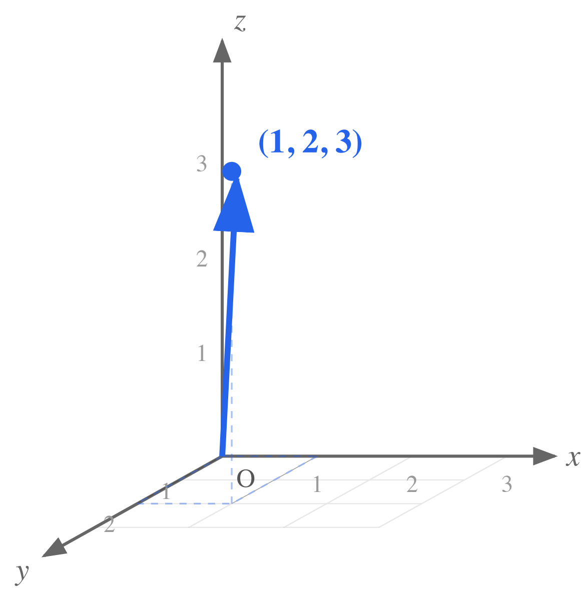
Problem: Write $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ in $\mathbb{R}^3$, identify its components, and describe the geometry.

Step 1: Components.

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \qquad v_1 = 1, \quad v_2 = 2, \quad v_3 = 3$$

Step 2: Geometric picture.

The arrow starts at the origin and ends at the point $(1, 2, 3)$ in 3-space.



$$\boxed{\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in \mathbb{R}^3}$$

Example 3: Vector Equality — Order and Dimension Matter

Problem: Determine which of the following pairs are equal:

(a) $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$

(b) $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$

(c) $\begin{pmatrix} \pi \\ e \end{pmatrix}$ and $\begin{pmatrix} \pi \\ e \end{pmatrix}$

Step 1: Apply the definition.

Vectors are equal iff they have the same dimension and every corresponding component matches.

Step 2: Check each pair.

(a) $v_1 = 1 \neq 2 = w_1$. **Not equal.** Order matters.

(b) The first is in $\mathbb{R}^2$, the second in $\mathbb{R}^3$. Different dimensions $\Rightarrow$ **not equal**, even though the first two components match.

(c) Same dimension ($\mathbb{R}^2$), and $\pi = \pi$, $e = e$. **Equal.**

$$\boxed{\text{(a) not equal} \qquad \text{(b) not equal} \qquad \text{(c) equal}}$$

Example 4: Solving for Components via Vector Equality

Problem: Find x, y, z such that

$$\begin{pmatrix} x + 2y \\ 3z \\ x - z \end{pmatrix} = \begin{pmatrix} 7 \\ 9 \\ 1 \end{pmatrix}$$

Step 1: Equate corresponding components.

$$x + 2y = 7, \qquad 3z = 9, \qquad x - z = 1$$

Step 2: Solve the system.

From the second equation: $z = 3$.

Substituting into the third: $x = 1 + z = 4$.

Substituting into the first: $4 + 2y = 7$, so $y = \frac{3}{2}$.

$$\boxed{x = 4, \quad y = \tfrac{3}{2}, \quad z = 3}$$

Verify: $\begin{pmatrix} 4 + 3 \\ 9 \\ 4 - 3 \end{pmatrix} = \begin{pmatrix} 7 \\ 9 \\ 1 \end{pmatrix}$ ✓

The Zero Vector

The **zero vector $\mathbf{0}$** $\in \mathbb{R}^n$ has every component equal to zero:

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n$$

It is the **additive identity**: $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v}$ (we'll prove this in Chapter 2). Geometrically, it is the "zero-length arrow" sitting at the origin — a vector with no direction and no magnitude. In the point interpretation, it is simply the origin itself.

Note that there is a different zero vector for each n. The zero vector in $\mathbb{R}^2$ is $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, in $\mathbb{R}^3$ it is $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, etc. When we write $\mathbf{0}$ without specifying the dimension, the space should be clear from context.

The zero vector plays a role in linear algebra analogous to the number 0 in ordinary arithmetic. Just as $a + 0 = a$ for any real number, $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for any vector. And just as $0 \cdot a = 0$, we will see in Chapter 2 that $0 \cdot \mathbf{v} = \mathbf{0}$ — scaling any vector by the scalar zero produces the zero vector. Despite these parallels, be careful: $\mathbf{0}$ (bold, a vector) and 0 (not bold, a scalar) are different types of objects. Confusing them is a common source of errors in proofs.

The zero vector also has a special status in the theory of linear independence (Chapter 4): it is *never* considered linearly independent, and it is always in the span of any set of vectors. Any equation of the form $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$ will be central to the definition of independence.

Position Vectors and Displacement Vectors

There are two important and fundamentally different ways to use vectors in geometry:

- A **position vector** $\overrightarrow{OP}$ points from the origin O to a point P. It names a *location* in space.
- A **displacement vector** $\overrightarrow{AB} = B - A$ encodes the *change* from point A to point B. It names a *movement* — a direction and a distance.

The key relationship connecting them is:

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$$

To understand the distinction, think of a GPS. Your current coordinates are a position vector — they pin you to a specific point on Earth's surface. The instruction "drive 3 km north and 2 km east" is a displacement vector — it tells you how to move, but not where to start. Two people in different cities could follow the same displacement and end up in completely different places.

Algebraically, the displacement from A to B is computed by subtracting component-wise: if $A = (a_1, a_2)$ and $B = (b_1, b_2)$, then $\overrightarrow{AB} = \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \end{pmatrix}$. The result is a vector in its own right — it lives in $\mathbb{R}^n$ just like position vectors do, but its meaning is different.

An important consequence: displacement vectors are **translation-invariant**. If you slide both A and B by the same amount, the displacement $\overrightarrow{AB}$ does not change. Position vectors, by contrast, depend entirely on where the origin is. This is why physicists prefer to work with displacements (forces, velocities) rather than positions — the physics should not depend on an arbitrary choice of origin.

One beautiful property of displacements is the **chain rule for points**: for any sequence of points $P_0, P_1, \ldots, P_k$, the displacements telescope:

$$\overrightarrow{P_0P_1} + \overrightarrow{P_1P_2} + \cdots + \overrightarrow{P_{k-1}P_k} = \overrightarrow{P_0P_k}$$

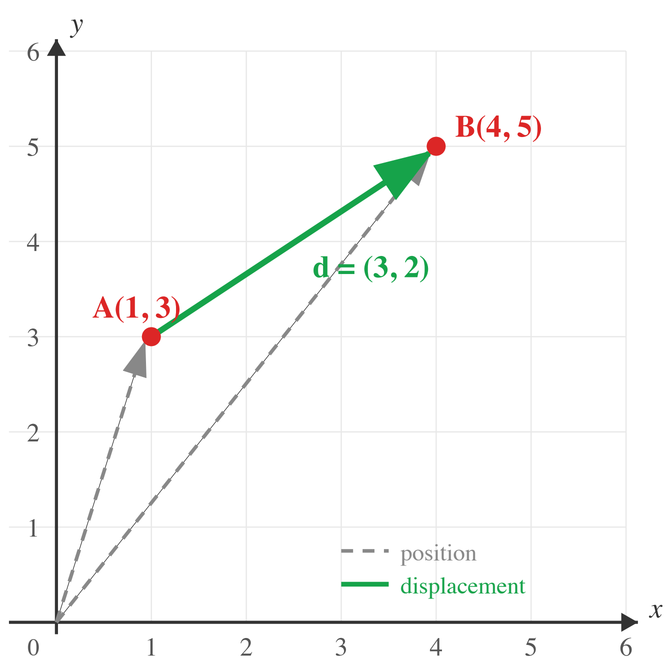
If the path is closed ($P_k = P_0$), the sum is $\mathbf{0}$. We will see this in Example 6.

Example 5: Computing Displacement Vectors

Problem: Points $A = (1, 3)$ and $B = (4, 5)$ are given. Find the displacement $\overrightarrow{AB}$ and verify that $A + \overrightarrow{AB} = B$.

Step 1: Compute the displacement.

$$\overrightarrow{AB} = B - A = \begin{pmatrix} 4-1 \\ 5-3 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$



$$\boxed{\overrightarrow{AB} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}}$$

Verify: $A + \overrightarrow{AB} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix} = B$. ✓

Example 6: Displacement Vectors Form a Closed Loop

Problem: Let $A = (0, 0)$, $B = (3, 0)$, $C = (1, 4)$. Compute $\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CA}$.

Step 1: Compute each displacement.

$$\overrightarrow{AB} = B - A = \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \quad \overrightarrow{BC} = C - B = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$$

$$\overrightarrow{CA} = A - C = \begin{pmatrix} -1 \\ -4 \end{pmatrix}$$

Step 2: Add.

$$\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CA} = \begin{pmatrix} 3 - 2 - 1 \\ 0 + 4 - 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0}$$

$$\boxed{\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CA} = \mathbf{0}}$$

This always holds: traversing a closed path returns you to the start, so the total displacement is zero. This is a consequence of the **telescoping property**: $(B - A) + (C - B) + (A - C) = \mathbf{0}$.

Vectors in High Dimensions

Nothing in our definitions restricts n. A vector in $\mathbb{R}^{1000}$ is just a column of 1000 real numbers. We lose geometric visualization for $n \geq 4$, but the algebra — equality, components, displacement — works identically. Every definition and theorem in this chapter applies to $\mathbb{R}^{784}$ (an image) just as well as to $\mathbb{R}^2$ (a point in the plane).

This is what makes linear algebra so powerful: the same framework handles 2D geometry, 3D physics, and data with millions of features. The formulas do not change; only the number of components grows.

In practice, most interesting data is high-dimensional. A color image at 256×256 resolution has $256 \times 256 \times 3 = 196{,}608$ components. A vocabulary of 50,000 words, each represented by a 300-dimensional embedding, gives a matrix with 15 million entries. Modern language models work with vectors in $\mathbb{R}^{12288}$ or larger. None of this requires new theory — it is the same dot products, the same matrix multiplications, the same eigenvalues you will learn in this book, scaled up.

The lesson is: **do not let the inability to draw a picture stop you from reasoning about high-dimensional vectors.** The geometric language — "length," "angle," "distance," "perpendicular" — all have precise algebraic definitions (Chapter 3) that work in any dimension. When we say two document vectors in $\mathbb{R}^{50000}$ are "close," we mean their distance (computed by a formula) is small. When we say two feature vectors are "perpendicular," we mean their dot product is zero. Geometry is a metaphor; the algebra is the reality.

Example 7: Data as Vectors in R^n

Problem: A grayscale image has $28 \times 28 = 784$ pixels, each with intensity from 0 (black) to 255 (white). Represent this image as a vector.

Step 1: Flatten the pixel grid into a column.

Reading pixel intensities row by row:

$$\mathbf{x} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{784} \end{pmatrix} \in \mathbb{R}^{784}$$

Step 2: Identify the space.

$$\boxed{\mathbf{x} \in \mathbb{R}^{784}}$$

Each image is a single point in a 784-dimensional space. Two images are "close" if the corresponding pixel vectors are close. This is exactly how handwritten digit recognition (MNIST) works: classify points in $\mathbb{R}^{784}$.

Example 8: Standard Basis Vectors

Problem: Write the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in $\mathbb{R}^3$.

Step 1: Definition.

The i-th standard basis vector $\mathbf{e}_i$ has a 1 in position i and 0 elsewhere.

Step 2: Write them out.

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Step 3: Express an arbitrary vector using these.

Any $\mathbf{v} \in \mathbb{R}^3$ can be written:

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$$

$$\boxed{\mathbf{e}_i \text{ has a 1 in position } i \text{ and 0s elsewhere}}$$

Verify: $2\mathbf{e}_1 + 3\mathbf{e}_2 + 5\mathbf{e}_3 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$. ✓

Example 9: General Standard Basis in Rⁿ

Problem: Write $\mathbf{e}_3$ in $\mathbb{R}^5$ and verify that

$$\begin{pmatrix} 4 \\ -1 \\ 7 \\ 0 \\ 2 \end{pmatrix} = 4\mathbf{e}_1 - \mathbf{e}_2 + 7\mathbf{e}_3 + 0\mathbf{e}_4 + 2\mathbf{e}_5$$

Step 1: Write $\mathbf{e}_3$ in $\mathbb{R}^5$.

$$\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \in \mathbb{R}^5$$

Step 2: Expand the linear combination.

$$4\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + (-1)\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + 7\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$+ 0\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ 7 \\ 0 \\ 2 \end{pmatrix}$$

$$\boxed{\text{In general: } \mathbf{v} = \sum_{i=1}^{n} v_i\, \mathbf{e}_i \text{ for any } \mathbf{v} \in \mathbb{R}^n}$$

Verify: Each component matches. ✓

Example 10: The Row Vector / Column Vector Distinction

Problem: Explain the difference between the row vector $\mathbf{v}^\top = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$ and the column vector $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Step 1: Shapes.

- $\mathbf{v}$ is a 3×1 matrix (column vector)
- $\mathbf{v}^\top$ is a 1×3 matrix (row vector)

Step 2: Why it matters.

In matrix multiplication (Chapter 6), shape determines what operations are legal. The product $\mathbf{v}^\top \mathbf{w}$ (a 1×3 times 3×1) gives a scalar — this is the dot product. The product $\mathbf{v}\mathbf{w}^\top$ (a 3×1 times 1×3) gives a 3×3 matrix — the outer product.

$$\boxed{\mathbf{v} \in \mathbb{R}^{n \times 1} \text{ (column)}, \quad \mathbf{v}^\top \in \mathbb{R}^{1 \times n} \text{ (row)}}$$

Convention: In this book, vectors are always columns. We write $\mathbf{v}^\top$ explicitly when a row is needed.

Common Mistakes

1. **Ignoring order.** $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \neq \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Vectors are **ordered** lists.

2. **Mixing dimensions.** $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}^2$ and $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \in \mathbb{R}^3$ are not equal — they live in different spaces.
3. **Confusing a vector with its transpose.** A column vector $\mathbf{v} \in \mathbb{R}^{n \times 1}$ and the row vector $\mathbf{v}^\top \in \mathbb{R}^{1 \times n}$ are different objects with different shapes. This matters the moment you multiply.
4. **Thinking you need to visualize.** Vectors in $\mathbb{R}^{784}$ are perfectly well-defined. Geometry is a crutch for low dimensions, not a requirement.
5. **Confusing the zero vector with the scalar 0.** $\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ is a vector; 0 is a number. They are different types of objects.

🌍 Use Case: Data Is Vectors

Almost every kind of data is naturally a vector:

Data	Dimension
RGB pixel	$\mathbb{R}^3$
Grayscale image (28×28)	$\mathbb{R}^{784}$
Audio (1 sec, 44.1 kHz)	$\mathbb{R}^{44100}$
Word embedding	$\mathbb{R}^{300}$
Stock portfolio (500 stocks)	$\mathbb{R}^{500}$

Once data is a vector, all the tools of linear algebra — norms, projections, transformations, decompositions — apply automatically.

🧠 Why This Matters for Neural Networks

A neural network's input is **always** a vector. An image is flattened into $\mathbf{x} \in \mathbb{R}^{784}$. A word becomes an embedding vector in $\mathbb{R}^{300}$. A sentence becomes a sequence of such vectors.

When you hear "a neural network processes an input," what actually happens is: a column vector $\mathbf{x} \in \mathbb{R}^n$ enters, gets multiplied by weight matrices, and a new column vector $\mathbf{y} \in \mathbb{R}^m$ comes out. The entire architecture is a chain of operations on column vectors and matrices.

The standard basis vectors $\mathbf{e}_i$ also appear directly: a "one-hot encoding" of a class label (e.g., digit 3 out of 10 classes) is exactly $\mathbf{e}_4 \in \mathbb{R}^{10}$ (using 0-indexed classes).

🔗 Connections

- **Chapter 2** defines addition and scalar multiplication — the two operations that make $\mathbb{R}^n$ a vector space.
- **Chapter 3** introduces the dot product, giving vectors a notion of length, distance, and angle.
- **Chapter 4** asks: when does a set of vectors "span" all of $\mathbb{R}^n$? The standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is the simplest example.
- **Chapter 5** stacks vectors into matrices — the row vs. column distinction from Example 10 becomes critical.

🔧 **Troubleshooting**

Problem	Likely Cause	Fix
"Is my vector a row or column?"	Ambiguous notation	Default to column. Write $\mathbf{v}^\top$ explicitly for rows.
"What dimension is $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$?"	Confusing components with dimension	Count entries: 3 entries $\Rightarrow \mathbb{R}^3$
"Can I add vectors of different sizes?"	Dimension mismatch	No — vectors must have the same dimension for any operation
"Is $\mathbf{0}$ the number 0?"	Confusing scalar and vector	$\mathbf{0}$ is a vector of zeros; 0 is a scalar
"How do I visualize $\mathbb{R}^{100}$?"	Expecting geometry	You don't — the algebra works without pictures

---

Summary

Concept	Definition
$\mathbb{R}^n$	The set of all n-component column vectors

Concept	Definition
Vector $\mathbf{v}$	A column $\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$
Component v_i	The i-th entry of $\mathbf{v}$
Zero vector $\mathbf{0}$	Column of all zeros; the additive identity
Equality	$\mathbf{u} = \mathbf{v}$ iff same dimension and $u_i = v_i$ for all i
Standard basis $\mathbf{e}_i$	Column with 1 in position i, 0 elsewhere
Position vector	Arrow from origin to a point
Displacement vector	$\overrightarrow{AB} = B - A$; encodes the change from A to B
Column vs. row	$\mathbf{v} \in \mathbb{R}^{n\times 1}$ (column) vs. $\mathbf{v}^\top \in \mathbb{R}^{1\times n}$ (row)

Exercises

Components and Equality

1. Write $\mathbf{v} = \begin{pmatrix} 3 \\ -1 \\ 0 \\ 7 \end{pmatrix}$ and identify v_3.

2. Find all values of a and b such that $\begin{pmatrix} a^2 \\ 2b-1 \end{pmatrix} = \begin{pmatrix} 9 \\ 5 \end{pmatrix}$.

3. Find x, y, z satisfying $\begin{pmatrix} 2x-y \\ x+z \\ y-3z \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ -7 \end{pmatrix}$.

4. True or false: a column vector $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and a row vector $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$ represent the same vector. Explain.

5. True or false: $\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. Explain.

Standard Basis

6. Write the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$ in $\mathbb{R}^4$.

7. Express $\begin{pmatrix} 5 \\ -2 \\ 0 \\ 7 \end{pmatrix}$ as a linear combination of the standard basis vectors in $\mathbb{R}^4$.

8. In $\mathbb{R}^n$, what is $\mathbf{e}_i + \mathbf{e}_j$ when $i \neq j$? When $i = j$?

9. How many standard basis vectors does $\mathbb{R}^{100}$ have?

Displacement Vectors

10. Points $P = (2, -1, 3)$ and $Q = (5, 0, -1)$ are given. Find $\overrightarrow{PQ}$ and $\overrightarrow{QP}$ as column vectors. What is their relationship?

11. A particle is at position $\begin{pmatrix} 1 \\ 4 \end{pmatrix}$. It undergoes displacement $\mathbf{d}_1 = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$ and then $\mathbf{d}_2 = \begin{pmatrix} -1 \\ 5 \end{pmatrix}$. Find the final position and the single displacement that achieves the same result.

12. Prove that for any three points A, B, C in $\mathbb{R}^n$: $\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CA} = \mathbf{0}$.

Data as Vectors

13. A color image has 64×64 pixels, each with R, G, B channels (values 0–255). What space does the flattened image vector live in?

14. A recommender system represents each user by their ratings of 1000 movies (0 if unrated). What is the dimension of the user vector?

15. Two audio clips are each 3 seconds long, sampled at 44,100 Hz. As vectors, what space do they live in? Under what condition are they the same clip?

Conceptual

16. Explain why the zero vector in $\mathbb{R}^2$ and the zero vector in $\mathbb{R}^3$ are different objects.

17. Give an example showing that the map $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$ does not satisfy $f(\mathbf{v}) = \mathbf{v}$, even though it "looks like" it preserves the vector.

18. The "one-hot" encoding of category k out of K categories is $\mathbf{e}_k \in \mathbb{R}^K$. How many nonzero components does a one-hot vector have? Why might this representation be useful?

19. If $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^m$ with $n \neq m$, can $\mathbf{v} = \mathbf{w}$? Can $\mathbf{v} + \mathbf{w}$ be defined? Justify.

20. Write down a vector $\mathbf{v} \in \mathbb{R}^6$ that could represent a student's performance across 6 courses. What does each component mean? What would $\mathbf{v} = \mathbf{0}$ mean in this context?

Vector Arithmetic

Chapter 1 defined vectors as ordered lists of numbers. This chapter introduces the three operations you can perform on them: **addition**, **subtraction**, and **scalar multiplication**. These three operations combine into one master idea – the **linear combination** – which is arguably the single most important concept in all of linear algebra.

Every operation is defined component-by-component. That makes the algebra mechanical: once you can add and scale individual numbers, you can add and scale vectors of any dimension. The geometry, meanwhile, gives you a powerful visual check. Addition is "tip-to-tail" placement of arrows. Scalar multiplication stretches or shrinks an arrow (and may flip its direction). Together, they let you reach any point that a given set of vectors can "build."

📋 **Quick Reference: Vector Arithmetic**

Vector Addition. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$:

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix}$$

Vector Subtraction. $\mathbf{u} - \mathbf{v} = \begin{pmatrix} u_1 - v_1 \\ u_2 - v_2 \\ \vdots \\ u_n - v_n \end{pmatrix}$

Scalar Multiplication. For $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$:

$$c\,\mathbf{v} = \begin{pmatrix} c\,v_1 \\ c\,v_2 \\ \vdots \\ c\,v_n \end{pmatrix}$$

Linear Combination. Given scalars $c_1, c_2, \ldots, c_k$ and vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \sum_{i=1}^{k} c_i\mathbf{v}_i$$

Key Properties (for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and scalars c, d):

Property	Statement
Commutativity	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
Associativity	$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
Additive identity	$\mathbf{v} + \mathbf{0} = \mathbf{v}$
Additive inverse	$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$, where $-\mathbf{v} = (-1)\mathbf{v}$
Scalar distributivity	$c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
Vector distributivity	$(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$
Scalar associativity	$c(d\mathbf{v}) = (cd)\mathbf{v}$
Multiplicative identity	$1\,\mathbf{v} = \mathbf{v}$

Geometric Rules:

- Addition: place arrows **tip-to-tail**; the sum is the diagonal of the parallelogram
- Subtraction: $\mathbf{u} - \mathbf{v}$ points **from the tip of $\mathbf{v}$ to the tip of $\mathbf{u}$**
- Scalar $c > 0$: stretches by factor c, same direction

> - Scalar $c < 0$: stretches by $|c|$, **reverses** direction
> - Scalar $c = 0$: collapses to $\mathbf{0}$

Vector Addition: The Geometric Picture

Vector addition has a clean geometric interpretation called the **tip-to-tail rule**. To add $\mathbf{u} + \mathbf{v}$:

1. Draw $\mathbf{u}$ starting at the origin.
2. Draw $\mathbf{v}$ starting at the **tip** of $\mathbf{u}$.
3. The sum $\mathbf{u} + \mathbf{v}$ is the arrow from the origin to the tip of the relocated $\mathbf{v}$.

Equivalently, $\mathbf{u} + \mathbf{v}$ is the diagonal of the parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$. To see this, place both vectors at the origin. Complete the parallelogram by drawing a copy of $\mathbf{v}$ starting at the tip of $\mathbf{u}$, and a copy of $\mathbf{u}$ starting at the tip of $\mathbf{v}$. The diagonal from the origin to the far corner is $\mathbf{u} + \mathbf{v}$. This is called the **parallelogram rule** for addition.

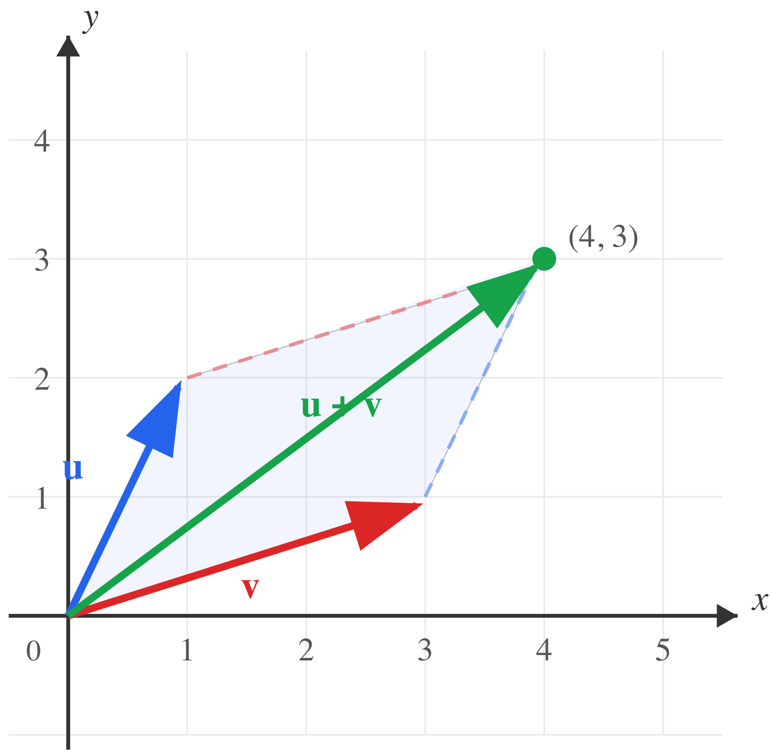
Why the geometric picture matches the algebra. Consider two vectors in $\mathbb{R}^2$: $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$. Algebraically, $\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \end{pmatrix}$. Geometrically, if you walk from the origin to the tip of $\mathbf{u}$ (moving u_1 in the x-direction and u_2 in the y-direction), and then walk along $\mathbf{v}$ (moving an additional v_1 in x and v_2 in y), you arrive at $(u_1 + v_1,\ u_2 + v_2)$. The tip-to-tail rule and the component-wise formula describe the same destination.

Commutativity and the parallelogram. The tip-to-tail rule immediately shows that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. If you go $\mathbf{u}$ then $\mathbf{v}$, you trace one path around the parallelogram; if you go $\mathbf{v}$ then $\mathbf{u}$, you trace the other path. Both paths end at the same corner. This is commutativity made visible: the order of addition does not matter.

Associativity. When adding three or more vectors, you can group them in any order. The tip-to-tail picture extends naturally: lay all the vectors end-to-end in sequence. The sum is the single arrow from the start of the first to the tip of the last. Since the final destination does not depend on the order or grouping, vector addition is both commutative and associative. This means expressions like $\mathbf{a} + \mathbf{b} + \mathbf{c}$ need no parentheses.

The additive identity and inverse. Adding the zero vector $\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ to any vector $\mathbf{v}$ leaves it unchanged: $\mathbf{v} + \mathbf{0} = \mathbf{v}$. Geometrically, the zero vector has no length and no direction, so "tip-to-tailing" it adds nothing. Every vector $\mathbf{v}$ also has an **additive inverse** $-\mathbf{v} = (-1)\mathbf{v}$, which points in the opposite direction with the same length. Adding a vector to its inverse gives the zero vector: $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$. Geometrically, you walk forward and then walk exactly the same distance backward, returning to the origin.

Higher dimensions. Everything above works identically in $\mathbb{R}^3$, $\mathbb{R}^4$, and beyond. In $\mathbb{R}^3$ you can still draw the parallelogram rule; in $\mathbb{R}^n$ for $n > 3$ you cannot draw a picture, but the algebra (adding component by component) is the same, and every property — commutativity, associativity, identity, and inverse — carries over unchanged.



Example 1: Vector Addition – Geometric (Tip-to-Tail)

Problem: Add $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ using the tip-to-tail method.

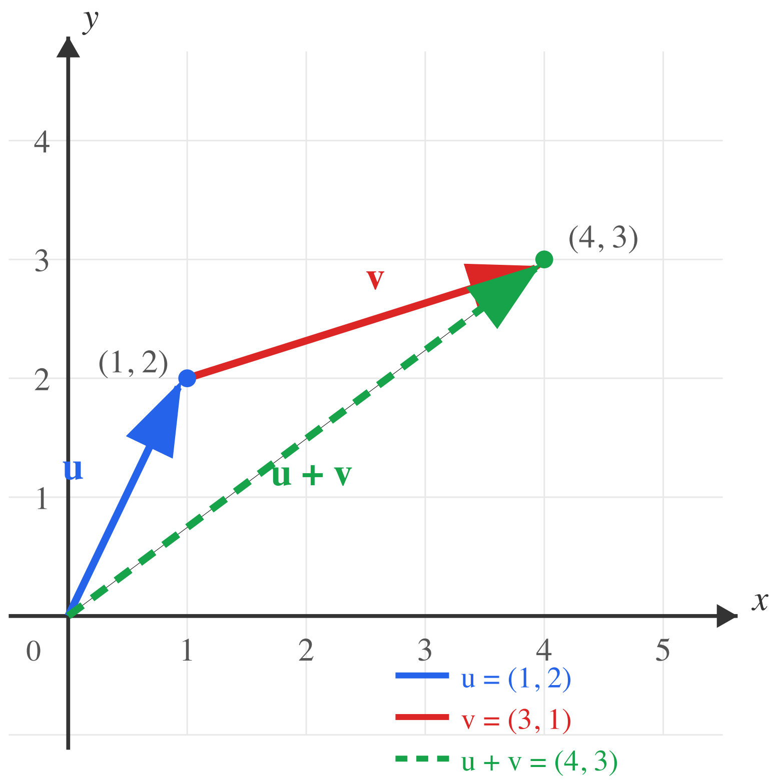
Step 1: Draw $\mathbf{u}$ from the origin.

The arrow goes from $(0, 0)$ to $(1, 2)$.

Step 2: Place $\mathbf{v}$ at the tip of $\mathbf{u}$.

Starting at $(1, 2)$, move 3 right and 1 up, arriving at $(1 + 3,\ 2 + 1) = (4, 3)$.

Step 3: Draw the sum from the origin to the new tip.



$$\boxed{\mathbf{u} + \mathbf{v} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}}$$

Verify: The parallelogram diagonal connects $(0, 0)$ to $(4, 3)$. Both paths around the parallelogram ($\mathbf{u}$ then $\mathbf{v}$, or $\mathbf{v}$ then $\mathbf{u}$) arrive at the same point, confirming commutativity. ✓

Example 2: Vector Addition – Algebraic (Component-Wise)

Problem: Compute $\mathbf{u} + \mathbf{v}$ algebraically for $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$.

Step 1: Apply the definition – add corresponding components.

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1+3 \\ 2+1 \end{pmatrix}$$

Step 2: Simplify.

$$\boxed{\mathbf{u} + \mathbf{v} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}}$$

Verify: This matches the geometric result from Example 1. The algebraic and geometric approaches always agree – that is the whole point. ✓

Example 3: Adding Three Vectors in R³

Problem: Compute $\mathbf{a} + \mathbf{b} + \mathbf{c}$ where

$$\mathbf{a} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} 0 \\ 5 \\ 1 \end{pmatrix}$$

Step 1: Add component-by-component.

$$\mathbf{a} + \mathbf{b} + \mathbf{c} = \begin{pmatrix} 1 + 3 + 0 \\ \\ 0 + (-1) + 5 \\ \\ -2 + 4 + 1 \end{pmatrix}$$

Step 2: Simplify each component.

$$\boxed{\mathbf{a} + \mathbf{b} + \mathbf{c} = \begin{pmatrix} 4 \\ \\ 4 \\ \\ 3 \end{pmatrix}}$$

Verify: Check each component independently: $1 + 3 + 0 = 4$, $0 - 1 + 5 = 4$, $-2 + 4 + 1 = 3$. ✓

Note: addition is associative, so we can group in any order: $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$. No parentheses are needed.

Vector Subtraction: The Displacement Interpretation

Subtraction $\mathbf{u} - \mathbf{v}$ computes the vector **from the tip of $\mathbf{v}$ to the tip of $\mathbf{u}$**. Think of it as: “What displacement takes me from $\mathbf{v}$ to $\mathbf{u}$?”

Algebraically, $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$. You negate every component of $\mathbf{v}$ and then add. Component by component:

$$\mathbf{u} - \mathbf{v} = \begin{pmatrix} u_1 - v_1 \\ u_2 - v_2 \\ \vdots \\ u_n - v_n \end{pmatrix}$$

The geometric picture. Draw both $\mathbf{u}$ and $\mathbf{v}$ starting from the origin. The difference $\mathbf{u} - \mathbf{v}$ is the arrow that connects the tip of $\mathbf{v}$ to the tip of $\mathbf{u}$. This makes sense because if you start at the tip of $\mathbf{v}$ and add the displacement $\mathbf{u} - \mathbf{v}$, you arrive at the tip of $\mathbf{u}$: $\mathbf{v} + (\mathbf{u} - \mathbf{v}) = \mathbf{u}$.

Notice that subtraction is closely tied to the parallelogram from addition. The parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$ has two diagonals. One diagonal is $\mathbf{u} + \mathbf{v}$; the other diagonal, connecting the tips of $\mathbf{v}$ and $\mathbf{u}$, is $\mathbf{u} - \mathbf{v}$ (or $\mathbf{v} - \mathbf{u}$, depending on direction).

Order matters in subtraction. Unlike addition, subtraction is **not commutative**: $\mathbf{u} - \mathbf{v} \neq \mathbf{v} - \mathbf{u}$ in general. Instead, $\mathbf{u} - \mathbf{v} = -(\mathbf{v} - \mathbf{u})$. Geometrically, these two differences are the same arrow but pointing in opposite directions. The displacement from $\mathbf{v}$ to $\mathbf{u}$ is the exact reverse of the displacement from $\mathbf{u}$ to $\mathbf{v}$.

Why subtraction is not a "new" operation. Subtraction is defined entirely in terms of addition and scalar multiplication: $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$. This means every property of vector arithmetic can be derived from just two primitives (addition and scalar multiplication). Subtraction is simply a convenient shorthand.

Displacement in applications. Subtraction appears constantly in practice. If $\mathbf{p}$ and $\mathbf{q}$ are position vectors of two points, then $\mathbf{q} - \mathbf{p}$ is the displacement from $\mathbf{p}$ to $\mathbf{q}$ — the direction and distance you would need to travel. The length of this displacement vector gives the distance between the points (Chapter 3).

Example 4: Vector Subtraction – The Displacement Interpretation

Problem: Compute $\mathbf{u} - \mathbf{v}$ where $\mathbf{u} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, and interpret geometrically.

Step 1: Subtract component-by-component.

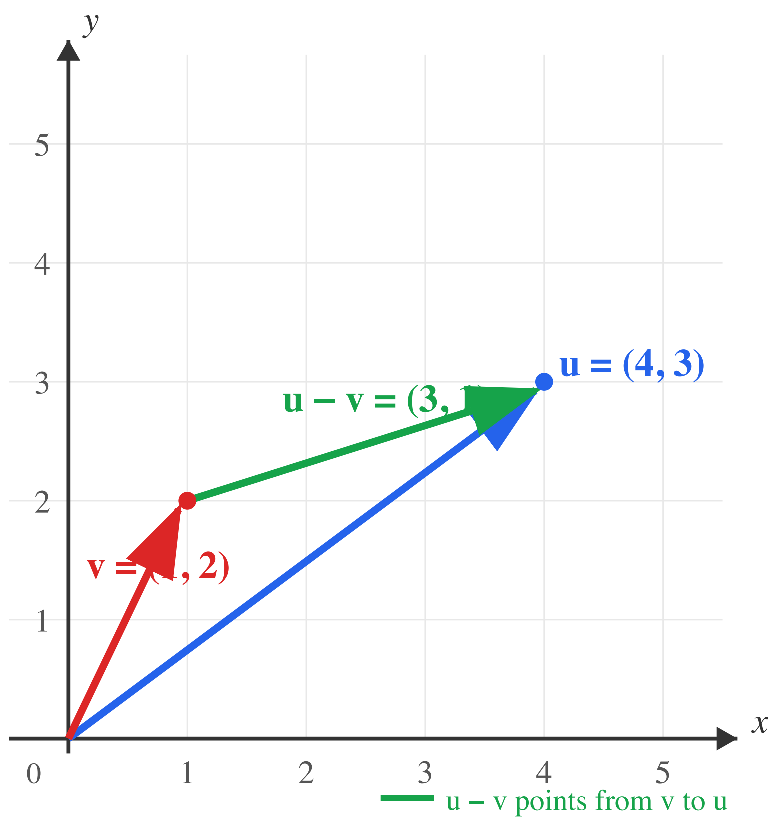
$$\mathbf{u} - \mathbf{v} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 - 1 \\ 3 - 2 \end{pmatrix}$$

Step 2: Simplify.

$$\boxed{\mathbf{u} - \mathbf{v} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}}$$

Step 3: Geometric interpretation.

The vector $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ points from the tip of $\mathbf{v}$ at $(1, 2)$ to the tip of $\mathbf{u}$ at $(4, 3)$. It answers the question: “How do I get from $\mathbf{v}$ to $\mathbf{u}$?”



Verify: $\mathbf{v} + (\mathbf{u} - \mathbf{v}) = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} = \mathbf{u}$. Adding the difference to $\mathbf{v}$ recovers $\mathbf{u}$. ✓

Scalar Multiplication

Multiplying a vector by a scalar c scales every component by c:

$$c\,\mathbf{v} = c\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} c\,v_1 \\ c\,v_2 \\ \vdots \\ c\,v_n \end{pmatrix}$$

The word “scalar” here just means a real number — a single value, not a vector. The idea is that a scalar *scales* the vector: it changes the vector’s size (and possibly its direction) without changing the line it lies on.

Geometric effects by case. The scalar c controls both the length and the direction of the result:

- If $c > 1$, the arrow **stretches** (gets longer), same direction.
- If $0 < c < 1$, the arrow **shrinks** (gets shorter), same direction.
- If $c = 1$, the vector is unchanged: $1\,\mathbf{v} = \mathbf{v}$.
- If $c = 0$, the result is the zero vector $\mathbf{0}$, regardless of $\mathbf{v}$.
- If $-1 < c < 0$, the arrow **reverses direction** and shrinks.
- If $c = -1$, the arrow **reverses direction** with the same length: $-\mathbf{v}$.
- If $c < -1$, the arrow **reverses direction** and stretches.

In every case, the result $c\mathbf{v}$ lies on the same line through the origin as $\mathbf{v}$. Scalar multiplication can move you forward or backward along that line, and closer to or farther from the origin, but it can never move you off the line.

The length of a scaled vector. When you multiply by c, the length of the vector gets multiplied by $|c|$ (the absolute value of c). That is, $\|c\,\mathbf{v}\| = |c|\ \|\mathbf{v}\|$. The absolute value appears because direction reversal does not change length. Multiplying by -3 triples the length and flips the direction, just as multiplying by $+3$ triples the length without flipping.

Key algebraic properties. Scalar multiplication interacts cleanly with addition:

- **Scalar distributivity:** $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$. Scaling a sum is the same as scaling each piece separately.

- **Vector distributivity:** $(c+d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$. Two scalars applied to the same vector can be combined.
- **Scalar associativity:** $c(d\,\mathbf{v}) = (cd)\mathbf{v}$. Scaling by c and then by d is the same as scaling once by cd.
- **Multiplicative identity:** $1\,\mathbf{v} = \mathbf{v}$. The scalar 1 changes nothing.

These properties might look obvious, but they are exactly the rules that make "linear" algebra work. They ensure that scaling and adding can be freely rearranged, which is what makes linear combinations (next section) so powerful.

A subtle but important point. Scalar multiplication is *not* the same as multiplying two vectors. There is no general "vector times vector" operation that produces another vector of the same type. (The dot product in Chapter 3 multiplies two vectors, but it produces a scalar, not a vector.) When you see $c\mathbf{v}$, the scalar c and the vector $\mathbf{v}$ play fundamentally different roles.

Example 5: Scalar Multiplication – Stretching

Problem: Compute $3\mathbf{v}$ where $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

Step 1: Multiply each component by 3.

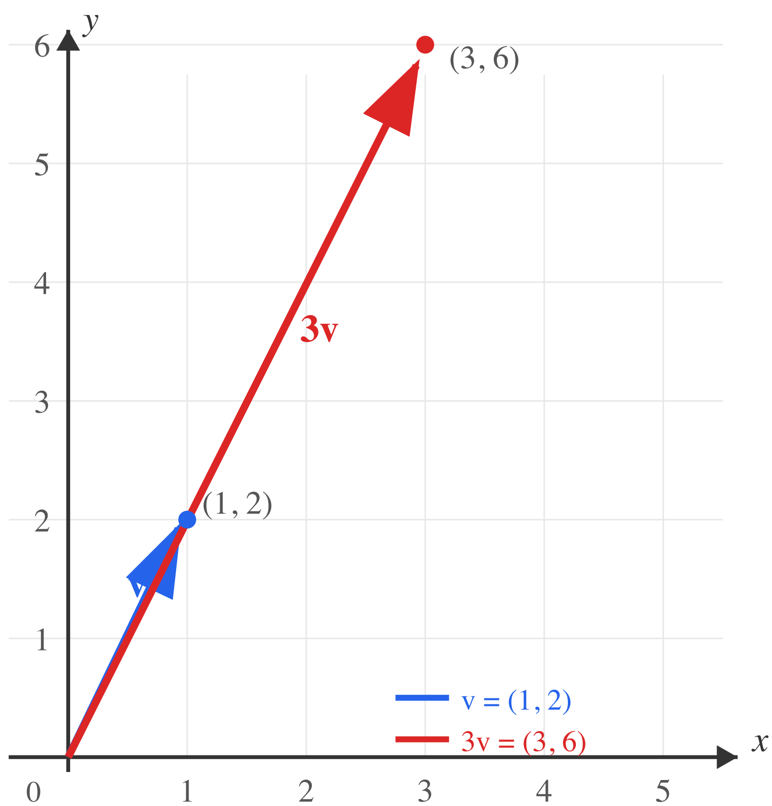
$$3\mathbf{v} = 3\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \cdot 1 \\ 3 \cdot 2 \end{pmatrix}$$

Step 2: Simplify.

$$\boxed{3\mathbf{v} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}}$$

Step 3: Geometric interpretation.

The arrow $3\mathbf{v}$ points in the same direction as $\mathbf{v}$ but is 3 times as long. Both vectors lie on the same line through the origin.



Verify: $\frac{3}{6} = \frac{1}{2}$ and $\frac{3}{6} = \frac{1}{2}$ – the components of $3\mathbf{v}$ are in the same ratio as those of $\mathbf{v}$, confirming the direction is unchanged. ✓

Example 6: Scalar Multiplication – Flipping Direction

Problem: Compute $-2\mathbf{v}$ where $\mathbf{v} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$.

Step 1: Multiply each component by -2.

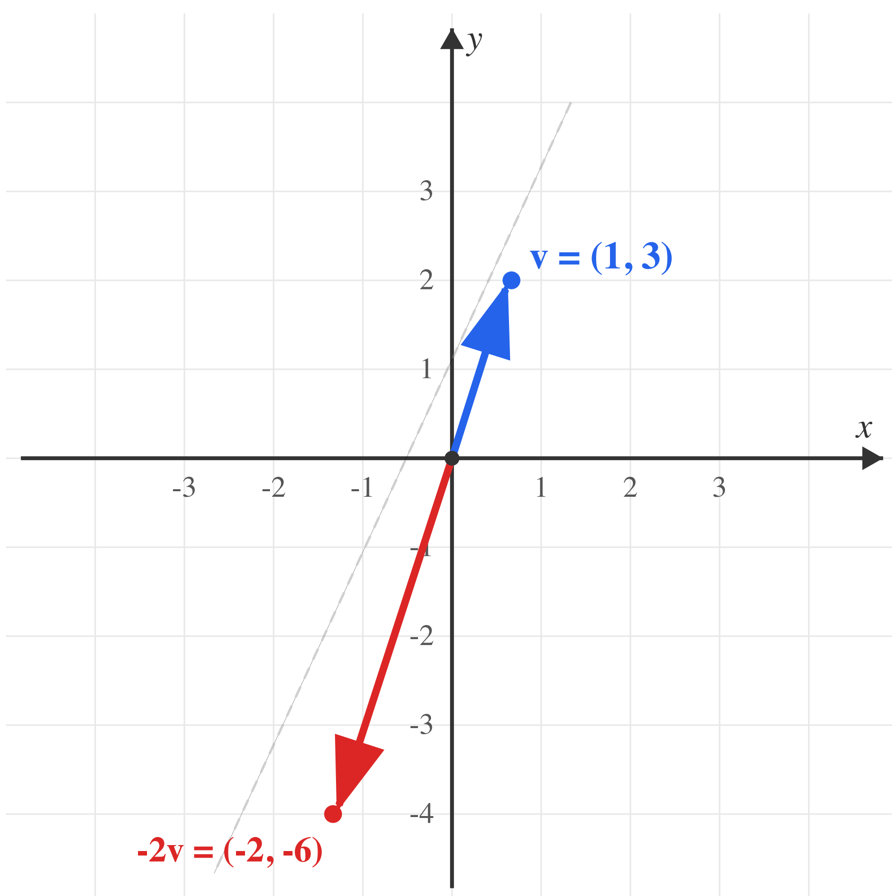
$$-2\mathbf{v} = -2\begin{pmatrix}1\\3\end{pmatrix} = \begin{pmatrix}-2\cdot 1\\-2\cdot 3\end{pmatrix}$$

Step 2: Simplify.

$$\boxed{-2\mathbf{v} = \begin{pmatrix}-2\\-6\end{pmatrix}}$$

Step 3: Geometric interpretation.

The negative scalar flips the direction: $-2\mathbf{v}$ points opposite to $\mathbf{v}$. The factor of 2 doubles the length. The result is an arrow twice as long as $\mathbf{v}$, pointing in the opposite direction, but still on the same line through the origin.



Verify: $-2\mathbf{v} + 2\mathbf{v} = \begin{pmatrix} -2 \\ -6 \end{pmatrix} + \begin{pmatrix} 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0}$. Opposite vectors cancel to zero. ✓

Linear Combinations

A **linear combination** of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is any expression of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

where $c_1, c_2, \ldots, c_k$ are scalars (real numbers). The scalars are called the **weights** or **coefficients** of the combination.

In compact notation using the summation symbol:

$$\sum_{i=1}^{k} c_i\mathbf{v}_i = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

Why this concept is central. Linear combinations are the central building block of linear algebra. Almost every major concept — span, linear independence, basis, matrix-vector products, solutions to linear systems — is defined in terms of linear combinations. When we ask "Can I reach this target vector from these building blocks?" we are asking whether the target is a linear combination of the building blocks. When we ask "Is this vector redundant?" we are asking whether it is a linear combination of the others. The language of linear combinations unifies all of these questions.

The geometry of a linear combination. A linear combination works in two stages: first **scale**, then **add**. You take each ingredient vector $\mathbf{v}_i$, stretch or shrink it by the scalar c_i (possibly flipping its direction if $c_i < 0$), and then add all the resulting vectors using the tip-to-tail rule. The final arrow points to the result.

By varying the scalars, you sweep out different results. The set of *all possible* linear combinations of a given collection of vectors is called the **span** of those vectors (Chapter 4). For example, all linear combinations of a single nonzero vector trace out a line through the origin. All linear combinations of two independent vectors fill out a plane through the origin. All linear combinations of three independent vectors in $\mathbb{R}^3$ fill out all of 3-space.

Special cases worth noting.

- If all coefficients are zero ($c_1 = c_2 = \cdots = c_k = 0$), the linear combination gives the zero vector $\mathbf{0}$. The zero vector is always a linear combination of any set of vectors.
- If exactly one coefficient is 1 and the rest are 0, the linear combination just picks out one of the ingredient vectors.
- If the coefficients sum to 1 ($c_1 + c_2 + \cdots + c_k = 1$) and are all non-negative, the result is called a **convex combination** — it lies "between" the ingredient vectors. This special case is important in optimization and geometry, but in this course we allow any real-valued coefficients.

The standard basis as a universal example. Every vector in $\mathbb{R}^n$ is a linear combination of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, and the coefficients are simply the components of the vector: $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n$. This is the simplest example of a linear combination, and it shows that the concept of "components" and the concept of "coefficients in a linear combination" are really the same idea, just expressed in different language.

Checking whether a vector is a linear combination. Given a target vector $\mathbf{b}$ and ingredient vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, asking "Is $\mathbf{b}$ a linear combination of the $\mathbf{v}_i$?" amounts to asking whether there exist scalars $c_1, \ldots, c_k$ satisfying $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{b}$. This is a system of equations (one equation per component). If the system has a solution, the answer is yes; if it has no solution, the answer is no. We will see many examples of this procedure in the worked problems that follow.

Example 7: A Linear Combination of Standard Basis Vectors

Problem: Compute $2\mathbf{e}_1 + 3\mathbf{e}_2$ in $\mathbb{R}^2$.

Step 1: Recall the standard basis vectors.

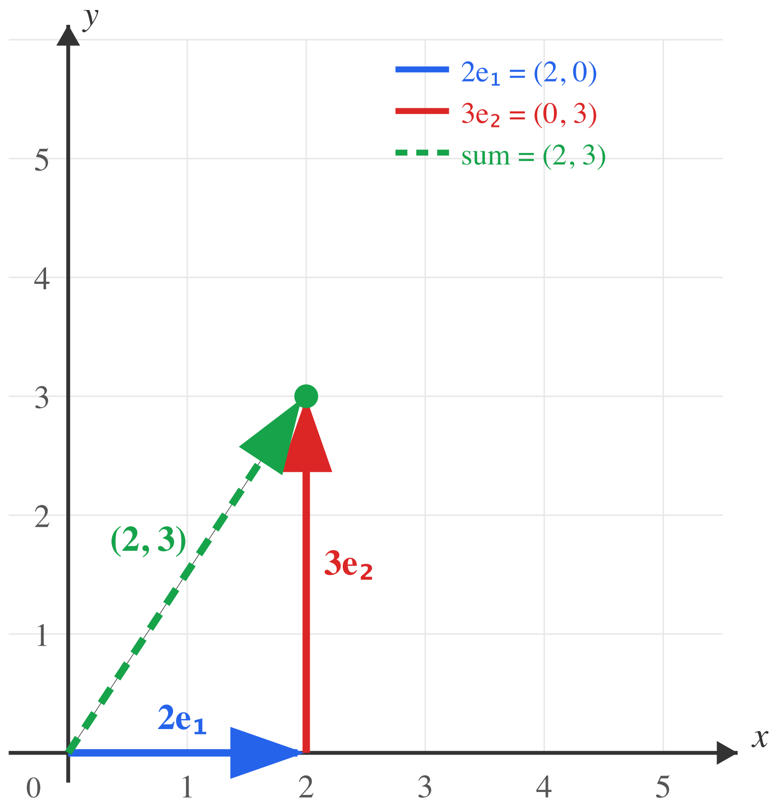
$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Step 2: Scale each vector.

$$2\mathbf{e}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad 3\mathbf{e}_2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

Step 3: Add.

$$2\mathbf{e}_1 + 3\mathbf{e}_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$



$$\mathbf{2e}_1 + \mathbf{3e}_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Verify: The vector $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ has first component 2 (from $2\mathbf{e}_1$) and second component 3 (from $3\mathbf{e}_2$). This is exactly how the standard basis works: the coefficient of $\mathbf{e}_i$ becomes the i-th component. ✓

Example 8: Solving for Coefficients in a Linear Combination

Problem: Find scalars a and b such that

$$a \begin{pmatrix} 1 \\ 1 \end{pmatrix} + b \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

Step 1: Expand the left side.

$$\begin{pmatrix} a \\ a \end{pmatrix} + \begin{pmatrix} b \\ -b \end{pmatrix} = \begin{pmatrix} a+b \\ a-b \end{pmatrix}$$

Step 2: Set up the system by equating components.

$$a + b = 3$$

$$a - b = 1$$

Step 3: Solve the system.

Add the two equations: $2a = 4$, so $a = 2$.

Substitute back: $2 + b = 3$, so $b = 1$.

$$\boxed{a = 2, \quad b = 1}$$

Verify: $2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$. ✓

Example 9: Is a Vector a Linear Combination of Two Others?

Problem: Determine whether $\begin{pmatrix} 5 \\ 3 \end{pmatrix}$ is a linear combination of $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Step 1: Set up the equation.

We need scalars c_1, c_2 such that

$$c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$$

Step 2: Write the component equations.

$$c_1 + 2c_2 = 5$$
$$c_1 + c_2 = 3$$

Step 3: Solve.

Subtract the second equation from the first: $c_2 = 2$.

Substitute back: $c_1 + 2 = 3$, so $c_1 = 1$.

$$\boxed{\begin{pmatrix} 5 \\ 3 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ 1 \end{pmatrix}}$$

Yes, $\begin{pmatrix} 5 \\ 3 \end{pmatrix}$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.

Verify: $1\begin{pmatrix}1\\1\end{pmatrix}+2\begin{pmatrix}2\\1\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}+\begin{pmatrix}4\\2\end{pmatrix}=\begin{pmatrix}5\\3\end{pmatrix}$. ✓

Example 10: A Linear Combination in R³

Problem: Is $\begin{pmatrix}1\\1\\0\end{pmatrix}$ a linear combination of $\mathbf{v}_1=\begin{pmatrix}1\\0\\0\end{pmatrix}$ and $\mathbf{v}_2=\begin{pmatrix}0\\1\\0\end{pmatrix}$?

Step 1: Set up the equation.

$$c_1\begin{pmatrix}1\\0\\0\end{pmatrix}+c_2\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}1\\1\\0\end{pmatrix}$$

Step 2: Write the component equations.

$$c_1=1,\quad c_2=1,\quad 0=0$$

Step 3: Solve.

The first two equations give $c_1=1$ and $c_2=1$. The third equation $0=0$ is automatically satisfied.

$$\boxed{\begin{pmatrix}1\\1\\0\end{pmatrix}=1\begin{pmatrix}1\\0\\0\end{pmatrix}+1\begin{pmatrix}0\\1\\0\end{pmatrix}=\mathbf{e}_1+\mathbf{e}_2}$$

Yes, it is a linear combination. The target vector lies in the xy-plane, which is exactly where combinations of $\mathbf{e}_1$ and $\mathbf{e}_2$ can reach.

Verify: $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. ✓

Example 11: Expressing a Vector in the Standard Basis

Problem: Write $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ as a linear combination of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in $\mathbb{R}^3$.

Step 1: Recall the general formula.

Any vector $\mathbf{v} \in \mathbb{R}^n$ can be written $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n$. The coefficient of $\mathbf{e}_i$ is the i-th component.

Step 2: Read off the coefficients.

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\boxed{\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1\,\mathbf{e}_1 + 2\,\mathbf{e}_2 + 3\,\mathbf{e}_3}$$

Verify: $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. ✓

This is always trivial for the standard basis – the components are the coefficients. For other bases (Chapter 4), finding the coefficients requires solving a system of equations.

Example 12: A Linear Combination That Fails

Problem: Determine whether $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ is a linear combination of $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

Step 1: Set up the equation.

$$c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Step 2: Write the component equations.

$$c_1 = 1, \quad c_2 = 2, \quad 0 = 3$$

Step 3: Check for consistency.

The third equation $0 = 3$ is a contradiction. No choice of c_1, c_2 can produce a nonzero third component from vectors whose third components are both zero.

$$\boxed{\text{No solution} - \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ is NOT a linear combination of } \mathbf{v}_1 \text{ and } \mathbf{v}_2}$$

Why it fails: The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ span only the xy-plane ($z = 0$). The target vector has $z = 3$, so it lies outside this plane. No amount of scaling and adding $\mathbf{v}_1$ and $\mathbf{v}_2$ can ever produce a nonzero z-component. We would need a third vector with a nonzero third component (such as $\mathbf{e}_3$).

Verify: Any combination $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ has the form $\begin{pmatrix} c_1 \\ c_2 \\ 0 \end{pmatrix}$, which always has third component 0. ✓

Common Mistakes

1. **Adding vectors of different dimensions.** $\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$ is **undefined**.

   Vectors must have the same number of components. Always check dimensions before operating.

2. **Multiplying two vectors component-wise and calling it "vector multiplication."** The expression $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \end{pmatrix}$ is not a standard linear algebra operation. The dot product (Chapter 3) produces a **scalar**, not a vector.

3. **Forgetting to distribute the scalar to every component.** A common error: $3 \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix}$. Wrong! Every component gets multiplied: $\begin{pmatrix} 3 \\ 6 \\ 12 \end{pmatrix}$.

4. **Confusing $\mathbf{u} - \mathbf{v}$ with $\mathbf{v} - \mathbf{u}$.** These point in opposite directions: $\mathbf{u} - \mathbf{v} = -(\mathbf{v} - \mathbf{u})$. The order determines which way the displacement points.

5. **Thinking a negative scalar only changes the sign.** $-2\mathbf{v}$ both **reverses direction** and **doubles the length**. The minus sign flips; the 2 scales.

6. **Assuming every vector can be written as a linear combination of any two vectors.** Example 12 shows this is false. Two vectors in $\mathbb{R}^3$ span only a plane, not all of 3-space.

> 🌍 **Use Case: Linear Combinations in the Real World**

Linear combinations appear everywhere:

Domain	Vectors	Scalars	Combination
Physics	Force vectors	Magnitudes	Net force = sum of individual forces
Color mixing	RGB channels $\begin{pmatrix} R \\ G \\ B \end{pmatrix}$	Intensities	Any color = $r\,\mathbf{red} + g\,\mathbf{green} + b\,\mathbf{blue}$
Finance	Asset return vectors	Portfolio weights	Portfolio return = weighted sum of asset returns
Audio	Sound waveforms	Volumes	Mixed track = $v_1 \cdot \text{guitar} + v_2 \cdot \text{drums} + \cdots$
Chemistry	Molecular formulas	Reaction coefficients	Balanced equation = lin. comb. that yields $\mathbf{0}$

In each case, the pattern is the same: take a collection of “ingredient” vectors, choose how much of each to use (the scalars), and combine. The question “Can I build this target from these ingredients?” is always a linear combination problem.

🧠 Why This Matters for Neural Networks

A single neuron computes exactly one linear combination and then applies a nonlinear activation:

$$z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b$$

Here $x_1, \ldots, x_n$ are the inputs, $w_1, \ldots, w_n$ are the **weights** (the scalars in the linear combination), and b is a bias. In vector notation:

$$z = \mathbf{w}^\top \mathbf{x} + b$$

Training a neural network means adjusting the scalars $w_1, \ldots, w_n$ so that the linear combination (followed by the activation) produces the desired output. Every gradient descent update is itself a linear combination: $\mathbf{w}_{\text{new}} = \mathbf{w}_{\text{old}} - \eta \nabla L$, where η is the learning rate scalar and ∇L is the gradient vector.

A layer with m neurons computes m linear combinations simultaneously – one per neuron. This is exactly matrix-vector multiplication (Chapter 6): $\mathbf{z} = \mathbf{W}\mathbf{x} + \mathbf{b}$.

🔗 **Connections**

- **Chapter 1** defined vectors. This chapter defines what you can *do* with them.
- **Chapter 3** introduces the dot product, which measures the angle and “alignment” between vectors. The dot product is itself a linear combination of components: $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots$.
- **Chapter 4** asks the fundamental question: given a set of vectors, what is the set of *all* linear combinations you can form? That set is the **span**. When no vector in the set is a linear combination of the others, the set is **linearly independent**.
- **Chapter 6** shows that matrix-vector multiplication $\mathbf{A}\mathbf{x}$ is a linear combination of the columns of $\mathbf{A}$, with coefficients from $\mathbf{x}$.

- **Chapter 7** reveals that solving $\mathbf{A}\mathbf{x} = \mathbf{b}$ is equivalent to asking: “Is $\mathbf{b}$ a linear combination of the columns of $\mathbf{A}$?”

🔧 **Troubleshooting**

Problem	**Likely Cause**	**Fix**
“I get a dimension mismatch error”	Adding vectors from different $\mathbb{R}^n$	Check that all vectors have the same number of components
“My linear combination gives the wrong answer”	Forgot to distribute the scalar	Multiply the scalar into *every* component of the vector
“I can’t express the target as a linear combination”	The target is outside the span	You need more/different vectors (see Chapter 4)
“Is subtraction a separate operation?”	Conceptual confusion	No – $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$
“Does the order of addition matter?”	Worried about commutativity	No – vector addition is commutative: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
“The system has no solution”	Target not in span of given vectors	Verify with the contradiction method (Example 12)

Summary

Concept	Definition
Vector addition	$\mathbf{u}+\mathbf{v}=\begin{pmatrix}u_1+v_1\\ \vdots\\ u_n+v_n\end{pmatrix}$; tip-to-tail geometrically
Vector subtraction	$\mathbf{u}-\mathbf{v}=\begin{pmatrix}u_1-v_1\\ \vdots\\ u_n-v_n\end{pmatrix}$; displacement from $\mathbf{v}$ to $\mathbf{u}$
Scalar multiplication	$c\mathbf{v}=\begin{pmatrix}cv_1\\ \vdots\\ cv_n\end{pmatrix}$; stretches/shrinks/flips the arrow
Linear combination	$c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k$; scalars choose “how much” of each vector
Additive identity	$\mathbf{v}+\mathbf{0}=\mathbf{v}$
Additive inverse	$\mathbf{v}+(-\mathbf{v})=\mathbf{0}$
Commutativity	$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
Scalar distributivity	$c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$

Exercises

Addition and Subtraction

1. Compute $\begin{pmatrix} 2 \\ -3 \end{pmatrix} + \begin{pmatrix} 5 \\ 1 \end{pmatrix}$.

2. Compute $\begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ 3 \\ 7 \end{pmatrix}$.

3. Compute $\begin{pmatrix} 6 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 4 \end{pmatrix}$ and draw the result as a displacement vector.

4. Given $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} 4 \\ -1 \\ 0 \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} -2 \\ 3 \\ 1 \end{pmatrix}$, compute $\mathbf{u} + \mathbf{v} + \mathbf{w}$.

5. Is $\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$ defined? Why or why not?

Scalar Multiplication

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

7. Compute $-3\begin{pmatrix}-1\\2\end{pmatrix}$. Does the result point in the same or opposite direction as the original?

8. For what value of c does $c\begin{pmatrix}2\\5\end{pmatrix} = \begin{pmatrix}6\\15\end{pmatrix}$?

9. Compute $\frac{1}{2}\begin{pmatrix}8\\-6\\4\end{pmatrix}$. Geometrically, what does multiplying by $\frac{1}{2}$ do?

10. Show that $0 \cdot \mathbf{v} = \mathbf{0}$ for any $\mathbf{v} \in \mathbb{R}^3$ by writing out the component calculation.

Linear Combinations

11. Compute $3\begin{pmatrix}1\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix}$.

12. Find scalars a, b such that $a\begin{pmatrix}1\\2\end{pmatrix} + b\begin{pmatrix}3\\1\end{pmatrix} = \begin{pmatrix}7\\8\end{pmatrix}$.

13. Is $\begin{pmatrix} 4 \\ 6 \end{pmatrix}$ a linear combination of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$? Find the coefficients or prove no solution exists.

14. Write $\begin{pmatrix} 5 \\ -3 \\ 7 \end{pmatrix}$ as a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$.

15. Is $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ a linear combination of $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$? Justify your answer.

Properties and Proofs

16. Verify the commutative property $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for $\mathbf{u} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ by computing both sides.

17. Verify the distributive property $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ for $c = 5$, $\mathbf{u} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$ by computing both sides.

18. Prove that for any $\mathbf{v} \in \mathbb{R}^n$: $\mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$.

Conceptual

19. The vectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ point in the same direction. Can every vector in $\mathbb{R}^2$ be written as a linear combination of these two? Explain why or give a counterexample.

20. A neural network layer computes $\mathbf{z} = w_1\mathbf{x}_1 + w_2\mathbf{x}_2 + w_3\mathbf{x}_3$ where each $\mathbf{x}_i \in \mathbb{R}^4$. If $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$, and $w_1 = 0.5, w_2 = -1.0, w_3 = 2.0$, compute $\mathbf{z}$. Can this layer ever produce a vector with a nonzero fourth component? Why or why not?