Algebra I: Matrices and Applications

Gudfit

Contents

Chapter 1

Ideas & Motivations

Welcome to Algebra I: Matrices and Applications by me (Gudfit). The goal here is to build a clean bridge between the mechanics of solving equations and the geometry of high-dimensional space. Instead of jumping straight into abstraction, we build the theory constructively.

I aim for each set of notes to be max 50 pages, as rigorous as possible, and far-reaching too. That means I'll cover the axioms and proofs of the most interesting stuff plus I'll pull in concepts we've touched on in my informal logic and earlier notes to show how math builds on itself like funky Lego.

It'll be a mix of algorithmic recipes and deep structural theorems. The original idea was a dry, fully axiomatic intro, but that felt too grindy. Why slog through abstract definitions without motivation when we can derive them from solving actual problems? So this will be efficient, blending computation with deduction, assuming you've got some mathematical maturity. Either way, let's dive in and enjoy!

Chapter 2

Linear Equations

The central problem of linear algebra is the resolution of systems of linear equations. Practically every problem in the field eventually reduces to solving a single matrix equation:

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

Here, A is a known $m \times n$ matrix (an array of numbers with m rows and n columns), $\mathbf{b}$ is a known column of constants, and $\mathbf{x}$ is a column of unknowns that we seek to determine. Our primary objectives are threefold: to determine the *existence* of a solution, to establish its *uniqueness* (or lack thereof), and to provide a constructive *algorithm* for finding the solution set.

2.1 Systems of Linear Equations

A system of linear equations is a collection of linear constraints on a set of variables.

Definition 2.1.1. ***System of Linear Equations***. A system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ is a set of equations of the form:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{2.1}$$

where $a_{ij} \in \mathbb{R}$ are the coefficients and $b_i \in \mathbb{R}$ are the constant terms. If all $b_i = 0$, the system is termed *homogeneous*; otherwise, it is *inhomogeneous*.

The scalar a_{ij} denotes the coefficient in the i-th equation associated with the j-th variable. It is cumbersome to manipulate these equations in algebraic form. We therefore adopt matrix notation to handle the data efficiently.

Notation 2.1.1. Matrix Representation We define the coefficient matrix A, the column of unknowns $\mathbf{x}$, and the column of constants $\mathbf{b}$ as:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

The system (2.1) is concisely written as $A\mathbf{x} = \mathbf{b}$. To capture the entire system including the right-hand side, we form the *augmented matrix* $[A \mid \mathbf{b}]$:

$$[A \mid \mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

Remark. What is a solution? The *solution set* of a system of linear equations is the set of all ordered lists of numbers $(c_1, \ldots, c_n)$ that, when substituted for $(x_1, \ldots, x_n)$, satisfy every equation in the system simultaneously. If $S = \emptyset$, the system is inconsistent. If $|S| = 1$, the solution is unique. If $|S| > 1$, there are infinitely many solutions.

Geometric Interpretation

Before formalising the solution method, it is instructive to visualise the problem in low dimensions. Consider a system of two equations in two unknowns ($m = n = 2$):

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned}$$

Each equation defines a line in the Cartesian plane. The solution set consists of points (x_1, x_2) that lie on both lines simultaneously. There are exactly three possibilities:

(i) **Unique Solution:** The lines are not parallel and intersect at exactly one point. The system is consistent and independent.
(ii) **No Solution:** The lines are distinct but parallel. They never intersect. The system is *inconsistent*.
(iii) **Infinitely Many Solutions:** The lines are coincident (one equation is a scalar multiple of the other). Every point on the line is a solution. The system is consistent but dependent.

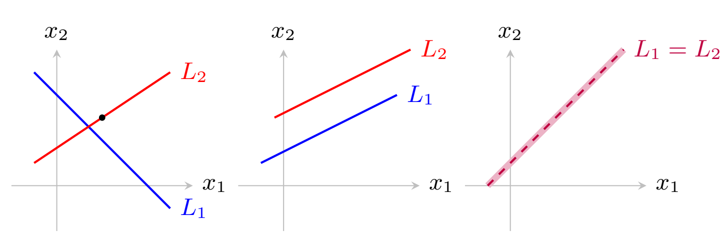
Example 2.1.1. Intersection of Lines. Consider the system:

$$\begin{aligned} 2x_1 + 3x_2 &= 1 \\ 3x_1 + 2x_2 &= 2 \end{aligned}$$

The slopes are $-2/3$ and $-3/2$. Since the slopes differ, the lines are not parallel and must intersect at a unique point. Solving yields $x_1 = 4/5, x_2 = -1/5$. Contrast this with:

$$\begin{aligned} 2x_1 + 3x_2 &= 1 \\ 4x_1 + 6x_2 &= 2 \end{aligned}$$

Here, the second equation is simply twice the first. The lines are identical, yielding infinitely many solutions. Finally, if the second equation were $4x_1 + 6x_2 = 5$, the lines would be parallel with distinct intercepts, resulting in no solution ($S = \emptyset$).



(i) Unique Solution (ii) No Solution (iii) Infinite Solutions

Figure 2.1: Geometric interpretation of linear systems in 2D. The solution set corresponds to the intersection of the lines defined by the equations.

While this geometric intuition is powerful in 2D or 3D, it fails to scale to higher dimensions or practical computational problems. We require an algebraic method that systematically simplifies the system without altering its solution set.

2.2 Gaussian Elimination

The standard algorithm for solving linear systems is Gaussian elimination. The strategy is to transform the system into a simpler form (specifically, an upper triangular form) where the solution can be found by back-substitution.

Remark. (Upper Triangular Structure). While the term "upper triangular" strictly applies to square matrices ($m = n$), Gaussian elimination generalises this structure to matrices of any shape ($m \times n$). We aim for a "staircase" pattern where the non-zero entries (pivots) cascade downwards and to the right.

- **Square Case** ($m = n$): The matrix reduces to a full triangle.

$$\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}$$

- **Overdetermined Case** ($m > n$): The matrix reduces to an upper triangle followed by zero rows.

$$\begin{bmatrix} * & * \\ 0 & * \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

In all cases, we seek a form where all entries below the pivots are zero.

The Logic of Equivalent Systems

We achieve this upper triangular form by manipulating the equations. However, we must ensure that these manipulations do not alter the set of valid solutions.

Definition 2.2.1. ***Linear Combination of Equations***. A *linear combination* of a set of equations $E_1, \ldots, E_m$ is a new equation formed by the sum:

$$\sum_{i=1}^{m} \lambda_i E_i = (\lambda_1 E_1 + \cdots + \lambda_m E_m),$$

where λ_i are scalars.

Two systems of equations are considered *equivalent* if they have the exact same solution set. A sufficient condition for equivalence is *mutual derivability*: if every equation in System A is a linear combination of equations in System B, and every equation in System B is a linear combination of equations in System A, the systems are equivalent.

Example 2.2.1. Equivalence vs. Extraneous Roots. To understand why Gaussian elimination works, we compare a valid linear transformation against a flawed algebraic one.

1. Valid Linear Combination (Reversible) Consider System I:

$$\begin{aligned} x + y &= 3 \quad (E_1) \\ x - y &= 1 \quad (E_2) \end{aligned}$$

If we replace E_2 with the sum $E_1 + E_2$, we obtain System II:

$$\begin{aligned} x + y &= 3 \quad (E_1) \\ 2x &= 4 \quad (E_3 = E_1 + E_2) \end{aligned}$$

Any solution to I is clearly a solution to II. Crucially, we can *reverse* the process to recover System I from System II. Since $E_2 = E_3 - E_1$, the information from the original system is preserved. Thus, System II is equivalent to System I (Solution: $x = 2, y = 1$).

2. Invalid Non-Linear Operation (Irreversible) Consider the equation:

$$\sqrt{x} = -1 \quad (E_{bad})$$

In the real numbers, this has no solution. If we square both sides (a non-linear operation), we derive:

$$x = 1 \quad (E_{new})$$

This new equation has the solution $x = 1$. However, checking back, $\sqrt{1} = 1 \neq -1$. The solution $x = 1$ is an *extraneous root*.

Remark. Gaussian elimination relies strictly on linear combinations (scaling and adding). Because these operations are reversible (we can always subtract the row back), we are guaranteed never to create extraneous roots or lose valid solutions. This is why we do not need to "check our answers" for existence in Linear Algebra in the same way we do for radical or rational equations.

To apply this logic systematically to matrices, we define operations that correspond exactly to these linear combinations.

Definition 2.2.2. ***Elementary Row Operations***. Let A be a matrix. The three types of elementary row operations are:

Type 1: **Scaling:** Multiply a row R_i by a non-zero scalar $\alpha \neq 0$.

$$R_i \to \alpha R_i.$$

Type 2: **Replacement:** Replace a row R_i with the sum of itself and a scalar multiple of another row R_j.

$$R_i \to R_i + \alpha R_j \quad (i \neq j).$$

Type 3: **Interchange:** Swap the positions of two rows R_i and R_j.

$$R_i \leftrightarrow R_j.$$

Two matrices A and B are said to be *row-equivalent*, denoted $A \sim B$, if one can be obtained from the other by a finite sequence of elementary row operations.

***Theorem* 2.2.1. Invariance of Solution Set**. If the augmented matrices of two linear systems are row-equivalent, then the systems have the exact same solution set.

Proof. Each elementary row operation is reversible.

- If $R_i' = \alpha R_i$ (with $\alpha \neq 0$), then $R_i = (1/\alpha) R_i'$.
- If $R_i' = R_i + \alpha R_j$, then $R_i = R_i' - \alpha R_j$ (since R_j is unchanged).
- If $R_i \leftrightarrow R_j$, applying the swap again restores the original order.

Since the operations are reversible, any solution satisfying the initial system satisfies the transformed system, and vice versa. ■

Example 2.2.2. Gaussian Elimination. Solve the system associated with the augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 2 & 4 & 2 & 16 \\ -1 & 5 & -4 & -3 \end{array}\right]$$

We seek to eliminate entries below the first pivot $a_{11} = 1$.

1. Replace R_2 with $R_2 - 2R_1$:
$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 2 & 0 & 4 \\ -1 & 5 & -4 & -3 \end{array}\right]$$
2. Replace R_3 with $R_3 + R_1$:
$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 2 & 0 & 4 \\ 0 & 6 & -3 & 3 \end{array}\right]$$
3. Now we focus on the sub-matrix beginning at $a_{22} = 2$. Eliminate the entry below it (6) by replacing R_3 with $R_3 - 3R_2$:
$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & -3 & -9 \end{array}\right]$$

The matrix is now in Row Echelon Form (upper triangular). We solve by back-substitution:

- From R_3: $-3x_3 = -9 \implies x_3 = 3$.
- From R_2: $2x_2 + 0x_3 = 4 \implies 2x_2 = 4 \implies x_2 = 2$.
- From R_1: $x_1 + x_2 + x_3 = 6 \implies x_1 + 2 + 3 = 6 \implies x_1 = 1$.

The unique solution is $S = \{(1, 2, 3)\}$.

The Gauss-Jordan Algorithm

While Gaussian elimination reduces a matrix to *Row Echelon Form* (REF), the *Gauss-Jordan algorithm* extends this process to produce a canonical form known as the *Reduced Row Echelon Form* (RREF). This form is unique for any given matrix, removing the ambiguity associated with the choice of operations in the standard REF.

Definition 2.2.3. ***Reduced Row Echelon Form***. A matrix is in Reduced Row Echelon Form (RREF) if it satisfies the following conditions:

1. It is in Row Echelon Form (all non-zero rows are above any rows of all zeros, and the leading entry of each row is strictly to the right of the leading entry of the row above).
2. The leading entry (or pivot) in each non-zero row is equal to 1.
3. Each pivot is the only non-zero entry in its column.

The algorithm proceeds in two phases: a *forward pass* to reach REF, and a *backward pass* to reach RREF.

Definition 2.2.4. ***Gauss-Jordan Algorithm***. Let A be an $m \times n$ matrix.

Step 1: **Forward Pass (Elimination):**

(i) Identify the leftmost non-zero column. This is a *pivot column*.

(ii) Select a non-zero entry in this column as the *pivot*. If necessary, perform a row interchange to move this entry to the top of the active sub-matrix.

(iii) Use row replacement operations to create zeros in all positions below the pivot.

(iv) Ignore the row containing the pivot and repeat the process on the remaining sub-matrix until no non-zero rows remain.

Step 2: Backward Pass (Reduction):

(i) Working from the rightmost pivot upwards and to the left: scale the pivot row so the pivot entry becomes 1.

(ii) Use row replacement operations to create zeros in all positions *above* each pivot.

The resulting matrix is denoted $\text{rref}(A)$.

Unlike the Row Echelon Form, which depends on the specific sequence of row operations chosen, the Reduced Row Echelon Form is an invariant of the matrix.

***Theorem* 2.2.2. Uniqueness of RREF**. Every matrix A is row-equivalent to a unique matrix in reduced row echelon form.

Proof. Suppose, for the sake of contradiction, that A is row-equivalent to two distinct reduced row echelon matrices R and S. Since row equivalence is an equivalence relation, R is row-equivalent to S. By Theorem 2.2.1, the homogeneous systems $R\mathbf{x} = \mathbf{0}$ and $S\mathbf{x} = \mathbf{0}$ share the exact same solution set. We show $R = S$ by induction on the number of columns n of A.

Base Case ($n = 1$): If A has one column, the RREF is either the zero column (if $A = \mathbf{0}$) or a column with a 1 in the top position and 0s elsewhere (if $A \neq \mathbf{0}$). In either case, R and S are identical.

Inductive Step: Assume the claim holds for any matrix with k columns. Consider a matrix with $k + 1$ columns. Let R_k and S_k denote the sub-matrices consisting of the first k columns of R and S respectively. These are RREFs of the first k columns of A. By the induction hypothesis, $R_k = S_k$.

We need only check the $(k + 1)$-th column. Let $\mathbf{r}$ and $\mathbf{s}$ be the $(k + 1)$-th columns of R and S.

- **Case 1:** The system $R_k\mathbf{y} = \mathbf{r}$ is consistent. This implies $\mathbf{r}$ is a linear combination of the pivot columns of R_k. The coefficients of this combination are unique and determined entirely by the solution set of $R\mathbf{x} = \mathbf{0}$. Since S has the same solution set and $S_k = R_k$, $\mathbf{s}$ must be the same linear combination of the same pivot columns. Thus $\mathbf{r} = \mathbf{s}$.
- **Case 2:** The system $R_k\mathbf{y} = \mathbf{r}$ is inconsistent. Then $\mathbf{r}$ is a new pivot column. In RREF, a new pivot column must take the form of a column with a 1 in the $(p + 1)$-th position (where p is the number of non-zero rows in R_k) and zeros elsewhere. The same logic applies to S, so $\mathbf{s}$ is the same column. Thus $\mathbf{r} = \mathbf{s}$.

Since all columns match, $R = S$. ■

Computation Examples

We illustrate the algorithm with specific examples. Note that while the forward pass targets entries below the diagonal, the backward pass clears the entries above.

Example 2.2.3. Unique Solution. Calculate $\text{rref}(A)$ for:

$$A = \begin{bmatrix} 1 & 2 & -3 & 1 \\ 2 & 4 & 0 & 7 \\ -1 & 3 & 2 & 0 \end{bmatrix}.$$

Forward Pass: Eliminate below the first pivot $a_{11} = 1$:

$$R_2 \to R_2 - 2R_1, \quad R_3 \to R_3 + R_1 \implies \begin{bmatrix} 1 & 2 & -3 & 1 \\ 0 & 0 & 6 & 5 \\ 0 & 5 & -1 & 1 \end{bmatrix}.$$

The second column has zeros below the first row, so we move to the third column. However, to maintain the echelon structure (pivots moving right), we require a non-zero entry in the $(2, 2)$ position if possible.

Noticing the non-zero entry in the second column of R_3, we swap $R_2 \leftrightarrow R_3$:

$$\begin{bmatrix} 1 & 2 & -3 & 1 \\ 0 & 5 & -1 & 1 \\ 0 & 0 & 6 & 5 \end{bmatrix}.$$

This is a Row Echelon Form.

Backward Pass: Scale R_3 to make the pivot 1 ($R_3 \to \frac{1}{6}R_3$):

$$\begin{bmatrix} 1 & 2 & -3 & 1 \\ 0 & 5 & -1 & 1 \\ 0 & 0 & 1 & 5/6 \end{bmatrix}.$$

Clear entries above this pivot ($R_2 \to R_2 + R_3$, $R_1 \to R_1 + 3R_3$):

$$\begin{bmatrix} 1 & 2 & 0 & 1+15/6 \\ 0 & 5 & 0 & 1+5/6 \\ 0 & 0 & 1 & 5/6 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 21/6 \\ 0 & 5 & 0 & 11/6 \\ 0 & 0 & 1 & 5/6 \end{bmatrix}.$$

Scale R_2 to make the pivot 1 ($R_2 \to \frac{1}{5}R_2$):

$$\begin{bmatrix} 1 & 2 & 0 & 7/2 \\ 0 & 1 & 0 & 11/30 \\ 0 & 0 & 1 & 5/6 \end{bmatrix}.$$

Clear the entry above the second pivot ($R_1 \to R_1 - 2R_2$):

$$\begin{bmatrix} 1 & 0 & 0 & 7/2-22/30 \\ 0 & 1 & 0 & 11/30 \\ 0 & 0 & 1 & 5/6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 83/30 \\ 0 & 1 & 0 & 11/30 \\ 0 & 0 & 1 & 5/6 \end{bmatrix}.$$

Remark. It is permissible to deviate slightly from the strict algorithmic order (e.g., scaling rows earlier) to avoid arithmetic with fractions, provided the final structure adheres to the definition.

***Proposition* 2.2.1.** *Zero Columns.* If a column of a matrix consists entirely of zeros, it remains a zero column throughout the Gauss-Jordan process. Consequently, $\text{rref}([A \mid \mathbf{0}]) = [\text{rref}(A) \mid \mathbf{0}]$.

Proof. Let $\mathbf{c}_k$ be the k-th column of the matrix A. Suppose $\mathbf{c}_k = \mathbf{0}$, meaning every entry $a_{ik} = 0$ for all rows i. We examine the effect of the three elementary row operations on the entries of this column:

(i) **Scaling:** $R_i \to \alpha R_i$. The new entry in the column is $\alpha \cdot 0 = 0$.
(ii) **Interchange:** $R_i \leftrightarrow R_j$. Since the entries in both row i and row j of this column are 0, swapping them leaves 0 in both positions.
(iii) **Replacement:** $R_i \to R_i + \alpha R_j$. The new entry is $a_{ik} + \alpha a_{jk} = 0 + \alpha(0) = 0$.

Since no elementary row operation can change a zero entry in a zero column to a non-zero value, the column remains a zero column throughout the entire Gaussian elimination process. Consequently, for a homogeneous system $[A \mid \mathbf{0}]$, the last column (the constant column) is a zero column. Since row operations do not alter this column, the RREF will also have a zero column in the last position. Thus $\text{rref}([A \mid \mathbf{0}]) = [\text{rref}(A) \mid \mathbf{0}]$. ■

***Theorem* 2.2.3. RREF of Augmented Columns**. Suppose that $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times p}$. Then the first n columns of $\text{rref}(A)$ and $\text{rref}([A \mid B])$ are identical.

Proof. The Gauss-Jordan algorithm processes a matrix from left to right (column by column). The decisions to swap rows, scale rows, or add multiples of rows to eliminate entries are determined entirely by the entries in the current column and the columns to its left. Therefore, when reducing $[A \mid B]$, the operations performed on the first n columns depend exclusively on the entries within those first n columns (which form A). The presence of the extra columns B to the right does not influence the choice of pivots or the sequence of row operations for the columns of A. ■

2.3 Classification of Solutions

The Reduced Row Echelon Form provides immediate insight into the structure of the solution set of a linear system $A\mathbf{x} = \mathbf{b}$. By identifying the pivot positions in $\text{rref}([A \mid \mathbf{b}])$, we can classify the system's behaviour.

- **Basic Variables:** Variables corresponding to columns *containing* pivots.
- **Free Variables:** Variables corresponding to columns *without* pivots.

When a system has infinitely many solutions, we express the basic variables in terms of the free variables.

Definition 2.3.1. ***General and Particular Solutions***. The presence of free variables implies the system has infinitely many solutions. We distinguish between two ways of describing them:

- **General Solution:** A formula expressing the solution components $(x_1, \ldots, x_n)$ in terms of the free variables (parameters). This single expression covers the entire solution set.
- **Particular (or Special) Solution:** A specific solution obtained by assigning numerical values to the free variables in the general solution (e.g., setting all free variables to 0).

Remark. One might ask why specific variables are chosen as "free". For example, if $x - y = 0$, why is y free and x basic? Algebraically, one could define y in terms of x. However, we require a systematic choice to avoid circular definitions in complex systems. We customarily choose the pivot variables as the "dependent" or "basic" variables and the non-pivot variables as the "free" variables (or parameters). This distinction is analogous to identifying independent variables in calculus, though here the choice is dictated by the algorithm's echelon structure.

Theorem **2.3.1. Trichotomy of Solutions**. Given a system of m linear equations and n unknowns, the solution set S falls into exactly one of the following cases:

1. $S = \emptyset$ (The system is inconsistent).
2. $|S| = 1$ (The solution is unique).
3. S is infinite (The system has infinitely many solutions).

Proof. We analyze the position of pivots in the reduced augmented matrix.

1. If the last column (the constants) contains a pivot, we have a row $[0 \ldots 0 \mid 1]$. This implies $0 = 1$, so $S = \emptyset$.
2. If the last column does not contain a pivot, the system is consistent.
   - If every variable column has a pivot (no free variables), each variable is determined uniquely by the constants. Thus $|S| = 1$.
   - If at least one variable column lacks a pivot (at least one free variable), we can assign any real number to that free variable. Since $\mathbb{R}$ is infinite, there are infinitely many solutions.

There is no case where the algorithm produces exactly two, or exactly seventeen solutions. ■

Theorem **2.3.2. Unique Solutions for Square Systems**. Given n linear equations in n unknowns $A\mathbf{x} = \mathbf{b}$, a unique solution $\mathbf{x}$ exists if and only if $\text{rref}([A \mid \mathbf{b}]) = [I \mid \mathbf{x}]$. Moreover, if $\text{rref}(A) \neq I$, then there is no unique solution to the system.

Proof.

($\Rightarrow$) Suppose a unique solution exists. By the previous theorem, there must be no free variables. Since there are n columns and n variables, there must be n pivots. In an $n \times n$ matrix, n pivots in RREF implies the matrix is the identity I. Thus $\text{rref}([A \mid \mathbf{b}])$ looks like $[I \mid \mathbf{c}]$, where $\mathbf{c}$ is the column of solution values.

($\Leftarrow$) Suppose $\text{rref}([A \mid \mathbf{b}]) = [I \mid \mathbf{c}]$. This explicitly gives the equations $x_1 = c_1, \ldots, x_n = c_n$, which is a unique solution.

For the second part: If $\text{rref}(A) \neq I$ (for a square matrix), then there are fewer than n pivots. This means there is at least one free variable (infinite solutions) or a row of zeros resulting in inconsistency (no solution). In either case, the solution is not unique. ■

Example 2.3.1. System with Free Variables. Consider the augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 1 & 1 \end{array}\right].$$

Forward Pass: $R_2 \to R_2 - R_1$, $R_3 \to R_3 + R_1$:

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 1 \\ 0 & -2 & 0 & -1 & 0 \\ 0 & 1 & 2 & 2 & 2 \end{array}\right].$$

Swap $R_2 \leftrightarrow R_3$ to simplify the pivot to 1 immediately (optional but convenient):

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 \\ 0 & -2 & 0 & -1 & 0 \end{array}\right].$$

Eliminate below the new pivot ($R_3 \to R_3 + 2R_2$):

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 \\ 0 & 0 & 4 & 3 & 4 \end{array}\right].$$

Backward Pass: Scale $R_3 \to \frac{1}{4}R_3$. Then eliminate above:

$$\text{rref}(A) = \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 3/4 & 1 \end{array}\right].$$

The pivots are in columns 1, 2, and 3 (x_1, x_2, x_3), while column 4 has no pivot (x_4 is free). From row 1: $x_1 = 0$. From row 2: $x_2 + \frac{1}{2}x_4 = 0 \implies x_2 = -\frac{1}{2}x_4$. From row 3: $x_3 + \frac{3}{4}x_4 = 1 \implies x_3 = 1 - \frac{3}{4}x_4$. We express the basic variables in terms of the free ones, thus the solution set is $S = \{(0, -\frac{1}{2}x_4, 1 - \frac{3}{4}x_4, x_4) \mid x_4 \in \mathbb{R}\}$.

Example 2.3.2. Inconsistent System. Solve the system:

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

The augmented matrix reduces as follows:

$$\left[\begin{array}{cc|c} 1 & -1 & 1 \\ 3 & -3 & 0 \\ 2 & -2 & -3 \end{array}\right] \xrightarrow{R_2 - 3R_1, R_3 - 2R_1} \left[\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 0 & -3 \\ 0 & 0 & -5 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array}\right].$$

The second row reads $0x + 0y = 1$. The system has no solution ($S = \emptyset$).

Example 2.3.3. Infinite Solutions (Choice of Variables). Solve:

$$\begin{aligned} x - y + z &= 0 \\ 3x - 3y &= 0 \\ 2x - 2y - 3z &= 0 \end{aligned} \quad \xrightarrow{\text{Augmented}} \quad \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 3 & -3 & 0 & 0 \\ 2 & -2 & -3 & 0 \end{array}\right].$$

Gaussian elimination yields the RREF:

$$\text{rref}([A \mid \mathbf{b}]) = \left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The pivots are in columns 1 and 3, corresponding to variables x and z. Thus, x and z are basic variables, and y is a free variable. From row 2: $z = 0$. From row 1: $x - y = 0 \implies x = y$. The solution set is $S = \{(y, y, 0) \mid y \in \mathbb{R}\}$.

Example 2.3.4. Parametric Solutions. Solve the system:

$$\begin{aligned} x_1 + x_3 &= 0 \\ 2x_2 &= 0 \\ 3x_3 + x_4 &= 0 \\ 3x_1 + 2x_2 &= 0 \end{aligned}$$

Form the augmented matrix and reduce:

$$\left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 \\ 3 & 2 & 0 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right].$$

Here, every variable column has a pivot. The unique solution is the trivial solution ($x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0$).

Example 2.3.5. Hyperplanes in $\mathbb{R}^5$. Consider the system corresponding to the augmented matrix:

$$[A \mid \mathbf{b}] = \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ 4 & 4 & 4 & 0 & 0 & 1 \end{array}\right].$$

Applying Gauss-Jordan elimination yields:

$$\operatorname{rref}([A \mid \mathbf{b}]) = \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & -1/2 & 1/4 \end{array}\right].$$

Thus, the solution set is $S = \left\{\left(-x_4,\ x_4 - \frac{1}{2}x_5,\ \frac{1}{4} + \frac{1}{2}x_5,\ x_4,\ x_5\right) \mid x_4, x_5 \in \mathbb{R}\right\}$.

Remark. The choice of which variables are free is determined by the algorithm (non-pivot columns). While one could algebraically solve for x_4 in terms of x_1, the convention of expressing pivot variables in terms of non-pivot variables is standard because it guarantees a consistent dependency structure without cyclic definitions.

Note. It is crucial to distinguish between reducing the coefficient matrix A and the augmented matrix $[A \mid \mathbf{b}]$.

- $\operatorname{rref}(A)$ reveals the dependency structure of the variables (which are basic vs. free) but ignores the target values $\mathbf{b}$. It effectively solves only the homogeneous equation $A\mathbf{x} = \mathbf{0}$.
- $\operatorname{rref}([A \mid \mathbf{b}])$ applies the operations to both the coefficients and the constants. This preserves the equality of the system at every step, allowing us to determine the specific solution set for $A\mathbf{x} = \mathbf{b}$.

To find the solution set of a general linear system, one must *always* use the augmented matrix.

Example 2.3.6. Importance of Augmentation. Consider the system:

$$\begin{aligned} x + y &= 1 \\ x + y &= 2 \end{aligned}$$

This is obviously inconsistent (parallel lines).

1. **Reducing Coefficient Matrix A:**

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

   Looking only at this, one might mistakenly conclude the system has a free variable (infinite solutions), as there is a non-pivot column.

2. **Reducing Augmented Matrix** $[A \mid \mathbf{b}]$**:**

$$[A \mid \mathbf{b}] = \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 1 & 2 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 0 & 1 \end{array}\right].$$

The second row reveals $0x + 0y = 1$, which is a contradiction.

By failing to augment, we missed the inconsistency in the constants. Thus, the solution set is $S = \emptyset$.

2.4 Exercises

1. **Parametric Analysis.** Consider the system of linear equations in unknowns x, y, z dependent on a parameter $k \in \mathbb{R}$:

$$\begin{aligned} x + y + z &= 1 \\ x + 2y + 4z &= k \\ x + 4y + 10z &= k^2 \end{aligned}$$

Use Gaussian elimination to determine the values of k for which the system has:

(a) A unique solution.
(b) No solution.
(c) Infinitely many solutions.

In the case of infinitely many solutions, express the solution set in terms of a free variable.

2. **The Fundamental Theorem of Homogeneous Systems.** A homogeneous system is one where the constant terms are zero, i.e., $A\mathbf{x} = \mathbf{0}$. Such a system always possesses the trivial solution. Prove that if a homogeneous system has strictly fewer equations than unknowns ($m < n$), it must possess a non-trivial solution (a solution where not all x_i are zero).

Remark. Consider the Reduced Row Echelon Form of the matrix. How many pivots can there be at most? What does this imply about the existence of free variables?

3. **The Affine Structure of Solutions.** Let A be an $m \times n$ matrix and let $\mathbf{b}$ be a column of constants. Let S_h denote the solution set of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. Suppose the inhomogeneous equation $A\mathbf{x} = \mathbf{b}$ is consistent and let $\mathbf{p}$ be one specific solution (called a particular solution). Prove that the solution set S of the inhomogeneous system is exactly the set of all solutions of the form $\mathbf{p}+\mathbf{h}$, where $\mathbf{h} \in S_h$.

$$S = \{\mathbf{p} + \mathbf{h} \mid \mathbf{h} \in S_h\}$$

4. **Intersection of Planes.** In a spatial Cartesian coordinate system, three planes are given by the equations:

$$\begin{aligned} 9x - 3y + z &= 20 \\ x + y + z &= 0 \\ -x + 2y + z &= -10 \end{aligned}$$

(a) Use Gaussian elimination to determine the set of common points (the intersection) of these three planes.
(b) Geometrically, do these planes meet at a point, a line, or nowhere?

5. **Rationality of Solutions.** One of the strengths of Gaussian elimination is that it relies solely on the arithmetic operations of addition, subtraction, multiplication, and division. Prove that if the coefficients a_{ij} and the constants b_i of a linear system are all rational numbers (i.e., elements of $\mathbb{Q}$), and if the system has a unique solution, then the solution components $x_1, \dots, x_n$ must consist entirely of rational numbers.

Remark. Contrast this with finding roots of polynomials, where $x^2 - 2 = 0$ has rational coefficients but irrational solutions.

6. Polynomial Interpolation. Complex problems can often be reduced to finding coefficients that satisfy a set of linear constraints.

(a) Suppose we wish to find a parabola with equation $y = ax^2 + bx + c$ that passes through the three specific points $(-3, 20)$, $(1, 0)$, and $(2, 10)$. Substitute the coordinates of these points into the equation to obtain a system of three linear equations in the unknowns a, b, c. Solve this system to determine the explicit equation of the parabola.

(b) Now, consider the general case. Let $(t_1, y_1), (t_2, y_2), \ldots, (t_n, y_n)$ be n points in the plane with distinct t-coordinates. We seek a polynomial of degree at most $n-1$, defined by $P(t) = c_0 + c_1 t + \cdots + c_{n-1}t^{n-1}$, that passes through all these points. Write down the system of n linear equations that the coefficients $c_0, \ldots, c_{n-1}$ must satisfy.

(c) Show that the system derived in (b) is equivalent to the single matrix equation $V\mathbf{c} = \mathbf{y}$, where the columns $\mathbf{c}$ and $\mathbf{y}$ are $\mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ and V is the *Vandermonde matrix*:

$$V = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{bmatrix}.$$

(d) Given that the homogeneous system $V\mathbf{c} = \mathbf{0}$ has only the trivial solution (a fact you may assume), what can you conclude about the existence and uniqueness of such a polynomial for any set of n distinct points?

(e) **Overdetermined Systems:** Suppose we are given $n = 4$ points, but we still wish to fit a quadratic polynomial (degree 2, which has only 3 coefficients). Write down the associated matrix equation. What is the shape of the matrix V? Based on your knowledge of linear systems, is it guaranteed that such a parabola exists?

7. Curve Fitting and Consistency. Suppose we suspect a functional relationship $y = f(x)$ between two variables is quadratic (i.e., of the form $y = ax^2 + bx + c$). We collect the following data:

x	1	2	3	4
y	2	7	16	29

(a) By substituting the data points into the quadratic equation, set up a system of 4 linear equations in the unknowns a, b, c.

(b) Form the augmented matrix and solve the system. Is there a unique set of coefficients a, b, c that satisfies all four data points perfectly, or is the system inconsistent?

8. ⋆ Redundancy of Row Interchange. The definition of Elementary Row Operations includes three types: Scaling, Replacement, and Interchange. Prove that the Interchange operation is theoretically redundant. Specifically, show that the row swap $R_1 \leftrightarrow R_2$ can be achieved by a sequence of Scaling and Replacement operations alone.

Remark. Attempt to construct the swap using the operations $R_1 \to R_1 + R_2$, $R_2 \to R_2 - R_1$, etc. Be careful with signs.

9. ⋆ Simultaneous Systems and Matrix Algebra. Suppose we wish to solve two linear systems that share the same coefficient matrix but have different right-hand sides:

$$A\mathbf{x} = \mathbf{b} \quad \text{and} \quad A\mathbf{y} = \mathbf{c}.$$

Instead of performing Gaussian elimination twice, we can augment the matrix with both columns: $[A \mid \mathbf{b} \;\; \mathbf{c}]$. Row operations applied to this array effectively solve both systems simultaneously.

Consider the system where $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Find the columns $\mathbf{u}$ and $\mathbf{v}$ such that $A\mathbf{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $A\mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Form the matrix X whose columns are $\mathbf{u}$ and $\mathbf{v}$, i.e., $X = [\mathbf{u} \;\; \mathbf{v}]$. What is the significance of the matrix X in relation to A?

Chapter 3

Vectors and Matrices

In the preceding chapter, we treated the linear system $A\mathbf{x} = \mathbf{b}$ primarily as an algebraic obstacle to be overcome via Gaussian elimination. We found solutions by manipulating arrays of numbers. However, to understand the deeper structure of these solutions we must formalise the objects we are manipulating.

But before formalising, what about some revision of the previous notes. A *field* F is a set of elements (in our case scalars elements) equipped with two binary operations, addition $(+)$ and multiplication $(\cdot)$, satisfying the following axioms for all $x, y, z \in F$:

1. **Commutativity:** $x + y = y + x$ and $xy = yx$.
2. **Associativity:** $x + (y + z) = (x + y) + z$ and $x(yz) = (xy)z$.
3. **Identities:** There exist distinct elements 0 (additive identity) and 1 (multiplicative identity) such that $x + 0 = x$ and $x \cdot 1 = x$.
4. **Inverses:**
   - For every $x \in F$, there is a unique $(-x)$ such that $x + (-x) = 0$.
   - For every $x \in F \setminus \{0\}$, there is a unique x^{-1} such that $xx^{-1} = 1$.
5. **Distributivity:** $x(y + z) = xy + xz$.

Remark. The most familiar fields are the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$. The integers $\mathbb{Z}$ do *not* form a field because distinct non-zero integers (like 2) generally lack multiplicative inverses within the set. Throughout this text, unless specified otherwise, F denotes the field $\mathbb{R}$.

3.1 The Vector Space $\mathbb{R}^n$

Last notes we introduced vectors in 2D and 3D now we generalise these vectors to n-dimensions.

Definition 3.1.1. ***n*-Dimensional Vector**. Let n be a positive integer. An n-dimensional vector over a field F is an ordered n-tuple $(x_1, x_2, \ldots, x_n)$ where each component $x_i \in F$. The set of all such vectors is the *vector space* F^n.

$$F^n = \{(x_1, \ldots, x_n) \mid x_i \in F\}.$$

Two vectors are equal if and only if their corresponding components are identical.

Remark. (Universality of Gaussian Elimination). In Chapter 1, we solved linear systems where the coefficients were real numbers. However, the algorithm of Gaussian Elimination relies solely on the four arithmetic operations (addition, subtraction, multiplication, and division by non-zero elements). Since these operations are exactly what define a field F, the entire theory of Chapter 1, including Row Echelon Form, Pivot variables, and the distinction between unique/infinite solutions, holds true for linear systems over *any* field F. For example, we can solve systems with complex coefficients ($F = \mathbb{C}$) or rational coefficients ($F = \mathbb{Q}$) using the exact same steps.

Notation 3.1.1. Column Vector Convention. There is a bijection between row tuples and column arrays. To ensure compatibility with matrix multiplication (defined later), we adopt the standard convention that vectors in F^n are *column vectors*.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

To conserve vertical space in text, we often write $\mathbf{x} = (x_1, \ldots, x_n)$ or use the transpose notation $\mathbf{x} = [x_1, \ldots, x_n]^T$.

Algebraic Structure

The set $\mathbb{R}^n$ is not merely a collection of points; it possesses an algebraic structure induced by the underlying field. We define two fundamental operations: vector addition and scalar multiplication.

Definition 3.1.2. ***Vector Operations***. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

1. **Addition:** We add vectors component-wise.

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_n + v_n \end{bmatrix}.$$

2. **Scalar Multiplication:** We scale a vector by multiplying every component by the scalar.

$$\lambda \mathbf{u} = \lambda \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} \lambda u_1 \\ \vdots \\ \lambda u_n \end{bmatrix}.$$

The zero vector, denoted $\mathbf{0}$, is the vector consisting entirely of zeros. It serves as the additive identity. These operations inherit the properties of the field F.

***Theorem* 3.1.1. Vector Space Axioms**. For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and scalars $r, s \in \mathbb{R}$:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
(iii) $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
(iv) $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$, where $-\mathbf{u} = (-1)\mathbf{u}$.
(v) $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$ and $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$.
(vi) $r(s\mathbf{u}) = (rs)\mathbf{u}$.
(vii) $1\mathbf{u} = \mathbf{u}$.

Proof. The proofs follow directly from the properties of the underlying field. For instance, commutativity of vector addition relies on the commutativity of addition in $\mathbb{R}$: the i-th component of $\mathbf{u} + \mathbf{v}$ is $u_i + v_i$, which equals $v_i + u_i$, the i-th component of $\mathbf{v} + \mathbf{u}$. ■

Remark. (Abstract Vector Spaces). In this chapter, we defined "vectors" concretely as ordered lists of numbers in $\mathbb{R}^n$. However, in broader mathematics, a **Vector Space** is defined abstractly as *any* set V equipped with addition and scalar multiplication that satisfies the eight axioms listed in the theorem above. This generalization allows us to treat diverse mathematical objects as vectors, including:

- The set of all polynomials (e.g., $1 + x + x^2$).
- The set of all continuous functions on an interval.
- The set of all $m \times n$ matrices (as noted in Section 3).

While this text focuses primarily on $\mathbb{R}^n$, we might briefly revisit these general spaces in later chapters and definitely introduce it in later notes to show how linear algebra unifies geometry and analysis.

3.2 Subspaces and Linear Combinations

A central concept in linear algebra is the construction of new vectors from a given set using only the operations of addition and scaling.

Definition 3.2.1. ***Linear Combination***. A vector $\mathbf{w}$ is a *linear combination* of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ if there exist scalars $c_1, \ldots, c_k$ such that:

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

In the previous chapter, solving $A\mathbf{x} = \mathbf{b}$ was equivalent to asking: "Is $\mathbf{b}$ a linear combination of the columns of A?". The set of all possible linear combinations of a set of vectors generates a structure known as a subspace.

Subspaces

Often we are interested in subsets of $\mathbb{R}^n$ that act like smaller vector spaces tucked inside the larger one. For a subset to preserve the algebraic structure, one must not be able to "escape" the set via addition or scaling.

Definition 3.2.2. ***Subspace***. A subset S of $\mathbb{R}^n$ is a *subspace* if it satisfies three conditions:

1. The zero vector $\mathbf{0}$ is in S.
2. **Closure under Addition:** If $\mathbf{u} \in S$ and $\mathbf{v} \in S$, then $\mathbf{u} + \mathbf{v} \in S$.
3. **Closure under Scaling:** If $\mathbf{u} \in S$ and $c \in \mathbb{R}$, then $c\mathbf{u} \in S$.

The first condition ensures the set is non-empty and contains the additive identity. The latter two can be compressed into a single criterion.

***Theorem* 3.2.1. Subspace Criterion**. A non-empty subset $S \subseteq \mathbb{R}^n$ is a subspace if and only if for all $\mathbf{u}, \mathbf{v} \in S$ and scalars $c, d \in \mathbb{R}$, the linear combination $c\mathbf{u} + d\mathbf{v}$ is in S.

Proof.

(⇒) If S is a subspace, it is closed under scaling ($c\mathbf{u}, d\mathbf{v} \in S$) and addition ($c\mathbf{u} + d\mathbf{v} \in S$).
(⇐) Suppose S is closed under linear combinations.

- Taking $c = 1, d = 1$, $\mathbf{u} + \mathbf{v} \in S$ (Closure under Addition).
- Taking $d = 0$, $c\mathbf{u} \in S$ (Closure under Scaling).
- Taking $c = 0, d = 0$, $\mathbf{0} \in S$.

Thus S satisfies the definition.

■

Example 3.2.1. Trivial Subspaces. For any n, $\mathbb{R}^n$ has two trivial subspaces:

1. The set $\mathbb{R}^n$ itself (the largest subspace).
2. The set $\{\mathbf{0}\}$ (the zero subspace, the smallest subspace).

Example 3.2.2. Solution Spaces. Let A be an $m \times n$ matrix. The set of solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$. *Verification:* Let $S = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$.

1. $A\mathbf{0} = \mathbf{0}$, so $\mathbf{0} \in S$.
2. If $\mathbf{u}, \mathbf{v} \in S$, then $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$. Thus $\mathbf{u} + \mathbf{v} \in S$.
3. If $\mathbf{u} \in S$, then $A(c\mathbf{u}) = c(A\mathbf{u}) = c\mathbf{0} = \mathbf{0}$. Thus $c\mathbf{u} \in S$.

Note that the solution set to an *inhomogeneous* system $A\mathbf{x} = \mathbf{b}$ (with $\mathbf{b} \neq \mathbf{0}$) is not a subspace, as it fails to contain the zero vector.

Remark. (Foreshadowing Dimension). In the earlier Chapter, we observed that the solution set is determined by the *free variables* (the columns without pivots). If a system has n variables and the elimination process yields r pivots, then there must be $n-r$ free variables left over. Geometrically, this count determines the "size" of the solution set. If $n-r=1$, the solution is a line; if $n-r=2$, it is a plane. One might be tempted to define the "dimension" of a subspace simply as this number $n-r$. However, to formalise this rigorously, we need the tools of *Linear Independence* and *Basis*, which we will develop in the next chapter.

Operations on Subspaces

Just as we operate on vectors, we can operate on the sets containing them.

***Theorem* 3.2.2. Intersection of Subspaces**. If U and V are subspaces of $\mathbb{R}^n$, then their intersection $U \cap V$ is also a subspace of $\mathbb{R}^n$.

Proof. Since $\mathbf{0} \in U$ and $\mathbf{0} \in V$, $\mathbf{0} \in U \cap V$. Let $\mathbf{x}, \mathbf{y} \in U \cap V$ and $c \in \mathbb{R}$. Since $\mathbf{x}, \mathbf{y} \in U$ (a subspace), $c\mathbf{x}+\mathbf{y} \in U$. Since $\mathbf{x}, \mathbf{y} \in V$ (a subspace), $c\mathbf{x}+\mathbf{y} \in V$. Therefore, $c\mathbf{x}+\mathbf{y} \in U \cap V$. By the Subspace Criterion, $U \cap V$ is a subspace. ∎

Note. The *union* of two subspaces $U \cup V$ is generally not a subspace. Consider $\mathbb{R}^2$. Let U be the x-axis (generated by $\mathbf{e}_1$) and V be the y-axis (generated by $\mathbf{e}_2$). Both are subspaces. However, $\mathbf{e}_1 \in U \cup V$ and $\mathbf{e}_2 \in U \cup V$, but their sum $\mathbf{e}_1 + \mathbf{e}_2 = (1,1)$ is in neither U nor V. The union is not closed under addition.

Instead of the union, the smallest subspace containing both U and V is their *sum*.

Definition 3.2.3. *Sum of Subspaces*. Let U and V be subspaces of $\mathbb{R}^n$. The sum $U+V$ is defined as:

$$U + V = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in U, \mathbf{v} \in V\}.$$

It is straightforward to verify that $U+V$ is a subspace.

3.3 Matrix Arithmetic

We have used matrices as shorthand for linear systems. We now formally define them and their internal structure.

Definition 3.3.1. *Matrix*. An $m \times n$ matrix A is a rectangular array of scalars with m rows and n columns. The entry in the i-th row and j-th column is denoted a_{ij} (or A_{ij}).

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

The set of all $m \times n$ matrices with real entries is denoted $\mathbb{R}^{m \times n}$.

Notation 3.3.1. Submatrices A matrix is effectively a collection of vectors.

- The j-th **column vector** of A, denoted $\text{col}_j(A)$, is the $m \times 1$ vector $\begin{bmatrix} a_{1j} & \dots & a_{mj} \end{bmatrix}^T$.
- The i-th **row vector** of A, denoted $\text{row}_i(A)$, is the $1 \times n$ vector $\begin{bmatrix} a_{i1} & \dots & a_{in} \end{bmatrix}$.

We may write A as a row of columns or a column of rows:

$$A = [\mathbf{c}_1 \mid \mathbf{c}_2 \mid \cdots \mid \mathbf{c}_n] = \begin{bmatrix} \mathbf{r}_1 \\ \hline \vdots \\ \hline \mathbf{r}_m \end{bmatrix}.$$

Definition 3.3.2. ***Transpose***. The *transpose* of an $m \times n$ matrix A, denoted A^T, is the $n \times m$ matrix obtained by swapping rows and columns. Formally, $(A^T)_{ij} = A_{ji}$.

Example 3.3.1. Transpose. If $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$. Notice that $\text{row}_i(A) = (\text{col}_i(A^T))^T$.

Basic matrix arithmetic (addition and scalar multiplication) follows the same component-wise logic as vector arithmetic.

- **Equality:** $A = B$ if they have the same dimensions and $A_{ij} = B_{ij}$ for all i, j.
- **Addition:** $(A + B)_{ij} = A_{ij} + B_{ij}$. Defined only if dimensions match.
- **Scalar Multiplication:** $(\lambda A)_{ij} = \lambda A_{ij}$.

Consequently, $\mathbb{R}^{m\times n}$ itself forms a vector space of dimension mn. The arithmetic of *multiplying* two matrices is more complex and corresponds to the composition of linear maps, which we shall explore in the subsequent chapter.

The Zero Matrix

Just as the number 0 serves as the additive identity in the field of real numbers, there exists a matrix that performs the same role in the vector space of matrices.

Definition 3.3.3. ***Zero Matrix***. The *zero matrix* in $\mathbb{R}^{m\times n}$, denoted $\mathbf{0}_{m\times n}$ (or simply $\mathbf{0}$ when dimensions are clear), is the matrix for which every entry is zero:

$$(\mathbf{0})_{ij} = 0 \quad \text{for all } i, j.$$

For any matrix $A \in \mathbb{R}^{m\times n}$, the *additive inverse* $-A$ is the matrix obtained by negating every entry of A. It satisfies $A + (-A) = \mathbf{0}$.

***Theorem* 3.3.1. Properties of Zero Matrix**. For any $A \in \mathbb{R}^{m\times n}$:

1. $A + \mathbf{0} = A$.
2. $A - A = \mathbf{0}$.
3. $0 \cdot A = \mathbf{0}$ (scalar zero times matrix A).

3.4 Matrix Multiplication

While addition and scalar multiplication are defined component-wise, matrix multiplication is structurally distinct. It is not obtained by multiplying corresponding entries. Instead, it is defined to facilitate the composition of linear transformations, a connection we will formalise in later chapters.

Definition 3.4.1. ***Matrix Product***. Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. The product AB is the $m \times p$ matrix defined by:

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

for all $1 \le i \le m$ and $1 \le j \le p$.

Note. The product AB is defined if and only if the number of columns of A equals the number of rows of B. The resulting matrix has the number of rows of A and the number of columns of B.

The Dot Product Perspective

The summation in the definition of matrix multiplication corresponds exactly to the Euclidean *dot product* (or scalar product) of vectors.

Definition 3.4.2. ***Dot Product***. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The dot product $\mathbf{u} \cdot \mathbf{v}$ is defined as:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{k=1}^{n} u_k v_k = \mathbf{u}^T \mathbf{v}.$$

Using this notation, the entry $(AB)_{ij}$ is the dot product of the i-th row of A and the j-th column of B.

$$AB = \begin{bmatrix} \text{row}_1(A) \cdot \text{col}_1(B) & \cdots & \text{row}_1(A) \cdot \text{col}_p(B) \\ \vdots & \ddots & \vdots \\ \text{row}_m(A) \cdot \text{col}_1(B) & \cdots & \text{row}_m(A) \cdot \text{col}_p(B) \end{bmatrix}.$$

Example 3.4.1. Non-Commutativity. Matrix multiplication is generally **not** commutative. That is, $AB \neq BA$ in general, even when both products are defined. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$.

$$AB = \begin{bmatrix} 1(5) + 2(7) & 1(6) + 2(8) \\ 3(5) + 4(7) & 3(6) + 4(8) \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}.$$

$$BA = \begin{bmatrix} 5(1) + 6(3) & 5(2) + 6(4) \\ 7(1) + 8(3) & 7(2) + 8(4) \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}.$$

Clearly $AB \neq BA$.

The Identity Matrix

Just as 1 is the multiplicative identity for real numbers, the identity matrix acts as the neutral element for matrix multiplication.

Definition 3.4.3. ***Identity Matrix***. The identity matrix of size n, denoted I_n (or simply I), is the $n \times n$ square matrix with 1s on the main diagonal and 0s elsewhere.

$$(I_n)_{ij} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Proposition **3.4.1.** *Multiplicative Identity.* For any $A \in \mathbb{R}^{m \times n}$:

$$I_m A = A \quad \text{and} \quad A I_n = A.$$

Matrix-Vector Products and Linear Combinations

A particularly important case of matrix multiplication arises when B is a column vector $\mathbf{x} \in \mathbb{R}^{n \times 1}$. The product $A\mathbf{x}$ yields a vector in $\mathbb{R}^m$.

Algebraically, the i-th component is $(A\mathbf{x})_i = \sum_{j=1}^{n} A_{ij} x_j$. However, we can reinterpret this sum geometrically. Note that A_{ij} is the i-th component of the j-th column of A.

Theorem **3.4.1. Linear Combination of Columns**. For any $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$, the product $A\mathbf{x}$ is a linear combination of the columns of A, with weights given by the entries of $\mathbf{x}$.

$$A\mathbf{x} = x_1 \,\text{col}_1(A) + x_2 \,\text{col}_2(A) + \cdots + x_n \,\text{col}_n(A).$$

Proof. Let $\mathbf{c}_j = \text{col}_j(A)$. The i-th entry of the linear combination $\sum_{j=1}^{n} x_j \mathbf{c}_j$ is:

$$\left(\sum_{j=1}^{n} x_j \mathbf{c}_j \right)_i = \sum_{j=1}^{n} x_j (\mathbf{c}_j)_i = \sum_{j=1}^{n} x_j A_{ij} = \sum_{j=1}^{n} A_{ij} x_j.$$

This matches the definition of $(A\mathbf{x})_i$. ■

This perspective is crucial. It transforms the problem "Does $A\mathbf{x} = \mathbf{b}$ have a solution?" into "Is $\mathbf{b}$ in the subspace spanned by the columns of A?".

***Proposition* 3.4.2.** *Concatenation.* Multiplication distributes over column concatenation. If $B = [\mathbf{b}_1 \mid \cdots \mid \mathbf{b}_p]$, then:

$$AB = A[\mathbf{b}_1 \mid \cdots \mid \mathbf{b}_p] = [A\mathbf{b}_1 \mid A\mathbf{b}_2 \mid \cdots \mid A\mathbf{b}_p].$$

In other words, the j-th column of the product AB is the product of A and the j-th column of B:

$$\mathrm{col}_j(AB) = A\,\mathrm{col}_j(B).$$

Example 3.4.2. Column Concatenation. Consider the matrix A and two vectors $\mathbf{b}_1, \mathbf{b}_2$:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

We construct the matrix B by concatenating these vectors: $B = [\mathbf{b}_1 \mid \mathbf{b}_2] = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$. To find AB, we can compute the product column by column:

- First column: $A\mathbf{b}_1 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1(1) + 2(0) \\ 3(1) + 4(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.
- Second column: $A\mathbf{b}_2 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1(-1) + 2(2) \\ 3(-1) + 4(2) \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

Concatenating these results yields the final matrix:

$$AB = [A\mathbf{b}_1 \mid A\mathbf{b}_2] = \begin{bmatrix} 1 & 3 \\ 3 & 5 \end{bmatrix}.$$

Properties of Matrix Algebra

Despite the lack of commutativity, matrix operations satisfy most standard arithmetic laws.

***Theorem* 3.4.2. Matrix Arithmetic Laws**. Let A, B, C be matrices of appropriate sizes such that the operations below are defined, and let λ be a scalar.

1. **Associativity of Addition:** $(A + B) + C = A + (B + C)$.
2. **Associativity of Multiplication:** $(AB)C = A(BC)$.
3. **Distributivity:** $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$.
4. **Scalar Commutativity:** $\lambda(AB) = (\lambda A)B = A(\lambda B)$.
5. **Transpose of Product:** $(AB)^T = B^T A^T$.

Proof. We prove associativity of multiplication, as it is non-trivial. Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{p\times q}$. The (i, l)-th entry of $(AB)C$ is:

$$((AB)C)_{il} = \sum_{k=1}^{p} (AB)_{ik} C_{kl} = \sum_{k=1}^{p} \left(\sum_{j=1}^{n} A_{ij} B_{jk} \right) C_{kl} = \sum_{k=1}^{p} \sum_{j=1}^{n} A_{ij} B_{jk} C_{kl}.$$

Since scalar multiplication is associative and sums are finite, we can swap the order of summation:

$$\sum_{j=1}^{n} A_{ij} \left(\sum_{k=1}^{p} B_{jk} C_{kl} \right) = \sum_{j=1}^{n} A_{ij} (BC)_{jl} = (A(BC))_{il}.$$

Thus $(AB)C = A(BC)$. ■

3.5 The Standard Basis

The structure of $\mathbb{R}^n$ is most easily analysed by decomposing vectors into fundamental building blocks.

Definition 3.5.1. ***Kronecker Delta***. The *Kronecker delta* δ_{ij} is a function of two integers i and j defined by:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Definition 3.5.2. ***Standard Basis Vectors***. The standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in $\mathbb{R}^n$ are the columns of the identity matrix I_n. That is, the j-th component of $\mathbf{e}_i$ is δ_{ij}.

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, 0, \ldots, 0)^T \\ \mathbf{e}_2 &= (0, 1, 0, \ldots, 0)^T \\ &\vdots \\ \mathbf{e}_n &= (0, 0, 0, \ldots, 1)^T \end{aligned}$$

***Theorem* 3.5.1. Standard Basis Expansion**. Every vector $\mathbf{v} = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$ can be uniquely expressed as a linear combination of the standard basis vectors:

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_n \mathbf{e}_n.$$

Proof. By the definition of vector addition and scalar multiplication:

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} v_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ v_n \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

■

Example 3.5.1. Any vector in $\mathbb{R}^n$ can be written as a sum of these basic vectors. For example,

$$\begin{aligned} v = (1, 2, 3) &= (1, 0, 0) + (0, 2, 0) + (0, 0, 3) \\ &= 1(1, 0, 0) + 2(0, 1, 0) + 3(0, 0, 1) \\ &= e_1 + 2e_2 + 3e_3 \\ &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \end{aligned}$$

Example 3.5.2. Vector Equations. The system of linear equations

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

can be decomposed into a single vector equation involving the standard basis of variables or the columns of the coefficient matrix:

$$x \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + y \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + z \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}.$$

This explicitly demonstrates that solving the system is equivalent to finding coefficients x, y, z that express the target vector as a linear combination of the column vectors.

3.6 The Standard Matrix Basis

We have established that any vector in $\mathbb{R}^n$ can be decomposed into a linear combination of the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$. This concept extends naturally to the space of matrices $\mathbb{R}^{m\times n}$, allowing us to trade complex matrix equations for elegant scalar operations involving indices.

Definition 3.6.1. ***Standard Basis Matrix***. The ij-th standard basis matrix for $\mathbb{R}^{m\times n}$, denoted E_{ij}, is the matrix with a 1 in the (i,j)-th position and 0 elsewhere. Formally,

$$(E_{ij})_{kl} = \delta_{ik}\delta_{jl}.$$

Just as the vectors $\mathbf{e}_k$ form the building blocks of $\mathbb{R}^n$, the matrices E_{ij} constitute the fundamental units of $\mathbb{R}^{m\times n}$.

***Proposition* 3.6.1.** *Matrix Decomposition.* Every matrix $A \in \mathbb{R}^{m\times n}$ can be uniquely expressed as a linear combination of the basis matrices E_{ij}:

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}E_{ij}.$$

Proof. Let $B = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}E_{ij}$. We examine the (k,l)-th entry of B:

$$B_{kl} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}(E_{ij})_{kl} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}\delta_{ik}\delta_{jl}.$$

The factor δ_{ik} is zero unless $i = k$, and δ_{jl} is zero unless $j = l$. Thus, the double sum collapses to the single term where $i = k$ and $j = l$:

$$B_{kl} = A_{kl}\cdot 1\cdot 1 = A_{kl}.$$

Since $B_{kl} = A_{kl}$ for all k, l, we have $B = A$. ■

Selection and Construction

The interaction between the standard vector basis $\{\mathbf{e}_k\}$ and the matrix basis $\{E_{ij}\}$ provides powerful tools for extracting specific components of a matrix. This notation is frequently employed in proofs to reduce matrix algebra to index arithmetic (a precursor to tensor calculus).

***Proposition* 3.6.2.** *Column and Row Selection.* Let $A \in \mathbb{R}^{m\times n}$.

1. Multiplication by a basis vector on the right extracts a column:

$$A\mathbf{e}_j = \mathrm{col}_j(A).$$

2. Multiplication by a transposed basis vector on the left extracts a row:

$$\mathbf{e}_i^T A = \mathrm{row}_i(A).$$

Proof. For the first claim, consider the k-th component of the vector $A\mathbf{e}_j$:

$$(A\mathbf{e}_j)_k = \sum_{p=1}^{n} A_{kp}(\mathbf{e}_j)_p = \sum_{p=1}^{n} A_{kp}\delta_{pj} = A_{kj}.$$

Thus, the resulting vector consists of the entries $A_{1j}, A_{2j}, \ldots, A_{mj}$, which is exactly the j-th column of A. The proof for the row extraction is analogous. ■

Combining these operations allows us to isolate a single scalar entry from a matrix.

***Corollary* 3.6.1.** Entry Selection

For any matrix A, the (i, j)-th entry is given by:

$$A_{ij} = \mathbf{e}_i^T A \mathbf{e}_j.$$

Proof. Using the previous proposition, $A\mathbf{e}_j$ is the j-th column of A. Premultiplying this column by $\mathbf{e}_i^T$ extracts its i-th component, which is A_{ij}. ■

Furthermore, we can construct the basis matrices E_{ij} directly from the basis vectors using the *outer product.*
***Proposition* 3.6.3.** *Outer Product Construction.* For $\mathbf{e}_i \in \mathbb{R}^m$ and $\mathbf{e}_j \in \mathbb{R}^n$:

$$E_{ij} = \mathbf{e}_i \mathbf{e}_j^T.$$

Proof. The (k, l)-th entry of the product $\mathbf{e}_i \mathbf{e}_j^T$ is:

$$(\mathbf{e}_i \mathbf{e}_j^T)_{kl} = (\mathbf{e}_i)_k (\mathbf{e}_j^T)_l = (\mathbf{e}_i)_k (\mathbf{e}_j)_l = \delta_{ik}\delta_{jl}.$$

This matches the definition of $(E_{ij})_{kl}$. ■

Algebra of Basis Matrices

When manipulating sums involving E_{ij}, it is often necessary to multiply two basis matrices. The result is elegantly determined by the "inner indices".
***Lemma* 3.6.1.** *Product of Basis Matrices.* Let $E_{ij} \in \mathbb{R}^{m\times n}$ and $E_{kl} \in \mathbb{R}^{n\times p}$. Then:

$$E_{ij} E_{kl} = \delta_{jk} E_{il}.$$

That is, the product is zero unless the column index of the first matches the row index of the second ($j = k$), in which case the result is the basis matrix E_{il}.

Proof. We compute the (r, s)-th entry of the product:

$$(E_{ij}E_{kl})_{rs} = \sum_{t=1}^{n} (E_{ij})_{rt}(E_{kl})_{ts} = \sum_{t=1}^{n} (\delta_{ir}\delta_{jt})(\delta_{kt}\delta_{ls}).$$

The term is non-zero only if $t = j$ (from the first delta) and $t = k$ (from the second delta). Thus, if $j \neq k$, the sum is zero. If $j = k$, the sum has exactly one non-zero term (where $t = j = k$):

$$(E_{ij}E_{kl})_{rs} = \delta_{ir}\delta_{ls}\delta_{jk}.$$

This is exactly the (r, s)-th entry of $\delta_{jk}E_{il}$. ■

Example 3.6.1. Matrix Manipulation with Basis Matrices. The basis matrices can be used to perform elementary row and column operations algebraically.

1. **Right Multiplication (AE_{ij}):** Let $A = \sum_{k,l} A_{kl}E_{kl}$. Then:

$$AE_{ij} = \left(\sum_{k,l} A_{kl}E_{kl}\right) E_{ij} = \sum_{k,l} A_{kl}(E_{kl}E_{ij}) = \sum_{k,l} A_{kl}\delta_{li}E_{kj}.$$

   The sum over l collapses to $l = i$, yielding:

$$AE_{ij} = \sum_{k} A_{ki}E_{kj}.$$

   This matrix has the entries of column i of A moved to column j, with zeros everywhere else.

2. **Left Multiplication ($E_{ij}A$):** Similarly,

$$E_{ij}A = E_{ij}\sum_{k,l} A_{kl}E_{kl} = \sum_{k,l} A_{kl}\delta_{jk}E_{il} = \sum_{l} A_{jl}E_{il}.$$

This matrix consists of the j-th row of A moved to the i-th row, with zeros elsewhere.

For instance, if $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then:

$$AE_{12} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix}.$$

Here, the first column of A $\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$ has been moved to the second column.

***Theorem* 3.6.1. Summary of Basis Identities.** For matrices $A \in \mathbb{R}^{m\times n}$ and standard basis vectors $\mathbf{e}_i$:

$$\begin{aligned} \mathbf{e}_i^T\mathbf{e}_j &= \delta_{ij} \\ E_{ij} &= \mathbf{e}_i\mathbf{e}_j^T \\ A &= \sum_{i,j} A_{ij}E_{ij} \\ E_{ij}E_{kl} &= \delta_{jk}E_{il} \end{aligned}$$

3.7 Exercises

1. **Parameter Analysis (Two Parameters).** For what values of a, b does the following system have a solution? Find the solution in these cases.

$$\begin{aligned} 3x_1 + 2x_2 + ax_3 + x_4 - 3x_5 &= 4 \\ 5x_1 + 4x_2 + 3x_3 + 3x_4 - x_5 &= 3 \\ x_1 + x_2 + 3x_3 + 2x_4 + x_5 &= 1 \\ x_2 + 2x_3 + 2x_4 + 6x_5 &= -3 \\ x_3 + bx_4 + x_5 &= 1 \end{aligned}$$

2. **Parameter Analysis (Cyclic System).** Discuss for what values of λ the following system has a unique solution, infinite solutions, or no solution. In the cases where solutions exist, find the general solution.

$$\begin{aligned} \lambda x_1 + x_2 + x_3 &= 1 \\ x_1 + \lambda x_2 + x_3 &= \lambda \\ x_1 + x_2 + \lambda x_3 &= \lambda^2 \end{aligned}$$

3. **Lattice Properties of Subspaces.** Let X, Y, A and B be subspaces of $\mathbb{R}^n$. The operations of intersection and sum allow us to order subspaces by containment. Prove the following fundamental containment laws:
   (a) If $A \subseteq X$ and $A \subseteq Y$, prove that $A \subseteq X \cap Y$.
   (b) If $X \subseteq B$ and $Y \subseteq B$, prove that $X + Y \subseteq B$.

4. **Subspace Absorption.** Let X and Y be subspaces of $\mathbb{R}^n$. Prove that $X + Y = X$ if and only if $Y \subseteq X$.

   Remark. This mirrors the arithmetic property of sets where $A \cup B = A \iff B \subseteq A$, replacing the union with the subspace sum.

5. Finite Intersections. We have established that the intersection of two subspaces is a subspace.

(a) By induction, prove that the intersection of any finite collection of subspaces $W_1, \dots, W_k$ is a subspace of $\mathbb{R}^n$.

(b) Is the complement $\mathbb{R}^n \setminus W$ ever a subspace? Prove that if W is a proper subspace (i.e., $W \neq \mathbb{R}^n$), its complement is never a subspace.

6. Symmetric Decomposition. A matrix A is called *symmetric* if $A^T = A$ and *skew-symmetric* if $A^T = -A$.

(a) Let A be any square matrix. Prove that $S = A + A^T$ is symmetric and $K = A - A^T$ is skew-symmetric.

(b) Deduce that any square matrix A can be uniquely expressed as the sum of a symmetric matrix and a skew-symmetric matrix.

(c) **Subspace Property:** Prove that the set of all $n \times n$ symmetric matrices is a subspace of $\mathbb{R}^{n\times n}$.

7. Commuting with Diagonals. Let D be an $n \times n$ *diagonal* matrix with *distinct* entries on the main diagonal (i.e., $D_{ii} \neq D_{jj}$ for $i \neq j$). Let A be an $n \times n$ matrix.

(a) Compute the (i, j)-th entry of the product AD and the product DA in terms of A_{ij} and the diagonal entries D_{kk}.

(b) Prove that if A commutes with D (i.e., $AD = DA$), then A must itself be a diagonal matrix (i.e., $A_{ij} = 0$ for all $i \neq j$).

8. Matrix Arithmetic and Commutativity. Let A and B be the following matrices:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

(a) Compute the products AB and BA. Are they equal?

(b) Compute the vector $A\mathbf{x}$ where $\mathbf{x} = [1, -2, 1]^T$.

(c) Find the matrix X such that $2(A - X) + 3B = \mathbf{0}$.

9. Nilpotency and Matrix Powers. A matrix N is called *nilpotent* if there exists a positive integer k such that $N^k = \mathbf{0}$. The smallest such k is called the index of nilpotency. Consider the matrix:

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

(a) Calculate N^2 and N^3. What is the index of nilpotency for N?

(b) Compute $(I - N)(I + N + N^2)$.

(c) Generalise the result from (b). If $N^k = \mathbf{0}$, prove that the matrix $(I - N)$ is invertible and find an expression for its inverse in terms of powers of N.

10. The Trace of a Matrix. The *trace* of a square matrix $A \in \mathbb{R}^{n\times n}$, denoted $\operatorname{tr}(A)$, is the sum of its diagonal entries:

$$\operatorname{tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

Using the summation definition of matrix multiplication (or index notation), prove the following properties:

(a) $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.

(b) $\operatorname{tr}(cA) = c \operatorname{tr}(A)$ for any scalar c.

(c) **Cyclic Property:** $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, even though $AB \neq BA$ in general.

(d) Show that for any distinct matrices A, B, it is impossible to have $AB - BA = I_n$.

Remark. This result has profound implications in quantum mechanics, demonstrating that bounded operators for position and momentum cannot satisfy the canonical commutation relation.

11. **The Centre of the Matrix Ring.** A matrix $A \in \mathbb{R}^{n\times n}$ is said to be in the *centre* of the algebra if it commutes with *every* matrix in $\mathbb{R}^{n\times n}$. That is, $AX = XA$ for all $X \in \mathbb{R}^{n\times n}$. Prove that A must be a scalar multiple of the identity matrix, i.e., $A = \lambda I_n$ for some $\lambda \in \mathbb{R}$.

    Remark. Hint: Test the condition $AE_{ij} = E_{ij}A$ using the standard basis matrices. Recall that right-multiplication AE_{ij} moves column i to column j, while left-multiplication $E_{ij}A$ moves row j to row i.

12. **Idempotent Matrices.** A matrix $P \in \mathbb{R}^{n\times n}$ is called *idempotent* if $P^2 = P$. Such matrices usually represent projection operators.

    (a) Prove that if P is idempotent, then $I - P$ is also idempotent.

    (b) Show that $P(I - P) = (I - P)P = \mathbf{0}$.

    (c) If P is idempotent and invertible, prove that $P = I$.

13. **The Fixed Point Subspace.** Let $A \in \mathbb{R}^{n\times n}$ be a fixed matrix. We are interested in vectors that are not changed by the transformation A. Let $W = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{x}\}$.

    (a) Prove that the condition $A\mathbf{x} = \mathbf{x}$ is equivalent to the homogeneous system $(A - I)\mathbf{x} = \mathbf{0}$.

    (b) Use the Subspace Criterion (or the result from Exercise 1) to prove that W is a subspace of $\mathbb{R}^n$.

    (c) Let $A = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$. Find a non-zero vector in W.

14. **Triangular Matrices.**

    (a) Let A and B be two $n \times n$ upper triangular matrices. Use the summation definition of the matrix product to prove that $C = AB$ is also upper triangular.

        Remark. Consider the entry $C_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$. Split the range of the sum or analyse when the terms must be zero.

    (b) Prove that the diagonal entries of the product are simply the products of the corresponding diagonal entries: $(AB)_{ii} = A_{ii}B_{ii}$.

15. $\star$ **The Adjoint Property.** The transpose allows us to move a matrix from one side of a dot product to the other. Prove that for any matrix $A \in \mathbb{R}^{m\times n}$ and vectors $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m$:

    $$A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^T\mathbf{y}.$$

    Remark. Recall that the dot product $\mathbf{u} \cdot \mathbf{v}$ can be written as the matrix product $\mathbf{u}^T\mathbf{v}$. Use the property $(AB)^T = B^T A^T$.

16. $\star$ **The Geometry of Subspaces in $\mathbb{R}^2$.** We intuitively understand that subspaces are "flat" sheets passing through the origin. Prove that the *only* non-trivial subspaces of $\mathbb{R}^2$ are lines passing through the origin. That is, any subspace $L \subset \mathbb{R}^2$ where $L \neq \{\mathbf{0}\}$ and $L \neq \mathbb{R}^2$ must be of the form:

    $$L = \{[x, y]^T \in \mathbb{R}^2 \mid ax + by = 0\}$$

    for some fixed scalars a, b, not both zero.

17. $\star$ **Counter-examples in Topology and Algebra.** To test the boundaries of the definition of a subspace, determine (with justification) whether the following sets X are subspaces of their respective parent spaces.

    (a) The set of vectors with rational components: $X = \{[r, s, t]^T \in \mathbb{R}^3 \mid r, s, t \in \mathbb{Q}\}$.

    (b) The set of vectors bounded within a circle: $X = \{[x, y]^T \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$.

    (c) The set of discrete integers: $X = \{\dots, -2, -1, 0, 1, 2, \dots\} \subseteq \mathbb{R}$.

    (d) The union of the axes: $X = \{[x, y]^T \in \mathbb{R}^2 \mid xy = 0\}$.

Chapter 4

Linear Spaces

In the previous chapters, we introduced the machinery of vectors and matrices. We now return to the central problem armed with these new tools. This chapter formalises the connection between algebraic solutions and geometric concepts such as spanning and linear independence. These ideas culminate in the "Column Correspondence Property", a powerful tool for understanding the structure of vector spaces.

4.1 Systems of Linear Equations Revisited

A system of linear equations can be compactly written as a single matrix equation $A\mathbf{x} = \mathbf{b}$. This shift from a list of equations to a single object allows us to treat the entire system as a relationship between vectors.

***Proposition* 4.1.1.** *Matrix Form of a Linear System.* Consider a system of m equations in n variables $x_1, \ldots, x_n$:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

This is equivalent to the matrix equation $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

The set of all vectors $\mathbf{x}$ satisfying this equation is the *solution set*, denoted $\mathrm{Sol}(A, \mathbf{b})$

$$\mathrm{Sol}(A, \mathbf{b}) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b}\}.$$

Example 4.1.1. Vector Form of Solutions. Consider the system:

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 4 \\ x_1 - x_2 + 4x_3 &= 1 \end{aligned}$$

Using Gaussian elimination, we find the general solution:

$$x_1 = 2 - 3x_3, \quad x_2 = 1 + x_3, \quad x_3 \text{ is free.}$$

In vector notation, the solution set is:

$$\mathbf{x} = \begin{bmatrix} 2 - 3x_3 \\ 1 + x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}.$$

Geometrically, this describes a line in $\mathbb{R}^3$ passing through the point $(2,1,0)$ and parallel to the vector $(-3,1,1)$.

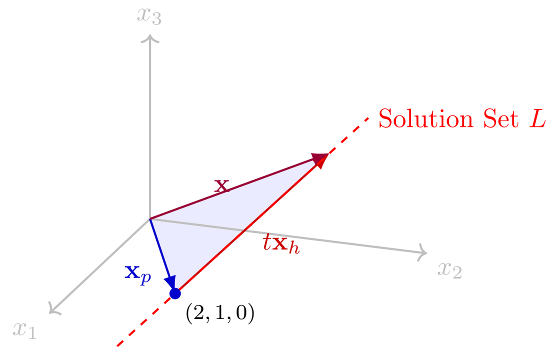


Figure 4.1: The geometry of the general solution. The solution set forms a line (red). Any solution vector $\mathbf{x}$ (purple) is formed by the vector addition triangle of the particular solution $\mathbf{x}_p$ (blue) and a scaled homogeneous vector $t\mathbf{x}_h$ (red segment).

Simultaneous Resolution of Systems

Often in applications (such as finding a matrix inverse), we need to solve multiple systems $A\mathbf{x} = \mathbf{b}_1, A\mathbf{x} = \mathbf{b}_2, \ldots, A\mathbf{x} = \mathbf{b}_k$ that share the same coefficient matrix but have different right-hand sides. Rather than performing Gaussian elimination k times, we can solve them all simultaneously.

***Proposition* 4.1.2.** *Simultaneous Solution.* Let $A \in \mathbb{R}^{m\times n}$. The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are solutions to the systems $A\mathbf{v}_i = \mathbf{b}_i$ (for $i = 1, \ldots, k$) if and only if the matrix $V = [\mathbf{v}_1 \mid \cdots \mid \mathbf{v}_k]$ satisfies $AV = B$, where $B = [\mathbf{b}_1 \mid \cdots \mid \mathbf{b}_k]$.

Proof. This follows directly from the definition of matrix multiplication. If $V = [\mathbf{v}_1 \mid \cdots \mid \mathbf{v}_k]$, then

$$AV = A[\mathbf{v}_1 \mid \cdots \mid \mathbf{v}_k] = [A\mathbf{v}_1 \mid \cdots \mid A\mathbf{v}_k].$$

Thus $AV = B$ if and only if $A\mathbf{v}_i = \mathbf{b}_i$ for all i. ■

Remark. To compute this efficiently, we form the "super-augmented" matrix $[A \mid \mathbf{b}_1 \ldots \mathbf{b}_k]$ and reduce it to $[\text{rref}(A) \mid \mathbf{c}_1 \ldots \mathbf{c}_k]$. If A is invertible (a concept we will define precisely later), the left side becomes I, and the right side contains the unique solutions.

Example 4.1.2. Simultaneous Solutions. To solve $A\mathbf{x} = \mathbf{b}_1$ and $A\mathbf{y} = \mathbf{b}_2$ for

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 3 \\ 7 \end{bmatrix},$$

we reduce the augmented matrix $[A \mid \mathbf{b}_1 \mid \mathbf{b}_2]$:

$$\left[\begin{array}{cc|c|c} 1 & 1 & 2 & 3 \\ 2 & 3 & 5 & 7 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|c|c} 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 \end{array}\right] \xrightarrow{R_1 - R_2} \left[\begin{array}{cc|c|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \end{array}\right].$$

Thus $\mathbf{x} = (1,1)$ and $\mathbf{y} = (2,1)$.

4.2 Elementary Matrices

Gaussian elimination relies on three elementary row operations. We can represent these operations algebraically as matrix multiplications. This perspective is crucial for theoretical developments, particularly in defining determinants and understanding matrix factorisations.

Definition 4.2.1. ***Elementary Matrix***. An *elementary matrix* is a matrix obtained by applying a single elementary row operation to the identity matrix I. The three types correspond to the three row operations:

1. **Row Replacement** ($R_i \to R_i + \alpha R_j$): The matrix $E_{ij}(\alpha)$ has 1s on the diagonal, α in position (i, j), and 0s elsewhere.
2. **Scaling** ($R_i \to \alpha R_i, \alpha \neq 0$): The matrix $D_i(\alpha)$ is diagonal with α in position (i, i) and 1s elsewhere.
3. **Interchange** ($R_i \leftrightarrow R_j$): The matrix P_{ij} is the identity with rows i and j swapped.

Theorem **4.2.1. Row Operations as Multiplication**. Performing an elementary row operation on a matrix A is equivalent to multiplying A on the *left* by the corresponding elementary matrix E.

Proof. Let E be an elementary matrix obtained by applying an operation ρ to I. That is, $E = \rho(I)$. Recall that $A = IA$. Since row operations are linear (row combination), applying ρ to the product IA is equivalent to applying it to the first factor:

$$\rho(A) = \rho(IA) = \rho(I)A = EA.$$

This can also be verified explicitly using the standard basis notation E_{ij} introduced in the previous chapter. ∎

Example 4.2.1. Elementary Matrices. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

- **Replacement:** Add 3 times row 1 to row 2.

$$E = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \implies EA = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 3a + c & 3b + d \end{bmatrix}.$$

- **Scaling:** Scale row 2 by 7.

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 7 \end{bmatrix} \implies EA = \begin{bmatrix} 1 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 7c & 7d \end{bmatrix}.$$

- **Swap:** Swap row 1 and row 2.

$$E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \implies EA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}.$$

Remark. Multiplying by elementary matrices on the *right* performs the corresponding *column* operations. For example, $AE_{ij}(\alpha)$ adds α times column i to column j.

Since every row operation corresponds to an elementary matrix, the entire Gaussian elimination process can be expressed as a product of matrices.

Proposition **4.2.1.** *Matrix Decomposition of RREF.* For any matrix A, there exists a sequence of elementary matrices $E_1, E_2, \ldots, E_k$ such that:

$$E_k \ldots E_2 E_1 A = \text{rref}(A).$$

Since elementary row operations are reversible, each E_i is invertible. This implies $A = E_1^{-1} \ldots E_k^{-1}\, \text{rref}(A)$.

4.3 Spanning and Linear Independence

The two central questions in linear algebra concerning a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ are:

1. *Existence:* Can every vector in the space be built from these vectors? (Spanning)
2. *Uniqueness:* Is the construction of a vector from these building blocks unique? (Linear Independence)

Spanning Sets

We formally define the set of all possible vectors that can be constructed from a given collection.

Definition 4.3.1. ***Span***. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. The *span* of S, denoted $\text{span}(S)$, is the set of all possible linear combinations of the vectors in S:

$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \{c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \mid c_i \in \mathbb{R}\}.$$

If $W = \text{span}(S)$, we say that S *spans* or *generates* W.

Recall from Theorem 3.4.1 that $A\mathbf{x}$ is a linear combination of the columns of A. This leads to a crucial connection between spanning and linear systems.

Proposition **4.3.1.** *Span and Consistency.* A vector $\mathbf{b}$ lies in $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ if and only if the linear system with augmented matrix $[\mathbf{v}_1 \ldots \mathbf{v}_k \mid \mathbf{b}]$ is consistent.

Proof. The vector equation $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{b}$ is equivalent to the matrix equation $A\mathbf{c} = \mathbf{b}$, where $A = [\mathbf{v}_1 \ldots \mathbf{v}_k]$. A solution $\mathbf{c}$ exists if and only if the system is consistent. ■

Example 4.3.1. Checking the Span. Let $\mathbf{v}_1 = (1, 1, 0)^T$, $\mathbf{v}_2 = (0, 1, 1)^T$, and $\mathbf{b} = (2, 3, 1)^T$. Is $\mathbf{b} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$? We form the augmented matrix and reduce:

$$\left[\begin{array}{cc|c} 1 & 0 & 2 \\ 1 & 1 & 3 \\ 0 & 1 & 1 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{array}\right] \xrightarrow{R_3 - R_2} \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right].$$

The system is consistent ($c_1 = 2, c_2 = 1$). Thus $\mathbf{b} = 2\mathbf{v}_1 + \mathbf{v}_2$, and $\mathbf{b} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

Theorem **4.3.1. Spanning $\mathbb{R}^m$.** Let A be an $m \times n$ matrix. The columns of A span $\mathbb{R}^m$ if and only if A has a pivot position in every row.

Proof. If A has a pivot in every row, then for any $\mathbf{b}$, the augmented matrix $[A \mid \mathbf{b}]$ cannot have a pivot in the last column (since the pivot is already in the A part of that row). Thus, the system is always consistent. Conversely, if there is a zero row in $\text{rref}(A)$, say row m, we can choose a $\mathbf{b}$ such that $[A \mid \mathbf{b}]$ reduces to a form with $[0 \ldots 0 \mid 1]$ in the last row, making the system inconsistent. ■

Linear Independence

The second fundamental question asks whether any vectors in our set are redundant. If a vector can be built from the others, it adds nothing to the span.

Definition 4.3.2. ***Linear Independence***. A set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is *linearly independent* if the only solution to the homogeneous equation

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

is the trivial solution $c_1 = c_2 = \cdots = c_k = 0$. If there exists a non-trivial solution (where at least one $c_i \neq 0$), the set is *linearly dependent*.

Remark. Geometric Intuition:

- Two vectors are dependent if they lie on the same line (collinear).
- Three vectors are dependent if they lie on the same plane (coplanar).
- In general, a set is dependent if one vector lies within the span of the others.

Theorem **4.3.2. Characterisation of Dependence**. An indexed set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ with $k \geq 2$ is linearly dependent if and only if at least one vector $\mathbf{v}_j$ is a linear combination of the others.

Proof.

($\Rightarrow$) If dependent, there exist c_i not all zero such that $\sum c_i \mathbf{v}_i = \mathbf{0}$. Let $c_j \neq 0$. We can rearrange:

$$c_j \mathbf{v}_j = -\sum_{i \neq j} c_i \mathbf{v}_i \implies \mathbf{v}_j = \sum_{i \neq j} \left(-\frac{c_i}{c_j}\right) \mathbf{v}_i.$$

($\Leftarrow$) If $\mathbf{v}_j = \sum_{i \neq j} a_i \mathbf{v}_i$, then $\left(\sum_{i \neq j} a_i \mathbf{v}_i\right) - 1\mathbf{v}_j = \mathbf{0}$. The coefficient of $\mathbf{v}_j$ is $-1 \neq 0$, so the relation is non-trivial.

■

***Proposition* 4.3.2.** *Independence of Matrix Columns.* The columns of a matrix A are linearly independent if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This occurs if and only if A has a pivot in every column (no free variables).

Proof. By definition, $A\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. If the only weights that produce $\mathbf{0}$ are all zero, the columns are independent. In Gaussian elimination, free variables correspond to non-trivial solutions. Thus, independence requires zero free variables. ■

Example 4.3.2. Checking Independence. Are the vectors $\mathbf{v}_1 = (1, 2, 3)^T$, $\mathbf{v}_2 = (4, 5, 6)^T$, $\mathbf{v}_3 = (2, 1, 0)^T$ independent? We reduce the matrix $A = [\mathbf{v}_1 \mathbf{v}_2 \mathbf{v}_3]$:

$$\begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 3 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & -6 & -6 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

There is no pivot in column 3. Thus, the variable x_3 is free, and the columns are *linearly dependent.* Specifically, the RREF implies $x_2 + x_3 = 0 \implies x_2 = -x_3$, and $x_1 + 4x_2 + 2x_3 = 0 \implies x_1 = -4(-x_3) - 2x_3 = 2x_3$. Setting $x_3 = 1$, we find a dependency relation: $2\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$.

The Column Correspondence Property

When we row-reduce a matrix A to its reduced row echelon form R, we drastically change the columns. However, the *linear relationships* between the columns are miraculously preserved. This phenomenon is known as the Column Correspondence Property (CCP).

***Theorem* 4.3.3. Column Correspondence Property**. Let A be a matrix and let $R = \text{rref}(A)$.

1. The linear dependence relations among the columns of A are identical to those among the columns of R. That is, $A\mathbf{c} = \mathbf{0}$ if and only if $R\mathbf{c} = \mathbf{0}$.
2. A column $\mathbf{a}_j$ is a linear combination of the pivot columns of A with coefficients c_i if and only if the corresponding column $\mathbf{r}_j$ is the *same* linear combination of the pivot columns of R.

Proof. Recall from the previous chapter that $R = EA$ for some invertible matrix E (a product of elementary matrices).

$$A\mathbf{c} = \mathbf{0} \iff (EA)\mathbf{c} = E\mathbf{0} \iff R\mathbf{c} = \mathbf{0}.$$

Since the homogeneous equations share the same solution set, the coefficients $\mathbf{c}$ that create a dependency in A are exactly those that create a dependency in R. ■

This property allows us to inspect the simple matrix R to deduce facts about the complex matrix A.

Example 4.3.3. Using CCP to Prune a Spanning Set. Suppose we want to find a linearly independent subset of $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ that spans the same space, where

$$A = [\mathbf{v}_1 \mathbf{v}_2 \mathbf{v}_3 \mathbf{v}_4] = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 4 & 7 & 3 \\ 1 & 2 & 4 & 2 \end{bmatrix}.$$

The RREF is:

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

The pivots are in columns 1 and 3.

1. **Identification:** By the CCP, the pivot columns of A (columns 1 and 3) form a linearly independent set. Thus $\{\mathbf{v}_1, \mathbf{v}_3\}$ is linearly independent.
2. **Dependencies:** The non-pivot columns of R reveal the dependencies:
   - Column 2 of R is $2\mathbf{e}_1$. Thus $\mathbf{r}_2 = 2\mathbf{r}_1$. By CCP, $\mathbf{v}_2 = 2\mathbf{v}_1$.
   - Column 4 of R is $-2\mathbf{e}_1 + 1\mathbf{e}_2$ (in terms of pivot columns $\mathbf{r}_1, \mathbf{r}_3$). Thus $\mathbf{r}_4 = -2\mathbf{r}_1 + \mathbf{r}_3$. By CCP, $\mathbf{v}_4 = -2\mathbf{v}_1 + \mathbf{v}_3$.

Therefore, $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_3\}$, and $\{\mathbf{v}_1, \mathbf{v}_3\}$ is a basis for this span.

***Proposition* 4.3.3.** *Pivot Columns as a Basis.* The pivot columns of a matrix A form a linearly independent set that spans the *column space* of A (the set of all linear combinations of its columns). The non-pivot columns are redundant.

Proof. In R, the pivot columns are standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots$ (potentially permuted). These are clearly independent. By CCP, the corresponding columns in A are independent. Furthermore, every non-pivot column in R has entries only in rows with pivots, meaning it is a linear combination of the pivot columns to its left. By CCP, the same holds for A. Thus, removing non-pivot columns does not shrink the span. ■

4.4 Bases and Dimension

We have seen that a set of vectors can span a subspace, but it might contain redundant information (linear dependencies). Conversely, a set can be linearly independent but fail to span the entire subspace. The "Goldilocks" set (a set that spans just enough and has no redundancies), is called a basis.

Definition 4.4.1. *Basis*. A *basis* for a vector space (or subspace) V is a set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in V that satisfies two conditions:

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

If a vector space V has a basis consisting of n vectors, then every basis for V has exactly n vectors. This invariant number is intrinsic to the space.

Definition 4.4.2. *Dimension*. The *dimension* of a finite-dimensional vector space V, denoted $\dim(V)$, is the number of vectors in any basis for V.

- The zero subspace $\{\mathbf{0}\}$ has dimension 0.
- $\mathbb{R}^n$ has dimension n.

Matrix Subspaces

Associated with every $m \times n$ matrix A are three fundamental subspaces. Their dimensions reveal the solvability of the system $A\mathbf{x} = \mathbf{b}$.

Definition 4.4.3. *Fundamental Subspaces*. Let $A \in \mathbb{R}^{m \times n}$.

1. **Column Space** (Col(A)): The subspace of $\mathbb{R}^m$ spanned by the columns of A.

   $$\text{Col}(A) = \text{span}\{\text{col}_1(A), \ldots, \text{col}_n(A)\}.$$

   The dimension of the column space is called the *rank* of A, denoted rank(A).
2. **Row Space** (Row(A)): The subspace of $\mathbb{R}^n$ (represented as row vectors) spanned by the rows of A.

   $$\text{Row}(A) = \text{span}\{\text{row}_1(A), \ldots, \text{row}_m(A)\}.$$

3. **Null Space** (Null(A)): The set of all solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

   $$\text{Null}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}.$$

   The dimension of the null space is called the *nullity* of A, denoted nullity(A).

Computing Bases

We can find bases for these subspaces systematically using Gaussian elimination.

Basis for the Column Space We cannot simply use the columns of rref(A) as a basis for Col(A) because row operations change the column space (e.g., they can zero out a row). However, the Column Correspondence Property ensures that the *positions* of the pivot columns remain invariant.

***Proposition* 4.4.1.** *Basis for* Col(A). The pivot columns of the original matrix A form a basis for Col(A).

Proof. By the CCP, the linear independence of the pivot columns in rref(A) (which are standard basis vectors) implies the independence of the corresponding columns in A. Furthermore, the non-pivot columns are linear combinations of the pivot columns, so removing them does not change the span. ■

Example 4.4.1. Column Space Basis. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 7 \end{bmatrix}$. Reducing A yields $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. The pivots are in columns 1 and 3. Thus, a basis for Col(A) is $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 7 \end{bmatrix} \right\}$. Note that column 2 is $2\times$ column 1, which is redundant.

Basis for the Row Space Unlike the column space, the row space *is* preserved under row operations.

***Proposition* 4.4.2.** *Basis for* Row(A). The non-zero rows of any row-echelon form (REF or RREF) of A form a basis for Row(A).

Proof. Row operations are linear combinations of rows, so they do not escape the span of the original rows. Thus Row(A) = Row(rref(A)). The non-zero rows of the RREF are linearly independent (due to the staircase pivot structure), so they form a basis. ■

Basis for the Null Space The null space is explicitly found by solving $A\mathbf{x} = \mathbf{0}$.

***Proposition* 4.4.3.** *Basis for* Null(A). To find a basis for Null(A):

1. Solve $A\mathbf{x} = \mathbf{0}$ to find the general solution in terms of free variables.
2. Decompose the general solution into a linear combination of vectors weighted by the free variables.
3. These vectors form a basis for Null(A).

Example 4.4.2. Null Space Basis. Let $A = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{bmatrix}$. This is already in RREF. Basic variables: x_1, x_3. Free variables: x_2, x_4. Equations: $x_1 + 2x_2 + 3x_4 = 0 \implies x_1 = -2x_2 - 3x_4$, and $x_3 + 4x_4 = 0 \implies x_3 = -4x_4$. Vector form:

$$\mathbf{x} = \begin{bmatrix} -2x_2 - 3x_4 \\ x_2 \\ -4x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

The basis is $\left\{(-2, 1, 0, 0)^T, (-3, 0, -4, 1)^T\right\}$.

The Rank-Nullity Theorem

There is a beautiful conservation law linking the dimensions of these subspaces.

***Theorem* 4.4.1. Rank-Nullity Theorem.** For any matrix $A \in \mathbb{R}^{m \times n}$:

$$\text{rank}(A) + \text{nullity}(A) = n.$$

That is, the number of pivot columns plus the number of free variable columns equals the total number of columns.

Proof. The columns of A are partitioned into pivot columns and non-pivot columns.

- rank(A) is the number of pivot columns (dimension of Col(A)).
- nullity(A) is the number of free variables, which corresponds exactly to the number of non-pivot columns.

Since every column is either a pivot column or a non-pivot column, the sum is n. ■

***Corollary* 4.4.1.** *Row Rank equals Column Rank.* For any matrix A, $\dim(\text{Col}(A)) = \dim(\text{Row}(A))$.

Proof. The dimension of Col(A) is the number of pivots. The dimension of Row(A) is the number of non-zero rows in the RREF. By the definition of RREF, each non-zero row has exactly one leading pivot. Thus, the number of pivots equals the number of non-zero rows. ■

General Linear Systems Theory

We can now fully characterize the solution set of any linear system $A\mathbf{x} = \mathbf{b}$ using the language of subspaces.

***Proposition* 4.4.4.** *Structure of Solutions.* Let $A\mathbf{x} = \mathbf{b}$ be a consistent linear system. The general solution $\mathbf{x}$ can be written as:

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h,$$

where $\mathbf{x}_p$ is a *particular solution* (any specific vector satisfying $A\mathbf{x}_p = \mathbf{b}$) and $\mathbf{x}_h$ is a solution to the homogeneous equation $A\mathbf{x}_h = \mathbf{0}$ (i.e., $\mathbf{x}_h \in \text{Null}(A)$).

Proof. If $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x}_p = \mathbf{b}$, then $A(\mathbf{x} - \mathbf{x}_p) = A\mathbf{x} - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Thus $\mathbf{x} - \mathbf{x}_p = \mathbf{x}_h \in \text{Null}(A)$, so $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$. Conversely, if $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, then $A\mathbf{x} = A\mathbf{x}_p + A\mathbf{x}_h = \mathbf{b} + \mathbf{0} = \mathbf{b}$. ■

This structure explains why the general solution involves "parameters". The parameters are simply the coefficients of the basis vectors of the null space.

***Corollary* 4.4.2.** *Degrees of Freedom.* If a consistent system $A\mathbf{x} = \mathbf{b}$ has a solution, the number of free parameters in the general solution is exactly nullity(A).

Example 4.4.3. Summary Example. Consider $A\mathbf{x} = \mathbf{b}$ where A is 3×5 and has rank 3.

- Since rank is 3 (max possible for 3 rows), the columns span $\mathbb{R}^3$. A solution exists for *all* $\mathbf{b}$.
- By Rank-Nullity, nullity$(A) = 5 - 3 = 2$. The solution set is a 2-dimensional "plane" (shifted by $\mathbf{x}_p$) in $\mathbb{R}^5$.

4.5 Rank and Elementary Transformations

In this final section, we investigate the relationship between the row rank, the column rank, and elementary transformations. We will establish the fundamental result that row rank and column rank are equal for any matrix.

Invariance of Rank We begin by showing that elementary operations do not change the dimension of the row or column spaces.

***Theorem* 4.5.1. Rank Invariance Under Row Operations**. Let A be a matrix and let B be the result of applying an elementary row operation to A. Then:

1. $\text{row\_rank}(A) = \text{row\_rank}(B)$.
2. $\text{col\_rank}(A) = \text{col\_rank}(B)$.

Proof.

1. **Row Rank:** The rows of B are linear combinations of the rows of A. Thus $\text{Row}(B) \subseteq \text{Row}(A)$. Since elementary row operations are reversible, the rows of A are linear combinations of the rows of B, so $\text{Row}(A) \subseteq \text{Row}(B)$. Therefore, the subspaces are identical, and their dimensions are equal.
2. **Column Rank:** An elementary row operation corresponds to left-multiplication by an invertible matrix E: $B = EA$. Let $\mathbf{c}_1, \ldots, \mathbf{c}_k$ be a subset of the columns of A. If these are linearly dependent, there exist scalars x_i not all zero such that $\sum x_i\mathbf{c}_i = \mathbf{0}$. Multiplying by E gives $\sum x_i(E\mathbf{c}_i) = \mathbf{0}$, so the corresponding columns of B are dependent. Conversely, if $\sum x_i(E\mathbf{c}_i) = \mathbf{0}$, multiplying by E^{-1} gives $\sum x_i\mathbf{c}_i = \mathbf{0}$. Thus, linear independence relationships among columns are preserved. Consequently, the dimension of the column space (maximum number of independent columns) remains unchanged.

■

The analogous result holds for elementary *column* operations (which correspond to right-multiplication by elementary matrices).

***Corollary* 4.5.1.** *Rank Invariance Under Column Operations.* Elementary column operations do not alter the row rank or column rank of a matrix.

Proof. This follows from the previous theorem by considering the transpose A^T. ■

Equality of Row and Column Rank We are now ready to prove one of the most surprising and important theorems in linear algebra.

***Theorem* 4.5.2. Equality of Ranks**. For any $m \times n$ matrix A, the row rank and the column rank are equal $\dim(\text{Row}(A)) = \dim(\text{Col}(A))$. This common value is simply called the **rank** of A.

Proof. Let r be the row rank of A. We can transform A into its reduced row echelon form R using elementary row operations. By rank invariance, $\text{row\_rank}(R) = r$ and $\text{col\_rank}(R) = \text{col\_rank}(A)$. In R, the non-zero rows are linearly independent and form a basis for the row space. Thus, there are exactly r non-zero rows. By the definition of RREF, each non-zero row has a leading pivot (1). These pivots appear in distinct columns. These r pivot columns are standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r$ (up to permutation of rows) and are clearly linearly independent. The non-pivot columns are linear combinations of the pivot columns. Thus, the dimension of the column space of R is exactly r. Therefore, $\text{col\_rank}(A) = \text{col\_rank}(R) = r = \text{row\_rank}(A)$. ■

4.6 Dimension and Basis Construction

We conclude this chapter by summarising how to construct bases for subspaces of $\mathbb{R}^n$ and stating some fundamental properties of dimension.

***Theorem* 4.6.1. Existence of Basis**. Every subspace V of $\mathbb{R}^n$ has a dimension d such that $0 \leq d \leq n$. If $d > 0$, V has a basis consisting of d vectors.

Proof. If $V = \{\mathbf{0}\}$, its dimension is 0. If V contains a non-zero vector $\mathbf{v}_1$, let $S = \{\mathbf{v}_1\}$. If S spans V, we are done. If not, pick $\mathbf{v}_2 \in V \setminus \text{span}(S)$. The set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent. We repeat this process. Since any set of $n+1$ vectors in $\mathbb{R}^n$ is dependent, the process must terminate at some $d \leq n$. ■

***Proposition* 4.6.1.** *Monotonicity of Dimension.* If U and V are subspaces of $\mathbb{R}^n$ such that $U \subseteq V$, then $\dim(U) \leq \dim(V)$. Furthermore, if $\dim(U) = \dim(V)$, then $U = V$.

Proof. Let $\mathcal{B}_U$ be a basis for U. These vectors are linearly independent in V. We can extend $\mathcal{B}_U$ to a basis for V. Thus the size of the basis for U cannot exceed the size of the basis for V. If the sizes are equal, the basis for U already spans V, so $U = V$. ■

Example 4.6.1. Finding a Basis. Find a basis for the subspace V of $\mathbb{R}^4$ generated by:

$$\mathbf{v}_1 = (1, -1, 2, 3), \quad \mathbf{v}_2 = (4, 5, 8, -9), \quad \mathbf{v}_3 = (2, 1, 4, -1), \quad \mathbf{v}_4 = (2, -5, 4, 13).$$

Construct a matrix A with these vectors as *rows*:

$$A = \begin{bmatrix} 1 & -1 & 2 & 3 \\ 4 & 5 & 8 & -9 \\ 2 & 1 & 4 & -1 \\ 2 & -5 & 4 & 13 \end{bmatrix}$$

Reduce to REF to find a basis for the row space:

$$A \sim \begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 9 & 0 & -21 \\ 0 & 3 & 0 & -7 \\ 0 & -3 & 0 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 3 & 0 & -7 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The non-zero rows $\mathbf{r}_1 = (1, -1, 2, 3)$ and $\mathbf{r}_2 = (0, 3, 0, -7)$ form a basis for V. Thus $\dim(V) = 2$. Note that these basis vectors are *not* the original vectors. If a basis consisting of a subset of the original vectors is required, one would form a matrix with the vectors as *columns* and select the pivot columns.

4.7 Exercises

1. **Bases and Parameters.** Consider the matrix A_k dependent on a scalar parameter k:

$$A_k = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & k & 3 \\ 2 & 3 & 4 & k^2 \end{bmatrix}$$

(a) Determine the rank and nullity of A_k for all possible values of k.
(b) For the case $k = 3$, find a basis for the column space consisting of column vectors of A_k, and a basis for the null space.

2. Basis Transformation. Let $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be a basis for $\mathbb{R}^3$. Determine whether the following sets are also bases for $\mathbb{R}^3$. Give a proof or a counterexample.

(a) The set of cumulative sums: $\{\mathbf{v}_1,\ \mathbf{v}_1 + \mathbf{v}_2,\ \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3\}$.
(b) The set of circular sums: $\{\mathbf{v}_1 + \mathbf{v}_2,\ \mathbf{v}_2 + \mathbf{v}_3,\ \mathbf{v}_3 + \mathbf{v}_1\}$.
(c) **Generalisation:** Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$. Let $\mathbf{u}_i = \mathbf{v}_i + \mathbf{v}_{i+1}$ for $1 \leq i < n$, and $\mathbf{u}_n = \mathbf{v}_n + \mathbf{v}_1$. For which values of n is the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ linearly dependent?

3. Rank Constraints and Geometry. The dimensions of a matrix impose strict limits on the geometry of its row and column spaces.

(a) Can a 3×4 matrix have all its column vectors linearly independent? Can it have all its row vectors linearly independent? Justify your answer using the definition of rank.
(b) If A is a 5×4 matrix and $\text{rank}(A) = 3$, can the columns of A be linearly independent? Can the rows?
(c) Prove that for any $m \times n$ matrix where $m \neq n$, it is impossible for both the row vectors to be linearly independent *and* the column vectors to be linearly independent.

4. The Inverse Problem: Constructing Constraints. Typically, we are given a matrix A and asked to find its null space. Consider the reverse problem: given a subspace, find a system of equations defining it. Let W be the subspace of $\mathbb{R}^4$ spanned by the vectors $\mathbf{w}_1 = (1, 2, 0, -1)$ and $\mathbf{w}_2 = (2, 3, 1, 0)$. Construct a matrix A such that $\text{Null}(A) = W$.

Remark. Hint: Use the property that $\mathbf{a} \cdot \mathbf{x} = 0$ for all $\mathbf{x} \in W$ if and only if $\mathbf{a}$ is orthogonal to both $\mathbf{w}_1$ and $\mathbf{w}_2$. The rows of A must be such vectors.

5. Preservation of Independence. Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ be a linearly independent set of vectors in $\mathbb{R}^n$.

(a) Let $\mathbf{y}$ be a vector such that $\mathbf{y} \notin \text{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$. Prove that the set $\{\mathbf{x}_1 + \mathbf{y}, \mathbf{x}_2 + \mathbf{y}, \ldots, \mathbf{x}_m + \mathbf{y}\}$ is linearly independent.
(b) Suppose instead that $\mathbf{y} = \sum_{i=1}^{m} \mathbf{x}_i$. Prove that $\{\mathbf{x}_1 + \mathbf{y}, \ldots, \mathbf{x}_m + \mathbf{y}\}$ is linearly independent if and only if $m + 1 \neq 0$. (Note: in $\mathbb{R}$, this is always true, but consider why this condition might matter in fields with finite characteristic).

6. The Vandermonde Rank. Let $x_1, x_2, \ldots, x_n$ be pairwise distinct real numbers. Consider the $n \times n$ Vandermonde matrix V:

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

Using elementary column and row operations, prove that $\text{rank}(V) = n$.

Remark. Hint: Use elementary column operations to create zeros in the first row, then factor out common terms to reduce the problem to a Vandermonde matrix of size $(n-1) \times (n-1)$. Proceed by induction.

7. Linear Maps on Pairs. Let $\mathbf{u}$ and $\mathbf{v}$ be linearly independent vectors in $\mathbb{R}^n$. Consider two new vectors defined by linear combinations:

$$\mathbf{x} = a\mathbf{u} + b\mathbf{v}, \quad \mathbf{y} = c\mathbf{u} + d\mathbf{v},$$

where $a, b, c, d \in \mathbb{R}$. Prove that $\{\mathbf{x}, \mathbf{y}\}$ is linearly independent if and only if $ad - bc \neq 0$.

Remark. This result establishes that the determinant determines invertibility for 2×2 coordinate transformations.

8. Rank of a Product. Let A be an $m \times n$ matrix and B be an $n \times p$ matrix.

(a) Prove that $\text{Col}(AB) \subseteq \text{Col}(A)$.
(b) Deduce that $\text{rank}(AB) \leq \text{rank}(A)$.
(c) By considering row spaces, prove that $\text{rank}(AB) \leq \text{rank}(B)$.

9. ⋆ **One-Sided Inverses and Rank.** A matrix A is *invertible* if there exists B such that $AB = I$ and $BA = I$. If A is not square, it cannot be invertible, but it may possess a one-sided inverse.

(a) **Left Inverse:** Prove that an $m \times n$ matrix A has a left inverse C (such that $CA = I_n$) if and only if $\text{rank}(A) = n$.

Remark. This requires $m \geq n$. Recall that $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$ for full column rank matrices.

(b) **Right Inverse:** Prove that an $m \times n$ matrix A has a right inverse D (such that $AD = I_m$) if and only if $\text{rank}(A) = m$.
(c) Construct 2×2 matrices A and B such that $\text{rank}(A) = \text{rank}(B) = 1$ but $\text{rank}(AB) = 0$.

10. Reversibility of Elementary Operations. We stated that every elementary row operation is reversible.

(a) For a row replacement operation $R_i \to R_i + kR_j$, write down the matrix E that performs this operation and the matrix E' that performs the inverse operation. Verify that $EE' = I$.
(b) For a row interchange $R_i \leftrightarrow R_j$, show that the corresponding elementary matrix P is its own inverse, i.e., $P^2 = I$.

11. The Basis Extension Theorem. Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a linearly independent set in $\mathbb{R}^n$ (where $k < n$).

(a) Prove that there exist vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n-k}$ such that the set $S \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_{n-k}\}$ forms a basis for $\mathbb{R}^n$.
(b) *Constructive Application:* Extend the independent set $\{(1, 0, 1), (0, 1, 1)\}$ to a basis for $\mathbb{R}^3$.

12. Affine Dimension. Let X be a subspace of $\mathbb{R}^n$. Let $\mathbf{u}$ be a non-zero vector in $\mathbb{R}^n$ such that $\mathbf{u} \notin X$. Define the set $Y = \{a\mathbf{u} + \mathbf{v} \mid a \in \mathbb{R}, \mathbf{v} \in X\}$.

(a) Prove that Y is a subspace of $\mathbb{R}^n$.
(b) Prove that $\dim(Y) = \dim(X) + 1$.
(c) Use this result to prove that if X is a subspace of $\mathbb{R}^n$ and $\dim(X) = n$, then $X = \mathbb{R}^n$.

13. Grassmann's Formula. Let U and W be subspaces of $\mathbb{R}^n$.

(a) Prove that $U \cap W = \{\mathbf{0}\}$ if and only if for every non-zero $\mathbf{u} \in U$ and $\mathbf{w} \in W$, the vectors $\mathbf{u}$ and $\mathbf{w}$ are linearly independent.
(b) Prove the dimension formula for the sum and intersection of subspaces $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$.

Remark. Hint: Start with a basis for $U \cap W$, extend it to a basis for U, and separately extend it to a basis for W. Show that the union of these vectors is a basis for $U + W$.

(c) Deduce that if U and W are planes through the origin in $\mathbb{R}^3$ (dimension 2), their intersection must contain a line (dimension at least 1).

14. ⋆ **The Left Null Space and Consistency.** We have defined three fundamental subspaces of a matrix $A \in \mathbb{R}^{m \times n}$. There is a fourth: the *left null space*, defined as $\text{Null}(A^T)$. This subspace lies in $\mathbb{R}^m$.

(a) Prove that $\mathbf{y} \in \text{Null}(A^T)$ if and only if $\mathbf{y}^T A = \mathbf{0}^T$.
(b) Prove the *Consistency Theorem*: The system $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{y}^T\mathbf{b} = 0$ for every $\mathbf{y} \in \text{Null}(A^T)$.

Remark. For the forward direction, multiply $A\mathbf{x} = \mathbf{b}$ by $\mathbf{y}^T$. For the reverse, the proof requires orthogonality concepts we will formalise later, but you may argue using row reduction: if the system is inconsistent, a row $[0 \ldots 0 \mid 1]$ appears. Interpret this row as a vector $\mathbf{y}$.

(c) Verify this theorem for the system in Figure 4.1 where $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$. Find a basis for $\text{Null}(A^T)$ (is it trivial?) and check the dot product condition.

Chapter 5

Inverse Matrices

We continue our exploration of matrix algebra by investigating the multiplicative inverse of a matrix. Just as the reciprocal $1/x$ allows us to solve $ax = b$ via $x = a^{-1}b$, the matrix inverse A^{-1} provides a powerful tool for solving linear systems $A\mathbf{x} = \mathbf{b}$.

5.1 Invertible Matrices

While every non-zero real number has a multiplicative inverse, not every non-zero matrix can be inverted. We restrict our attention to square matrices, as invertibility requires the matrix to map a space to itself bijectively.

Definition 5.1.1. ***Invertible Matrix***. A square matrix $A \in \mathbb{R}^{n\times n}$ is *invertible* (or *non-singular*) if there exists a matrix $B \in \mathbb{R}^{n\times n}$ such that:

$$AB = I_n \quad \text{and} \quad BA = I_n.$$

If such a matrix B exists, it is unique and is denoted by A^{-1}. If no such matrix exists, A is called *singular*.

Remark. Why "singular"? In the landscape of all $n \times n$ matrices, the non-invertible ones are rare—they form a "singular" subset of lower dimension, much like lines in a plane. Most random matrices are invertible.

Elementary Matrices are Invertible

Recall the elementary matrices introduced in the previous chapter. Since every elementary row operation is reversible, every elementary matrix must be invertible.

***Proposition* 5.1.1.** *Inverses of Elementary Matrices.* Every elementary matrix E is invertible, and its inverse E^{-1} is the elementary matrix of the same type that reverses the operation.

- **Replacement:** If E adds αR_j to R_i, then E^{-1} subtracts αR_j from R_i.

$$(E_{ij}(\alpha))^{-1} = E_{ij}(-\alpha).$$

- **Scaling:** If E scales R_i by $\alpha \neq 0$, then E^{-1} scales R_i by $1/\alpha$.

$$(D_i(\alpha))^{-1} = D_i(1/\alpha).$$

- **Interchange:** If E swaps R_i and R_j, then E^{-1} swaps them back (it is its own inverse).

$$(P_{ij})^{-1} = P_{ij}.$$

Example 5.1.1. Inverse of an Elimination Step. The matrix $E = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$ adds $3 \times \text{Row}_1$ to Row_2. Its inverse must subtract $3 \times \text{Row}_1$ from Row_2:

$$E^{-1} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}.$$

Verification:

$$EE^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3-3 & 1 \end{bmatrix} = I.$$

Uniqueness of Solutions

The existence of an inverse is inextricably linked to the uniqueness of solutions for linear systems.

***Theorem* 5.1.1. Invertibility and Unique Solutions**. Let $A \in \mathbb{R}^{n \times n}$. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$ if and only if A is invertible.

Proof.

($\Rightarrow$) Suppose $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Then $\text{rref}(A) = I$. By the decomposition theorem, $A = E_1^{-1} \dots E_k^{-1} I$. A product of invertible matrices is invertible, so A is invertible.

($\Leftarrow$) Suppose A is invertible. If $A\mathbf{x} = \mathbf{0}$, left-multiply by A^{-1}:

$$A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{0} \implies (A^{-1}A)\mathbf{x} = \mathbf{0} \implies I\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}.$$

■

***Proposition* 5.1.2.** *Left and Right Inverses.* For square matrices, a one-sided inverse implies a two-sided inverse.

1. If $BA = I$, then $AB = I$ (so $B = A^{-1}$).
2. If $AB = I$, then $BA = I$ (so $B = A^{-1}$).

Proof. Suppose $BA = I$. If $A\mathbf{x} = \mathbf{0}$, then $B(A\mathbf{x}) = B\mathbf{0} = \mathbf{0} \implies (BA)\mathbf{x} = \mathbf{0} \implies I\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$. Since $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, A is invertible. Let A^{-1} be its inverse. Then $B = B(AA^{-1}) = (BA)A^{-1} = IA^{-1} = A^{-1}$. Thus $AB = AA^{-1} = I$. ■

Properties of the Inverse

Matrix inversion interacts predictably with other matrix operations, though order matters.

***Theorem* 5.1.2. Algebra of Inverses**. Let $A, B \in \mathbb{R}^{n \times n}$ be invertible matrices and $c \neq 0$ be a scalar.

1. **Inverse of Inverse:** $(A^{-1})^{-1} = A$.
2. **Product:** $(AB)^{-1} = B^{-1}A^{-1}$. (Note the reversal of order).
3. **Scaling:** $(cA)^{-1} = \frac{1}{c}A^{-1}$.
4. **Transpose:** $(A^T)^{-1} = (A^{-1})^T$.

Proof. We verify the product property (item 2). We must show that $(B^{-1}A^{-1})$ is the inverse of (AB):

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = A(I)A^{-1} = AA^{-1} = I.$$

Similarly, $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}(I)B = B^{-1}B = I$. ■

5.2 Computing the Inverse

The 2x2 Formula

For 2×2 matrices, there is a simple explicit formula.

***Proposition* 5.2.1.** *Inverse of a 2x2 Matrix.* The matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $ad - bc \neq 0$. If so:

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

The quantity $ad - bc$ is called the *determinant* of A.

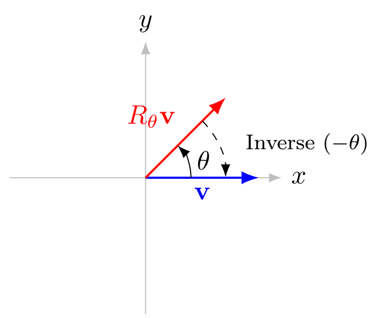


Figure 5.1: Visualising the inverse of a rotation matrix.

Example 5.2.1. Rotation Matrices. The rotation matrix $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ has determinant $\cos^2\theta - (-\sin^2\theta) = 1$. Its inverse is:

$$R_\theta^{-1} = \frac{1}{1} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = R_{(-\theta)}.$$

Geometrically, the inverse of rotating by θ is rotating by $-\theta$.

The General Algorithm (Gauss-Jordan)

For larger matrices, we use Gaussian elimination. The problem "Find X such that $AX = I$" is equivalent to solving n linear systems simultaneously:

$$A[\mathbf{x}_1 \mid \cdots \mid \mathbf{x}_n] = [\mathbf{e}_1 \mid \cdots \mid \mathbf{e}_n].$$

This can be done by reducing the super-augmented matrix $[A \mid I]$.

***Theorem* 5.2.1. Algorithm for Inversion.** Let $A \in \mathbb{R}^{n \times n}$. Form the augmented matrix $[A \mid I]$.

1. Apply the Gauss-Jordan algorithm to reduce $[A \mid I]$.
2. If A is invertible, the result will be $[I \mid A^{-1}]$.
3. If the left side reduces to a matrix with a zero row (i.e., $\text{rref}(A) \neq I$), then A is singular (not invertible).

Example 5.2.2. Computing a 3x3 Inverse. Find the inverse of $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ 4 & 4 & 4 \end{bmatrix}$. Form $[A \mid I]$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ 4 & 4 & 4 & 0 & 0 & 1 \end{array}\right].$$

Eliminate below pivots:

$$R_2 \to R_2 - 2R_1,\ R_3 \to R_3 - 4R_1 \implies \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & -2 & 1 & 0 \\ 0 & 4 & 4 & -4 & 0 & 1 \end{array}\right].$$

$$R_3 \to R_3 - 2R_2 \implies \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & -2 & 1 & 0 \\ 0 & 0 & 4 & 0 & -2 & 1 \end{array}\right].$$

Scale to identity:

$$R_2 \to \frac{1}{2}R_2,\ R_3 \to \frac{1}{4}R_3 \implies \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & -1/2 & 1/4 \end{array}\right].$$

Thus,

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1/2 & 0 \\ 0 & -1/2 & 1/4 \end{bmatrix}.$$

This algorithm provides a constructive proof that A is invertible if and only if $\text{rref}(A) = I$.

5.3 Matrices with Special Shapes

We have already encountered diagonal and triangular matrices in the context of Gaussian elimination. In this section, we formalise these concepts and introduce other structural properties—such as symmetry—that play a crucial role in both pure mathematics and applications like physics and engineering.

Symmetric and Antisymmetric Matrices

A matrix is symmetric if it is equal to its transpose. This implies the matrix must be square.

Definition 5.3.1. ***Symmetry***. Let $A \in \mathbb{R}^{n\times n}$.

1. A is *symmetric* if $A^T = A$. In components, $A_{ij} = A_{ji}$.
2. A is *antisymmetric* (or *skew-symmetric*) if $A^T = -A$. In components, $A_{ij} = -A_{ji}$.

Remark. For an antisymmetric matrix, the diagonal entries must satisfy $A_{ii} = -A_{ii}$, which implies $A_{ii} = 0$. Thus, all diagonal elements of an antisymmetric matrix are zero.

Example 5.3.1. Examples of Symmetry.

- Symmetric: I, $\mathbf{0}_{n\times n}$, E_{ii}, $\begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$.
- Antisymmetric: $\mathbf{0}_{n\times n}$, $\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$.
- Neither: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

The transpose operation interacts with matrix arithmetic in predictable ways, often referred to as the "socks-shoes" property for products (you put them on in one order, take them off in the reverse).

Proposition **5.3.1.** *Properties of Transpose.* Let $A, B \in \mathbb{R}^{n\times n}$ and $c \in \mathbb{R}$.

1. $(A^T)^T = A$.
2. $(AB)^T = B^T A^T$ (The reverse order law).
3. $(cA)^T = cA^T$.

4. $(A + B)^T = A^T + B^T$.
5. If A is invertible, $(A^T)^{-1} = (A^{-1})^T$.

***Theorem* 5.3.1. Symmetric-Antisymmetric Decomposition**. Every square matrix A can be uniquely expressed as the sum of a symmetric matrix S and an antisymmetric matrix K.

$$A = S + K, \quad \text{where } S = \frac{1}{2}(A + A^T) \text{ and } K = \frac{1}{2}(A - A^T).$$

Proof. We verify the properties:

- $S^T = \frac{1}{2}(A^T + (A^T)^T) = \frac{1}{2}(A^T + A) = S$ (Symmetric).
- $K^T = \frac{1}{2}(A^T - (A^T)^T) = \frac{1}{2}(A^T - A) = -K$ (Antisymmetric).
- $S + K = \frac{1}{2}(A + A^T + A - A^T) = A$.

■

Example 5.3.2. Decomposition Example. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Then $A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$.

$$S = \frac{1}{2}\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2.5 \\ 2.5 & 4 \end{bmatrix}.$$

$$K = \frac{1}{2}\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\right) = \begin{bmatrix} 0 & -0.5 \\ 0.5 & 0 \end{bmatrix}.$$

***Proposition* 5.3.2.** *Symmetry of $A^T A$.* For any matrix $A \in \mathbb{R}^{m \times n}$ (not necessarily square), the product $A^T A$ is always a symmetric $n \times n$ matrix.

Proof. $(A^T A)^T = A^T (A^T)^T = A^T A$. ■

Matrix Exponents

For a square matrix A, we can define powers just as we do for scalars.

Definition 5.3.2. *Matrix Powers*. Let $A \in \mathbb{R}^{n \times n}$.

- $A^0 = I_n$.
- $A^k = \underbrace{AA \dots A}_{k \text{ times}}$ for $k \in \mathbb{N}$.
- If A is invertible, $A^{-k} = (A^{-1})^k$.

Standard exponent laws hold: $A^p A^q = A^{p+q}$ and $(A^p)^q = A^{pq}$. However, due to non-commutativity, $(AB)^k \neq A^k B^k$ in general. Instead, $(AB)^2 = ABAB$. The binomial theorem $(A + B)^2 = A^2 + 2AB + B^2$ only holds if A and B commute ($AB = BA$).

***Proposition* 5.3.3.** *Powers of Symmetric Matrices.* If A is symmetric, then A^k is symmetric for all integer $k \geq 1$.

Proof. By induction. Base case $k = 1$ is true. If $(A^k)^T = A^k$, then $(A^{k+1})^T = (AA^k)^T = (A^k)^T A^T = A^k A = A^{k+1}$. ■

Diagonal and Triangular Matrices

Matrices with many zeros are computationally desirable. The most structured forms are diagonal and triangular.

Definition 5.3.3. ***Structured Matrices***. Let $A \in \mathbb{R}^{n\times n}$.

1. **Diagonal:** $A_{ij} = 0$ for all $i \neq j$.
2. **Upper Triangular:** $A_{ij} = 0$ for all $i > j$ (all entries below the diagonal are zero).
3. **Lower Triangular:** $A_{ij} = 0$ for all $i < j$ (all entries above the diagonal are zero).

Remark. A diagonal matrix is simultaneously upper and lower triangular.

***Proposition* 5.3.4.** *Multiplication of Structured Matrices.* Let A and B be $n \times n$ matrices.

1. If A, B are diagonal, then AB is diagonal, and $(AB)_{ii} = A_{ii}B_{ii}$.
2. If A, B are upper triangular, then AB is upper triangular.
3. If A, B are lower triangular, then AB is lower triangular.

Proof. We prove (2). Let A, B be upper triangular. Then $A_{ik} = 0$ if $i > k$ and $B_{kj} = 0$ if $k > j$. The (i, j) entry of the product is $\sum_k A_{ik}B_{kj}$. For a term to be non-zero, we need $i \leq k$ (from A) and $k \leq j$ (from B). Thus, we need $i \leq k \leq j$. If $i > j$, no such k exists, so the sum is zero. Thus AB is upper triangular. ∎

Block Matrices

It is often useful to partition a large matrix into smaller submatrices (blocks). We can treat these blocks as single algebraic units, provided the dimensions are compatible.

Definition 5.3.4. ***Block Multiplication***. Let A and B be partitioned matrices:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

If the column partitions of A match the row partitions of B (i.e., width of A_{11} = height of B_{11}, etc.), then the product is given by:

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

This rule generalises to partitions of any size. It essentially says "matrix multiplication works block-wise", treating the blocks as non-commutative scalars.

Example 5.3.3. Block Diagonal Multiplication. If A and B are block diagonal:

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}, \quad B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix}.$$

Then:

$$AB = \begin{bmatrix} A_1B_1 + 0 & 0 + 0 \\ 0 + 0 & 0 + A_2B_2 \end{bmatrix} = \begin{bmatrix} A_1B_1 & 0 \\ 0 & A_2B_2 \end{bmatrix}.$$

This property is fundamental in quantum mechanics and signal processing, where we decompose systems into independent subsystems.

Example 5.3.4. Inverse of a Block Triangular Matrix. Suppose we wish to invert $M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$, assuming A and C are invertible. Let $M^{-1} = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix}$. We require $MM^{-1} = I$:

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix} \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} = \begin{bmatrix} AX + BZ & AY + BW \\ CZ & CW \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

From the bottom row:

1. $CZ = 0 \implies Z = 0$ (since C is invertible).

2. $CW = I \implies W = C^{-1}$.

Substituting $Z = 0, W = C^{-1}$ into the top row:

1. $AX = I \implies X = A^{-1}$.
2. $AY + BC^{-1} = 0 \implies AY = -BC^{-1} \implies Y = -A^{-1}BC^{-1}$.

Thus,

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BC^{-1} \\ 0 & C^{-1} \end{bmatrix}.$$

5.4 LU Factorisation

We have seen that Gaussian elimination can be described as a sequence of matrix multiplications. This perspective leads to a powerful tool known as the *LU Factorisation*, which decomposes a matrix A into the product of a lower triangular matrix L and an upper triangular matrix U. This factorisation is particularly useful for solving linear systems $A\mathbf{x} = \mathbf{b}$ efficiently, especially when multiple right-hand sides $\mathbf{b}$ must be processed.

The Factorisation

The forward pass of Gaussian elimination transforms a matrix A into a row echelon form U. If we restrict ourselves to row replacement operations (adding a multiple of one row to a lower row), this process can be represented as:

$$E_k \dots E_2 E_1 A = U,$$

where each E_i is a lower triangular elementary matrix. Since the inverse of a lower triangular elementary matrix is also lower triangular, and the product of lower triangular matrices is lower triangular, we can invert the operations:

$$A = (E_1^{-1} E_2^{-1} \dots E_k^{-1})U = LU.$$

Here, L is a lower triangular matrix with 1s on the diagonal (unit lower triangular), and U is the upper triangular row echelon form of A.

***Theorem* 5.4.1. LU Decomposition**. Let $A \in \mathbb{R}^{n\times n}$. If A can be reduced to row echelon form without row interchanges, then A admits a factorisation $A = LU$, where L is unit lower triangular and U is upper triangular.

Remark. The entries of L below the diagonal are precisely the multipliers used during elimination. Specifically, if we eliminate the entry A_{ij} by the operation $R_i \to R_i - \ell_{ij}R_j$, then the (i,j)-th entry of L is ℓ_{ij}.

Example 5.4.1. Calculating LU. Let $A = \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix}$.

1. Eliminate below $A_{11} = 2$:
   - $R_2 \to R_2 - 2R_1$ (multiplier $\ell_{21} = 2$).
   - $R_3 \to R_3 - (-1)R_1$ (multiplier $\ell_{31} = -1$).

   Matrix becomes: $\begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 5 \end{bmatrix}$.
2. Eliminate below $A_{22} = 1$:
   - $R_3 \to R_3 - 1R_2$ (multiplier $\ell_{32} = 1$).

   Matrix becomes $U = \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix}$.

3. Construct L from the multipliers:

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

One can verify that $LU = A$.

PLU Factorisation

If row interchanges (permutations) are required during elimination to avoid a zero pivot, we cannot write $A = LU$ directly. Instead, we perform the permutations first (or track them). This leads to the $A = PLU$ factorisation.

***Theorem* 5.4.2. PLU Decomposition**. For any square matrix A, there exists a permutation matrix P (a reordered identity matrix), a unit lower triangular matrix L, and an upper triangular matrix U such that:

$$A = PLU$$

or equivalently $P^T A = LU$. The matrix P accounts for the row swaps needed to proceed with elimination.

Example 5.4.2. PLU Example. Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$. Elimination fails immediately because the pivot is 0. Swap $R_1 \leftrightarrow R_2$: $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

$$PA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = U.$$

In this trivial case, $L = I$. Thus $A = P^T IU = PIU$ (since $P^T = P$ here).

Solving Systems with LU

The primary application of LU factorisation is solving systems $A\mathbf{x} = \mathbf{b}$. The system becomes $L(U\mathbf{x}) = \mathbf{b}$. We define $\mathbf{y} = U\mathbf{x}$ and solve in two steps:

1. **Forward Substitution:** Solve $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$. Since L is lower triangular, this is immediate (solve y_1, then y_2, etc.).
2. **Backward Substitution:** Solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$. Since U is upper triangular, this is the standard back-substitution step.

This method is computationally superior to computing A^{-1} because triangular systems are cheap to solve.

5.5 Applications

Matrix arithmetic appears across diverse fields. We highlight a few illustrative examples.

Stochastic Matrices and Markov Chains Processes that evolve probabilistically over time—such as population dynamics or web page rankings — are often modelled by *Markov chains*. The transitions are described by a matrix.

Definition 5.5.1. *Stochastic Matrix*. A matrix $P \in \mathbb{R}^{n \times n}$ is a *stochastic matrix* if all its entries are non-negative ($P_{ij} \geq 0$) and the sum of the entries in each column is 1.

If $\mathbf{x}_k$ is a probability vector representing the state of a system at step k (where components sum to 1), the state at the next step is $\mathbf{x}_{k+1} = P\mathbf{x}_k$. The long-term behaviour is governed by the *steady-state vector* $\mathbf{x}^*$, satisfying $P\mathbf{x}^* = \mathbf{x}^*$ (an eigenvector problem, which we will study later).

Example 5.5.1. Disease Propagation. Consider a population where individuals are either Healthy (H) or Infected (I).

- An infected individual stays infected with probability 0.9, recovers with 0.1.
- A healthy individual stays healthy with probability 0.8, gets infected with 0.2.

The transition matrix is:

$$P = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}.$$

If initially $\mathbf{x}_0 = \begin{bmatrix} 100 \\ 900 \end{bmatrix}$ (100 infected), then $\mathbf{x}_1 = P\mathbf{x}_0 = \begin{bmatrix} 90 + 180 \\ 10 + 720 \end{bmatrix} = \begin{bmatrix} 270 \\ 730 \end{bmatrix}$.

Cryptography Matrices can perform linear transformations on data, serving as a basic form of encryption (Hill ciphers).

Encoding: Convert a message into a sequence of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. Choose an invertible matrix A. The encoded vectors are $\mathbf{c}_i = A\mathbf{v}_i$. **Decoding:** The receiver, possessing the key A^{-1}, reconstructs the original message via $\mathbf{v}_i = A^{-1}\mathbf{c}_i$.

Example 5.5.2. Simple Cipher. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$. Since $\det(A) = -1$, A is invertible. Message vector $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Encoded: $\mathbf{c} = A\mathbf{v} = \begin{bmatrix} 4 \\ 11 \end{bmatrix}$. Decoded: $\mathbf{v} = A^{-1}\mathbf{c} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 11 \end{bmatrix} = \begin{bmatrix} -20 + 22 \\ 12 - 11 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

5.6 The Symmetry of Row Equivalence

In Chapter 2, we defined two matrices to be *row-equivalent* ($A \sim B$) if B could be obtained from A via row operations. At the time, we claimed this relationship was reversible. With the tool of matrix inversion, we can now prove this rigorously.

***Theorem* 5.6.1. Symmetry of Row Equivalence**. *If $A \sim B$, then $B \sim A$. Furthermore, if $A \sim B$ and $B \sim C$, then $A \sim C$.*

Proof. If $A \sim B$, there exist elementary matrices $E_1, \dots, E_k$ such that $B = E_k \dots E_1 A$. Since every elementary matrix is invertible, we can multiply by their inverses in reverse order:

$$A = (E_1^{-1} \dots E_k^{-1})B.$$

Since the inverse of an elementary matrix is also an elementary matrix (as proven earlier), A is obtained from B by a sequence of row operations. Thus $B \sim A$. ■

Example 5.6.1. Reversibility of Reduction. Consider the matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$. We reduce A to the identity I via the sequence:

1. $R_3 \to R_3 - R_1$: $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}$.
2. $R_3 \to R_3 + R_2$: $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$.
3. $R_3 \to \frac{1}{2}R_3$: $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

4. $R_2 \to R_2 - R_3$: $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

5. $R_1 \to R_1 - R_2$: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$.

Because every step is reversible, we can reconstruct A from I by applying the inverse operations in reverse order: $R_1 \to R_1 + R_2$, $R_2 \to R_2 + R_3$, $R_3 \to 2R_3$, and so forth.

5.7 Exercises

1. **The Neumann Series.** A matrix N is called *nilpotent* if there exists a positive integer k such that $N^k = \mathbf{0}$.

   (a) Prove that if N is nilpotent, then the matrix $I - N$ is invertible.

   (b) Show that the inverse is given by the finite series:

   $$(I - N)^{-1} = I + N + N^2 + \cdots + N^{k-1}.$$

   (c) Use this result to compute the inverse of $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$ by writing $A = I - N$.

2. **The Trace and the Commutator.** The *trace* of a square matrix A, denoted $\text{tr}(A)$, is the sum of its diagonal entries: $\text{tr}(A) = \sum_{i=1}^{n} A_{ii}$.

   (a) Prove that for any $A, B \in \mathbb{R}^{n \times n}$, $\text{tr}(AB) = \text{tr}(BA)$.

   (b) The *commutator* of A and B is defined as $[A, B] = AB - BA$. Prove that it is impossible to find matrices A and B such that $[A, B] = I$.

   Remark. This result is significant in quantum mechanics, implying that the position and momentum operators cannot be represented by finite-dimensional matrices.

   (c) Is it possible to find A and B such that $AB - BA$ is diagonal with non-zero entries?

3. **The Isomorphism of Complex Numbers.** Let $\mathcal{C}$ be the set of 2×2 matrices of the form $M(a, b) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ where $a, b \in \mathbb{R}$.

   (a) Prove that this set is closed under addition and multiplication. That is, the sum and product of matrices in $\mathcal{C}$ are also in $\mathcal{C}$.

   (b) Show that if $M(a, b) \neq \mathbf{0}$, then $M(a, b)$ is invertible and its inverse lies in $\mathcal{C}$.

   (c) Establish a correspondence between $M(a, b)$ and the complex number $z = a + bi$. Show that matrix multiplication corresponds exactly to complex number multiplication.

4. **Symmetric and Skew-Symmetric Forms.** Let A be an $n \times n$ matrix.

   (a) Prove that the matrix $B = A^T A$ is always symmetric.

   (b) Prove that if A is skew-symmetric ($A^T = -A$) and n is odd, then A is not invertible.

   Remark. Hint: Use the properties of the determinant, specifically $\det(A^T) = \det(A)$ and $\det(-A) = (-1)^n \det(A)$, assuming the reader is familiar with basic determinant properties, or prove it via characteristic equations if preferred.

   (c) Let S be a symmetric matrix and K be a skew-symmetric matrix. Prove that $\text{tr}(SK) = 0$.

5. The Sherman-Morrison Formula. In computational physics, one often needs to update the inverse of a matrix after a small modification. Let A be an invertible matrix, and let $\mathbf{u}, \mathbf{v}$ be column vectors. Prove that if $1 + \mathbf{v}^T A^{-1}\mathbf{u} \neq 0$, then the matrix $A + \mathbf{u}\mathbf{v}^T$ is invertible and its inverse is given by:

$$(A + \mathbf{u}\mathbf{v}^T)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}^T A^{-1}}{1 + \mathbf{v}^T A^{-1}\mathbf{u}}.$$

Remark. Verify the result by multiplying $(A + \mathbf{u}\mathbf{v}^T)$ by the proposed inverse to obtain I.

6. Block Matrix Inversion (The Schur Complement). Consider a partitioned matrix $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ where A is invertible.

(a) Verify the following factorisation of M:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix},$$

where $S = D - CA^{-1}B$ is called the *Schur complement* of A in M.

(b) Prove that M is invertible if and only if S is invertible.

(c) Derive the formula for M^{-1} assuming S is invertible.

7. Uniqueness of the LDU Decomposition. We established the existence of the LU decomposition. A more symmetric form involves a diagonal matrix.

(a) Prove that if a square matrix A admits an LU factorisation where L is unit lower triangular and U is upper triangular with non-zero diagonal entries, then A can be written uniquely as $A = LDU'$, where:
- L is unit lower triangular,
- D is a diagonal matrix,
- U' is unit upper triangular (1s on the diagonal).

(b) Use this to prove that if A is a symmetric matrix that allows LDU' factorisation, then $L = (U')^T$, and thus $A = LDL^T$.

8. Stochastic Matrices. Let P and Q be $n \times n$ stochastic matrices (matrices with non-negative entries where each column sums to 1).

(a) Prove that the product PQ is also a stochastic matrix.

(b) Let $\mathbf{e} = [1, 1, \ldots, 1]^T$. Show that $P^T\mathbf{e} = \mathbf{e}$.

(c) Use part (b) to show that $\lambda = 1$ is always an eigenvalue of P (i.e., $P - I$ is singular).

9. Orthogonal Matrices. A square matrix Q is called *orthogonal* if $Q^TQ = I$.

(a) Show that if Q is orthogonal, then Q is invertible and $Q^{-1} = Q^T$.

(b) Prove that orthogonal matrices preserve the dot product: for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $(Q\mathbf{x}) \cdot (Q\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$.

(c) Conclude that orthogonal matrices preserve lengths (magnitudes) of vectors and angles between them.

10. Polynomial Interpolation via Inversion. We wish to find the unique polynomial $f(t) = c_0 + c_1t + c_2t^2$ passing through three points $(t_1, y_1), (t_2, y_2), (t_3, y_3)$ with distinct t_i.

(a) Set up the system $V\mathbf{c} = \mathbf{y}$ where V is the Vandermonde matrix.

(b) For the case $t_1 = 0, t_2 = 1, t_3 = 2$, compute V^{-1} using the Gauss-Jordan algorithm.

(c) Use V^{-1} to find the polynomial passing through $(0, 1), (1, 0), (2, 4)$.