0 Preliminaries

In this chapter we will review a few preliminary notions that are necessary for our formal treatment of linear algebra in subsequent chapters. We will briefly discuss properties of vectors in $\mathbb{R}^n$, matrices, complex numbers, fields, polynomials, and mathematical induction. Some of the material (on vectors, matrices, and polynomials) will be used as motivation for our axiomatic development of vector spaces, while other material (on fields and induction) will occasionally be needed during our later discussions.

0.1 Sets, Numbers, and Functions

- We very briefly recall a few basic notions about sets, numbers, and functions.

- Definition: A set is a well-defined collection of distinct elements.
  
  - The elements of a set can be essentially anything: integers, real numbers, other sets, people.
  - Sets are generally denoted by capital or script letters, and when listing the elements of a set, curly brackets \{\} are used.
  - Sets do not have to have any elements: the empty set $\emptyset = \{\}$ is the set with no elements at all.
  - Two sets are the same precisely if all of their elements are the same. The elements in a set are not ordered, and no element can appear in a set more than once: thus the sets \{1, 4\} and \{4, 1\} are the same.
  - We write $x \in S$ when $x$ is an element of the set $S$, and $x \notin S$ when $x$ is not an element of the set $S$.
  - Example: For $S = \{1, 2, 5\}$ we have $1 \in S$ and $5 \in S$ but $3 \notin S$ and $\pi \notin S$.

- There are two primary ways to describe a set.
  
  - One way is to list all the elements: for example, $A = \{1, 2, 4, 5\}$ is the set containing the four numbers $1, 2, 4,$ and $5$.
  - The other way to define a set is to describe properties of its elements:\footnote{It is possible to run into trouble by trying to define sets in this “naïve” way of specifying qualities of their elements. In general, one must be more careful when defining arbitrary sets, although we will not worry about this.}: for example, the set $S$ of one-letter words in the English alphabet has two elements: $S = \{a, I\}$.

- We often employ “set-builder” notation for sets: for example, the set $S$ of real numbers between 0 and 5 is denoted $S = \{x : x$ is a real number and $0 \leq x \leq 5\}$.
• Some authors use a vertical pipe | instead of a colon : but this distinction is irrelevant.

• **Definition:** If $A$ and $B$ are two sets with the property that every element of $A$ is also an element of $B$, we say $A$ is a **subset** of $B$ (or that $A$ is contained in $B$) and write $A \subseteq B$. \(^2\)

  ○ **Example:** If $A = \{1, 2, 3\}$, $B = \{1, 4, 5\}$, and $C = \{1, 2, 3, 4, 5\}$, then $A \subseteq C$ and $B \subseteq C$ but neither $A$ nor $B$ is a subset of the other.

• **Definition:** If $A$ and $B$ are two sets, then the intersection $A \cap B$ is the set of all elements contained in both $A$ and $B$. The union $A \cup B$ is the set of all elements contained in either $A$ or $B$ (or both).

  ○ **Example:** If $A = \{1, 2, 3\}$ and $B = \{1, 4, 5\}$, then $A \cap B = \{1\}$ and $A \cup B = \{1, 2, 3, 4, 5\}$.

• Certain sets of numbers are extremely important, and arise often.

  ○ **Definition:** The **positive integers** $(1, 2, 3, \ldots)$ are obtained by adding 1 to itself some number of times; the other integers are 0 and the negatives of the positive integers $(-1, -2, -3, \ldots)$. Integers can be added, subtracted, multiplied, but only sometimes divided. The set of positive integers is denoted $\mathbb{N}$ (“naturals”) and the set of all integers is denoted $\mathbb{Z}$ (“Zahlen”, German for “numbers”).

  ○ Examples of integers: $3, 0, -666, 1337, 10^{10^{10}}$.

  ○ **Definition:** The **rational numbers** are numbers of the form $a/b$ where $a$ and $b$ are integers and $b$ is not zero. Rational numbers can be added, subtracted, multiplied, and always divided (except by 0). The set of rational numbers is denoted $\mathbb{Q}$ (“quotients”).

    ○ Examples of rational numbers: $\frac{1}{2}, -\frac{225}{1037}, 11, 0$.

    ○ The basic operations are $\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$ and $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$.

  ○ **Definition:** The **real numbers** are harder to define simply, but roughly speaking, the real numbers are obtained by filling in the “gaps” between the rational numbers. A common way to think of real numbers is as (infinte) decimal sequences.\(^3\) The set of real numbers is denoted $\mathbb{R}$ (“reals”). Real numbers can be added, subtracted, multiplied, and always divided (except by 0).

    ○ Examples of real numbers: $\pi, \sqrt{2}, \frac{\pi + \sqrt{2}}{5}, 11, 0, 0.12131415161718192021\ldots$.

  • We can also speak of functions in this general context.

  ○ **Definition:** A **function** is a relation between a set of inputs (called the **domain** of the function) and a set of outputs (called the **range** of the function): to each element of the domain, the function associates a single value in the range. Two functions are the same if they have the same domain and associate the same value in the range to each element of their common domain.

    ○ **Example:** Consider $f(x) = x^3$, with domain and range both the set of real numbers. This function $f$ sends each real number $x$ to its cube $x^3$; thus $f(2) = 8, f(0) = 0$, and $f(-1) = -1$.

    ○ If the function is $f$, we write $f : S \to T$ to indicate that $S$ is the domain of $f$ and that the range of $f$ is contained in $T$.\(^4\)

    ○ **Example:** Consider the function $g : \{1, 2\} \to \{1, 2, 3\}$ with $g(1) = 2$ and $g(2) = 3$. The range of $g$ is the set $\{2, 3\}$.

\(^2\)Subset notation is not universally agreed-upon: the notation $A \subset B$ is also commonly used to say that $A$ is a subset of $B$. The difference is not terribly relevant except for when $A$ can be equal to $B$: some authors intend $A \subset B$ to include the possibility that $A$ could be equal to $B$, while others insist that $A \subset B$ means that $A$ is a subset of $B$ which cannot be all of $B$.

\(^3\)Some real numbers have two decimal sequences: $1.000\ldots = 0.999\ldots$ are two different ways of writing the positive integer 1. A similar ambiguity occurs with any other decimal number ending in an infinite string of 9’s, but these are the only real numbers with two decimal expansions.

\(^4\)Technically, the set $T$ in the notation $f : S \to T$ is called the codomain (or “target”) of $f$, since the range of $f$ is only the set of values of the form $f(x)$ for some $x$ in the domain of $f$. For $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$, the range of $f$ is only the set $[0, \infty)$ and not all of $\mathbb{R}$.
○ In general, unless specified, the domain of a function is the largest possible set for which the definition of the function makes sense.

○ The notation \(f(g(x))\) symbolizes the result of applying \(f\) to the value \(g(x)\): this is called function composition, and is well-defined provided that the range of \(g\) is a subset of the domain of \(f\). We use the notation \(f \circ g\) to refer to the composite function itself, so that \((f \circ g)(x) = f(g(x))\).

○ Example: For \(g : \{1, 2\} \rightarrow \{2, 3\}\) having \(g(1) = 3\) and \(g(2) = 2\), and \(f : \{2, 3\} \rightarrow \{3, 4\}\) having \(f(2) = 3\) and \(f(3) = 4\), the composition \(f \circ g : \{1, 2\} \rightarrow \{3, 4\}\) has \((f \circ g)(1) = f(g(1)) = f(3) = 4\) and \((f \circ g)(2) = f(g(2)) = f(2) = 3\).

○ One can also compute other compositions, like \(g \circ f\) or \(f \circ f\). In general, it will be the case that \(f \circ g\) and \(g \circ f\) are completely unrelated functions! (In the example above, \(g \circ f\) is not even defined.)

○ Definition: A function \(f\) is one-to-one (or injective) if \(f(a) = f(b)\) implies \(a = b\), or equivalently, if \(a \neq b\) then \(f(a) \neq f(b)\). In other words, \(f\) is one-to-one if unequal elements in the domain are sent to unequal elements in the range.

○ For functions whose domain and range are (subsets of) the real numbers, one-to-one functions satisfy the “horizontal line test”: a horizontal line can intersect the graph of the function at most once.

○ Definition: A one-to-one function \(f(x)\) has an inverse function \(f^{-1}(x)\) defined so that \(f^{-1}(f(x)) = x\) for every \(x\) in the domain of \(f\), and \(f(f^{-1}(y)) = y\) for every \(y\) in the range of \(f\).

○ To compute the inverse function of \(f\), simply solve the equation \(y = f(x)\) for \(y\) in terms of \(x\): this will give \(x = f^{-1}(y)\).

0.2 Vectors in \(\mathbb{R}^n\)

○ A vector, as we typically think of it, is a quantity which has both a magnitude and a direction. This is in contrast to a scalar, which carries only a magnitude.

○ Real-valued vectors are extremely useful in just about every aspect of the physical sciences, since just about everything in Newtonian physics is a vector: position, velocity, acceleration, forces, etc. There is also “vector calculus” (namely, calculus in the context of vector fields) which is typically part of a multivariable calculus course; it has many applications to physics as well.

○ We often think of vectors geometrically, as a directed line segment (having a starting point and an endpoint).

○ Algebraically, we write a vector as an ordered tuple of coordinates: we denote the \(n\)-dimensional vector from the origin to the point \((a_1, a_2, \ldots, a_n)\) as \(v = (a_1, a_2, \ldots, a_n)\), where the \(a_i\) are real-number scalars.

○ Some vectors: \((1, 2), (3, 5, -1), \left(\frac{\pi}{2}, 273, \frac{4}{\pi}, 0, 0, -1\right)\).

○ Notation: We use angle brackets \((\cdot)\) rather than parentheses \((\cdot)\) so as to draw a visual distinction between a vector and the coordinates of a point in space. We also draw arrows above vectors (as \(\vec{v}\)) or typeset them in boldface (as \(v\)) in order to set them apart from scalars. Boldface is hard to produce without a computer, so it is highly recommended to use the arrow notation \(\vec{v}\) when writing by hand.

○ We often think of, and draw, vectors as directed line segments.

○ However, technically speaking, vectors are slightly different from directed segments, because we don’t care where a vector starts: we only care about the difference between the starting and ending positions. Thus: the directed segment whose start is \(0, 0\) and end is \((1, 1)\) and the segment starting at \((1, 1)\) and ending at \((2, 2)\) represent the same vector \((1, 1)\).

○ We can add vectors (provided they are of the same length!) in the obvious way, one component at a time: if \(v = (a_1, \ldots, a_n)\) and \(w = (b_1, \ldots, b_n)\) then \(v + w = (a_1 + b_1, \ldots, a_n + b_n)\).
○ We can justify this using our geometric idea of what a vector does: \( \mathbf{v} \) moves us from the origin to the point \( (a_1, \ldots, a_n) \). Then \( \mathbf{w} \) tells us to add \( (b_1, \ldots, b_n) \) to the coordinates of our current position, and so \( \mathbf{w} \) moves us from \( (a_1, \ldots, a_n) \) to \( (a_1+b_1, \ldots, a_n+b_n) \). So the net result is that the sum vector \( \mathbf{v} + \mathbf{w} \) moves us from the origin to \( (a_1+b_1, \ldots, a_n+b_n) \), meaning that it is just the vector \( (a_1+b_1, \ldots, a_n+b_n) \).

○ Geometrically, we can think of vector addition using a parallelogram whose pairs of parallel sides are \( \mathbf{v} \) and \( \mathbf{w} \) and whose diagonal is \( \mathbf{v} + \mathbf{w} \):

![](image)

\[ \begin{align*}
\mathbf{w} & = \begin{bmatrix} 1 & 2 & 2 \end{bmatrix} \\
\mathbf{v} & = \begin{bmatrix} -1 & 2 & 2 \end{bmatrix} \\
\mathbf{v} + \mathbf{w} & = \begin{bmatrix} 0 & 0 & -2 \end{bmatrix} \\
\mathbf{v} - 2\mathbf{w} & = \begin{bmatrix} -7 & 2 & 10 \end{bmatrix}
\end{align*} \]

The arithmetic of vectors in \( \mathbb{R}^n \) satisfies several algebraic properties that follow more or less directly from the definition:

○ Addition of vectors is commutative and associative.

○ There is a zero vector (namely, the vector with all entries zero), and every vector has an additive inverse.

○ Scalar multiplication distributes over addition of both vectors and scalars.

With vectors in \( \mathbb{R}^n \), we also have a quantity that resembles a product, the dot product:

**Definition:** The dot product of two vectors \( \mathbf{v}_1 = (a_1, \ldots, a_n) \) and \( \mathbf{v}_2 = (b_1, \ldots, b_n) \) is defined to be the scalar \( \mathbf{v}_1 \cdot \mathbf{v}_2 = a_1b_1 + a_2b_2 + \cdots + a_nb_n \).

○ Note: The dot product of two vectors is a scalar, not a vector! (For this reason, the dot product is sometimes called the “scalar product” of two vectors.)

○ Example: The dot product \((1, 2) \cdot (3, 4)\) is \((1)(2) + (3)(4) = 14\).

○ Example: The dot product \((-1, 2, 2) \cdot (3, 0, -4)\) is \((-1)(3) + (2)(0) + (2)(-4) = -11\).

The dot product obeys several very nice properties reminiscent of standard multiplication. For any vectors \( \mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \), and any scalar \( r \), we can verify the following properties directly from the definition:

○ The dot product is commutative: \( \mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v} \).

○ The dot product distributes over addition: \( (\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = (\mathbf{v}_1 \cdot \mathbf{w}) + (\mathbf{v}_2 \cdot \mathbf{w}) \).

○ Scalars can be factored out of a dot product: \( (r \mathbf{v}) \cdot \mathbf{w} = r (\mathbf{v} \cdot \mathbf{w}) \).

One of the chief motivations of the dot product is to provide a way to measure angles between vectors. First, we must define the length of a vector:

**Definition:** The norm (length, magnitude) of the vector \( \mathbf{v} = (a_1, \ldots, a_n) \) is \( \|\mathbf{v}\| = \sqrt{(a_1)^2 + \cdots + (a_n)^2} \).
○ This is just an application of the distance formula: the norm of the vector \( \langle a_1, \ldots, a_n \rangle \) is just the length of the line segment joining the origin \((0, \ldots, 0)\) to the point \((a_1, \ldots, a_n)\).

○ Example: For \( v = \langle -1, 2, 2 \rangle \) and \( w = \langle 3, 0, -4, 5 \rangle \), we have \( ||v|| = \sqrt{(-1)^2 + 2^2 + 2^2} = \sqrt{3} \) and \( ||w|| = \sqrt{3^2 + 0^2 + (-4)^2 + 5^2} = \sqrt{50} \).

○ If \( r \) is a scalar, we can see immediately from the definition that \( ||rv|| = |r||v|| \), since we can just factor out a \( \sqrt{r^2} = |r| \) from each term under the square root.

○ Observe also that the dot product of a vector with itself is the square of the norm: \( v \cdot v = ||v||^2 \).

- Using the dot product, we can give a formula for the angle between two vectors:

- **Theorem** (Dot Product): For vectors \( v_1 \) and \( v_2 \) forming an angle \( \theta \) between them, \( v_1 \cdot v_2 = ||v_1|| \cdot ||v_2|| \cdot \cos(\theta) \).

- Using the Dot Product Theorem, we can compute the angle between two vectors.

- **Proof:** Consider the triangle formed by \( v_1, v_2, \) and \( v_2 - v_1 \): applying the Law of Cosines in this triangle yields

\[
||v_2 - v_1||^2 = ||v_1||^2 + ||v_2||^2 - 2 ||v_1|| \cdot ||v_2|| \cdot \cos(\theta).
\]

Since the square of the norm is the dot product of a vector with itself, and the dot product is distributive, we can write

\[
||v_2 - v_1||^2 = (v_2 - v_1) \cdot (v_2 - v_1) = (v_2 \cdot v_2) - (v_1 \cdot v_2) - (v_2 \cdot v_1) + (v_1 \cdot v_1) = ||v_2||^2 - 2(v_1 \cdot v_2) + ||v_1||^2.
\]

Now by comparing to the Law of Cosines expression, we must have \( ||v_1|| \cdot ||v_2|| \cdot \cos(\theta) = v_1 \cdot v_2 \), as claimed.

- **Example:** Compute the angle between the vectors \( v = \langle 2\sqrt{2}, 1, \sqrt{3} \rangle \) and \( w = \langle 0, \sqrt{3}, 1 \rangle \).

- We compute \( v \cdot w = (2\sqrt{2})(0) + (1)(\sqrt{3}) + (\sqrt{3})(1) = 2\sqrt{3}, \) and \( ||v|| = \sqrt{(2\sqrt{2})^2 + 1^2 + (\sqrt{3})^2} = 2\sqrt{3} \) and \( ||w|| = \sqrt{\sqrt{3})^2 + 0^2 + 1^2} = 2 \).

- Then by the Dot Product Theorem, the angle \( \theta \) between the vectors satisfies \( 2\sqrt{3} = 2 \cdot 2\sqrt{3} \cdot \cos(\theta) \), meaning that \( \theta = \cos^{-1} \left( \frac{1}{2} \right) = \frac{\pi}{3} \).

- **Example:** Compute the angle between the vectors \( v = \langle 0, 2, 1, 2 \rangle \) and \( w = \langle 3, 4, 0, -12 \rangle \).

- We compute \( v \cdot w = 0 + 8 + 0 - 24 = -16, ||v|| = 3, \) and \( ||w|| = 13 \).

- Then by the Dot Product Theorem, the angle \( \theta \) between the vectors satisfies \( -16 = 3 \cdot 13 \cdot \cos(\theta) \), so

\[
\theta = \cos^{-1} \left( -\frac{16}{39} \right)
\]

### 0.3 Complex Numbers, Fields

- **Definitions:** A complex number is a number of the form \( a + bi \), where \( a \) and \( b \) are real numbers and \( i \) is the "imaginary unit", defined so that \( i^2 = -1 \). The set of all complex numbers is denoted \( \mathbb{C} \) ("complex").

- **Notation:** Sometimes, \( i \) is written as \( \sqrt{-1} \). In certain areas (especially electrical engineering), the letter \( j \) can be used to denote \( \sqrt{-1} \), rather than \( i \) (which is used to denote electrical current).

- The real part of \( z = a + bi \), denoted \( \text{Re}(z) \), is the real number \( a \).

- The imaginary part of \( z = a + bi \), denoted \( \text{Im}(z) \), is the real number \( b \).

- The complex conjugate of \( z = a + bi \), denoted\(^5\) as \( \bar{z} \), is the complex number \( a - bi \).

\(^5\)The notation for the complex conjugate varies among disciplines. In physics and computer science, the notation \( z^* \) often denotes the complex conjugate (among other reasons, because it is easier to typeset).
The modulus (also called the absolute value, magnitude, or length) of \( z = a + bi \), denoted \(|z|\), is the real number \( \sqrt{a^2 + b^2} \). Observe that \(|z|^2 = a^2 + b^2 = z \overline{z} \).

**Example:** Re\((4 - 3i)\) = 4, Im\((4 - 3i)\) = -3, \(4 - 3i = 4 + 3i, |4 - 3i| = 5.\)

- Two complex numbers are added (or subtracted) simply by adding (or subtracting) their real and imaginary parts: \((a + bi) + (c + di) = (a + c) + (b + d)i\).

  **Example:** The sum of \(1 + 2i\) and \(3 - 4i\) is \((1 + 2i) + (3 - 4i) = 4 - 2i\). The difference is \((1 + 2i) - (3 - 4i) = -2 + 6i\).

- Two complex numbers are multiplied using the distributive law and the fact that \(i^2 = -1\): \((a + bi)(c + di) = ac + adi + bci + bdi^2 = (ac - bd) + (ad + bc)i\).

  **Example:** The product of \(1 + 2i\) and \(3 - 4i\) is \((1 + 2i)(3 - 4i) = 3 + 6i - 4i - 8i^2 = 11 + 2i\).

For division, we rationalize the denominator: \(\frac{a + bi}{c + di} = \frac{(a + bi)(c - di)}{(c + di)(c - di)} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i\).

  **Example:** The quotient of \(2i\) by \(1 - i\) is \(\frac{2i}{1 - i} = \frac{2i(1 + i)}{(1 - i)(1 + i)} = \frac{-2 + 2i}{2} = -1 + i\).

A key property of the conjugate is that it is multiplicative: if \(z = a + bi\) and \(w = c + di\), then \(\overline{zw} = \overline{z} \cdot \overline{w}\). (This is easy to see just by multiplying out the relevant quantities.) From this we see that the modulus is also multiplicative: \(|zw| = |z| \cdot |w|\).

  **Example:** If \(z = 1 + 2i\) and \(w = 3 - i\), then \(\overline{z} = 1 - 2i\) and \(\overline{w} = 3 + i\). We compute \(zw = 5 + 5i\) and \(\overline{zw} = 5 - 5i\), so indeed \(\overline{zw} = zw\). Furthermore, we have \(|z| = \sqrt{5}, |w| = \sqrt{10}, \) and \(|zw| = \sqrt{50} = |z| \cdot |w|\).

  **Example:** This is the underlying reason for why division works in general: we write \(\frac{z}{w} = \frac{z \cdot \overline{w}}{w \cdot \overline{w}} = \frac{z \cdot \overline{w}}{|w|^2}\), where the denominator is now the real number \(|w|^2 = c^2 + d^2\).

For reference, here are a few more properties of complex numbers (each of which can be verified by writing everything out in terms of real and imaginary parts):  

- For any \(z\) and \(w\), \(\overline{z + w} = \overline{z} + \overline{w}\), \(\overline{zw} = \overline{z} \cdot \overline{w}\), and \(\overline{\overline{z}} = z\).
- For any \(z\), \(z = \overline{\overline{z}}\) if and only if \(z\) is real, and \(z = -\overline{z}\) if and only if \(z\) is purely imaginary (of the form \(ri\) where \(r\) is real).
- For any \(z\) and \(w\), \(\text{Re}(z) \leq |z|, \text{Im}(z) \leq |z|, |zw| = |z| \cdot |w|, \) and \(|z + w| \leq |z| + |w|\). (The latter follows by observing \(|z + w|^2 = (z + w)(\overline{z + w}) = |z|^2 + |w|^2 + 2\text{Re}(zw) \leq |z|^2 + |w|^2 + 2|zw| = (|z| + |w|)^2\).)

The real numbers and the complex numbers are both examples of fields: sets of numbers that can be added, subtracted, multiplied, and divided (except by zero) and possess various algebraic relations involving these operations. More formally:

**Definition:** A field is an ordered triple \((F, +, \cdot)\) consisting of a set of numbers \(F\) together with two binary operations\(^6\), \(+\) (addition) and \(\cdot\) (multiplication), satisfying the following axioms:

[A1] Addition is commutative: \(a + b = b + a\) for any \(a\) and \(b\) in \(F\).

[A2] Addition is associative: \((a + b) + c = a + (b + c)\) for any \(a, b,\) and \(c\) in \(F\).

[A3] There exists a zero element, 0, with \(0 + a = a = a + 0\) for any \(a\) in \(F\).

[A4] Every element \(a\) has an additive inverse \(-a\), such that \(a + (-a) = 0 = (-a) + a\).

[M1] Multiplication is commutative: \(a \cdot b = b \cdot a\) for any \(a\) and \(b\) in \(F\).

[M2] Multiplication is associative: \((a \cdot b) \cdot c = a \cdot (b \cdot c)\) for any \(a, b,\) and \(c\) in \(F\).

[M3] There exists a multiplicative identity, 1, with \(1 \cdot a = a = a \cdot 1\) for any \(a\) in \(F\).

\(^6\) The result of applying these operations to elements \(a\) and \(b\) is denoted by \(a + b\) and \(a \cdot b\) (or simply \(ab\)), respectively. The definition of “binary operation” means that \(a + b\) and \(a \cdot b\) are also numbers in \(F\).
[M4] Every nonzero element \( a \) has a multiplicative inverse \( a^{-1} \), with \( a \cdot a^{-1} = 1 = a^{-1} \cdot a \).

[D] Multiplication distributes over addition: \( a \cdot (b + c) = a \cdot b + a \cdot c \) for any \( a, b, \) and \( c \) in \( F \).

- Subtraction and division are not explicitly part of the axioms of a field. They are instead obtained in terms of addition and multiplication, per the axioms.
  - Explicitly, in a field we define \( a - b = a + (-b) \) and \( a/b = a \cdot (b^{-1}) \), the latter whenever \( b \neq 0 \).

- It will not be necessary for us to work with this “axiomatic” description of a field: we will only need to use these properties in the context of vector spaces.

- We are primarily interested in the fields of real numbers and complex numbers, but for completeness, here are a few more examples of fields:
  - The set of rational numbers \( \mathbb{Q} \) is a field.
  - The set \( \mathbb{Q}(\sqrt{2}) \) of real numbers of the form \( x + y\sqrt{2} \), where \( x \) and \( y \) are rational numbers, is a field.
  - The set \( \{0, 1\} \) forms a field, under the operations \( 0+0 = 1+1 = 0, 1+0 = 0+1 = 1, 0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0, 1 \cdot 1 = 1 \). (It is cumbersome to verify all of the properties, but they do all hold.) Unlike the other examples above, in which the field has infinitely many elements, this field only has two elements. It is denoted as \( F_2 \) or \( \mathbb{Z}/2\mathbb{Z} \).

### 0.4 Matrices and Systems of Linear Equations

- Let \( F \) be a field.

- **Definition:** An \( m \times n \) matrix is an array of numbers (from \( F \)) with \( m \) rows and \( n \) columns. A square matrix is one with the same number of rows and columns: that is, an \( n \times n \) matrix for some \( n \). The set of all \( m \times n \) matrices with elements in \( F \) is denoted \( M_{m \times n}(F) \).

  - **Examples:** \[
  \begin{bmatrix}
  4 & 1 & -1 \\
  3 & 2 & 0
  \end{bmatrix}
  \] is a \( 2 \times 3 \) real matrix, and \[
  \begin{bmatrix}
  \pi & 0 & 0 \\
  0 & 4i & 0 \\
  0 & 0 & 6
  \end{bmatrix}
  \] is a \( 3 \times 3 \) complex matrix.

- **Definition:** If \( A \) is a matrix, the entry in the \( i \)th row and \( j \)th column of \( A \) is called the \( (i, j) \)-entry of \( A \), and will be denoted \( a_{i,j} \).

  - **Warning:** It is easy to mix up the coordinates. Remember that the first coordinate specifies the row, and the second coordinate specifies the column.

  - **Example:** If \( A = \begin{bmatrix} 2 & -1 & 4 \\ 3 & 0 & 5 \end{bmatrix} \), then the \( (2,2) \)-entry is \( a_{2,2} = 0 \) and the \( (1,3) \)-entry is \( a_{1,3} = 4 \).

- **Definition:** In an \( n \times n \) square matrix, the \( (i, i) \)-entries for \( 1 \leq i \leq n \) form the diagonal of the matrix. The trace of a square matrix is the sum of its diagonal entries.

  - **Example:** The diagonal entries of \[
  \begin{bmatrix}
  4 & 1 \\
  3 & 2
  \end{bmatrix}
  \] are 4 and 2, so the trace of this matrix is 6.

### 0.4.1 Matrix Arithmetic

- Like with vectors, we can add and subtract matrices of the same dimension, and we can also multiply a matrix by a scalar. Each of these operations is done “componentwise”: to add or subtract, we just add or subtract the corresponding entries of the two matrices. To multiply by a scalar, we just multiply each entry by that scalar.

  - **Example:** If \( A = \begin{bmatrix} 1 & 6 \\ 2 & 2 \end{bmatrix} \) and \( B = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \), then \( A + B = \begin{bmatrix} 1+3 & 6+0 \\ 2+0 & 2+2 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 2 & 4 \end{bmatrix} \). \( 2A = \begin{bmatrix} 2 \cdot 1 & 2 \cdot 6 \\ 2 \cdot 2 & 2 \cdot 2 \end{bmatrix} = \begin{bmatrix} 2 & 12 \\ 4 & 4 \end{bmatrix} \), and \( A - \frac{1}{3}B = \begin{bmatrix} 0 & 6 \\ 2 & \frac{4}{3} \end{bmatrix} \).
• We also have a transposition operation, where we interchange the rows and columns of the matrix:

**Definition (Matrix Transpose):** If \( A \) is an \( n \times m \) matrix, then the **transpose** of \( A \), denoted \( A^T \), is the \( m \times n \) matrix whose \((i,j)\)-entry is equal to the \((j,i)\)-entry of \( A \).

- **Example:** If \( A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \), then \( A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \).

- A symmetric matrix is one with \( A = A^T \), and a skew-symmetric matrix is one with \( A = -A^T \). (For either of these properties to hold, \( A \) must be a square matrix.)

• We can also multiply matrices in certain situations. However, matrix multiplication, unlike matrix addition, is NOT performed componentwise.

- The definition of matrix multiplication seems strange and arbitrary at first. However, there is a very good reason for this peculiar definition, and we will explain it when we discuss linear transformations.

**Definition (Matrix Product):** If \( A \) is an \( m \times n \) matrix and \( B \) is an \( n \times q \) matrix, then the matrix **product** \( A \cdot B \), often written simply as \( AB \), is the \( m \times q \) matrix whose \((i,j)\)-entry is the sum \( (AB)_{i,j} = \sum_{k=1}^{n} a_{i,k}b_{k,j} \), the sum of products of corresponding entries from the \( i \)th row of \( A \) with the \( j \)th column of \( B \).

- **Important Note:** In order for the matrix product to exist, the number of columns of \( A \) must equal the number of rows of \( B \). In particular, if \( A \) and \( B \) are the same size, their product exists only if they are square matrices. Also, if \( AB \) exists, then \( BA \) may not necessarily exist.

- More compactly, the \((i,j)\) entry of the matrix product \( AB \) is the dot product of the \( i \)th row of \( A \) with the \( j \)th column of \( B \) (each thought of as vectors). These two vectors must have the same length for the dot product to exist, which is why the length of each row of \( A \) (the number of columns of \( A \)) must equal the length of each column of \( B \) (the number of rows of \( B \)).

- This product is sometimes called the **row-column product** to emphasize the fact that it is a product involving the rows of \( A \) with the columns of \( B \).

**Example:** If \( A = \begin{bmatrix} -1 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix} \) and \( B = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ 3 & 3 \end{bmatrix} \), find \( AB \) and \( BA \).

- Since \( A \) is a \( 2 \times 3 \) matrix and \( B \) is a \( 3 \times 2 \) matrix, we see that \( AB \) is defined and will be a \( 2 \times 2 \) matrix.
- The \((1,1)\) entry of \( AB \) is \((-1)(1) + (1)(2) + (2)(3) = 7\).
- The \((1,2)\) entry of \( AB \) is \((-1)(2) + (1)(0) + (2)(3) = 4\).
- The \((2,1)\) entry of \( AB \) is \((0)(1) + (1)(2) + (1)(3) = 5\).
- The \((2,2)\) entry of \( AB \) is \((0)(2) + (1)(0) + (1)(3) = 3\).

- Putting all of this together gives \( AB = \begin{bmatrix} 7 & 4 \\ 5 & 3 \end{bmatrix} \).

- We see that \( BA = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} -1 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix} \) is also defined and will be a \( 3 \times 3 \) matrix.
- The \((1,1)\) entry of \( BA \) is \((1)(-1) + (2)(0) = -1\), the \((1,2)\) entry is \((1)(1) + (2)(1) = 3\), and the \((1,3)\) entry is \((1)(2) + (2)(1) = 4\).

- In a similar way we can compute the other six entries: the result is \( BA = \begin{bmatrix} -1 & 3 & 4 \\ -2 & 2 & 4 \\ -3 & 6 & 9 \end{bmatrix} \).

• If we restrict our attention to square matrices, then matrices under addition and multiplication obey some, but not all, of the algebraic properties that real numbers do.
In general, matrix multiplication is NOT commutative: $AB$ typically isn’t equal to $BA$, even if $A$ and $B$ are both square matrices.

* **Example:** $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$. Then $AB = \begin{bmatrix} 10 & 13 \\ 22 & 29 \end{bmatrix}$ while $BA = \begin{bmatrix} 11 & 16 \\ 19 & 28 \end{bmatrix}$.

* Matrix multiplication distributes over addition, on both sides: $(A + B)C = AC + BC$ and $A(B + C) = AB + AC$.

  * This property can be derived from the definition of matrix multiplication, along with some arithmetic.

  * Matrix multiplication is associative: $(AB)C = A(BC)$, if $A, B, C$ are of the proper dimensions.

  * In particular, taking the $n$th power of a square matrix is well-defined for every positive integer $n$.

  * This property can also be derived from the definition, but the arithmetic is very cumbersome.

  * The transpose of the product of two matrices is the product of their transposes in reverse order: $(AB)^T = B^TA^T$.

  * This property can likewise be derived from the definition of matrix multiplication.

  * The trace distributes over addition: $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$. Also, $\text{tr}(cA) = c\text{tr}(A)$ for any constant $c$.

**Definition:** If $A$ is an $n \times n$ matrix, then there is a zero matrix $Z_n$ which has the properties $Z_n + A = A$ and $Z_nA = AZ_n = Z_n$. This matrix $Z_n$ is the matrix whose entries are all zeroes.

  * **Example:** The $2 \times 2$ zero matrix is $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

  * **Remark:** In contrast to real (or complex) numbers, where $x^2 = 0$ implies $x = 0$, there exist nonzero matrices whose square is nonetheless the zero matrix. One such matrix is $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$: it is easy to check that $A^2$ is the zero matrix, but of course $A$ itself is nonzero.

**Definition:** If $A$ is an $n \times n$ matrix, then there is an $n \times n$ identity matrix $I_n$ which has the property that $I_nA = AI_n = A$. This matrix $I_n$ is the matrix whose diagonal entries are 1s and whose other entries are 0s.

  * **Example:** The $2 \times 2$ identity matrix is $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and the $3 \times 3$ identity matrix is $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

  * Observe that $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ for any $2 \times 2$ matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

### 0.4.2 Systems of Linear Equations

- We can use matrices to solve systems of linear equations, like the system $\begin{cases} x + \frac{y}{2x} = \frac{7}{2y} = -2 \end{cases}$. (The braces are only there to emphasize that the equations are to be considered together.)

- Recall that a linear equation is an equation of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, for some constants $a_1, \ldots, a_n$, $b$, and variables $x_1, \ldots, x_n$.

- When we seek to find the solutions to a system of linear equations, this means finding all possible values of the variables such that all equations are satisfied simultaneously.

- The traditional method for solving a system of linear equations (likely familiar from basic algebra) is by elimination: we solve the first equation for one variable (say, $x$) in terms of the others, and then plug in the result to all the other equations to obtain a reduced system involving one fewer variable. Eventually, the system will simplify either to a contradiction (e.g., $1 = 0$), a unique solution, or an infinite family of solutions.

- **Example:** Solve the system $\begin{cases} x + \frac{y}{2x} = \frac{7}{2y} = -2 \end{cases}$.
We can solve the first equation for \( x \) to obtain \( x = 7 - y \).

Plugging in this relation to the second equation gives \( 2(7 - y) - 2y = -2 \), or \( 14 - 4y = -2 \), so that \( y = 4 \). Then since \( x = 7 - y \) we obtain \( x = 3 \).

Another way to perform elimination is to add and subtract multiples of the equations, so to eliminate variables (and remove the need to solve for each individual variable before eliminating it).

In the example above, instead of solving the first equation for \( x \), we could multiply the first equation by \(-2\) and then add it to the second equation, so as to eliminate \( x \) from the second equation.

This yields the same overall result, but is less computationally difficult.

This procedure of elimination can be simplified even more, because we don’t really need to write the variable labels down every time. We only need to keep track of the coefficients, which we can do by putting them into a “coefficient matrix”.

For example, \( \begin{cases} x + y + 3z = 4 \\ 2x + 3y - z = 1 \\ -x + 2y + 2z = 1 \end{cases} \) has associated coefficient matrix \( \begin{bmatrix} 1 & 1 & 3 & 4 \\ 2 & 3 & -1 & 1 \\ -1 & 2 & 2 & 1 \end{bmatrix} \).

When working with a coefficient matrix, we will draw a line to separate the coefficients of the variables from the constant terms. This type of matrix is often called an augmented matrix.

When doing elimination, each step involves one of the three elementary row operations on the rows of the coefficient matrix:

1. Interchange two rows.
2. Multiply all entries in a row by a nonzero constant.
3. Add a constant multiple of one row to another row.

Each of these elementary row operations leaves unchanged the solutions to the associated system of linear equations. The idea of elimination is to apply these elementary row operations to the coefficient matrix until it is in a simple enough form that we can simply read off the solutions to the original system of equations.

When we use elementary row operations on a matrix, we will indicate the type of operation along with an arrow.

**Example**: Solve the system \( \begin{cases} x + y + 3z = 4 \\ 2x + 3y - z = 1 \\ -x + 2y + 2z = 1 \end{cases} \) using elimination.

The associated coefficient matrix for this system is \( \begin{bmatrix} 1 & 1 & 3 & 4 \\ 2 & 3 & -1 & 1 \\ -1 & 2 & 2 & 1 \end{bmatrix} \).

We apply elementary row operations to clear out the first column:

\[
\begin{bmatrix}
1 & 1 & 3 & 4 \\
2 & 3 & -1 & 1 \\
-1 & 2 & 2 & 1 \\
\end{bmatrix} \xrightarrow{R_2 - 2R_1 \leftrightarrow R_3} \begin{bmatrix}
1 & 1 & 3 & 4 \\
0 & 1 & -7 & -7 \\
-1 & 2 & 2 & 1 \\
\end{bmatrix} \xrightarrow{R_3 + R_1 \leftrightarrow R_3} \begin{bmatrix}
1 & 1 & 3 & 4 \\
0 & 1 & -7 & -7 \\
0 & 3 & 5 & 5 \\
\end{bmatrix} .
\]

Now we are done with the first column and can focus on the other columns:

\[
\begin{bmatrix}
1 & 1 & 3 & 4 \\
0 & 1 & -7 & -7 \\
0 & 3 & 5 & 5 \\
\end{bmatrix} \xrightarrow{R_2 - 3R_2 \leftrightarrow R_2} \begin{bmatrix}
1 & 1 & 3 & 4 \\
0 & 1 & -7 & -7 \\
0 & 26 & 26 & 26 \\
\end{bmatrix} \xrightarrow{\frac{1}{26}R_3 \leftrightarrow R_3} \begin{bmatrix}
1 & 1 & 3 & 4 \\
0 & 1 & -7 & -7 \\
0 & 0 & 1 & 1 \\
\end{bmatrix} .
\]

We can now solve the system from the bottom up: the bottom row yields \( z = 1 \). The middle row then says \( y - 7z = -7 \) from which we see \( y = 0 \). Finally, the top row says \( x + y + 3z = 4 \), so that \( x = 1 \).

Thus, the system has the unique solution \( x = 1, y = 0, z = 1 \).
• We can further simplify our approach to systems of linear equations by rewriting a system of linear equations as a matrix equation (using matrix multiplication).

  ○ For example, \[
  \begin{bmatrix}
  1 & 1 & 3 \\
  2 & 3 & -1 \\
  -1 & 2 & 2
  \end{bmatrix}
  \cdot
  \begin{bmatrix}
  x \\
  y \\
  z
  \end{bmatrix}
  =
  \begin{bmatrix}
  x + y + 3z \\
  2x + 3y - z \\
  -x + 2y + 2z
  \end{bmatrix},
  \]
  so the system \[
  \begin{cases}
  x + y + 3z = 4 \\
  2x + 3y - z = 1 \\
  -x + 2y + 2z = 1
  \end{cases}
  \]
  is equivalent to the single matrix equation
  \[
  \begin{bmatrix}
  1 & 1 & 3 \\
  2 & 3 & -1 \\
  -1 & 2 & 2
  \end{bmatrix}
  \cdot
  \begin{bmatrix}
  x \\
  y \\
  z
  \end{bmatrix}
  =
  \begin{bmatrix}
  4 \\
  1 \\
  1
  \end{bmatrix}.
  \]

0.5 Polynomials

• Let \( F \) be a field.

• Definition: A polynomial with coefficients in \( F \) is an expression of the form \( p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \), where \( n \) is a nonnegative integer and each of the coefficients \( a_i \) is in \( F \).

  ○ The set of all polynomials with coefficients in \( F \) is sometimes denoted as \( P(F) \).
  ○ Two polynomials are equal if and only if their coefficients are all equal.

• Definition: If \( p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \), and \( a_n \neq 0 \), we say that \( p \) is a polynomial of \textit{degree} \( n \). (The degree of the zero polynomial, whose coefficients are all zero, is defined to be \(-\infty\).)

  ○ Example: The polynomial \( p(x) = 3x^{11} - 5x^4 \) has degree 11, while the polynomial \( q(x) = 6 \) has degree 0.
  ○ Observe that \( \deg(pq) = \deg(p) \deg(q) \) and that \( \deg(p + q) \leq \max(\deg(p), \deg(q)) \).

• Polynomials can be added, subtracted, and multiplied in the usual way (using the distributive law to expand out a product): for example, \((x^2 + 1) + (x^3 + 4) = x^3 + x^2 + 5\), while \((x^2 + 1)(x^3 + 4) = x^5 + x^3 + x^2 + 4\).

• Division is marginally more delicate.

• Definition: We say that \( a(x) \) \textit{divides} \( b(x) \) if there exists a polynomial \( q(x) \) with \( b(x) = a(x) \cdot q(x) \).

  ○ Example: The polynomial \( x - 1 \) divides \( x^2 - 1 \), since \( x^2 - 1 = (x - 1)(x + 1) \).

• In a similar way to the procedure with integers, there is a “long division algorithm” for polynomials:

• Theorem (Division Algorithm): If \( b(x) \) is a polynomial and \( a(x) \) is a polynomial of degree \( m \geq 0 \), then there exist unique polynomials \( q(x) \) and \( r(x) \) such that
  \[
  b(x) = q(x)a(x) + r(x)
  \]
  and where the degree of \( r(x) \) is less than \( m \).

  ○ The idea of the algorithm is simply to perform polynomial “long division”: starting with \( b(x) \), subtract an appropriate multiple of \( a(x) \) to remove this largest-degree term of \( b(x) \), and repeat this procedure until a polynomial with degree lower than \( a(x) \) is obtained.
  ○ We will omit further details of the proof, and content ourselves instead with giving an example.

• Example: Apply the division algorithm to \( a(x) = x^2 + 3x \) and \( b(x) = x^3 - 5 \).

  ○ For \( a(x) = x^2 + 3x \) and \( b(x) = x^3 - 5 \), subtracting \( xa(x) \) will remove the leading \( x^3 \) term from \( b(x) \): \( b(x) - xa(x) = (x^3 - 5) - x(x^2 + 3x) = -3x^2 - 5 \).
  ○ Then adding \( 3a(x) \) will remove the leading \(-3x^2 \) from the remaining portion: \( b(x) - xa(x) + 3a(x) = (-3x^2 - 5) + 3(x^2 + 3x) = -5 + 9x \).
  ○ The resulting polynomial now has degree less than the degree of \( a(x) \), so we are done. Rearranging, we see that \( b(x) = (x - 3)a(x) + (-5 + 9x) \), so the quotient is \( q(x) = x - 3 \) and the remainder is \( r(x) = -5 + 9x \).

• By invoking the division algorithm with a linear polynomial, we obtain a useful property of factorization:

11
Theorem (Remainder Theorem): For any \( a \) in \( F \), the remainder obtained by dividing \( p(x) \) by \( x - a \) is the value \( p(a) \). In particular, \( x - a \) divides \( p(x) \) if and only if \( p(a) = 0 \).

- **Proof:** Since the degree of \( x - a \) is 1, applying the division algorithm to \( p(x) \) and \( x - a \) yields a remainder \( r(x) \) of degree less than 1: in other words, a constant polynomial.

- Thus, we obtain \( p(x) = (x - a)q(x) + r \). Now setting \( x = a \) yields \( p(a) = (a - a)q(a) + r = r \), so the remainder is equal to \( p(a) \).
- The other statement follows immediately: \( x - a \) divides \( p(x) \) if and only if the remainder \( p(a) \) obtained by dividing \( p(x) \) by \( p(a) \) is equal to zero.

**Definition:** If \( x - r \) divides \( p(x) \), we say that \( r \) is a root of the polynomial.

When working with polynomials over the real or complex numbers, we often factor the polynomial to make it easier to study.

- For example, over the real numbers we can write \( x^4 - 1 = (x - 1)(x + 1)(x^2 + 1) \).
- Over the complex numbers this polynomial factors further: \( x^4 - 1 = (x - 1)(x + 1)(x - i)(x + i) \).

Theorem (Fundamental Theorem of Algebra): Over the complex numbers, any polynomial can be completely factored into a product of linear terms. In other words, for any \( p(x) = a_n x^n + \cdots + a_0 \) with \( a_n \neq 0 \), there exist complex numbers \( r_1, \ldots, r_n \) for which \( p(x) = a_n(x - r_1)(x - r_2)\cdots(x - r_n) \).

- Although the Fundamental Theorem of Algebra guarantees the existence of this factorization, it is often difficult to compute the roots \( r_1, \ldots, r_n \).
- If the degree of the polynomial is equal to 2, 3, or 4, there exist algebraic formulas involving radicals for the roots, but for general polynomials of larger degree, it is a theorem of Abel that no such formulas exist. (Of course, any particular polynomial of large degree might still have nice roots.)

- For example, by completing the square, we can obtain the famous quadratic formula: the polynomial \( az^2 + bz + c = 0 \) has the two solutions \( z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \) over the complex numbers. When \( b^2 - 4ac > 0 \) the polynomial has two unequal real roots, when \( b^2 - 4ac = 0 \) the polynomial has a double real root, and when \( b^2 - 4ac < 0 \) the polynomial has two complex-conjugate roots.

- However, since there are no such formulas for polynomials of large degree, unless a nice factorization happens to exist, one must use a numerical approximation procedure such as Newton’s method to compute the roots.

When factoring polynomials by hand, the following result is often useful:

Theorem (Rational Root Test): If \( p(x) = a_n x^n + \cdots + a_0 \) is a polynomial with integer coefficients, and \( \frac{a}{b} \) is a rational root (in lowest terms), then \( b \) divides \( a_n \) and \( a \) divides \( a_0 \). In particular, if the leading coefficient \( a_n \) is equal to 1, then the only possible rational roots are integers dividing the constant term \( a_0 \).

- This result follows by expanding out \( p(a/b) \), clearing denominators, rearranging, and factoring appropriately: one eventually sees that \( b \) divides \( a_n a^n \) and that \( a \) divides \( a_0 b^n \), and since \( a/b \) is in lowest terms, \( b \) must divide \( a_n \) and \( a \) must divide \( a_0 \).
- This test cuts down on the amount of trial and error necessary for finding rational roots of polynomials, since we can write down all the possible rational roots. (Of course, a generic polynomial will not have a rational root, but our examples will usually be set up to have nice factorizations.)

**Example:** Factor \( p(x) = x^3 - x^2 + 4x - 4 \) over the complex numbers.

- By the rational root test, if the polynomial has a rational root then it must be an integer dividing \(-4\): that is, one of \( \pm 1, \pm 2, \pm 4 \). Testing the possibilities reveals that \( x = 1 \) is a root, and then we get the factorization \( p(x) = (x - 1)(x^2 + 4) \).
- The roots of the quadratic (by the quadratic formula) are \( x = \pm 2i \), so the full factorization is \( p(x) = (x - 1)(x + 2i)(x - 2i) \).
Mathematical Induction

- Mathematical induction is a useful tool for proving certain kinds of results, typically, results that hold “for all positive integers”.

- The principle of mathematical induction is as follows: suppose we have a sequence of statements $P(1)$, $P(2)$, $P(3)$, and so forth. If $P(1)$ is true, and $P(n)$ implies $P(n + 1)$ for every $n \geq 1$, then $P(k)$ is true for every positive integer $k$.

  - A useful analogy for understanding the inductive principle is of climbing a ladder: if we can get on the first rung of the ladder, and we can always climb from one rung to the next, then we can eventually climb to any rung of the ladder (no matter how high).
  
  - We often refer to the step of showing that $P(1)$ is true as the base case, and the step of showing that $P(n)$ implies $P(n + 1)$ for every $n \geq 1$ as the inductive step.

- For example, suppose we wish to show that $1 + 2 + 3 + 4 + \cdots + n = \frac{1}{2}n(n + 1)$ for every positive integer $n$.

  - Some quick numerical experimentation will suggest that this formula is correct, but is not likely to suggest a proof.

- To prove that $1 + 2 + 3 + 4 + \cdots + n = \frac{1}{2}n(n + 1)$ for every positive integer $n$, we can use the principle of mathematical induction.

  - If we take $P(n)$ to be the statement “$1 + 2 + 3 + 4 + \cdots + n = \frac{1}{2}n(n + 1)$”, then by the inductive principle, all we need to do is show that $P(1)$ is true and that $P(n)$ implies $P(n + 1)$ for each $n \geq 1$.

  - The statement $P(1)$ simply reads $1 = \frac{1}{2} \cdot 1 \cdot 2$, which is clearly true.

  - The statement $P(n)$ says that $1 + 2 + 3 + 4 + \cdots + n = \frac{1}{2}n(n + 1)$, while the statement $P(n + 1)$ says that $1 + 2 + 3 + 4 + \cdots + n + (n + 1) = \frac{1}{2}(n + 1)(n + 2)$.

  - To prove that $P(n)$ implies $P(n + 1)$, we need to start from the statement $1+2+3+4+\cdots+n = \frac{1}{2}n(n+1)$ and use it (somehow) to show that $1 + 2 + 3 + 4 + \cdots + n + (n + 1) = \frac{1}{2}(n + 1)(n + 2)$.

  - We can do this as follows: observe that

    
    \[
    1 + 2 + 3 + 4 + \cdots + n + (n + 1) = [1 + 2 + 3 + 4 + \cdots + n] + (n + 1)
    \]
    
    \[
    = \frac{1}{2}n(n + 1) + (n + 1)
    \]
    
    \[
    = \frac{1}{2}n^2 + \frac{1}{2}n + n + 1
    \]
    
    \[
    = \frac{1}{2}(n^2 + 3n + 2) = \frac{1}{2}(n + 1)(n + 2).
    \]

  where we applied the “inductive hypothesis” piece of information that $1 + 2 + 3 + 4 + \cdots + n = \frac{1}{2}n(n + 1)$ to go from the first line to the second, and then simply did algebra to rearrange the result into the desired expression.

  - Since we have proven the two required pieces, namely that $P(1)$ is true and that $P(n)$ implies $P(n + 1)$, by the principle of mathematical induction, $P(k)$ is true for every $k \geq 1$.

- Induction arguments are useful because they can convert difficult direct proofs into (often) comparatively routine exercises.

  - The base case is usually an easy example where the result is obvious or almost obvious, while the inductive step gives a clear hypothesis to start with and an equally clear goal to reach. Generally, most of the work in the proof goes into the proof of the inductive step.
• Here are a few more examples of proofs by induction, written in a more typical style.

• **Example:** Prove that $1 + 3 + 5 + \cdots + (2n - 1) = n^2$ for every positive integer $n$.
  - We prove this by induction on $n$.
  - For the base case $n = 1$, we must show that $1 = 1$ which is clearly true.
  - For the inductive step, we are given that $1 + 3 + 5 + \cdots + (2n - 1) = n^2$ and must show that $1 + 3 + 5 + \cdots + (2n - 1) + (2n + 1) = (n + 1)^2$.
  - By the inductive hypothesis, we can write
    \[
    1 + 3 + 5 + \cdots + (2n - 1) + (2n + 1) = [1 + 3 + 5 + \cdots + (2n - 1)] + (2n + 1)
    = n^2 + 2n + 1 = (n + 1)^2
    \]
  - and therefore we see $1 + 3 + 5 + \cdots + (2n - 1) + (2n + 1) = (n + 1)^2$, as required.
  - By induction, $1 + 3 + 5 + \cdots + (2n - 1) = n^2$ for every positive integer $n$.

• **Example:** Prove that $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n \cdot (n + 1)} = \frac{n}{n + 1}$ for every positive integer $n$.
  - We prove this by induction on $n$.
  - For the base case $n = 1$, we must show that $\frac{1}{1 \cdot 2} = \frac{1}{2}$ which is clearly true.
  - For the inductive step, we are given that $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n \cdot (n + 1)} = \frac{n}{n + 1}$ and must show that
    \[
    \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n \cdot (n + 1)} + \frac{1}{(n + 1) \cdot (n + 2)} = \frac{n + 1}{n + 2}
    \]
  - By the inductive hypothesis, we can write
    \[
    \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n \cdot (n + 1)} + \frac{1}{(n + 1) \cdot (n + 2)} = \left[ \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n \cdot (n + 1)} \right] + \frac{1}{(n + 1) \cdot (n + 2)}
    = \frac{n}{n + 1} + \frac{1}{(n + 1) \cdot (n + 2)}
    = \frac{n(n + 2) + 1}{(n + 1)(n + 2)} = \frac{(n + 1)^2}{(n + 1)(n + 2)} = \frac{n + 1}{n + 2}
    \]
  - and therefore we see $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n \cdot (n + 1)} + \frac{1}{(n + 1) \cdot (n + 2)} = \frac{n + 1}{n + 2}$, as required.
  - By induction, $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n \cdot (n + 1)} = \frac{n}{n + 1}$ for every positive integer $n$.

• There are various modifications to this “basic” form of induction. The procedure for any induction problem is essentially the same, however: we establish a base case, and prove an inductive step.
  - We often want to start at a different base case than $n = 1$: frequently, we instead start at $n = 0$ or $n = 2$.
  - As long as we establish the base case and inductive step, the inductive principle will still work.
  - If, for example, our base case is $n = 2$, then we would prove $P(2)$ is true and that $P(n)$ implies $P(n + 1)$, with the conclusion being that $P(k)$ is true for all integers $k \geq 2$.

• **Example:** Show that $2^n > n^2$ for all integers $n \geq 5$.
  - We prove this by induction on $n$.
  - For the base case $n = 5$, we must show that $2^5 > 5^2$, or $32 > 25$, which is clearly true.
  - For the inductive step, we are given that $2^n > n^2$ and $n \geq 5$, and must show that $2^{n + 1} > (n + 1)^2$.
  - By the inductive hypothesis, we can write $2^{n + 1} = 2 \cdot 2^n > 2n^2$.
  - Furthermore, since $n \geq 5$, $2n^2 = n^2 + n^2 \geq n^2 + 5n \geq n^2 + 2n + 1 = (n + 1)^2$. 
  -
Putting the inequalities together, we see that $2^{n+1} > 2n^2 \geq (n+1)^2$, so $2^{n+1} > (n+1)^2$ as required.

Therefore, by induction, $2^n > n^2$ for all integers $n \geq 5$.

- Another flavor of induction is called “complete induction” or “strong induction”: rather than assuming the immediately previous case, we assume all of the previous cases: the inductive step is now that $P(1), P(2), \ldots, P(n)$ collectively imply $P(n+1)$.

- It may seem like we are assuming extra information, but in fact complete induction and (regular) induction are logically equivalent.

- **Example**: Prove that every positive integer $n \geq 2$ can be factored into a product of prime numbers, where we consider a product involving only one term to be a product. (Recall that a number greater than 1 is prime if its only divisors are 1 and itself.)

  - We will show this by complete induction on $n$.
  - For the base case, we take $n = 2$, which is already prime.
  - For the inductive step, suppose $n \geq 3$ and that all integers less than $n$ can be factored into a product of prime numbers.
  - If $n$ is prime, then we are already done.
  - Otherwise, $n$ necessarily has a divisor not equal to 1 or itself, say $n = ab$.
  - Then $a$ and $b$ are both between 1 and $n$, and so by the strong induction hypothesis, both $a$ and $b$ can be written as a product of prime numbers. Hence, so can $n$ (simply multiply the two products).
  - In all cases, $n$ can be factored into a product of prime numbers, which establishes the inductive step.
  - Since both the base case and inductive step hold, by (complete) induction, every positive integer $n \geq 2$ can be factored into a product of prime numbers.

- In some situations, it may be necessary to have multiple base cases depending on the structure of the induction argument:

  - **Example**: Let $a_0 = 2$, $a_1 = 5$, and, for $n \geq 2$, let $a_n = 5a_{n-1} - 6a_{n-2}$. Prove that $a_n = 2^n + 3^n$ for all $n \geq 0$.

    - We will show this by complete induction on $n$.
    - For $n = 0$ and $n = 1$, the result is obvious, since $a_0 = 2^0 + 3^0$ and $a_1 = 2^1 + 3^1$.
    - Now suppose $n \geq 2$. By the complete induction hypothesis and the fact that $n \geq 2$, we have $a_{n-1} = 2^{n-1} + 3^{n-1}$ and $a_{n-2} = 2^{n-2} + 3^{n-2}$, and we want to show that $a_n = 2^n + 3^n$.
    - By the recursion and the induction hypotheses,
      
      \[
      a_n = 5a_{n-1} - 6a_{n-2} \\
      = 5(2^{n-1} + 3^{n-1}) - 6(2^{n-2} + 3^{n-2}) \\
      = 4 \cdot 2^{n-2} + 9 \cdot 3^{n-2} = 2^n + 3^n
      \]

    - and therefore $a_n = 2^n + 3^n$ as claimed.
    - By (complete) induction, we conclude that $a_n = 2^n + 3^n$ for all integers $n \geq 0$.