Linear algebra: the theory of vector spaces and linear transformations

Linear algebra: the theory of vector spaces and linear transformations

Aaron Greicius
Northwestern University

Contents

Chapter 1

Foundations

1.1 Sets and functions

We will gradually make our way to definitions of the vector spaces and linear transformations mentioned in this text's title. For now it will suffice to observe that a vector space is a certain kind of *set*, and a linear transformation is a special type of *function*. Accordingly we gather here some notions about sets and functions that will come in handy once we meet the two main players of linear algebra.

1.1.1 Sets

Definition 1.1.1 Sets. A **set** is a collection of objects. An object x is a **member** (or **element**) of a set A if A contains x. In this case, we write $x \in A$. If x is not a member of A, we write $x \notin A$. ◊

We use curly braces to describe the contents of a set. For example, $A = \{1, 2, 3\}$ is the set containing the first three positive integers, and $B = \{1, 2, 3, \dots\}$ is the set of all positive integers. The defining property of sets is that they are completely determined by their members, and nothing more. In particular, when describing sets as above, it does not matter in what order the elements are listed, nor if they are repeated: e.g., $\{1, 2, 3\}$, $\{2, 1, 3\}$, and $\{2, 1, 1, 3, 2\}$ are three descriptions of the same set. This somewhat slippery notion is made perfectly clear by specifying exactly what it means for two sets to be equal, as we do below.

Definition 1.1.2 Set equality. Sets A and B are **equal**, denoted $A = B$, if they have precisely the same elements: i.e., if for any object x, we have $x \in A$ if and only if $x \in B$. ◊

Set membership naturally extends to a notion of one set "lying" within another.

Definition 1.1.3 Set inclusion (subsets). A set A is a **subset** of a set B, denoted $A \subseteq B$, if every member of A is a member of B: i.e., $x \in A$ implies $x \in B$ for any object x. The relation $A \subseteq B$ is called **set inclusion**. ◊

Remark 1.1.4 The definitions of set equality and the subset relation make use of two important logical operations: namely, the "if and only if" (or "iff" for short) and "if-then" operations. We describe these notions in more detail in Logic, and we outline techniques for proving "if and only if" and "if-then" statements, including statements about set equality and the subset relation, in Proof techniques.

With the fundamental notions of membership, equality, and subset in place, we now introduce means of building new sets from existing ones. The first is a manner of carving out a subset of a given set using a specified property.

Definition 1.1.5 Set-builder notation. Let A be a set, and let P be a property that elements of A either satisfy or do not satisfy. For an element $x \in A$, let $P(x)$ denote the statement that x satisfies P. The set of all elements of A satisfying P is denoted

$$\{x \in A \colon P(x)\}.$$

◊

Remark 1.1.6 Set builder notation '$\{x \in A \colon P(x)\}$' is parsed from left to right as follows:

- '\{' is read as "the set of"
- '$x \in A$' is read as "elements of A"
- ': ' is read as "such that"
- '$P(x)$' is read as "$P(x)$ is true".

Taken altogether we get: "The set of elements of A such that $P(x)$ is true".

Example 1.1.7 Let $A = \{0, 1, 2, \dots\}$ be the set of nonnegative integers. The subset $B = \{0, 2, 4, \dots\}$ of *even* positive integers can be described using set-builder notation as

$$B = \{x \in A \colon x \text{ is even}\},$$

or alternatively,

$$B = \{x \in A \colon x = 2k \text{ for some integer} k\}.$$

□

Next we use set builder notation, the set membership relation, and some basic logic to define the union, intersection, and difference of sets.

Definition 1.1.8 Basic set operations. Let A and B be subsets of a common set X.

- *Set union.*

  The **union** $A \cup B$ of A and B is defined as

  $$A \cup B = \{x \in X \colon x \in A \text{ or } x \in B\}.$$

  More generally, the union $A_1 \cup A_2 \cdots \cup A_n$ of a collection of subsets of $A_i \subseteq X$ is defined as

  $$A_1 \cup A_2 \cdots \cup A_n = \{x \in X \colon x \in A_i \text{ for some } 1 \leq i \leq n\}.$$

- *Set intersection.*

  The **intersection** $A \cap B$ of A and B is defined as

  $$A \cap B = \{x \in X \colon x \in A \text{ and } x \in B\}.$$

  More generally, the intersection $A_1 \cap A_2 \cdots \cap A_n$ of a collection of subsets of $A_i \subseteq X$ is defined as

  $$A_1 \cap A_2 \cdots \cap A_n = \{x \in X \colon x \in A_i \text{ for all } 1 \leq i \leq n\}.$$

- *Set difference.*

  The **difference** $A - B$ is defined as

  $$A - B = \{x \in X : x \in A \text{ and } x \notin B\}.$$

$\Diamond$

With the help of these set operations, we can now describe some common sets used in mathematics.

Definition 1.1.9 Common mathematical sets. We denote by $\mathbb{R}$ the set of all real numbers. The **integers** $\mathbb{Z}$ and **rational numbers** $\mathbb{Q}$ are the subsets of $\mathbb{R}$ defined as

$$\mathbb{Z} = \{0, 1, 2, 3, \dots\} \cup \{-1, -2, -3, \dots\}$$
$$\mathbb{Q} = \{x \in \mathbb{R} : x = \tfrac{m}{n} \text{ for some } m, n \in \mathbb{Z}\}.$$

The **complex numbers** $\mathbb{C}$ are defined as the set of all formal expressions of the form $a + bi$, where $a, b \in \mathbb{R}$. Identifying a real number a with the complex number $a + 0i$, we have the following chain of subsets:

$$\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

The **empty set** is the set containing no objects, denoted $\{\ \}$ or $\varnothing$. $\Diamond$

Lastly, we define the *cartesian product* of sets, which is a formal description of an ordered collection of objects.

Definition 1.1.10 Cartesian product. An n**-tuple** (or **sequence**) of the sets $A_1, A_2, \dots, A_n$ is an ordered list $(a_1, a_2, \dots, a_n)$ such that $a_i \in A_i$ for all $1 \leq i \leq n$. We define two n-tuples $(a_1, a_2, \dots, a_n)$, and $(a'_1, a'_2, \dots, a'_n)$ to be equal, denoted $(a_1, a_2, \dots, a_n) = (a'_1, a'_2, \dots, a'_n)$, if $a_i = a'_i$ for all $1 \leq i \leq n$. We call n the **length** of the sequence $(a_1, a_2, \dots, a_n)$, and we call a_i its i**-th entry** for all $1 \leq i \leq n$.

The **(Cartesian) product** of the sets $A_1, A_2, \dots, A_n$, denoted $A_1 \times A_2 \times \dots \times A_n$ or $\prod_{i=1}^{n} A_i$, is the set of all n-tuples: i.e.,

$$\prod_{i=1}^{n} A_i = A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n) : a_i \in A_i \text{ for all } 1 \leq i \leq n\}.$$

Given a set A, its n**-ary Cartesian product** A^n is defined as

$$A^n = \prod_{i=1}^{n} A = \underbrace{A \times A \times \dots \times A}_{n \text{ times}}.$$

$\Diamond$

Remark 1.1.11 We have more suggestive names for n-tuples when n is small: a 2-tuple (a, b) is called a pair, a 3-tuple (a, b, c) is called a triple, a 4-tuple (a, b, c, d) is called a quadruple, etc.. We will use the generic term "tuple" to designate a n-tuple of unspecified length.

Remark 1.1.12 Observe how tuples capture the notion of an *ordered* collection of object. For example, whereas the sets $\{1, 1, 2, 3\}$ and $\{1, 2, 2, 3\}$ are all equal to one another, the 4-tuples $(1, 1, 2, 3)$ and $(1, 2, 2, 3)$ are not: they differ in their second entries.

What about the tuples $(1,1,1)$ and $(1,1,1,1)$? Are these equal? Technically our definition of equality only applies to tuples living in the same fixed Cartesian product. In particular, for the question of equality to make sense, the tuples must have the same length. As such we will officially avoid writing things like $(1,1,1) \neq (1,1,1,1)$, although unofficially we consider these two objects as completely different. You should think of $(1,1,1)$ and $(1,1,1,1)$ as creatures living on two different planets in the universe of tuples.

1.1.2 Functions

Definition 1.1.13 Functions. Let X and Y be two sets. A **function from** X **to** Y, denoted $f\colon X \to Y$, is a rule which, given any **input** $x \in X$, returns an **output** $y \in Y$. In this case we write $y = f(x)$ and call y the **image of** x **under** f, or the **value of** f **at** x. Similarly, we say f **maps** (or **sends**) the input x to the output y.

The set X is called the **domain** of f; the set Y is called the **codomain** of f.

When we wish to indicate what rule defines the function, we use the elaborated notation

$$\begin{aligned} f\colon X &\to Y \\ x &\mapsto f(x). \end{aligned}$$

This is read as "The function f with domain X and codomain Y maps an input x to the output $f(x)$". ◊

Example 1.1.14 Consider the function

$$\begin{aligned} f\colon \mathbb{Z} &\to \mathbb{Z} \\ x &\mapsto x^2. \end{aligned}$$

This function has domain and codomain equal to $\mathbb{Z}$ and maps an integer to its square. □

Example 1.1.15 Arithmetic operations as functions. Our familiar arithmetic operations can be expressed as functions whose inputs are pairs of numbers: addition is the function

$$\begin{aligned} a\colon \mathbb{R}\times\mathbb{R} &\to \mathbb{R} \\ (x,y) &\mapsto x+y \end{aligned}$$

and multiplication is the function

$$\begin{aligned} m\colon \mathbb{R}\times\mathbb{R} &\to \mathbb{R} \\ (x,y) &\mapsto xy \end{aligned}$$

□

Remark 1.1.16 Invoking the notion of a rule in the definition of a function is admittedly somewhat vague. A more precise definition can be given using the Cartesian product. Namely, given sets X and Y, we define a function $f\colon X \to Y$ to be a subset $f \subseteq X \times Y$ satisfying the following property: for all $x \in X$ there is a unique element $(x,y) \in f$. The uniqueness of the pair (x,y) then allows us to define the value $f(x)$ of f at x as y, denoted $f(x) = y$.

As with sets and tuples, we are obliged to specify what we mean for two functions to be equal. The definition below makes clear how the the domain and codomain, as well as the rule that takes inputs to outputs, are all essential ingredients of a function.

Definition 1.1.17 Function equality. Functions f and g are **equal** if

i they have the same domain X and codomain Y, and

ii for all $x \in X$, we have $f(x) = g(x)$.

$\Diamond$

Definition 1.1.18 Image of a set. Given a function $f\colon X \to Y$ and a subset $A \subseteq X$, the **image of** A **under** f, denoted $f(A)$, is defined as

$$f(A) = \{y \in Y \colon f(a) = y \text{ for some } a \in A\}.$$

In other words, $f(A)$ is the set of all outputs $f(a)$, where $a \in X$.

The **image of** f, denoted $\operatorname{im} f$, is the set of *all* outputs of f: i.e.,

$$\operatorname{im} f = f(X) = \{y \in Y \colon f(x) = y \text{ for some } x \in X\}.$$

$\Diamond$

Definition 1.1.19 Injective, surjective, bijective. Let $f\colon X \to Y$ be a function.

- The function f is **injective** (or **one-to-one**) if for all $x, x' \in X$, if $f(x) = f(x')$, then $x = x'$: equivalently, if $x \neq x'$, then $f(x) \neq f(x')$.
- The function f is **surjective** (or **onto**) if for all $y \in Y$, there is an $x \in X$ such that $f(x) = y$: equivalently, $\operatorname{im} f = Y$.
- The function f is **bijective** (or a **one-to-one correspondence**) if it is injective and surjective.

$\Diamond$

Example 1.1.20 Role of domain and codomain in injectivity and surjectivity. Consider the following three functions

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} & \qquad g\colon [0,\infty) &\to \mathbb{R} & \qquad h\colon [0,\infty) &\to [0,\infty) \\ x &\mapsto x^2 & x &\mapsto x^2 & x &\mapsto x^2 \end{aligned}.$$

Although all three functions are defined using the same rule ($x \mapsto x^2$), they are not equal thanks to their differing domains and/or codomains. The choice of domain and codomain in these examples also plays a crucial role in determining whether the function is injective and/or surjective. The function f is neither injective ($f(-2) = f(2) = 4$) nor surjective ($f(X) = [0,\infty) \neq \mathbb{R}$); the function g is injective but not surjective; the function h is both injective and surjective, hence bijective. $\square$

Definition 1.1.21 Function composition. Suppose $f\colon Z \to W$ and $g\colon X \to Y$ are functions satisfying $Y \subseteq Z$. The **composition of** f **and** g is the function $f \circ g\colon X \to W$ defined as $f \circ g\,(x) = f(g(x))$, for all $x \in X$. $\Diamond$

Definition 1.1.22 Identity and inverse functions. For any set X the **identity function** on X is the function $\mathrm{id}_X\colon X \to X$ defined as $\mathrm{id}_X(x) = x$ for all $x \in X$.

A function $f\colon X \to Y$ is **invertible** if there is a function $g\colon Y \to X$

satisfying $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$: i.e.,

$$g(f(x)) = x \text{ for all } x \in X,$$
$$f(g(y)) = y \text{ for all } y \in Y.$$

The function g in this case is called the **inverse** of f, denoted $g = f^{-1}$. ♢

Theorem 1.1.23 Invertible is equivalent to bijective. *A function $f\colon X \to Y$ is invertible if and only if it is bijective.*

Proof. The proof of this theorem is left as an example of proving "if and only if" statements. See Example 1.3.3. ■

1.2 Logic

When dealing with mathematical statements and arguments, we must pay close attention to logical structure. This section addresses Different logical connectors give rise to different proof approaches. For the rest of this section, the symbols $\mathcal{P}$ and $\mathcal{Q}$ will stand for propositions.

1.2.1 Propositional logic

A *proposition* is a sentence that is either true or false. We build *compound propositions* from component propositions using *logical operators* (or *logical connectors*); the truth value of the compound proposition is defined as a function of the truth values of the component propositions.

Definition 1.2.1 Logical operators.

- *Negation.*

  Given a proposition $\mathcal{P}$, the **negation of** $\mathcal{P}$ is the proposition "Not $\mathcal{P}$", denoted $\neg\mathcal{P}$ in logical notation, the truth value of which is defined as follows: $\neg\mathcal{P}$ is true exactly when $\mathcal{P}$ is false.

- *Conjunction (logical and).*

  Given propositions $\mathcal{P}$ and $\mathcal{Q}$, their **conjuction** is the proposition "$\mathcal{P}$ and $\mathcal{Q}$", denoted $\mathcal{P} \wedge \mathcal{Q}$ in logical notation, the truth value of which is defined as follows: $\mathcal{P} \wedge \mathcal{Q}$ is true when both $\mathcal{P}$ and $\mathcal{Q}$ are true, and false otherwise.

- *Disjunction (logical or).*

  Given propositions $\mathcal{P}$ and $\mathcal{Q}$, their **disjunction** is the proposition "$\mathcal{P}$ or $\mathcal{Q}$", denoted $\mathcal{P} \vee \mathcal{Q}$ in logical notation, the truth value of which is defined as follows: $\mathcal{P} \vee \mathcal{Q}$ is true when at least one of $\mathcal{P}$ and $\mathcal{Q}$ is true, and false otherwise.

- *Logical implication (if-then).*

  Given propositions $\mathcal{P}$ and $\mathcal{Q}$, the proposition "If $\mathcal{P}$, then $\mathcal{Q}$", denoted $\mathcal{P} \implies \mathcal{Q}$ in logical notation, is called an **implication**, and its truth value is defined as follows: $\mathcal{P} \implies \mathcal{Q}$ is false when $\mathcal{P}$ is true and $\mathcal{Q}$ is false, and true otherwise.

- *Logical equivalence (if and only if).*

  Given propositions $\mathcal{P}$ and $\mathcal{Q}$, the proposition "$\mathcal{P}$ if and only if $\mathcal{Q}$", denoted $\mathcal{P} \iff \mathcal{Q}$ in logical notation, is called an **equivalence**, and its truth value is defined as follows: $\mathcal{P} \iff \mathcal{Q}$ is true when $\mathcal{P}$ and $\mathcal{Q}$ have the same truth value, and false otherwise. We say $\mathcal{P}$ and $\mathcal{Q}$ are **logically equivalent** when $\mathcal{P} \iff \mathcal{Q}$ is true.

◊

Remark 1.2.2 A *truth table* of a compound proposition $\mathcal{P}$ is a concise way of displaying how the truth value of $\mathcal{P}$ depends on the truth values of its component propositions. It contains one row for each possible truth assignment of the component propositions. As illustration, we give the truth tables for the logical operators defined above:

$\mathcal{P}$	$\neg\mathcal{P}$
T	F
F	T

$\mathcal{P}$	$\mathcal{Q}$	$\mathcal{P} \wedge \mathcal{Q}$
T	T	T
T	F	F
F	T	F
F	F	F

$\mathcal{P}$	$\mathcal{Q}$	$\mathcal{P} \vee \mathcal{Q}$
T	T	T
T	F	T
F	T	T
F	F	F

$\mathcal{P}$	$\mathcal{Q}$	$\mathcal{P} \implies \mathcal{Q}$
T	T	T
T	F	F
F	T	T
F	F	T

$\mathcal{P}$	$\mathcal{Q}$	$\mathcal{P} \iff \mathcal{Q}$
T	T	T
T	F	F
F	T	F
F	F	T

Example 1.2.3 Use a truth table to find all truth value assignments of $\mathcal{P}$ and $\mathcal{Q}$ making the compound proposition

$$\neg\mathcal{P} \implies (\mathcal{Q} \implies P)$$

false.

Solution. We construct a truth table with columns for $\mathcal{P}$, $\mathcal{Q}$, $\neg\mathcal{P}$, $\mathcal{Q} \implies \mathcal{P}$, and $\neg\mathcal{P} \implies (\mathcal{Q} \implies P)$:

$\mathcal{P}$	$\mathcal{Q}$	$\neg\mathcal{P}$	$\mathcal{Q} \implies \mathcal{P}$	$\neg\mathcal{P} \implies (\mathcal{Q} \implies \mathcal{P})$
T	T	F	T	T
T	F	F	T	T
F	T	T	F	F
F	F	T	T	T

.

We conclude that $\neg\mathcal{P} \implies (\mathcal{Q} \implies P)$ is false exactly when $\mathcal{P}$ is false and $\mathcal{Q}$ is true. (It follows that $\neg\mathcal{P} \implies (\mathcal{Q} \implies P)$ is equivalent to $Q \implies P$.) ☐

Remark 1.2.4 Our definitions of the logical operators above are chosen to agree with their usage in a very particular type of discourse: namely, *mathematical* discourse. They do not always agree with their use in natural language: hence the modifier "logical" in their titles.

For example, disjunctions in natural language of the form "$\mathcal{P}$ or $\mathcal{Q}$" are often understood to be true when $\mathcal{P}$ is true or $\mathcal{Q}$ is true, *but not both.* This notion of disjunction is called the *exclusive or* in logic, in contrast to the standard logical or.

Similarly, according to our definition, the implication "If the President of the US is a dog, then the Vice President of the US is a cat" is true, since the proposition "The President of the US is a dog" is false. (In logic we say the implication is *vacuously true* in this case.) However, working outside of our logical definitions of truth value, we may have been inclined to treat this implication as false, since both "The President of the US is a dog" and "The Vice President of the US is a cat" are false.

1.2.2 Predicate logic

Propositions like “All humans are mortal” and “Every positive real number has a square-root” are modeled in logic in the form “For all x, $P(x)$” and “For all r, there exists an s such that $Q(r,s)$”, where $P(x)$ stands for the phrase “x is mortal” and $Q(r,s)$ stands for the phrase “s is a square-root of r”. Observe that $P(x)$ and $Q(r,s)$ on their own are not propositions; there is no truth value to “x is mortal” or “s is a square-root of r”. Instead, we think of $P(x)$ and $Q(r,s)$ as *functions* which return propositions when evaluated at a specific choice for x, or for r and s. For example, evaluating $P(x)$ at $x =$ Aaron Greicius yields the proposition “Aaron Greicius is mortal”, which happens to be true at the time of writing. Similarly evaluating $Q(r,s)$ at $r = 2, s = 11$ yields the proposition “11 is a square-root of 2”, which happens to be false. In logic $P(x)$ and $Q(r,s)$ are called *propositional functions* (also called *predicates*): functions whose outputs are propositions.

The propositions “For all x, $P(x)$” and “For all r, there exists an s such that $Q(r,s)$” are obtained from $P(x)$ and $Q(r,s)$ by *bounding* them with a series of *quantifiers* (e.g., “for all x”, “there exists an s”). The definition below allows us to assign truth values to such propositions.

Definition 1.2.5 Let D be a set, and let P be a propositional function that assigns to all elements $d \in D$ the proposition $P(d)$. The set D is called the **domain of discourse** of P.

- *Universal quantifier.*

  The proposition “For all x in D, $P(x)$”, denoted $\forall x P(x)$ in logical notation, is called a **universal quantification**, and its truth value is defined as follows: $\forall x P(x)$ is true if for all elements d of D, the proposition $P(d)$ is true; it is false if there is some $d \in D$ such that $P(d)$ is false.

- *Existential quantifier.*

  The proposition “There exists an x in D such that $P(x)$”, denoted $\exists x P(x)$ in logical notation, is called an **existential quantification**, and its truth value is defined as follows: $\exists x P(x)$ is true if there is some element d of D for which the proposition $P(d)$ is true; it is false if $P(d)$ is false for all $d \in D$.

◊

Remark 1.2.6 Truth depends on domain of discourse. Just as a function is not properly defined before its domain is specified, we do not have a well-defined propositional function, nor well-defined truth values of propositions built from this propositional function, until its domain of discourse is given.

For example, let $P(x)$ be “x>0”. If we declare $D = (0,\infty)$, then the proposition $\forall x P(x)$ is true, since by definition every $d \in (0,\infty)$ is positive. On the other hand, if we declare $D = \mathbb{R}$, the proposition $\forall x P(x)$ is false since not all elements of $\mathbb{R}$ are positive: indeed, -1 is negative, making $P(-1)$ false.

Because of the important role played by the domain of discourse D, we sometimes include it in our quantifier expressions: e.g., $\forall x \in D P(x)$, $\exists x \in D P(x)$. Using this convention allows us to see more immediately that $\forall\, x \in (0,\infty)\, P(x)$ is true and $\forall\, x \in \mathbb{R}\, P(x)$ is false.

Example 1.2.7 Modeling “Every positive number has a square-root”. Model the sentence “Every positive real number has a square-root” in the form $\forall x P(x)$, where P is a propositional function with domain of discourse $D = \mathbb{R}$. Determine the truth value of $\forall x\, P(x)$ using Definition 1.2.5.

Solution. Fix our domain of discourse to be $D = \mathbb{R}$. For any $r, s \in \mathbb{R}$, let $Q(r, s)$ be the proposition that s is a square-root of r. Define $P(x)$ to be the propositional function

$$x > 0 \implies \exists y\, Q(x, y).$$

Thus for any $r \in \mathbb{R}$, $P(r)$ is the proposition that if r is positive, then r has a square-root. It follows that $\forall x \in \mathbb{R}\, P(x)$ is the proposition that for all real numbers x, if x is positive, then x has a square-root. This is clearly equivalent to the proposition that every positive real number has a square-root, as desired.

Lastly, we use Definition 1.2.5 to show $\forall x \in \mathbb{R}\, P(x)$ is true. Take any $r \in \mathbb{R}$. The real number r is either positive or not positive. If r is not positive, then $r > 0$ is false and hence $P(r)$, which is the implication $r > 0 \implies \exists y\, Q(r, y)$, is true vacuously. If r is positive, then $r > 0$ is true, and $\exists y\, Q(r, y)$ is true, since r has a square-root in this case: namely, $\sqrt{r}$. But if $r > 0$ is true and $\exists y\, Q(r, y)$ is true, then the implication $r > 0 \implies \exists y\, Q(r, y)$ is true. We have proved $P(r)$ is true for all $r \in \mathbb{R}$. Thus $\forall x \in \mathbb{R}\, P(x)$ is true. □

Warning 1.2.8 Order of quantifiers matters. As Example 1.2.7 illustrates, we can take a propositional function $Q(x, y)$ in two variables and quantify one of the two variables to obtain a propositional function in one variable: e.g., $P(x) = \exists y\, Q(x, y)$ or $R(y) = \forall x\, Q(x, y)$. Proceeding in this manner we engender propositions involving sequences of quantifiers. Mark well that the order of the quantifiers in such sequences is important!

For example, letting $Q(x, y)$ be "y is a square-root of y" with domain of discourse $D = (0, \infty)$. The proposition $\forall x \in \mathbb{R}\, \exists y \in \mathbb{R}\, Q(x, y)$ says that every positive real number has a positive square-root (true); the proposition $\exists y \in \mathbb{R}\, \forall x \in \mathbb{R}\, Q(x, y)$ says that there is a positive real number that is the square-root of all real numbers (false).

In general to "unpack" a sequence of quantifiers we start from the leftmost quantifier and proceed to the right.

Remark 1.2.9 Negating quantifiers. Let $P(x)$ be a propositional function with domain of discourse D. According to Definition 1.2.5 a universal quantification $\forall x\, P(x)$ is false if it is *not the case* that $P(d)$ is true for all for all $d \in D$; that is, if there is some $d \in D$ such that $P(d)$ is false. Letting $\neg P(x)$ be the propositional function defined as $\neg P(d)$ for all $d \in D$, we see that $\forall x\, P(x)$ is false if and only if $\exists x\, \neg P(x)$ is true. This proves the equivalence

$$\neg \forall x\, P(x) \iff \exists x\, \neg P(x).$$

Similarly, we have

$$\neg \exists x\, P(x) \iff \forall x\, \neg P(x).$$

These equivalences can be understood as rules of negating quantifier statements. They can be summarized as follow: "swap quantifiers and negate the predicate."

The example below taken from calculus nicely illustrates how to negate a proposition that involves a sequence of quantifiers.

Example 1.2.10 The limit does not exist. Let $f(x)$ be a function with domain $\mathbb{R}$, and let $c \in \mathbb{R}$ be a point of this domain. By definition, the proposition that $\lim\limits_{x \to c} f(x)$ exists is equivalent to the following proposition:

$$\exists L \in \mathbb{R}\, \forall \epsilon > 0\, \exists \delta > 0\, \forall x \in \mathbb{R}\ (|x - c| < \delta \implies |f(x) - L| < \epsilon). \tag{1.2.1}$$

(We made a number of shortcuts in our logical notation above (e.g. $\forall \epsilon > 0$, $\exists \delta > 0$) in order to simplify the expression somewhat; the intended meaning should still be clear.)

Use the negation rules described in Negating quantifiers to derive a similar proposition equivalent to the statement that $\lim_{x \to c} f(x)$ do not exist.

Solution. Let $\mathcal{P}$ be the proposition in (1.2.1). Using the negation rules in series, we derive the *chain of equivalences* below (see Chains of implications/ equivalences). We've added parentheses to emphasize what is being negated at each step. Note how a quantifiers are "swapped" each time we pass the negation to the right.

$$\begin{aligned}
\neg\mathcal{P} &\iff \forall L \in \mathbb{R}\, \neg\, (\forall \epsilon > 0\, \exists \delta > 0\, \forall x \in \mathbb{R}\ (|x-c| < \delta \implies |f(x) - L| < \epsilon)) \\
&\iff \forall L \in \mathbb{R}\, \exists \epsilon > 0\, \neg\, (\exists \delta > 0\, \forall x \in \mathbb{R}\ (|x-c| < \delta \implies |f(x) - L| < \epsilon)) \\
&\iff \forall L \in \mathbb{R}\, \exists \epsilon > 0\, \forall \delta > 0\, \neg\, (\forall x \in \mathbb{R}\ (|x-c| < \delta \implies |f(x) - L| < \epsilon)) \\
&\iff \forall L \in \mathbb{R}\, \exists \epsilon > 0\, \forall \delta > 0\, \exists x \in \mathbb{R}\ \neg(|x-c| < \delta \implies |f(x) - L| < \epsilon) \\
&\iff \forall L \in \mathbb{R}\, \exists \epsilon > 0\, \forall \delta > 0\, \exists x \in \mathbb{R}\ (|x-c| < \delta \text{ and } |f(x) - L| \not< \epsilon).
\end{aligned}$$

(The last link in our chain uses the fact that $\neg(\mathcal{Q} \implies \mathcal{R})$ is equivalent to $\mathcal{Q} \wedge \neg\mathcal{R}$, as a truth table easily shows.) Translating back into English, we conclude that the limit not existing ($\neg\mathcal{P}$) is equivalent to the following: for every $L \in \mathbb{R}$ there is an $\epsilon > 0$ such that for all $\delta > 0$ there exists an $x \in \mathbb{R}$ satisfying $|x - c| < \delta$ and $|f(x) - L| \not> \epsilon$. Quite a mouthful! $\square$

1.3 Proof techniques

Proof writing is an art, a technical skill you will hone and refine throughout your career; and like any art, proof writing has many tricks of the trade. We gather a few here in the form of a collection of general proof techniques. Part of mastering these techniques involves understanding the circumstances where they can be of use. When selecting a technique, we are guided in part by the logical structure and particular mathematical content of the proposition under consideration. The proof techniques below are organized under this guiding principle.

1.3.1 Logical structure

1.3.1.1 Implication

By Definition 1.2.1 the only time an implication $\mathcal{P} \implies \mathcal{Q}$ is false is when $\mathcal{P}$ is true and $\mathcal{Q}$ is false. Accordingly, the *direct approach* to proving an implication $\mathcal{P} \implies \mathcal{Q}$ is to assume $\mathcal{P}$ is true, and use this assumption to prove $\mathcal{Q}$ is true.

A common *indirect approach* used to prove an implication $\mathcal{P} \implies \mathcal{Q}$ is to prove its *contrapositive* $\neg\mathcal{Q} \implies \neg\mathcal{P}$, which is logically equivalent to the original implication. In this case we assume $\mathcal{Q}$ is not true, and show $\mathcal{P}$ is not true. (Exercise: use a truth table to show the contrapositive is logically equivalent to the original implication.)

Warning 1.3.1 The *converse* of an implication $\mathcal{P} \implies \mathcal{Q}$ is the implication $\mathcal{Q} \implies \mathcal{P}$; the *inverse* of $\mathcal{P} \implies \mathcal{Q}$ is the implication $\neg\mathcal{P} \implies \neg\mathcal{Q}$. Neither the converse nor the inverse is equivalent to the original implication, and thus neither of these can be used to give an indirect proof of $\mathcal{P} \implies \mathcal{Q}$. (Exercise: use a truth table to show that neither the converse nor the inverse implication is logically equivalent to the original implication.)

1.3.1.2 Disjunction

Again, using Definition 1.2.1, to directly show a disjunction $\mathcal{P}$ or $\mathcal{Q}$ is true, we need only show one the two component propositions is true.

Alternatively, we can prove either of the implications $\neg\mathcal{P} \implies \mathcal{Q}$ or $\neg Q \implies \mathcal{P}$, both of which are logically equivalent to $\mathcal{P}$ or $\mathcal{Q}$. (Exercise: use a truth table to show these three propositions are indeed equivalent!)

1.3.1.3 Equivalence

The equivalence $\mathcal{P} \iff \mathcal{Q}$ is logically equivalent to the conjunction

$$(\mathcal{P} \implies \mathcal{Q}) \text{ and } (\mathcal{Q} \implies \mathcal{P}).$$

Accordingly, the standard way of proving $\mathcal{P} \iff \mathcal{Q}$ is to prove the two implications $\mathcal{P} \implies \mathcal{Q}$ and $\mathcal{Q} \implies \mathcal{P}$ separately. (Exercise: use a truth table to show these propositions are indeed equivalent!)

1.3.1.4 Chains of implications/equivalences

The relation "$\mathcal{P}$ implies $\mathcal{Q}$" is transitive: i.e., if $\mathcal{P} \implies \mathcal{Q}$ and $\mathcal{Q} \implies \mathcal{R}$, then $\mathcal{P} \implies \mathcal{R}$. Similarly, the relation "$\mathcal{P}$ is equivalent to $\mathcal{Q}$" is transitive. This allows us to prove an implication or equivalence via a *chain of implications* or *chain of equivalences*. When writing up a proof using this technique, use a vertically aligned format like the example below, treating one implication or equivalence per line, and adding a brief justification to the right:

$$\begin{aligned}
\mathcal{P} &\iff \mathcal{P}_\infty && \text{(justification)}\\
&\iff \mathcal{P}_\in && \text{(justification)}\\
&\vdots\\
&\iff \mathcal{P}_\backslash && \text{(justification)}\\
&\iff \mathcal{Q} && \text{(justification)}.
\end{aligned}$$

It is also possible to build an argument as a hybrid chain of equivalences and implications. In this case the chain is only as strong as its weakest link. For example, a chain of the form

$$\begin{aligned}
\mathcal{P} &\iff \mathcal{Q} && \text{(justification)}\\
&\implies \mathcal{R} && \text{(justification)}\\
&\iff \mathcal{S} && \text{(justification)}
\end{aligned}$$

only shows that $\mathcal{P} \implies \mathcal{S}$. Indeed, we will have $\mathcal{P} \iff \mathcal{S}$ if and only if the intervening implication $\mathcal{Q} \implies \mathcal{R}$ is in fact an equivalence (i.e., the arrow goes both ways).

Warning 1.3.2 It is often tempting, for the sake of space, to try and prove an equivalence $\mathcal{P} \iff \mathcal{Q}$ via a chain of equivalences, as opposed to showing $\mathcal{P} \implies \mathcal{Q}$ and $\mathcal{Q} \implies \mathcal{P}$ separately. When proceeding in this manner, make doubly sure that each $\iff$ is indeed an equivalence: i.e., that the implication arrow really goes both ways at each step. Even if each step in your chain truly is an equivalence, you should consider whether this will be obvious to your reader.

The example below provides the proof that a function is invertible if and only if it is bijective (Theorem 1.1.23). The proof nicely illustrates some of the different techniques used for proving implications and equivalences. Additionally, it is a nice example of how to separate out cases of the argument into clearly distinguished steps.

Example 1.3.3 Proof: invertible is equivalent to bijective. Let $f\colon X \to Y$ be a function. Prove: f is invertible if and only if f is bijective.

Solution. Let $\mathcal{P}$ be the proposition that f is invertible, and let $\mathcal{Q}$ be the proposition that f is bijective. We prove the equivalence $\mathcal{P} \iff \mathcal{Q}$ by proving the two implications $\mathcal{P} \implies \mathcal{Q}$ and $\mathcal{Q} \implies \mathcal{P}$.

Proof of $\mathcal{P} \implies \mathcal{Q}$. We must show that if f is invertible, then f is bijective. Assume f is invertible. Then f has an inverse f^{-1}. We show separately that f is injective and surjective, hence bijective.

f *is injective.* We show $f(x) = f(x') \implies x = x'$ via a chain of implications:

$$\begin{aligned} f(x) = f(x') &\implies f^{-1}(f(x)) = f^{-1}(f(x')) \\ &\implies x = x' && (f^{-1} \circ f = \mathrm{id}_X). \end{aligned}$$

f *is surjective.* Let b be an element of Y. We must show that there is an $x \in X$ such that $f(x) = y$. Letting $x = f^{-1}(y)$, we have

$$\begin{aligned} f(x) &= f(f^{-1}(y)) \\ &= y && (f \circ f^{-1} = \mathrm{id}_Y). \end{aligned}$$

Proof of $\mathcal{Q} \implies \mathcal{P}$. We must show that if f is bijective, then f is invertible. Assume f is bijective. First we define a function $g\colon Y \to X$ as follows: for all $y \in Y$, let $g(y)$ be the unique element $x \in X$ such that $f(x) = y$. Note that our definition of g uses both that f is surjective (there is *some* element x such that $f(x) = y$)) and injective (there is *exactly one* element x such that $f(x) = y$).

We now prove that g is the inverse of f, showing $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$ separately.

$g \circ f = \mathrm{id}_X$. Take any $x \in X$ and let $y = f(x)$. By definition of g, we have $g(y) = x$ and hence $g(f(x)) = g(y) = x$. This proves $g \circ f = \mathrm{id}_X$.

$f \circ g = \mathrm{id}_Y$. Take any $y \in Y$. By definition of g, $g(y)$ is the unique $x \in X$ such that $f(x) = y$. Thus $f(g(y)) = f(x) = y$. This proves $f \circ g = \mathrm{id}_Y$. □

1.3.1.5 Proof by contradiction

The technique of *proof by contradiction* (or *reductio ad absurdum*) proves a proposition $\mathcal{P}$ by (a) assuming the negation $\neg\mathcal{P}$ is true, and then (b) using this assumption to derive a proposition $\mathcal{Q}$ known to be false. The choice of falsehood $\mathcal{Q}$ is completely up to the person providing the proof. However, in order for the proof to be convincing, it should be clear, either logically or because of theory assumed to be known, that $\mathcal{Q}$ is indeed false.

Example 1.3.4 Proof by contradiction. Prove by contradiction that 0 has no multiplicative inverse in the reals: i.e., there is no $r \in \mathbb{R}$ such that $r \cdot 0 = 1$.

Solution. We prove the claim by contradiction. Assume there is an $r \in \mathbb{R}$ such that $r \cdot 0 = 1$. Since $r \cdot 0 = 0$ for any $r \in \mathbb{R}$ (a property of multiplication by 0), we have $1 = r \cdot 0 = 0$: a contradiction since $1 \neq 0$. We conclude that there is no $r \in \mathbb{R}$ such that $r \cdot 0 = 1$. □

Remark 1.3.5 Proof by contradiction resembles, but is not quite the same thing as proving an implication via its contrapositive. Letting F denote an arbitrary falsehood (the $\mathcal{Q}$ described above) what we do in a proof by contradiction is show that the implication $\neg\mathcal{P} \implies F$ is true. Since F is false, and the implication is true, $\neg\mathcal{P}$ must be false: equivalently, $\mathcal{P}$ must be true.

1.3.2 Equalities

Equality is not as simple as it may seem. In general an equality is a mathematical statement of the form

$$\text{LHS} = \text{RHS}. \tag{1.3.1}$$

Here "LHS" and "RHS" stand for left- and right-hand side, respectively. What exactly such an equality means depends very much on what kind of mathematical objects the two sides of the equation are: e.g., numbers, sets, functions, etc. Below we discuss equality for objects of a particular type in detail. (See Subsection 1.3.3 and Subsection 1.3.4.) In all settings, the notion of equality will be *transitive*: i.e., if $x = y$ and $y = z$, then $x = z$. We use transitivity implicitly when proving an equality via a *chain of equalities* as described below.

1.3.2.1 Chain of equalities

Often to prove an equality as in (1.3.1) we proceed in a sequence of intervening bite-size equalities, each of which is easy for the reader to digest. As with chains of implications/equivalences, we present such an argument in a vertically aligned format, with brief justifications to the right:

$$\begin{aligned}
\text{LHS} &= \text{something} && \text{(justifiction)}\\
&= \text{something} && \text{(justifiction)}\\
&\ \vdots \\
&= \text{RHS} && \text{(justifiction).}
\end{aligned}$$

Warning 1.3.6 Never attempt to prove an equality by starting off with the equality you wish to prove, and then deduce a series of further equalities ending in some inanity: e.g.,

$$\begin{aligned}
\text{LHS} &= \text{RHS}\\
\text{something} &= \text{something}\\
&\ \vdots\\
0 &= 0.
\end{aligned}$$

What this suggests is that you are in fact proving an implication: namely, *if* the desired equality is true, *then* $0 = 0$. Clearly this is not what we set out to prove! This type of fallacious argument is called "begging the question" (*petitio principippii* in Latin), as we assume that which was to be proven.

1.3.3 Basic set properties

1.3.3.1 Set inclusion

Let A and B be sets. By Definition 1.1.3, to prove $A \subseteq B$ we must show that for all elements x we have

$$x \in A \implies x \in B.$$

This requires proving the implication above for a general element x, and we may use any of the techniques described in Implication and Chains of implications/equivalences to do so.

1.3.3.2 Set equality

Let A and B be sets. To prove $A = B$ directly using Definition 1.1.2 we must show that for all elements x we have

$$x \in A \iff x \in B.$$

To prove this universal equivalence, we must give an argument for the equivalence that holds for a general element x.

Alternatively, you can prove $A = B$ by proving the two set inclusions $A \subseteq B$ and $B \subseteq A$ separately. This is equivalent to proving the two implications $x \in A \implies x \in B$ and $x \in B \implies x \in A$ separately.

1.3.4 Basic function properties

1.3.4.1 Function equality

By Definition 1.1.17, in order to show functions f and g are equal we must

i show that f and g have the same domain X and codomain Y, and

ii show that $f(x) = g(x)$ for all $x \in X$.

By Definition 1.1.17 The universal quantifier "for all $x \in X$" of item (ii) gives this subtask the feel of proving an *identity*: we must show that equality $f(x) = g(x)$ holds *for all*$x \in X$. By the same token, to show (ii) does not hold, it suffices to show that $f(x) \neq g(x)$ for some $x \in X$.

1.3.4.2 Injective, surjective, bijective

Let $f\colon X \to Y$ be a function.

Injectivity. To show f is injective, we must show that the implication

$$f(x) = f(x') \implies x = x'$$

holds for all $x, x' \in$. Frequently it will be convenient to prove the (universal) contrapositive:

$$x \neq x' \implies f(x) \neq f(x')$$

for all $x, x' \in X$.

Similarly, to show f is not injective, we simply have to find $x, x' \in X$ satisfying $x \neq x'$ and $f(x) = f(x')$.

Surjectivity. To prove f is surjective, we must prove the universal quantification:

$$\text{for all } y \in Y, \text{ there exists an } x \in X \text{ such that } f(x) = y.$$

To prove f is *not* surjective, we must prove the negation of this proposition (Remark 1.2.9): i.e., there exists a $y \in Y$ for which there is no $x \in X$ with $f(x) = y$.

Bijectivity. To show f is bijective directly using Definition 1.1.19, we must show that f is injective and surjective. This is equivalent to showing that for $y \in Y$ there exists a *unique* element $x \in X$ such that $f(x) = y$.

Alternatively, using Theorem 1.1.23 we can show that f is bijective by providing an inverse function $f^{-1}\colon Y \to X$.

1.3.5 Mathematical induction

Mathematical induction is a technique for proving universal quantifications of the form

$$\text{For all integers } n \geq b,\ P(n),$$

where b is a fixed starting integer, called the **base**, and P is a predicate defined on the integers. If the setting makes clear that n ranges over integers, we write such propositions using logical notation as

$$\forall n \geq b\, P(n).$$

1.3.5.1 Proof by induction

Suppose P is a predicate of integers. To prove the proposition $\forall n \geq b\, P(n)$ by **induction** (sometimes called **weak induction**), we proceed in two steps.

Base step. Show that $P(b)$ is true.

Induction step. Prove the universal implication

$$\forall n \geq b\, P(n) \implies P(n+1).$$

In practice, if proving the implication $P(n) \implies P(n+1)$ directly, this means we assume $P(n)$ is true (the **induction hypothesis**), and use this assumption to show $P(n+1)$ is true.

Remark 1.3.7 "Step 0" of induction. When meeting a proposition in the wild that we wish to prove by induction, you should first take care to model the proposition in the form

$$\forall n \geq b\, P(n).$$

Make explicit the predicate P in question, as well as the base case b. We illustrate this preparatory "Step 0" in the examples below.

Example 1.3.8 Weak induction. Prove the identity

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

for all $n \geq 1$. Recall:

$$\sum_{k=1}^{n} k = 1 + 2 + \cdots + n.$$

Solution. We prove the proposition by induction.

Step 0: preparation. The proposition is modeled logically as $\forall n \geq 1\, P(n)$, where $P(n)$ is the proposition that

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Base step: $n = 1$. The proposition $P(1)$ is the statement that

$$1 = \frac{1(1+1)}{2},$$

which is clearly true.

Induction step. We must show the universal implication

$$\forall n \geq 1 \, P(n) \implies P(n+1).$$

Let $n \geq 1$, and assume $P(n)$ is true: i.e.,

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

The proposition $P(n+1)$ states that

$$\sum_{k=1}^{n+1} k = \frac{(n+1)(n+2)}{2}.$$

We prove this, assuming $P(n)$, via a chain of equalities:

$$\begin{aligned}
\sum_{k=1}^{n+1} k &= \sum_{k=1}^{n} k + (n+1) \\
&= \frac{n(n+1)}{2} + n + 1 && \text{(induction hyp.)} \\
&= \frac{n(n+1) + 2(n+1)}{2} \\
&= \frac{(n+1)(n+2)}{2},
\end{aligned}$$

as desired. Having completed our base and induction steps, our proof is now finished. □

So why does proof by induction work? In other words, why is it a valid proof technique? Imagine our propositions $P(n)$ as forming an infinite ladder that we wish to ascend. Cautious climbers that we are, we only will step on a rung if we know the corresponding proposition is true. Knowing $P(b)$ is true (the base step) allows us to step onto the first rung. The universal implication $\forall n \geq b \, P(n) \implies P(n+1)$ (induction step) gives us a *rule* that says if rung n is secure (i.e., true), then so is rung $n+1$. Since this rule holds for all rungs (i.e., for all $n \geq b$), we can safely ascend the entire ladder!

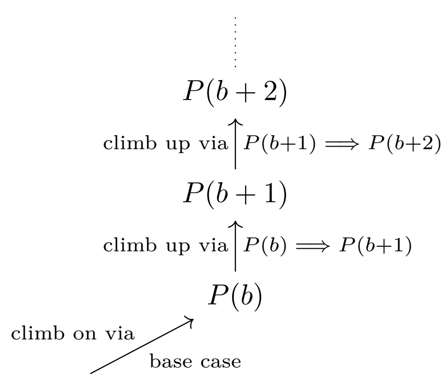

Figure 1.3.9 Mathematical induction as ladder of propositions

1.3.5.2 Proof by strong induction

Suppose P is a predicate of integers. To prove the proposition $\forall n \geq b P(n)$ by **strong induction**, we proceed in two steps.

Base step: $n = b$. Show that $P(b)$ is true.

Strong induction step. Prove the universal implication

$$\forall n \geq b \ (P(b) \wedge P(b+1) \cdots \wedge P(n)) \implies P(n+1).$$

This technique is called strong induction, as now the induction hypothesis is much stronger: to prove this implication directly we assume $P(k)$ is true for all $1 \leq k \leq n$ (not just $k = n$ as in weak induction), and use this assumption to show $P(n+1)$ is true. In fact, strong induction is, logically speaking, no stronger than weak induction. Both techniques apply to propositions of the form $\forall n \geq b\, P(n)$, and you are free to choose which form of induction to use each time. We typically use strong induction out of convenience, when the nature of the predicate P is such that we can prove $P(n+1)$ most elegantly by assuming $P(b), P(b+1), \ldots, P(n)$, as opposed to just $P(n)$. The following example is characteristic in this regard.

Example 1.3.10 Strong induction. Prove that every integer $n \geq 2$ can be written as a product of primes.

Solution. We prove the statement by induction.

Step 0: preparation. The proposition is modeled logically as $\forall n \geq 2\, P(n)$, where $P(n)$ is the proposition that n is a product of primes.

Base step: $n = 2$. The proposition $P(2)$ asserts that 2 is a product of primes. This is true since 2 is itself prime, hence a product of one prime number.

Strong induction step. We must show the universal implication

$$\forall n \geq 2 \ (P(1) \wedge P(2) \wedge \cdots \wedge P(n)) \implies P(n+1).$$

Let $n \geq 2$, and assume $P(k)$ is true for all $2 \leq k \leq n$: i.e., for all such k we assume k can be written as a product of primes. We use this assumption to prove $P(n+1)$: i.e., that $n+1$ is a product of primes. We proceed in two cases, depending on whether $n+1$ is itself prime.

Case 1: $n+1$ is prime. If $n+1$ is prime, then it is trivially a product of one prime number, just as with the base case.

Case2 : $n+1$ is not prime. If $n+1$ is not prime, then we can factor $n+1$ nontrivially as $n+1 = k_1k_2$. Here "nontrivially" means that we have $2 \leq k_1, k_2 \leq n$. Using the strong induction hypothesis, we may assume that k_1 and k_2 are both products of primes: i.e., we have

$$k_1 = p_1p_2 \cdots p_r \qquad\qquad k_2 = q_1q_2 \cdots q_s,$$

where p_i and q_j are prime for all $1 \leq i \leq r$ and $1 \leq j \leq s$. It follows that

$$n+1 = k_1k_2 = p_1p_2 \cdots p_rq_1q_2 \cdots q_s,$$

and hence that $n+1$ is also a product of primes, as desired. Having completed the base and induction steps, our proof by induction is now finished. □

Chapter 2

Systems of linear equations

2.1 Systems of linear equations

We begin in a deceptively simple manner with the study of systems of linear equations and their solutions.

Definition 2.1.1 Linear equations. A **linear expression** in the n unknowns (or variables) $x_1, x_2, \ldots x_n$ is an expression of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n,$$

where $a_1, a_2, \ldots, a_n$ are fixed real numbers.

A **linear equation** in the unknowns $x_1, x_2, \ldots, x_n$ is an equation that can be simplified, using only addition and subtraction, to an equation of the form

$$a_1x_1 + a_2x_2 + \cdots a_nx_n = b, \tag{2.1.1}$$

which we call its **standard form**. An equation in the unknowns $x_1, x_2, \ldots, x_n$ is **nonlinear** if it cannot be simplified to the form (2.1.1) using only addition and subtraction.

Given a linear equation with standard form (2.1.1), the equation is called **homogeneous** if $b = 0$, and **nonhomogeneous** if $b \neq 0$. ◊

Example 2.1.2 Linear and nonlinear equations.

1. Consider $\sqrt{3}x + \sin(5) = 2z - e^4y$. This is a linear equation in the unknowns x, y, z. Its standard form is $\sqrt{3}x + e^4y - 2z = -\sin(5)$. Since the right-hand side is nonzero, we see that the equation is nonhomogeneous.

2. The equation $x^2 + y^2 = 1$ is a *nonlinear* equation in the unknowns x and y.

□

Definition 2.1.3 Systems of linear equations. A **system of linear equations** (or **linear system**) is a set of linear equations.

A **homogeneous** linear system is a set of homogeneous linear equations. ◊

When displaying a system of m equations in the n unknowns $x_1, x_2, \ldots x_n$, we typically write each equation in standard form and align the corresponding terms of each equation into columns:

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

A homogeneous system is thus typically written as:

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & +\cdots+ & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & +\cdots+ & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & +\cdots+ & a_{mn}x_n & = & 0 \end{array}.$$

Remark 2.1.4 You will want to get comfortable with the double-indexing used to display linear systems as quickly as possible. Here is a good way to orient yourself:

- The i appearing in a_{ij} and b_i indicates the i-th row in our displayed system, or equivalently, the i-th equation.
- The j appearing in a_{ij} indicates the j-th column in our displayed system, which is associated to the j-th variable, for $1 \leq j \leq n$.

Definition 2.1.5 Solutions to linear systems. A **solution to a linear equation**

$$a_1x_1 + a_2x_2 + \cdots a_nx_n = b$$

is an ordered sequence $(s_1, s_2, \ldots, s_n)$ of real numbers for which the variable assignment $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ makes the equation true. We say $(s_1, \ldots, s_n)$ **solves the equation** in this case.

A **solution to a system of linear equations**

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

is a sequence $(s_1, s_2, \ldots, s_n)$ that is a solution to each of the system's m equations. We say $(s_1, s_2, \ldots, s_n)$ **solves the system** in this case. ♢

Given a linear system we will seek to find the set of *all* solutions to the system. As we will soon see, this set of solutions will take one of three qualitative forms:

1. The solution set is empty; i.e., there are no solutions. We say the system is **inconsistent** in this case. Otherwise a system is called **consistent**.
2. The solution set contains a single element; i.e., there is exactly one solution.
3. The solution set contains infinitely many elements; i.e., there are infinitely solutions.

Example 2.1.6 Solutions to elementary systems. For each system, compute the set of solutions.

1.

$$\begin{array}{ccccc} x & - & y & = & 0 \\ x & - & y & = & 1 \end{array}$$

2.
$$\begin{array}{rcrcl} x & - & y & = & 0 \\ x & + & y & = & 1 \end{array}$$

3.
$$\begin{array}{rcrcl} x & - & y & = & 1 \\ 2x & - & 2y & = & 2 \end{array}$$

Solution.

1. The first equation implies $x = y$. Substituting y for x in the second equation, we would then have $0 = 1$, a contradiction. Thus there are no solutions: i.e., S is the empty set, denoted $S = \{\ \}$ or $S = \varnothing$.

2. The first equation implies $x = y$. Making this substitution in the second equation yields $2x = 1$, or $x = 1/2$. Thus $(x, y) = (1/2, 1/2)$ is the unique solution, and $S = \{(1/2, 1/2)\}$.

3. The second equation is just twice the first. It follows that both equations have the exact same set of solutions, and so we need only find all solutions to $x - y = 1$. Note that we can set $x = t$ for *any* real number $t \in \mathbb{R}$. Solving for y in terms of t we get $(x, y) = (t, t - 1)$ for any $t \in \mathbb{R}$, and thus $S = \{(t, t - 1) \colon t \in \mathbb{R}\}$, an infinite set!

□

In Example 2.1.6 we were able to combine some simple logic and arithmetic to solve each system completely. (A more geometric approach is discussed in Exercise 2.1.1.) Of course, things get more complicated as the number of equations and/or unknowns in the system increases.

For example, consider the system

$$\begin{array}{rcrcrcl} 2x & - & y & - & z & = & 3 \\ x & & & - & z & = & 2 \\ x & - & y & & & = & 1 \end{array} \tag{2.1.2}$$

Observe first that the sequence $(5, 2, 5)$ is not a solution to the system; one easily checks that is satisfies the first and second equations, but not the third. Similarly, it is easy to verify that $(4, 2, 3)$ and $(0, -1, 2)$ are both solutions to the system. The question remains: How do we find *all* solutions to the linear system in a systematic way?

Remark 2.1.7 Observe the somewhat funny spacing in (2.1.2). This is in force in order to align the unknowns in separate columns. In general, when handed a linear system in the wild, your first step should be to write each equation in standard form, and make sure the unknowns are aligned vertically in this manner.

Some systems are easier to solve than others. Below you find two systems of three equations in three unknowns.

System L	System L'
$2x + 3y + -z = 18$	$x + y + z = 10$
$x + 2y - 2z = 8$	$y - 3z = -2$
$-\frac{1}{2}x + -\frac{1}{2}y + \frac{1}{2}z = -3$	$z = 2$

The staircase pattern of L' allows us to solve easily by *backwards substitution*. In more detail:

1. Equation 3 in L' tells us that $z = 2$.

2. Now substitute $z = 2$ into Equation 2 of L' and solve for y to get $y = 4$.

3. Finally, substitute $y = 4$ and $z = 2$ into Equation 1 of $L'$and solve for x to get $x = 4$. We conclude that $(x, y, z) = (4, 4, 2)$ is a solution to L'.

4. Furthermore, since at each step we had no choice in the matter (z must be equal to 2, y then must be equal to 4, etc.), we see that in fact $(4, 4, 2)$ is the *only* solution to L'.

Our method for solving a more complicated system, like L above, will be to *transform* it to a simpler system like L'.

Key point. In order for this method to work, we need to make sure that the transformed system has *exactly* the same solutions as the original system! To this end we will restrict our permissible transformations to the *elementary operations* define below.

Definition 2.1.8 Elementary operations on linear systems. An **elementary operation** (or **elementary row operation**) is one of the three types of operations on linear systems described below.

Scalar multiplication	Multiply an equation by a *nonzero* number $c \neq 0$: i.e., replace the i-th equation e_i of the system with $c \cdot e_i$ for $c \neq 0$. Note: $c \cdot e_i$ is the result of multiplying the left and right sides of equation e_i by c.
Equation swap	Swap the i-th and j-th equations of the system, $i \neq j$: i.e., replace equation e_i of the system with e_j, and replace equation e_j with e_i.
Equation addition	Add a multiple of one equation to another: i.e., replace e_i with $e_i + ce_j$ for some c, i, and j.

The act of transforming a system of equations using elementary operations is called **reduction** (or **row reduction**). ◊

After reducing a linear system L using elementary operations, we are left with a new system L'. The systems L and L' will generally look very different from one another. Note, however, that the two systems will have the same *number* of equations, and the same *number* of variables. More importantly, the two systems will have identical sets of solutions; that is, the set of solutions of the new system L' is identical to the set of solutions to L!

To convince ourselves of this last assertion, it suffices to show that the application of any one of these elementary row operations produces a new system L' with exactly the same set of solutions as L; if this is so, then performing any finite sequence of elementary row operations must also preserve the set of solutions in this sense. Theorem 2.1.12 will make these claims official. For now let's look at an example.

Complete example. Consider again the linear system

$$\begin{array}{rcrcrcr} 2x & + & 3y & + & -z & = & 18 \\ x & + & 2y & - & 2z & = & 8 \\ -\frac{1}{2}x & + & -\frac{1}{2}y & + & \frac{1}{2}z & = & -3 \end{array}$$

We transform the system as as follows:

$$\begin{array}{rcrcrcr} 2x & + & 3y & + & -z & = & 18 \\ x & + & 2y & - & 2z & = & 8 \\ -\frac{1}{2}x & + & -\frac{1}{2}y & + & \frac{1}{2}z & = & -3 \end{array} \xrightarrow{2e_3} \begin{array}{rcrcrcr} 2x & + & 3y & + & -z & = & 18 \\ x & + & 2y & - & 2z & = & 8 \\ -x & + & -y & + & z & = & -6 \end{array}$$

$$\xrightarrow{e_1 - e_2} \begin{array}{rcrcrcr} x & + & y & + & z & = & 10 \\ x & + & 2y & - & 2z & = & 8 \\ -x & + & -y & + & z & = & -6 \end{array}$$

$$\xrightarrow{e_2 - e_1} \begin{array}{rcrcrcr} x & + & y & + & z & = & 10 \\ & & y & - & 3z & = & -2 \\ & & & & 2z & = & 4 \end{array}$$

$$\xrightarrow{e_3 + e_1} \begin{array}{rcrcrcr} x & + & y & + & z & = & 10 \\ & & y & - & 3z & = & -2 \\ & & & & 2z & = & 4 \end{array}$$

$$\xrightarrow{\frac{1}{2}e_3} \begin{array}{rcrcrcr} x & + & y & + & z & = & 10 \\ & & y & - & 3z & = & -2 \\ & & & & z & = & 2 \end{array}$$

Now put the logic together. Our original linear system L was transformed by a sequence of elementary row operations to a new system L':

$$\begin{array}{rcrcrcr} 2x & + & 3y & + & -z & = & 18 \\ x & + & 2y & - & 2z & = & 8 \\ -x & + & -y & + & z & = & -6 \end{array} \xrightarrow{\text{row op.'s}} \begin{array}{rcrcrcr} x & + & y & + & z & = & 10 \\ & & y & - & 3z & = & -2 \\ & & & & z & = & 2 \end{array}$$

We saw already that the second system has exactly one solution, namely the triple $(x, y, z) = (4, 4, 2)$.

Since transforming a system by row operations preserves solutions, the first and second systems have *exactly the same solutions*.

Thus $(x, y, z) = (4, 4, 2)$ is the only solution to the original system!

Remark 2.1.9 Notation. As we will see later, keeping track of the exact sequence of row operations is important in some situations. The notation used in the last example is useful for this bookkeeping. Let's explicate the notation somewhat.

The notation

$$L \xrightarrow{c\,e_i} L'$$

indicates that system L' is obtained from L by replacing equation e_i with $c\,e_i$.

The notation

$$L \xrightarrow{e_i \leftrightarrow e_j} L'$$

indicates that system L' is obtained from L by swapping rows e_i and e_j.

The notation

$$L \xrightarrow{e_i + c\,e_j} L'$$

indicates that system L' is obtained from L by replacing equation e_i with $e_i + c\,e_i$.

Remark 2.1.10 Mandate. You may be tempted to do multiple operations in a single step during this procedure. Resist this temptation for now until we have a better understanding of things.

Furthermore, if later on you do give in to this temptation, make sure that the two (or more row operations) you perform are independent of one another. For example, do not swap e_2 with e_1 and replace e_3 with $e_3 - 5e_2$ in the same step.

We end by making official a claim made earlier. First we introduce the notion of *equivalent linear systems*, mainly to spare ourselves from the mouthful that is "obtained by applying a finite sequence of elementary row operations".

Definition 2.1.11 Equivalence of linear systems. We say two systems of linear equations are **equivalent** (or **row equivalent**) if the one can be obtained from the other by a finite sequence of elementary operations. $\Diamond$

Theorem 2.1.12 Row equivalence theorem. *Equivalent systems of linear equations have identical sets of solutions.*

Proof. We prove by induction that if system L' is the result of applying n elementary operations to system L, $n \geq 0$, then L and L' have the same set of solutions.

Base case: $n = 0$. In this case $L = L'$ (we have applied no operations) and the statement is obvious.

Induction step. Assume that applying a sequence of n elementary operations to a linear system preserves the set of solutions.

Suppose L' is the result of applying $n + 1$ elementary operations to the system L. Let L'' be the result of applying the first n of these operations. By the induction hypothesis, systems L and L'' have the same set of solutions.

Since the system L' is obtained from L'' by applying exactly one elementary operation, Exercise 2.1.2 implies L' and L'' have the same set of solutions.

We conclude that L and L' have the same set of solutions, as desired. $\blacksquare$

Exercises

1. Geometry of linear systems. Recall that the *graph* of an equation is the set of all solutions to the equation. Graphs of linear equations correspond to some familiar geometric objects:

- *Lines.*

  A *line* in $\mathbb{R}^2$ is the graph ℓ of a linear equation of the form $ax+by = c$, where $a, b, c \in \mathbb{R}$ are fixed constants, and at least one of a and b is nonzero: i.e.,

  $$\ell = \{(x, y) \in \mathbb{R}^2 \colon ax + by = c\}.$$

- *Planes.*

  A *plane* in $\mathbb{R}^3$ is the graph $\mathcal{P}$ of a linear equation of the form $ax + by + cz = d$, where $a, b, c, d \in \mathbb{R}$ are fixed constants, and at least one of a, b, c is nonzero: i.e.,

  $$\mathcal{P} = \{(x, y, z) \in \mathbb{R}^3 \colon ax + by + cz = d\}.$$

We often use the abbreviated notation

$$\ell \colon ax + by = c \qquad\qquad \mathcal{P} \colon ax + by + cz = d$$

to introduce the line ℓ with defining equation $ax + by = c$ or plane $\mathcal{P}$ with definining equation $ax + by + cz = d$.

(a) Fix $m > 1$ and consider a system of m linear equations in the two unknowns x and y. What does a solution (x, y) to this *system* of linear equations correspond to geometrically?

(b) Use your interpretation in (a) to give a *geometric* argument that a system of m equations in two unknowns will have either (i) zero solutions, (ii) exactly one solution, or (iii) infinitely many solutions.

(c) Use your geometric interpretation to help produce explicit examples of systems in two variables satisfying these three different cases (i)-(iii).

(d) Now repeat (a)-(b) for a system of linear equations in three variables x, y, z.

Solution. (a) Geometrically, each equation in the system represents a line $\ell_i\colon a_ix + b_iy = c_i$. A solution (x, y) to the i-th equation corresponds to a point on ℓ_i. Thus a solution (x, y) to the system corresponds to a point lying on *all* of the lines: i.e., a point of intersection of the lines.

(b) First of all to prove the desired "or" statement it suffices to prove that if the number of solutions is greater than 1, then there are infinitely many solutions.

Now suppose there is more than one solution. Then there are at least two different solutions: $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$. Take any of the two lines ℓ_i, ℓ_j. By above the intersection of ℓ_i and ℓ_j contains P_1 and P_2. But two *distinct* lines intersect in at most one point. It follows that ℓ_i and ℓ_j must be equal. Since ℓ_i and ℓ_j were arbitrary, it follows *all* of the lines ℓ_i are in fact the same line ℓ. But this means the common intersection of the lines is ℓ, which has infinitely many points. It follows that the system has infinitely many solutions.

(c) We will get 0 solutions if the system includes two different parallel lines: e.g., $\ell_1\colon x + y = 5$ and $\ell_2\colon x + y = 1$.

We will get exactly one solution when the slopes of each line in the system are distinct.

We will get infinitely many solutions when *all* equations in the system represent the *same line*. This happens when all equations are multiples of one another.

(d) Now each equation in our system defines a plane $\mathcal{P}_i\colon a_ix + b_iy + c_iz = d_i$. A solution (x, y, z) to the system corresponds to a point $P = (x, y, z)$ of intersection of the planes. We recall two facts from Euclidean geometry:

(a) *Fact 1.*

Given two distinct points, there is a unique line containing both of them.

(b) *Fact 2.*

Given any number of distinct planes, they either do not intersect, or intersect in a line.

We proceed as in part (b) above: that is show that if there are two distinct solutions to the system, then there are infinitely many solutions. First, for simplicity, we may assume that the equations $\mathcal{P}_i\colon a_ix+b_iy+c_iz = d_i$ define *distinct* planes; if we have two equations defining the same plane, we can delete one of them and not change the set of solutions to the system.

Now suppose $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$ are two distinct solutions to the system. Let ℓ be the unique line containing P and Q (Fact 1). I claim that ℓ is precisely the set of solutions to the system. To see this, take any two equations in the system $\mathcal{P}_i\colon a_ix + b_iy + c_iz = d_i$ and $\mathcal{P}_j\colon a_jx + b_jy + c_iz = d_j$. Since the two corresponding planes are distinct, and intersect in at least the points P and Q, they must intersect in a line (Fact 2); since this line contains P and Q, it must be the line ℓ (Fact 1).

Thus any two planes in the system intersect in the line ℓ. From this it follows that: (a) a point satisfying the system must lie in ℓ; and (b) all points on ℓ satisfy the system (since we have shown that ℓ lies in all the planes). It follows that ℓ is precisely the set of solutions, and hence that there are infinitely many solutions.

2. **Row operations preserve solutions.** We made the claim that each of our three row operations

(a) scalar multiplication ($e_i \mapsto c \cdot e_i$ for $c \neq 0$),

(b) swap ($e_i \leftrightarrow e_j$),

(c) addition ($e_i \mapsto e_i + c \cdot e_j$ for some c)

do not change the set of solutions of a linear system. To prove this claim, let L be a general linear system

$$\begin{array}{rccccccc} e_1: & a_{11}x_1 & + & a_{12}x_2 & +\cdots+ & a_{1n}x_n & = & b_1 \\ e_2: & a_{21}x_1 & + & a_{22}x_2 & +\cdots+ & a_{2n}x_n & = & b_2 \\ & \vdots & & \vdots & & \vdots & \vdots & \\ e_m: & a_{m1}x_1 & + & a_{m2}x_2 & +\cdots+ & a_{mn}x_n & = & b_m \end{array}.$$

Now consider each type of row operation separately, write down the new system L' you get by applying this row operation, and prove that an n-tuple $s = (s_1, s_2, \ldots, s_n)$ is a solution to the original system L if and only if it is a solution to the new system L'.

Solution. Let L be the original system with equations $e_1, e_2, \ldots, e_m$. For each specified row operation, we will call the resulting new system L' and its equations $e'_1, e'_2, \ldots, e'_m$.

Row swap. In this case systems L and L' have exactly the same equations, just written in a different order. Thus the n-tuple s satisfies L if and only if it satisfies each of the equations e_i, if and only if it satisfies each of the equations e'_i, since these are the same equations! It follows that s is a solution of L if and only if it is a solution to L'.

Scalar multiplication. In this case $e_j = e'_j$ for all $j \neq i$, and $e'_i = c \cdot e_i$ for some $c \neq 0$. Since only the i-th equation has changed, it suffices to show that s is a solution to e_i if and only if s is a solution to $c \cdot e_i$. Let's prove each direction of this if and only if separately.

If s satisfies e_1, then $a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n = b_i$. Multiplying both sides by c we see that

$$ca_{i1}s_1 + ca_{i2}s_2 + \cdots + ca_{in}s_n = cb_i,$$

and hence that s is also a solution of $c\,e_i = e'_i$.

For the other direction, if s satisfies $ce_i = e'_i$, then

$$ca_{ii}s_1 + ca_{i2}s_2 + \cdots ca_{in}s_n = cb_i.$$

Now, since $c \neq 0$, we can multiply both sides of this equation by $1/c$ to see that

$$a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n = b_i$$

and hence that s is a solution to e_i.

Row addition. The only equation of L' that differs from L is

$$e'_i = e_i + ce_j.$$

Writing this equation out in terms of coefficients gives us

$$e_i' : a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + c(a_{j1}x_1 + a_{j2}x_2 + \cdots + a_{jn}x_n) = b_i + cb_j.$$

Now if s satisfies L, then it satisfies e_i and e_j, in which case evaluating the RHS of the e_i' above at s yields

$$a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n + c(a_{j1}s_1 + a_{j2}s_2 + \cdots + a_{jn}s_n) = b_i + cb_j$$

showing that s satisfies e_i'. Now suppose $s = (s_1, s_2, \ldots, s_n)$ satisfies L'. Since s satisfies $e_j' = e_j$, we have

$$a_{j1}s_1 + a_{j2}s_2 + \cdots + a_{jn}s_n = b_j \tag{2.1.3}$$

Since s satisfies e_i', we have

$$a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n + c(a_{j1}s_1 + a_{j2}s_2 + \cdots + a_{jn}s_n) = b_i + cb_j$$

Substituting (2.1.3) into the equation above we get

$$a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n + c(b_j) = b_i + cb_j,$$

and hence

$$a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n = b_i.$$

This shows that s satisfies e_i. It follows that s satisfies L.

3. Nonlinear systems. A *nonlinear* system of equations is a collection of equations, at least one of which is nonlinear. Our definition of a solution to a linear system generalizes easily to any system of equations.

(a) Consider the following nonlinear system in the unknowns x, y:

$$\begin{array}{ccccc} x^2 & + & y^2 & = & 1 \\ x & + & y & = & \frac{1}{2} \end{array}.$$

i. Sketch the graphs of each of the two equations in the system on a common coordinate system.

ii. Describe geometrically what a solution to the system is in terms of your sketch. Explain your reasoning. How many solutions to the system are there, according to your sketch.

iii. Compute the set of all solutions to the system algebraically.

(b) Now consider a more general system

$$\begin{array}{ccccc} x^2 & + & y^2 & = & 1 \\ ax & + & by & = & c, \end{array} \tag{2.1.4}$$

where $a, b, c \in \mathbb{R}$ are fixed constants and at least one of a or b is nonzero.

i. Explain geometrically what a solution to the system corresponds to in terms of the graphs of its two equations.

ii. Use your geometric interpretation in (i) to argue that the system (2.1.4) has either 0, 1, or 2 solutions. Give explicit examples of such systems corresponding to each of these three cases.

4. Not all arithmetic operations preserve solutions. In this exercise we investigate how the operation of squaring both sides of an equation changes the set of solutions. Let

$$F(x_1, x_2, \ldots, x_n) = G(x_1, x_2, \ldots, x_n) \tag{2.1.5}$$

represent a general equation (linear or nonlinear) in the unknowns $x_1, x_2, \ldots, x_n$, let

$$(F(x_1, x_2, \ldots, x_n))^2 = (G(x_1, x_2, \ldots, x_n))^2 \tag{2.1.6}$$

be the equation obtained by squaring both sides of the (2.1.5), let S_1 be the set of solutions to (2.1.5), and let S_2 be the set of solutions to (2.1.6).

(a) Show that $S_1 \subseteq S_2$.

(b) Given an explicit example of an equation of the form (2.1.5) in two variables for which $S_2 \not\subseteq S_1$.

2.2 Gaussian elimination

In Section 2.1 we sketched a procedure for solving a linear system L. That procedure involved applying a sequence of row operations to L to obtain a "simpler" system L'.

We will fill out this sketch in the next two sections. Specifically, we will

1. describe precisely what we mean by a "simpler" system,
2. provide an algorithm (or recipe) that decides exactly what sequence of row operations to apply to obtain this simpler system,
3. explain how to find *all* solutions of the resulting simpler system.

Before we embark on these steps, we introduce a useful bit of notation. As you may have noticed, when performing row operations on a system of equations, we essentially treat the unknowns, as well as the plus and equals symbols, as placeholders; the only things that actually change in a given step are the coefficients in the equations. The *augmented matrix associated to a linear system* is a formal way of extracting just the information of the coefficients from a linear system.

Definition 2.2.1 Augmented matrix. Let L be the linear system

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

The **augmented matrix associated to L** is the matrix

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

◊

Remark 2.2.2 As defined more thoroughly in Definition **??**, a matrix is just a rectangular array of numbers.

2.2.1 Row echelon matrices

Here is our precise formulation of a "simple" linear system; it is a system whose associated augmented matrix is in *row echelon form*, as described below.

Definition 2.2.3 Row echelon form. A **zero row** of a matrix, is a row whose entries are all equal to zero; a **nonzero row** is a row that contains at least one nonzero entry.

A matrix is in **row echelon form** if the following conditions hold.

- **(i)** In any nonzero row the first (i.e., leftmost) nonzero entry is equal to one. A **leading one** of a matrix is such an entry: i.e., an entry of a row that is equal to one, and that is also the first nonzero entry of that row.
- **(ii)** All zero rows are grouped together at the bottom of the matrix.
- **(iii)** Given any two nonzero rows in the matrix, the leading one of the lower row occurs strictly to the right of the leading one of the row above it.

A matrix is in **reduced row echelon form** if in addition to conditions (i)-(iii) it also satisfies the following condition.

- **(iv)** Any column of the matrix that contains a leading one has zeros elsewhere. In other words, if a column contains a leading one, then that is the only nonzero entry of that column.

A linear system L is in **row echelon form** (resp. **reduced row echelon form**) if its augmented matrix is in row echelon form (resp. reduced row echelon form). ♢

In practice to decide whether a matrix is in in (reduced) row echelon form, follow these steps:

1. First verify whether all zero rows are at the bottom.
2. For each nonzero row, determine whether the first nonzero entry is a 1, and put a box around it.
3. Make sure your boxes make a staircase pattern.
4. (For reduced row echelon form only.) Decide whether each column with a box has 0's everywhere else.

Example 2.2.4 Row echelon versus reduced row echelon form. For each matrix decide (a) whether it is in row echelon form, and (b) whether it is in reduced row echelon form.

1.

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

2.

$$\begin{bmatrix} 1 & 0 & 0 & -3 & -7 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Solution.

1. Below you find the matrix with leading ones boxed. This matrix fails to be in row echelon form (and hence also reduced row echelon form) for a variety of reasons: the zero rows are not all grouped at the bottom; the first row is nonzero, but does not have a leading one; the leading one of the fourth row is to the left of the leading one of the leading one in the row above it.

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

2. Below you find the matrix with leading ones boxed. This matrix is in row echelon form: the zero rows (rows 4 and 5) are grouped at the bottom; each nonzero row has a leading one (boxed in the matrix below); the leading ones step strictly to the right as we move down the rows.

$$\begin{bmatrix} \boxed{1} & 0 & 0 & -3 & -7 \\ 0 & 0 & \boxed{1} & 2 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The matrix is *not* in reduced row echelon form, as the last column contains a leading one in its third row, and a nonzero entry in its first row.

□

2.2.2 Gaussian elimination

We will now describe a systematic procedure, called *Gaussian elimination*, that allows us to reduce a given linear system L to a system L'in row echelon form. In keeping with the foregoing discussion, we will identify a system L with its augmented matrix A. Furthermore, reducing a linear system using elementary operations on equations is now cast as performing *elementary row operations* on matrices. At the risk of redundancy we now officially translate a number of our former notions regarding reduction of linear systems to the setting of matrices.

Definition 2.2.5 Elementary row operations on matrices. An **elementary row operation** is one of the three following types of matrix operations. Let A be a given $m \times n$ matrix, and denote by r_i the i-th row of A.

Scalar multplication Multiply a row by a *nonzero* number $c \neq 0$: i.e., replace r_i with $c\,r_i$, the result of multiplying all entries of the row by c.

Row swap Swap two rows of A.

Row addition Add a multiple of one row to another: i.e., replace r_i with $r_i + cr_j$ for some c, i, and j.

The act of transforming a matrix using elementary row operations is called **row reduction**

Two matrices are **row equivalent** if the one can be obtained from the other by performing a finite sequence of elementary row operations. ◊

Remark 2.2.6 Notation. Our former elementary operation notation easily transfers to row operations on matrices. The expressions

$$A \xrightarrow{c\,r_i} B \qquad A \xrightarrow{r_i \leftrightarrow r_j} B \qquad A \xrightarrow{r_i + c\,r_j} B$$

denote the operations that replace row r_i in A with $c\,r_i$, swap rows r_i and r_j in A, and replace r_i in A with $r_i + c\,r_j$, respectively.

At last we are ready to define Gaussian elimination. This is a procedure, or *algorithm*, that takes an input matrix A and row reduces it to a matrix B in row echelon form.

Mantra 2.2.7 Gaussian elimination is the workhorse of linear algebra. *Simplifying linear systems is but one of the many useful applications of Gaussian elimination, and you should think of it as having a life independent of its role in solving systems of linear equations. In its essence Gaussian elimination is simply a certain procedure performed on matrices: one that we will come back to again and again, and to diverse ends.*

Definition 2.2.8 Gaussian elimination. Gaussian elimination is the algorithm described below. It takes as an input a matrix A and returns as an output a row equivalent matrix B in row echelon form.

Step 1 Find the leftmost nonzero column and perform a row swap to move the row with this nonzero entry to the top of the matrix.

Step 2 Scale the new top row to produce a leading one in the row. Call this new row r.

Step 3 For each row r_i below r perform a row operation of the form $r_i + c\,r$ to replace all entries below the leading one of r with zeros.

Step 4 Begin again with Step 1 applied to the matrix consisting of all rows below r. Continue until the matrix is in row echelon form.

◊

2.2.3 Model example

Use the following example as a model of how to both perform and annotate the steps in Gaussian elimination. When first learning this procedure, make sure to follow it *to the letter*. In particular, in the spirit of the mandate issued in Remark 2.1.10, you should only perform one row operation at a time, and only in the sequence prescribed in Steps 1-4 of Definition 2.2.8.

$$\begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix} \xrightarrow{r_1 \leftrightarrow r_2} \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{2}r_1} \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

$$\xrightarrow{r_3-2r_1} \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{bmatrix} \qquad \text{(now done with first row)}$$

$$\xrightarrow{-\frac{1}{2}r_2} \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -7/2 & -6 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{bmatrix}$$

$$\xrightarrow{r_3+(-5)r_2} \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -7/2 & -6 \\ 0 & 0 & 0 & 0 & 1/2 & 1 \end{bmatrix} \qquad \text{(now done with 2nd row)}$$

$$\xrightarrow{2r_3} \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -7/2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

The matrix outputed by Gaussian elimination is guaranteed to be in row echelon form, but it may not be in *reduced* row echelon form. *Gauss-Jordan elimination* describes a systematic way to continue reducing to this even simpler state.

Definition 2.2.9 Gauss-Jordan elimination. Gauss-Jordan elimination is the algorithm described below. It takes as an input a matrix A and returns a row equivalent matrix B in reduced row echelon form.

Steps 1-4 Apply Gaussian elimination to transform A to a matrix in row echelon form.

Step 5 Find the rightmost column of the matrix containing a leading one. Let r_i be the row containing this leading one. For each row r_j above r_i perform a row operation of the form $r_i + c\,r_j$ to replace all entries above the leading one with zeros.

Step 6 Begin again with Step 5 applied to the next column to the left that contains a leading one. Continue until the matrix is in reduced row echelon form.

◊

Continuing our previous example:

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -7/2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}. \xrightarrow{r_2+7/2r_3} \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

$$\xrightarrow{r_1-6r_3} \begin{bmatrix} 1 & 2 & -5 & 3 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

$$\xrightarrow{r_1+5r_2} \begin{bmatrix} 1 & 2 & 0 & 3 & 0 & 7 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

Definition 2.2.8 and Definition 2.2.9 are really theorems in disguise, and we make this official in Theorem 2.2.10.

Theorem 2.2.10 Row equivalent matrix forms.

Row echelon forms exist *Any matrix A is row equivalent to a matrix in row echelon form. Indeed, Gaussian elimination row reduces any matrix to a matrix in row echelon form.*

Reduced row echelon forms exist *Any matrix A is row equivalent to a matrix in reduced row echelon form. Indeed, Gauss-Jordan elimination row reduces any matrix to a matrix in reduced row echelon form.*

Reduced row echelon forms are unique *Any matrix A is row equivalent to a* unique *matrix in reduced row echelon form.*

We will make heavy use of the first two results of Theorem 2.2.10. The proofs of these statements are not difficult, but not especially illuminating. Accordingly we omit them here, and point the interested reader to Robert Beezer's *A First Course in Linear Algebra* (Theorem REMEF).

The third statement of Theorem 2.2.10, that every matrix is row equivalent to a *unique* matrix in reduced row echelon form, does in fact have an enlightening proof. We will postpone this proof, however, until we have a little more theory at our disposal. (See Corollary **??**.) Until then we will conscientiously *not* make use of this result in developing any of our further theory.

Example 2.2.11 Row echelon form is not unique. Show that a matrix A may be row equivalent to two or more matrices in row echelon form.

Solution. Take $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. This row reduces to $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ using Gaussian elimination; and it row reduces further to $C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ using Gauss-Jordan elimination. Thus we see that A is row equivalent to two different matrices in row echelon form. (According to Theorem 2.2.10, the matrix C is the only matrix in *reduced* row echelon that is row equivalent to A.) □

2.2.4 Exercises

Exercise Group. Explain why each of the following matrices fails to be in row echelon form.

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

Solution. The first nonzero term in the second row is not a one.

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

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

Exercise Group. For each of the given linear systems, find an equivalent system in row echelon form. Use augmented matrices and follow the Gaussian elimination algorithm to the letter.

4.

$$\begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & = & x_3 & + & x_4 & + & 3 \\ 3x_1 & + & 6x_2 & = & 2x_3 & - & 4x_4 & + & 8 \\ -x_1 & + & 2x_3 & = & 2x_2 & - & x_4 & - & 1 \end{array}$$

Solution. First bring the system into standard form:

$$L\colon \begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & - & x_3 & - & x_4 & = & 3 \\ 3x_1 & + & 6x_2 & - & 2x_3 & + & 4x_4 & = & 8. \\ -x_1 & - & 2x_2 & + & 2x_3 & + & x_4 & = & -1 \end{array}$$

Then perform Gaussian elimination on the associated augmented matrix:

$$\begin{aligned} \begin{bmatrix} 1 & 2 & -1 & -1 & 3 \\ 3 & 6 & -2 & 4 & 8 \\ -1 & -2 & 2 & 1 & -1 \end{bmatrix} &\xrightarrow{r_2-3r_1} \begin{bmatrix} 1 & 2 & -1 & -1 & 3 \\ 0 & 0 & 1 & 7 & -1 \\ -1 & -2 & 2 & 1 & -1 \end{bmatrix} \\ &\xrightarrow{r_3+r_1} \begin{bmatrix} 1 & 2 & -1 & -1 & 3 \\ 0 & 0 & 1 & 7 & -1 \\ 0 & 0 & 1 & 0 & 2 \end{bmatrix} \\ &\xrightarrow{r_3-r_2} \begin{bmatrix} 1 & 2 & -1 & -1 & 3 \\ 0 & 0 & 1 & 7 & -1 \\ 0 & 0 & 0 & -7 & 3 \end{bmatrix} \\ &\xrightarrow{-\frac{1}{7}r_3} \begin{bmatrix} 1 & 2 & -1 & -1 & 3 \\ 0 & 0 & 1 & 7 & -1 \\ 0 & 0 & 0 & 1 & -\frac{3}{7} \end{bmatrix}. \end{aligned}$$

The corresponding equivalent system is

$$L'\colon \begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & - & x_3 & - & x_4 & = & 3 \\ & & & & x_3 & + & 7x_4 & = & -1. \\ & & & & & & x_4 & = & -\frac{3}{7} \end{array}$$

5.

$$\begin{array}{rcrcrcrcr} x_1 & + & x_2 & - & x_3 & + & x_4 & = & 1 \\ -2x_1 & - & 2x_2 & + & 2x_3 & - & 2x_4 & = & -2 \\ x_1 & + & x_2 & + & x_3 & + & 2x_4 & = & 3 \end{array}$$

6.

$$\begin{array}{rcrcrcr} 2x_1 & + & 2x_2 & + & 2x_3 & = & 0 \\ -2x_1 & + & 5x_2 & + & 2x_3 & = & 1 \\ 8x_1 & + & x_2 & + & 4x_3 & = & -1 \end{array}$$

7.

$$\begin{array}{rcrcrcr} & & -2b & + & 3c & = & 1 \\ 3a & + & 6b & - & 3c & = & -2 \\ 6a & + & 6b & + & 3c & = & 5 \end{array}$$

8.

$$\begin{array}{rcrcrcrcrcr} & & & & T_3 & + & T_4 & + & T_5 & = & 0 \\ -T_1 & - & T_2 & + & 2T_3 & - & 3T_4 & + & T_5 & = & 0 \\ T_1 & + & T_2 & - & 2T_3 & & & - & T_5 & = & 0 \\ 2T_1 & + & 2T_2 & - & T_3 & & & + & T_5 & = & 0 \end{array}$$

2.3 Solving linear systems

Let's continue with our model example from Section 2.2. Summarizing the various steps, we have

$$\begin{array}{rcl} -2x_3+7x_5 &=& 12 \\ 2x_1+4x_2-10x_3+6x_4+12x_5 &=& 28 \\ 2x_1+4x_2-5x_3+6x_4-5x_5 &=& -1 \end{array} \xrightarrow{\text{aug. mat.}} \begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

$$\xrightarrow{\text{Gauss. elim.}} \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

$$\xrightarrow{\text{new system}} \begin{array}{rcl} x_1+2x_2-5x_3+3x_4+6x_5 &=& 14 \\ x_3-\frac{7}{2}x_5 &=& -6 \\ x_5 &=& 2 \end{array}.$$

The new system in row echelon form is undoubtedly simpler, but describing *all* of its solutions still requires some subtle analysis.

2.3.1 Model example

We begin by illustrating the general method for solving a linear system using our model example; a careful description of the procedure, along with a proof of its validity, is found in Theorem 2.3.5.

A key first step involves separating the variables of the system into *free* and *leading* variables.

Definition 2.3.1 Free and leading variables. Let L be a linear system in the unknowns $x_1, x_2, \ldots, x_n$, and let A be its associated augmented matrix. Assume L (and hence A) is in row echelon form.

The unknown x_j is a **leading variable** if the corresponding column in A (i.e., the i-th column) contains a leading one; it is a **free variable** if the corresponding column in A does not contain a leading one. ◊

Example 2.3.2 Let L be the linear system in the unknowns x_1, x_2, x_3, x_4 with augmented matrix

$$A = \left[\begin{array}{cccc|c} \boxed{1} & 1 & 3 & 2 & 4 \\ 0 & \boxed{1} & 1 & 2 & 3 \\ 0 & 0 & 0 & \boxed{1} & -5 \end{array}\right].$$

Then x_1, x_2, x_4 are leading variables, since the first, second, and fourth columns of A have leading ones, as indicated by the boxes. The variable x_3 is free since the third column of A has no leading one. □

In our model example we transformed the original system L to the equivalent system L' below:

$$\begin{array}{rcrcrcrcrcr} x_1 &+& 2x_2 &-& 5x_3 &+& 3x_4 &+& 6x_5 &=& 14 \\ &&&& x_3 &-& && \frac{7}{2}x_5 &=& -6. \\ &&&&&&&& x_5 &=& 2 \end{array}$$

The free variables of L' are x_2 and x_4; the leading variables are x_1, x_3, and x_5. Observe that if we assign $x_2 = s$ and $x_4 = t$, where r and s are any real numbers, then we are left with a system L'' in three unknowns (x_1, x_3, x_5) of the form

$$\begin{array}{rcrcrcr} x_1 &-& 5x_3 &+& 6x_5 &=& 14-2s-3t \\ && x_3 &-& \frac{7}{2}x_5 &=& -6. \\ &&&& x_5 &=& 2 \end{array}$$

Using back-substitution, we see that the unknowns x_1, x_3, x_5 are then uniquely expressed in terms of s and t as

$$x_5 = 2, x_3 = 1, x_1 = 7 - 2s - 3t. \tag{2.3.1}$$

Thus for any choice of real numbers s and t we get a unique solution of the form

$$(x_1, x_2, x_3, x_4, x_5) = (7 - 2s - 3t, s, 1, t, 2).$$

We conclude the set S of solutions to L is given as

$$S = \{(7 - 2s - 3t, s, 1, t, 2) \colon s, t \in \mathbb{R}\}. \tag{2.3.2}$$

Alternatively, we can describe the solutions to L with the *parametric equations*

$$x_1 = 7 - 2s - 3t, x_2 = s, x_3 = 1, t = t, x_5 = 2, \ t, s \in \mathbb{R}. \tag{2.3.3}$$

Remark 2.3.3 Mandate. Get used to describing solutions to linear systems using either the set notation format of (2.3.2) or the parametric equation format of (2.3.3).

Note also the distinct roles played by free and leading variables in the description of solutions. We assign each free variable *freely* to any choice of real parameters (s and t in our example), and then solve for the leading variables in terms of these parameters in a unique manner. In particular, the leading variables are completely determined by our assignment of free variables.

2.3.2 General method for solving linear systems

Before describing a precise algorithm for computing the set of solutions to a linear system, we must address the possibility that there are no solutions to the system whatsoever. Such a system is called *inconsistent*.

Definition 2.3.4 Consistent and inconsistent systems. A linear system is **consistent** if it has at least one solution; it is **inconsistent** if it has no solutions. $\Diamond$

We are now in a position to describe an algorithm for computing the set of solutions of a linear system.

Theorem 2.3.5 Solving linear systems. *Let L be a linear system in the unknowns $x_1, x_2, \ldots, x_n$, and let S be the set of all solutions of L. We compute S as follows.*

Step 1 *Apply Gaussian elimination to reduce L to an equivalent system L' in row echelon form.*

Step 2 *Let U be the augmented matrix associated to L'. If the last column of U has a leading one, then L is inconsistent: i.e., $S = \{\}$ is the empty set. Otherwise, proceed to the next step.*

Step 3 *Determine which if any of the unknowns are free variables of L'.*

Step 4 *If there are no free variables, solve for each unknown using back-substitution. In this case, there is a unique solution to L: i.e.,* $S = \{(s_1, s_2, \ldots, s_n)\}$ *contains exactly one element.*

Otherwise, let $x_{i_1}, x_{i_2}, \ldots, x_{i_r}$ *be the leading variables of* L'*, and let* $x_{j_1}, x_{j_2}, \ldots, x_{j_s}$ *be the free variables. Back-substitution allows us to express each leading variable in terms of the free variables. In other words, we can write*

$$\begin{aligned} x_{i_1} &= F_1(x_{j_1}, x_{j_2}, \ldots, x_{j_s}) \\ x_{i_2} &= F_2(x_{j_1}, x_{j_2}, \ldots, x_{j_s}) \\ &\vdots \\ x_{i_r} &= F_s(x_{j_1}, x_{j_2}, \ldots, x_{j_s}), \end{aligned}$$

where each $F_i(x_{j_1}, x_{j_2}, \ldots, x_{j_s})$ *is a linear expression in the free variables. Each solution of L thus corresponds to a* unique *variable assignment of the form*

$$\begin{aligned} \textit{Free variables: } & x_{j_1} = t_1,\ x_{j_2} = t_2, \ldots, x_{j_s} = t_s \\ \textit{Leading variables: } & x_{i_1} = F_1(t_1, t_2, \ldots, t_s) \\ & x_{i_2} = F_2(t_1, t_2, \ldots, t_s) \\ & \vdots \\ & x_{i_r} = F_s(t_1, t_2, \ldots, t_s), \end{aligned}$$

where $t_1, t_2, \ldots, t_s$ *are any real numbers.*

Proof. First recall that L and L' have the same set of solutions (Theorem 2.1.12). So it suffices to show that the algorithm returns the correct set of solutions to L'.

Regarding consistency: if the last column of the augmented matrix U associated to L' has a leading one in the i-th row, then the corresponding equation in L' is

$$0x_1 + 0x_2 + \cdots 0x_n = 1.$$

Clearly this equation has no solutions, and hence the set of solutions to L' is empty.

Suppose now that the last column of U does not have a leading one.

Case 1: no free variables. Suppose in Step 3 you determine that there are no free variables. Then each of the first n columns of U has a leading one in it. If follows that for each $1 \leq i \leq n$ the i-th equation of L' is of the form

$$x_i + c_{i,i+1}x_{i+1} + \cdots c_{i,n}x_n = b_i. \tag{2.3.4}$$

Since U does not have a leading one in the last column, it follows that all equations beyond the n-th equation are of the form $0x_1 + 0x_2 + \cdots 0x_n = 0$, and as such may be disregarded since they are satisfied by any choice of the x_i. The remaining system of n equations in n unknowns can be solved by back-substitution, yielding a *unique* solution $(x_1, x_2, \ldots, x_n)$ of the form

$$\begin{aligned} x_n &= b_n \\ x_{n-1} &= b_{n-1} - c_{n-1,n}x_n = b_{n-1} - c_{n-1,n}b_n \\ x_{n-2} &= b_{n-2} - c_{n-2,n}x_n - c_{n-2,n-1}x_{n-1} \end{aligned}$$

$$= b_{n-2} - c_{n-2,n}b_n - c_{n-2,n-1}(b_{n-1} - c_{n-1,n}b_n)$$
$$\vdots$$

Do not concern yourself overly with the exact formulas. The important point here is that once we know there is a unique assignment of the variables $x_n, x_{n-1}, \dots, x_{i+1}$ that satisfies the system, (2.3.4) allows us to solve for x_i in terms of the the x_j, $j > i$. As such working our way up from the last equation, we find there is a unique solution to the system.

Case 2: free variables. Suppose now that $x_{i_1}, x_{i_2}, \dots, x_{i_r}$ are the leading variables of L', and $x_{j_1}, x_{j_2}, \dots, x_{j_s}$ are the free variables. Again, since U does not have a leading one in the last column, there are exactly r nonzero equations in L': one for each leading variable. After bringing any terms involving free variables to the right, the k-th such equation takes the form

$$x_{i_k} + c_{k,k+1}x_{i_{k+1}} + \cdots c_{k,r}x_{i_r} = d_k - G_k(x_{j_1}, x_{j_2}, \dots, x_{j_s}).$$

As in the previous case, back-substitution now allows us to solve for each leading variable as a function of the free variables:

$$\begin{aligned}
x_{i_r} &= d_r - G_r(x_{j_1}, x_{j_2}, \dots, x_{j_s}) \\
&= F_s(x_{j_1}, x_{j_2}, \dots, x_{j_s}) \\
x_{i_{r-1}} &= d_{r-1} - c_r x_{i_r} - G_{r-1}(x_{j_1}, x_{j_2}, \dots, x_{j_s}) \\
&= d_{r-1} - c_r F_r(x_{j_1}, x_{j_2}, \dots, x_{j_s}) - G_{r-1}(x_{j_1}, x_{j_2}, \dots, x_{j_s}) \\
&= F_{r-1}(x_{j_1}, x_{j_2}, \dots, x_{j_s}) \\
&\vdots \\
x_1 &= F_1(x_{j_1}, x_{j_2}, \dots, x_{j_s})
\end{aligned}$$

This new system of equations clearly has the same set of solutions as L' (and L), since it was obtained from L' by deleting zero rows of U and using only addition and subtraction operations. Furthermore, it is clear that any assignment of the free variables

$$x_{j_1} = t_1, x_{j_2} = t_2, \dots, x_{j_s} = t_s$$

extends uniquely to the solution of L' that further assigns

$$x_{i_1} = F_1(t_1, t_2, \dots, t_s), x_{i_2} = F_2(t_1, t_2, \dots, t_s), \dots, x_r = F_r(t_1, t_2, \dots, t_s).$$

The idea behind uniqueness here, is that once you assign values to the free variables, the values of the leading variables are completely determined by the equations $x_{i_k} = F_k(x_{j_1}, x_{j_2}, \dots, x_{j_s})$.

Lastly, we show that every solution of L' (and L) is obtained in this way. Suppose $u = (u_1, u_2, \dots, u_n)$ is a solution. According to the discussion above u must be the unique solution to L' corresponding to the free variable assignment

$$x_{j_1} = u_{j_1}, x_{j_2} = u_{j_2}, \dots, x_{j_s} = u_{j_s}$$

and corresponding leading variable assignment

$$x_{i_1} = F_1(u_{j_1}, x_{j_2}, \dots, x_{i_2}), x_{i_2} = F_2(u_{j_1}, x_{j_2}, \dots, x_{j_s}), \dots, x_{i_r} = F_r(u_{j_1}, x_{j_2}, \dots, x_{j_s}).$$

■

Interesting in its own right is the following corollary of Theorem 2.3.5, which tells us that a linear system has either no solutions, exactly one solutions, or