

1.5 Left, Right and Two-Sided Inverses

Section 1.4 sorted functions by two questions. Does f ever send two different inputs to the same output? Does f reach every point of its codomain? The answers were the words injective and surjective (Definition 1.4.3, Definition 1.4.4). This section asks a question that sounds different: can the work of f be **undone**? It turns out that it asks the same two questions again, one for each side of f . Undoing f after it has acted forces f to be injective, and when X is non-empty injectivity is also enough. Undoing it before, so that f finishes the job, forces f to be surjective. The one new idea in the section is an inverse on one side; an inverse on both sides, written f^{-1} , is simply a map that is both.

Throughout, $f: X \rightarrow Y$ is a function in the sense of Definition 1.4.1, and composites are formed as in Definition 1.4.2: $g \circ f$ is defined when the codomain of f is the domain of g , and $(g \circ f)(x) = g(f(x))$.

1.5.1 Undoing a function

Start with a function we can compute in our heads. Let $d: \mathbb{N} \rightarrow \mathbb{N}$ be the doubling map, $d(n) = 2n$. Suppose someone applies d to a secret number and tells us the result, say 14. We recover the secret by halving: it was 7. Halving undoes doubling.

But halving is not quite a function from $\mathbb{N}$ to $\mathbb{N}$: 7 has no half in $\mathbb{N}$. If we want a function $h: \mathbb{N} \rightarrow \mathbb{N}$ that undoes d , we must give h a value at every odd number as well. Those values do not matter for the undoing, because d never produces an odd number. Any choice will do. So one answer is

$$h(m) := \begin{cases} m/2 & \text{if } m \text{ is even,} \\ 1 & \text{if } m \text{ is odd,} \end{cases}$$

and another is $h'(m) := m/2$ for even m and $h'(m) := m$ for odd m . Both are functions $\mathbb{N} \rightarrow \mathbb{N}$: every $m \in \mathbb{N}$ is even or odd and not both, and an even m is $2k$ for exactly one $k \in \mathbb{N}$, which is $m/2$. Both undo d : for every $n \in \mathbb{N}$, the number $2n$ is even, so

$$h(d(n)) = h(2n) = \frac{2n}{2} = n,$$

and the same computation holds for h' . In the language of composition, $h \circ d = \text{id}_{\mathbb{N}}$ and $h' \circ d = \text{id}_{\mathbb{N}}$.

Now try to undo in the other order. Is there a function $k: \mathbb{N} \rightarrow \mathbb{N}$ with $d \circ k = \text{id}_{\mathbb{N}}$, so that d undoes k ? Such a k would give $d(k(1)) = 1$, that is $2k(1) = 1$, and 1 is odd, not twice a natural number. So there is no such k . The obstacle is visible: 1 is not in the range of d , so no input to d , however chosen, comes out as 1.

Two things happened here, and they are the whole section in miniature. Undoing d **afterwards** was possible, in more than one way, and the reason is that d loses no information: different inputs give different outputs. Undoing it **beforehand** was impossible, and the reason is that d misses a point of its codomain. The two sides behave differently, so they get two names.

1.5.2 Left and right inverses

A left inverse undoes f after it acts; a right inverse is undone by f .

Definition 1.5.1. Left and Right Inverses

Let $f: X \rightarrow Y$ be a function, and let $g: Y \rightarrow X$ be a function in the opposite direction.

- (a) g is a **left inverse** of f if $g \circ f = \text{id}_X$, that is, $g(f(x)) = x$ **for every** $x \in X$.
- (b) g is a **right inverse** of f if $f \circ g = \text{id}_Y$, that is, $f(g(y)) = y$ **for every** $y \in Y$.

In words: a left inverse is written on the **left** of f in the composite, and the composite $g \circ f$ does nothing to points of X . A right inverse is written on the right, and $f \circ g$ does nothing to points of Y . The two phrasings in each clause say the same thing. Two functions $X \rightarrow X$ are equal exactly when they agree at every point (Proposition 1.4.1), and $g \circ f$ and id_X have the same domain and codomain X ; so $g \circ f = \text{id}_X$ means $(g \circ f)(x) = \text{id}_X(x)$ for every $x \in X$, which is $g(f(x)) = x$.

The phrase “ g is a left inverse of f ” and the phrase “ f is a right inverse of g ” say the same thing: both say $g \circ f = \text{id}_X$. Which word applies depends on which function we are talking about.

Check the definition against the doubling map. With $d(n) = 2n$ on $\mathbb{N}$, the functions h and h' above are left inverses of d : the composite $h \circ d$ sends every n back to n . And d has **no** right inverse, because clause (b) fails at the single point $y = 1$. For a right inverse the “for every $y \in Y$ ” is the demanding part: one bad y is enough to rule it out.

Example. Left inverses on small sets

Let $X = \{1, 2\}$, $Y = \{1, 2, 3\}$, and $e: X \rightarrow Y$, $e(1) = 1$, $e(2) = 2$: the inclusion of X into Y .

- (a) Show that $g: Y \rightarrow X$ with $g(1) = 1$, $g(2) = 2$, $g(3) = 2$ is a left inverse of e , and that e is a right inverse of g .
- (b) Is g a right inverse of e ?
- (c) Let $c: X \rightarrow Y$ be constant, $c(1) = c(2) = 3$. Does c have a left inverse?

Solution.

- (a) Clause (a) of [Definition 1.5.1](#) asks for $g(e(x)) = x$ at each of the two points of X : $g(e(1)) = g(1) = 1$ and $g(e(2)) = g(2) = 2$. So $g \circ e = \text{id}_X$, and g is a left inverse of e . Read with the roles exchanged, the same equation $g \circ e = \text{id}_X$ is clause (b) for $g: Y \rightarrow X$ and $e: X \rightarrow Y$: e is a right inverse of g .
- (b) No. Clause (b) would need $e(g(y)) = y$ for every $y \in Y$, and at $y = 3$ we get $e(g(3)) = e(2) = 2 \neq 3$. The equation $e \circ g = \text{id}_Y$ fails at one point, which is enough.
- (c) No. A left inverse g would need $g(c(1)) = 1$ and $g(c(2)) = 2$, that is $g(3) = 1$ and $g(3) = 2$. A function has one value at 3, and $1 \neq 2$. The information “which input was it” is gone once both inputs land on 3, and nothing applied afterwards can bring it back.

Part (c) is the degenerate case of the example, and it is the one to remember. A left inverse must recover x from $f(x)$, so it cannot exist when two inputs share an output. Part (b) shows the other side: g undoes e , but e does not undo g , because e never produces 3.

Warning. A left inverse need not be a right inverse. For the doubling map $d(n) = 2n$ on $\mathbb{N}$ and its left inverse h above, $h(d(n)) = n$ for every n , but $d(h(1)) = d(1) = 2 \neq 1$. The equation $h \circ d = \text{id}_{\mathbb{N}}$ says nothing about $d \circ h$: composition is not commutative (Section 1.4), and here the two composites are different functions.

Quick check. Let $s: \mathbb{N} \rightarrow \mathbb{N}$ be the successor map, $s(n) = n + 1$. Does s have a left inverse? Does it have a right inverse?

Solution. It has a left inverse, for instance $p(m) := m - 1$ for $m \geq 1$ and $p(0) := 0$. This is a function $\mathbb{N} \rightarrow \mathbb{N}$, since $m - 1 \in \mathbb{N}$ when $m \geq 1$, and $p(s(n)) = p(n + 1) = n$ because $n + 1 \geq 1$. It has no right inverse: a right inverse $k: \mathbb{N} \rightarrow \mathbb{N}$ of s would need $s(k(0)) = 0$, that is $k(0) + 1 = 0$, which no $k(0) \in \mathbb{N}$ satisfies. The number 0 is not reached by s .

1.5.3 A left inverse exists exactly for injective functions

The examples point to a pattern. The doubling map, the inclusion e , and the successor map are injective, and each has a left inverse. The constant map c is not injective, and it has none. Here is the pattern as a theorem. Its hypothesis $X \neq \emptyset$ looks like a technicality; after the proof we will see that it is not.

We need the **range** of f from Section 1.4, and from now on we write it as $f(X)$:

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

Section 1.6 extends this notation to the image of any subset of X .

Theorem 1.5.1. Left Inverses and Injectivity

Let $f: X \rightarrow Y$ be a function.

- (a) If f has a left inverse, then f is injective.
- (b) If $X \neq \emptyset$ and f is injective, then f has a left inverse.

Consequently, if $X \neq \emptyset$, then f has a left inverse **if and only if** f is injective.

Part (a) carries no hypothesis on X , on purpose: the last theorem of this section uses it for every X , the empty set included. The hypothesis $X \neq \emptyset$ sits in (b) only, where it is used. As in Proposition 1.4.2 (b), “ $X \neq \emptyset$ ” is read as “ X is non-empty: it has an element”, and the proof uses it in that form. (From an element, $X \neq \emptyset$ follows at once, since the empty set has none (Definition 1.1.4). The passage back, from “ X is not the empty set” to “ X has an element”, is the classical law “not all fail, so one holds” of *Logic and Proof*, Chapter 6; reading the hypothesis as “has an element” avoids it.)

Idea. For (a), start from the goal. Injectivity asks: from $f(x_1) = f(x_2)$, get $x_1 = x_2$. A left inverse g turns each $f(x)$ back into x , so apply g to both sides.

For (b), build g point by point, the way halving was extended to the odd numbers. ① A point y of the range has a preimage, and injectivity says it has only one; let $g(y)$ be that one. ② A point outside the range has no preimage; send it to a fixed point $x_0 \in X$. This is where $X \neq \emptyset$ is needed. ③ Check that ① and ② together define a function, and that $g(f(x)) = x$.

Before reading on: where is $X \neq \emptyset$ used, and what would go wrong at that line if X were empty?

Proof.

- (a) Let $g: Y \rightarrow X$ be a left inverse of f . Let $x_1, x_2 \in X$ with $f(x_1) = f(x_2)$. Since g is a function, equal inputs give equal outputs, so $g(f(x_1)) = g(f(x_2))$. By [Definition 1.5.1 \(a\)](#), $g(f(x)) = x$ for every $x \in X$, so

$$x_1 = g(f(x_1)) = g(f(x_2)) = x_2.$$

Since x_1 and x_2 were arbitrary, f is injective by Definition 1.4.3.

- (b) Suppose $X \neq \emptyset$ and f is injective. Since X has an element, fix one and call it x_0 $[\exists E]$. Let G be the subset of $Y \times X$ given by specification (Definition 1.1.3):

$$G := \{ (y, x) \in Y \times X : (y \in f(X) \text{ and } f(x) = y) \text{ or } (y \notin f(X) \text{ and } x = x_0) \}.$$

We show that for every $y \in Y$ there is exactly one $x \in X$ with $(y, x) \in G$, by checking the two clauses (Fn1) (existence) and (Fn2) (uniqueness) into which Section 1.4 splits the condition of Definition 1.4.1 (Logic and Proof, Proposition 5.8.1). Then G , with codomain X , is a function $g: Y \rightarrow X$. Let $y \in Y$.

Existence (Fn1). By excluded middle (Logic and Proof, Proposition 2.5.3), $y \in f(X)$ or $y \notin f(X)$. *Case 1:* $y \in f(X)$. Then $f(x_1) = y$ for some $x_1 \in X$; name it x_1 $[\exists E]$. The pair (y, x_1) satisfies the first alternative, so $(y, x_1) \in G$. *Case 2:* $y \notin f(X)$. Then (y, x_0) satisfies the second alternative, so $(y, x_0) \in G$. In either case some $x \in X$ has $(y, x) \in G$.

Uniqueness (Fn2). Let $x, x' \in X$ with $(y, x) \in G$ and $(y, x') \in G$. Each pair satisfies one of the two alternatives. If both satisfy the first, then $f(x) = y = f(x')$, so $x = x'$ because f is injective. If both satisfy the second, then $x = x_0 = x'$. If one satisfies the first and the other the second, then $y \in f(X)$ and $y \notin f(X)$, a contradiction, so this case cannot occur. Hence $x = x'$.

So $g: Y \rightarrow X$ is a function, and $g(y) = x$ means $(y, x) \in G$. It remains to show $g \circ f = \text{id}_X$. Let $x \in X$. Then $f(x) \in f(X)$, with x as the witness $[\exists I]$, and $f(x) = f(x)$, so the pair $(f(x), x)$ satisfies the first alternative: $(f(x), x) \in G$, that is, $g(f(x)) = x$. The functions $g \circ f$ and id_X have domain and codomain X and agree at every point, so they are equal by Proposition 1.4.1. Hence g is a left inverse of f .

The last sentence of the theorem combines (a) and (b). This proves the theorem. ■

Look at where each hypothesis did its work. Injectivity was used once, in the uniqueness step:

it is what makes “**the** preimage of y ” a single point. The hypothesis $X \neq \emptyset$ was used once, in the first line of (b), to have a point x_0 for the second alternative. The case split on $y \in f(X)$ used excluded middle, a classical step: the proof does not tell us, for a given y , which case we are in.

Neither hypothesis of (b) can be dropped.

- **Without injectivity.** The constant map c of [Example \(Left inverses on small sets\)](#) (c) sends 1 and 2 to 3, and it has no left inverse. In the proof, the uniqueness step breaks: $(3, 1)$ and $(3, 2)$ both lie in G , so G is not a function.
- **Without $X \neq \emptyset$.** The empty function $\emptyset \rightarrow \{1\}$ is injective: the condition “for all $x_1, x_2 \in \emptyset \dots$ ” holds vacuously. A left inverse would be a function $\{1\} \rightarrow \emptyset$, and there is none, since the point 1 would need a value in the empty set (Proposition 1.4.2). In the proof, the first line fails: there is no x_0 to fix, and the point 1, which lies outside the range, has nowhere to go.

The proof of (b) describes g explicitly in terms of f and x_0 , so on a small example the construction can be run by hand.

Example. Running the construction

Let $f: \{1, 2\} \rightarrow \{a, b, c\}$ be given by $f(1) = a$ and $f(2) = c$, where a, b, c are three different objects. Run the proof of [Theorem 1.5.1](#) (b) with $x_0 = 1$, then with $x_0 = 2$. How many left inverses does f have?

Solution. First the hypotheses: $\{1, 2\} \neq \emptyset$, and f is injective because its two inputs have the different outputs a and c . The range is $f(\{1, 2\}) = \{a, c\}$.

With $x_0 = 1$: the points a and c of the range go back to their unique preimages, $g(a) = 1$ and $g(c) = 2$. The point b lies outside the range, so $g(b) = x_0 = 1$. Check: $g(f(1)) = g(a) = 1$ and $g(f(2)) = g(c) = 2$, so $g \circ f = \text{id}_{\{1, 2\}}$.

With $x_0 = 2$: $g'(a) = 1$, $g'(c) = 2$ as before, and $g'(b) = 2$. The same check shows $g' \circ f = \text{id}_{\{1, 2\}}$.

These are all the left inverses. Any left inverse k must satisfy $k(a) = k(f(1)) = 1$ and $k(c) = k(f(2)) = 2$, so its values on the range are **forced**. Its value at b is free, and there are two points of $\{1, 2\}$ to choose from. So f has exactly 2 left inverses, and the construction with $x_0 = 1$ and with $x_0 = 2$ produces both.

The example shows that a left inverse is **not unique** in general. The values on the range are forced by the equation $g(f(x)) = x$; the values off the range are free, and they may go to any point of X . (They are free but not always varied: if X has only one point, there is only one place to send them.)

1.5.4 A right inverse forces surjectivity

The doubling map has no right inverse because it misses 1. The same obstacle works for every function.

Proposition 1.5.2. Right Inverses Give Surjectivity

Let $f: X \rightarrow Y$ be a function. If f has a right inverse, then f is surjective.

Proof. Let $g: Y \rightarrow X$ be a right inverse of f , and let $y \in Y$. By [Definition 1.5.1](#) (b), $f(g(y)) = y$. So $x := g(y)$ is an element of X with $f(x) = y$ $[\exists I]$. Since y was arbitrary, f is surjective by [Definition 1.4.4](#). ■

The proof names its witness: the right inverse **is** a rule that hands us, for each y , a preimage $g(y)$. No hypothesis on X or Y is needed. If $Y = \emptyset$, then f is surjective vacuously; and $X = \emptyset$ as well, since there is no function from a non-empty set to $\emptyset$ ([Proposition 1.4.2](#)).

Right inverses are not unique either. Here is a small surjection with several.

Example. All right inverses of a small surjection

Let $q: \{1, 2, 3, 4\} \rightarrow \{0, 1\}$ send each number to 0 if it is even and to 1 if it is odd: $q(1) = 1$, $q(2) = 0$, $q(3) = 1$, $q(4) = 0$. Find all right inverses of q .

Solution. A right inverse is a function $k: \{0, 1\} \rightarrow \{1, 2, 3, 4\}$ with $q(k(0)) = 0$ and $q(k(1)) = 1$. The first condition says that $k(0)$ is a preimage of 0, so $k(0) \in \{2, 4\}$. The second says $k(1) \in \{1, 3\}$. The two choices are independent, and each combination gives a function $\{0, 1\} \rightarrow \{1, 2, 3, 4\}$ satisfying both conditions. So q has exactly 4 right inverses:

$$(k(0), k(1)) = (2, 1), \quad (2, 3), \quad (4, 1), \quad (4, 3).$$

Each one picks one preimage of each point of the codomain. By [Proposition 1.5.2](#), the fact that one exists shows again that q is surjective.

1.5.5 Does every surjection have a right inverse?

[Theorem 1.5.1](#) went both ways: for $X \neq \emptyset$, a left inverse exists exactly when f is injective. Does [Proposition 1.5.2](#) go both ways too? Suppose $f: X \rightarrow Y$ is surjective. Must f have a right inverse?

Read what a right inverse must do. For **every** $y \in Y$, $g(y)$ must be a preimage of y : an x with $f(x) = y$. Surjectivity promises that each y has at least one preimage. It does not say which, and there may be many. So the right inverse must **choose** one preimage for every y , all at once, and assemble the choices into one set of pairs.

Now look at what our rules can do.

- **One y .** For a single y , the statement “there is $x \in X$ with $f(x) = y$ ” lets us name a preimage by $\exists E$ (*Logic and Proof*, Section 5.6). That is one choice, made inside one proof.
- **Finitely many y .** For a particular finite $Y = \{y_1, \dots, y_n\}$, with n a fixed number, we can apply $\exists E$ n times, once for each point, and write down the n pairs. (To get this for every finite Y at once, with n arbitrary, one inducts on n , *Logic and Proof*, Axiom 4.2.1.)

Example (All right inverses of a small surjection) picked one preimage per point in the same spirit, by listing the preimage sets and choosing from them.

- **Infinitely many y .** A proof is a finite object. It cannot contain one $\exists E$ for each of infinitely many points, and no rule of *Logic and Proof*, Chapter 5 makes infinitely many choices in one step.

Sometimes the difficulty does not arise, because a **rule** picks the preimage. The squaring map from $\mathbb{R}$ onto $[0, \infty)$ is surjective, and each $y \geq 0$ has the preimages $\sqrt{y}$ and $-\sqrt{y}$. The rule “take the non-negative one” defines $g(y) := \sqrt{y}$, and g is a right inverse: $(\sqrt{y})^2 = y$. Likewise the projection $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$, $(m, n) \mapsto m$, is surjective, and $m \mapsto (m, 1)$ is a right inverse. In each case the specification of Definition 1.1.3 writes g down, because the rule names one preimage for each y .

For a surjection with no such rule, nothing proved so far produces a right inverse. The missing ingredient is a new principle of set formation, which says that a choice of one element from each set of a family can be made. Chapter 3 states it and proves, with it, that every surjection has a right inverse. Until then, we use only the implication proved above: a right inverse gives surjectivity.

1.5.6 Two-sided inverses

The two halves meet in a single fact about composition: when a function has an inverse on each side, the two inverses coincide.

Proposition 1.5.3. Left and Right Inverses Agree

Let $f: X \rightarrow Y$ be a function. If $g: Y \rightarrow X$ is a left inverse of f and $h: Y \rightarrow X$ is a right inverse of f , then $g = h$.

Proof. By Proposition 1.4.3 (associativity and the identity laws),

$$g = g \circ \text{id}_Y = g \circ (f \circ h) = (g \circ f) \circ h = \text{id}_X \circ h = h,$$

where the first and last equalities are the identity laws, the second uses that h is a right inverse, the third is associativity, and the fourth uses that g is a left inverse. This proves the proposition. ■

The chain is worth reading twice. The composite $g \circ f \circ h$ can be simplified from either end: the left pair $g \circ f$ collapses to id_X , the right pair $f \circ h$ collapses to id_Y . Associativity says both simplifications are of the same function.

Definition 1.5.2. Inverse Function

Let $f: X \rightarrow Y$ be a function. A function $g: Y \rightarrow X$ is a **two-sided inverse**, or an **inverse**, of f if it is both a left inverse and a right inverse of f :

$$g \circ f = \text{id}_X \quad \text{and} \quad f \circ g = \text{id}_Y.$$

A function with an inverse is **invertible**. Its inverse is **unique**: if g and g' are both inverses

of f , then g is a left inverse and g' a right inverse, so $g = g'$ by [Proposition 1.5.3](#). We call it **the** inverse of f and write it $f^{-1}: Y \rightarrow X$.

In words: f^{-1} undoes f on both sides. It sends $f(x)$ back to x , and f sends $f^{-1}(y)$ back to y . The article “the” was earned by [Proposition 1.5.3](#); without it, f^{-1} could name several functions, as “a left inverse” does.

The definition is symmetric in f and f^{-1} . The two equations that say f^{-1} is an inverse of f also say that f is an inverse of f^{-1} . So f^{-1} is invertible, and by uniqueness its inverse is f :

$$(f^{-1})^{-1} = f.$$

This is the pattern for every identity about inverses: to show that some function is **the** inverse of f , check the two equations of [Definition 1.5.2](#), and uniqueness does the rest.

Example. Inverses found and checked

Show that each function is invertible by finding its inverse and checking both composites.

- (a) $a: \mathbb{R} \rightarrow \mathbb{R}$, $a(x) = 3x - 2$.
- (b) $r: (0, \infty) \rightarrow (0, \infty)$, $r(x) = 1/x$.
- (c) $s: \mathbb{R} \rightarrow \mathbb{R}$, $s(x) = x|x|$, that is, $s(x) = x^2$ for $x \geq 0$ and $s(x) = -x^2$ for $x < 0$.

Solution. Idea. Solve $y = f(x)$ for x to find a candidate, then check both equations of [Definition 1.5.2](#). The solving is scratch work; the two checks are the proof.

- (a) Solving $y = 3x - 2$ gives $x = (y + 2)/3$. Let $b(y) := (y + 2)/3$, a function $\mathbb{R} \rightarrow \mathbb{R}$. For every $x \in \mathbb{R}$, $b(a(x)) = (3x - 2 + 2)/3 = x$, and for every $y \in \mathbb{R}$, $a(b(y)) = 3 \cdot (y + 2)/3 - 2 = y$. So $b = a^{-1}$.
- (b) For $x > 0$, $1/x > 0$, so r does map $(0, \infty)$ to $(0, \infty)$. Here $r(r(x)) = 1/(1/x) = x$ for every $x > 0$. This one equation is both checks at once, with r in the roles of f and g . So $r^{-1} = r$: the function is its own inverse.
- (c) Solve by cases on the sign of y , since $s(x)$ has the sign of x . For $y \geq 0$ the candidate is $\sqrt{y}$; for $y < 0$ it is $-\sqrt{-y}$. Let

$$t(y) := \begin{cases} \sqrt{y} & \text{if } y \geq 0, \\ -\sqrt{-y} & \text{if } y < 0. \end{cases}$$

First check, $t(s(x)) = x$, by cases on x . If $x \geq 0$, then $s(x) = x^2 \geq 0$, so $t(s(x)) = \sqrt{x^2} = x$, the non-negative square root of x^2 being x . If $x < 0$, then $s(x) = -x^2 < 0$, so $t(s(x)) = -\sqrt{-x^2} = -(-x) = x$, since the non-negative square root of x^2 is $-x$.

Second check, $s(t(y)) = y$, by cases on y . If $y \geq 0$, then $t(y) = \sqrt{y} \geq 0$, so $s(t(y)) = (\sqrt{y})^2 = y$. If $y < 0$, then $t(y) = -\sqrt{-y} < 0$, so $s(t(y)) = -(\sqrt{-y})^2 = -(-y) = y$.

Both checks hold, so $t = s^{-1}$. Notice that each check needs its own case split: the first splits on the sign of the input x , the second on the sign of y .

Now the theorem that connects inverses with the words of Section 1.4.

Theorem 1.5.4. Bijections Are the Invertible Functions

A function $f: X \rightarrow Y$ is bijective if and only if it has a two-sided inverse.

Idea. $(\Leftarrow)$ is already proved in two halves: an inverse is a left inverse, which gives injectivity by [Theorem 1.5.1 \(a\)](#), and a right inverse, which gives surjectivity by [Proposition 1.5.2](#). Neither half asks anything of X .

$(\Rightarrow)$ looks like the open question of the last subsection, since we must pick a preimage of every y . But for a bijection there is nothing to pick. Surjectivity gives each y a preimage and injectivity says there is only one. So the set of pairs (y, x) with $f(x) = y$ already has the $\forall\exists!$ property of Definition 1.4.1.

Proof. $(\Leftarrow)$ Let g be a two-sided inverse of f . Then g is a left inverse of f , so f is injective by [Theorem 1.5.1 \(a\)](#); and g is a right inverse, so f is surjective by [Proposition 1.5.2](#). Hence f is bijective by Definition 1.4.5.

$(\Rightarrow)$ Let f be bijective, and let

$$G := \{(y, x) \in Y \times X : f(x) = y\},$$

a subset of $Y \times X$ by specification (Definition 1.1.3). Let $y \in Y$. Since f is surjective, there is $x \in X$ with $f(x) = y$, so some x has $(y, x) \in G$. If $(y, x) \in G$ and $(y, x') \in G$, then $f(x) = y = f(x')$, so $x = x'$ since f is injective. So G satisfies (Fn1) and (Fn2) (Logic and Proof, Proposition 5.8.1): for every $y \in Y$ there is exactly one $x \in X$ with $(y, x) \in G$, and G with codomain X is a function $g: Y \rightarrow X$ (Definition 1.4.1). By construction, $g(y) = x$ holds exactly when $f(x) = y$.

Now let $y \in Y$. Then $g(y) = x$ for the x with $(y, x) \in G$, so $f(g(y)) = y$. And let $x \in X$. The pair $(f(x), x)$ lies in G , since $f(x) = f(x)$; so $g(f(x)) = x$. By Proposition 1.4.1, $f \circ g = \text{id}_Y$ and $g \circ f = \text{id}_X$. So g is a two-sided inverse of f . This proves the theorem. ■

Compare the two constructions of this section. The left inverse of [Theorem 1.5.1 \(b\)](#) needed a fixed point x_0 and a case split by excluded middle, because some y have no preimage. The inverse of a bijection needs neither. Every y has **exactly one** preimage, so the definition “ $g(y)$ is the preimage of y ” names one element for each y without any choosing. This is why no new principle is needed here, while the open question about surjections needs one. The difference between the two is exactly the difference between “exactly one” and “at least one”.

The theorem also covers the degenerate case. A function $f: \emptyset \rightarrow Y$ is surjective only if $Y = \emptyset$, since a point of Y would need a preimage in $\emptyset$. The empty function $\emptyset \rightarrow \emptyset$ is bijective, and it is its own inverse: both composites are the empty function $\emptyset \rightarrow \emptyset$, which is $\text{id}_\emptyset$. The proof of $(\Leftarrow)$ runs at $X = \emptyset$ without change, because it cites only part (a) of [Theorem 1.5.1](#), which has no hypothesis on X .

Remark. The notation f^{-1} is reserved, as a function, for a bijection f . For a function that is not bijective, “ $f^{-1}: Y \rightarrow X$ ” names nothing. Section 1.6 uses the symbol in a second way,

$f^{-1}(B)$ for a **set** B , which makes sense for every function; it shows that for a bijection the two readings give the same set.

Warning. Do not write f^{-1} for a left inverse. The doubling map d has the two different left inverses h and h' of the first subsection, so “ d^{-1} ” would not say which one is meant. And neither is an inverse: $d(h(1)) = 2 \neq 1$. The notation is earned by [Proposition 1.5.3](#), which needs an inverse on **both** sides.

Quick check. Does $x \mapsto x^3$ from $\mathbb{R}$ to $\mathbb{R}$ have a two-sided inverse? Does $x \mapsto x^2$ from $[0, \infty)$ to $\mathbb{R}$?

Solution. The first does: $y \mapsto \sqrt[3]{y}$, the real cube root, which exists and is unique for every real y (a school fact). Indeed $\sqrt[3]{x^3} = x$ and $(\sqrt[3]{y})^3 = y$. The second does not. It is injective (Example (Squaring, three ways) (b)), but it is not surjective, since -1 has no preimage; by [Theorem 1.5.4](#) it has no two-sided inverse. (It has left inverses, by [Theorem 1.5.1](#), for instance $y \mapsto \sqrt{y}$ for $y \geq 0$ and $y \mapsto 0$ for $y < 0$.)

1.5.7 What the section proved

For a function $f: X \rightarrow Y$, the section proved four facts and left one question open.

- A left inverse makes f injective ([Theorem 1.5.1](#) (a)).
- If $X \neq \emptyset$, an injective f has a left inverse ([Theorem 1.5.1](#) (b)). The empty function $\emptyset \rightarrow \{1\}$ shows that the hypothesis is needed.
- A right inverse makes f surjective ([Proposition 1.5.2](#)).
- A two-sided inverse exists exactly when f is bijective ([Theorem 1.5.4](#)), and it is unique ([Proposition 1.5.3](#)).
- **Open:** a surjective f has a right inverse. Chapter 3 proves it, with the new principle.

1.5.8 Exercises

1.5.8.1 A. Check your understanding

Exercise. A1

These parts are about [Theorem 1.5.1](#) and its proof.

- At which line of the proof of (b) is $X \neq \emptyset$ used, and for what?
- The proof shows that the set G is a function by proving two things about the case split into $y \in f(X)$ and $y \notin f(X)$. What are they, and which step of the proof uses each?
- Run the construction on $f: \{1, 2, 3\} \rightarrow \{1, 2, 3, 4\}$, $f(k) = k + 1$, with $x_0 = 3$. Write out the left inverse g and check $g \circ f = \text{id}_{\{1, 2, 3\}}$.
- Give the idea of the proof of (b) in two sentences.
- Let $f: \{1, 2\} \rightarrow \{1\}$ be the only function between these sets. Which step of the proof

of (b) fails for f , and why?

Solution.

- (a) In the first line of (b): “fix one and call it x_0 ”. The point x_0 is the value $g(y)$ for every y outside the range $f(X)$; without it, the second alternative in the definition of G would have nothing to send such y to.
- (b) The two cases **exhaust** every y : by excluded middle, $y \in f(X)$ or $y \notin f(X)$; this is used in the existence step. And they **do not overlap**: a y cannot satisfy both, since $y \in f(X)$ and $y \notin f(X)$ contradict each other; this is used in the uniqueness step, to rule out a pair from each alternative. (Uniqueness inside the first case uses injectivity; inside the second, it is automatic.)
- (c) The hypotheses hold: $\{1, 2, 3\} \neq \emptyset$, and f is injective, since $k + 1 = k' + 1$ gives $k = k'$. The range is $f(\{1, 2, 3\}) = \{2, 3, 4\}$. On the range, g returns the unique preimage: $g(2) = 1$, $g(3) = 2$, $g(4) = 3$. The point 1 is outside the range, so $g(1) = x_0 = 3$. Check: $g(f(1)) = g(2) = 1$, $g(f(2)) = g(3) = 2$, $g(f(3)) = g(4) = 3$. So $g \circ f = \text{id}_{\{1,2,3\}}$.
- (d) A point of the range has exactly one preimage, by injectivity, and g sends it there; a point outside the range is sent to a fixed point x_0 of the non-empty domain. Then $g(f(x))$ is the unique preimage of $f(x)$, which is x .
- (e) The uniqueness step fails. The point $y = 1$ lies in the range and has the two preimages 1 and 2, so $(1, 1)$ and $(1, 2)$ both lie in G , and G is not a function. The step “ $f(x) = y = f(x')$, so $x = x'$ ” needs injectivity, which this f lacks. (By part (a) of the theorem, f has no left inverse at all.)

1.5.8.2 B. Practice

Exercise. B1: Counting one-sided inverses

- (a) Find all left inverses of $f: \{1, 2\} \rightarrow \{1, 2, 3\}$, $f(k) = k$.
- (b) Find all right inverses of $g: \{1, 2, 3\} \rightarrow \{1, 2\}$, $g(1) = g(2) = 1$, $g(3) = 2$.

Solution.

- (a) A left inverse $h: \{1, 2, 3\} \rightarrow \{1, 2\}$ must satisfy $h(f(1)) = 1$ and $h(f(2)) = 2$, that is $h(1) = 1$ and $h(2) = 2$. The point 3 is outside the range $\{1, 2\}$, so $h(3)$ is free: 1 or 2. Each choice satisfies both equations. So there are exactly 2 left inverses: $h(3) = 1$ and $h(3) = 2$, with $h(1) = 1$, $h(2) = 2$ in both.
- (b) A right inverse $k: \{1, 2\} \rightarrow \{1, 2, 3\}$ must satisfy $g(k(1)) = 1$ and $g(k(2)) = 2$. The preimages of 1 under g are 1 and 2, so $k(1) \in \{1, 2\}$; the only preimage of 2 is 3, so $k(2) = 3$. So there are exactly 2 right inverses: $k(1) = 1$ or $k(1) = 2$, with $k(2) = 3$ in both.

Exercise. B2: The inverse of a composite

Let $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ be bijections. Prove that $g \circ f$ is invertible and that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Hint: check the two equations of Definition 1.5.2 for the candidate $f^{-1} \circ g^{-1}$.

Solution. The candidate $f^{-1} \circ g^{-1}: Z \rightarrow X$ is defined, since $g^{-1}: Z \rightarrow Y$ and $f^{-1}: Y \rightarrow X$ exist by Theorem 1.5.4. We check the two equations of Definition 1.5.2, using Proposition 1.4.3 to regroup and for its identity laws $k \circ \text{id} = k = \text{id} \circ k$:

$$\begin{aligned}(f^{-1} \circ g^{-1}) \circ (g \circ f) &= f^{-1} \circ (g^{-1} \circ g) \circ f = f^{-1} \circ \text{id}_Y \circ f = f^{-1} \circ f = \text{id}_X, \\ (g \circ f) \circ (f^{-1} \circ g^{-1}) &= g \circ (f \circ f^{-1}) \circ g^{-1} = g \circ \text{id}_Y \circ g^{-1} = g \circ g^{-1} = \text{id}_Z.\end{aligned}$$

So $f^{-1} \circ g^{-1}$ is a two-sided inverse of $g \circ f$. By the uniqueness of inverses (Proposition 1.5.3, as in Definition 1.5.2), it is **the** inverse: $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. Notice the order: to undo “first f , then g ”, undo g first.

Exercise. B3: A bijection between an interval and the line

Let $h: (-1, 1) \rightarrow \mathbb{R}$, $h(x) = x/(1 - |x|)$, and $k: \mathbb{R} \rightarrow (-1, 1)$, $k(y) = y/(1 + |y|)$. Show that h and k are well defined (each denominator is non-zero, and k takes values in $(-1, 1)$), and that k is the inverse of h , by checking both composites by cases on the sign. Hence h is a bijection.

Solution. Well defined. For $x \in (-1, 1)$, $|x| < 1$, so $1 - |x| > 0$. For $y \in \mathbb{R}$, $1 + |y| \geq 1 > 0$, and $|k(y)| = |y|/(1 + |y|) < 1$ because $|y| < 1 + |y|$; so $k(y) \in (-1, 1)$.

First check, $k(h(x)) = x$ for $x \in (-1, 1)$. If $0 \leq x < 1$, then $h(x) = x/(1 - x) \geq 0$, so

$$k(h(x)) = \frac{x/(1 - x)}{1 + x/(1 - x)} = \frac{x/(1 - x)}{1/(1 - x)} = x.$$

If $-1 < x < 0$, then $|x| = -x$ and $h(x) = x/(1 + x) < 0$, so $|h(x)| = -x/(1 + x)$ and

$$k(h(x)) = \frac{x/(1 + x)}{1 - x/(1 + x)} = \frac{x/(1 + x)}{1/(1 + x)} = x.$$

Second check, $h(k(y)) = y$ for $y \in \mathbb{R}$. If $y \geq 0$, then $k(y) = y/(1 + y) \in [0, 1)$, so

$$h(k(y)) = \frac{y/(1 + y)}{1 - y/(1 + y)} = \frac{y/(1 + y)}{1/(1 + y)} = y.$$

If $y < 0$, then $k(y) = y/(1 - y) \in (-1, 0)$, so $|k(y)| = -y/(1 - y)$ and

$$h(k(y)) = \frac{y/(1 - y)}{1 + y/(1 - y)} = \frac{y/(1 - y)}{1/(1 - y)} = y.$$

Both composites are identities, so $k = h^{-1}$, and h is a bijection by Theorem 1.5.4.

1.5.8.3 C. Going deeper

Exercise. C1: One composite equal to the identity

Let $f: X \rightarrow Y$ and $g: Y \rightarrow X$ be functions with $g \circ f = \text{id}_X$.

- (a) Prove that f is injective and g is surjective.
- (b) Show by an example that neither f nor g need be bijective.
- (c) Prove that if, in addition, f is surjective, then $g = f^{-1}$.

Solution (Hint 1). Read the equation $g \circ f = \text{id}_X$ twice: once as a statement about f , once as a statement about g .

Solution (Hint 2). For g : let $x \in X$; you need a point of Y that g sends to x .

Solution.

- (a) The equation says that g is a left inverse of f , so f is injective by [Theorem 1.5.1](#) (a). The same equation says that f is a right inverse of g , so g is surjective by [Proposition 1.5.2](#). (Directly: for $x \in X$, the point $f(x) \in Y$ satisfies $g(f(x)) = x$.)
- (b) Let $f: \{1\} \rightarrow \{1, 2\}$, $f(1) = 1$, and $g: \{1, 2\} \rightarrow \{1\}$, $g(1) = g(2) = 1$. Then $g(f(1)) = 1$, so $g \circ f = \text{id}_{\{1\}}$. But f is not surjective (2 has no preimage) and g is not injective ($g(1) = g(2)$ with $1 \neq 2$), so neither is bijective.
- (c) If f is also surjective, it is bijective by (a), so f^{-1} exists by [Theorem 1.5.4](#), and it is in particular a right inverse of f . Also g is a left inverse of f . By [Proposition 1.5.3](#), $g = f^{-1}$.

Marking scheme. (10 points)

- 2: f injective, from the left-inverse equation (citing [Theorem 1.5.1](#) (a) or arguing directly from $f(x_1) = f(x_2)$).
- 2: g surjective, either by citing [Proposition 1.5.2](#) (f is a right inverse of g) or directly, naming the witness $f(x)$ for a given $x \in X$.
- 3: an example with explicit sets and both failures checked (a point missed by f , two points identified by g).
- 3: (c) obtains a right inverse of f (from bijectivity) and applies uniqueness of one-sided inverses, or checks $f \circ g = \text{id}_Y$ directly (for $y = f(x)$, $f(g(y)) = f(g(f(x))) = f(x) = y$, using surjectivity to write $y = f(x)$).

Fatal errors (0 for the step): assuming $f \circ g = \text{id}_Y$ in (a) or (c) without proof; in (c), writing f^{-1} before showing f bijective.

Exercise. C2: When is a left inverse unique?

Let $f: X \rightarrow Y$ be injective, with $X \neq \emptyset$.

- (a) Prove that if f is surjective, or X has exactly one element, then f has exactly one left inverse.
- (b) Prove that if some $y_1 \in Y$ is not in $f(X)$, and X has two different elements $x_0 \neq x_1$, then f has at least two different left inverses.

- (c) Let $f: \{1, 2\} \rightarrow \{1, 2, 3, 4\}$, $f(k) = 2k$. List all left inverses of f , and explain in one sentence why there are $2^2 = 4$ of them.

Solution (Hint 1). For (a), a left inverse exists by [Theorem 1.5.1](#) (b); only “at most one” needs proof. For (b), run the construction of that proof twice.

Solution (Hint 2). For (a), let g and g' be left inverses of f . If f is surjective, write a given $y \in Y$ as $y = f(x)$ and compare $g(y)$ with $g'(y)$.

Solution.

- (a) A left inverse exists by [Theorem 1.5.1](#) (b). For “at most one”, let g and g' be left inverses of f . *Case 1:* f is surjective. Then f is bijective (Definition 1.4.5), so it has the inverse f^{-1} ([Theorem 1.5.4](#)), which is in particular a right inverse of f . By [Proposition 1.5.3](#), $g = f^{-1}$ and $g' = f^{-1}$, so $g = g'$. *Case 2:* X has exactly one element x_0 . For every $y \in Y$, both $g(y)$ and $g'(y)$ are elements of X , so $g(y) = x_0 = g'(y)$; hence $g = g'$ by Proposition 1.4.1. In either case any two left inverses are equal, so f has exactly one.
- (b) Run the proof of [Theorem 1.5.1](#) (b) once with the fixed point x_0 and once with x_1 ; call the resulting left inverses g_0 and g_1 . The point y_1 is outside the range, so the second alternative in the definition of G applies to it: $g_0(y_1) = x_0$ and $g_1(y_1) = x_1$. Since $x_0 \neq x_1$, the functions g_0 and g_1 differ at y_1 , so $g_0 \neq g_1$.
- (c) The range is $\{2, 4\}$. A left inverse $g: \{1, 2, 3, 4\} \rightarrow \{1, 2\}$ must satisfy $g(f(1)) = 1$ and $g(f(2)) = 2$, that is $g(2) = 1$ and $g(4) = 2$. The values $g(1)$ and $g(3)$ are free, each in $\{1, 2\}$, and every choice gives a left inverse. The four are

$$(g(1), g(2), g(3), g(4)) = (1, 1, 1, 2), (1, 1, 2, 2), (2, 1, 1, 2), (2, 1, 2, 2).$$

The values on the range are forced, and each of the two points 1, 3 outside the range may go to either of the two points of the domain, independently: $2 \cdot 2 = 4$ choices. This is the situation of (b): f is not surjective and its domain has two elements.

Marking scheme. (10 points)

- 1: (a) existence of a left inverse, citing [Theorem 1.5.1](#) (b).
- 2: (a) surjective case: an inverse exists, and uniqueness of one-sided inverses ([Proposition 1.5.3](#)) forces every left inverse to equal it (or a direct argument: for $y = f(x)$, $g(y) = x = g'(y)$).
- 2: (a) one-element case: any two functions into a one-element set agree at every point, so they are equal.
- 3: (b) two left inverses that take the values x_0 and x_1 at y_1 (by the construction or defined directly), each checked to be a left inverse, and shown to differ at y_1 .
- 2: (c) the forced values $g(2) = 1$, $g(4) = 2$, the four left inverses, and the count $2 \cdot 2$.

Fatal errors (0 for the step): in (a), writing f^{-1} before showing f bijective; in (b), concluding $g_0 \neq g_1$ without exhibiting a point where they differ.