HW 1 solutions and an addendum
Comments are welcome

A well written solution, besides being precise, correct and complete, should also be properly organized so that is as easy to follow as possible. For whom? Write for a well intentioned but skeptical reader who knows as much as you did before you solved the problem, maybe even a little less. There are multiple ways to achieve this. Find a style that works for you. Here are some suggested good practices. (i) It is better to write more and give a complete argument than to try to be elegantly concise and leave gaps. (ii) Do not assume that your reader can easily fill details that you had to think through. (iii) Do not try to impress the reader, at least at this stage of your career. (iv) Simpler is always better — as long as it is not wrong.

I have tried to illustrate the preceding discussion to some extent for 1(a) and 1(b) below. Parts in blue indicate setting up the problem, which is free once you understand the meaning of the problem. This is *always* desirable to do unless the solution is clear to you *and* you have enough experience for your confidence to be a reliable indicator of your understanding. Solutions to some problems (1c, 1d and 3) are sketched. You may need to work through them and fill in details to make them complete.

---

1. A function $f : S \to T$ is called *invertible* or an *isomorphism of sets* if there is a function $g : T \to S$, called *an* inverse of f, such that $f \circ g = id_T$ and $g \circ f = id_S$. (a) Show that f is invertible if and only if f is both injective (i.e., one-to-one) and surjective (i.e., onto).

Note that being an *if and only if* statement, part (a) requires two arguments. One of the two is so immediate that it may be confusing to articulate it precisely. Try to do the rest too as concisely and incisively as you can. It may help to explore the consequence of just one of the two required equations and prove a subclaim. Does $f \circ g = id_T$ imply something about f? About g?

Follow-your-nose solution to (a). First suppose f is invertible with g an inverse, so $f \circ g = id_T$ and $g \circ f = id_S$. We are asked to show that f is injective and surjective.

(i) To show f is injective, suppose $f(x_1) = f(x_2)$ where $x_1, x_2 \in S$. It is enough to show $x_1 = x_2$. Applying g to both sides of $f(x_1) = f(x_2)$ gives $g(f(x_1)) = g(f(x_2))$. As $g \circ f = id_S$, we have $g(f(x_1)) = x_1$ and $g(f(x_2)) = x_2$. Putting together the three relevant equalities shows that $x_1 = x_2$.

(ii) To show f is surjective, let $y \in T$. It is enough to find $x \in S$ such that $f(x) = y$. As we know that $f \circ g = id_T$, we have $f(g(y)) = y$. So taking $x = g(y)$ gives $f(x) = f(g(y)) = y$ as required.

Conversely, suppose that f is injective and surjective. We are required to show existence of a function $g : T \to S$ such that $f \circ g = id_T$ and $g \circ f = id_S$. To define g, take $y \in T$. By surjectivity of f there exists at least one element in S, say x, such that $f(x) = y$. Moreover, such x is unique by injectivity of f because for any element $x' \in S$ such that $f(x') = f(x)$, injectivity of f forces $x = x'$. So for a given $y \in T$, we may define a function g by saying $g(y) =$ the *unique* element, say x, in S with $f(x) = y$. Now for any $s \in S$, we have $g(f(s)) = s$ by the definition of g. For any $t \in T$, we have $f(g(t)) = t$, again by the definition of g. This completes the proof of (a).

Another solution to first half of (a). The ideas in the above proof give the following sharper lemma, as suggested by questions in the note. *Lemma*: If $f \circ g = id_T$ then f is surjective and g is injective. *Proof*: To show injectivity of g, suppose $g(x_1) = g(x_2)$. Apply f to both sides to get $f(g(x_1)) = f(g(x_2))$. Using $f \circ g = id_T$, this gives $x_1 = x_2$, showing that g is injective. To show f is surjective, let $y \in T$. We will be done by exhibiting some $x \in S$ such that $f(x) = y$. Define $x = g(y)$. Now $f(x) = f(g(y)) = y$, as needed.

Returning to the given problem, if f is invertible with g an inverse, then the lemma applies to $f \circ g = id_T$ to give that f is surjective (and g injective). The same lemma applied to $g \circ f = id_S$ gives that f is injective (and g surjective).

A shorter rendering of the converse: Suppose f is surjective and injective. Given $y \in T$, define $g(y) =$ the unique x for which $f(x) = y$. Such x exists by surjectivity of f and is unique by injectivity of f. By definition of g, we immediately get $f(g(y)) = y$ for any $y \in T$ and $g(f(x)) = x$ for any $x \in S$.

(b) Show that if f is invertible then it has a unique inverse.

Solution to (b). Suppose g_1, g_2 are inverses to f. Consider $h := g_2 \circ f \circ g_1$, where the RHS does not need parentheses because of associativity of function composition. Observe that $h = (g_2 \circ f) \circ g_1 = g_1$ and simultaneously $h = g_2 \circ (f \circ g_1) = g_2$, showing $g_1 = g_2$.

Notes: (1) Notice that we only used $f \circ g_1 = id_T$ and $g_2 \circ f = id_S$. So the tiny calculation above shows more than what was asked, namely the following: if a function has a left inverse and a right inverse, then the two inverses must be the same and this common function is a two-sided inverse, perforce unique. This is an utterly standard tidbit that you should see once and then swat away whenever it appears again, as it will in linear/abstract algebra. (2) Do you have an itch to see the tiny calculation written as $g_1 = (g_2 \circ f) \circ g_1 = g_2 \circ (f \circ g_1) = g_2$? Congratulations on sharing that judgment with most textbook authors. Yes, it is a bit cuter that way, but now the key expression $g_2 \circ f \circ g_1$ is buried in the middle. It is produced by magic, as it were, instead of being honestly revealed as the initial brainwave that makes the proof tick. Faux efficiency at the cost of a bit of transparency?

(c) *Results mining the same vein.* A function $f : S \to T$ is called left invertible if it has a left inverse, i.e., a function $g : T \to S$ with $gf = id_S$. (We're dropping $\circ$ from now on.) Which functions have a left inverse? If a left inverse exists, must it be unique? Repeat for right inverses. What can you say about a function having a left inverse and a right inverse?

Now call f right cancellable if for any functions h_1, h_2, the equation $h_1 f = h_2 f$ implies $h_1 = h_2$. Can you see an immediate formal connection between being left/right cancellable and left/right invertible? Which functions are right cancellable? Left cancellable? Both?

Note: This part gets one's toes wet in formulating statements purely in terms of functions and their composition as opposed to using elements. Analogues of such statements and this style of doing business *might* become relevant for some of you later on when you deal with categories of objects other than plain sets. Unless and until that happens, don't get too fond of such things.

Solution sketch for (c). The lemma in (a) shows that a function with a left inverse must be injective and a function with a right inverse must be surjective. Converses to both are true by language appropriately modeled on the last paragraph in the first solution to (a) with the following additional ingredient: appropriate choices are involved for defining a desired inverse function on certain input elements (the ones not in the image of an injective f and the ones that have multiple preimages for a surjective f). Uniqueness is false for both cases. Examples are easy to construct using the additional ingredient. If f has a left inverse g_1 and a right inverse g_2 then $g_1 = g_2$ is the unique two-sided inverse by note (1) in the solution to (b) above.

If f has a right inverse g, then $h_1 f = h_2 f$ implies $h_1 = h_1 f g = h_2 f g = h_2$ (ha!), showing f is right cancellable. This part was formal. The converse is also true. (Proof is left for you. How will you set it up?) All in all we have the following equivalences for functions between sets: surjective $\Leftrightarrow$ right invertible $\Leftrightarrow$ right cancellable. Similarly injective $\Leftrightarrow$ left invertible $\Leftrightarrow$ left cancellable. For objects with more structure than sets and appropriate morphisms between them, these issues have to be examined again.

(d) *Two-out-of-three property.* See informally for yourself that any two of the following properties for a function between *finite* sets implies the third: injective, surjective, domain and codomain have the same finite cardinality. In particular a function between finite sets of equal cardinality is injective if and only if it is surjective. This is not so exciting for sets but its analogue for vector spaces is quite useful, as you will see. What happens if we drop the word *finite* everywhere?

Solution sketch for (d). The claim is "intuitively obvious". A proper discussion of the proof will require making precise the meaning of cardinality and some way to deal with finiteness, possibly going down to developing natural numbers from some logical foundation. For a start, $|S| = |T|$ *means by definition* that there exists a bijection between the sets S and T. So a function $f : S \to T$ being injective and surjective directly gives $|S| = |T|$, finite cardinality or not. For infinite sets, the other two claims are false. Let $\mathbb{N}$ be the set of non-negative integers. The function $f(x) = x + 1$ from $\mathbb{N}$ to itself is injective but not surjective as 0 is not in the image. Consider a function, again from $\mathbb{N}$ to itself, defined by $g(x) = x - 1$ for $x > 0$ and $g(0) =$ some k. Then g is surjective but not injective as $g(0) = g(k+1)$.

To go a bit further, what should be the meaning of $|S| \leq |T|$? One answer is that there exists an injection from S to T. Can we equivalently use a surjection instead? Next, $|S| \leq |T|$ and $|T| \leq |S|$ should imply $|S| = |T|$. Proof? Read more if you are interested.

---

2. Consider the commonly made statement "Defining an equivalence relation on a set S is the same as defining a partition of S". Make the preceding vague statement precise. Then prove the precise statement. Both underlined notions are defined below. Hint: I almost made this a part of the previous problem.

A *partition* of S is defined to be a set of pairwise disjoint subsets of S whose union is S, i.e., a partition of S is a set $\{S_\alpha | \alpha \in I\}$, where α ranges over some index set I and each $S_\alpha \subset S$, such that $\bigcup_{\alpha \in I} S_\alpha = S$ and for all distinct $\alpha, \beta \in I$, one has $S_\alpha \bigcap S_\beta = \varnothing$.

Recall that a relation on a set S means a relation from S to itself, i.e., a subset of $S \times S$. A relation $\sim$ on a set S is called an equivalence relation if it is

- reflexive (i.e., for each $x \in S$, one has $x \sim x$),
- symmetric (i.e., whenever $x \sim y$, one must also have $y \sim x$), and
- transitive (i.e., whenever $x \sim y$ and $y \sim z$, one must also have $x \sim z$).

As an aside, recall that changing *symmetric* to *antisymmetric* gives the notion of a partial order, a completely different kind of beast.

Solution. Fix a set S. Claim: there is a bijection between (1) $E =$ the set of equivalence relations on S and (2) $P =$ the set of partitions of S. (This is the least that is required to justify the vague but potent "is the same as" in the question.[1]) To prove the claim we will define functions $f : E \to P$ and $g : P \to E$ such that $fg = id_P$ and $gf = id_E$. This is enough by problem 1(a). Note that just defining f and g is not enough.

Definition of f. Given an equivalence relation $\sim$ on S, for any $x \in S$, define the "equivalence class of x" to be $S_x = \{y \in S | x \sim y\}$. Define $f(\sim) =$ the collection $\{S_x | x \in S\}$ of subsets of S. (Remember that a set has no repeats, so the cardinality of $f(\sim)$ may look like $|S|$ but instead it is the number of *distinct* subsets S_x.)

To see that $f(\sim)$ forms a partition of S, first observe that by reflexivity of $\sim$, we have $x \in S_x$. So $\bigcup_{x \in S} S_x = S$. It remains to show that distinct equivalence classes are disjoint. We show the equivalent contrapositive: $z \in S_x \bigcap S_y$ implies that $S_x = S_y$. This can be deduced from symmetry and transitivity of $\sim$ as follows. We have $x \sim z$ and $y \sim z$. By symmetry $z \sim y$. Using transitivity with $x \sim z$ gives $x \sim y$ and so $y \sim x$. Now for any $w \in S_x$, we have $x \sim w$. Together with $y \sim x$ we get $y \sim w$, showing $S_x \subset S_y$. Similarly $S_y \subset S_x$.

Definition of g. Given a partition $\{S_\alpha | \alpha \in I\}$ of S, define g(this partition) to be the relation $\equiv$ specified as follows. Declare $x \equiv y$ exactly when x, y belong to the same part, i.e., when there exists $\alpha \in I$ such that $x \in S_\alpha$ and $y \in S_\alpha$. Now see that $\equiv$ is reflexive (because $\bigcup_{\alpha \in I} S_\alpha = S$), symmetric (immediate), and transitive (because $x \equiv y, y \equiv z$ forces x, y, z to belong to the same unique part as distinct parts are disjoint).

It is "easy to see" that (i) $fg = id_P$ and (ii) $gf = id_E$. Proof: (i) Given a partition, first g creates a relation in which elements are declared related to each other iff they are in the same part and then f simply clubs together the elements related to each other, recreating the original parts. (ii) Given an equivalence relation, f assigns to it the set of corresponding equivalence classes. By construction these classes have the property that all elements in a class are mutually related and no element in a class is related to one in a different class. Now g declares a relation by the exact same recipe, thus recreating the original relation. (*Is this "proof" valid? My judgment is that once the definitions of f and g are digested, the preceding descriptive proof is (1) acceptable in our context and (2) maybe even preferable to a proof couched in terms of notation because the latter may only obscure the simplicity of what is going on. If you disagree, by all means write a more technical proof. For a start, you will have to name a given partition, which I avoided above.*)

[1] In fact "is the same as" should be interpreted as asserting the existence of a bijection that is *natural* in a precise sense. The technical term, which you may learn about later, is functorial isomorphism.

3. Let $f : S \to T$ be a function. Recall two definitions: for $A \subset S$, we have $f(A) = \{f(x)|x \in A\}$, a subset of T. For $B \subset T$, we have $f^{-1}(B) = \{x|f(x) \in B\}$, a subset of S. The latter notation does not require f^{-1} to be a function from T (but can be thought of as defining a function from the power set of T to that of S). Answer the following with proofs/counterexamples as necessary. While all questions are phrased as asking for a YES/NO answer, supply the strongest results that you can for all parts.

(a) Is $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$, i.e., does f preserve finite unions? (Why "i.e."?) Arbitrary unions?

(b) Does f preserve finite intersections? Arbitrary intersections?

(c) Does f preserve complements, i.e., is $f(A^c) = f(A)^c$? More precisely, is $f(S \setminus A) = T \setminus f(A)$?

(d) Repeat the previous questions for f^{-1}.

(e) Is $f^{-1}(f(A)) = A$? Is $f(f^{-1}(B)) = B$?

(f) For each NO answer, does it change to YES for some functions f other than isomorphisms?

Solution sketch. All unexplained answers can be obtained by following one's nose.

(a) Yes for both, the first being a special case. (The "i.e." can be explained formally by induction.)

(b) No for both, e.g., take a constant function and two disjoint subsets of the domain. However $f(\bigcap_\alpha A_\alpha) \subset \bigcap_\alpha f(A_\alpha)$ is always true. Equality holds for all possible collections $\{A_\alpha\}$ if and only if f is injective.

(c) No, equality always fails for any A when f is not surjective. For surjective f, we have $f(A^c) \supset f(A)^c$, but equality will still fail if $f(A)$ and $f(A^c)$ intersect, which will happen if $f(x) = f(x')$ with $x \in A$ and $x' \notin A$. So equality will hold precisely when f is surjective and A is a union of fibers of f.

(d) Yes, yes, yes.

(e) No, e.g., take a constant function and A a nonempty proper subset of S. However $f^{-1}(f(A)) \supset A$ always holds. Equality holds for all A if and only if f is injective. In general $f^{-1}(f(A)) =$ union of fibers over elements of $f(A)$, so for a given f equality holds for those A that are unions of fibers of f.

No, e.g., take a non-surjective function and $B = T$. In general $f(f^{-1}(B)) \subset B$ as $f(f^{-1}(B)) = B \bigcap f(S)$. So for a given f equality holds for $B \subset f(S)$. Equality holds for all B if and only if f is surjective.

Remark about validity of some assertions above only for unions of fibers of f: such considerations become relevant when one studies quotient topology. Unions of fibers of f are called *f-saturated*. Can something interesting be said about an arbitrary f preserving intersections involving f-saturated sets?

Addendum on basic set theory

We will not discuss any more set theory in class for its own sake. Look at Rudin 2.1 to 2.11 to confirm your familiarity with the basics. You may need to understand the definitions in 2.4, especially of (un)countability, and the discussion until 2.8. The results in 2.12 to 2.14 on these notions are very interesting and of major historical/foundational significance, e.g., the number of rational numbers is the same as the number of integers but is strictly less than the number of real numbers. Why? But such things are NOT of primary importance for us in *this* course. We will discuss (un)countability in class if and when it arises. Do not *dwell* on the set theory material. Our approach to it is strictly utilitarian. Another reference: appendices F, O and perhaps P of Erdmann's ProblemText.

Here is a small exercise in language to work out. See Rudin 2.9 for comparison. A "family of subsets of a set S indexed by a set I" means, formally speaking, a function from I to the set of subsets of S. In this context the subset of S that is the image of $i \in I$ is simply denoted S_i.

(a) Define $\bigcup_{i \in I} S_i = \{x \in S \mid \exists\, i \in I \text{ with } x \in S_i\}$. What is the meaning of this when I is the empty set?
(b) Similarly define $\bigcap_{i \in I} S_i$. Deduce its meaning when I is the empty set.