ONE-ONE FUNCTIONS AND INVERSES

MATH 152, SECTION 55 (VIPUL NAIK)

Corresponding material in the book: Section 7.1.

What students should definitely get: The definition of one-to-one function, the computational and checking procedures for checking that a function is one-to-one, computing the inverse of such a function, and relating the derivative of a function to that of its inverse.

What students should hopefully get: The subtleties of domain and range issues, the distinction between the algebraic and the calculus approaches.

Executive summary

0.1. **Vague generalities.** Words...

(1) Old hat: Given two sets A and B, a function $f : A \to B$ is something that takes inputs in A and gives outputs in B. The *domain* of a function is the set of possible inputs, while the *range* of a function is the set of possible outputs. The notation $f : A \to B$ typically means that the domain of the function is A. However, the whole of B need not be the range; rather, all we know is that the range is a *subset* of B. One way of thinking of functions is that *equal inputs give equal outputs*.
(2) A function f is one-to-one if $f(x_1) = f(x_2) \implies x_1 = x_2$. In other words, *unequal inputs give unequal outputs*. Another way of thinking of this is that *equal outputs could only arise from equal inputs*. Or, *knowledge of the output allows us to determine the input uniquely*. One-to-one functions are also called one-one functions or injective functions.
(3) Suppose f is a function with domain A and range B. If f is one-to-one, there is a *unique* function g with domain B and range A such that $f(g(x)) = x$ for all $x \in B$. This function is denoted f^{-1}. We further have that g is also one-to-one, and that $f = g^{-1}$. Note that f^{-1} differs from the reciprocal function of f.
(4) Suppose $f : A \to B$ and $g : B \to C$ are one-to-one functions. Then $g \circ f$ is also one-to-one, and its inverse is the function $f^{-1} \circ g^{-1}$.

Actions ...

(1) To determine whether a function is one-to-one, solve $f(x) = f(a)$ for x in terms of a. If, for every a in the domain, the only solution is $x = a$, the function is one-to-one. If, on the other hand, there are some values of a for which there is a solution $x \neq a$, the function is not one-to-one.
(2) To compute the inverse of a one-to-one function, solve $f(x) = y$ and the expression for x in terms of y is the inverse function.

0.2. **In graph terms.** Thousand words ...

(1) A picture in a coordinatized plane is the graph of a function if every vertical line intersects the picture at most once. The vertical lines that intersect it exactly once correspond to the x-values in the domain. This is known as the *vertical line test*.
(2) A function is one-to-one if and only if its graph satisfies the *horizontal line test*: every horizontal line intersects the graph at most once. The horizontal lines that intersect the graph exactly once correspond to y-values in the range.
(3) For a one-to-one function, the graph of the inverse function is obtained by reflecting the graph of the function about the $y = x$ line. In particular, a function equals its own inverse iff its graph is symmetric about the $y = x$ line.
(4) Many of the results on inverse functions and their properties have graphical interpretations. For instance, the fact that the derivative of the inverse function is the reciprocal of the derivative corresponds to the geometrical fact that reflection about the $y = x$ line inverts slopes of tangent lines.

Similarly, the results relating increase/decrease and concave up/down for a function and its inverse function can all be deduced graphically.

0.3. **In the real world.** Words... (from now on, we restrict ourselves to functions whose domain and range are both subsets of the real numbers)

(1) An increasing function is one-to-one. A decreasing function is one-to-one.
(2) A *continuous* function on an *interval* is one-to-one if and only if it is either increasing throughout the interval or decreasing throughout the interval.
(3) If the derivative of a continuous function on an interval is of constant sign everywhere, except possibly at a few isolated points where it is either zero or undefined, then the function is one-to-one on the interval. Note that we need the function to be continuous *everywhere* on the interval, even though it is tolerable for the derivative to be undefined at a few isolated points.
(4) In particular, a one-to-one function cannot have local extreme values.
(5) A continuous one-to-one function is increasing if and only if its inverse function is increasing, and is decreasing if and only if its inverse function is decreasing.
(6) If f is one-to-one and differentiable at a point a with $f'(a) \neq 0$, with $f(a) = b$, then $(f^{-1})'(b) = 1/f'(a)$. This agrees with the previous point and also shows that the rates of relative increase are inversely proportional.
(7) Two extreme cases of interest are: $f'(a) = 0$, $f(a) = b$. In this case, f has a horizontal tangent at a and f^{-1} has a vertical tangent at b. The horizontal tangent is typically also a point of inflection. It is definitely *not* a point of local extremum. Similarly, if $(f^{-1})'(b) = 0$, then f^{-1} has a horizontal tangent at b and f has a vertical tangent at a.
(8) A slight complication occurs when f has one-sided derivatives but is not differentiable. If both one-sided derivatives of f exist and are nonzero, then both one-sided derivatives of f^{-1} (at the image point) exist and are nonzero. When f is increasing, the left hand derivative of f^{-1} is the reciprocal of the left hand derivative of f, and the right hand derivative of f^{-1} is the reciprocal of the right hand derivative of f. When f is decreasing, the right hand derivative of f^{-1} is the reciprocal of the left hand derivative of f, and the left hand derivative of f^{-1} is the reciprocal of the right hand derivative of f.
(9) The second derivative of f^{-1} at $f(a)$ is $-f''(a)/(f'(a))^3$. In particular, the second derivative of the inverse function at the image point depends on the values of both the first and the second derivatives of the function at the point.
(10) If f is increasing, the sense of concavity of f^{-1} is opposite to that of f. If f is decreasing, the sense of concavity of f^{-1} is the same as that of f.

Actions ...

(1) For functions on intervals, *to check if the function is one-to-one*, we can compute the derivative and check if it has constant sign everywhere except possibly at isolated points.
(2) In order to find $(f^{-1})'$ at a particular point, given an explicit description of f, it is *not* necessary to find an explicit description of f^{-1}. Rather, it is enough to find f^{-1} at that particular point and then calculate the derivative using the above formula. The same is true for $(f^{-1})''$, except that now we need to compute the values of both f' and f''.
(3) The idea can be extended somewhat to finding $(f^{-1})'$ when f satisfies a differential equation that expresses $f'(x)$ in terms of $f(x)$ (with no direct appearance of x).

0.4. **In graph terms.** Thousand words ...

(1) A picture in a coordinatized plane is the graph of a function if every vertical line intersects the picture at most once. The vertical lines that intersect it exactly once correspond to the x-values in the domain. This is known as the *vertical line test*.
(2) A function is one-to-one if and only if its graph satisfies the *horizontal line test*: every horizontal line intersects the graph at most once. The horizontal lines that intersect the graph exactly once correspond to y-values in the range.

(3) For a one-to-one function, the graph of the inverse function is obtained by reflecting the graph of the function about the $y = x$ line. In particular, a function equals its own inverse iff its graph is symmetric about the $y = x$ line.
(4) Many of the results on inverse functions and their properties have graphical interpretations. For instance, the fact that the derivative of the inverse function is the reciprocal of the derivative corresponds to the geometrical fact that reflection about the $y = x$ line inverts slopes of tangent lines. Similarly, the results relating increase/decrease and concave up/down for a function and its inverse function can all be deduced graphically.

1. Warm-up

1.1. **What is/was a function?** Let's recall some of the terminology associated with the concept of functions. A *function* is some thing that allowed you to take certain kind of inputs and spit out certain kinds of outputs, with the main constraint being that *equal inputs give equal outputs*.

The set of permissible inputs for a function is called the *domain* of the function. If the input fed into the function is in the domain, the fnuction processes it and give an output. If the input is not in the domain, the function cannot process it. The set of possible values that the function could spit out was called the *range* of the function. We think of a function as a *black box* that takes an input at one end and emits the output at the other end.

When we say that $f : A \to B$ is a function, what we mean is that the domain of f is A (i.e., f takes as inputs precisely the elements of A) and the range is a *subset* of B. In other words, the notation $f : A \to B$ does *not* imply that everything in B is in the range. This is a useful notational convenience because we would often like to define functions that takes values in some large set (such as the real numbers) without trying to locate the *precise* range.

Another thing that we saw long ago is that a function is *not* the same thing as an expression for the function. There are two aspects to this:

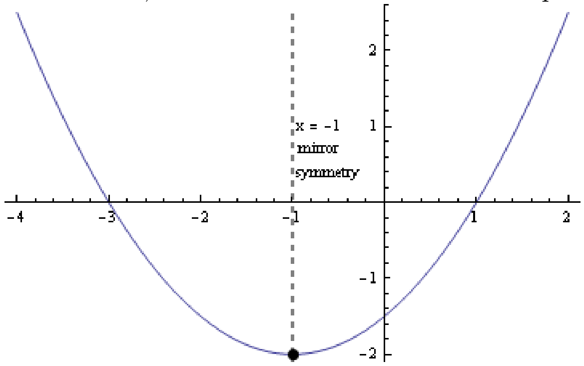
(1) The same *expression* could define different functions, depending on the domain where we are considering the function. For instance, the expression x^2 could be considered on the positive reals, on the negative reals, on the positive integers, on the negative integers, or on the open interval $(0, 1)$. These are all different functions in the technical sense. To avoid this confusion, we posited that if the domain is not explicitly specified or otherwise clear from the context, it is taken to be the largest subset of the real numbers where the expression makes sense (this is the *maximal possible domain*).
(2) Different expressions could specify the same function. For instance, x^2 is the same, as a function, as $2x(x/2)$, even though the literal expressions are different. Similarly, $\sin(\pi x)$ and 0 are the same as functions when restricted to the set of integers.

1.2. **The role of expressions.** There is a fundamental difference between thinking of *functions* and thinking of *expressions*. When we are thinking of a function, we are thinking of a very specific input-output relationship, which may be expressed using an algebraic expression, a table of values, or a graph. The algebraic expression has the advantage of being compact, succinct, and unambiguous, as well as easy to manipulate for many purposes. The graphical expression allows us to use our visual instincts. The table method is something we have been giving short shrift for good reason: most of our functions have infinite domains, and tables just don't work. If you were taking discrete mathematics rather than calculus, we might have been using tables instead of graphs because we were dealing with finite domains.

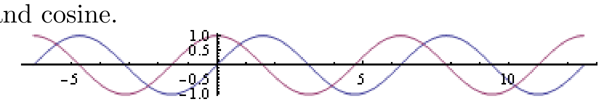
Expressions are useful for a multitude of reasons. With the algebraic expressions, we are able to formally differentiate the function once, twice, and more times. We can calculate its value, the value of its derivatives, find the domain, find the critical points, find the points of inflection, etc., all just starting from a compact formal expression. Another point worth noting is that some of the formal manipulations of expressions can be done even without having any idea of how the graph of the function looks like.

1.3. **Thinking of inverting the function.** Equal inputs for a function give equal outputs, but unequal inputs may give the same output. The extreme example of this is the constant function, whose output is completely indifferent to the input. But there are other examples. For instance:

(1) For *even functions* f, such as the absolute value function, the square function, and the cosine function, we have the relation $f(x) = f(-x)$. Thus, two different inputs give rise to the same output.
(2) More generally, consider a function f with *mirror symmetry* about the line $x = c$. This is a function whose graph is symmetric about the vertical line $x = c$. In particular, we have $f(c+h) = f(c-h)$ for all h. Thus, we have different values of the input giving the same value of the output.



(3) For a *periodic function* with period h, we have the relation $f(x+h) = f(x)$. Thus, two different inputs give rise to the same output. The typical examples are trigonometric functions, such as sine and cosine.



(4) If we have a continuous function with a *local maximum* or a *local minimum*, then there are multiple inputs close to the point of attainment of the local extreme value where the function values are equal.

2. Getting into one-to-one functions

2.1. **One-to-one functions.** A function $f : A \to B$ is termed *one-to-one*, *one-one*, or *injective* if $f(x) = f(y) \implies x = y$. In other words, it is a function having the property that unequal inputs give unequal outputs. Equivalently, we can *reverse* the function in the sense that knowing the output allows us to deduce the input.

Note that this general definition is set-theoretic, and makes sense for functions between arbitrary sets; however, in this course, all functions that we consider are between subsets of reals. So, we are looking at functions $f : A \to \mathbb{R}$ where $A \subseteq \mathbb{R}$.

Next note: Whether a function is one-to-one depends on what domain we are considering for that function. For instance, the squaring function is one-to-one on $[0, \infty)$ but not on the whole real line. Similarly, the greatest integer function is one-to-one when restricted to integers but not on the whole real line. The sine function is one-to-one on the interval $(-\pi/2, \pi/2)$ but not on the whole real line. Of course, if we are just given an *expression* and asked whether the function corresponding to that expression is one-to-one, we consider the domain to be the maximum possible subset of the real line.

2.2. **The horizontal line test.** Consider a function f and now consider the graph of $y = f(x)$. The *horiozontal line test* says that f is one-to-one if and only if every horizontal line intersects the graph of f at most once. Further, the horizontal lines for which the intersection occurs once are precisely those corresponding to the range. This makes sense, because a horizontal line corresponds to a particular value of y, and the intersections with the graph correspond to the values x such that $f(x) = y$.

Remember the *vertical line test*? This states that a given picture arises as the graph of a function if and only if its intersection with every vertical line has at most one point. Further, the vertical lines that intersect it at one point are the vertical lines corresponding to the domain. The rationales behind the vertical line test and horizontal line test are similar.

2.3. **How do we find out if a function is one-to-one? The purely algebraic way.** To determine whether a function is one-to-one, we can use a purely algebraic way – except that it usually doesn't work. We pick two letters, x and a, then write $f(x) = f(a)$ and try to solve algebraically to see if we get a solution with $x \neq a$. For instance, consider the function $f(x) := x^2$. The general equation would be:

$$x^2 = a^2$$

This simplifies to:

$$(x - a)(x + a) = 0$$

This has two solutions: $x = a$ and $x = -a$. The two solutions coincide when $a = 0$ and are distinct otherwise. Thus, the function is *not* one-to-one.

Now consider the function $f(x) := x^3$.

The general expression would be:

$$x^3 = a^3$$

This simplifies to:

$$(x - a)(x^2 + ax + a^2) = 0$$

The second quadratic factor has negative discriminant, so has no real solution, and the only solution is $x = a$. Thus, the function is one-to-one.

What about something more complicated, such as $f(x) := x^3 + x$? We again set $f(x) = f(a)$ and simplify:

$$x^3 + x = a^3 + a$$

Moving everything to one side:

$$(x^3 - a^3) + (x - a) = 0$$

This simplifies to:

$$(x - a)(x^2 + ax + a^2) + (x - a)(1) = 0$$

We combine terms:

$$(x - a)(x^2 + ax + a^2 + 1) = 0$$

The quadratic factor has negative discriminant, so the only solution is $x = a$.

This purely algebraic approach works for quadratic and cubic functions, but it starts getting tedious for more complicated functions. For instance, how do we handle functions such as $f(x) := x - \sin x$? It is hard to solve:

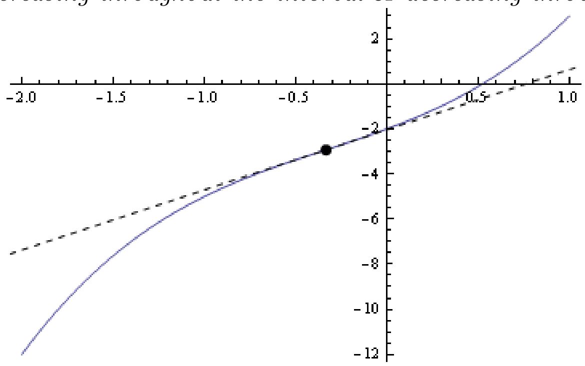
$$x - \sin x = a - \sin a$$

Even for algebraic functions, the approach could be less tractable when the functions are more complicated. This suggests that we need to supplement the *algebraic approach* (with its attendant focus on looking at points of the domain separately) with the *calculus approach* (with its attendant focus/stress on moving along the real line and thinking in terms of limits and continuity). What can the calculus approach tell us that mere algebra cannot?

2.4. **Remember the range computations.** It might be useful to draw a parallel and remember a related generic problem we tackled a while ago. That was the problem of *finding the range*. The *algebraic method* of finding the range of a function f is to set $a = f(x)$ and *solve for* x. We are not interested in actually *finding* a solution – we are interested in determining the conditions on a such that *at least* one solution exists. For instance, for linear and quadratic polynomials and rational functions of small degree, this often reduces to a condition on the discriminant of a quadratic polynomial.

That was the algebraic approach, and it was limited to a small number of functions that were algebraically tractable. But then we saw the calculus approach, which essentially allowed us to graph any reasonably nice function. Once we have the entire graph, we can find the range. For a continuous function, this is simply the interval between the minimum value of the function and the maximum value of the function. This allowed us to determine the range of a much larger class of functions, particularly those that are continuous and once or twice differentiable.

2.5. **The calculus interpretation of one-to-one.** Consider a *continuous* function f on a (possibly open, closed, half-open, half-closed, or infinite) interval I. Continuous means that the function cannot jump about suddenly. Under what conditions is the function one-to-one? Clearly, if it changes direction somewhere, i.e., has a local extreme value, then there are horizontal lines close by that intersect the graph at two points. Thus, there are no local extreme values. Or, another way of putting this is that the function must either be *increasing throughout the interval* or *decreasing throughout the interval*.



Thus, for a continuous function on an interval, being one-to-one is equivalent to being increasing throughout or decreasing throughout. *If you change direction, you repeat points.* Remember that both parts of the statement: *continuous* and *interval*, are needed in order to conclude that a one-to-one function *must* be increasing throughout or decreasing throughout. However, the *other direction of implication* is always true: a function that is increasing throughout on its domain is one-to-one, and so is a function that is decreasing throughout on its domain.

Let's see some counterexamples:

(1) Can you think of a discontinuous function on an interval that is one-to-one but not increasing or decreasing?
(2) Can you think of a function on a set that is a union of two or more intervals that is continuous on each piece and is one-to-one but is not increasing or decreasing when viewed on the whole domain? The problem is that although we can deduce that the function is increasing throughout or decreasing throughout separately on each of the intervals, we cannot compare across intervals.

2.6. **The proof that a continuous function on an interval is one-to-one iff its increasing or decreasing.** Let us now try to prove the statement that a continuous function on an interval is one-to-one if and only if it is either increasing throughout on the interval or decreasing throughout on the interval. Note that increasing throughout or decreasing throughout obviously implies one-to-one, so we concentrate on proving the other direction of implication.

Suppose f is a continuous function on an interval I and there are points x_1, x_2, x_3 with $x_1 < x_2 < x_3$ and $f(x_1) < f(x_2) > f(x_3)$. Then, suppose M is a number greater than both $f(x_1)$ and $f(x_3)$ but less than $f(x_2)$. By the intermediate value theorem, there exists $x_4 \in (x_1, x_2)$ and $x_5 \in (x_2, x_3)$ such that $f(x_4) = f(x_5) = M$. This forces f to *not be one-to-one*, a contradiction. Thus, we cannot have a situation where $x_1 < x_2 < x_3$ but $f(x_1) < f(x_2) > f(x_3)$. A similar argument shows that we cannot have $x_1 < x_2 < x_3$ but $f(x_1) > f(x_2) < f(x_3)$. This forces f to be increasing throughout or decreasing throughout.

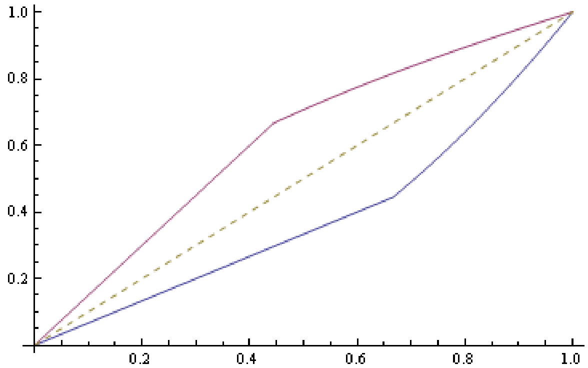
Note that to apply the intermediate value theorem, we applied both the continuity of f and the fact that the domain is an interval, hence contains $[x_1, x_2]$ and $[x_2, x_3]$.

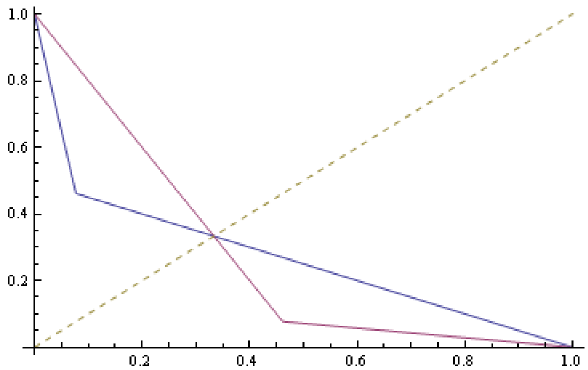
3. A two-way street: inverse functions

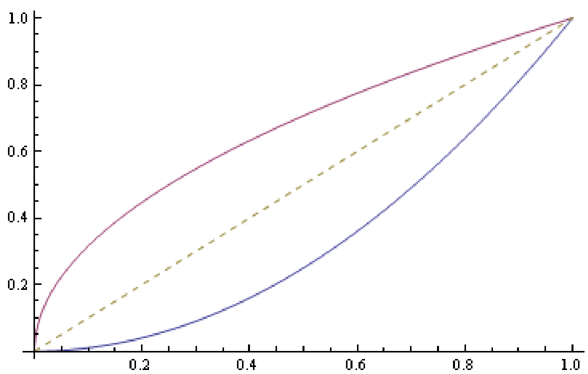
3.1. **The inverse function.** Suppose we have a function f with domain A and range B. If f is one-to-one, we can define an *inverse function* g such $f \circ g$ is the identity map on B. Moreover, this function g is unique. This function is called the *inverse function* to f, since it *reverses* or *inverts* the action of f. Note that here, we really do need B to be precisely the range, because we cannot define the function g on points outside the range.

This function g is denoted as f^{-1} and is termed the *inverse function* to f. Note that this is not the same as the *pointwise multiplicative inverse* of f, which is the function $1/f(x)$. The latter may be denoted as

$[f(x)]^{-1}$ or as $1/f$, as opposed to $f^{-1}(x)$. I might also use the word *reciprocal function* for the pointwise multiplicative inverse.







What happens if f is not one-to-one? In this case, there are many different candidates for g that work, and it is not clear which one to pick. To avoid confusion, we do not talk of *the inverse function* any more. For instance, when f is the squaring function on the reals, we could take g to be the positive square root or the negative squareroot, or to sometimes be the positive squareroot and sometimes the negative squareroot. For instance, one candidate would be a function that is the positive square root for nonnegative rationals and the negative square root for positive irrationals.

This is a very rich and deep question that we shall return to later, when we study things such as inverse trigonometric functions.

3.2. **The inverse operation is involutive.** The operation of taking the inverse is an *involutive* operation, in that it has the following two properties:

(1) $(f^{-1})^{-1} = f$. In other words, if $g = f^{-1}$, the $f = g^{-1}$.
(2) $(f_1 \circ f_2)^{-1} = f_2^{-1} \circ f_1^{-1}$. In other words, the inverse of the composite is the composite of the inverses, but the sequence of composition flips over.

You'll be asked to show this in a forthcoming homework.

3.3. **Finding the inverse function: like finding the range.** As we just recalled, to find the range of a function f, we consider the equation $y = f(x)$ and solve for x in terms of y. When we were trying to compute the range, our sole purpose was to find the set of y for which there exists at least one value of x that solves the equation. When our goal is to determine the inverse function, we are interested in the actual *expression* for x in terms of y since that is the inverse function.

Note that in this process, we can discard the values of y for which *no solution exists*. But if we find that for some y, there are multiple values of x, then we've gone down a bad path: the function wasn't one-to-one, so we shouldn't have been trying to find an inverse at all.

3.4. **Situations where the algebraic procedure works.** It works for nonconstant linear functions. Given a function $y = mx + c, m \neq 0$, we can rewrite $x = (y - c)/m$. It also works for functions of the form $y = x^{p/q}$ where both p and q are odd integers. The inverse function to $y = x^{p/q}$ is $y = x^{q/p}$. And it works for functions that are obtained by composing such power functions and linear functions. For instance, see Example 3 in the book.

3.5. **Graphical interpretation of inverse function.** Suppose g is the inverse function of a one-to-one function f. For every point (x, y) in the graph of f, we have $y = f(x)$, hence $x = g(y)$. Thus, the point (y, x) is in the graph of g. In other words, the graph of g is obtained by taking the graph of f and sending each point to the point obtained by interchanging its coordinates. The *coordinate interchange operation* is equivalent to the geometric operation of reflection about the $y = x$ line. Thus, the graph of g is obtained by reflecting the graph of f about the $y = x$ line. This geometrical interpretation is useful for understanding the relationship between derivatives.

It also helps us identify a new kind of symmetry that some functions possess. The graph of a function f is symmetric about the $y = x$ line iff $f = f^{-1}$. Examples of such functions are $y = x$, $x + y = C$ for some constant C, and *implicit* functions given by $p(x) + p(y) = C$ where p is a one-to-one function from $\mathbb{R}$ to $\mathbb{R}$. For instance, $x^3 + y^3 = 1$ is an implicit description of y as a function of x – the explicit description is $y = (1 - x^3)^{1/3}$, and the inverse is exactly what we'd expect.

3.6. **The inverse of a continuous function is continuous.** From the previous result about reflection, it should be reasonably intuitive that the inverse of a continuous function is continuous. Proving this using the $\epsilon - \delta$ definition is a nice exercise, but not one that we shall undertake in class. This is Theorem 7.1.7 of the book, and you are encouraged to read the proof for your understanding and also to refresh your memories of ϵs and δs.

There is something qualitative that we can say, though, that will shed some light. Remember that the general $\epsilon - \delta$ definition is *not* symmetric in the roles of x and $f(x)$. The skeptic starts with choosing an $\epsilon > 0$ which determines an open interval around the claimed limit for the $f(x)$-value. The prover then has to come up with an interval around the domain value (of radius δ) such that the function value is within the ϵ-interval for every input value in the δ-interval. This definition is asymmetric, because we insist that if the x-value is really close, the $f(x)$-value is also really close, but we do not insist that if the $f(x)$-value is really close, the x-value is also really close.

For one-to-one functions, however, this inherent definitional asymmetry is automatically overcome, because a given $f(x)$-value can be realized only by one x-value. This is the reason that, even though the definition is not inherently symmetrical, the one-to-one nature allows us to show that it is in effect symmetric in the roles of domain and range.

4. Bijective functions and infinite sets

This is a concept you will see in somewhat more detail if you take higher mathematics courses, so I'll briefly mention it here.

A function $f : A \to B$ is termed a *bijection* or *bijective function* from A to B if its range is precisely B and it is one-one. Bijective functions are the same thing as one-one functions considered as functions *to their range* and ignoring the rest of the stuff. The notion of inverse function that we introduced is best viewed in the context of a bijective function, because the inverse is defined only on the range of the function.

Note that the theorem proved earlier basically states that for a continuous bijective function, the inverse is also continuous.

An interesting question now might be: can we define continuous bijective functions between various typical infinite subsets of $\mathbb{R}$? We note some obvious positive results in this direction:

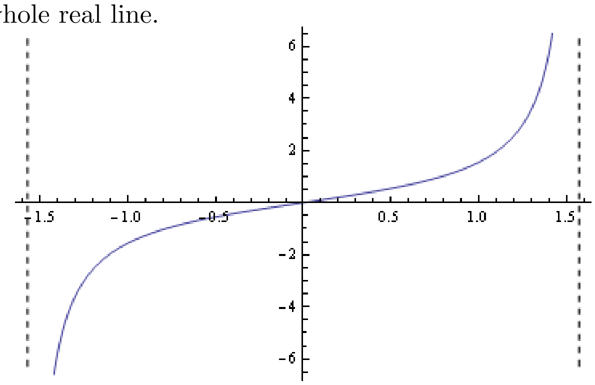
(1) For the open intervals (a, b) and (c, d), where $a < b$ and $c < d$ are real numbers, there is a *linear* bijection between the open intervals. Namely, there is a unique linear function that sends a to c and b to d, and this function gives a bijection of the intervals in between. We could also pick the linear map that would send a to d and b to c.

   In particular, this bijection is continuous and infinitely differentiable, since it involves only translation and scaling.

(2) Interestingly, there is a bijection between any finite open interval and an open interval going to infinity in one or both directions. Since (1) above shows that any two finite open intervals look the same, we just give one bijection of each kind.

   The map $x \mapsto \tan x$ gives a bijection between the open interval $(0, \pi/2)$ and the one-sided infinite open interval $(0, \infty)$.

   The map $x \mapsto \tan x$ also gives a bijection between the finite open interval $(-\pi/2, \pi/2)$ and the whole real line.



Thus, for infinite sets, small finite open intervals can be in bijection with larger finite open intervals and even with infinite open intervals. The fact that a subset of a set can be in bijection with the whole set perturbed the mathematician Cantor when he first discovered and pondered about it. He eventually went crazy, but before doing so, made a key observation: being in bijection with a proper subset is a *defining characteristic* of infinite sets.

5. A new process and new responsibilities

So far, we have seen the following processes for creating new functions from old:

(1) Pointwise addition, subtraction, multiplication, and division.
(2) Function composition.
(3) Piecing together different definitions (using piecewise definitions).

We have now added a new process: inverting a one-to-one function. Hence, it is our job to now describe how to do all the things we used to do in the past for new functions created from old functions using this new process.

5.1. The derivative of a one-to-one function and its inverse. uppose we have a one-to-one function f on an interval I with inverse function g. The intermediate value theorem tells us that the range of f (and hence the domain of g) is also an interval. If the domain of f is a closed bounded interval, the extreme value theorem tell us that the range of f (and hence the domain of g) is also closed and bounded.

Since f is increasing throughout or decreasing throughout, what can we say about its derivative (assuming it exists everywhere)? If f is differentiable, then:

(1) f is increasing throughout iff f' is positive everywhere except possibly at isolated points, where it can be zero.
(2) f is decreasing throughout iff f' is negative everywhere except possibly at isolated points, where it can be zero.

Thus, a differentiable f is one-to-one iff f' is of constant sign throughout except possibly at isolated points, where it can be zero. Examples are x^3 and $x - \sin x$.

More generally, if f' has constant sign everywhere except at isolated points where it is either zero or undefined, f is one-to-one.

Let us now bring g into the picture. It turns out that the following is true: if f is one-to-one, and g is its inverse, then if $f(a) = b$, and $f'(a)$ exists and is nonzero, then $g'(b) = 1/f'(a)$. Thus, we have the general formula:

$$g'(x) = \frac{1}{f'(f^{-1}(x))}$$

Equivalently:

$$g'(f(x)) = \frac{1}{f'(x)}$$

This can be seen in three ways:

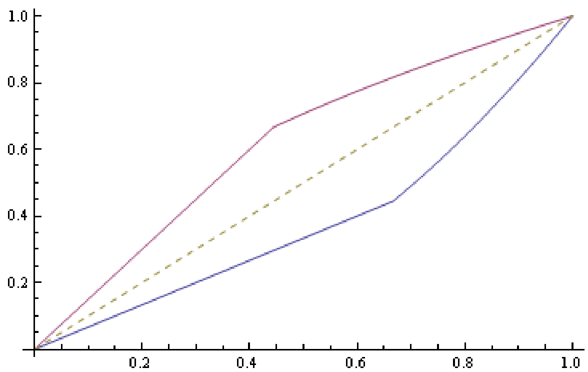
(1) *From first principles*: The difference quotient whose limit gives the value of f' is the reciprocal of the difference quotient whose limit gives the value of g'.
(2) *Using the chain rule*: Use that $f \circ g$ is the identity map and hence obtain that $(f' \circ g) \cdot g' = 1$.
(3) *Graphically*: We know that the graphs of f and g are reflections of each other about the line $y = x$, and the points we are interested in are images of each other under this reflection. The tangent lines through these points are thus also reflections of each other about $y = x$. The slopes of these tangent lines are thus reciprocals of each other.

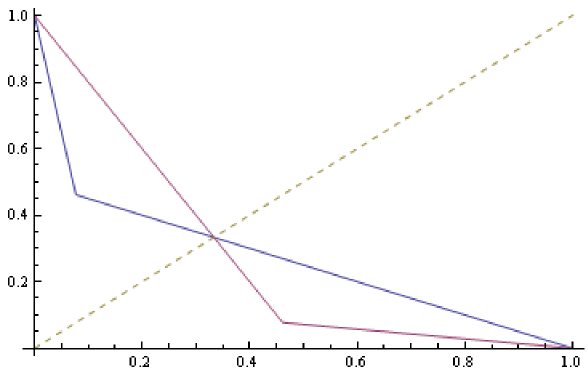
Let $f(a) = b$. There are the following cases of interest:

(1) $f'(a)$ exists and is nonzero. This happens if and only if $g'(b)$ exists and is nonzero, and $f'(a)$ and $g'(b)$ are multiplicative inverses of each other. Pictorially, this means that the tangent lines for the graphs of f and g are neither vertical nor horizontal. $f'(a)$ is positive iff $g'(b)$ is positive, in which case both f and g are locally increasing. $f'(a)$ is negative iff $g'(b)$ is negative.
(2) $f'(a)$ exists and is equal to zero: In this case, $g'(b)$ is undefined. The graph of f has a horizontal tangent at the point a, and the graph of g has a vertical tangent. Moreover, we can deduce from the one-to-one nature of f that the horizontal tangent for f cannot be a local extreme value type – hence (with suitable further differentiability assumptions) it must be the *point of inflection* type.[1]
(3) $g'(b)$ exists and is equal to zero: In this case, $f'(a)$ is undefined. The remarks of the previous point apply with the roles of f and g interchanged, as well as the roles of a and b.
(4) Both the left-hand and the right-hand derivative for f exist at a, but they are not equal: In this case, the left-hand derivative and the right-hand derivative exist for g. Further, if f is increasing, so is g, in which case the left-hand derivative of g is the multiplicative inverse for the left-hand derivative of f, and the right-hand derivative of g is the multiplicative inverse of the right-hand derivative of f. If f is decreasing, so is g, in which case the left-hand derivative of g is the multiplicative inverse of the right-hand derivative of f, and the right-hand derivative of g is the multiplicative inverse of the left-hand derivative of f.

Re-read that last point a few times till you understand it. The interplay between one-sidedness and increase/decrease behavior is extremely important and potentially confusing. Here are some pictures:

[1] The tangent line cutting through the graph is the typical geometric description of a point of inflection; however, it is not strictly correct since there do exist weird situations where we have a point that is not a point of inflection but the tangent line still cuts through the graph. Nonetheless, this is the typical case to keep in mind.





5.2. **Full details of the difference quotient derivation.** To understand this proof, it is helpful to recall that there are two different ways of thinking about functions and about differentiation. The first, which is the typical way, is to think about a function as a machine that takes in an input and gives out an output. There is another, slightly different, way of thinking about functions. Here, the focus is not on the function but on the input and the output. We think of the function as the process relating the input quantity and the output quantity. For instance, we may think of the position x of a particle as a function of time t. Here, the *output quantity* is viewed as a function of the input quantity.

When we switch back and forth between these two ideas of functions, there is a slight abuse of notation. For instance, when we are trying to write the position of a particle as a function of time, we often use the same letter x for the position *function* $x(t)$ and for the actual position variable. This is a bit like saying that instead of writing a function $y = f(x)$, we write $y = y(x)$. This really is an abuse of notation, but it is an abuse that comes with some advantages. For instance, in the book's description of the u-substitution, the book used the letter u both for the function and the variable name.

Recall now that for a function f, the *difference quotient* between the input values x_1 and x_2 is the value:

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

If we write $y_1 = f(x_1)$ and $y_2 = f(x_2)$, we can rewrite this as:

$$\frac{y_2 - y_1}{x_2 - x_1}$$

which can be written in shorthand as:

$$\frac{\Delta y}{\Delta x}$$

With the interpretation as a relationship between quantities, we are interested in the question of how much a specific change in x-values leads to a change in the y-values. The limit of this is at $x = x_0$ is defined as the derivative $f'(x_0)$. With this notation, we also see that:

$$\frac{\Delta y}{\Delta x} = \frac{1}{\frac{\Delta x}{\Delta y}}$$

Thus, we see the intuitive reason why, when we pass to the limits, we should get that dy/dx and dx/dy are multiplicative inverses at any particular pair (x, y).

Let us make this formal. Suppose f is one-to-one. Then, for any a with $f'(a)$ finite and with $f(a) = b$, we have:

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{y \to f(a)} \frac{y - f(a)}{g(y) - a} = \lim_{y \to b} \frac{y - b}{g(y) - g(b)} = \frac{1}{g'(b)}$$

This explains why $f'(a)$ and $g'(b)$ are multiplicative inverses of each other.

The multiplicative inverse relationship can also be verified using the chain rule. Here, we use the fact that $f \circ g$ is the identity map, and apply the chain rule to get:

$$(f \circ g)'(b) = f'(g(b))g'(b) \implies 1 = f'(a)g'(b)$$

5.3. **Higher derivatives.** Recall that we can compute higher derivatives as well for various ways of creating new functions from old. For sums, differences, and scalar multiples, the rule is simple: since differentiation is a linear operator, the k^{th} derivative of the sum/difference/scalar multiple is the sum/difference/scalar multiple of the k^{th} derivatives. For products, the k^{th} derivative, as we saw in some quizzes, has a binomial formula, which we can discover by iteration. In particular, for instance:

$$\begin{aligned} (f \cdot g)' &= (f' \cdot g) + (f \cdot g') \\ (f \cdot g)'' &= (f'' \cdot g) + 2(f' \cdot g') + (f \cdot g'') \\ (f \cdot g)''' &= (f''' \cdot g) + 3(f'' \cdot g') + 3(f' \cdot g'') + (f \cdot g''') \end{aligned}$$

The story is trickiest for composites, where, in order to compute the second derivative of a composite, we need to use the chain rule *and* the product rule. The formula we get, which you saw in a past quiz, was:

$$(f \circ g)'' = (f'' \circ g) \cdot (g')^2 + (f' \circ g) \cdot (g'')$$

We have to do something similar to calculate the second derivative of the inverse function. However, this time we need to use the *quotient rule*. Note that:

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$$

To find the second derivative, we must differentiate both sides and use the quotient rule or equivalently, the rule for differentiating a reciprocal function. The upshot is that we get:

$$(f^{-1})''(x) = \frac{-f''(f^{-1}(x))}{(f'(f^{-1}(x)))^3}$$

You'll be working out the full details of this in a homework problem.