Chapter 13

The Differentiation Rules via Limits

13.1 The Limit Rules (Theorems)

> *"In the old days when people invented a new function they had something useful in mind. Now, they invent them deliberately just to invalidate our ancestors' reasoning, and that is all they are ever going to get out of them."*
>
> – Henri Poincaré (1854 – 1912)

In this section we will state several theorems about limits which we will need in the sections following. The limit concept is very subtle and our understanding of it is still quite intuitive. We are not yet quite prepared to prove these theorems so we will leave these theorems unproven for now. Our immediate goal is simply to understand what they say and learn how to use them. In the next section we will begin using these theorems to show how the limit in Definition #15 allows us to recapture all of the major results we used in Part I of this text.

In Chapter 16 we will finally discard our intuitive definition of a limit (Definition #12) and formally define both a limit at infinity (Theorem 26) and a limit at a point (Theorem 27). Then we will return to the theorems in this section and (finally) prove rigorously that they are, in fact, true. Until then any result which relies on the theorems in this section should be considered contingent.

In Chapter 11 we stated the following three theorems about limits "at infinity."

1. the limit of a sum is the sum of the limits (Theorem 8),
2. the limit of a product is the product of the limits (Theorem 9) and,
3. the limit of a quotient is the quotient of the limits (Theorem 10).

All three of these theorems remain true if x is approaching some finite number, a, instead of infinity.

Theorem 15: **[The Limit of a Sum is the Sum of the Limits]**
Let a be some real number. Suppose that the functions $f(x)$ and $g(x)$ are defined on some open interval about a except, possibly, at a itself.

Then if $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$, then

$$\lim_{x\to a} [f(x) + g(x)] = L + M = \lim_{x\to a} f(x) + \lim_{x\to a} g(x).$$

Theorem 16: **[The Limit of a Product is the Product of the Limits]**
Let a be some real number. Suppose that the functions $f(x)$ and $g(x)$ are defined on some open interval about a except, possibly, at a itself.

If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$ then

$$\lim_{x\to a} (f(x) \cdot g(x)) = \left(\lim_{x\to a} f(x)\right) \cdot \left(\lim_{x\to a} g(x)\right) = L \cdot M.$$

Theorem 17: **[The Limit of a Quotient is the Quotient of the Limits]**
Let a be some real number. Suppose that the functions $f(x)$ and $g(x)$ are defined on some open interval about a except, possibly, at a itself.

Then if $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M \neq 0$ then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{L}{M}.$$

Notice that in addition to changing ∞ to some real number, a, we have added two qualifications to the statement of each of these theorems from Chapter 11:

1. "Suppose that the functions $f(x)$ and $g(x)$ are defined on some open interval about a" and,

2. "except, possibly, at a itself"

To see why these are necessary recall that we're going to use limits to define the derivative as in Theorem 15 so we'll need to evaluate the limit $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$. Clearly the expression $\frac{f(x+h)-f(x)}{h}$ is not defined at $h = 0$. But we're only interested in its value **in the limit** as $h \to 0$ which means that h must be able to get close to 0. That is, there must be an open interval around 0 where the expression $\frac{f(x+h)-f(x)}{h}$ is defined.

But we don't care if $\frac{f(x+h)-f(x)}{h}$ is defined at $h = 0$ or not. It is irrelevant to our purpose. So we state explicitly that we do not consider whether $h = 0$.

Ok, but why did we insert the word "possibly"? Wouldn't it be enough to simply say "except at a"?

We need to say "possibly" because these theorems, like all theorems, are stated with as much generality as possible. For example, consider the function $f(x) = 2x$. Had we not included "possibly" in the conditions of our theorems the limit: $\lim\limits_{x\to 3} 2x$, which is clearly equal to 6, would have to be considered undefined because $f(x) = 2x$ is defined at $x = 3$. This distinction[1] may seem like a very fussy, and unimportant detail right now, but it will be important when we discuss the meaning of continuity in Section #13.1.

Example #82:

Suppose

$$f(x) = \begin{cases} 2x & \text{if } x \neq 3 \\ 10 & \text{if } x = 3 \end{cases}.$$

Then

$$\lim_{x\to 3} f(x) = 6.$$

In particular the limit is not 10.

Here is how we would evaluate this limit using the tools we currently have at our disposal.

We're interested in the limit as $x \to 3$ so in particular we do not need to consider the case when $x = 3$. But as long as $x \neq 3$ we have $f(x) = 2x$ so

$$\lim_{x\to 3} f(x) = \lim_{x\to 3} 2x.$$

As x gets close to 3 it is clear that $2x$ gets close to 6. Therefore

$$\lim_{x\to 3} f(x) = 6.$$

Notice that our reasoning is a little vague in the last step because we had to resort to the phase "gets close to," and we know from our work in Chapter 11.4 that this is not a precise phrase. This is the best we can do now because we have not yet rigorously defined a limit. We will do that in Chapter 16.

End Of Example #82

Problem #486:

By reasoning in a manner similar to Example #82 show that $\lim\limits_{x\to 3} f(x) = 9$ for each function.

[1] In the eighteenth century there was a, public, protracted, and vitriolic argument between Benjamin Robins and James Jurin over exactly this point. Jurin would have claimed that the statement $\lim\limits_{x\to 3} 2x = 6$ is meaningless. Robins would have said it has meaning because it is obviously true. The point here is not that either man was right or wrong, but rather that it depends on how we define limits. By one definition Robins was correct, by another Jurin was. Their controversy was the result of the incomplete understanding of limits that prevailed at the time.

(a) $f(x) = x^2$

(b) $f(x) = \begin{cases} x^2 & \text{if } x \neq 3 \\ \text{undefined} & \text{if } x = 3 \end{cases}$

It will be tedious to write (and to read) the phrase "Suppose that $f(x)$ is defined on some open interval about a except, possibly, at a itself" every time we need it so it is customary to say something more like "Suppose $f(x)$ is defined **near** a" instead. Because we are trying to be as precise, and rigorous as possible we will formalize this by redefining the word **near**.

Definition 16: Near
We say that $f(x)$ has some property **near**[2] $x = a$ if $f(x)$ has that property on an open interval about $x = a$, except possibly at a itself.

We have the following theorem.

Theorem 18: [The Limit of a Constant is the Constant]
Suppose a and L are real numbers, and $f(x) = L$ near a. Then

$$\lim_{x \to a} f(x) = L.$$

Notice that Theorem 18 would be considerably less useful had we not required that $f(x) = L$ for x near a, rather than $f(x) = L$ on its entire domain. For example, as stated Theorem 18 allows us to conclude that if

$$H(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

then $\lim\limits_{x \to 5} H(x) = 1$ and $\lim\limits_{x \to -5} H(x) = -1$ because there are open intervals about 5 and -5 where $H(x) = 1$, and $H(x) = -1$, respectively.

Since there is no such interval about $x = 0$, $\lim\limits_{x \to 0} H(x)$ is undefined.

Drill #487:

(a) Find an open interval about 5 where $H(x) = 1$.

(b) Find an open interval about -5 where $H(x) = -1$.

[2]Notice that this is not what "near" means in ordinary speech. This is one of the things that makes it difficult to read mathematics. We routinely co-opt words from natural languages (like English) and redefine them to fit our needs. In this case our purpose requires that we change the definition of "near" slightly as you've seen. Because "near" is a common word and you have a lifetime of experience using it, it can be very difficult to cast off your preconceptions. The familiar definition you learned in childhood will intrude and cause confusion. It is hard to overcome this. Refer back to the definition frequently until you have internalized the mathematical definition.

(**Comment:** There are many to choose from in both parts. Choose only one for each.)

Drill #488:
Suppose

$$f(x) = \begin{cases} 1; & \text{if } x > 1 \text{ or if } x = -2 \\ -2; & \text{if } x \leq 1 \text{ and } x = -2. \end{cases}$$

Determine whether the following statements are true or false.

(a) $f(x) = 1$ near $x = 4$.

(b) $f(x) = 1$ near $x = 1$.

(c) $f(x) = -2$ near $x = 1$.

(d) $f(x) = 1$ near $x = -2$.

(e) $f(x) = -2$ near $x = -2$.

(f) $f(x) = 0$ near $x = 0$.

Problem #489:
Explain, that the following statements are true by citing Theorem 15 through Theorem 18 as needed.

(a) $\displaystyle\lim_{x\to 0}\left(\frac{5x}{x} + \frac{\pi x}{x}\right) = 5 + \pi$

(b) $\displaystyle\lim_{x\to 2}\left(\frac{x-2}{x-2} + \frac{3x-6}{x-2}\right) = 4$

(c) $\displaystyle\lim_{x\to 5}\left(2x + 3x^2\right) = 85$

(d) $\displaystyle\lim_{x\to -1}\left(\frac{x^2-1}{x+1} + \frac{x^2+3x+2}{x+1}\right) = -1$

Problem #490:
Notice that neither $\lim_{x\to 0}\frac{1}{x}$ nor $\lim_{x\to 0}\left(\frac{-1}{x}\right)$ exists. However their sum,

$$\lim_{x\to 0}\left(\frac{1}{x} + \frac{-1}{x}\right) = \lim_{x\to 0} 0 = 0$$

does exist. Explain why this does not contradict Theorem 15.

Problem #491:
Suppose $g(x) \neq 0$ near $x = a$ and $\lim_{x\to a}\frac{f(x)}{g(x)}$ exists. Use Theorem #16 to show that if $\lim_{x\to a} g(x) = 0$ then $\lim_{x\to a} f(x) = 0$.
(**Hint:** Consider $f(x) = \frac{f(x)}{g(x)} \cdot g(x)$ for x near a.)

(**Comment:** This problem shows that if $\lim_{x\to a} g(x) = 0$ then the only way that $\lim_{x\to a} \frac{f(x)}{g(x)}$ can exist is if we have a L'Hôpital Indeterminate.)

The following Corollary says that if $f(x)$ is approaching L_f and we multiply $f(x)$ by a number, k, then the product $kf(x)$ approaches kL_f. It follows from Theorem 18 and Theorem 16.

Corollary 19: If $\lim_{x\to a} f(x) = L_f$ and k is a real number then $\lim_{x\to a} kf(x) = k \lim_{x\to a} f(x) = kL_f$.

Drill #492: Prove Corollary 19.

The Limit of a Composition and Continuity at a Point

The concept of **continuity** is essential to Calculus, but you may have noticed that we have carefully avoided it as much as possible until now. This is because defining continuity is similar to defining the line tagent to a curve (Definition #2). We need to think carefully about what we want the term **continuous** to mean, and then craft our definition to capture that meaning. This would have been very difficult to do without a fairly sophisticated understanding of the limit concept.

So stop and think about this for a moment. What do we mean when we say a curve is **continuous**? A first, intuitive definition usually goes something like this: "A function is continuous if you can draw its graph without lifting your pencil from the paper," but this is unsatisfactory for a number of reasons. In particular, it is impossible to apply in most cases. Think about it. Have often have you seen the *entire* graph of any function? Usually we just draw the part neat the origin and put arrowheads on both ends of the graph. We need something more precise.

At the end of Example #79 we remarked that it is only when $f(x)$ is continuous at $x = g(a)$ that $\lim_{x\to a} f(g(x))$ is equal to $f\left(\lim_{x\to a} g(x)\right)$, but we did not discuss the matter any further. It is time for that discussion.

First, notice that when you think closely about the statement "$f(x)$ is continuous at $g(a)$" it appears to be nonsense, because $g(a)$ is the value of f at the single value $g(a)$ but . Does it make sense to you that a curve can be continuous at a single value of its domain? In ordinary usage the concept of continuity requires an interval to be continuous on, doesn't it?

Since we need the concept of "continuity at a point," we define it.

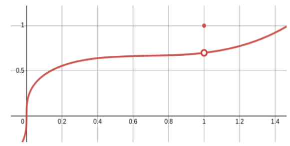
Definition 17: Continuity at a Point

A function f, whose domain is an interval in $\mathbb{R}$, is continuous at $x = a$ in the interval, if and only if $\lim_{x\to a} f(x) = f(a)$, (alternatively, if $\lim_{h\to 0} f(a+h) = f(a)$).

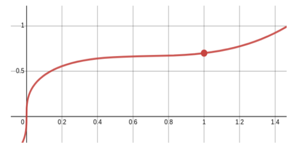
If f is continuous at every point in its domain we'll just call it a continuous function.

The sketch at the right shows that Definition #17 recovers the intuitive notion that a function is continuous if we can draw its graph without lifting pen from page. Both of the functions, $f(x)$ and $g(x)$ are identical everywhere except at $x = 1$. Clearly, to draw $g(x)$, which is discontinuous at $x = 1$, we must lift our pen from the page. This is not true of the graph of $f(x)$, which is continuous.

The following lemma is true and the proof will be valid once the limit theorems have been proven in Chapter 16.



$g(x)$



$f(x)$

Lemma 20: [**Differentiability Implies Continuity**]
If $f(x)$ is differentiable at $x = a$ then $f(x)$ is also continuous at $x = a$.

Proof: We assume that $f(x)$ is differentiable at $x = a$. Then

$$\begin{aligned}
\lim_{h\to 0}(f(a+h) - f(a)) &= \lim_{h\to 0}\left[(f(a+h) - f(a)) \cdot \frac{h}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{f(a+h) - f(a)}{h} \cdot h\right] \\
&= \left[\lim_{h\to 0}\frac{f(a+h) - f(a)}{h}\right] \cdot \left[\lim_{h\to 0} h\right] \\
&= f'(a) \cdot 0 \\
&= 0.
\end{aligned}$$

Therefore $\lim\limits_{h\to 0} f(a+h) = f(a)$ and therefore f is continuous at a. ■

Problem #493:
Justify each step in the proof of Lemma 20 by citing the appropriate Theorem or by filling in the appropriate algebraic steps.

When we were studying horizontal asymptotes in section 11.4 we encountered Theorem 11 (the Squeeze Theorem "at" Infinity). But the Squeeze Theorem is also valid if $x \to a$, where a is a real number.

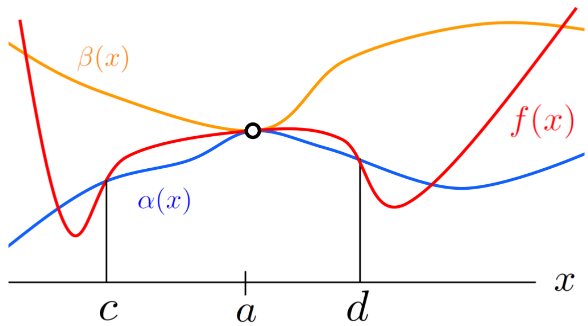
Theorem 21: [**The Squeeze Theorem (The Finite Case)**]

If $\alpha(x) \le f(x) \le \beta(x)$ for x near a and

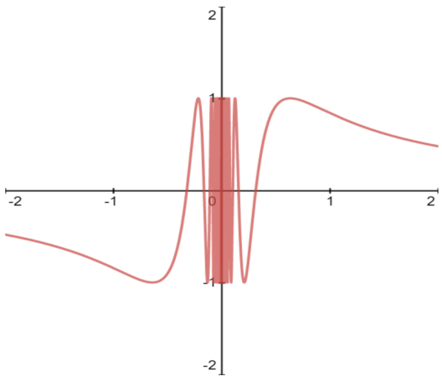
$$\lim_{x\to a} \beta(x) = \lim_{x\to a} \alpha(x) = L$$

then $\lim_{x \to a} f(x) = L$ also.

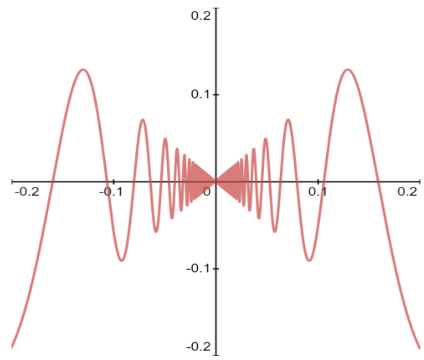
Theorem 21 is illustrated below, but a formal proof will not be given until Chapter 16.



Problem #494:
Consider the two functions defined in the sketch below:



$$T(x) = \begin{cases} \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \end{cases}$$



$$U(x) = \begin{cases} x\sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

(a) Use Theorem 21 to show that $U(x)$ is continuous at $x = 0$.
(**Hint:** What functions is $U(x)$ caught between?)

(b) Use Definition 17 to show that $T(x)$ is not continuous at $x = 0$.
(**Hint:** Try the substitution $z = \frac{1}{x}$ for $x \neq 0$. What would $\lim_{x \to 0} T(x)$ look like in terms of z?)

Theorem 22: **[The Limit of a Composition is the Composition of the Limits]**
Suppose $\lim_{x \to a} g(x) = L_g$ and that $f(x)$ is continuous at L_g. Then

$$\lim_{x \to a} f(g(x)) = f\left(\lim_{x \to a} g(x)\right) = f(L_g).$$

Essentially this says that we can interchange the function f and the "lim" symbols if f is continuous at $g(a)$.

13.2 The General Differentiation Theorems, via Limits

". . . one way in math to take care of destabilizing problems is to legislate them out of existence . . . by loading theorems with stipulations and exclusions designed to head off crazy results."

– David Foster Wallace (1962 - 2008)

Since we will now be proving the the differentiation rules rigorously we will call them what they really are: Theorems.

Because limits are much less intuitive than differentials we'll want to be as efficient as possible when using them. The sooner we can build up some tools to make things easier, the better.

Also, in this section we will add a new differentiation rule (theorem): The Chain Rule. Or rather, we will give a name to an already familiar technique and elevate it's status by providing a formal proof. Proving the Chain, Product, and Quotient Differentiation Rules using limits will require a good deal of cleverness. These proofs will also uncover some unexpected subtleties along the way.

Before we begin there is one more point that needs to be clear. Because differentiation is now defined via a limit and limits are defined *at a point* we can only differentiate a function *at a point*. We usually say that limit evaluation and differentiability are **local properties**. If we don't specify the "at x" the convention is that the function is differentiable at every point in its domain.

The proofs of the Constant, Sum, and Constant Multiple Differentiation Rules are all completely straightforward so we will leave them as exercises[3] for you.

Theorem 23: [**The Constant Rule for Differentiation**]
If L is some number and $f(x) = L$ for all real values of x near[4] L, then $f'(x) = 0$ at every real number x near[5] L .
Proof: By Definition #12.3

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

[3]Frequently students will simply ignore problems that are described as straightforward. Don't make that mistake. Straightforward does not mean easy, and it does not mean unimportant. We are leaving these problems for you so you can gain experience using limits in the simplest cases, not because they are unimportant.

[4]"Near" means on an open interval about L. Recall Definition 16.

[5]On the same open interval.

But $f(x+h) = L = f(x)$ so

$$f'(x) = \lim_{h\to 0} \frac{L-L}{h} = \lim_{h\to 0} \frac{0}{h} = 0.$$

■

Theorem 24: **[The Sum Rule for Differentiation]**
If $\alpha(x)$ and $\beta(x)$ are differentiable at x and $f(x) = \alpha(x) + \beta(x)$, then $f(x)$ is also differentiable at x and

$$f'(x) = \alpha'(x) + \beta'(x).$$

Problem #495: Use Definition 15 to prove the Sum Rule for Differentiation.

Recall that when we first established the General Differentiation Rules using differentials in Chapter 3 we said that the Constant Multiple, Power and Quotient Rules for differentiation were just conveniences because they depend on the other rules.

This is still true of course (except for the caveat in the point of rigor below) which means that we don't have to prove any of them using limits. For example, since the Constant Rule and the Product Rule are now established theorems we can use these to prove the Constant Multiple Rule directly, without having to resort to using limits.

Theorem 25: **[The Constant Multiple Rule for Differentiation]**
If $f(x)$ is differentiable at x and K is a constant then $\alpha(x) = Kf(x)$ is also differentiable and

$$\alpha'(x) = Kf'(x).$$

Problem #496: Use the Definition #15 to prove the Constant Multiple Rule.

13.3 The Chain Rule

To understand the Chain Rule we will need to slightly blur the distinction between function and variable.

Example #83:

Here's what we mean: The formula $y = (2x^2 - 6x)^3$, is given entirely in terms of the variables x, and y. To differentiate using differentials we would make the (variable) substitution $z = 3x^2 + 6x$ so that $y = z^3$. In that case, $\mathrm{d}y = 3z\,\mathrm{d}z = 3\left(3x^2 + 6x\right)^2(6x+6)\,\mathrm{d}x$, and dividing through by $\mathrm{d}x$ gives us the derivative of y with respect to x,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 3z\,\mathrm{d}z = 3\left(3x^2 + 6x\right)^2(6x+6). \tag{13.1}$$

But Definition #15 requires that we think about functions, not variables so let's translate this problem into the language of functions. If $y = \left(2x^2 - 6x\right)^3$, clearly y is a function of (depends on) x. Naming that function f, we have $y = f(x)$. Replacing y with $f(x)$, we get $f(x) = (2x^2 - 6x)^3$.

Similarly, if $z = 3x^2 + 6x$ then z is also a function of (depends on) x, and naming that function β we have $z = \beta(x)$. Replacing z with $\beta(x)$ we have $f(x) = (\beta(x))^3$. If we suppress the "(x)" part of $\beta(x)$, we see that f can also be thought of as a function of (depends on) β so that

$$f(\beta) = \beta^3$$

is also a valid representation of our function. If we now define $\alpha(\beta) = \beta^3$ we see that

$$f(x) = \alpha(\beta).$$

Looking again at Equation 13.1, and mixing the differential and functional notations a bit we see that

$$f'(x) = \frac{\mathrm{d}y}{\mathrm{d}x} = 3z\,\mathrm{d}z = \underbrace{3\left(3x^2 + 6x\right)^2}_{\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}=\alpha'(\beta)}\underbrace{(6x+6)}_{\frac{\mathrm{d}\beta}{\mathrm{d}x}=\beta'(x)} = \alpha'(\beta)\cdot\beta'(x).$$

Thus if $f(x) = \alpha(\beta(x))$ is the composition of $\alpha(x)$ and $\beta(x)$ then

$$f'(x) = \alpha'(\beta(x))\beta'(x).$$

This is the Chain Rule. We have expressed the Chain Rule in this form so that we can prove it rigorously, not so that we can use it. The substitution process using differentials still works so there is no reason to stop using substitution when you are actually computing derivatives.

End Of Example #83

Theorem 26: **[The Chain Rule]**

Suppose that $\beta(x)$ is differentiable at x, that $\alpha(x)$ is differentiable at $\beta(x)$ and that $\Delta\beta \neq 0$ near x. Then the composition,

$$f(x) = \alpha(\beta(x))$$

is also differentiable, and

$$f'(x) = \alpha'(\beta(x))\cdot\beta'(x). \tag{13.2}$$

Digression #21: The Origins of the Chain Rule

Before the invention of Calculus, arithmetic primers gave the name "The Chain Rule" to the computational technique that is used to, among other things, convert money from one currency to another. For example if we need to convert 30 American dollars (\$) to British pounds (£) and we know that[a]

$$1 \text{ dollar} = 0.86 \text{ euros, and that } 1 \text{ euro} = 0.9 \text{ pounds.}$$

Then the conversion is

$$\$30 = 30\,\cancel{\text{dollars}} \times \frac{0.86}{1}\frac{\cancel{\text{euros}}}{\cancel{\text{dollars}}} \times \frac{0.9}{1}\frac{\text{pounds}}{\cancel{\text{euros}}} = 30 \times 0.86 \times 0.9 \text{ pounds} = 23.22\pounds$$

We've actually seen this type of conversion before. We used in in Section 5 when we converted angular velocity to linear velocity via the formula:

$$\left(\frac{3}{1}\frac{\cancel{\text{revolution}}}{\cancel{\text{minute}}}\right)\cdot\left(\frac{2\pi}{1}\frac{\cancel{\text{radians}}}{\cancel{\text{revolution}}}\right)\cdot\left(\frac{10}{1}\frac{\text{meters}}{\cancel{\text{radians}}}\right)\cdot\left(\frac{1}{60}\frac{\cancel{\text{minute}}}{\text{second}}\right) = \frac{\pi}{1}\frac{\text{meters}}{\text{second}} \approx 3.14\frac{\text{meters}}{\text{second}}.$$

A similar chain of cancellations will occur when we differentiate a function composition of the form $\alpha(t) = \alpha(\beta(y(x(t))))$. We think of

α as a function of β (so that $\alpha'(\beta) = \frac{\mathrm{d}\alpha}{\mathrm{d}\beta}$)
β as a function of y (so that $\beta'(y) = \frac{\mathrm{d}\beta}{\mathrm{d}y}$),
y as a function of x (so that $y'(x) = \frac{\mathrm{d}y}{\mathrm{d}x}$),
and x as a function of t (so that $y'(x) = \frac{\mathrm{d}x}{\mathrm{d}t}$).

Putting this all together we see that

$$\alpha'(t) = \frac{\mathrm{d}\alpha}{\mathrm{d}\cancel{\beta}}\cdot\frac{\mathrm{d}\cancel{\beta}}{\mathrm{d}\cancel{y}}\cdot\frac{\mathrm{d}\cancel{y}}{\mathrm{d}\cancel{x}}\cdot\frac{\mathrm{d}\cancel{x}}{\mathrm{d}t} = \frac{\mathrm{d}\alpha}{\mathrm{d}t}.$$

The substitutions we used to make things "easier on your eyes" in Section 1.2 is equivalent this chain of cancellations. With the invention of Calculus the older Chain Rule for unit conversion was extended to the differentiation by substitution technique using differentials. Eventually it became the only Chain Rule. When the limit was used to provide rigor to Calculus the name was also applied to Equation (13.2).

[a]These numbers were accurate on the day this passage was written. They are almost certainly wrong on the day you are reading it. Don't use them to convert currency.

End Of Digression #21

Understanding the Chain Rule in this form requires that we blur the distinction between function and variable a bit. When we compute $\frac{\mathrm{d}\alpha}{\mathrm{d}\beta} = \alpha'(\beta)$ (the derivative of α with with respect to β) we view β as a variable, but when we compute $\frac{\mathrm{d}\beta}{\mathrm{d}x} = \beta'(x)$ (the derivative of β with respect to x) we view it as a function.

As far as the Chain Rule is concerned it is both.

Proof: Before we begin take specific notice of the assumption "$\Delta\beta \neq 0$ near x" in the statement of the Chain Rule. We will have a few comments about this in Digression #22 after the proof is completed.

We will first establish that

$$\lim_{h\to 0} \Delta\beta = 0. \tag{13.3}$$

Suppose[6] that

$$\beta(x+h) - \beta(x) = \Delta\beta. \tag{13.4}$$

Then

$$\lim_{h\to 0} \beta(x+h) = \lim_{h\to 0}(\beta(x) + \Delta\beta).$$

By Theorem 15 we have

$$\lim_{h\to 0} \beta(x+h) = \lim_{h\to 0} \beta(x) + \lim_{h\to 0} \Delta\beta,$$

and since $\beta(x)$ is differentiable at x we see from Lemma 20 that

$$\beta(x) = \beta(x) + \lim_{h\to 0} \Delta\beta$$

from which Equation (13.3) follows.

To prove the Chain Rule recall that

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{\alpha(\beta(x+h)) - \alpha(\beta(x))}{h}. \end{aligned}$$

Multiplying by 1 in the form[7] $\frac{\Delta\beta}{\Delta\beta}$ gives

$$f'(x) = \lim_{h\to 0}\left(\frac{\alpha(\beta(x+h)) - \alpha(\beta(x))}{\Delta\beta} \cdot \frac{\Delta\beta}{h}\right). \tag{13.5}$$

From Equation (13.4) we see that $\Delta\beta = \beta(x+h) - \beta(x)$ so

$$= \lim_{h\to 0}\left(\frac{\alpha(\beta(x+h)) - \alpha(\beta(x))}{\Delta\beta} \cdot \frac{\beta(x+h) - \beta(x)}{h}\right).$$

By Theorem 16 we have:

$$= \lim_{h\to 0}\left(\frac{\alpha(\beta(x+h)) - \alpha(\beta(x))}{\Delta\beta}\right) \cdot \lim_{h\to 0}\left(\frac{\beta(x+h) - \beta(x)}{h}\right).$$

[6]Because asserting the equality of non-existing objects would be meaningless we assume, implicitly that all limits in this argument exist.

[7]In the past we have called this "uncancelling" $\Delta\beta$. Also, notice that this is where we use the assumption, "$\Delta\beta \neq 0$ near x."

Equation (13.3) says that $h \to 0$ is equivalent to $\Delta\beta \to 0$ so we have

$$f'(x) = \lim_{\substack{\cancel{h\to 0}\\ \Delta\beta\to 0}} \left(\frac{\alpha(\beta+\Delta\beta)-\alpha(\beta)}{\Delta\beta}\right) \cdot \lim_{h\to 0}\left(\frac{\beta(x+h)-\beta(x)}{h}\right),$$

$$f'(x) = \underbrace{\lim_{\Delta\beta\to 0} \left(\frac{\alpha(\beta+\Delta\beta)-\alpha(\beta)}{\Delta\beta}\right)}_{=\alpha'(\beta)} \cdot \underbrace{\lim_{h\to 0}\left(\frac{\beta(x+h)-\beta(x)}{h}\right)}_{=\beta'(x)}. \tag{13.6}$$

$$f'(x) = \alpha'(\beta)\cdot\beta'(x). \tag{13.7}$$

In Equation (13.7 β is first used as a variable in $\alpha'(\beta)$, and then as the function $\beta(x)$. While this is correct, it is also poor form because it accentuates the dual use of β. To avoid this we usually express the Chain Rule as

$$f'(x) = \alpha'(\beta(x))\cdot\beta'(x)$$

to emphasize that x, not β, is the variable. ∎

Digression #22: Why Assume that $\Delta\beta \neq 0$ near Zero?

Do you see why we had to assume that $\Delta\beta \neq 0$ near x?

Observe that in Equation (13.6) $\Delta\beta$ plays the same role the h plays in Definition (15). In Definition (15) we were careful to insist that h could never equal zero,so if we are going to interpret $\lim\limits_{\Delta\beta\to 0}\left(\dfrac{\alpha(\beta+\Delta\beta)-\alpha(\beta)}{\Delta\beta}\right)$ as the derivative of α with respect to β, as we did in Equation (13.6), we need to know that $\Delta\beta \neq 0$ when h is near zero.

Our imposition of that constraint means that Theorem #26 does not apply to any function $f(x) = \alpha(\beta(x))$ where $\Delta\beta$ might be equal to zero no matter how close h is to zero. Fortunately, functions of that sort are generally the kinds of "pathological functions" that Poincaré is complained about in the quote at the beginning of this chapter. A valid proof of the Chain Rule without that constraint is possible, but since it would have very little relevance to anything we'll be doing we have chosen to prove only this weaker form of the Chain Rule[a]

If you are unsatisfied with this proof and want to see a proof of the stronger version of the Chain Rule, consider majoring in mathematics. You'll see that and much, much more. In the meantime try working through the following problem.

Problem #497:

(a) Show that the function $\beta(x) = \sin\left(\frac{1}{x}\right)$ does not satisfy the constraint $\Delta\beta \neq 0$ when x is near zero.
(**Hint:** Recall Definition #16.)

(b) As a result of part (a) Theorem 26 does not apply to any of the following functions at $x = 0$. Nevertheless one of them is differentiable at $x = 0$. Use Definition #15 to find out which one.

(i) $T(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}.$

(ii) $U(x) = \begin{cases} x\sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}.$

(iii) $V(x) = \begin{cases} x^2\sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}.$

---

[a]You may be wondering if our choice to go with a weaker form of the Chain Rule means we've given Bishop Berkeley cause for compliant. The answer is no, we haven't. If we'd left off the condition that $\Delta\beta \neq 0$, then our proof would not have been rigorous because we'd have ended by claiming more than we'd proved. As it is, we've only claimed what we have proved.
Rigorous does not mean perfect, it means logical.

End Of Digression #22

Example #84:

Suppose that $f(x) = (\sin(x) + \cos(x))^2$. To use the Chain Rule to compute the derivative of $f(x)$ we need to recognize that $f(x)$ is the composition of $\alpha(x) = x^2$, and $\beta(x) = \sin(x)+\cos(x)$ and then apply Theorem 26 as follows.

$$\begin{aligned} f'(x) &= \alpha'(\beta(x)) \cdot \beta'(x) \\ &= \alpha'(\beta(x)) \cdot (\cos(x) - \sin(x)) \\ &= \alpha'(\sin(x) + \cos(x)) \cdot (\cos(x) - \sin(x)) \\ f'(x) &= 2(\sin(x) + \cos(x)) \cdot (\cos(x) - \sin(x)). \end{aligned}$$

End Of Example #84

In our opinion the Chain Rule leaves a lot to be desired as a computational technique. But we don't have to use it that way since Theorem 26 validates the substitutions we have always used.

Drill #498:

Suppose $y = f(x) = (\sin(x) + \cos(x))^2$. Compute the differential $\mathrm{d}y$ and then divide through by $\mathrm{d}x$ to find the derivative $\frac{\mathrm{d}y}{\mathrm{d}x}$. Confirm that it is the same as the derivative we found in Example #84.

Problem #499:

Compute $\frac{\mathrm{d}y}{\mathrm{d}x}$ for each of the following functions by identifying $\alpha(x)$ and $\beta(x)$ such that $y(x) = \alpha(\beta(x))$ and applying the Chain Rule. You may have to do this more than once for a given problem.

In each case confirm that your computation is correct with an appropriate differential substitution.

(a) $y = (3x+5)^6$

(b) $y = \sec(\tan(x))$

(c) $y = \sqrt[7]{\frac{1}{x} + x^3}$

(d) $y = \left(\frac{x - x^{\frac{1}{2}}}{x^3 - 1}\right)^2$

(e) $y = e^{x - \cos^2(x)} + (2x^2 - 3)^{\frac{1}{5}}$

(f) $y = \sqrt{x + \sqrt[3]{2 + \sqrt[4]{3 - x^2}}}$

13.4 The Product Rule

A rigorous proof of the Product Rule is also fairly complex, but it does not suffer from the kind of technical problems we encountered in the proof of the Chain Rule.

Theorem 27: [**The Product Rule for Differentiation**]
If $\alpha(x)$ and $\beta(x)$ are differentiable at x then $f(x) = \alpha(x) \cdot \beta(x)$ is differentiable and

$$f'(x) = \alpha(x) \cdot \beta'(x) + \beta(x) \cdot \alpha'(x). \tag{13.8}$$

Proof: We start with the two observations. The first is that

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{\alpha(x+h)\beta(x+h) - \alpha(x)\beta(x)}{h} \tag{13.9}$$

and the second is that, in limit form, Equation (13.8) is

$$f'(x) = \alpha(x)\left(\lim_{h \to 0} \frac{\beta(x+h) - \beta(x)}{h}\right) + \beta(x)\left(\lim_{h \to 0} \frac{\alpha(x+h) - \alpha(x)}{h}\right). \tag{13.10}$$

It appears then that our goal is to simply reorganize Equation (13.9) until it looks like Equation (13.10). We say "simply" but it will only appear to be simple after we have succeeded. We will proceed slowly.

Observe that if we subtract $\alpha(x+h)\beta(x)$ from the blue part of the numerator in Equation (13.9) we get

$$\alpha(x)\beta(x) - \alpha(x+h)\beta(x) = -\beta(x)\left(\alpha(x+h) - \alpha(x)\right),$$

whereas if we *add* $\alpha(x+h)\beta(x)$ to the red part of the numerator in Equation (13.9) we get

$$\alpha(x+h)\beta(x+h) + \alpha(x+h)\beta(x) = \alpha(x+h)\left(\beta(x+h) - \beta(x)\right).$$

This suggests that we should both add and subtract the expression $\alpha(x+h)\beta(x)$ to the numerator of Equation (13.9). Doing this and factoring as we've indicated above we get

$$f'(x) = \lim_{h \to 0} \frac{\alpha(x+h)\left(\beta(x+h) - \beta(x)\right) - \left[-\beta(x)\left(\alpha(x+h) - \alpha(x)\right)\right]}{h},$$

By Theorem 15 we can separate this into the limit of the two fractions as follows:

$$f'(x) = \lim_{h\to 0}\left(\frac{\alpha(x+h)\,(\beta(x+h)-\beta(x))}{h}\right) + \lim_{\substack{h\to 0\\ h>0}}\left(\frac{\beta(x)\,(\alpha(x+h)-\alpha(x))}{h}\right),$$

and by Theorem 16 we see that

$$f'(x) = \underbrace{\left[\lim_{h\to 0}\alpha(x+h)\right]}_{=\alpha(x)}\underbrace{\left[\lim_{h\to 0}\left(\frac{\beta(x+h)-\beta(x)}{h}\right)\right]}_{=\beta'(x)} + \underbrace{\left[\lim_{h\to 0}\beta(x)\right]}_{\beta(x)}\underbrace{\left[\lim_{h\to 0}\left(\frac{\alpha(x+h)-\alpha(x)}{h}\right),\right]}_{\alpha'(x)}$$

and therefore

$$f'(x) = \alpha(x)\beta'(x) + \beta(x)\alpha'(x).$$

■

13.5 The Other General Differentiation Rules

Theorem 28: [**The Quotient Rule for Differentiation**]

We assume that $\alpha(x)$, $\beta(x)$, and $f(x) = \frac{\alpha(x)}{\beta(x)}$ are all differentiable functions[8]. Assume further that $\beta(x) \neq 0$. Then

$$f'(x) = \frac{\beta(x)\alpha'(x) - \alpha(x)\beta'(x)}{[\beta(x)]^2}.$$

Proving this directly by using limits would be unpleasant, but as we observed in Chapter 3 the Quotient Rule can be viewed as a rearranged version of the Product Rule.

Problem #500: Use the Product Rule to derive the Quotient Rule.
(**Hint:** First solve $f(x) = \frac{\alpha(x)}{\beta(x)}$ for $\alpha(x)$.)

With the Product Rule for Differentiation in place we now have the tools needed to prove the Power Rule for Positive Integer Exponents. The method of proof we outline in the following problem is called **Mathematical Induction**[9] and it can be used in other contexts as well. In fact, most of the "Find the Pattern" problems in this text require an Induction argument for full rigor.

Problem #501: The Power Rule for Positive Integer Exponents

Assume that $\alpha(x) = x^n$ is differentiable at x for any positive integer n.

Part 1: Assume that $n = 1$. Use the limit definition to show that $\alpha'(x) = nx^{n-1}$.
(**Comment:** This says, "The Power Rule holds for $k = 1$.")

[8] Notice that we have explicitly assumed that the quotient

[9] We mentioned **Mathematical Induction** In part (d) of problem #47 back in Chapter 3.

Part 2: Now assume that the Power Rule for Positive Integer Exponents holds for $n = k$, where k is an arbitrary, fixed positive integer.

Let $\beta(x) = x^{k+1}$ and show that $\beta'(x) = (k+1)x^k$.
(**Comment:** This says, "If the Power Rule holds for k then it must also hold for $k+1$.")

Do you see how this proves that the Power Rule holds for *any* positive integer, n? Write a short paragraph explaining the logic behind this.

With the **Power Rule for Positive Integer Exponents** in place we can extend it to both negative and rational exponents in the same way we did it in Chapter 3. The following problem is essentially a repeat of problems 56 and 57, using Lagrange's prime notation, and function notation, rather than differentials.

Problem #502: The Power Rule for Rational and Negative Exponents

(a) Assume n is a positive integer and that $\alpha(x) = x^{-n}$ is differentiable. Show that

$$\alpha'(x) = -nx^{-(n+1)}.$$

(**Hint:** Rewrite $\alpha(x) = x^{-n}$ as $\frac{1}{x^n}$ and use the Quotient Rule for Differentiation and the Power Rule for positive integers.)

(b) Assume that q is a non-zero integer and that $\alpha(x) = x^{1/q}$ is differentiable at x. Show that

$$\alpha'(x) = (1/q)x^{(1/q-1)}.$$

(**Hint:** Rewrite $\alpha(x) = x^{1/q}$ as

$$[\alpha(x)]^q = x$$

and use the Chain Rule and the Power Rule for positive integers.)

(c) Assume that p and q are integers, $q \neq 0$, and that $\alpha(x) = x^{p/q}$ is differentiable at x. Show that

$$\alpha'(x) = (p/q)x^{(p/q-1)}.$$

(**Hint:** Rewrite $\alpha(x) = x^{p/q}$ as $\alpha(x) = \left(x^{1/q}\right)^p$ and use the Chain Rule and part (b).)

Together the previous two problems prove the Power Rule for *rational* exponents:

Theorem 29: **[The Power Rule for Rational Exponents]**
Assume that p and q are integers, $q \neq 0$, and that $\alpha(x) = x^{p/q}$ is differentiable at x. Then

$$\alpha'(x) = (p/q)x^{(p/q-1)}.$$

In the statement of Theorem #28 we explicitly assumed that the quotient, $\frac{\alpha(x)}{\beta(x)}$, is differentiable at x. This has the effect that the theorem does not necessarily apply to all possible quotients, in the same way that when we add $\Delta\beta \neq 0$ to the statement of the Chain Rule, the theorem applies to fewer compositions. And just like the Chain Rule the functions that Theorem #28 does not apply to are mostly pathological, and of no use to us right now.

We added the same assumption to Theorem #29 for similar reasons.

Digression #23: Are You a Mathematician?

If leaving these theorems incomplete in this way is troubling to you then you are almost certainly a mathematician by temperament. If you haven't decided on a major yet, consider mathematics. You obviously like it. Why not learn more?

If you find that you simply don't care about completing all of the details and you are not majoring in mathematics, congratulations! You've made the right choice.

Problem #503 will lead you through the steps necessary to prove the Quotient Rule for Differentiation without the assumption that $\frac{\alpha(x)}{\beta(x)}$ is differentiable. Have fun!

Problem #503:

Assume that $\alpha(x)$ and $\beta(x)$ are differentiable and that $\beta(x) \neq 0$, but we make no assumption about the differentiability of $f(x) = \frac{\alpha(x)}{\beta(x)}$.

(a) First prove the special case of the Quotient Rule where $f(x) = \frac{1}{\beta(x)}$.

- **(i)** Use the limit definition to show that $\beta'(x) = \lim_{h \to 0} \dfrac{\beta(x) - \beta(x+h)}{h\beta(x)\beta(x+h)}$.
- **(ii)** Now evaluate the limit in part (i) to show that $f'(x) = \frac{-\beta'(x)}{[\beta(x)]^2}$.

(b) Use the Product Rule for Differentiation and the Chain Rule (along with the result of part a) to show that $f(x) = \frac{\alpha(x)}{\beta(x)}$ is differentiable at x and that

$$f'(x) = \frac{\beta(x)\alpha'(x) - \alpha(x)\beta'(x)}{[\beta(x)]^2}.$$

Problem #505 will lead you through the steps necessary to prove the Product Rule for Rational Exponents without the assumption that $x^{\frac{p}{q}}$ is differentiable. It relies on the result of Problem #504. Have fun!

Problem #504:

To prove Theorem 29 we will first focus on the special case of $\beta(x) = x^{\frac{1}{q}}$, q is a non-negative integer.

The key to proving this special case is a generalization of the difference of squares formula: $(a-b)(a+b) = a^2 - b^2$.

(a) Show that $(a-b)(a^2 + ab + b^2) = a^3 - b^3$.

(b) Show that $(a-b)(a^3 + a^2b + ab^2 + b^3) = a^4 - b^4$.

(c) Use **Mathematical Induction** to show that

$$(a-b)(a^{q-1}+a^{q-2}b+a^{q-3}b^2+\cdots+ab^{q-2}+b^{q-1})=a^q-b^q.$$

Problem #505:

Assume that p and q are integers and that $q \neq 0$. If we apply Definition #15 to $f(x)=x^{\frac{1}{q}}$, we get

$$f'(x)=\lim_{h\to 0}\frac{(x+h)^{\frac{1}{q}}-x^{\frac{1}{q}}}{h}.$$

(a) Use the substitutions $a=(x+h)^{\frac{1}{q}}$, $b=x^{\frac{1}{q}}$, and part (c) of the previous problem to show that

$$f'(x)=\lim_{a\to b}\frac{a-b}{a^q-b^q}=\frac{1}{qb^{q-1}}.$$

(b) Substitute $b=x^{\frac{1}{q}}$ into the result of part a to obtain

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

(c) Use the Chain Rule to show that for $\alpha(x)=x^{\frac{p}{q}}$

$$\alpha'(x)=\frac{p}{q}x^{\frac{p}{q}-1}.$$

End Of Digression #23

13.6 Derivatives of the Trigonometric Functions, via Limits

See also the TRIUMPHS Primary Source Project in Appendix A.7.

Theorem 30: **[Derivative of sin(x)]**
Suppose $\alpha(\theta)=\sin(\theta)$. Then $\alpha'(\theta)=\cos(\theta)$.

Proof: Showing that the derivative of $\sin(\theta)$ is $\cos(\theta)$ is mostly straightforward but we're going to hit a snag partway through. We'll proceed for a bit to see where the trouble is.

Start with the limit definition:

$$\alpha'(\theta) = \lim_{h\to 0} \frac{\sin(\theta + h) - \sin(\theta)}{h}$$

In the numerator we see the expression $\sin(\theta + h)$. Recall the sum formula for the Sine:

$$\sin(A + B) = \sin(A)\cos(B) + \cos(A)\sin(B).$$

Taking $A = \theta$ and $B = h$ we have:

$$\alpha'(x) = \lim_{h\to 0} \frac{\sin(\theta)\cos(h) + \cos(\theta)\sin(h) - \sin(\theta)}{h}$$

Next, if we factor $\sin(\theta)$ out of the terms where it appears and rearrange the numerator a bit we have:

$$\begin{aligned} &= \lim_{h\to 0} \frac{\sin(\theta)(\cos(h) - 1) + \cos(\theta)\sin(h)}{h}. \\ &= \lim_{h\to 0} \left(\frac{\sin(\theta)(\cos(h) - 1)}{h} + \frac{\cos(\theta)\sin(h)}{h}\right). \end{aligned}$$

By Theorem 15:

$$= \lim_{h\to 0} \frac{\sin(\theta)(\cos(h) - 1)}{h} + \lim_{h\to 0} \frac{\cos(\theta)\sin(h)}{h}$$

and by Corollary 19:

$$= \sin(\theta) \underbrace{\left(\lim_{h\to 0} \frac{(\cos(h) - 1)}{h}\right)}_{=0} + \cos(\theta) \underbrace{\left(\lim_{h\to 0} \frac{\sin(h)}{h}\right)}_{=1}.$$

If the values of the two limits are 0 and 1 respectively as we've indicated we can conclude that $\alpha'(\theta) = \cos(\theta)$.

Unfortunately this proof cannot be considered complete until we have shown that these last two limits are what we claim they are. We will do this via the two lemmas below.

■

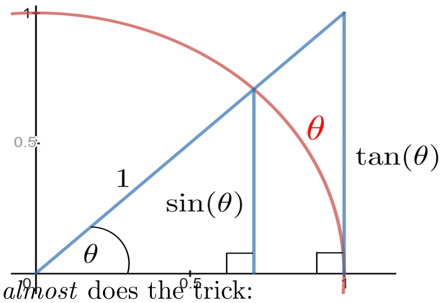
It is tempting to use L'Hôpital's Rule to evaluate these limits, especially since it is so very easy to do.

Drill #506: Use L'Hôpitals Rule to show that $\lim_{h\to 0} \frac{(\cos(h)-1)}{h} = 0$ and that $\lim_{h\to 0} \frac{\sin(h)}{h} = 1$.

Sadly, using Drill #506 to finish the proof of Theorem #30 is an example of circular reasoning. We can't use the fact that the derivative of $\sin(x)$ is $\cos(x)$ to prove that the derivative of $\sin(x)$ is $\cos(x)$. So we will have to find a way to evaluate these limits without using L'Hôpital's Rule.

Lemma 31: $\lim_{h \to 0} \frac{\sin(h)}{h} = 1$

Proof:
There are two cases:
Case 1, $\theta \geq 0$: We will use the Squeeze Theorem. Recall that in Section 5.1 we observed that the lengths of certain line segments associated with the unit circle in the first quadrant are equal to the trigonometric functions. The figure at the right shows the relationship between θ, $\sin(\theta)$, and $\tan(\theta)$. Notice in particular that



$$\sin(\theta) \leq \theta \leq \tan(\theta).$$

Dividing each expression in the inequality by $\sin(\theta)$ *almost* does the trick:

$$1 \leq \frac{\theta}{\sin(\theta)} \leq \frac{1}{\cos(\theta)}.$$

In the center we now have the reciprocal of what we need, so we need to invert each expression.

However, keep in mind that these are not equations they are inequalities. When we invert an inequality we must reverse its sense. This gives

$$1 \geq \frac{\sin(\theta)}{\theta} \geq \frac{\cos(\theta)}{1},$$

and this is true on the interval $\left[0, \frac{\pi}{2}\right]$. Since

$$\lim_{\substack{\theta \to 0 \\ \theta > 0}} 1 = \lim_{\substack{\theta \to 0 \\ \theta > 0}} \cos(\theta) = 1$$

the Squeeze Theorem applies, and we conclude that

$$\lim_{\substack{\theta \to 0 \\ \theta > 0}} \frac{\sin(\theta)}{\theta} = 1.$$

Case 2, $\theta < 0$: For this case notice that $\sin(-\theta) = -\sin(\theta)$ so that $\frac{\sin(-\theta)}{-\theta} = \frac{\sin(\theta)}{\theta}$. We make the substitution $\theta = -\phi$ where $\phi > 0$. Therefore when $\theta < 0$ we have

$$\lim_{\substack{\theta \to 0 \\ \theta < 0}} \frac{\sin(\theta)}{\theta} = \lim_{\substack{\phi \to 0 \\ \phi > 0}} \frac{\sin(-\phi)}{-\phi} = \lim_{\substack{\phi \to 0 \\ \phi > 0}} \frac{\sin(\phi)}{\phi} = 1$$

by Case 1. ■

Problem #507: Show that $\lim_{h \to 0} \frac{(\cos(h)-1)}{h} = 0$.
(**Hint:** It is tempting to model this proof on the proof of Lemma 31. While this can be done,

it is delicate. It is simpler to multiply by 1 in the form $\frac{\cos(h)+1}{\cos(h)+1}$. Try that instead.)

Once Problem (507) has been solved the proof that $\frac{\mathrm{d}(\sin(x))}{\mathrm{d}x} = \cos(x)$ is complete.

Problem #508: Prove that $\frac{\mathrm{d}(\cos(\theta))}{\mathrm{d}\theta} = -\sin(\theta)$, using the proof of Theorem 30 as a guide.

Assuming that $\tan(\theta)$, $\cot(\theta)$, $\sec(\theta)$, and $\csc(\theta)$ are differentiable we can now use Theorem 503 to find their derivatives as well. Since this is exactly what we did in Section 5.4 we have the derivatives of all of the trigonometric functions.

13.7 Inverse Functions

Although we have worked with the inverses of some specific functions we have not formally defined what we mean by an inverse. We will remedy that now. We have seen that not all functions can be inverted (see for example, Digression #9) so the first step is to define which functions are invertible.

Informally a function that never takes the same value twice is called a **one-to-one function**[10]. Formally we have the following.

Definition 18: One-To-One Functions
A function, $f(x)$, defined on a domain, D, is said to be one-to-one if, whenever x_1 and x_2 are in D and $x_1 \neq x_2$ then, $f(x_1) \neq f(x_2)$.

Recall that when we tried to invert $\tan(x)$ (which is not one-to-one) in Section (5.6) we got the multifunction $\arctan(x)$. We had to restrict the domain of the tangent function to $\frac{-\pi}{2} \leq x \leq \frac{\pi}{2}$, in order to find an inverse. That restriction gave us a one-to-one function which we could invert because one-to-one functions are the only functions with inverses.

Definition 19: Inverse Functions
Suppose $f(x)$, with domain D and range, R is a one-to-one function. Then the inverse of $f(x)$ is the function $f^{-1}(x)$ with domain R and range[11] D which satisfies the following properties:

1. $f\left(f^{-1}(x)\right) = x$
2. $f^{-1}\left(f(x)\right) = x$

for every value of x in the domain of f (equivalently, in the range of f^{-1}).

Loosely speaking, Definition #19 says that two functions are mutually inverse if they "undo" each other.

Our next task is to show that the derivatives of the in inverse trigonometric functions are what we expect them to be. Given that we have now obtained the derivatives of all of the

[10] They are also called **injective**.
[11] Notice that the domain and range have been swapped.

trigonometric functions it appears that we could proceed just as we did in Section 5.7 and Section 5.8.

But that would require that we explicitly assume that each of the inverse trigonometric functions is differentiable, similar to the way we found the derivative of a quotient. This is a valid approach of course, but proceeding in that manner would mask some issues that will be of interest to us later. So we will approach the derivatives of inverse functions abstractly by (rigorously) finding a formula for the derivative of the inverse of a generic, invertible function. After that we'll only need to apply the formula to each of the inverse trigonmetric formulas.

Digression #24: Inverse and Derivative Notation

As we saw in Digression 8 there are some difficulties with the notation we use to indicate inverse functions. These problems only get worse when we mix the standard derivative notations with the inverse function notation. Lagrange's prime notation is especially problematic.

For example if $f(x)$ is an invertible function the derivative of $f^{-1}(x)$ could be denoted either as:

$$\frac{\mathrm{d}(f^{-1})}{\mathrm{d}x} \quad \text{or} \quad {f^{-1}}'(x).$$

But both of these are somewhat awkward. Mathematicians also sometimes use the **operator** notation:

$$\mathbf{D}(f(x)) = f'(x) = \frac{\mathrm{d}f}{\mathrm{d}x}$$

and in this situation it minimizes the awkwardness a bit.

As we've seen there can also be some vagueness involving the distinction between functions and variables. For example suppose we want to sketch a graph of this *relation* between x and y:

$$y - x^3 = 0.$$

The simplest thing to do is to choose a value for either x or y and then figure out what the corresponding y or x is. This is simpler to do that if we rearrange the *relation* so that we have one variable strictly in terms of ("as a function of") the other. For this particular relation it is easiest to choose a value for x and compute the corresponding y value so we would normally rearrange it as

$$y(x) = x^3. \tag{13.11}$$

Equation 13.11 defines y *as a function of* x.

But we only solved for y because we could see it was a little easier to do. Otherwise our choice was completely arbitrary. We could also have solved for x giving,

$$x(y) = \sqrt[3]{y}. \tag{13.12}$$

In this case we have x *as a function of* y.

The two functions, $y(x)$, ("cube") and $x(y)$ ("cube root"), clearly contain the same information as the original relation $y - x^3 = 0$. But they are different, related, functions. They are in fact mutually inverse.

For example suppose we choose $x = 2$ and use Equation (13.11) to find $y = 8$. If we then take $y = 8$ and use Equation (13.12) we find that $x = 2$. That is, $x(y)$ has "undone" $y(x)$ for the single pair $(2, 8)$. Drill #509 asks you to show that it is true for every pair $(x, y(x))$. This "undoing" makes $y(x)$ and $x(y)$ a pair of mutually inverse functions.

But in function notation the variable (frequently x or t) is a placeholder. For example, each of $y(x) = x^3$, $y(t) = t^3$, $f(\alpha) = \alpha^3$, or even $f(\blacksquare) = \blacksquare^3$ defines *exactly* the same function: The function which cubes its input. It doesn't matter what we call the variable. It just holds a place in the formula that tells us what the input is and what to do with it. *Since it doesn't matter* what we call the variable we usually call it x unless there is some compelling reason to use something else.

To avoid confusing variable names with function names we usually denote $y(x)$ as $f(x)$. It's inverse, $x(y)$ should probably be denoted as $f^{-1}(y)$. But sadly, recognizing that the variable is just a placeholder in function notation we use the same variable name in both the function and it's inverse. So we denote the inverse of $f(x)$ as $f^{-1}(x)$, even though it would probably make it easier for beginners to use $f^{-1}(y)$, as a reminder that both *functions* come from the same original *relation*.

Drill #509:

Prove that $f(x) = x^3$ and $f^{-1}(x) = \sqrt[3]{x}$ are mutually inverse by showing that they satisfy the conditions stated in Definition #19

The notation for inverse functions is not great. It can be very confusing, especially for beginners. Be careful with it.

End Of Digression #24

Our next task is to show that if $f(x)$ is invertible and differentiable, then f^{-1} is also differentiable[12]. We do this by showing that the limit

$$\mathbf{D}\left(f^{-1}(x)\right) = \lim_{h\to 0} \frac{f^{-1}(x+h) - f^{-1}(x)}{h} \tag{13.13}$$

exists.

In general this is true but there is one exception that has to be addressed. When f is differentiable at a and $f'(a) = 0$ then the limit in Equation (13.13 does not exist. Hence f^{-1} is not differentiable at $f(a)$. More formally, we have the following lemma.

Lemma 32:

If f is an invertible function, $f(a) = b$, f is differentiable at $x = a$, and $f'(a) = 0$, then f^{-1} is not differentiable at $x = b$. That is $\mathbf{D}\left(f^{-1}(b)\right)$ does not exist.

The following proof of this lemma is very challenging to read and understand for several reasons.

First, it is quite abstract. We don't have a particular function to think about so we can't simply write down formulas for the function and its inverse. Instead we have only the generic function, f and its inverse f^{-1}, and we'll need to remember what these symbols represent.

Second, we need to think about the functions f and f^{-1} as well as their derivatives.

[12] f^{-1} is obviously invertible

Third, instead of using the differential notation, $\frac{\mathrm{d}f}{\mathrm{d}x}$ that we've grown very comfortable with we'll be using the less familiar Lagrange prime notation and the operator notation we just introduced.

Finally, the nature of the problem forces us to mix these last two notations, using one here and the other there. This can make for difficult reading.

Read slowly. Remember that each symbol has meaning. Take time to understand that meaning and what each formula as a whole is telling you.

We include this proof in its full abstraction for two reasons:

1. To be as precise and as rigorous and we can.

2. We want to give you practice with higher level abstract reasoning in this (fairly) simple case.

The strategy behind the following proof follows the same general scheme as the Sherlock Holmes Maxim that we referred to in Problem #354. We will eliminate the impossible so that "whatever remains, however improbable, must be the truth."

There are two possibilities: Either the derivative of $f^{-1}(b)$ exists or it does not exist. There are two steps:

(1) Assume that the derivative of f^{-1} *does* exist at $x = b$ and calculate what $\mathbf{D}\left(f^{-1}(b)\right)$ must be.

(2) Show that our computed value is impossible. Then *á la* Holme's Maxim the only possibility left will be that the derivative of f^{-1} does *not* exist at $x = b$.

Proof of Lemma #32: Assume that f^{-1} is differentiable[13] at $x = b$. Because f and f^{-1} are mutually inverse we know that

$$f\left(f^{-1}(x)\right) = x.$$

Therefore

$$\mathbf{D}\left(f\left(f^{-1}(x)\right)\right) = \mathbf{D}\left(x\right).$$

On the right we have

$$\mathbf{D}(x) = 1.$$

On the left apply the Chain Rule:

$$f'\left(f^{-1}(x)\right) \cdot \mathbf{D}\left(f^{-1}(x)\right) = 1. \tag{13.14}$$

But when $x = b$ we find that $f'\left(f^{-1}(b)\right) = f'\left(a\right) = 0$, so that

$$0 = \underbrace{f'\left(f^{-1}(b)\right)}_{=0} \cdot \mathbf{D}\left(f^{-1}(b)\right) = 1$$

[13] We don't really believe this assumption. Be sure you are very clear on this point. We make this assumption so that we can use it to derive an absurd result; a contradiction. If there are no errors in our reasoning then the only possible conclusion will be that this assumption is false: f^{-1} is not differentiable at $x = b$.

or

$$0 = 1$$

which is ridiculous or in Holmes' word, impossible.

Therefore our assumption cannot true so f^{-1} is not differentiable at $x = b$. ■

While valid and correct, this proof is not very enlightening. A well chosen sketch would be much more convincing, if less rigorous.

Problem #510:
Choose a function whose derivative is equal to zero at some point and sketch the graph of your function and its inverse on the same set of axes. Be sure to include the point where the derivative is zero.

Use your graph to explain why the derivative of the inverse of your function does not exist.

We now understand what conditions are necessary for an arbitrary function, $f(x)$, to have a differentiable inverse.

Also, from Equation (13.14) we know what the derivative of the inverse will be *if it exists*:

$$\mathbf{D}\left(f^{-1}(x)\right) = \frac{1}{f'\left(f^{-1}(x)\right)}.$$

Drill #511:
Let $y = f^{-1}(x)$ and explain how the formula above is equivalent to

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{\frac{\mathrm{d}x}{\mathrm{d}y}} \tag{13.15}$$

The only thing left is to show that under the conditions on f in Lemma 32 the derivative (that is, the limit which defines the derivative) of the inverse does in fact exist.

Theorem 33: **[The Derivative of Inverse Functions]**
Suppose that

1. f is differentiable at $x = a$,
2. $f(a) = b$,
3. $f'(a) \neq 0$,
4. f^{-1} is continuous at[14] $x = b$.

Then the inverse of f is differentiable at $x = b$ and

$$\mathbf{D}\left(f^{-1}(b)\right) = \frac{1}{f'\left(f^{-1}(b)\right)}.$$

[14] In fact, this follows from the continuity of f at $x = a$. We do not have all of the tools necessary to prove this so we must include it in the assumptions of our theorem.

Reading and understanding the notation in Theorem 33 presents the same difficulties we saw in the proof of Lemma 32. Read it carefully. Be patient with yourself and do not rush.

Proof: We want to show that the limit

$$\mathbf{D}\left(f^{-1}(b)\right) = \lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h} = \frac{1}{f'\left(f^{-1}(b)\right)}.$$

Since $f(a) = b$ we know that $f^{-1}(b) = a$ so that

$$\lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h} = \lim_{h\to 0} \frac{f^{-1}(b+h) - a}{h}$$

Observe that if $b+h$ is in the domain of f^{-1} then it is in the range of f. Thus there is some number in the domain of f (for convenience we'll call it $a+k$) such that $b+h = f(a+k)$. Thus

$$\lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h} = \lim_{h\to 0} \frac{f^{-1}(f(a+k)) - a}{h}.$$

Again since f and f^{-1} are mutually inverse they "undo" each other, so $f^{-1}(f(a+k)) = a+k$ so that

$$\lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h} = \lim_{h\to 0} \frac{k}{h}.$$

Solving $b+h = f(a+k)$ for h gives $h = f(a+k) - b$ so

$$\lim_{h\to 0} \frac{k}{h} = \lim_{h\to 0} \frac{k}{f(a+k) - b}$$

and since $b = f(a)$ we have

$$\begin{aligned}
&= \lim_{h\to 0} \frac{k}{f(a+k) - f(a)} \\
&= \lim_{h\to 0} \frac{1}{\frac{f(a+k)-f(a)}{k}}. \\
&= \frac{1}{\displaystyle\lim_{h\to 0} \frac{f(a+k) - f(a)}{k}}.
\end{aligned}$$

The expression $\displaystyle\lim_{h\to 0} \frac{f(a+k) - f(a)}{k}$ would be $f'(a)$ if only we had $k \to 0$ instead of $h \to 0$. What we need to show now is that if $h \to 0$ then $k \to 0$. Then we could write

$$\mathbf{D}\left(f^{-1}(b)\right) = \frac{1}{\displaystyle\lim_{k\to 0} \frac{f(a+k) - f(a)}{k}} = \frac{1}{f'(a)} \tag{13.16}$$

and our proof would be complete

Written a little more carefully, what we need to show is that $\lim_{h\to 0} k = 0$. Recall that $a = f^{-1}(b)$, and that $a + k = f^{-1}(b+h)$ so we need to show that

$$\lim_{h\to 0} k = \lim_{h\to 0} [(a+k) - a] = \lim_{h\to 0} \left[f^{-1}(b+h) - f^{-1}(b)\right] = 0$$

or But we assumed that f^{-1} is continuous at $x = b$ which means that

$$\lim_{h\to 0} \left[f^{-1}(b+h) - f^{-1}(b)\right] = 0,$$

and the proof is complete.

One last point: On the left side of formula (13.16) the variable is b and on the right it is a. While this is not strictly wrong it is a more useful theorem if we state it in terms of b alone.

Since $f(a) = b$ we see that $f^{-1}(b) = a$ so

$$\mathbf{D}\left(f^{-1}(b)\right) = \frac{1}{f'(f^{-1}(b))}$$

and the proof is complete. ■

Using Theorem 33 we can now show that the derivatives of the inverse trigonometric functions and the natural logarithm are exactly what we expect them to be. The difference is that now there is no uncertainty or vagueness in our foundations. No modern Bishop Berkeley can step in and sew doubt.

Example #85: The Derivative of the Inverse Sine

Suppose $f(x) = \sin(x)$. Then $f^{-1}(x) = \sin^{-1}(x)$ so

$$\begin{aligned}\mathbf{D}\left(f^{-1}(x)\right) &= \mathbf{D}\left(\sin^{-1}(x)\right)\\ &= \frac{1}{f'(f^{-1}(x))}\\ &= \frac{1}{\cos(\sin^{-1}(x))}\\ \mathbf{D}\left(f^{-1}(x)\right) &= \frac{1}{\sqrt{1-x^2}}.\end{aligned}$$

End Of Example #85

Problem #512:

Use Theorem 33 to show that each of the following differentiation rules is correct:

(a) $\mathbf{D}\left(\cos^{-1}(x)\right) = \frac{-1}{\sqrt{1-x^2}}$

(b) $\mathbf{D}\left(\tan^{-1}(x)\right) = \frac{1}{1+x^2}$

(c) $\mathbf{D}\left(\cot^{-1}(x)\right) = \frac{-1}{1+x^2}$

(d) $\mathbf{D}\left(\sec^{-1}(x)\right) = \frac{1}{|x|\sqrt{x^2-1}}$

(e) $\mathbf{D}\left(\csc^{-1}(x)\right) = \frac{-1}{|x|\sqrt{x^2-1}}$

(f) $\mathbf{D}\left(\ln^{-1}(x)\right) = \frac{1}{x}$

Wait a minute! Did we forget one? What about the natural exponential function? Don't we also have to show that $\mathbf{D}(e^x) = e^x$?

Drill #513:
Look back at Definition 8 and explain why it is not necessary to use limits to show that $\mathbf{D}(e^x) = e^x$.