Calculus Teacher's Edition - Common Errors

CK-12 Foundation

December 11, 2009

CK-12 Foundation is a non-profit organization with a mission to reduce the cost of textbook materials for the K-12 market both in the U.S. and worldwide. Using an open-content, web-based collaborative model termed the "FlexBook," CK-12 intends to pioneer the generation and distribution of high quality educational content that will serve both as core text as well as provide an adaptive environment for learning.

flexbook
next generation textbooks

Contents

Chapter 1

Calculus TE - Common Errors

1.1 Functions, Limits, and Continuity

This Calculus Common Errors FlexBook is one of seven Teacher's Edition FlexBooks that accompany the CK-12 Foundation's Calculus Student Edition.

To receive information regarding upcoming FlexBooks or to receive the available Assessment and Solution Key FlexBooks for this program please write to us at teacher-requests@ck12.org.

Lesson 1: Equations and Graphs

To begin the study of calculus, it is helpful to review some important properties of equations and functions, and how to graph different kinds of functions on an $x - y$ coordinate system. A solid understanding of analytic geometry is essential to developing the techniques of differentiation and integration presented in this textbook. The ability to identify analytic solutions to the points where graphs intersect the x and y axes (e.g. the intercepts), as well as finding the exact points where two graphs or curves cross each other, will be necessary to evaluate limits, derivatives and integrals.

An important technique students will need throughout the study of calculus is evaluating functions by substituting in a value for a function's argument. In simple cases, the argument is given as a number. Given a function $f(x) = x^2$, to find $f(4)$ we substitute the 4 in for x, and get $f(4) = 4^2 = 16$.

Students must soon become comfortable with substituting entire algebraic expressions in for the argument of a function and evaluating the output. For instance, when calculating the derivative of a function, students will need to evaluate expressions like $f(x + a)$ for a variety of functions. For instance, if $f(x) = x^2 + 2x + 3$, to calculate $f(x + a)$ we must substitute $x + a$ for the value of x in the original function. This gives us:

$$\begin{aligned} f(x) &= (x + a)^2 + 2(x + a) + 3 \\ f(x + a) &= (x + a)(x + a) + 2x + 2a + 3 \\ f(x + a) &= x^2 + 2ax + a^2 + 2x + 3 \end{aligned}$$

Because polynomials are usually grouped into like terms, and the letter "a" in this case is a constant (i.e. not a variable), we would rewrite this expression as:

$$f(x+a) = x^2 + (2a + a^2)x + a^2 + 3$$

In this process, we have done nothing more than apply the rules of algebra to our function, but the process of evaluating functions with algebraic expressions as arguments will be unfamiliar to many students. Many will try to use some sort of shortcut to avoid expanding terms as necessary.

An important distinction should be drawn between the terms "function" and "equation", and how the graphical representation of a function can help us to solve an equation. For example, the table on page 1 displays the output values for $f(x) = x^2$ when evaluated for different values of the input variable x. This enables us to graph the function on the $x - y$ plane for any value of x, as seen on the top of page 2.

Alternatively, when we consider an equation with x^2 in it, for instance $x^2 = 4$, we are asking for the specific point on the graph of $f(x) = x^2$ that equals 4. Instead of the expression containing the dependent variable y (or $f(x)$), we are substituting a specific numerical value for the dependent variable, and determining what value or values of the independent variable x that satisfy this condition. Whereas a function represents the general rule to calculate the value of $f(x)$ for any input value, an equation asks for the value or values of the input that yields a specific output value for y. It therefore usually only has a finite set of answers. On a graph, an equation corresponds to particular points on the curve we have drawn, whereas a function refers to the entire curve.

In the case of $x^2 = 4$, there are two points on the parabola where $f(x)$ or y equals 4, so there exist two answers to this equation: $x = +2$, and $x = -2$.

Similarly, when calculating the $y-$intercept of a function, such as $y = 2x + 3$, we are asking for the $y-$value when $x = 0$. So we would substitute the value 0 for x, and arrive at the equation

$$2(0) + 3 = y$$

which tells us the value of y when x equals 0, called the y-intercept. In this case, the $y-$intercept is 3.

The graph on page 3 illustrates the relationships between graphs and equations, by setting the values of two functions equal to each other. If one function is represented by $f(x) = 2x + 3$, and the other is represented by $g(x) = x^2 + 2x - 1$, to find the points where the graphs of these two curves intersect entails finding the place where $f(x) = g(x)$ for a given value of x. To determine these values, we write the equation

$$2x + 3 = x^2 + 2x - 1$$

and solve for the values of x where this equation is true. This example requires using techniques for solving quadratic equations, as shown on Page 3. Again it turns out that there are two answers for x, corresponding to the two points of intersection for the graphs of the functions $f(x)$ and $g(x)$.

Although much of the notation introduced in this and subsequent lessons is very formal, it is important to stress that functions are important because they enable us to model a number of real world phenomena. In the exercises for this chapter, the relationship between the independent variable x, and the example of modeling costs using both linear and nonlinear functions, is emphasized. By using functions to model real world phenomena, we find that properties of functions like slopes and intercepts correspond to actual real world phenomena, like break even points, fixed and variable costs, as well as velocity and acceleration.

Lesson 2: Relations and Functions

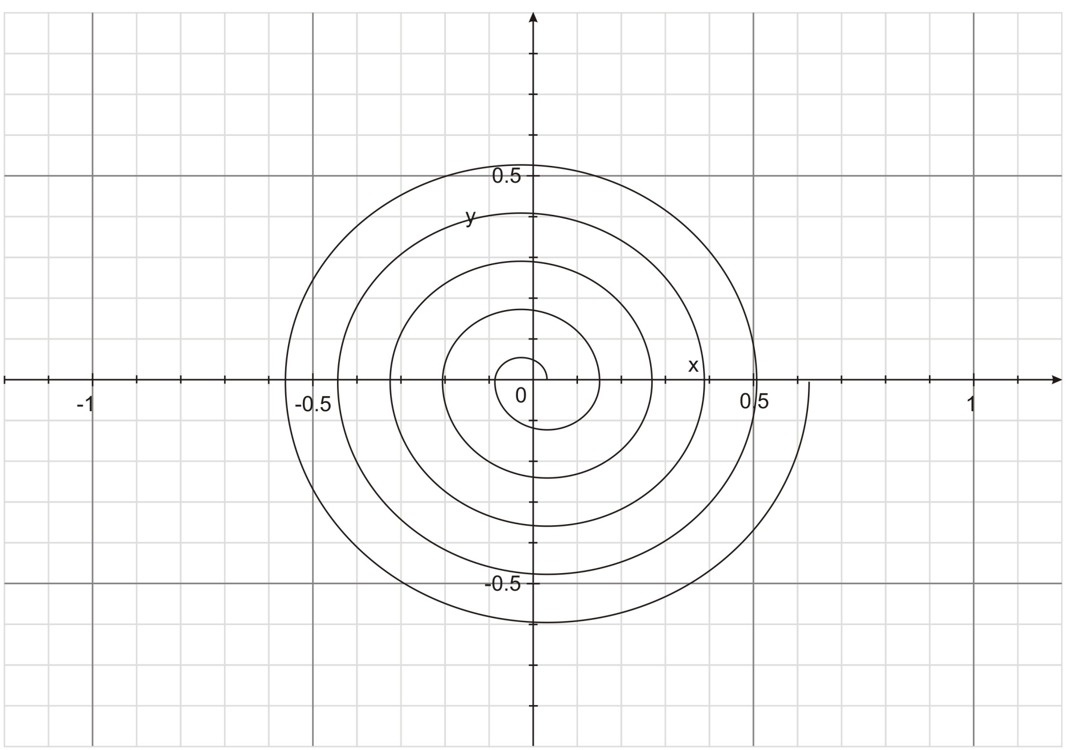
In Lesson 2, the more formal definition of a function is introduced, as are the topics of Domain and Range which provide useful information for analyzing and graphing functions. The classic definition of a function when displayed graphically is that it is a curve on the $x - y$ plane that must satisfy the "vertical line test", e.g. if you draw a vertical line through the function, it touches the graph in at most one point. This test ensures that for any value of x, there is at most only one value that our function evaluates to for this input. This is sometimes referred to as being "onto" or "surjective".

The example given in this Lesson of a common graphical representation which is NOT a function is the graph of a circle. There are clearly places where if we were to draw a vertical line on the coordinate plane, it would cross the circle twice. Although the rationale behind this isn't explained in this Lesson, it may be helpful for students to be shown why an equation like $x^2 + y^2 = 4$ will not be a function, whereas an equation like $x^2 + y = 4$ does turn out to be a function. The answer becomes clear if we were to isolate the variable y:

$$\begin{aligned} x^2 + y^2 &= 4 \\ y^2 &= 4 - x^2 \\ y &= \pm\sqrt{(4 - x^2)} \end{aligned}$$

By isolating y, we see that for a given input value of x, there can be two values of y due to the plus/minus in the square root. Because there are two output values for only one input value, this is NOT a function, and thus does not pass the vertical line test.

When treating functions in the context of the $x - y$ plane, it often appears that the variety of curves that are functions is very limited, since there are a number of interesting curves which don't satisfy the vertical line test. These include the circle graphed in the text, and the spiral graphed below. Can the techniques we develop to analyze functions be applied to these non-functions?



Although it is outside the scope of this textbook, most students will have been introduced to the concept of a “parametrized curve” in their pre-calculus course. Parametrizing a curve enables us to consider curves that are not functions, like the circle or spiral, and represent them AS functions so that we can analyze them using function-based techniques. This entails creating a new variable, or parameter, and re-writing our expressions for the $x-$ and $y-$coordinates of our curve using this parameter. For instance, if we were to create a new variable named t, referred to as our parameter, we could describe the circle in this Lesson using the equations:

$$x = \cos(t), y = \sin(t), 0 \le t \le 2\pi$$

In this case, both of our “parameterized” equations ARE functions: $\cos(t)$ and $\sin(t)$. By using techniques like parameterization, we can transform curves that are not functions into representations which ARE functions. This dramatically increases the class of curves and graphs which we can analyze.

The bulk of this lesson is devoted to reviewing the topic of a function’s Domain and Range, which define the values of x and y over which a given function extends.

Determining the domain of a function is usually much easier for students than finding its range, since there are only a finite number of situations where we cannot evaluate a function at a given $x-$value. The two most common are dividing by zero and taking the square root of a negative number. In looking at a function to determine its domain, most often we are simply looking for cases where a particular value of x will lead us into one of these conditions of undefinedness, and exclude those values.

Take, for instance, the example of the rational function $f(x) = \frac{1}{x}$ given on page ###. In determining the values of x for which this function exists, clearly we must exclude the value $x = 0$ since we are not allowed to divide by zero. Since there are no other opportunities for our equation to be undefined through either dividing by zero or taking the square root of a negative number, this is the only point excluded in our domain. We can therefore define the domain as:

$$D = \{x | x \neq 0\}$$

Similarly, if we were to look at the rational function:

$$f(x) = \frac{1}{(x-2)(x+3)}$$

the denominator in this expression will be equal to zero when the product $(x-2)(x+3) = 0$. This happens when $x = 2$ or $x = -3$. In this case, our domain includes all values of x except for $x = 2$ and $x = -3$.

The determination of a function’s range is much more complicated, since it often requires a great deal of intuition into the behavior of algebraic expressions to understand which values a complicated function can and cannot take. For instance, terms in polynomials which raise x to an even power will always be positive, and the sine or cosine of a variable will always range between -1 and 1. An excellent process to help students identify the range of a function, particularly one that has many terms, is to look at the range of the individual terms, and combine them through logical reasoning to determine the range of the entire functions.

For example, consider the following function:

$$f(x) = x^2 + \cos(x)$$

Can we determine what values $f(x)$ will take by just looking at this expression? Looking at the first term, we know that the range of x^2 is always greater than or equal 0, since x^2 can never be negative. Moving to the next term, $\cos(x)$, which we know cosine is always between +1 and −1. Combining these facts, we see that $f(x)$ can never get less than −1, but can grow positively as large as we want due to the term, x^2. We can therefore say that the Range of this function is at the very least $f(x) > -1$, since $f(x)$ can only get as small as −1. Though it turns out that the range is actually more restricted than this, this type of reasoning provides students with an example of bounding a range to a particular interval.

To many students, understanding the domain and range often becomes formulaic, with little or no motivation as to why these terms are important. In keeping with the importance of understanding both the practical applications of functions, and being able to graph functions and identify functions from their graphs, there are two ways to stress the usefulness of determining the domain and range. First, in many situations in physics, engineering and the natural sciences we derive equations for quantities like cost, weight or distance utilizing functions and algebraic expressions. Understanding the properties of an answer we attain for such quantities, like its domain and range, enables us to check the validity of our solution through physical intuition.

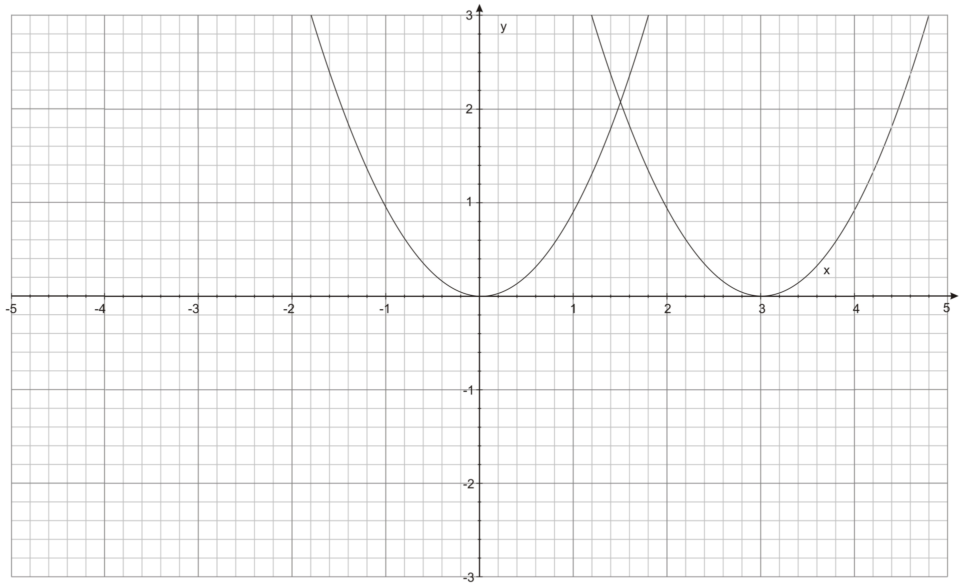
For instance, when using a function to calculate a quantity that must be strictly positive, like height or weight, if we are using a function whose range contains negative values, we should be wary. In some instances, this is a sign that we have improperly modeled the physical situation at hand. In others, it is a sign that the function we have computed is only valid on certain intervals, for example those values of x which make the function positive. In order to remain consistent with the reality of the physical situation when our range extends to values that seem impossible, we often "restrict the domain", meaning that we exclude the values of the input variable which lead to the impossible values of the range.

This Lesson contains a showcase of the graphs of many important types of functions that we will encounter throughout this textbook. A student should be able to identify these graphs quite easily in the first few weeks of class. They should be able to determine the intercepts and locations of important features of a function, such as the focus of a parabola, the center of a circle, and the domain of the logarithmic function. Most, if not all, of these topics should be review, but a strong understanding of these fundamentals will be important to developing the more complicated topics in this book.

In anticipating the next few lessons on limits and derivatives, it might be helpful to have students recognize that the function $y = |x|$ is unique amongst the functions showcased. All of the other functions except $|x|$ are smooth, meaning that they have no sharp corners or breaks in them. The absolute value of x has the sharp corner at $x = 0$, which is an example of a function having a point where the slope approaching from one side isn't equal to the slope when approaching from the other side. If one were to graph the slope of the function $f(x) = |x|$, soon to be referred to as its derivative, we would find that the value $x = 0$ would be excluded, making the derivative of $|x|$ a discontinuous function.

The final topic brought up in this chapter is function transformation. This is an important technique for interpreting functions that arise in modeling physical situations to understand the behavior of systems without graphing them. Transformations allow us to take a prototypical function, like one of the 8 showcased in the textbook, and alter their shape to get many different versions of them on the $x - y$ plane.

For instance, consider the parabola given by the formula $f(x) = x^2$. What if we wanted to move the graph of our parabola to the right by 3 units? As explained in this lesson, a rightward shift of 3 would be enacted by subtracting 3 from our variable, so instead of $f(x) = x^2$, we would get $f(x) = (x - 3)^2$. The graphs of these functions are shown below.



An important example of where the shift transformation arises in a physical contexts the solution to the Wave Equation in two dimensions. In that case, if we were to start a wave on the middle of a string that had a particular shape $f(x)$, we would get two copies of that wave, each half in amplitude, that move in opposite directions. This can be written as

$$v(x) = \frac{1}{2}f(x+ct) + \frac{1}{2}f(x-ct)$$

This expression tells us that we have two copies of our original function $f(x)$, divided in half in amplitude, with one copy shifted to the left by the product of c, the wave speed, and t, the time elapsed, and the other copy shifted similarly to the right. As time gets bigger, this shift grows larger, representing the wave moving away from its original position, and traveling along the string.

Transformations can create tremendous confusion for students because they appear in some ways the opposite of what one would expect. Take, for instance the shift of the function $f(x)$ to $f(x-c)$. Many students will think that because we are subtracting c, this corresponds to a shift in the negative direction. However, as we see above, by subtracting a constant c, we actually shift the function in the positive direction.

Similarly, if we consider the transformation $f(x)$ to $f\left(\frac{x}{2}\right)$, we might expect our original graph to be *compressed* by a factor of 2, since we are dividing by 2; conversely, if we consider the transformation of $f(x)$ to $f(2x)$, we might expect our graph to be *expanded* by a factor of 2.

In each of these 3 cases of function transformation, the opposite to what seems immediately apparent turns out to be true. If we transform our function $f(x)$ to $f(x-c)$, we are shifting our function to the *right* by the value c. Transforming $f(x)$ to $f\left(\frac{x}{2}\right)$ *expands* our original function by a factor of 2, and transforming $f(x)$ to $f(2x)$ compresses our original function by a factor of 2. These caveats should be emphasized at this stage to ensure that a student is able to easily identify how to graph common functions which have been transformed through these standard operations (called a shift, dilation and contraction, respectively). The rationale for these operations can be deduced algebraically.

Lesson 3: Modeling Data with Functions

In this lesson, students use their graphing calculators to find curves which best approximate a set of data points on a scatterplot. This technique is often referred to as "regression" or "curve-fitting". Unlike traditional treatments of regression in statistics classes, which often focus exclusively on the topic of linear regression, Lesson 3 shows students that different sets of data are often best fit by a variety of different functions, depending on the visual character of the scatterplot. Though a linear approximation is sometimes the best approximation (and most often, the simplest), illustrating that we can also model data using higher order polynomials, trigonometric functions and transcendental functions may be new to many students.

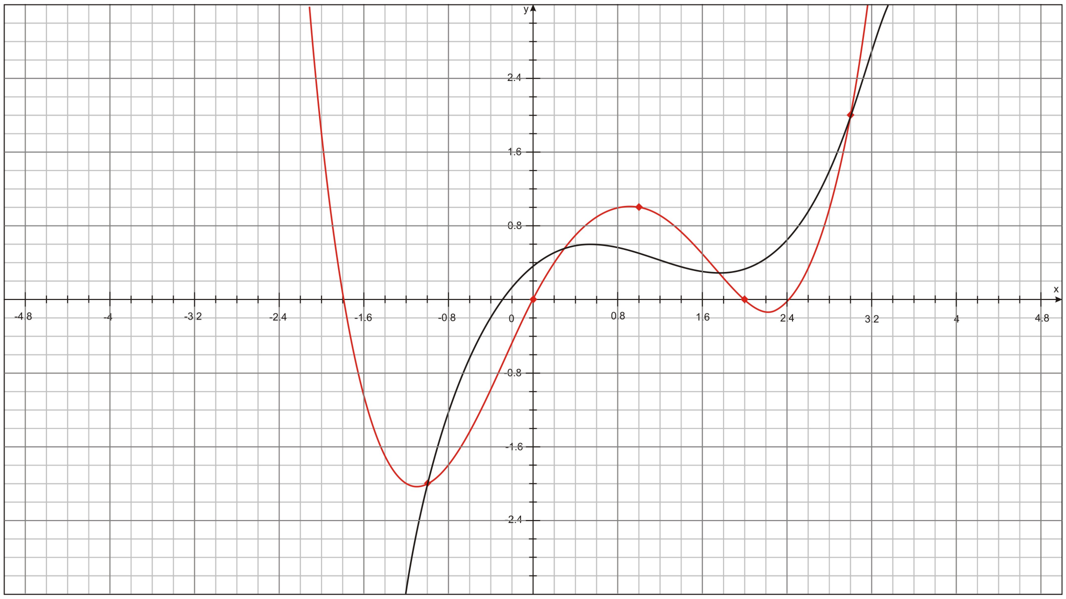
The handling of real world data, even sets as small as provided in this Lesson, is usually handled by a computer or calculator since the calculations involved in determining the curve of best fit can be quite cumbersome. In this Lesson, calculating curves to fit the data is performed through both a graphing calculator as well as using Excel, and both are skills that a student should become comfortable with. It is important, however, to ensure that students understand the underlying reasoning their calculator is using to calculate curves of best fit since the criteria we can use to measure "best fit" can be interpreted very differently.

In the examples given, the lines of best fit are calculated by minimizing the least square error between the curve and the data points. This meaning that if we were to add up the distance squared between the curve selected by our curve fitting technique, and all of the data points, the curve that is selected will provide the smallest value for the sum of the squared error.

$$\text{error} = \sum_{i=1}^{M} (f(x) - y_i)^2$$

This raises two important points that are hidden to the student if they exclusively rely on technology to find their curve. First, in some cases the use of the Least Squares approximation does not suit the purpose we are trying to achieve by fitting the data with a curve. Secondly, the Least Squares Approximation is so widely used because it is a Quadratic function, and thus is guaranteed to have a unique extreme value as illustrated in Lesson 2. This is helpful to only have a measurement of error that has only one unique maximum or minimum.

Take, for instance, the curves below which are both trying to approximate the same data set:



Clearly the red curve has the lower least square error, since the curve runs exactly through all of the points. Thus its total error is zero. But if we are trying to capture the trend of the data, however, the black curve is much better since it captures the undulations of the data, as well as the likely trend of the data at the end points. This is true even though it doesn't exactly fit most off the points. This is an example of where using the Least Square Error to fit our data may not be in our best interest, and is referred to as "over-fitting".

In other cases, the presence of outlying and anomolous data may seriously affect the calculation of our regression line if we use the Least Squares Error as our criterion for curve fitting. If a student were to rely exclusively on the answers provided by their calculator, they may arrive at curves which are not appropriate to model the data they are interpreting.

An additional source of confusion for many students in understanding curve fitting is the difference between parameters and variables in our original function. For example, consider the standard equation of a line:

$$y = mx + b$$

In this equation, the variables are y and x, and the parameters are the values m and b. For any particular line, m and b are fixed and determine the character of the line we are graphing. In the case of curve fitting, however, we now treat our values of m and b as variables, and try to find the values of these variables which optimize our problem in some sense (for instance, minimizing the Least Square Error). Once we find these values, we plug them into the values of m and b in the equation above, thus providing constant values for these parameters. For instance, on Page 29 the regression line found by the graphing calculator is $y = 0.76x + 14$. To arrive at the values of $m = 0.76$ and $b = 14$ we had to allow these values to be variables and solve for them through a Least Squares technique. Once we have solved for them, they become constant values in our function.

Lesson 4: Introduction to The Calculus

Perhaps the most important topic in all of calculus is the concept of a limit, which is introduced in this Lesson through two of its most common uses: finding the slope of a tangent line to a curve at a given point (the derivative), and finding the area under a curve (the integral).

The tremendous usefuleness of a limit can be seen most clearly in the definition of the derivative, or finding the slope of a curve that is continuously changing, using what are referred to as secant line approximations. As discussed in this Lesson, the equation for the slope of a line is given by:

$$\text{slope} = m = \frac{y_2 - y_1}{x_2 - x_1}$$

In determining the slope of the line that is tangent to the curve at a single point, we can interpret this as moving the two points between which we draw our secant line closer and closer to the point where we are looking for the tangent line to. To determine the slope at the point of interest, we can calculate the slopes of the secant line as we move its two endpoints closer and closer together.

A problem arises, however, when those two endpoints are brought together until their distance in the $x-$direction, i.e. $x_2 - x_1$, becomes zero, since it will lead us to divide by zero in the slope equation above. In most cases, this would make our slope fraction undefined, meaning it didn't have a slope. However just as the denominator becomes zero, the numerator also becomes zero, since now both values in the difference expression in the numerator, $f(x_2)$ and $f(x_1)$ become the same. We know that any fraction with a zero in it is equal to zero, but any fraction with a zero in the denominator is undefined; when both are true, we are left with an indeterminate situation. This is called "an indeterminate form", and is what motivates most of the techniques developed in calculus.

It is problems like this which gave rise to the study of limits. Instead of considering the the equation above when $x_2 = x_1$, or when the secant endpoints' $x-$values become exactly the same, we consider what happens in the limit as x_2 approaches x_1. This is written as:

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

Similarly, when we try to approximate the area under the curve as a set of rectangles, known as calculating Riemann sums, we must make our rectangles narrower and narrower to better approximate what the area under the curve actually looks like. This corresponds to making our rectangles "infinitisemally wide" (i.e. their width approaches zero), and so the expression for adding up the areas of these rectangles ALSO becomes indeterminate. The process of calculating this area in the presence of such indeterminancy is referred to as integration. Chapter 4 introduces the limiting operations needed to perform integration.

This lesson gives us a concrete example of how the technique of finding a derivative works by providing a table of values for a set of secant line approximations, and the sum of rectangular area approximations, as we move the endpoints of our interval closer and closer together. In these cases, the value that we are trying to determine (the slope of the secant lines, and the area of the sum of rectangles) start to converge towards a particular value. If this continues to happen as the distance between the endpoints gets smaller and smaller, then we say that the limit exists, and it is given by the value that these approximations approach.

It might be helpful for some students to show them examples where the above techniques don't work, particularly in cases where the derivative of a function does not exist, and thus the limit above does not exist. Take, for instance, the absolute value function $|x|$ we looked at in Lesson 2. If we were to construct a table of values for this situation, we would find that the slopes of our secant lines DON't approach the same value depending on how we move our two endpoints closer and closer together. This is a case where the limit of the above slope equation does NOT exist. These caveats will be discussed more thoroughly in Chapter 3.

Lesson 5: Finding Limits

The beginning of this Lesson takes students through the exercise of finding a limit using a table of values, but this time by using their graphing calculator on functions which would be much more difficult to evaluate by hand. In particular, we consider the function:

$$f(x) = \frac{x+3}{x^2+x-6} = \frac{x+3}{(x-2)(x+3)}$$

An important point to raise here is the danger of relying on results from a graphing calculator or computer program, particularly when it comes to using limits. For the equation above, the calculator claims the value of the function is undefined at the two places where the denominator becomes zero, which may lead some students to believe that it doesn't have a limit as it approaches these points. But as we learned in the last lesson, this function actually can be evaluated in the limit as x appoaches -3, even though it would appear as if the zero in the denominator would be catastrophic.

The reason again that this function does have a limit as x approaches -3 is that even though the denominator becomes zero, the numerator in this case also becomes zero. Therefore we have the indeterminate form $\frac{0}{0}$, which may or may not have a finite value. It is important for students to understand that this is why we can have a limit at $x = -3$, but we do not have a limit as we approach the other point where the denominator becomes zero, $x = 2$. When $x = 2$, the expression becomes $\frac{5}{0}$ which is clearly undefined.

As mentioned in the previous Lesson, it can be helpful for many students to be shown examples of when limits do NOT exist in order to understand how and when they are useful. In the case above, we considered the case where the derivative of a function does not exist at a particular point as motivated by the formula for the derivative which incorporates the limit. In that case, $|x|$, the function DID exist at the point $x = 0$, but the derivative didn't. In the function above, at the point $x = 2$ the function itself does not exist, an example of a function being discontinuous. This can be seen very clearly from the graph of the function, since as we approach the point of interest, $x = 2$, from both sides, the value of our function approaches very different values. As we move towards $x = 2$ from the negative side, our function approaches negative infinity, whereas from the positive side, it approaches positive infinity.

This example offers a good opportunity to present the formal definition of a limit provided at the end of this Lesson. This definition presents difficulty for even advanced students of mathematics, and is primarily used for the topic of real analysis. Intuitively, we can understand the formal definition of the limit in the context of our function which does not have a limit near the point $x = 2$. The definition of a limit tells us that if we are in an arbitrarily small interval around $x = 2$, then we should be able to find an arbitrarily small interval in the $y-$direction in which our function must be. Clearly if we were to look at an interval around $x = 2$, however, our function is not contained in an arbitrarily small interval, since it approaches negative infinity on one side of $x = 2$, and positive infinity on the other side of $x = 2$. There would not be an interval in the $y-$direction around the function at $x = 2$ which would be able to bound the entire function. Hence the limit does not exist.

Lesson 6: Evaluating Limits

As mentioned in the last Lesson, the use of the formal definition of a limit can be quite cumbersome, and is rarely used in most situations that a student will encounter. It will be helpful in understanding theorems and concepts later in Calculus and Analysis, but the actual task of finding a limit is usually much more straightforward. Sometimes it can be as simple as performing direct substitution into our algebraic expression as we saw in Lesson 4. But in most real applications, more complicated techniques are required.

In this Lesson, the student is introduced to some important properties of limits which will enable them to use the technique of substitution with more complicated algebraic expressions. Although the properties of limits being additive should seem intuitive, it should be noted that the truth of these properties in the presence of multiplication, division and raising expressions to exponents is much more subtle. In each of these cases, it can be seen that the operation of taking a limit is commutative, distributive and associative over multiplication and addition. This enables us to use substitution to find the limit for arbitrary polynomial expressions by substituting our value of interest for the variable in each term.

The next two techniques to finding limits will appear much more strange to students, especially since the motivation of how and why they work is reserved for later chapters when the concept of a derivative has been introduced more thoroughly. The important point to stress here is that the simple technique of substitution is often insufficient to finding limits of many important functions, and more sophisticated techniques exist in many of these cases.

The first such technique entails finding limits to rational expressions when they take on an indeterminate form. Indeterminate forms are examples of where substitution is not sufficient to finding limits of expressions, since when we have a zero divided by another zero, or infinity divided by another infinity we cannot be sure what the value of our substituted algebraic expression is. This lesson exposes students to handling such situations more formally, and introduces them to the concept of a "removeable singularity". Even though the denominator does become zero at the point of interest, because the numerator also becomes zero in exactly the same way, we can remove the denominator by factoring the numerator and cancelling.

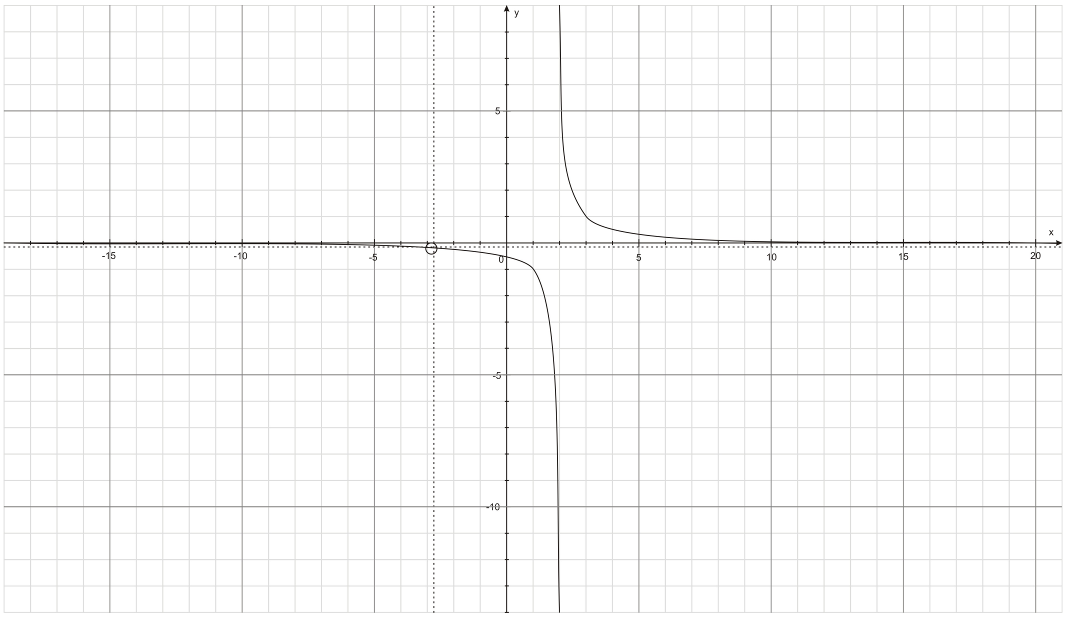
The second technique introduced in this Lesson as a way to calculate more complicated limits is called the Squeeze Theorem. The Squeeze Theorem states that if there exist two functions which bound our function of interest, meaning that our function is always in between the two other functions, and those functions both converge to the same limit at a point, then our function must also converge to that limit. What the Squeeze Theorem is saying is that we are squeezing our function in between two functions that are approaching the same value. Since they always bound the function of interest, the function of interest must necessarily also be that value, since there are no other values for it to be in between the functions that are squeezing it.

Lesson 7: Continuity

This Lesson builds on the previous discussion of limits to introduce the notion of a function being continuous or discontinuous. What we find is that even if a function has a limit at a particular point, that does not mean that it is defined at that point. This idea is illustrated by returning to the example of a rational function given above:

$$f(x) = \frac{x+3}{x^2+x-6} = \frac{x+3}{(x-2)(x+3)}$$

As students learned in Lesson 5, the above expression does have a limit as x approaches -3, even though the denominator is equal to zero at this point. This is seen clearly in the factored form of the expression, where the terms $x+3$ will cancel each other leaving no term in the denominator equal to zero. This type of discontinuity is referred to as a removeable singularity. On the other hand, this function does NOT approach a limit as x approaches $+2$, for reasons discussed above.



Near the point $x = -3$, the function behaves rather smoothly, because it does have a limit as it approaches -3 even though at that point it is undefined. This is represented graphically by a curve with a hole punched out at the point of discontinuity - this is called a point discontinuity. This example illustrates that just because a limit exists as we approach a certain point, it does not mean that the function itself has to exist at that point.

At the point $x = 2$, however, our discontinuity appears very different. In this case, we have a discontinuity that is manifested as our curve approaching positive infinity and negative infinity in opposite directions depending on the side of $x = 2$ that we are on. At $x = 2$, there exists a "vertical asymptote" at the point of discontinuity. This is also a very common occurence for rational expressions which appear in physical situations. In this case $x = 2$ is referred to as a pole.

The important point to note is that we can have different types of discontinuities, and we often use limits to determine which type we have. This helps us to understand the behavior of our function around points of interest, like the places where the denominator equals zero, without the trouble of graphing the function.

The final topic discussed in Lesson 7 is one-sided limits, which you may have already addressed in the context of the absolute value function, $f(x) = |x|$, in Lesson 2. In that discussion it was noted that as we considered the slope of our function as we approached $x = 0$ from the left, our slope was -1. If we approached $x = 0$ from the other side, our slope was $+1$. Because our slopes didn't match, and abruptly changed from -1 to $+1$, the slope of our line is undefined at $x = 0$. If, however, we defined the existence of a limit as only needing to be valid as we approached from one side, we would say that the slope DOES have two one-sided limits, just not a two-sided limit.

In cases like the square root function shown in the text, it is seen that because our function is only defined on one side of the point $x = 0$, we would have to conclude that the function square root of x does not have a limit there, since we cannot approach from the left. In such cases, the one-sided right hand limit (i.e. approaching from the right) does exist, but the one-sided left handed limit (i.e. approaching from the left) does not.

Lesson 8: Infinite Limits

The last type of limits treated in Chapter 1 are those where the value we are approaching is either infinity or negative infinity. This allows us to determine the behavior of our function as it extends infinitely in both directions along the $x-$axis. By understanding the behavior of our function as it approaches positive and negative infinity, we will be able to without having to plug in multiple values as x gets very large in the positive and negative directions.

Up until this point, the only example of an indeterminate form that students have been exposed to is $\frac{0}{0}$. We noted in those cases that we could not determine an answer for our expression by mere inspection, since most fractions with a zero in the numerator are equal to zero, whereas a most fractions with a zero in the denominator are undefined.

Which one of these facts wins out in such cases? This is where the techniques of finding limits proved useful.

Similarly, what if we are presented with a fraction where the numerator is approaching infinity, and the denominator is approaching infinity?? This is another situation in which we are presented with an indeterminate form. A fraction whose numerator gets bigger and bigger will grow larger and larger, but if its denominator is also growing larger and larger, then the value of the fraction will get smaller and smaller. These are important points to stress to students, since it may be one of the first times that they have considers what happens to an expression when the variable gets infinitely large. It is an essential tool to analyzing many important physical situations.

An important point to stress to students, and to make sure that they understand, is that as the denominator of a fraction approaches positive or negative infinity, the value of the overall fraction approaches zero. This is a point that can be stressed by looking at some examples of fractions with large numbers in the denominator. Take for instance, the sequence:

$$\frac{1}{10} = 0.1$$
$$\frac{1}{100} = 0.01$$
$$\frac{1}{1000} = 0.001$$
$$\dots$$

As the denominator gets larger, the value of our fraction gets smaller. We say that in the limit as our denominator becomes infinite, the value of our fraction becomes "infinitesimal". Being able to quickly understand rational expressions that have complicated algebraic terms in them, and to identify what happens when the independent variable gets very large, is an important tool in understanding physical situations.

An example of this is the determination of steady state behavior of a physical system. In many cases, the behavior of a physical system can be modeled as an exponential function, such as

$$f(x) = e^{ax}$$

where a is a constant. Let's consider the case where a is positive, and our independent variable "x" is a measure of time. For instance, suppose we pluck a string on a musical instrument, and listen to the sound over time. What happens to the sound as time goes on for longer and longer?

In the case where a is positive, if we tak3e the limit of the expression above as x goes to infinity, we find that our function also grows to infinity! This is true, since a positive value greater than 1, when raised to a large positive power, grows larger and larger as we raise it to larger and larger powers.

However, what happens if a is negative? Now, as x goes to infinity, the expression goes to zero, since:

$$\lim_{t \to \infty} e^{-at} = \lim_{t \to \infty} \frac{1}{e^{at}} \to \frac{1}{\infty} \to 0$$

These behaviors are observed in the graphs of these functions.

Often when we are designing mechanical or electrical systems, the type and quantity of materials we use will affect the value of the constant in the exponential term that describes our system. By designing the system in such a way that this constant is negative, we guarantee that after a long time, any input to our system eventually dies away. If we designed the system in such a way that the exponent was POSITIVE, we would find that any input to our system would make its output start to grow very very large, without any bound. This is referred to as an "unstable" system, and the task of design then becomes to make our exponents negative, the importance of which is seen by taking the limit of our expression as time goes to infinity.

1.2 Differentiation

Lesson 9: Tangent Lines and Rates of Change

It will be helpful to students in this Lesson if you have already introduced them to the formula for the derivative in the Lessons about limits above. In this lesson we formally define the equation of the tangent line to a curve by considering the slope of the secant lines as their two endpoints get closer and closer together. In many respects this equation, known as the first derivative in one dimension, forms the heart of a student's understanding of calculus for years to come. It is their first illustration of "the calculus", or how to calculate important quantities from indeterminate forms by understanding the infinite and the infinitesimal.

It will also be helpful in this Lesson if students have been introduced to substituting algebraic expressions into functions, such as $f(x+h)$, and not only substituting numbers. At this stage of their mathematical development, most students are not yet comfortable with substituting binomials into functions, and this technique is essential to understand most of the subsequent work we will do in understanding the derivative. The complicated mechanics of this process is described in Lesson 1 above. If you have not introduced them to this concept already, it is essential to do so at this time.

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

The power of expressing the equation of a function's tangent line using the equation above is that some very remarkable things happen once we start to substitute things in. If you remember, one motivation for finding limits was to deal with situations where a rational expression has both a zero in the numerator and the denominator. What happens in the slope equation above is that for some expressions, the h in the denominator gets cancelled. If you remember from Lesson 1, we calculated the expression $f(x+h)$ in the case that $f(x) = x^2 + 2x - 3$, and arrived at:

$$f(x+h) = x^2 + (2h+2)x + (h^2+3)$$

If we substitute this into the expression for $\frac{df}{dx}$, we will find that we can cancel the h in the denominator, thereby permitting u to substitute $h = 0$, and not have our expression become undefined. The discovery that algebraic expressions can be manipulated like this to determine their values even when they are indeterminate was the fundamental breakthrough of calculus.

Lesson 10: The Derivative

This Lesson ties together many of the topics brought up in this Teacher's Edition, and so tips on its content can be found sprinkled in the pages above. If you have not already introduced students to the idea of continuity and differentiability in the context of the absolute value function, or stressed the techniques of evaluating the formula for the derivative discussed above, then all of those concepts will need to be brought up in this lesson.

The technique of using the formula for the derivative developed in Lesson 9 is applied to more complicated functions, and the importance of the algebraic techniques utilized should be stressed heavily. In using the derivative formula to determine the slope of an unknown function, the greatest trouble students face is making algebraic mistakes and then trusting their answers, leading to an improper intuition of the underlying mathematical structures. This should be emphasized here, and their facility with making complicated algebraic substitutions into the derivative formula should be tested. As students will see in the next lesson, the expressions they are differentiating in this lesson are illustrating examples of "the power rule", a powerful technique to find derivatives of polynomial functions.

The different notations for expressing the derivative of a function are introduced in this Lesson, and students may wonder why there are so many different ways to express the derivative. There are a number of answers to this question. In general, there are three common ways in which a derivative is expressed:

$$f'(x), f(x), \frac{df}{dx}$$

The first and second of these, or prime and dot notation, are only useful for cases where we have a single independent variable, which is usually x or t. The prime notation is the generic form of a first derivative, whereas the dot notation is usually reserved for cases where we are taking the derivative with respect to time, such as calculating rates or velocities. These two forms are very useful when we want to keep our entire expression on one line, as we often will when solving differential equations. However, a great deal of meaning is lost in this notation as opposed to what is known as differential notation, or Leibniz notation.

This notation expressed the role of the derivative as a rate of change, or a slope. When the change in y (or $f(x)$) and x are finite, we denoted them with deltas in front of them.

$$\text{slope} = \frac{\triangle y}{\triangle x} = \frac{y_2 - y_1}{x_2 - x_1}$$

When these changes become infinitesimally small by taking the limit as the differences become zero, we denote them with a d in front of them, and they are referred to as "differentials". So, dy is referred to as "the differential of y", and dx is referred to as "the differential of x". The ratio of them is thus the change in y over the change in x, which is the equation for a slope.

The reason why differential notation is important is that it helps many students understand more complicated topics, like the product rule, chain rule and implicit differentiation. The differential form of the derivative allows us to the rules of fractions to these terms, and the rationale for topics such as related rates and implicit differentiation become much easier to understand.

The final topic in the Lesson concerns the definition of differentiability and how it relates to continuity. When we take the derivative of a function, what we get is another function (usually). If a function has a place where it is not differentiable, then the function of its derivative will have a place where it is not continuous. By definition, a function is said to be "differentiable" if the function for its derivative is continuous. Furthermore, on order for a function to have a derivative at a particular point, the function must exist at that point. So a function that has an open hole in it, a discussed in Lesson 7 on Continuity, will not have a derivative at that point, even if its slope could be estimated by using the Table of values method. This fact is by definition.

However there are a number of important functions that ARE continuous, but don't have a derivative, including the absolute value function as discussed above. Three other important examples of functions which are continuous but not differentiable are presented in this lesson: the "cusp", "vertical tangents" and "jump discontinuities". These are examples of functions which appear often in physical situations, and thus techniques to handle them analytically are important to understanding those systems. In the case of the jump discontinuity and cusp, if we were to use a one-sided limit in our definiton of the derivative, we could see that both of these examples' slopes have one-sided limits as they approach the point of indifferentiability, but being differentiable requires that we have a two sided limit as we approach the point of interest.

Lesson 11: Techniques of Differentiation

This Lesson forms a turning point for a student's understanding of derivatives, as now we are able to move away from the formal definition of limits, and its algebaraic complications, and move to more general rules for differentiating a number of important functions. We begin by treating the important class of functions known as polynomials. We also introduce two of the most important formulas in all of calculus: the definition of the product rule and the quotient rule.

In previous applications of the definition of the derivative, we have always been looking at a particular function, such as $(x) = \frac{1}{x}$ or $f(x) = x^2 + 6x + 3$. In this lesson we introduce a more powerful application of the derivative definition to generic formulas for important types of functions, like polynomials, trigonometric functions and exponentials. To do so, we need to have a notation for representing entire classes of functions, which introduces a much greater degree of complexity into a student's understanding of functions. Take, for instance, the representation of a polynomial in generic terms. In fact, there an infinite number of different types of polynomials, but we can write them all in a generic form using "sigma notation", which is introduced for the first time in the text.

In sigma notation, since any polynomial is some sum of a bunch of powers of a variable with numbers in front of it (called coefficients), we can write a generic polynomial as:

$$\sum_{i=0}^{n} a_n x^n = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^N$$

Because we now have to represent ANY type of polynomial, we cannot just use the letters of the alphabet, like we do when we represent the line as $y = mx + b$, or the parabola as $y = ax^2 + bx + c$. We would run out of letters after 26 terms in our polynomial! We therefore introduce subscript notation for a constant - the n^{th} power of x has the coefficient a_n.

As noted in the text, we do not prove the Power Rule, but it can be very helpful to do this as an exercise with the students, to help them gain comfort with the next stage of algebraic substitution into the derivative formula – the case of substituting arbitrary functions of a particular class, like the polynomial above. What we find is that by simply applying the rules of function substitution that we have been practicing throughout the preceding lessons, we are able to arrive at very interesting results for the derivatives of these classes of functions.

By applying the expression for the derivative function to all members of a particular class of function, we often discover that that entire class of functions has a rule which describes how to differentiate functions of that type. When we do this with the generic expression for a polynomial provided above, we arrive at the Power Rule:

$$\frac{d\left(\sum_{i=0}^{N} a_n x^n\right)}{dx} = \sum_{i=0}^{N-1} n a_n x^n$$

The Power Rule represents one of the most important and useful properties of derivatives: the derivative of a polynomial is another polynomial that is always one degree less than the degree of the original polynomial. So the derivative of a quadratic function, or second degree polynomial, is a line, which is a first degree polynomial. The derivative of a cubic function, a third order polynomial, is a quadratic function, a second order polynomial. This is an important point for students to understand, particularly in the context of measuring position, velocity and acceleration.

As will be discussed later in the text, the velocity of an object is the derivative of it position, and the acceleration of an object is the derivative of its velocity. This can be illustrated by the case of throwing a ball in the air, which behaves very similarly to a quadratic position. If you were to measured and graphed the position of the object, it would appear to be parabolic. If you were to graph the same object's velocity it would be a line, since that is a polynomial one degree less than a parabola. The acceleration, or derivative of the velocity, would then be a constant, since that is a "zeroth" order polynomial, which is one degree less than a line. In the case of throwing a ball in the air, this constant represents the constant gravitational pull of the Earth, or 9.8 m/s2 or 32 ft/s2.

The next part of this Lesson introduces many of the important properties of derivatives, which also can cause quite a degree of confusion for students. Each of these results can be proven by returning to the definition of the derivative, and applying the appropriate rules of algebra. This is an excellent exercise for the students to ensure that they understand the technique of substituting generic algebraic expressions into the derivative formula as discussed above in the case of polynomials. This is especially easy in the case of the constant rule and addition subtraction rule. In the case of the Product Rule and Quotient Rule, the reasoning is much more sophisticated, and presents a situation that most students do not expect and forget to use when confronted with complicated algebraic expressions.

Take for example the rational expression:

$$f(x) = \frac{x}{x^2+3x-4}$$
$$f'(x) \neq \frac{1}{2x+3}$$

Out of laziness, many students will calculate the derivative of the numerator and denominator, and express the derivative as the ratio of these derivatives. Same is true for the product of two expressions:

$$f(x) = (2x^2+4x)(x^2-3)$$
$$f'(x) \neq (4x+4)(2x)$$

What the product and quotient rules tell us is that these are NOT valid calculations. For many students, this is the first time they may have encountered such situations, since traditionally in algebra multiplication and division act similarly to addition and subtraction. When we take the derivative of a product and quotient of functions, the result does NOT behave as the derivative of the sum or difference of functions. The ability to recognize situations where the product or quotient rule are necessary, as well as the ability to execute the necessary calculations to apply them, are two separate but very important skills students need to master to understand how calculus works.

To be most precise, the formula we have been referring to as “the derivative” is more specifically called “the first derivative”. This explains the single prime in $f'(x)$ and corresponds to taking the “first order difference” in calculating the numerator and denominator in the slope equation. However besides just analyzing the slope of a function, we can also interpret a number of other important visual qualities of a function through the process of taking derivatives of derivatives, or higher order derivatives.

As was discussed above, the first derivative of a quadratic equation, which graphs as a parobola, is a first order polynomial, or a straight line. The second derivative of the parabola is defined as the first derivative of its first derivative, which in this case is a straight line. Since the first derivative of a straight line is a constant, we say that the *second* derivative of a parabola is a constant. The physical meaning of this will be discussed more thoroughly when we consider the First and Second Derivative Tests in Chapter 3, but for now the definition of higher order derivatives as derivatives of derivatives, as well as seeing how patterns emerge in the taking of higher order derivatives, is a sufficient introduction to this topic for students.

As an example of where using higher order derivatives is essential in understanding physical applications, this Lesson considers the example of a “second order differential equation”, which will be considered again in Lesson ??. Just as we often arrive at algebraic equations to solve when presented with a word problem or physical situation, in more advanced contexts we often arrive at what is called a “differential equation”, where the terms in our equation are not only unknown quantities, but also the derivatives of unknown quantities. Let’s say, for instance, that we wanted to know the function y which has a slope of 4. We would want to solve for y in the equation:

$$\frac{dy}{dx} = 4$$

This is an example of a “first-order differential equation”, since it only contains terms which have at most a first derivative. The techniques of calculus, particularly those of integration, will enable us to solve equations like this for the function $y(x)$. The examples in the text are examples of “second-order differential equations”, which provide students an opportunity to practice taking the second derivative of a function, using the context of an equation to check their answers.

Lesson 12: Derivatives of Trigonometric Functions

By the time they have arrived at this level of Calculus students should have already developed reasonable comfort with the three major trigonometric functions: sine, cosine and tangent. For sin and cosine, they should understand the idea of an amplitude, frequency and phase shift so that the expressions we develop for taking derivatives and integrals of trigonometric functions can be understood.

As introduced briefly above, the sin and cosine functions can also be part of algebraic expressions whose values at points of interest takes on an indeterminate form. The most classic example of an indeterminate trigonometric form appearing in physical applications is called the “sink” function:

$$f(x) = \frac{\sin(x)}{x}$$

If we were to look at the point $x = 0$, then we have an indeterminate form since both the numerator and denominator are zero. Similarly, since we know that the $\cos(x)$ is equal to "1" at $x = 0$, and equals "0" at $x = \frac{pi}{2}$, the following functions are also indeterminate forms:

$$f(x) = \frac{1 - \cos(x)}{x} @x = 0$$
$$f(x) = \frac{\cos(x)}{2x - \pi} @x = \frac{\pi}{2}$$

To understand how to take the limits of such expressions, first we must understand how to take the derivative of a trigonometric function.

The most important tool in finding the derivatives of a trigonometric function are the trigonometric identities. These often allow us to express the terms in the derivative expression more simply, which allows us to cancel and combine different terms. There are many important identities to remember, but the identities which refer to taking the sin or cosine of the sum or difference of angles prove most useful:

$$\sin(x \pm y) = \sin(x)\sin(y) \pm cos(x)\cos(y)$$
$$\cos(x \pm y) = \sin(x)\cos(y) \pm sin(y)\cos(x)$$

It is very helpful for students to consider the relationship between a sine function and its derivative, the cosine function, graphically. If we were to graph both sine and cosine on the same graph, what we would find is that the sine function, at $x = 0$, has a slope of 1, since that is the value of the cosine function at $x = 0$. As we move to the right, we see that although the sine function is increasing, the steepness of its increase is getting smaller and smaller. This happens all the way until $x = \frac{\pi}{2}$, when the slope of the sine function becomes zero, and turns to being negative until $x = \frac{3\pi}{2}$.

Knowing that the cosine function at a value of x is the slope of the $\sin(x)$, we see that at $x = \frac{\pi}{2}$ the cosine function becomes zero (corresponding the the sine function being "flat" at this point), and is now negative until $x = \frac{3\pi}{2}$. The fact that the slope of the sine function is equal to the cosine function is a remarkable property of these two functions, and is responsible for them appearing in a number of physical situations.

The remainder of this lesson shows students how to use additional trigonometric identities coupled with the product and quotient rules to take derivatives of complicated compositions of trigonometric functions. Although many students will later rely on memorizing the derivatives of each of these functions, the techniques of using identities and trigonometric properties is very helpful in developing facility with manipulating these functions, as well as understanding when and where the product and quotient rules are needed.

Lesson 13: The Chain Rule

The chain rule is one of the most important techniques for being able to understand the behavior of derivatives in engineering situations, since we often engineer machines to perform a sequence of mathematical operations on an input, which is represented mathematically as function composition. The chain rule enables us to determine the derivative of such function compositions, which do not behave in a manner that is immediately obvious or intuitive. The chain rule also provides an important example of where using the differential notation for the derivative provides information and the basis for intuition that the prime or dot notations do not.

There are many ways in which to interpret the chain rule, all of which provide intuition for its application in a range of situations. For this Lesson, we will focus on the using $u-$substitution in the formula for the derivative, since this is one of the most powerful techniques in calculus, and will appear many times throughout the course of this textbook.

The idea of u-substitution is similar to the idea of creating a dummy variable. Instead of dealing with the original expression we have been supplied with, we make up a new variable for that expression and see what happens. This introduces the ability to take derivatives of variables other than x which is the first time many of these students have considered this. This concept will continue to be discussed in the next lesson on implicit differentiation and in later lessons on integration techniques, but helping students develop intuition for (1) what it means to take the derivative with respect to a variable that ISN'T the independent variable x, and (2) knowing how and when to apply the product rule to generic algebraic expressions that may not seem in need of the product rule, are both ESSENTIAL skills for a student to grasp as we move into more complicated topics in calculus.

What the Chain Rule allows us to do is to take the derivative of a composition of functions, or $f-$circle$-g$. This enables us to apply the techniques we learned above for taking the derivatives of polynomials and other special functions to cases which don't fall neatly into one of these categories. To understand the chain rule and the use of $u-$substitution, we consider a simple example:

$$y = \sqrt{2x+3}$$

We learned above that if we have the expression:

$$y = \sqrt{x} = x^{\frac{1}{2}}$$

we can find its derivative using the power rule. However, now we have the case where

$$y = g(x)$$

A student's initial intution might be to simply apply the power rule to this expression, thereby arriving at the incorrect answer

$$y \neq \frac{1}{2} \cdot (2x+3)^{-\frac{1}{2}}$$

The reason that this is incorrect is that instead of taking the derivative of y with respect to "x", they have taken the derivative with respect to $g(x)$. This is seen more clearly if we introduce a new variable u, and set it equal to $g(x)$, or $2x+2$. Nowe can write y as:

$$y = \sqrt{u}$$

Now we see that if we were to take the derivative of y with respect to u, we would get

$$\frac{dy}{du} = \frac{1}{2}u^{-\frac{1}{2}}$$

Since we are looking for $\frac{dy}{dx}$, and not $\frac{dy}{du}$, we have not arrived at our answer yet. This is the error made in the expression above. We must multiply $\frac{dy}{du}$ by $\frac{du}{dx}$, since by the rules of fraction arithmetic, this yields the expression for $\frac{dy}{dx}$.

Since $\frac{du}{dx} = 2$ when $u = 2x + 3$, we must multiply our expression for $\frac{dy}{du}$ by 2 to arrive at the correct answer:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \left(\frac{1}{2}u^{-\frac{1}{2}}\right)(2)$$
$$\frac{dy}{dx} = u^{-\frac{1}{2}} = (2x+3)^{-\frac{1}{2}}$$

The bulk of this Lesson is spent practicing the technique above on increasingly complicated expressions, including instances where multiple applications of the chain rule are required, as well as the derivation of the "general power rule". Though the mechanic of applying the chain rule are important, it is equally important to ensure that students understand when to apply the chain rule, and the motivation behind finding derivatives of constituent parts of a function and combining them through the techniques discussed above.

Lesson 14: Implicit Differentiation

The technique of implicit differentiation offers another example where differential notation is helpful to illustrate the algebraic operations that underlie the calculation of derivatives. Implicit differentiation builds on the technique of the chain rule to enable us to take derivatives of complicated functions when we cannot solve for one variable in terms of the other. Up until this point, most students have primarily dealt with functions which can be expressed "explicitly", or when we can solve for our dependent variable y by itself. Even in instances where our primary form of the equation intermingled x's and y's , like in the definition of a circle, we were still able to solve for the dependent variable and apply the techniques of differentiation we learned above.

When a function is expressed "implicitly", this means that the function definition $f(x)$, or independent variable y, is sprinkled throughout our expression, and is not easily separated from the rest of the expression. In these cases, of which many real world applications fall into, we need a different way to solve for their derivatives. This technique is an extremely sophisticated application of both the chain rule AND the product rule, and thus will present difficulty for many students for many weeks.

Previously, when we would take a derivative we really only needed to consider operations to the independent variable x. We were differentiating a function of x with respect to x. But what happens when we are differentiating an expression with respect to x which both contains x's, but also y's? In these cases, the y represents $f(x)$, and is an unknown function of x, so we don't know what it's derivative will be exactly. But we DO know that its derivative with respect to x is $\frac{dy}{dx}$, which is ultimately what we will want to solve for.

Let's consider one of the terms from the example provided in the textbook.

$$2xy = 4$$

We remember from algebra that if we do something to both sides of an equation, then we usually still have equality. Taking the derivative of both sides of this expression:

$$\frac{d}{dx}(2xy) = \frac{d}{dx}(4)$$

In this simple expression, there are many important points a student will need to understand. First of all, the derivative of a CONSTANT is always equal to zero so the right hand side of our equation equals zero - this is a situation that arises commonly in implicit differentiation. But what about the left hand side?

Many student will want to apply the power rule to the expression $2xy$, and say that the answer is "$2y$", i.e. the coefficient of x. It is important to stress that y is also a function of x, so if y were equal to $\cos(x)$, then $2xy = 2x \cdot \cos(x)$. In this case, we would need to use the PRODUCT rule, since we are multiplying two functions of x.

SO for a general function of x, which we are calling y, applying the product rule yields

$$\begin{aligned}\frac{d}{dx}(2xy(x)) &= 2x\frac{d}{dx}(y(x)) + y(x)\frac{d}{dx}(2x) \\ &= 2xy' + 2y\end{aligned}$$

If we return to the original expression, we now find that taking the derivatives of both sides gives us the equation below:

$$\begin{aligned}2xy' + 2y = 0 \\ y' = -\frac{2y}{2x} = -\frac{y}{x}\end{aligned}$$

where we used basic algebra to solve for y', which is the derivative we are trying to find. This technique is extremely non-intuitive, but builds consistently on the reasoning of the product and chain rule, and offers a chance to illustrate both the calculation of derivatives in much more complicated scenarios, as well as to teach students to notice when and where the product and chain rules are necessary.

Lesson 15: Linearization and Newton's Method

Once we have a way to find the slope of the tangent line to a function, we may also want to know the equation of the line which describes that tangent line. For instance, if a car were traveling around a circular track, and all of a sudden lost friction with the ground, it would begin to travel along the tangent line to the circle. This line is determined by the technique of linearization discussed in this Lesson.

The difference between linear and nonlinear functions was discussed briefly in Chapter 1, but the difficulty of analyzing nonlinear functions was not stressed. In fact, many of the techniques for analyzing nonlinear functions involve considering an approximation to that function which IS linear. This is particularly important when trying to analyze fields which arise from solutions to differential equations that are highly nonlinear and often can only be expressed implicitly. In many of these cases, a multidimensional linear approximation is used, often referred to as the field's Jacobian matrix.

Our treatment of straight lines has focused on what is called slope-intercept form, or the equation $y = mx+b$. This was the equation we used to calculate the slopes of lines when we considered the taking the derivatives of polynomials and the Power Rule. Another way to express the equation of a line is point-slope form when we know the slope of our line, (in this case $f'(x_0)$ and a point on our line $(x_0, f(x_0))$. In this case, we can write the equation of the line as:

$$f(x) - f(x_0) = f'(x_0)(x - x_0)$$

where m is the slope of the line. Rearranging to solve for $f(x)$ gives:

$$f(x) = f'(x_0)(x) + (f(x_0)) - (f'(x_0))(x_0)$$

The usefulness of this equation is seen in the Lesson for the example of the square root function. If we take a point on the square root function x_0, and wanted to know the value of the square root function near the point x_0, often what we will do is use the linear approximation above, since then we can just plug in the value of x we are interested in for x in the above equation, and solve for $f(x)$. If we are very close to our original point, and our function does not change too abruptly, often the answer we get using the "linear approximation" to the function in the vicinity of x_0 will be accurate enough for our purposes. To get more accuracy, we would incorporate additional terms involving higher order derivatives, as will be discussed in Chapter 8.

The final topic discussed in Chapter 2 is known as Newton's method, which is a technique to find the roots of a polynomial when it does not easily factor. Newton's method is another approximation technique which used the concept of derivatives to estimate an unknown value when no analytic techniques are available. Unlike linear approximation, however, Newton's method is an example of a recursive procedure which can be repeated on the answer output at each iteration, causing the subsequent answers to converge towards the correct value. This technique is often implemented through computer programs trying to locate the roots of polynomials which describe the dynamics of complicated physical systems in biology and engineering.

1.3 Applications of Derivatives

Lesson 16: Related Rates

The technique of implicit differentiation learned in Chapter 2 has an immediate physical application which can help students to develop intuition for how differentiation works. This is introduced through the topic of Related Rates, which illustrate how to use implicit differentiation to compare derivatives of functions when their relationship is expressed implicitly. This topic is quite confusing since it involves understanding and calculating relationships between variables in a manner most students have not seen before. It also introduces the concept of differentiating every variable in an expression with respect to time, as opposed to using one of the variables as the independent variable like we did in Lesson 14.

In many physical situations, we can characterize the behavior of two or more variables using information contained in the problem and our knowledge of geometrical relationships. Take, for instance, the example of two people walking away from each other at a right angle developed in Example 1 of this Lesson. If one person is walking along the $x-$axis at 5 mph, and another person is walking along the $y-$axis at 3 mph, can we determine the rate at which the distance between the two walkers is changing?

To do so, we first represent the relationship for the distance between the two walkers using the Pythagorean theorem, where x denotes their distance in the $x-$direction, and y their distance in the $y-$direction:

$$x^2 + y^2 = z^2$$

We know that the distance between the two walkers at a given time is z, the length of the hypotenuse of the right triangle for which their motion forms the two legs. But this expression does not tell us the rate at which their distance is growing as they walk; to find this rate, we must use implicit differentiation.

Because we are given the rates at which x and y are changing with time, and we are trying to calculate the rate at which the distance between them, e.g. z, is changing with time, we must take the derivative of all of the terms with respect to time, t. This might seem confusing, since t does not appear anywhere explicitly in the equation above; it is important for students to realize that each of the variables is actually a function of time, and could have been written $x(t),\ y(t)$ and $z(t)$.

This problem offers an opportunity to illustrate the use of the chain rule. Let's consider taking the derivative of the term x^2 with respect to time. We know that the distance x is a function of time, $x(t)$, and so we must apply the chain rule to this term.

$$\begin{aligned}\frac{d}{dt}x(t)^2 &= \frac{d}{dx}x(t)^2 \times \frac{dx}{dt}\\ &= 2x(t)\frac{dx}{dt}\end{aligned}$$

Similarly, if we apply this technique to all of the terms in our function using implicit differentiation with respect to t, we get the expression:

$$2x(t)\frac{dx}{dt} + 2y(t)\frac{dy}{dt} = 2z(t)\frac{dz}{dt}$$

There are a number of important points to stress in problems like this. First of all, it is important for students to realize that we are solving for $\frac{dz}{dt}$- the change in distance between the walkers (z) with respect to time. Secondly, if we look at this expression, this value of $\frac{dz}{dt}$ will depend on the x and y value we are at, since x and y both still appear in our expression. Finally, it is important to realize that although a z still appears in the expression, we can use the original equation for the Pythagorean theorem to solve for z at a particular (x, y) value. Solving for $\frac{dz}{dt}$ and substituting for z gives:

$$\frac{dz}{dt} = \frac{x(t)\frac{dx}{dt} + y(t)\frac{dy}{dt}}{x^2 + y^2}$$

Each of these points are likely very new to the student, and will take some time and practice to become intuitive. The remainder of this Lesson provides more examples to practice these techniques in the context of additional physical applications.

Lesson 17: Extrema and the Mean Value Theorem

One of the most important applications of derivatives is their use in optimization, where understanding derivatives and their properties can help to determine where the maximum or minimum value of a function exists. This can enable us to calculate the optimal time to buy a stock, the optimal place to cut a roll of fabric, or the relative number of shirts and pants we should manufacture to optimize our profit.

The places where a function achieves its largest or smallest values are referred to as its extreme values, or extrema. As we will learn in the next Lesson, these extrema occur for differentiable functions only at places where the derivative is equal to zero. For non-differentiable functions, they will only occur at places where either the derivative is equal to zero or does not exist. Take, for instance, the absolute value function $|x|$. Its minimum value happens at the point $x = 0$, but at this point the function does not have a derivative, as discussed previously.

The remaining theorems developed in this Chapter dovetail closely with the rationale for the First Derivative Test discussed in the next Lesson. Rolle's Theorem and the First Derivative Test are both based on an intuitive and important result - if a differentiable function has to turn around, e.g. is moving upwards then moving downwards or vice versa, it must at some point in that transition have a slope of zero.

Rolle's Theorem, as discussed in this Lesson, states that if a differentible function on a closed interval $[a, b]$, is equal at the ends of the intervals, e.g. $f(a) = f(b)$, then there is at least one point in that interval where its slope must equal zero. This is because regardless of the direction the function is moving (up or down) from the left endpoint "a", it must at some point move in the opposite direction to get back to the value it was at "a", $f(a)$, when it gets to the point "b". In the extreme case, where the function is a constant value between a and b, then the graph is a horizontal line and its slope everywhere in this interval is equal to zero.

The second theorem introduced in this Lesson, the Mean Value Theorem, is a slightly more advanced application of the result above. Instead of a function's slope needing to go to zero in the interval, the Mean Value Theorem states that the slope of a function f will at some point be the average value between the slope at the left endpoint and the right endpoint. If we were to connect the two points at the end of the interval, as seen on *p.* 122, we will get a straight line with some slope m. The Mean Value Theorem states that somewhere in the interval between the endpoints, the function must have a slope of m.

Again, the salient point for students to understand is that if a function is smooth, its slope will go through every value between the value of the slope at the endpoints. The proof of this is detailed in the textbook. It can be helpful to show students how these theorems do not apply to functions that are discontinuous or not differentiable in this interval.

Lesson 18: The First Derivative Test

The theorems discussed in Lesson 17 help motivate a very useful result of differential calculus and analytical geometry - a function that is continuous and differentiable will be at a maximum or minimum value, **only** if its slope is equal to zero. However, as we will learn in the next Lesson, just because the slope of a function is equal to zero does not mean that it must be at an extreme value.

The motivation for the First Derivative Test is that for a smooth function to "turn around", meaning for it to transition from moving upwards to moving downwards, or conversely from moving downwards to moving upwards, the slope at some point in that transition must be equal to zero. When a function is moving upwards, its derivative is positive; when it is moving downwards, its derivative is negative. By looking at the first derivative of a function, and knowing where this derivative is positive or negative, we can identify regions where a function is increasing or decreasing.

For a function that is continuous to go from a positive to negative value, or negative to positive value, it must pass through the point zero. Otherwise it would have to somehow jump over that point (and thus not be continuous). When a function takes on an extreme value, the function must, in some sense, "turn around", since if it kept going in the same direction past the extreme value, there would be other values more extreme than the one we have picked.

This fact enables us to look at a very complicated function which we may not wish to graph, and determine (1) if it has any extreme values, and (2) where those extreme values are. An important point to stress, however, is that the types of extrema we are identifying right now are called relative extrema, meaning they are only extreme in a neighborhood. Usually the absolute extrema occur at one of the relative extrema, but we will consider cases where this isn't true in subsequent Lessons.

As noted in Lesson 17, being able to find the extreme values of a function forms the basis of optimization. If we can represent a value of interest like cost or quantity of material needed, in terms of a function, then by finding the extreme values of that function we can determine the cost or quantity of supplies needed to maximize or minimize our objective. Consider the polynomial function:

$$f(x) = x^3 + 2x^2 + 4x - 5$$

By taking the first derivative of this function, and setting it equal to zero, we can find places that are candidates for an extreme value.

$$f'(x) = 3x^2 - 4x + 4 = 0$$

In this particular case, we find that there are no values for which $f'(x) = 0$, since the discriminant of this quadratic function is negative. Consequently, we can say that this function does not have a maximum or minimum without needing to graph it. However, it is important to note that just because a function has places where its slope does equal zero, this does not guarantee that we are at a extreme value. This is discussed in the next Lesson.

Lesson 19: The Second Derivative Test

The First Derivative Test is an important example of a test generating "necessary but not sufficient conditions". As we saw in the case of $f(x) = x^3$, even at a place where the first derivative is equal to zero, we may not be at an extremal value. To be at an extreme value for a differentiable function, the first derivative must equal zero (i.e. is necessary) but just because it equals zero does not guarantee we are at an extremum (i.e. it is insufficient). To determine if we are at an actual extreme value, we must also look at the second derivative of the function, or the derivative of the first derivative, as discussed in this Lesson.

Just as the first derivative of a function measures how fast or slowly a function is changing (i.e. its slope, or rate of change), the second derivative of a function measures how fast or slowly this slope gets bigger or smaller. Visually, this represents the concavity of the graph, and is a measure of the graph's curvature. If a graph is curving upwards, we say it is concave upwards and has a positive second derivative. This corresponds to its first derivative growing larger, i.e. the graph getting steeper and steeper, causing it to curve upwards. If a graph is curving downwards, it is called concave down and has a negative second derivative, as the slope is becoming more and more negative, and thus decreasing. A decreasing first derivative implies a negative second derivative.

What about the situation where the second derivative equals zero? A straight line has a constant slope everywhere; if were to plot its first derivative we would get a horizontal line at the value of its slope. Since the slope of a horizontal line is zero, this means that the derivative of the first derivative (aka the second derivative of the straight line) is equal to zero. A second derivative of zero means that the function has no curvature, which is the case for a straight line. It can also mean that a function is transitioning from being concave upwards to concave downwards, or vice versa, Consequently, its second derivative is going from positive to negative, or negative to positive and so it will also pass through a point where it will equal zero. This point is referred to as a function's inflection point(s).

To understand how the second derivative of a function can help to determine extreme values, students should be reminded what the second derivative is telling us, and how this relates to the behavior of a graph near its extreme value. If a function is at a maximum value, then the values of the function around it are all less than the value of the function at its maximum. As the graph approaches from the right, the first derivative is positive, but getting smaller in magnitude until it approaches zero. To the right of the extreme point, the slope begins to get more and more negative.

If the first derivative started positive, became zero, and then became negative, then the first derivative is decreasing, meaning its slope (or its derivative) is negative. This means the second derivative of the original function is negative. Thus for a function to be at a maximum at a place where the first derivative is equal to zero, the second derivative must also be negative.

Now let's consider what happens to the first derivative near a minimum value. In this case, the slope starts negative, goes to zero, and then "turns around" and becomes positive.

near a maximum

$$f'(x) : + \to 0 \to -$$

e.g. $f'(x)$ is decreasing (going from positive to negative)

near a minimum

$$f'(x) : - \to 0 \to +$$

e.g. $f'(x)$ is increasing (going from negative to positive)

Since the first derivative is increasing, this means that its derivative (or the function's second derivative) must be positive. So for a function to be at a minimum, not only must its first derivative be zero, but also the second derivative must be positive. Graphically this is seen as the function being concave up.

Lesson 20: Limits at Infinity

This Lesson presents an application of the derivative as a tool to find limits of indeterminate functions such as we saw in Lessons 5 and 6. This technique is referred to as L'Hopital's rule, and is one of the most powerful tools developed by the calculus to comprehend the behavior of the infinite and infinitesimal. It also provides us with a greater understanding of what it means for an algebraic expression to be an indeterminate form, and why in some cases it has a limit, and in some cases it doesn't.

L'Hopitals rule tells us that for a rational expression of two functions $f(x)$ and $g(x)$, if their limit as x approaches a critical value x^* makes the expression indeterminate, then their limit is equal to the limit of the rational expression of the first derivatives of the two functions, or the rational expression of higher order derivatives. In other words:

$$\lim_{x \to x^*} \frac{f(x)}{g(x)} = \lim_{x \to x^*} \frac{f'(x)}{g'(x)} = \lim_{x \to x^*} \frac{f''(x)}{g''(x)}$$

In many cases, the original expression is indeterminate, but once we take the derivative of the numerator and denominator, sometimes the limit evaluates to an actual number. There are numerous examples of this in the text, and we have considered an important example previously of the "sink" function:

$$f(x) = \frac{\sin(x)}{x}$$

At the point $x = 0$, this expression becomes indeterminate. Using L'Hopitals rule, we find that the value at $x = 0$ actually equals one, since these two functions approach zero at about the same rate.

The rationale for L'Hopitals Rule is extremely helpful to show students how the rate at which a function approaches a value determines how "infinite" or "zero" it is. L'Hopitals Rule is an example of how different kinds of zero or infinity can behave differently, and how this can help us to evaluate limits of indeterminate forms.

Take, for instance, the rational expression

$$L = \lim_{x \to 0} \frac{x^2 + 3x}{x}$$

If we apply L'Hopitals rule once by taking the derivative of the numerator and denominator, we now have the following expression:

$$L = \lim_{x \to 0} \frac{2x + 3}{1} = \frac{2(0) + 3}{1} = 3$$

which we see is no longer indeterminate since we do not have a zero in the denominator. What this tells us is that even though both functions go to zero as x approaches 0, the numerator actually goes to zero more slowly than the denominator, because it is a higher degree polynomial. Consequently, the zero-ness of the numerator is more powerful than the zero-ness of the denominator, making the problem of dividing by zero go away. This doesn't mean, however, that this limit goes to zero; rather, it means that the limit becomes defined to some finite value.

L'Hopitals rule is, for many students, their first introduction to different kinds of infinity and zero, which proves to be one of the most important results of the mathematics they are learning. In addition, it illustrates how different kinds of infinity can be understood by understanding the rate of a function to tend to zero or infinity through analyzing its derivative.

A common error for students when using L'Hopitals rule is to apply it to situations where the original limit is not indeterminate. This error will lead them to calculate derivatives incorrectly, since L'Hopitals rule only is applicable to cases where we have an indeterminate form. The following example illustrates this error:

$$L = \lim_{x \to 0} \frac{2x + 3}{x} \neq \frac{2}{1} = 2$$
$$L = \lim_{x \to 0} \frac{2x + 3}{x} \neq \frac{3}{0} = \infty$$

Lesson 21: Analyzing a Function's Graph

Lesson 21 combines the results from previous Lessons to illustrate how to graph complicated functions by identifying important features and values using the techniques discussed above. In particular, by identifying a function's domain, zeros, regions where it is increasing and decreasing, behavior at infinity, relative extrema and inflection points, much of its graph can be drawn on the $x - y$ plane without needing to plug in a library of values. This Lesson provides a number of sophisticated functions on which students can practice the techniques they have been learning thus far in the textbook.

This Lesson presents a table that students can use as a template to calculate many of the features and values mentioned, and then use this table to construct the function's graph. As mentioned in Lesson 2, whereas the identification of a function's domain is usually quite straightforward, determining its range can be much more difficult, since it is not always clear from simply looking at the expression. Many of the functions in this Lesson can be used to apply the technique of finding a function's range by finding the range of its constituent terms and combining them. Consider Example 5 in this Lesson:

$$f(x) = -\sqrt{2x+6} + 3$$

It should be clear that the Domain of this function includes all values of x which do not make the expression $2x - 6$ negative. We write this as :

$$\text{Domain} = \{x|2x + 6 > 0\}$$
$$\text{Domain} = \{x|x > -3\}$$

However, how do we identify its range? The first step is to look at the term containing the square root. We know that the square root takes on a positive quantity when not otherwise specified, so this term must always be negative (due to the minus sign that appears in front of it). Knowing that this term must always be negative, when we add +3 to it, this causes the entire function to be shifted upwards by 3. So instead of our function always being negative, i.e. less than zero, now it is always less then or equal to 3. Consequently, we can identify its range as:

$$\text{Range} = \{y|y \leq 3\}$$

This technique is very useful to identifying the Range, but often cannot be applied as straightforwardly as in this example. This was seen in Lesson 2, where we were only able to make an estimate of the range using this technique, since the minimum and maximum values of the terms of our function did not all occur at the same place. In the example above, since we are applying a shift of +3 everywhere, this issue is alleviated.

Lesson 22: Optimization

Optimization is one of the most immediate and useful applications of differential calculus, and is based on the First and Second Derivative tests discussed in previous Lessons. In those lessons, we identified what we called relative maxima and minima, or places where a function is at its largest or smallest values in a neighborhood of that point. So, for instance, the following function has a relative maxima at ?? and minima at ??.

In Lesson 22 this result is extended to the situation where we want to find the absolute extreme values of the function, or the places where it takes on its largest and smallest values everywhere on its domain. To do so, we compare the values where the function achieves its relative maxima and minima, as well as the values that it assumes at its endpoints. In so doing, the places where our function is at its largest and smallest values, respectively, will be its absolute maxima and minima.

It is important to make sure that students understand both the reasoning for why the absolute extrema must occur at either relative extrema or at the endpoints of the function. Why are there not other points that we need to check? In Lesson 17 on the First Derivative Test, the reasoning for why the derivative must equal zero at a relative extrema was presented, which applies equally to the case of the absolute extrema If a function is not at a relative maximum, then this implies that there are values nearby where the function has a greater value. This implies that this cannot be the largest value for the function. Similar reasoning applies to minima.

The second scenario, when the absolute extrema occur at the endpoints, raises the important point that not all functions have absolute extrema (even if they have relative extrema). We remember that a function must "turn around" in order for it to be at a relative maxima or minima, which is true everywhere except at the endpoints of the interval (which in some cases are positive and negative infinity). At these points, because the function is not defined to the left or right of the endpoints, respectively, the derivative technically does not exist at these endpoints, thereby making them special cases. Similar to the case of points of discontinuity and indifferentiability being candidates for absolute or relaive extrema for non-smooth functions.

A technique students will need to grow comfortable as they apply the tools of optimization is manipulating difficult algebraic equations (and later entire differential equations) to eliminate variables and isolate the variable of interest. As we see in many physical cases, our expressions often have more than one variable, thus making the techniques we have thus far developed inapplicable. But, just like we could eliminate variables in the case of solving algebraic equations with two unknowns, we can also do so in more complicated physical scenarios. This technique is stressed in this Lesson, where the exercise of creating one equation from the statement of the problem, and another equation for an understanding of the geometrical or physical situation, allows us to isolate a single variable to optimize.

Lesson 23: Approximation Errors

Now that students have more familiarity with taking higher order derivatives, we can extend the discussion of linear approximations to the more general case of polynomial approximations to functions, often referred to as a function's Taylor Series expansion. Although this topic will be considered in much greater depth in Chapter 8, we use this Lesson as an opportunity to show students how using the second derivative can help us to approximate functions more accurately than using the first derivative alone. It is also an opportunity to show students how to use their graphing calculators to compare approximations to functions with the function themselves, and determine the region in which the linear approximation is within a given tolerance of the actual function.

We learned in Lesson 15 that in a small region around a function, often times a linear approximation to that function provides an answer that is close to the true value of the function. The intution for this face became apparent graphically, where is the function behaves smoothly, it only gradually starts to move away from the tangent line. Soon enough, however, the true function and our line grow far apart.

As will we will address more thoroughly in Chapter 8, many functions that are not polynomials can be written as polynomials that have an infinite number of terms. As we add more and more terms, or higher order terms like $x^\wedge 2, x^\wedge 3$, etc., our approximation usually starts to look much more like the function we are approximating. This is because these higher order terms can capture the non-linearity of the original function which linear approximations do not. In this Lesson, we consider the case of adding a quadratic term to our approximation, and show how it relates to the second derivative of our function in the vicinity of the point around which we are approximating our value.

The technique of using higher order derivatives in an expansions is actually a case of using a higher order derivative to approximate a lower order derivative. Just as we could use linear approximation on our actual function by knowing its first derivative, we can estimate the first derivative of a function at the two ends of an interval by using linear approximation on it, which entails calculating the original function's second derivative.

As we will learn in Chapter 8, we can continue to apply this technique to each term we arrive at, and in so doing will get an approximation that is a little better than it was before (usually). This forms the topic of infinite series, and introduces an important rule for understanding approximation errors: our error is at most the value of the term of the infinite polynomial we are at. So, if we were to use a second-order polynomial approximation to a function, then the value for the error is at most the value of the third term in our polynomial, or the term containing the second derivative. This term is necessarily larger than that for the third derivative, which, if available, will thus give us an even more strict bound on our approximation error.

1.4 Integration

Lesson 24: Indefinite Integrals

This Lesson introduces the definition of a function's anti-derivative, which is a typical way of introducing the subject of integration. The introduction of the "constant of integration", as well as understanding how to take a function's anti-derivative when it does not follow a simple rule of differentiation, are both sources of confusion for students of calculus. Because the differentiation of the product or composition of functions is not straightforward, as discussed in Chapter 3, taking the anti-derivative of products and compositions is also not trivial, and doesn't follow the rules of traditional operations like multiplication and division.

To find the "anti-derivative" of a function $f(x)$, we are asking for the function $F(x)$ which, when differentiated, will yield $f(x)$. If we were taking the derivative of $f(x)$, we would get $f'(x)$. If we were to take the derivative of $F(x)$, we would get $f(x)$. So, to find the antiderivative of the function x^2, we are looking for a function which when differentiated will yield x^2. Written symbolically, we are asking for the function $f(x)$ which satisfies the following equation:

$$\frac{df}{dx} = x^2$$

We are, in some sense, taking the opposite of the derivative. In the case of polynomials, this technique is somewhat intuitive since it builds upon the Power Rule for taking the derivative of a polynomial. If the derivative of our function is x^2, we know that our function is going to be a polynomial of degree 3, or have an x^3 in it. This follows from our discussion of the Power Rule above. Unfortunately, if we differentiate x^3, we get $3x^2$, not x^2. So, we must also include a constant in front of our new monomial to cancel out the power which is brought down as a coefficient. In this case:

$$f(x) = \frac{1}{3}x^3 + C$$

since the derivative of this expression yields x^2.

This raises an important point in the discussion of integration - the appearance of a constant of integration. As students learned in Chapter 2, the derivative of a constant is always equal to zero. Consequently, if we add a constant to our function above, which is written as C, and differentiate it, we would still get x^2. This is true for any constant we add - how do we decide which constant to pick? The reason why integrals in the form of antiderivatives are referred to as "indefinite" integrals is that we do not have a clear way to pick this constant, and thus must leave it generic.

The intuition for why we need to include a constant of integration is easiest for students to understand graphically, as illustrated on the top of Page 180. As is seen from this figure, all three of these curves have exactly the same derivative, but are clearly different functions. This illustrates that two or more curves can have the same derivative if they are shifted vertically by a constant.

By just knowing the derivative of the function, we are not able to determine which of these graphs, or functions, is actually the antiderivative of the function we are looking for. This explains why the antiderivative we are calculating is "indefinite", and that we need more information to be able to decide which of these curves to select. As we see from the curves in this Figure, if we were given just one point that we knew our curve to go through, we could select between these curves without ambiguity.

This Lesson introduces the antiderivatives for a number of functions, including the exponential and square root fuctions, as well as more complicated polynomial expressions. Although the motivation for these results is not developed at this time, it will be helpful for students to know how to apply the antiderivatives they are learning in this Lesson in order to understand and complete examples in subsequent Lessons. Understanding both the motivation and behavior of these functions' antiderivatives forms an important part of the remainder of the text.

Lesson 25: The Initial Value Problem

The concept of an anti-derivative offers an opportunity to continue the introduction to differential equations presented in Chapter 2. If we are looking for the function $F(x)$ which, when taken the derivative of, yields the function in question, $f(x)$, we are asking for the solution to the following equation:

$$\frac{dF(x)}{dx} = f(x)$$

In many cases, finding the anti-derivative of $f(x)$ will be much more complicated than simply taking the anti-derivative of each term. For now, we limit our discussion of differential equations to those which can be solved directly by integration. In these cases, the variable we are differentiating with respect to is usually time.

When we solve such equations in physical situations, and are trying to determine what value to pick for our constant of integration, we are often given the state of the system at time = 0. This is known as the system's initial value, and thus these are known as initial value problems. In order to predict the system's value at some later time, t, we must have knowledge of its value at some initial time, otherwise our answer will always be off by this constant. To solve for the constant of integration, we plug in the value of our expression at the value $t = 0$, leaving us with an equation with just unknown (the constant of integration). For instance, if the result of our integral is:

$$f(t) = \frac{1}{2}t^2 + C$$

then substituting $t = 0$, and being given $f(0)$, e.g. the system's initial value, we are able to solve for C as:

$$\begin{aligned} f(0) &= \frac{1}{2}(0)^2 + C = C \\ f(0) &= C \end{aligned}$$

Although in this example the constant of integration turns out to be the system's initial value (which happens often in many of the simpler integrals students will encounter), it should be stressed that this is not always the case, and algebraic manipulation will be necessary to solve for the constant of integration.

The introduction of variables into the limits of integration might also cause some students difficulty, and can be expressed in terms of the initial value problem:

$$y(t) = \int_{t_0}^{t} y'(\tau)d\tau$$

Here, we are saying that given an unknown time t, somewhere in the future (or at some time after t0, the initial time), we would like to find the value of some quantity at that time. The output of evaluating this definite integral is a function due to the variable in the limits, whereas in other definite integrals, we get out a numerical value. By putting variables in as our limit of integration, and then differentiating with respect to a different variable, we can arrive at a function. It is very confusing for students to need to go through this extra step, but it proves important (1) for integration to be consistent, and (2) to represent situations in real problems where an initial value is given.

If we wish to produce a function from a definite integral, we must place the dependent variable in the limits of integration, and introduce a new "dummy" variable into our integrand. Many students will want to use the same variable in the integrand as in the limits of integration, but this creates ambiguity which will lead them into trouble in the future. For instance, students might want to write the expression above as:

$$y(t) = \int_{t_0}^{t} y'(t)dt$$

which is technically incorrect, since the variable $'t'$ is being used in two different ways.

Lesson 26: The Area Problem

One of the most important visual representations of an integral, and the way it is often defined for one-dimensional functions, is that the integral of a function is the area between it and the $x-$axis over a given interval. Sometimes that interval is infinite, but the area under the curve can still be finite. How this can happen is due to the convergence of limits at infinity that we discussed earlier. Those functions which converge to zero faster than the rate at which they move towards infinity have a hope of having a finite area, whereas those that don't have no hope. Understanding indeterminate forms and how to calculate infinite limits thus plays an important role in integration.

In this Lesson, the textbook provides a formal introduction to "sigma notation", which is an essential technique for understanding operations on expressions with a very large numer of terms. The use of sigma notation was already discussed in Lesson ?? in the context of applying the formula for the derivative to the general expression for a polynomial. Using the sigma notation allowed us to find a general rule for the derivative of any polynomial, also expressed in this notation. The $\sum$ for sigma stands for sum, and as this sum becomes a sum of an infinite number of infinitesimal terms, the sigma turns into the symbol for integration, which looks like an S. In order to prove many useful results for integrals, taking limits on summation expressions in sigma notation will be an important skill, and thus a number of important properties of summations should be understood. Particularly the identities introfuced on the top of page 190.

In the lesson on Evaluating Limits (p 53) students were introduced to the technique of finding a limit using the Squeeze Theorem. That theorem states that if a function is bounded by two other functions, and those other functions both converge to the same value, then the function of interest also converges to that value in the limit of those two functions becoming the same. This fact is used to understand how a definite integral calculates the area between a function and the $x-$axis.

In this Lesson we present the notion of Upper and Lower sums in the context of RIemann integration, and remind students that there are two ways to consider the sums of rectangular approximations to a function's enclosed area. We could use the function value at the left endpoint OR the function value at the right endpoint. As we will learn in the Lesson about Numerical Integration, there are many ways that we can approximate the curve between two points, including drawing a trapezoid or parabola between them. Many of those techniques will often lead to better appriximations quickly, but create much more complicated expressions to both calculate values at each iteration and error bounds.

What is interesting about the comparison between the upper and lower sums created by rectangular approximation is that in the limit as we make the width of the rectangle narrower, those two sums will approach each other. And that if our function is well behaved between those values, the value of the ACTUAL area will always be in between them. The reason for this goes back to the Mean Value Theorem, and one reason why it is stressed earlier in the textbook.

What the area approximation shows us is how integration really is a summation of terms which are in some sense adding up the function as we move along. Unlike the derivative, which focused on the ratio of differentials, the integral focuses on products of differentials, as the width of our rectangle grows smaller and smaller, and the value of the function used in the lower and upper approximations become the same.

Lesson 27: Definite Integrals

This Lesson builds on the results of Lesson 26 to develop techniques to calculate the limits of upper and lower sums to determine the area under a curve on a finite interval $[a, b]$. This process is referred as a definite integral, since by specifying the endpoints of our integral, known as the limits of integration, we are now able to calculate an exact value for the integral without knowing the constant of integration. The definition of the definite integral introduces new notation, where now we specify the limits of the interval on the bottom and top of the integral sign.

The second key point which can be emphasized in this lesson is the differential term (in most cases, dx) which appears in the integral. It is important for students to realize that for any integral, there MUST be a differential element within the integrand, and that this term reflects the variable with which we are integrating with respect to. This fact can be emphasized in this Lesson in the context of the Riemann sum, since the differential term, dx, corresponds to the finite difference $\triangle x$ in the limit as this difference approaches zero. The dx thus corresponds to the width of our rectangles as we make them narrower and narrower. Many students will omit this term, which will cause endless confusion as they progress to more advanced integration techniques such as $u-$substitution and integration by parts. The integral is the adding up of a bunch of products.

When we integrate a function, we hopefully get out another function. When we do, sometimes we want to know the value of this new function at two ends of an interval. This is most easily seen in the case of taking the area under a function. In many cases, we want to integrate a function OVER a certain interval, often a region in space or interval in time. For instance, if we were to integrate the velocity of a function from the time a runner started until the time the runner stopped, we would be calculating the distance the runner traveled. In this case, we could write the expression as follows:

$$d(t) = \int_{t_0}^{t} v(\tau) d\tau$$

Unlike indefinite integrals, definite integrals always have an exact number to which they evaluate. This is because we are evaluating the function that results from the integral at two endpoints and subtracting those values from each other. Consequently, the constants of integration at each endpoint cancel out.

$$\int_{x_1}^{x_2} x^2 dx = \left[\frac{1}{3}x^3 + C\right]_{x_1}^{x_2} = \frac{1}{3}x_2^3 + C - \left(\frac{1}{3}x_1^3 + C\right) = \frac{1}{3}x_2^3 - \frac{1}{3}x_1^3$$

This is a nice fact, and shows how a definite integral operates on a function without needing to know its offset in the $y-$direction. This will become more clear when discussing the Fundamental Theorem of Calculus.

Example 2 in this section is an excellent opportunity to highlight the use of the important summation identitites discussed in Lesson 26, since these are often necessary when using the original definition of an integral to evaluate an integral. As seen in this Example, we apply the limiting condition to the Riemann sum by breaking up the interval into n pieces, and then letting the width of those pieces get smaller and smaller by making n larger and larger. In the context of applying the definition of the definite integral, we make use of the results from Lesson 26 which identify the actual sums for some important infinite series.

Lesson 28: Evaluating Definite Integrals

Lesson 28 considers many of the important properties of definite integrals, particularly as they entail breaking up domains into subdomains. With the knowledge that the integral of a function between two points must only be evaluated at the endpoints of the interval, a number of useful results follow. An important rule for students to remember, and which may not seem intuitive, is that

$$\int_a^b f(x)dx = -\int_b^a f(x)dx$$

This says that if we integrate from left to right, we will get the opposite answer then if we were to integrate from right to left. This is somewhat surprising, since if the integral is the area under the curve, does this area change sign when we move from right to left? Apparently it does in some sense, insofar as the value of a definite integral is defined as the difference in a function's antiderivative at the two limits of integration. By switching a function's limits of integration, we essentially negate our answer. This is an important and non-obvious result about integrals, and becomes increasingly important when dealing with multivariable integrals in later course, where important results rely on defining the orientation of the integral in a particular way.

Another interesting and useful fact about definite integrals is that they can be evaluated over subintervals of their total interval, and then have the answers to their sub integrals added together. This follows quite clearly from the interpretation of an integral as an area under a curve, since if we were to break this area up into two sub-areas, we would expect the total area to be the sum of these areas on each interval.

This interpretation also lends itself to the integral sign being distributive over the operations of addition and subtraction (but not to products and quotients!). Taking the integral of the difference of two functions corresponds to finding the area in between the two curves. What if instead of taking their difference, we calculated the area between the upper curve and the $x-$axis, and then subtracted the area between the lower curve and the $x-$axis? From the visual representation of area, we clearly get the same answer as indicated in this Lesson.

The proof of Theorem 4.2 is also an excellent opportunity to illustrate the utility of the Mean Value Theorem for differentiation, which was intoduced in Chapter 3. In this case, we have an analagous Mean Value

Theorem for integration, which says that if were to calculate the area under the curve on an interval $[a, b]$ for a continuously changing function, there is some point on that function in that interval which we could use to multiply the width of the interval to get the actual area. It should be emphasized that this technique works primarily for differentiable and continuous functions, and that the average value is usually not at the middle of the interval.

Lesson 29: The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus formally unites the two most powerful new concepts that calculus has introduced: the derivative and the integral. In essence, what the fundamental theorem of calculus tells us is that to calculate the area under a curve over an interval, ALL we need to know is the value of that function's integral at the two endpoints of the interval. This is a surprising and useful fact, though there are some conditions that the function must satisfy to make sure that this theorem holds true. This fact is interesting for both its applications, as well as its underlying reasoning.

The Fundamental Theorem of Calculus asserts that the derivative and integral are really inverse operations to each other. The derivative of an integral of a function returns the function, and the integral of the derivative of a function returns the function. (confused?) This is an important tool to solving differential equations when we try to "undo" the derivative by applying integration, as we will see in subsequent Lessons. The proof of this is provided at the end of this Lesson, and provides an excellent situation to use many important results discussed previously.

In particular, this proof illustrates the use of upper and lower sums to squeeze a function towards a limit, as discussed previously in Chapter 1. It also illustrates how to generate a derivative in an expression, as is the case in the Fundamental Theorem of Calculus, we appeal to the original definition of the derivative:

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

The point should be emphasized that though we don't often use this original definition in evaluating derivatives now that we have a slew of other techniques, it is still important to remember its definition to complete proofs like this one.

An interesting point arises in this Lesson regarding the relationship between an integral and the area under the curve. When integrating, the area under a curve is positive if the curve is above the $x-$axis, and negative when it is below the $x-$axis. Consequently, any function that is "odd", or symmetric about the origin, would always have its integral evaluate to zero! In this sense, the integral is NOT the area under the curve, since areas cannot be negative.

This Lesson illustrates how to handle such a case by recognizing that because the functions we are integrating between are odd, the area to the right of the $y-$axis will equal the are to the left. Consequently we can define the area as twice the area of the region to the left of the $y-$axis, which solves our problem of having our areas cancel each other out. Consider what happens if we were to integrate the function in Example 1 between -1 and 1.

$$\int_{-1}^{1} (x - x^3)dx = \left[\frac{x^2}{2} - \frac{x^4}{4}\right]_{-1}^{+1} =$$

$$\left[\frac{1}{2} - \frac{1}{4}\right] - \left[\frac{1}{2} - \frac{1}{4}\right] = 0$$

Because our original function has odd powers of x, its integral will have even powers of x which will be equal when evaluated at a positive and negative value. This yields an answer of zero due to the reasoning above, contrary to the interpretation of an integral as the area under or between two curves. This is a good time to discuss how to determine symmetries so that techniques like breaking the integral up into subintervals can be applied.

Lesson 30: Integration by Substitution and Integration by Parts

One of the difficulties in evaluating the integrals of more complicated functions is due to the product and chain rules for differentiation that werediscussed in Chapter 2 and 3. Just as the derivative of the product of two functions was not the product of the derivatives, so too the integral of the products of two functions is not the product of the integrals. By being able to make clever algebraic substitutions and observations, however, we can often use our understanding of the chain rule and product rule to calculate complicated integrals. The integration analogue of the chain rule is known as $u-$substitution; the analogue of the product rule is known as integration by parts.

Just as differential notation was useful in helping students understand the chain rule and implicit differentiation, so too is it useful in understanding $u-$substitutions. When we make a $u-$substitution, we are defining a new variable in terms of the old; in so doing, we must be very careful to also re-define the differential element in the integrand, as well as new limits of integration. Once we are done with the integration of our new, easier integrand, we must then reverse our substitution in order to recover a function of the original variable of interest.

$U-$substitution is most effective when the integrand of our expression contains a function multiplied by its derivative, since this enables us to recreate the integrand using new differential notation. In particular, it is important for students to understand that the dx which often appears in integrands is the same as the dx which appears in the denominator of derivative expressions when taking the derivative with respect to x. For instance, if given the function

$$u = 4x^2$$

when we calculate $\frac{du}{dx}$, we get:

$$\frac{du}{dx} = 8x$$

Now, given our choice of u, what does the term $'dx'$ mean, in terms of u. As pointed out earlier, the use of differential notation represents these terms as the algebraic numerator and denominator of a fraction, and so we can manipulate this expression just like a fraction. Consequently, we find throudh algebraic manipulation:

$$dx = \frac{du}{8x}$$

Recognizing the times where u-substitution is most valuable can be made more clear through the following identity:

$$du = \frac{du}{dx} \cdot dx$$

Since our goal is to have an easily integrable function of u in our integrand, with the appropriate differential element du, if our original integrand was:

$$\int u \frac{du}{dx} dx$$

Then $u-$substitution yields the very straightforward integral:

$$\int u du$$

Being able to quickly recognize when an integrand takes on a special form to which u-substitution or integration by parts can be applied is an important skill to master in order to make applying these techniques efficient and useful. Often multiple attempts at finding the right substitution are needed, since sometimes it is not clear what substitution to apply. This becomes more apparent in Chapter 7 when we consider trigonometric substitutions for very complicated integrals.

Integration by parts also is based on recognizing a special form for the integrand, namely

$$\int f(x)dx = \int u dv$$

This is a strange and sophisticated expression, and takes a great deal of time to get used to. What this form represents, though not obviously, is the multiplication of one function by another function's differential element. If the original function can be written as such a product, using the same requirements on replacing the dx by dv which we saw above with $u-$substitution, then by undoing the product rule, we can find this integral somewhat straightforwardly (in many cases). This is a strange operation whose meaning is actually very subtle, since it tells us that we can integrate those products of functions in this manner which are a function by the differential of another function.

This will be discussed in greater detail in subsequent Lessons, but it will be helpful for students to be shown the basic reasoning behind this which is presented in this Lesson. Again, an important point to stress is that solving complicated integrals is often predicated by identifying that the integrand fits into a certain template, and can thus be integrated through techniques established for those templates.

Lesson 31: Numerical Integration

We find that as the functions we are trying to integrate become more complicated, we cannot evaluate them to an exact analytical function. In these cases we can use approximation techniques like the rectangular approximations called Riemann sums which we used to motivate the integral. But instead of just using evenly spaced rectangles, there are a number of more powerful geometric approximations which we can use that provide better results. There are a few fundamental techniques that a student should be familiar with, as numerical integration is an extremely powerful tool to solving differential equations when integrals to more complicated functions cannot be solved for.

In this Lesson we consider two of the most important and useful techniques of numerical integration, which build easily upon the definition of an integral offered to students previously. The first is called the trapezoidal rule, and eliminates the problem of choosing between the left and right endpoints for the height of the rectangular approximation. In the case of the trapezoidal rule, we connect the two endpoints with a linear approximation, which will often approximate the behavior of the function much more accurately in that interval.

The discussion of error approximation is extremely important as students move forward to issues in scientific computing and numerical analysis, since although we are often not in a position to calculate our error exactly (since in that case, we could calculate our function exactly), we are often in a position to apply an upper bound to our error. This is extremely important in applications where a given tolerance is acceptable, but no more than that tolerance. As was discussed previously in the Lesson on Linear approximation, and which will be discussed more thoroughly in Chapter 8, the bound for our errors is usually dependent on the derivatives of the function we are integrating. Because the technique of linear approximation and the trapezoidal rule are first-order approximations, meaning that they only consider the first derivative of the function, the error bound is dependent on the next derivative which has not been incorporated which in this case is the second derivative. To help students understand this intuitively without delving deeply into topics like Taylor and power series, which will be explored in Chapter 8, we can consider the case of approximating a straight line with the trapezoidal rule. In this case, it is easy to see that our trapezoidal approximation will be a perfect approximation to the area. As we learned above, the second derivative of a straight line is zero, which implies that our error is bounded by zero in this case, which matches out intuition.

Example 2 shows how thihe error bound inroduced on p. 220 can also be used to determine how many subintervals we must break our integral up into in order to be guaranteed of an error smaller than a certain amount. This is the opposite of what we discussed above, ans students should become comfortable with both approaches to using these useful results for error bounds.

The second technique discussed is using Simpson's Rule, which uses parabolas to interpolate the function between 3 points. What we find is that as we start using higher order polynomials to approximate our function, we must use more estimate points on our function to calculate the polynomial. This is because although a straight line can be uniquely defined by two points it passes through, a parabola must have three points specified to be uniquely defined. Similar reasoning applies to higher and higher order polyomials.

Again, our error bound is dependent on the higher order derivatives of the functions we are approximating, but because we are making a higher order approximation when using Simpson's rule, and consequently incorporating estimates based on higher order derivatives, the bound in this case is dependent on the 4^{th} derivative of the original function. The reasoning about this will become clearer in Chapter 8 when we consider Taylor and Maclaurin series, but the important point to emphasize is that as we use higher order functions to approximate our function to estimate its integral, our error will usually become much smaller, and will be dependent on the order of the derivative which is first excluded.

Understanding this tradeoff between greater accuracy and the need to perform more complicated functions using more data is essential to designing efficient algorithms for computers to solve differential equations.

1.5 Applications of Integration

Lesson 32: Area Between Two Curves

Chapter 5 begins considering some more advanced applications of integrals which involve more intuition into both limits of integration as well as the definition of the integral. Lesson 33 considers a new situation where instead of integrating with respect to x, we integrate with respect to y. This is quite strange, because the graphs we consider are no longer functions with respect to "x", but ARE functions with respect to "y". In these instances, we will need to rewrite our expressions as functions of "y", and determine the appropriate limits of integration.

The situation of finding an area between two curves is quite common, from determining the area of a region which does not have straight edges, such as agriculture, clothing or vehicle design. By returning to the definition of the integral as the area between a curve and the $x-$axis (when differentiating with respect to x), it is easy to see that the area between two curves can be found by taking each curves respective integral, and subtracting the results as shown in this lesson. Some students might wonder if it matters which curve they put first, since the only way to know which curve is above the other is by graphing them in some cases. To this point, students should be led to the conclusion that if they pick the order "wrong", they will get a negative value for the area.

One of the most important techniques in applying integration to real world phenomena is understanding how to find the appropriate limits of integration. This was mentioned briefly in Lesson 1, where we were looking at the intersection points of a straight line and a parabola. In this case, we were able to set the two equations equal to each other, and solve for the values at which these curve crossed. This lesson illustrates why this technique is important in calculus, because to find the area between two curves which cross (like the line and parabola, and unlike the functions $f(x)$ and $g(x)$ discussed in this lesson), we will need to know the points at which they cross.

Students should be led to understand how this works not only in the case of integrating with respect to x, but also integrating with respect to y. For instance, if they find limits of integration with respect to x, they should be able to easily understand why and how to switch these limits to being with respect to the $y-$axis. The importance of this is stressed in this Lesson, where calculations are often made much easier by switching the orientation of the integral. It is important to stress the need to change the limits of integration to reflect our new variable of integration, since this is a point that many students will forget.

Lesson 33: Volumes

The transition to calculus in 3 dimensions involves a great deal of additional analytical geometry which is not covered in this textbook, including understanding new types of coordinate systems and vector-based differential elements. By learning how to perform integration on $3-$dimensional objects, we can start to calculate the areas and volumes of very unusual geometries, as well as compute the values of quantities which occupy these geometries, like heat, electric fields, and material properties.

For three dimensional objects with a uniform cross-section, it is easy to extend the formula for the area of a cross section to calculating the volume. The volume for a cylinder, for instance, can be found by multiplying the area of the circular cross section by the height of the cylinder. But what about cases where the cross section is not uniform?

These situations are analogous to the case of finding an area under a continuously changing curve instead of under a straight line. Instead of multiplying the cross sectional area by the height, as we would do for a cylinder, when the cross sectional area is changing continuously over the height of the solid, we must multiply the area at a given point by a differential element of height, which yields a differential volume. We then integrate this product over the whole height to get the true volume.

An important skill to help students understand in this Chapter is being able to move back and forth between integrating with respect to x and y depending on the geometry of the situation. In many cases, one or the other will prove most effective in solving the problem at hand, as is considered in this Lesson

One of the greatest difficulties students will have with applying this technique is determining the formula for the cross sectional area in terms of the appropriate variable. If we can write an expression to represent the area of the cross section at a given x (or y) value, then we can integrate this expression over the height of the pyramid to determine its volume. The difficulty lies in determining this formula, since it usually requires students to use geometrical intuition which is not supplied in the problem itself. Applying this procedure to the many various shapes provided in this Lesson will help students to practice this technique, which is extremely important to applying integration to physical applications.

Lesson 34: The Length of a Plane Curve

Beyond finding the area under a curve between two points, or the rate of change of a curve at a particular point, calculus also allows us to calculate the length of curves which traditional geometric techniques do not. This lesson introduces students to the formula for calculating this length, which forms the basis of many important formulas in physics and engineering. To calculate the length of a plane curve, we use the following formula:

$$\text{length} = \int \sqrt{1 + f'(x)} dx$$

Many students may wonder where this formula comes from, and how calculus is used to derive it. This is an excellent exercise in reviewing limits, the definition of an integral, as well as developing the intuition to apply calculus to physical situations. Being able to identify how to use differential elements to calculate quantities which are changing continuously is an essential skill.

In this case, to derive the formula of a plane curve we use the Pythagorean theorem on a small differential element of the curve, and apply the techniques of limits. If we consider the curve from the text, we can develop this reasoning. We see that the length of the segment, which we call ds, can be solved for as:

$$\begin{aligned} ds^2 &= dx^2 + dy^2 \\ \left(\frac{ds}{dx}\right)^2 &= 1 + \left(\frac{dy}{dx}\right)^2 \\ \frac{ds}{dx} &= \sqrt{\left(1 + \left(\frac{dy}{dx}\right)^2\right)} \\ ds &= dx \cdot \sqrt{\left(1 + \left(\frac{dy}{dx}\right)^2\right)} \end{aligned}$$

To find the length of the actual segment, we must integrate both sides,

$$s = \int ds = \int dx \sqrt{1 + \left(\frac{dy}{dx}\right)^2}$$

Giving students an understanding of how to derive formulas like this is an extremely useful skill for them to develop in applying calculus to unfamiliar physical situations. In addition, understanding how to manipulate expressions like the one above is made much more intuitive by using differential notation, and is a technique used widely in the derivation of the applications of integration.

Lesson 35: Area of a Surface of Revolution

Using integration to determine surface areas for objects created by revolving a plane curve around an axis is very similar to that of calculating volumes which we treated in Lesson 34. In this case, however, instead of multiplying a cross sectional area by a differential length element and integrating, we will multiply the length of the plane curve by a differential length element and integrate. This will give us the area of the surface.

This technique is applied to the case of a spherical shell in this Lesson, which is one of the most important 3−dimensional shapes a student will encounter in applications. In this case, we take the differential length of a region on the plane curve, which is given in Lesson 34, and then multiply that length by the circumference of the circle which that differential length creates when rotated about the appropriate axis.

If we were to consider the entire volume that this process yields, the volume would appear to be a wedge. Since we are only considering the outer surface, however, the area we are calculating is the outer surface of this wedge. By integrating the differential areas of these wedges over the entire outer surface of the object, we arrive at an expression for the entire surface area.

As mentioned in Lesson 34, the ability to look at geometric and physical situations and apply the technique of integration is a challenging yet important skill for a student to learn. Stressing that an integral is really adding up the product of a changing quantity with differential length element, and how to apply this understanding to physical situations is one of the most important skills in applying calculus.

Lesson 36: Applications from Physics, Engineering and Statistics

Integration is very useful in calculating physical quantities which are defined by a product, an in which one or both of the quantities in the product are changing continuously over time or space. A classic example of this, as discussed in this Lesson, is the calculation of work done by applying a force over a distance. The formula for work is defined as:

$$W = Fd$$

In most cases that a student has likely encountered, the Force being applied to an object is constant over the distance an object is being moved. In this case, we can simply calculate the above product by multiplying the Force and the distance. But what happens if the Force we apply is not the same over the entire distance? In this case, we need to consider the force at each point along the distance it is applied, and use integration.

At this point it is helpful to review the appearance of the differential element within the integral sign, and remind students that this actually represented an infinitesimal distance. In the case of calculating an area, we would multiply the height of our function by this infinitesimal distance to calculate the area of one our infinitesimal rectangles, and then add up all of those areas using integration. Similar reasoning holds in physical cases which are not areas. Because our force may be constantly changing over the course of our distance, we need to consider how our force is acting on an infinitesimal piece of the overall distance. To do so, we consider the term dW, or the differential of Work.

$$dW = F\ dx$$

What this says is that at a particular point along our curve of action, the amount of work we do along a differential element of distance (dW) is equal to the Force at that point times that differential distance (dx). To find the total work done over an interval of distance, we integrate dW over the interval on which the force is acting. This yields:

$$W = \int_{x_1}^{x_2} dW = \int_{x_1}^{x^2} F \ dx$$

Because the force is changing over our interval, we can express the force as a function, and integrating it over our distance will find the total work done. This reasoning can be applied to a number of physical situations as discussed in this Lesson, and illustrates that integration is not simply about finding the areas under curves. Rather, it is about adding up a bunch of products with an infinitesimal term, as in the case of the Work expression above.

1.6 Transcendental Functions

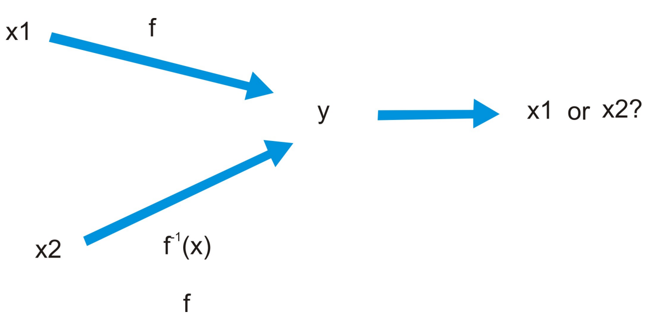
Lesson 37: Inverse Functions

This Lesson introduces students to the important concept of function invertibility. To be invertible, a function must not only pass the vertical line test, but also the horizontal line test, meaning that each and every input value $'x'$ has at most one output value $'y'$.

Many students might think this is the same thing as the definition of a function, and must be reminded that a function CAN have many inputs giving the same output; it just cannot have multiple outputs for the same input. Examples include the parabola and horizontal line, both of which are functions but which are not one-to-one, and thus not invertible.

Why is it that one-to-one-ness implies invertibility? This is a confusing and important topic to discuss, since many students may not have considered this before. Quite simply, a function is invertible if given the output of a function, we can uniquely figure out what input got us to that output. Because functions like the parabola have multiple input values that lead to the same output value, if we were just given the output value we would have no way of knowing which input value was used to get us there. This ambiguity means that our function is not invertible.

This can be illustrated with a diagram. In this diagram, we see that two $x-$values are mapped to the same $y-$value. If we were given the $y-$value, and asked to find the $x-$value that got us there, we would not be sure if it was $x1$ or $x2$, making our calculation ambiguous and leading us to say that the inverse does not exist.



The traditional definition of an inverse of a function, which is often written as f^{-1} (not to be confused with $\frac{1}{f}$), is that if we apply a function to an argument, and then apply the inverse function to that function's output, we should get out original argument back. Written symbolically:

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

The reason that we use the -1 notation is that multiplication and division are considered to be inverse functions. As noted in the text, this can cause a great deal of confusion as students might think that the inverse of a function is just the reciprocal of the function. For instance, a common mistake might be:

$$\text{if } f(x) = x^2, \quad f^{-1}(x) = \frac{1}{x^2}$$

This is most definitely not true, as we could see by applying $f(x)$ and $f-1(x)$ to the argument “4”:

$$f^{-1}(f(4)) = f^{-1}(4^2) = f^{-1}(16) = \frac{1}{16^2} = \frac{1}{256} \neq 4$$

This Lesson reminds students of the technique of finding a function’s inverse, as well as the geometric intuition of an inverse of a function being a reflection about the line $y = x$. Although the justification for these interpretations is not given, students should be familiar with these techniques from previous mathematics classes. It is helpful to remind them of the

Lesson 38: Exponential and Logarithmic Functions

Two of the most important functions appearing in physical applications are the exponential and logarithmic functions, which are inverses of each other. The reason for their importance is provided in the next two lessons. In this lesson, the definitions of exponential and logarithmic functions are discussed, as well as some important properties and identities for both classes of functions. Most students have learned logarithms in the context of base-10 number systems, but should become more comfortable with the usage of the natural logarithm, or a logarithm with a base of e. They should also be comfortable with the operation “log” referring to base$-e$ logarithms, as is the practice in most engineering applications.

In looking at the graphs of the exponential and logarithmic functions, it may be helpful for students to identify certain features. In particular, what happens to the logarithm function for negative values, and what is the logarithm of zero? Both of these are important cases as they represent places where the logarithm function is not defined for real numbers, and which may lead to instances where input values must be excluded from the domain. This is the third most important case of undefinedness, after division by zero and taking the square root of a negative number discussed in Lesson 2.

Take, for instance, the natural logarithm of zero. To what number can we raise the positive number “e” to get zero? There is no such number, and thus we see that the graph of the logarithm function is not defined at zero. Instead, the graph approaches negative infinity asymptotically, since if we take the limit of a positive number greater than 1 (like e) raised to a negative power that approaches infinity, this will give us zero, since raising a positive number to a large negative power makes the expression get smaller and smaller. However, it will never reach zero, therefore making the logarithm of zero undefined.

Similarly for negative numbers. There is no real number exponent to which we can raise a positive number that will give us a negative number. It can be helpful for students to be made clear on this point, and to convince themselves of this fact. In so doing, it will test their intuition and understanding of logarithmic functions, and enable them to remember the definition of a logarithm if they forget.

Finally, the properties of logarithm functions given in the box at the end of this lesson are extremely important tools to being able to apply the rules of calculus to them, and yield interesting and useful results. These properties are based heavily on the rules for multiplying and dividing exponential functions, where we learn that multiplication of terms leads to an addition of exponents, and the division of terms leads to a subtraction of exponents.

The most common mistakes is to think that the $ln(a)ln(b) = ln(a + b)$ or $\frac{ln(a)}{ln(b)} = ln(a - b)$. The invalidity of these identities can be made clear by picking actual numbers for a and b, and reviewing how the use of the exponential properties mentioned above lead to the identities for logarithms presented in this chapter.

Lesson 39: The Calculus of Exponential and Logarithmic Functions

The exponential and logarithmic functions play important roles in calculus, for reasons that most students have likely not encountered. If we return to the analogue of the power rule for integration, we see that although this rule works for most polynomials, it does not work for the case $\frac{1}{x}$. In this case, if we applied the reverse power rule, we would end up dividing by zero.

Though this might seem that $\frac{1}{x}$ therefore does not have an integral, this turns out to be a rash judgment. As explained in this lesson, the integral of $\frac{1}{x}$ is the natural logarithm of x, or $ln(x)$. This is likely to be an unexpected result for most students, and the source of some confusion. Though it is helpful to convince them of this fact using the sophisticated proof presented in this chapter, for the most part they will be able to just use this rule without needing to derive it ever again. Because many physical applications arise which involve taking the integral of functions like $\frac{1}{x}$, the natural logarithm appears in a wide range of formulas in chemistry, physics and biology.

Similarly, the exponential function ex assumes a special place in calculus for two reasons. One, it is an example of using limit theory to calculate a limit on new indeterminate form which we have not yet considered, in this case 1^{infinity}. Secondly, the exponential function is the only function which is its own derivative. This is a very surprising but useful result. Consider the following differential equation:

$$\frac{dy}{dx} = y$$

This is asking for the function $f(x)$ which when differentiated, will yield itself. We will consider this example more closely in Lesson 41, but it is interesting to note that the solution to such equations will always be an exponential function.

Of particular importance in applying the techniques of calculus to both the natural logarithm and the exponential function is the proper use of the chain rule when the situation calls for it. As noted in the text, the formal definition for the natural logarithm of a function $u(x)$ is given as:

In the case that $u(x) = x,\ u'(x) = 1$, which shows that the derivative of the natural logarithm is $\frac{1}{x}$. What about the case where $f(x) = \frac{1}{\sqrt{(2x+1)}}$? In this case, we can make the $u-$substitution that $u = 2x + 1$, and $\frac{du}{dx} = 2$. Given the expression:

$$\int \frac{1}{\sqrt{2x+1}} dx$$

We must solve for dx via the $u-$substitution we have selected. So, in that case:

$$du = 2dx$$
$$dx = \frac{du}{2}$$

When we make the appropriate $u-$substitution, we arrive at the following integral:

$$\int \frac{1}{2u} du = \frac{1}{2} \int \frac{1}{u} du$$

It is important for students to realize that in the process of making u-substitutions, we are really making algebraic substitutions, and so must follow the rules of equality when substituting the du for the dx term. The steps shown above where we isolate dx in order to make the substitution is very important, and something which many students will make errors with if they are not clear that the dx we see in the original expression must be completely accounted for. In many cases they will want to substitute the du directly for the dx, or solve for du and then make the substitution, which will get thim into trouble.

Lesson 40: Exponential Growth and Decay

One of the most commonly appearing mathematical phenomena is exponential growth and decay. It can describe the degradation of a sound or other wave over time and distance, the growth or decline of the price of a stock, and the multiplication of a bacteria in a Petri dish. The reason for this common occurence, and how to analyze these situations, is presented in this Lesson.

As mentioned above, the exponential function is the solution to a commonly occuring differential equation,

$$\frac{dy}{dx} = y$$

Because the exponential is the only function which is its own derivative, it will solve equations of the above form, and thus explains its appearance in formulas to calculate many important quantities.

The phenomena of exponential growth and decay refer to the cases where the exponential term has a positive or negative exponent, respectively. In both of these cases, when graphing the exponential function, at $x = 0$, it will intercept the $y-$axis at $y = 1$. It is at this point that $e^x = e^{-x}$, and is a helpful point of reference for students to generate the graphs of each.

In addition, the expression for e^{-1} is an important number for students to remember since it corresponds to the point where our function has decreased by one order of magnitude in a logarithmic sense. We often encounter exponential terms in physical formulas which look like:

$$f(t) = e^{-\frac{t}{\tau}}$$

In these cases, we refer to τ as the "time constant", and it corresponds to the amount of time that it takes for the function to decay to 0.37 of its original value, since if we were to substitute $t = \tau$, we would get e^{-1}.

The fact that the exponential and logarithmic functions are inverses of each other is developed in this Lesson through the use of examples where we are solving for quantities in the exponent of the exponential term. In order to solve for quantities in the exponent through traditional algebraic means, we must apply the natural logarithm to the exponential terms, which, because it is the exponential's inverse, will return whatever value is in the exponent. In so doing, we often must apply the properties of logarithms and exponentials to generate an expression to which this annihiliation can be implied, as is illustrated in the examples in this lesson.

Lesson 41: The Calculus of Inverse Trigonometric Functions

The appearance of inverse trigonometric functions in physical applications is largely due to the scenarios developed in this lesson. In particular, the interesting algebraic and rational expressions in their calculus leads to their appearance in formulas used to compute important physical quantities in unexpected situations. It can be helpful for many students to provide explicit graphical representations of each of the inverse trigonometric functions, since their graphs are usually unfamiliar and display properties that are uncommon.

In particular, if we were to allow many of these functions to range over their entire domain, we would find that many of them are NOT functions, since they do not pass the vertical line test. To make them into functions, we often need to restrict their range or only use what is referred to as their principal argument, ensuring that we have only one output for each input.

An important component of this Lesson is illustrating how inverse trigonometric functions often appear in solutions to integrals via the technique of $u-$substitution. What makes integration so challenging for may students is being able to quickly identify common forms in the integrand which can be simplified by techniques like $u-$substitution, integration by parts, and trigonometric substitution. It is important to stress that that developing this intuition will take time and practice, and will depend on their ability to recognize integrals which are easy to solve, such as polynomials, logarithms, exponentials and basic trigonometric functions. By finding a way to transform the existing integrand to a more suitable form, we can often find integrals of very complicated expressions.

It is important for students to be reminded that when using $u-$substitution, they must also change the limits of integration on definite integrals based on the original limits of the original variable. This is a step that many students forget and don't understand, and thus should be emphasized in every instance that u-substitution is presented.

Lesson 42: L'Hopitals Rule (again)

This Lesson reviews L'Hopitals rule that was originally presented in Chapter 2, but in the context of transcendental functions. This presents an opportunity for students to practice using the rules of differentiation for transcendental functions which they just learned, as well as applying the product and chain rules where applicable. As mentioned previously, L'Hopitals rule helps us to handle indeterminate forms in fractions. It does so by noting that the rate at which a function approaches the value which is causing indeterminancy (usually zero or infinity) affects the nature of the indeterminancy. So if a function in the numerator is approaching zero much faster than the denominator is approaching zero, the numerator's zero-ness will dominate, and cause the seemingly indeterminate form to be evaluable.

The examples presented in this Lesson involve a number of the functions introduced in this Chapter, including logarithms, exponentials and trigonometric functions. Knowing when these values approach zero or infinity is essential to being able to identify cases where a limit is indeterminate. In many cases involving trigonometric functions, the point at which we are trying to find the limit is not zero or infinity, as was the case often in previous examples, since trigonometric functions often approach zero at non traditional points, like $\frac{\pi}{2}$ or $\frac{3\pi}{2}$. Making sure that students are comfortable with the zero crossings (and tendencies to infinity) of trigonometric functions is important to help them recognize indeterminate forms, as well as apply the appropriate techniques.

1.7 Integration Techniques

Lesson 43: Integration by Substitution

Integration by substitution is one of the most important techniques to being able to find integrals of many complicated expressions. Over time, students will begin using integral tables to solve more complicated integrals, and $u-$substitution is a common technique to being able to represent the given problem into a form that has a computable integrable.

As mentioned above, being able to identify the appropriate substitution takes much time and practice, as well as being comfortable with the types of integrals that are computable. What students should be looking for is the appearance of a term in the integrand which appears to be the derivative of the other term. This is often not obvious, but will become more obvious with time. It is also important for students to understand not only how to make a substitution for the expression in the integrand, but also for the differential element in the original integrand. For instance, if we are faced with the integral:

$$\int \frac{1}{\sqrt{x+3}} dx$$

we might make the substitution $u = x + 3$. Once an expression is selected for u, the next step is to take the derivative of u with respect to x and then by using differential notation, we can solve for dx.

$$\begin{aligned}
\frac{du}{dx} &= 2 \\
du &= 2dx \\
dx &= \frac{du}{2}
\end{aligned}$$

In some cases, this term with the dx appears explicitly in our integral, and thus the substitution is immediate. In other cases, we must solve for the term dx in terms of the other terms, and then substitute in for dx. This case is considered in the Examples.

Once a substitution is made, sometimes the original variable of integration still appears in the integrand, which is unacceptable if we are to integrate with respect to our new term. In these cases, we can sometimes solve for the old variable (in the above case, x), in terms of the new variable (in this case, u). In so doing, we can often eliminate the original variable completely from our expression, and evaluate the integral. We must also be careful to change the limits of integration so that they reflect the limits on the new variable, or be sure to change our variables back to the original to apply the original limits.

Lesson 44: Integration by Parts

Integration by parts, like $u-$substitution, requires students to be clever in their identification of appropriate substitutions to transform the given integral into a form which can be solved. In particular, it requires us to identify the integrand as a product of a function $'u'$ and the differential element of another function, $'dv'$. By working the product rule backwards, we can often transform this integral into one which is possible to solve. In general the following is the order in which students should select the function for "u" in the expression $u\ dv$, due to the ease of subsequent calculations:

Logarithms

Inverse Trigonometric Functions

Polynomials

Exponential Functions

Trigonometric Functions

Integration by parts is thus executed by applying its formula with the appropriate substitutions, and solving for what is hopefully an easier integral:

$$\int vdu$$

Many students may find it difficult to understand how to interpret the expression under the integrand as $u - dv$, and may want to always equate the dv term with the dx term. Though this is valid in some limited cases, as illustrated in this Lesson in Example 3, in general this will not be sufficient to transform the integral as needed.

What is usually required is to include the dx term with another portion of the integrand, and use this to represent dv. By integrating both sides, we arrive at the expression for v (hopefully) which can then be used in the formula. For instance:

$$\int ln(x)e^x dx \to u = ln(x),\ dv = e^x dx$$
$$v = \int dv = \int e^x dx = e^x$$

Students should become comfortable with always writing out their choice of substitution in order to develop the proper technique of converting the original diferential element (usually dx) into the appropriate term for dv.

A second point to emphasize is the use of repeated integration by parts to solve integrals that don't integrate nicely even after being transformed via the substitution above. In some cases, the new integral vdu is not easily integrable, but can have integration by parts applied again. Sometimes this yields another integral which CAN be integrated if the terms are selected appropriately.

In other cases (usually for trigonometric functions), the new integral is of the same form as the original integral. By moving it to the left hand side of the equation, and combining it with the original term, we are creating an algebraic expression to be solved for. In this case, our unknown is the entire expression containing the integral. By noticing that we are only left with function evaluations on the right hand side, we can solve directly for the first expression. This is a very non-obvious and algebraically clever technique which students may struggle to recognize in future problems, but is an essential tool in applying integration by parts effectively.

Lesson 45: Integration by Partial Fraction Expansion

This Lesson introduces a powerful technique for evaluating integrals of rational expressions through the method of partial fraction expansion. As the previous two Lessons have illustrated, an important skill to being able to evaluate complicated integrals is to transform them into a form which has a simpler integral. This is also the basis of partial fraction expansion, which tries to transform a rational expression into a sum of fractions whose denominator is a polynomial of power 1, which can then be integrated easily.

From the previous chapter, students learned that the family of functions $\frac{1}{x+a}$ has a straightforward integral: $ln|x+a|$. Functions like $\frac{l}{(x+a)\hat{}2}$ also have straightforward integrals as given by the power rule for integration. Partial Fraction Expansion exploits these faces by attempting to decompose rational expressions into a sum of such terms. Once our expression is in the form of a sum of partial fractions, we can then integrate each term separately.

As will become clear in later mathematics that students will encounter, the most effective way to perform partial fraction expansion is called the residue method, although this is definitely the less obvious of the two ways to perform a PFE. Making sure that students are comfortable with both techniques will be helpful as they move forward into engineering and physical applications, and will be the source of some confusion.

In addition, it is important to stress the techniques of using PFE when the numerator of our rational expression is of degree equal or greater than the degree of our denominator, as well as the situation where the denominator has repeated roots. These cases require us to apply slightly different techniques to execute the PFE which many students may not yet be comfortable with.

When the rational expression in the integrand is improper, meaning that the degree of the polynomial in the numerator is greater than the degree in the denominator, the first step is applying long division to yield a polynomial and a proper rational expression. Reviewing long division of polynomials is important, since many students may not have encountered this technique since learning about it in Algebra 2. Once a plynomial and proper rational expression are derived, we can apply the inverse power rule to the polynomial term, and PFE to the rational expression to evaluate the integral.

The second point of confusion for many students with PFE is the case when the denominator has repeated roots. In this case, not only must a term in the expansion be written which has the repeated root in the denominator, but also one which has the root only appear once. This is not obvious, and the failure to do so will yield equations which cannot be solved for to yield the appropriate expansion, leading to frustration for many students.

Lesson 46: Trigonometric Integrals

The use of trigonometric identities and properties to transform complicated integrals into ones which can be solved is an important tool, as discussed in Lesson ???. Because of this, it is essential that students are comfortable with the primary identities and properties of trigonometric functions being introduced in this Lesson.

This Lesson provides a number of examples of how to apply trigonometric identities to unfamiliar integrals, which are not obvious and will likely take students a long time to master. Emphasizing the need to be clever in applying these identities will help students to overcome much of the frustration they will first experience.

A common error when integrating powers of expressions, like $\sin^2(x)$ or $\cos^2(x)$ is to apply the power rule to the integrand to yield something like:

$$\int \sin^2(x)dx \neq \frac{1}{3}\sin^3(x)$$

This is an opportunity to illustrate the use of the chain rule in $u-$substitution, since the reason this doesn't work is that if we were to make the u substitution, $u = \sin(x)$, then we would need there to be another term with $\cos(x)$ for us to substitute the du for the dx. Since this term does not appear, it is clear that the power rule does not apply here. On the other hand, if we consider the case where the product of a sin and cosine appear in the integrand, then this is an opportunity to try $u-$substitution, since the function and its derivative appear as a product, making it likely that we will be able to substitute du into the integrand.

A similar reasoning applies to the integral of products of powers of $\tan(x)$ and $\sec(x)$, because the derivative of $\tan(x)$ is $\sec^2(x)$. When the powers of these functions are appropriate, as displayed in the table in this Lesson, we can also use $u-$substitution to solve these integrals. Again, this is an excellent opportunity to stress the thoroughness required when applying $u-$substitution which entails solving for the term $\frac{du}{dx}$, and then noting that $du = \frac{du}{dx} * dx$. If $\frac{du}{dx}$ appears in our original integrand in a product with u, then $u-$substitution allows us to gain the easily integrable expression:

$$\int u\frac{du}{dx}dx = \int u\ du$$

The important point throughout this Lesson is that there are certain compositions and products of trigonometric functions which can be integrated because of the rules of differentiation we learned earlier about trigonometric functions.

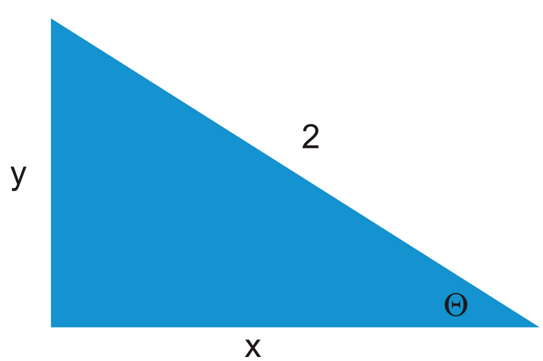
Lesson 47: Integration by Trigonometric Substitution

The techniques of trigonometric substitution and trigonometric integrals are quite different, and involve a different type of application. Both rely on trigonometric identities to transform complicated expressions into forms which can be integrated more easily. But unlike the examples from Lesson 46, the technique of trigonometric substitution is for situations where the properties of a trigonometric function can make its use as a $u-$substitution yield a solvable integral. This is often true even if no trigonometric functions are in the original integral.

The forms presented in this Lesson should be memorized by the student, since it is difficult to know which substitution to use without extensive practice. It should be stressed that the reason why we make these substitutions is because of the wonderful properties which some trigonometric functions satisfy, like $\sin^2 x + \cos^2 x = 1$, and has little to do with any underlying trigonometric significance of the functions in the integrand.

An important step in applying trigonometric substitution is to understand how to return to our original variable of integration. In these cases, we usually do not want to change our limits of integration into those of the substituted trigonometric function, but rather want to use the substitution to evaluate the newly transformed integral, and then transform our integral back to the original. This is done by using our knowledge of the definitions of trigonometric functions in terms of right triangles.

For instance, in example 2 we have made the substitution: $x = 2cos(\theta)$. Once we have performed the integral on our trigonometrically transformed function, we must then try to recover the original functions in the manner we did when using u-substitution. In the case of trigonometric $u-$substiutions, this often requires much more insight. Consider the triangle implied by the statement $x = 2\cos(\text{theta})$:



Since cosine is defined as the length of the adjacent side over the hypotenuse, we arrive at the triangle above. By using the Pythagorean theorem, we can solve for the side adjacent to theta, as follows:

$$x^2 + y^2 = 2^2$$
$$y = \sqrt{4 - x^2} = \text{opposite}$$

Now that we have an expression for all the sides of the triangle in terms of our original variable x, when we see the appearance of a trigonometric function in theta, we can transform it back to the original variable. In the case above,

$$\sin(\theta) = \frac{\sqrt{4 - x^2}}{2}$$
$$\tan(\theta) = \frac{\sqrt{4 - x^2}}{x}$$

By making appropriate $u-$substitutions involving trig functions, we can evaluate integrals which would otherwise be unsolvable, even when they don't involve trigonometric functions. Much of this has to do with the results discussed in this Chapter, which shows how the relationship between functions can often be emulated with substituted functions that are more amenable to analysis.

Lesson 48: Improper Integrals

Improper integrals tie together the notion of integration and limits in a new and powerful manner, and introduce the notion of convergence which students will encounter more thoroughly in the next chapter about infinite series. An improper integral is one which has limits of integration as infinity or negative infinity. If students wonder where this seemingly unrealistic situation might occur, make it clear that often we use positive and negative infinity as limits of integration to help us make approximations to situations that will go on for a long time or have been going on for a long time.

Though it might seem like an unnecessary step, to be consistent with our definition of limits and infinity, when an infinity is encountered in the limits of integration it is most appropriate to replace the infinity with a new variable, like L. We can then take the limit of the entire integral expression as L approaches infinity as seen in this Lesson. It turns out that even though we are adding up an infinite number of pieces, if the function decays to zero fast enough as it tends to infinity, then the result is still finite. This is another example of an indeterminate form, since it is essentially a product of zero with infinity.

Another form of improper integral discussed in this Lesson is the integration of functions with infinite discontinuities. The examples may cause some confusion for students, since it is claimed that unless our integrand is bounded on the entire interval, then its integral does not exist. If this is so, then how can we apply the techniques of integration to such functions. The easiest way to do this is to separate the integral into intervals where the integrand is finite, and place the points of discontinuity as endpoints. As seen in Example 4, when the point of infinite discontinuity is at an endpoint of the interval, it is possible to still get a finite answer for the integral. In these cases, instead of creating a dummy variable for the limit of integration and then taking the limit as it approaches infinity, we take the limit as the dummy variable approaches the point of infinite discontinuity.

Once again, this is a case where an indeterminate form exists, since the width of a point discontinuity as is seen in these examples is infinitesimal, and so its product with an infinte value is indeterminate. The techniques of applying limits allow us to sometimes find finite answers in these uncertain cases.

Lesson 49: Ordinary Differential Equations

The importance of the techniques of calculus to understanding physical phenomena is most evident in their application to differential equations. As mentioned previously, often when modeling a physical situation we are not able to write expressions solely in terms of our variable of interest; we must also include terms that express the rates of change of our variable, and rates of change of the rates of change. This was seen quite clearly when considering the phenomena of exponential growth and decay in Lesson 40.

This Lesson is an overview of techniques for solving ordinary differntial equations (as opposed to partial differential equations) using the techniques the student has already learned. It also serves to expand on the previous discussion of differential equations to more complicated expressions and techniques. In particular the consideration of a differential equation's slope field, sometimes referred to as its "phase portrait". This section is important to help students understand the intutive meaning of an ordinary differential equation - even if we canot solve for the function of interest directly, since we know the value of the derivative as a function of position, we can graph the solution by drawing a function which satisfies the expression for the functions's derivatives.

The final section of this lesson expands on the previous discussion of numerical techniques. Many of the differential equations encountered in physical applications cannot be solved by the analytical means developed in this text (or elsewhere), and thus must be subjected to numerical approximation to be solved. Even in these cases, however, an understanding of the calculus underlying these equations is essential to interpreting the results which a numerical approximation yields. The results we obtain from numerical techniques can often be skewed significantly by the approximation technique we employ. An understanding of the underlying calculus is essential to interpret our solutions appropriately and with caution.

1.8 Infinite Series

Lesson 50: Sequences

This Lesson introduces a number of concepts related to the topic of Infinite Series through their predecessor, infinite sequences. An infinite sequence is a progression of numbers that goes on forever, whereas a series is a sum of all of the numbers in a sequence. A sequence is said to converge if the values of its terms start to approach a particular finite value. It diverges if they approach infinity. This definition of convergence and divergence is quite different than that of series, which may be good to emphasize now.

This Lesson offers a number of formal definitions to define a sequence converging, but the most important tool is to understand how functions behave as a goes to positive or negative infinity. $\frac{1}{x}$, for example, starts to get very small as x goes to zero. If this term were to appear in the expression for our terms, then it would go to zero as n goes to infinity. Similarly, a positive number raised to a negative exponent will go to zero as the exponent goes to infinity. Understanding the long term or steady state behavior of functions as discussed in the early Chapters of this textbook is very helpful for quickly identifying what happens to terms in infinite sequences and series.

The topic of Picard iteration is also likely to cause students some confusion, since this is a case where we are considering a sequence of functions converging to another function, as opposed to a sequence of expressions converging to a number. The difference between these two phenomena is vast, and should be emphasized when discussing Picard's Theorem. As students will learn in subsequent Lessons, there are different conditions we apply to approximations to functions to make sure that a polynomial representation, like we see in the Example on p. 391, does actually converge to the function we are looking for.

The reason why we use Picard's method is that often the differential equation we are trying to solve is more difficult than any of the methods at our disposal. Picard's method is another example of a numerical iterative method we can use to get a function which approaches our true function, often perfectly. We are able to do this without solving the original equation, but rather we evaluate integrals of functions that we get by successive approximations. This point illustrates that the difference between a sequence of functions approaching a function, or a sequence of numbers approaching a value is that our functions must converge to our true function EVERYWHERE, and so for an infinite number of numbers (in some cases). This is a much more stringent requirement, but allows us to use the notion of infinite sequences to solve very difficult problems.

Lesson 51: Infinite Series

Beyond just representing the terms of an infinite progression of numbers, it is also helpful to represent the **sum** of an infinite progression of numbers. Calculus gives us the tools to do something like this. It turns out that there are series where the progression of the terms to zero is severe enough that they no longer affect our sum, and thus we can add up an infinite number of terms. It is essential that these terms converge to zero, since clearly adding up an infinite number of positive values will yield an infinite result. This Lesson illustrates a formal technique to consider the convergence of a series of partial sums that is helpful for a student to see, but not used very often to calculate actual convergence. It is more important for students to understand the terminology and caveats of understanding convergent series presented later in this Lesson and Chapter.

Besides an infinite series being the sum of terms in an infinite sequence, the relationship between sequences and series can also be considered the other way around through the topic of partial sums. The nth partial sum of a series is the sum of the first n terms. The sequence of partial sums is the sequence of these individual partial sums. This topic is important to determining the convergence of an infinite series since we can apply tests of sequence convergence to the partial sums, particularly if we can write the partial sums in a format that is a function of their index. In so doing, we can determine the convergence of a much wider class of infinite series than just the specific ones discussed in this and the subsequent lessons. This is illustrated by Example 8.

The topic of geometric series is presented in this Lesson, which students should already be comfortable with, and which will be discussed in further detail in the next lesson. In helping them to understand infinite series convergence, it may be helpful to show them how to prove the classic result for geometric series,

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}, \quad 0 < r < 1$$

since it involves interesting algebraic techniques to manipulate entire series. Another important result in geometric series whcih students should become familiar with, and which could cause some difficulty, is how to handle the case where the series is NOT infinite. For instance,

$$\sum_{i=0}^{N} r^i, \quad 0 < r < 1$$

In this case, we can represent the series as the difference of two infinite series, and then apply the result above to each to get a valid result for the sum. This is an interesting example since it stresses three important points that may cause confusion for many students: (1) that even if our series does not start at an index of $i = 0$, it can still be infinite, (2) Often dealing with infinite series is much easier than dealing with finite series, and (3) we must often manipulate our summand expressions using a change of variables to re-index our series. (3) is likely to cause great difficulty for students for some time, but is an essential technique to solving a wide range of differential equations using power series.

$$\sum_{i=0}^{N} r^i = \sum_{i=0}^{\infty} r^i - \sum_{i=N+1}^{\infty} r^i = \sum_{i=0}^{\infty} r^i - \sum_{i=0}^{\infty} r^{i+N+1} = \sum_{i=0}^{\infty} r^i - r^{N+1} \sum_{i=0}^{\infty} r^i$$
$$= (1 - r^{N+1}) \sum_{i=0}^{\infty} r^i = \frac{1 - r^{N+1}}{1 - r}$$

Lesson 52: Non-negative Series

This Lesson introduces students to a number of important types of series, including harmonic series, geometric series and $p-$series. These series have properties and tests for convergence which are unique and make for much easier analysis, and so are considered explicitly here.

The first point addressed in this Lesson is actually quite obvious, and its obviousness should be stressed to students. It is very important, however, to notice that these rules are applying to sequences, and not series. If the terms of a sequence are getting bigger and bigger, then it should be clear that the series of this terms does NOT converge, since if we continue to add up larger and larger numbers infinitely, then our sum will grow larger without bound. The point being made here, however, is for sequences, and this should be stressed. If the terms of a sequence are growing larger, but they are bounded, then eventually the terms must approach a limit, since they can't keep getting bigger without gettnig closer and closer to their bound. On the other hand, if they do not have a bound, then the terms will just keep growing larger and larger.

Geometric series are an extremely important tool for students, and they should have encountered them many times previously in their mathematics education. A geometric series is one in which each term is a fixed multiple of the previous term, and thus if we divide two adjacent terms, we will always get the same ratio. If this ratio is less than 1, then the geometric series converges and has a computable sum. If not, then it diverges and does now. Being able to identify the ratios between terms of a geometric series in non obvious cases is an important skill for students to master.

P series are related to harmonic series, but instead of each term being $\frac{1}{n}$, it will be $\frac{1}{n^p}$. The harmonic series is an example of a $p-$series having a $p-$value of 1, and as is stated in the Theorem on p. 410, $p-$series will only converge for $p-$values that are greater than 1, which confirms our result that the harmonic series diverges. It is interesting to note that we can have $p-$series with fractional exponents, and this Theorem tells us that only for fractional exponents greater than 1, the series will converge.

The remainder of this Lesson is focused on the most important examples of convergence tests which can be used to determine if an infinite series has a finite value. These are based on many of the limit theorems presented earlier in the text. There are a few important things to keep in mind when applying these tests: (1) the choice of test is based heavily on the situation encountered, (2) some tests will prove inconclusive about convergence, which does not mean the series diverges, and (3) the selection of a comparison series to create for the comparison tests requires a good deal of creativity to ensure a conclusive answer about convergence.

Lesson 53: Alternating Negative Series

There are many series which have terms whose signs alternate back and forth, from a positive to a negative value and back again. Undertstanding how to represent this phenomena in series notation, as well as the unique nature of limits of alternating series, is the central theme of this Lesson. The ways in which these considerations differ from the purely non-negative series we considered in the previous Lessons must be emphasized.

It is important in writing series in Sigma notation to understand how starting the index at 0 or 1 will affect our subsequent result, and how to operate on powers of numbers using this index. Students should be comfortable describing the sign of a term based on its oddness or evenness using the notation:

$(-1)^k$ or $(-1)^{k+1}$

For any infinite series, we can make its signs alternate by putting one of these terms into the term expression. It may be helpful for some students to know that these expressions only work for the case that we start with $k = 1$, and to make sure that the first term has the appropriate sign. In this case, if we would use the second expression above to ensure that the first term, i.e. $k = 1$, would yield a positive number. Alternatively, if we started the index of summation at $k = 0$, we would find that the opposite holds true. It is important to emphasize to students to check the first index of summation when using the expressions above to represent alternating signs.

A second important consideration when treating sequences or series with alternating negative signs is that we can no longer consider the terms approaching a limit as easily, since now our numbers will jump back and forth due to the sign changes. In these cases, we must apply different tests of convergence and limits, as discussed in this Lesson. For instance, what if we had an infinite series whose terms alternated in signs, but every two terms were the same? In this case, looking at this pattern implies that the sum of the entire series will be zero, regardless of whether the terms in the series get smaller and smaller. Cases like this can be helpful in introducing students to the greater subtleties involved when dealing with series that have alternating signs.

$$1 - 1 + \frac{1}{2} - \frac{1}{2} + \frac{1}{4} - \frac{1}{4} + \dots$$

Even though the magnitude of our terms do not go to zero, we can be certain that there will always be a finite value to this series if we can guarantee that each pair of terms cancel each other out. Situations like this make treatment of alternating series more subtle, as discussed in this Lesson.

Lesson 54: Convergence Tests

The convergence tests presented previously dealt with series which had only positive terms. Now that we have included the case of where the series can have terms with alternating signs, we must apply more rigorous tests to determine convergence. In so doing, we define two types of convergence of a series: (1) absolute convergence, where the series of absolute values of the terms converges, and (2) conditional convergence, where the series of terms converges regardless of its signage.

An important point is that if a series is absolutely convergent, than it is conditionally convergent. There are a number of ways to convince students of this, but perhaps is to stress that the value of the series without absolute values will always be smaller than the series with absolute values, since it translates the cases where we are adding negative values (which would make our sum smaller), to adding their absolute value.

However, it is important to note that just because the series is NOT absolutely convergent, this does not mean that it cannot be conditionally convergent. Indeed, conditional convergence is a weaker test of convergence. This is a confusing point, but as mentioned previously, many convergence tests provide sufficient but not necessary conditions, and so it must be stressed that failing a test of convergence does not always mean that the series does not converge.

The ratio test tells us that if the limit of two terms approaches a ratio that is less than 1, then the series will converge. This is somewhat non-obvious, since it doesn't tell us that the terms themselves goes to zero, rather to a finite value. This is a case where regardless of how our series starts, as we move farther along its sequence, we are starting to get a geometric series with a ratio less than one. So, if we were to cut the series into two parts, the initial finite part, and then the infinite part, the first part which may or may not be geometric (or have a ratio of terms less than 1), will add up to a finite value (since it is only a finite number of terms). The second half of our sequence, which is an infinite series, can be bounded by a geometric series whose ratio is less than the ratio bound of the terms (denoted as alpha) in the text.

Breaking up an infinite series into parts like this might be a confusing concept, but is also effective in representing series into parts that are susceptible to one of our tests of convergence.

Lesson 55: Power Series

Power series are one of the most powerful tools in mathematics as students will learn in the next Lesson. They are really nothing more than a polynomial with an infinite number of terms. the series we have been considering so far, a Power Series is actually a function with respect to a variable, as opposed to the summation of a sequence of nubers. In representing power series, we must usually find compact notation for the coefficients of each term in the polynomial, where a student's study of sequences and infinite series will prove useful.

The idea of a series being centered around a particular point is also very confusing to many students without any context of sense of how these series behave. We say that a power series is centered around a point when we write the series thus:

$$\sum_{i=0}^{N} a_i (x - x_0)^i$$

What does this mean and why do we care? These might be questions that arise in a student's mind, since a series should be a series, regardless of where we start. It is therefore important to stress that like in Picard's theorem, the series we are now representing are functions, and therefore their convergence to another function is based on their convergence to points everywhere on that function. So, the place where the power series approximation is centered might be a point where the original function and our power series approximation are exactly equal, but as we move away from this point, our power series approximation is no longer as valid.

If we think of the place where a series is centered, if we were to plug in that value of x, we would just be left with the first term in our series, since the other terms which contain $x - c$ will vanish. This first term is sometimes called the bias or offset of the power series. If we were to plug in other values for x, however, we also will get a series of numbers, whose sum may or may not converge. It turns out that usually there is a distance away from the center of our series for which values of x do lead to our series converging. This distance is referred to as the interval or radius of convergence, for real-valued and comple $x-$ valued series, respectively.

To determine a series' interval of convergence, a student must use the tests from previous lessons and apply it to the case where the terms in questions are variables. In so doing, no longer are we given exact ratios between terms, but rather these ratios are dependent on the values of x we plug in. By understanding what values of x will make our convergence test of choice yield the criterion of convergence, we are identifying the values of x which are inside the interval of convergence. For some power series, this interval is infinite, meaning that the power series converges for all values of x. These form an important case in the next few Lessons.

The wonderful thing about power series is that we can often differentiate them very easily, since they follow the Power Rule of differentiation. This proves to be very effective when solving differential equations, since if we can represent the solution of our function as a Power series, we can easily take any derivative of that function term by term. To use this technique to solve such equations, however, students must become comfortable with representing the derivatives of series as new series, where the coefficients are often a function of the index.

When we are using this technique on finite series, we must be very careful to adjust the limits of our summation to accommodate the lowering of our polynomial's degree by one, as well as often reducing the number of terms by 1. We can often accomplish this in multiple ways, by changing the starting index, the terminating index, and/or the expression for the index in the expression within the sigma sign (i.e. changing an n to an $n+1$ or $n-1$). In different contexts, students are likely to encounter each of these methods for representing the derivative of a finite power series, and so should be cautious.

In the case of infinite power series, this problem is often alleviated since we are not reducing the power of our original polynomial (since both it and its derivative will be infinite). This is not the only case where treating things infinitely is much easier than treating them finitely, and sometimes motivates us to use an infinite representation to approximate a finite one.

These same precautions apply to integrating a powers series term by term, with the additional complication caused by taking the integral of $\frac{1}{n}$, yielding a $ln(n)$. In some of these cases, by integrating our power series we seriously affect its properties and interval of convergence.

Lesson 56: Taylor and MacLaurin Series

Much of this chapter, and the discussion of Linear and Non-linear approximations to functions in previous Lessons, has been an introduction to one of the most powerful tools in mathematics: the Taylor and Maclaurin series. For many functions, we can represent them as a power series with an infinite number of terms for at least part of their domain. This series is usually referred to as the function's Taylor series approximation. This proves very useful when solving differential equations as well as in analyzing properties of unfamiliar and unwieldy functions.

This Lesson is likely to cause a great deal of confusion for students, as a number of new concepts and representations are being introduced. It might be helpful for them to review the chapters on linear and non-linear approximations, and show them how Taylor Series is really just an extension of those techniques. If we start adding up enough derivative terms, we can often approximate the function exactly ... especially if we can add up an infinite number of terms.

A tricky part of Taylor Series is based on the discussion in the Lesson on Power Series concerning the place at which we center our Taylor series. Just as in the case of the linear approximation, we needed to pick a place to evaluate our function and its derivative to construct a linear expression, and this linear expression was only accurate near the point of interest. This point of interest is called the "center" of our series, and can be selected in a number of ways. The series approximations we arrive at are usually more closely aligned with the original function near the point at which the series is centered.

The appearance of factorials and other combinatorial expressions may also be the source of confusion for some students, but is essentially derived from the Power rule for differentiation. If we were to continually differentiate a polynomial using the power rule, the exponents would start to multiply each other as we brought them in front of the term we were differentiating, as seen in these expressions for the first few derivatives of a generic power series.

$$f(x) = \sum_{i=0}^{N} a_i x^i$$

$$f'(x) = \sum_{i=1}^{N} i a_i x^{i-1}$$

$$f''(x) = \sum_{i=2}^{N} i(i-1) a_i x^{i-2}$$

For functions that are continuous and differentiable over the entire real axis, the Taylor series with an infinite number of terms will be a valid approximation to the function at all points on the domain. This will become evident by considering the interval of convergence of the Taylor series, which in these cases will be infinite. Even for functions which have discontinuities or sharp corners, we can still often use a Taylor series approximation to approximate our function for a particular region of the domain. What we find is that the interval of convergence for the Taylor series will be the distance from the center to the point of discontinuity or indifferentiability. This happens because the smooth functions generated by polynomial series approximations cannot mimic points of indifferentirability or discontinuity, and thus the series approximation at these points does not converge to the actual function. This is what motivated the discussion of intervals of convergence in the previous Lesson.

In most practical cases, particularly those involving computers to represent functions, it is not possible to retain an infinite number of terms. In these cases, we must determine how many terms to keep to get an approximation to our original function that is "good enough". To do so, we use the Remainder Estimation theorem, which provides an upper bound to the truncation error, or the error imposed by limiting the number of terms in the Taylor Series that we retain. It is important to emphasize that the value of this remainder depends on the point we are at, meaning that the truncation error will be different for different values of x. However, by identifying an upper bound to the derivatives of the original function on the appropriate interval, we can bound the error everywhere on that interval through this Theorem.

The power of Taylor Series approximations has already been mentioned in the context of ordinary differential equations, which is emphasized again in this Lesson's section on "Evaluating Nonelementary Integrals". Often a function we are trying to integrate or differentiate does not follow any of the rules we have. Sometimes there does not exist an analytic solution to the integral or derivative, and in these cases we use series approximations to represent the function, since we can easily differentiate or integrate them term by term, as mentioned in the last Lesson.

Lesson 57: Calculation with Series

This Lesson builds on the topics of the previous lessons, and offers a few more techniques to performing calculations with infinite and binomial series. There are a number of interesting techniques presented, including how to represent certain functions of the form

$$(1 + x)^r$$

in terms of a binomial series expansion using combinatorial representations.

Most students will have only encountered this representation where both r and k are whole numbers, since this form is usually reserved for dealing with combinatorial situations. For instance, the expression

$$\binom{n}{k}$$

can be interpreted as the number of ways to choose k items from a population of n items. If n is a fraction, this analogy seems to break down, which is why it is essential for students to know the formal definition for how to expand expressions like above.

The second important topic expressed in this Lesson is that of how to choose the point to center our Taylor Series around. As mentioned above, we are usually unable to keep an infinite number of terms in the series expansion of a function, and so must truncate our series at a particular point. Usually, the more terms we keep, the better our approximation will be. Students should recall from determining linear approximations to functions that another factor to consider the place that we have centered our series. Since the series will tend to be a much better approximation to our function near the point at which we center our expansion, to approximate the function near the series' center we will need far fewer terms to get a good approximation. This is analagous to the linear approximation where the linear representation of the function was really only valid in a small region around the point of tangency. The center of the series approximation is equivalent to the point of tangency emplyed when finding linear approximations.