Chapter 5

Using Derivatives to Analyze Functions; Applications

In this chapter we develop methods for analyzing functions extensively by exploiting the information contained in their derivatives. We also investigate higher-order derivatives, where the second derivative is the derivative of the derivative function, the third derivative is the derivative of the second derivative, and so on. We will pay particular attention to the theorems and intutions involving a function's second derivative.

Of course a very natural and general approach for analyzing functions is to consider their graphs in the Cartesian Plane, and much of this chapter is devoted to describing the graphs by analyzing their derivatives. While the plane is an abstract setting, the analysis will have useful and intuitive import to applied problems.

Fortunately we will be able to prove all of the results presented in this chapter, referring only to previously proved results and already mentioned facts we borrow from more advanced studies. However, we have already used some of the intuition without proof, for example when we noted that it seemed reasonable to believe

$$\begin{aligned} f'(x) > 0 \text{ on } (a,b) &\implies f(x) \text{ increasing on } (a,b), \\ f'(x) < 0 \text{ on } (a,b) &\implies f(x) \text{ decreasing on } (a,b) \end{aligned}$$

and stated this as Theorem 4.2.5, page 276. In this chapter we will finally prove this theorem, using the Mean Value Theorem. In fact we will have a slightly modified version in which (a,b) is replaced by a closed interval $[a,b]$.

While the Mean Value Theorem is interesting, its power is not immediately obvious. Nonetheless it is astonishingly useful as a cornerstone in proving many of the other intuitive, but difficult to prove, results dealing with the implications of the first derivative. Furthermore, it lends itself to many practical problems, particularly those involving average rates of change and inequalities.

The graphical significance of a function's first derivative as slope was discussed in the previous section, and should seem straighforward. The significance of the second derivative is nearly as straightforward, though it is often necessary to break that analysis into more cases which depend upon the value (or sign) of the first derivative. Beyond the second derivative, it becomes more difficult to see the significance, and it relies upon the values (or signs) of the first two derivatives, making it more complicated and usually out of reach graphically. Similarly for even higher-oder derivatives. Still, the second derivative is important enough for its own analysis, albeit coupled to that of the first derivative.

5.1 Extrema on Closed Intervals

In this section we see how to locate the extrema of a continuous function $f : [a, b] \longrightarrow \mathbb{R}$. The main theorem is Theorem 5.1.2, page 380. However, it is a long argument for why that theorem is valid, so the argument is given its own subsection directly below, and the theorem opens Subsection 5.1.2.

5.1.1 Argument for Main Theorem

We wish to find those points in $[a, b]$ at which a continuous function $f : [a, b] \longrightarrow \mathbb{R}$ achieves its maximum and minimum values for $x \in [a, b]$. Recall from Section 3.3 that $f([a, b])$ will itself be a finite, closed interval, and thus will have a maximum and minimum value, i.e., there will indeed exist some x_{min}, x_{max} in the interval $[a, b]$ so that $f([a, b]) = [f(x_{\text{min}}), f(x_{\text{max}})]$, i.e.,

$$(\exists x_{\text{min}}, x_{\text{max}} \in [a, b])\,(\forall x \in [a, b])\,[f(x_{\text{min}}) \le f(x) \le f(x_{\text{max}})]\,. \tag{5.1}$$

This was the essence of the Extreme Value Theorem (EVT), which was Corollary 3.3.1, page 177.[1] The next theorem is stated in a manner reflective of its proof. However, its equivalent forms given in the corollaries are more useful in applications.

Theorem 5.1.1 *Suppose $f : [a, b] \longrightarrow \mathbb{R}$ is continuous, and that $f(x)$ achieves its maximum or minimum value for the interval $[a, b]$ at an interior point $x_0 \in (a, b)$. Under these assumptions, if $f'(x_0)$ exists then $f'(x_0) = 0$, i.e.,*

$$f'(x_0) \text{ exists} \implies f'(x_0) = 0.$$

<u>Proof:</u> First we look at the case $f : [a, b] \longrightarrow \mathbb{R}$ continuous, $x_0 \in (a, b)$, and $f(x_0)$ is the maximum value of $f(x)$ on $[a, b]$. We need to show that $f'(x_0)$ exists implies $f'(x_0) = 0$. So we (further) suppose $f'(x_0)$ exists. This implies

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = f'(x_0) \text{ exists.}$$

Thus the left and right side limits exist and must agree with this two-sided limit:

$$f'(x_0) = \lim_{\Delta x \to 0^-} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}, \tag{5.2}$$

$$f'(x_0) = \lim_{\Delta x \to 0^+} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}. \tag{5.3}$$

Now we carefully consider each of these two limits, (5.2) and (5.3). In particular we look at the signs of these. In both limits above, since $f(x_0)$ is a maximum, we have $f(x_0 + \Delta x) \le f(x_0)$, implying $f(x_0 + \Delta x) - f(x_0) \le 0$. Thus the numerators in the limits are both nonpositive.

[1] It is very strongly suggested that the reader at least briefly review Section 3.3 at some point during the reading of this current section. It would likely be particularly helpful to reconsider the figures in that section to recall why it is necessary to have a continuous function *and* a closed interval for the conclusion of the Extreme Value Theorem to be guaranteed.

Next we look specifically at (5.2). Since $\Delta x \to 0^-$, we are looking at limits of fractions with denominators $\Delta x < 0$. Thus we have

$$f'(x_0) = \lim_{\Delta x \to 0^-} \frac{\overbrace{f(x_0 + \Delta x) - f(x_0)}^{\le 0}}{\underbrace{\Delta x}_{<0}} \ge 0. \tag{5.4}$$

On the other hand, when we instead look closely at the form of the limit in (5.3), we see

$$f'(x_0) = \lim_{\Delta x \to 0^+} \frac{\overbrace{f(x_0 + \Delta x) - f(x_0)}^{\le 0}}{\underbrace{\Delta x}_{>0}} \le 0. \tag{5.5}$$

So from (5.4) and (5.5) we have $(f'(x_0) \ge 0) \wedge (f'(x_0) \le 0)$, i.e., $0 \le f'(x_0) \le 0$, so we have to conclude $f'(x_0) = 0$.

The case where $f(x_0)$ is the minimum value of x on $[a, b]$ is similar, except that $f(x_0+\Delta x)-f(x_0) \ge 0$ in the limits (5.4) and (5.5), giving $(f'(x_0) \le 0)\wedge(f'(x_0) \ge 0)$, repectively, yielding again $f'(x_0) = 0$, q.e.d.[2]

It is useful to consider graphically why (5.4) and (5.5) should hold, as well as the conclusion of the theorem. This is left to the reader.

Now, recalling that $P \longrightarrow Q \iff (\sim P) \vee Q$ (see (1.29), page 22, though first proved as (1.8), page 18), we can rewrite the conclusion of Theorem 5.1.1:

$$[f'(x_0) \text{ exists }] \longrightarrow [f'(x_0) = 0] \iff [f'(x_0) \text{ does not exist}] \vee [f'(x_0) = 0].$$

The complete rewriting of Theorem 5.1.1 we state as a corollary:

Corollary 5.1.1 *Suppose $f : [a, b] \longrightarrow \mathbb{R}$ is continuous, and achieves its maximum or minimum at a point $x_0 \in (a, b)$, then*

$$[f'(x_0) = 0] \vee [f'(x_0) \textit{ does not exist}].$$

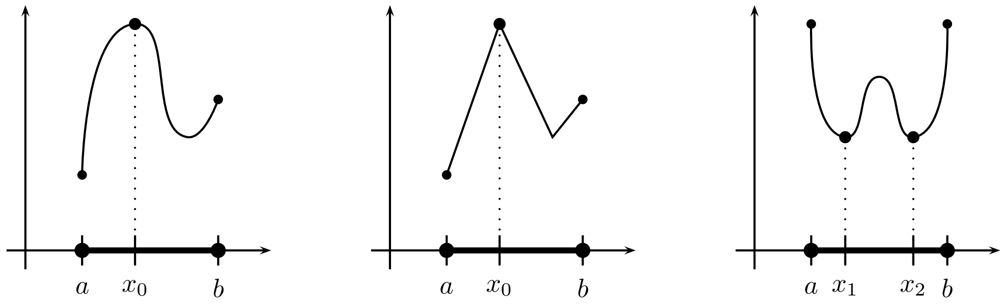
So any interior extremum of a continuous function on a closed interval $[a, b]$ must occur at points where f' is zero or does not exist. This is illustrated in Figure 5.1.

Because of the two possibilities for the derivative at an interior extremum, it is convenient to define those points where $f'(x) = 0$ or $f'(x)$ does not exist. We collectively call such points *critical points.*

Definition 5.1.1 *For $f(x)$ defined and continuous on $(a - \delta, a + \delta)$, for some $\delta > 0$, the point $x = a$ is called a* **critical point** *of $f(x)$ if and only if $f'(a) = 0$ or $f'(a)$ does not exist.*

It helps later to clarify now that critical points are always interior points, allowing for some "wiggle room" both to the left and to the right and still being in the interval in question. Using this definition, we can again rewrite Theorem 5.1.1, and Corollary 5.1.1:

[2] In the proof above it was necessary that $x_0 \in (a, b)$, so that there is room to the left and right of x_0 in the interval (a, b), so we can have $\Delta x \to 0^-$ and $\Delta x \to 0^+$ and eventually have $x + \Delta x \in (a, b)$ as well. In particular, if $x_0 \in \{a, b\}$ we can say nothing of its derivative.



x_0 an interior maximum, $f'(x_0) = 0$. No interior minima.

x_0 an interior maximum, $f'(x_0)$ does not exist. No interior minima.

x_1, x_2 interior minima, $f'(x_1), f(x_2) = 0$. No interior maxima.

Figure 5.1: Illustration that interior extrema can only occur at points where f' is zero or does not exist. In fact, the same is true of local extrema (not all labelled above).

For $f(x)$ continuous on $[a, b]$, any interior extremum must occur at a critical point.

This is still not quite the final and most useful version of this theorem. To arrive at the most complete statement, let us first recall (from Section 3.3) that if $f(x)$ is continuous on $[a, b]$, it will necessarily achieve its maximum and minimum values for that interval on that interval. Now we can import our theorem, which states that if one of these is achieved in the *interior* of $[a, b]$, i.e., on (a, b), then that point must be a critical point. Assume that x_0 is *any* point in $[a, b]$ at which f achieves either its maximum or minimum value for that interval $[a, b]$. Then

$$\begin{aligned} x_0 \in [a,b] &\iff (x_0 \in (a,b)) \vee (x_0 \in \{a,b\}) \\ &\implies (x_0 \text{ a critical point of } f) \vee (x_0 \in \{a,b\}). \end{aligned}$$

In other words, x_0 must be a critical point of f in (a, b), or an endpoint of the interval $[a, b]$. Thus, when looking for a maximum or minimum value of a continuous f on $[a, b]$, we need only look at the set of critical points in (a, b), and at the endpoints a and b, for our candidates for $x_{\min}$ and $x_{\max}$.

5.1.2 Main Theorem

We tie all this together in the following theorem, which includes for context the Extreme Value Theorem first given as Corollary 3.3.1, page 177.

Theorem 5.1.2 *For any function $f(x)$, continuous on a closed interval $[a, b]$,*

1. *(EVT) there exist $x_{min}, x_{max} \in [a, b]$ such that*

$$(\forall x \in [a,b])[f(x_{min}) \le f(x) \le f(x_{max})]; \tag{5.6}$$

2. *furthermore, any such $x_{\min}, x_{\max} \in [a, b]$ satisfying (5.6) must be a critical point of f, or an endpoint a, b of the interval $[a, b]$.*

In other words, as long as the function is continuous on $[a, b]$, we need only scrutinize the critical points—where f' is zero or does not exist—on (a, b), and the points a, b themselves. Checking the function values at these points exhausts all possibilities for the maximum and minimum values of $f(x)$. The analysis also yields their locations, as x-values.

Example 5.1.1 *Find the location and values of the maximum and minimum values of* $f(x) = x^3 - x$ *for* $x \in [0, 10]$.

<u>*Solution*</u>*: Clearly* $f(x)$ *is continuous on* $[0, 10]$ *(it is a polynomial), and* $f'(x) = 3x^2 - 1$ *exists throughout* $(0, 10)$. *We first find the critical points, and then check the values of* f *at any critical points in* $(0, 10)$, *and then the endpoints* $x = 0, 10$.

$$\begin{aligned} f'(x) = 0 &\iff 3x^2 - 1 = 0 \\ &\iff \quad 3x^2 = 1 \\ &\iff \quad x^2 = 1/3 \iff x = \pm 1/\sqrt{3}. \end{aligned}$$

Now $-1/\sqrt{3} \notin (0, 10)$, *so the only relevant critical point is* $x = 1/\sqrt{3} \approx 0.577350269 \approx 0.577$.[3] *Now we check the functional (output) values at the one critical point, and the two endpoints.*

critical point: $f\left(1/\sqrt{3}\right) = \left(1/\sqrt{3}\right)^3 - 1/\sqrt{3} = \dfrac{1}{3\sqrt{3}} - \dfrac{1}{\sqrt{3}} = \dfrac{1-3}{\sqrt{3}} = \dfrac{-2}{\sqrt{3}} \approx -1.154700538.$

endpoints: $f(0) = 0^3 - 0 = 0$, *and* $f(10) = 10^3 - 10 = 1000 - 10 = 990$.

From this we get that the minimum value of f *on* $[0, 10]$ *is* $-2/\sqrt{3} \approx -1.155$, *occurring at* $x = 1/\sqrt{3} \approx 0.577$, *and the maximum value is* 990 *occuring at* $x = 10$.

We could have graphed $y = f(x)$, but to have the vertical scale to accommodate a clear picture of both the minimum and the maximum would be difficult. Significantly, the method leaves us with only three values to check for this particular example.

Example 5.1.2 $f(x) = x - \sqrt{x}$

5.2 The Mean Value Theorem

The mean value theorem states the surprisingly useful fact that the average rate of change of a function over an interval must occur as an instantaneous rate of change somewhere in that interval. It is akin to stating that if our average velocity on a (one-dimensional) trip were 60 miles per hour (mph), then the instantaneous velocity had to be 60mph at some time during the trip. It seems reasonable, that if we were to drop below the 60mph average for a while, we would have to speed up to a velocity greater than 60mph to compensate, and in doing so we would pass through the 60mph as an actual speed at least once. Of course this assumes that the velocity is continuous, because then to pass from below 60mph to above 60mph we have to pass through 60mph.

Similarly for other functions, but we have to define what we mean by average rate of change of a function, and determine the necessary hypotheses.

5.2.1 Linear Interpolation and Average Rate of Change

Suppose we are given a function $f : [a, b] \longrightarrow \mathbb{R}$ which is continuous on $[a, b]$. We can connect the two points $(a, f(a))$ and $(b, f(b))$ by a line, which is itself a function we call the *linear interpolation* of f between the two points. Graphically it is also called a *secant line* of the curve

[3] The approximation $x = 1/\sqrt{3} \approx 0.577350269$ is given so that some idea of the actual value is visible, and also to illustrate that, indeed, $1/\sqrt{3} \in [0, 10]$. For the sake of illustration, the approximate values of both inputs and outputs of the functions will be given routinely, though the analysis should use the exact values before stating approximations. Indeed, we will refrain from using any approximations in actual calculations throughout this text.

$y = f(x)$, meaning literally a line connecting two points on a circle, but generalized to mean any graph.

The equation of the linear interpolation is readily calculated. It has slope $(f(b)-f(a))/(b-a)$, and so when we note that $(a, f(a))$ is one point on the graph, we can quickly write $y = l(x)$, where

$$l(x) = f(a) + \left[\frac{f(b) - f(a)}{b - a}\right](x - a). \tag{5.7}$$

The linear interpolation is useful if we wish to approximate $f(x)$ for some x between or near a and b, but only have knowledge of f evaluated at a and b. In fact, we can also approximate a value for x given the desired output value for $f(x)$ by solving for x in $y = l(x)$. While the accuracy of the interpolation varies too much to be reliable without further information on the behavior of f on (a, b), often there are tables of pairs (x, y) for functions which behave somewhat linearly between such pairs, and linear interpolation (using $y = l(x)$) is a fairly standard procedure.

At this point we might again define the *average rate of change of* $f(x)$ *over the interval* $[a, b]$ as we did briefly in Chapter 4, specifically (4.4), page 265. It is the net change in f over $[a, b]$, divided by the length of the interval, to give the averate change in $f(x)$ per unit length traversed in x, namely

$$\frac{f(b) - f(a)}{b - a}.$$

It is also the slope l' of the linear interpolation of f on the same interval, i.e., the slope of the line through $(a, f(a))$ and $(b, f(b))$.

5.2.2 Mean Value Theorem

Here we give the statement of the Mean Value Theorem, and then look at an important Lemma known as Rolle's Theorem, which we then use in the proof of the Mean Value Theorem. Recall that f being differentiable means that f' exists.

Theorem 5.2.1 (Mean Value Theorem) *Suppose* $f : [a, b] \longrightarrow \mathbb{R}$ *is continuous on* $[a, b]$ *and differentiable on* (a, b)*. Then there exists* $\xi \in (a, b)$ *such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}. \tag{5.8}$$

Before we begin to prove this, note that what this says is that there is a point between a and b on which the slope of the curve matches that of the linear interpolation from $(a, f(a))$ to $(b, f(b))$, i.e., matches the average rate of change of $f(x)$ over $[a, b]$.

See Figure 5.2 for an illustration of this theorem. Note the line $y = l(x)$, called the *interpolation line* connecting $(a, f(a))$ and $(b, f(b))$. It is also called a *secant line* to the graph because it connects two points on the graph.

Now we turn to the proof, but in two stages. First we state and prove what is known as Rolle's Theorem, which is perhaps a bit more intuitive than the Mean Value Theorem.

Theorem 5.2.2 (Rolle's Theorem) *Suppose that* $f : [a, b] \longrightarrow \mathbb{R}$ *is continuous on* (a, b) *and differentiable on* (a, b)*. If* $f(a) = f(b)$*, then there exists* $\xi \in (a, b)$ *such that* $f'(\xi) = 0$*.*

<u>Proof</u>: *It is useful to consider two cases. First, that* $f(x)$ *is constant, with the common value* $f(a) = f(b)$ *on all of* $[a, b]$*, in which case* $f'(\xi) = 0$ *for all* $\xi \in (a, b)$*, and the conclusion of the theorem holds true.*

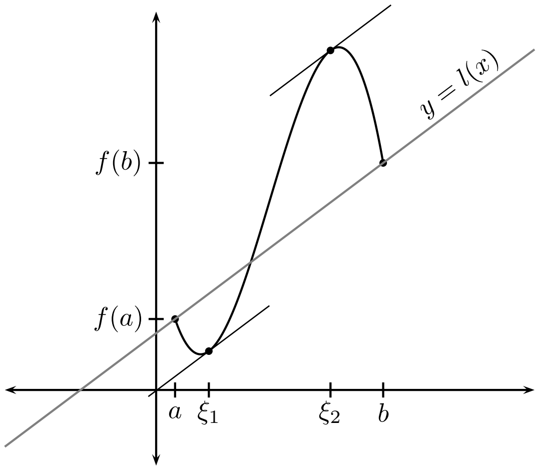


Figure 5.2: An illustration of the Mean Value Theorem, with the function $f(x)$ being continuous on $[a, b]$ and differentiable on (a, b). The graph $y = l(x)$ is the line which interpolates $(a, f(a))$ and $(b, f(b))$, and has slope equal to the average rate of change of $f(x)$ over $a \leq x \leq b$. According to the Mean Value Theorem, there exists $\xi \in (a, b)$ so that $f'(\xi)$ is also equal to the slope of $l(x)$. For this particular function there are, in fact, two such values, labeled ξ_1 and ξ_2.

For the other case, we suppose $f(x)$ is not a constant function on $[a, b]$, so it must achieve a maximum or minimum value for $x \in (a, b)$ (other than $f(a) = f(b)$), at some point $\xi \in (a, b)$. Thus ξ is a critical point of $f(x)$, and since $f'(\xi)$ exists, we must conclude $f'(\xi) = 0$, q.e.d.

5.2.3 Applications to Graphing Theorems

Recall that we "observed" in Chapter 4 that $f' > 0$ on an interval implied f was increasing on that interval. Similarly $f' < 0$ on the interval showed f was decreasing on that interval. With the Mean Value Theorem at our disposal, we are now in a position to prove these.

Theorem 5.2.3 *Suppose $a < b$, $f(x)$ is continuous on $[a, b]$ and $f'(x)$ exists on (a, b).*

(1) If $f'(x) > 0$ on $[a, b]$, then $f(x)$ is increasing on $[a, b]$.

(2) If $f'(x) < 0$ on $[a, b]$, then $f(x)$ is decreasing on $[a, b]$.

Proof: *First we prove (1). Let $a \leq x_1 < x_2 \leq b$. Then there exists $\xi \in (x_1, x_2)$ such that*

$$\frac{f(x_2) - f(x_2)}{x_2 - x_1} = f'(\xi) > 0 \implies f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) > 0$$
$$\implies f(x_2) > f(x_1).$$

This being true for all $x_1, x_2 \in [a, b]$ where $x_1 < x_2$ shows $f(x)$ is increasing on

$[x_1, x_2]$. *Case (2) is an easy modification of the argument to prove case (1):*

$$\frac{f(x_2) - f(x_2)}{x_2 - x_1} = f'(\xi) < 0 \implies f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) < 0$$
$$\implies f(x_2) < f(x_1).$$

Another useful and intuitive theorem is the following.

Theorem 5.2.4 *Suppose $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Moreover, suppose $f'(x) = 0$ for all $x \in (a, b)$. Then $f(x)$ is constant on $[a, b]$.*

Proof: *The proof follows a similar strategy as above. Let $a \le x_1 < x_2 \le b$. Then there exists $\xi \in (x_1, x_2)$ such that*

$$\frac{f(x_2) - f(x_2)}{x_2 - x_1} = f'(\xi) = 0 \implies f(x_2) - f(x_1) = 0 \implies f(x_2) = f(x_1).$$

While the theorem itself is interesting, it seems obvious enough to be almost worthless. But in fact it has a very nice corollary, which is that if two functions have the same derivative on an interval, they must differ by a constant.

Corollary 5.2.1 *Suppose $f(x)$ and $g(x)$ are continuous on $[a, b]$, differentiable on (a, b), and $f'(x) = g'(x)$ on (a, b). Then there exists $C \in \mathbb{R}$ such that $f(x) = g(x) + C$.*

Proof: *Consider the function $h(x) = f(x) - g(x)$, which is also continuous on $[a, b]$, differentiable on (a, b), and $h'(x) = f'(x) - g'(x) = 0$ on (a, b), so there exists $C \in \mathbb{R}$ such that, on $[a, b]$ we have*

$$h(x) = C \iff f(x) - g(x) = C \iff f(x) = g(x) + C.$$

Example 5.2.1 *Note how, on an interval such as $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ we have*

$$\begin{aligned} \frac{d}{dx}\tan^2 x &= 2\tan x \cdot \frac{d\tan x}{dx} = 2\tan x \sec^2 x x &&= 2\sec^2 x \tan^2 x, \\ \frac{d}{dx}\sec^2 x &= 2\sec x \cdot \frac{d\sec x}{dx} = 2\sec x \sec x \tan x &&= 2\sec^2 x \tan x. \end{aligned}$$

Thus $\frac{d}{dx}\tan^2 x = \frac{d}{dx}\sec^2 x$ on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, leading us to conclude $\tan^2 x = \sec^2 x + C$. Indeed, with $C = -1$ we have one of our basic trigonometric identities, namely $\tan^2 x = \sec^2 x - 1$. (Of course, this identity actually holds true anywhere $\cos x \neq 0$.)

Graphically, two functions with the same derivatives on an interval will be the same shape. One will just be a vertical translation of the other.

5.2.4 Numerical Applications

The existence of the value ξ with the prescribed derivative (equal to the average rate of change) can imply many things. Knowing bounds for the derivative over an interval means we know bounds for $f'(\xi)$, and thus the average rate of change. Consider the following example.

Example 5.2.2 *Suppose $f(x)$ is continuous on $[2,10]$, differentiable on $(2,10)$, and $f'(x) > 3$ on $(2,10)$. If $f(2) = 5$, find a lower bound for $f(10)$.*

Solution: There exists $\xi \in (2,10)$ such that $f'(\xi) = (f(10) - f(2))(10 - 2)$. Now

$$\frac{f(10) - f(2)}{10 - 2} = f'(\xi) > 3 \implies f(10) - f(2) = f'(\xi)(10 - 2) > 3(8) = 24$$
$$\implies f(10) > f(2) + 24 = 5 + 24 = 29.$$

We conclude that $f(10) > 29$.

Note that it was important that we multiplied the first equation by $10 - 2 > 0$.

The example is akin to asking, if $s = 5$ ft when $t = 2$ sec, and $v > 3$ ft/sec at all times $t \in (2\text{sec}, 10\text{sec})$, then where must $s(10\text{sec})$ be? If our starting position is at 5 ft, and our velocity is greater than 3 ft/sec for a time interval of length 8 sec, then we must have travelled more than 24 ft in that time, leaving us to the right of the position marked 29 ft.

In fact this method works for more general problems.

Example 5.2.3 *Supposing $f(x)$ is continuous on $[3,15]$, and differentiable on $(3,15)$ in such a way that $-2 \le f'(x) < 4$ on that interval, and $f(3) = 6$, find an interval containing $f(15)$.*

Solution: According to the Mean Value Theorem, there exists $\xi \in (3,15)$ such that $f(\xi) = (f(15) - f(3))/(15 - 3)$. We can use this to find bounds for $f(15)$ as follows:

$$\frac{f(15) - f(3)}{15 - 3} = f(\xi) \in [-2,4) \implies -2 \le \frac{f(15) - f(3)}{15 - 3} < 4$$
$$\implies -2 \le \frac{f(15) - 6}{12} < 4$$
$$\implies -12 \le f(15) - 6 < 48$$
$$\implies -6 \le f(15) < 54.$$

In fact, one can use a geometric interpretation of f' to construct lines of the slopes given by bounds on $f'(x)$ for $3 \le x \le 15$, and the values of $f(x)$ will fall between these lines. This is illustrated in Figure 5.3.

5.3 Differentials and the Linear Approximation Method

The main idea of this section is to show how the tangent at $(a, f(a))$ can be used to approximate the function $f(x)$ near $x = a$. But because differential calculus (calculus of derivatives) is not just about tangent lines, we will first show intuitively how the same idea is reasonable in other contexts. Finally, this will give yet another chance to interpret the Leibniz notation, this time giving dx and $df(x)$ numerical significance, consistent with but beyond the meaning we gave these before, when we only considered these symbols when put together formally in a fraction $\frac{df(x)}{dx} = f'(x)$.

We will begin this section with some intuitive examples to motivate the technique. We will then show how these are specific cases of *linear approximations*. At that point we will introduce differentials to show their geometric meaning.

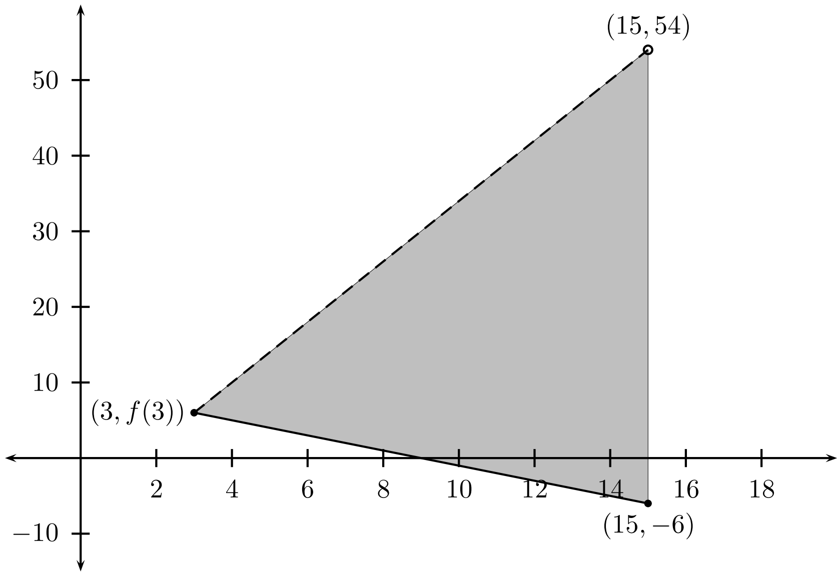


Figure 5.3: Illustration of the situation in Example 5.2.3. There, $f'(x) \in [-2, 4)$, $f(3) = 6$, and bounds for $f(15)$ are desired. Geometrically, the function is bounded by the two lines shown above, throughout all of $3 < x \leq 15$. The final bounds of $f(x)$ at $x = 15$ are labeled above. The upper line represents the "extreme" case $f'(x) = 4$ throughout $[3, 15]$, and the lower line the other "extreme" case where $f'(x) = -2$ on throughout all of $[3, 15]$.

5.3.1 Some Approximation Problems

The following examples should be intuitive, and eventually we will see that the underlying idea in each is the same. We will discuss this principle after we introduce the examples.

Example 5.3.1 *Suppose a car travels along a highway, and a passenger notices the speedometer reads 72 miles per hour. About how far will the car travel in the next second?*

<u>Solution</u>: *We have very little information here, but a reasonable assumption to make is that the car will not change speed very much in that one second. To the extent that that is true, we can say the distance traveled is given approximately by the following:*

$$\textit{Distance} \approx \frac{72 \text{ mile}}{\text{hour}} \cdot 1 \text{ second} = \frac{72 \text{ mile}}{1 \text{ hour}} \cdot \frac{5280 \text{ foot}}{1 \text{ mile}} \cdot \frac{1 \text{ hour}}{3600 \text{ second}} \cdot 1 \text{ second} = 105.6 \text{ foot}.$$

The idea in the above example is akin to the grade school formula (distance) = (rate) · (time), but we made an assumption here that the velocity would be approximately constant over that one second. The first form of the answer is technically consistent, but unit conversions were included to make the answer more intuitive. Still, the gist of the method was to assume that the rate of position change (72 mile/hour), at the time that rate was sampled, would be approximately the rate for the entire one-second time interval in question.[4]

[4]The difference between *interpolation* and *approximation* is subtle but important. Interpolation refers to taking some data and using it to approximate where other points may lie, particularly but not limited to those in between the given data.

Example 5.3.2 *Suppose a manufacturer's research shows that the profit from making x of a particular item should be*

$$P(x) = -0.004x^3 + 10x^2 - 1000.$$

Suppose further that the manufacturer is initially planning on a production run of 100 items. How much more profit would he make if he produced 101 items instead?

Solution 1: If the model is correct, the actual extra profit from making that 101st item would be the difference in profit from making 101 items and the profit from making 100 items:

$$\begin{aligned} P(101) - P(100) &= \left[-0.004(101)^3 + 10(101)^2 - 1000\right] - \left[-0.004(100)^3 + 10(100)^2 - 1000\right] \\ &= 96888.796 - 95000 \\ &= 1888.796. \end{aligned}$$

Of course this should be rounded to hundredths of a dollar (that is, to the nearest cent), so according to this model the manufacturer would make and extra \$1888.80 from that 101st item.

Of course, the model itself is likely an approximation based upon research. The apparent precision of the expected extra profit is open to further scruitiny (as with any economic model). In any event, another approach, this time **definitely** *an approximation, is given below:*

Solution 2: Note that

$$\begin{aligned} P'(x) &= \frac{d}{dx}\left[-0.004x^3 + 10x^2 - 1000\right] \\ &= -0.012x^2 + 20x. \end{aligned}$$

Furthermore, $P'(100) = -0.012(100^2) + 20(100) = 1880$ *(dollars/item). In other words, the profit is changing at* \$1880/*item when the number of items is 100. We can use this to approximate that the next (101st) item will cause a growth in the profit of approximately* \$1880.

In the above example, note that the second method offered a very good approximation for the difference $P(101) - P(100)$, by considering how quickly $P(x)$ was changing with x (quantified by $P'(x)$) when $x = 100$, and—assuming that $P(x)$ continued to change at approximately that rate for x near 100—used that to approximate the actual change in the value of $P(x)$ as x changes from 100 to 101. This is in the same spirit as in the previous example (Example 5.3.1). Below are two reasons why one may wish to use the approximating method of Solution 2 instead of the more exact, Solution 1 method[5]:

- It was easier (and faster) to compute $P'(100\text{ item}) \cdot (1\text{ item})$ than to compute the difference $P(101\text{ item}) - P(100\text{ item})$. The former (approximation) was a second-degree polynomial with two terms, while the latter (and actual) was a third-degree polynomial with three terms, evaluated twice.

- The original model was only an approximation, so this approximation to an approximation might not have lost too much confidence in accuracy to be useful.

Interpolation is thus a kind of approximation, but there are others. For instance, in this Section 5.3 we instead use a function's output and derivative at a given input, and assume the derivative is approximately constant to approximate function values elsewhere. So we are not "connecting" data points—as we do when we interpolate—to predict others, but look at the function's properties at a single datum point to predict others. There is an entire literature on the topic of approximation theory, and here we have two basic examples: linear interpolation from two data, and linear approximation from one datum, but including information about the derivative there.

[5]Of course the reasons given here are usually balanced by fact that, if the exact answer is easily available, it may be far preferable to an approximation. We will discuss this further as we progress through this section.

Note that since we have an approximation for the difference $P(101) - P(100) \approx P'(100) \cdot 1$, we can write this as an approximation for $P(101)$, rather than the difference:

$$P(101) \approx P(100) + P'(100)(1),$$

where the trailing 1 represents how many units away from $x = 100$ we traveled to get to $x = 101$. If we strayed too far from $x = 100$, the profit change may stray far from $P'(100)$, and our approximation will be less accurate.

In the next example we will use a similar strategy which again lets us use a simple function to approximate a computationally more complicated one.

Example 5.3.3 *Suppose a laser at ground level points to the base of a building 300 feet away. If the laser beam is then turned so that it still points to the building, but with an angle of elevation of* $5°$*, then approximately how high on the building is the point illuminated by the beam?*

Solution: The height h on the building is a function of θ, the angle of elevation of the beam, that is, the angle formed by the beam and the horizontal. Since $h/(300\text{ foot}) = \tan\theta$, we can write

$$h(\theta) = 300\text{ foot} \cdot \tan\theta. \tag{5.9}$$

If we use radian measure for θ, then

$$h'(\theta) = 300\text{ foot}\sec^2\theta.$$

For $\theta = 0$, we have $h'(0) = 300\text{ foot}$. It is instructive to note that the units of $dh(t)/dt$ are formally foot/radian (though we often omit the ultimately dimensionless unit of radian). With this in mind, we can note that $5° = 5° \cdot \frac{\pi(\text{rad})}{180°} = \frac{\pi}{36}\text{ (radian)}$, and so

$$\begin{aligned} h\left(\frac{\pi}{36}\right) &\approx h(0) + h'(0)\frac{\pi}{36}\text{ (radian)} \\ &= 0 + 300\frac{\text{foot}}{\text{radian}} \cdot \frac{\pi}{36}\text{ (radian)} \\ &= \frac{300\pi}{36}\text{ foot} \\ &\approx 26.18\text{ foot}. \end{aligned}$$

The actual height of the laser beam is $300\text{ foot} \cdot \tan 5° \approx 26.25\text{ foot}$.

If we look at the above example closely, we see that replacing $5°$ with any angle θ in *radian* measure, then we can claim the approximation

$$h(\theta) \approx 300\text{ foot} \cdot \theta. \tag{5.10}$$

From the example we see that this approximation is very good for $\theta = 5° = \pi/36$ (radian). In fact it compares well even for larger θ, but certainly not for all θ. Clearly from (5.9), $h(\theta) = 300\text{ ft} \cdot \tan\theta \longrightarrow \infty$ as $\theta \to \frac{\pi}{2}^-$, which does not happen with our approximation (5.10). What happened is that the rate of change $dh/d\theta = 300\sec^2\theta$ did not stay at all constant, and in fact blew up also as $\theta \to \frac{\pi}{2}^-$.

5.3.2 Linear Approximation

In all these examples, we approximated a "future" measurement of a function based upon its presently known value at a particular point, and how fast it was changing at that point. For a function $f(x)$, the known value was at some $x = a$, where we had data on $f(a)$ and on how fast $f(x)$ changes with respect to x at $x = a$; that is, we knew $f(a)$ and $f'(a)$. Based upon these, we could find an approximation of $f(a + \Delta x)$ by thinking of Δx as a "run," with $f(a + \Delta x) - f(a)$ being the resulting "rise." Hence $(f(a + \Delta x) - f(a))/\Delta x$ is "rise/run," which is assumed to be approximately $f'(a)$, which—upon multiplying by Δx gives us:

$$f(a + \Delta x) - f(a) \approx f'(a)\Delta x.$$

Solving for $f(a + \Delta x)$, we would have

$$f(a + \Delta x) \approx f(a) + f'(a)\Delta x. \tag{5.11}$$

For a more useful formula, we now let $x = a + \Delta x$, so that $\Delta x = x - a$, and then (5.11) gives us the following:

Definition 5.3.1 *For a function $f(x)$, and a real number a for which $f(a)$, $f'(a)$ exist, the* **linear approximation of** $f(x)$ **at** a *is given by*

$$f(x) \approx f(a) + f'(a)(x - a). \tag{5.12}$$

Note that the right-hand side of (5.12) is exactly the expression for the tangent line to $y = f(x)$ at the point $x = a$, given by (4.11), that is, $y = f(a) + f'(a)(x - a)$.[6] We can write (5.12) in a colloquial way as follows: Where is f at x? The approximate answer is, where it was at a, plus how fast it was changing at a multiplied by how far we traveled (+/−) from a.

Example 5.3.4 *Use the linear approximation of $f(x) = \sqrt[3]{x}$ to approximate $\sqrt[3]{8.5}$.*
Solution: Here $f(x) = x^{1/3}$, and $f'(x) = \frac{1}{3}x^{-2/3}$. Using $a = 8$ we get

$$f(8) = 8^{1/3} = 2,$$
$$f'(8) = \frac{1}{3}(8)^{-2/3} = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}.$$

Thus, for x near 8, we have

$$f(x) \approx f(8) + f'(8)(x - 8), \text{ i.e.,}$$
$$f(x) \approx 2 + \frac{1}{12}(x - 8).$$

Using this we get

$$\begin{aligned}\sqrt[3]{8.5} &= f(8.5)\\ &\approx 2 + \frac{1}{12}(8.5 - 8)\\ &= 2 + \frac{1}{12} \cdot \frac{1}{2}\\ &= 2 + \frac{1}{24} = 49/24 = 2.041666666\cdots.\end{aligned}$$

Thus $\sqrt[3]{8.5} \approx 2.0417$.

[6]That the approximation technique should give rise to the tangent line should not be surprising, since $f'(a)$ measures the (instantaneous) rate of change of y with respect to x, as well as the slope of the tangent line at $x = a$. In our examples, we used respectively how position changed with time, how profit changed with the number of items produced, and how one leg of a right triangle changed with its opposite angle. The connection to the derivative was apparent in each, and the derivative is geometrically the slope of the curve, which also defines the slope of the tangent line.

The actual value of $\sqrt[3]{8.5}$ is approximately 2.04082755, so our linear approximation is accurate to three significant digits when $x = 8.5$. Any linear approximation is a statement regarding x close to the point $x = a$ (at which the linear approximation is just the tangent line to the graph). Such an approximation is likely to worsen in accuracy as we leave the immediate vicinity of $x = a$, although the degree to which this happens depends upon the function—in particular, how closely the graph follows the trend of the tangent line at $x = a$, as we move away from $x = a$.

Below is a list of values of $\sqrt[3]{x}$ as approximated by this method, and than as computed directly (first 8 digits shown).

$$\begin{array}{rr}
f(8) \approx 2 & f(8) = 2 \\
f(8.5) \approx 2.0416667 & f(8.5) = 2.0408275\cdots \\
f(9) \approx 2.0833333 & f(9) = 2.0800838\cdots \\
f(10) \approx 2.1666667 & f(10) = 2.1544346\cdots \\
f(11) \approx 2.2500000 & f(11) = 2.2239800\cdots \\
f(12) \approx 2.3333333 & f(12) = 2.2894284\cdots \\
f(13) \approx 2.4166667 & f(13) = 2.3513346\cdots \\
\vdots & \vdots \\
f(20) \approx 3.0000000 & f(20) = 2.7144176\cdots \\
f(30) \approx 3.8333333 & f(30) = 3.1072325\cdots \\
\vdots & \vdots \\
f(64) \approx 6.6666667 & f(64) = 4\cdots \\
f(1000) \approx 84.6666667 & f(1000) = 10.
\end{array}$$

We see that the approximation based upon the behavior at $x = 2$ (i.e., the linear approximation at $x = 2$) stays reasonably close to the actual values of $f(x)$ until we stray far from $x = 8$.[7] The actual graph of $f(x) = \sqrt[3]{x}$, together with the tangent line emanating from $(8, f(8))$ are graphed in Figure 5.4, page 391.

Again looking at Figure 5.4, we see that the tangent line at $x = 2$ does stay somewhat close to the curve for a while as x increases past $x = 2$. However, we see a different behavior as x moves left of $x = 2$. Consider the following comparisons of the linear approximation and actual

[7] Of course, "close" and "far" are subjective measures of proximity, and acceptable tolerances differ from context to context.

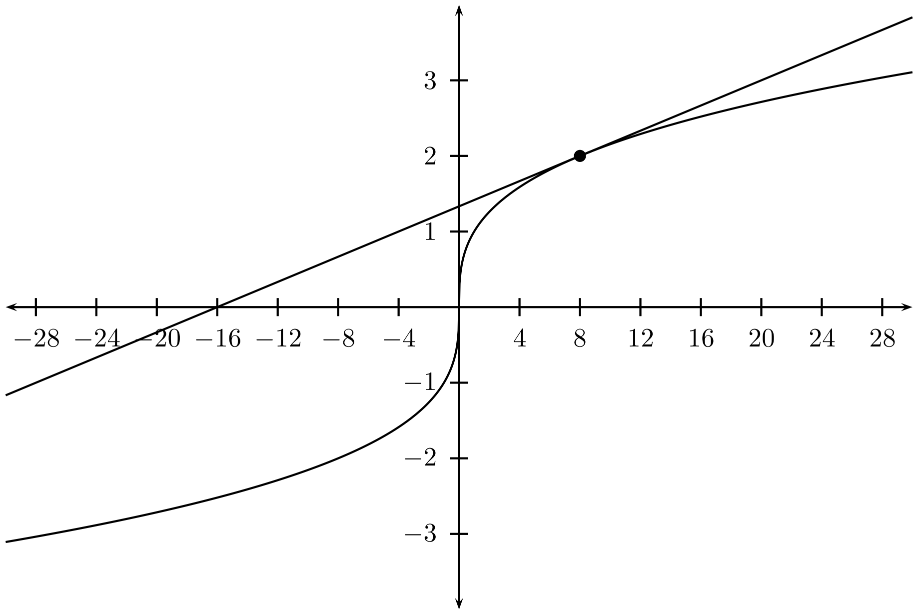


Figure 5.4: Partial graph of $f(x) = \sqrt[3]{x}$, along with the linear approximation (tangent line) at $x = 8$. The two graphs are very close to eachother near $x = 8$ (and coincide at $x = 8$), but part company as we stray farther from $x = 8$. They may (and in fact do) evenutually come together again, but that is coincidence, while the approximation is known to be useful near $x = 8$ by the nature of the tangent line.

value of $f(x)$ for $x < 2$:

$$\begin{aligned} f(7) &\approx 1.9166667 & f(7) &= 1.9129311\cdots \\ f(6) &\approx 1.8333333 & f(6) &= 1.8171205\cdots \\ f(5) &\approx 1.75 & f(5) &= 1.7099759\cdots \\ f(4) &\approx 1.6666667 & f(4) &= 1.5874010\cdots \\ f(3) &\approx 1.5833333 & f(3) &= 1.4422495\cdots \\ f(2) &\approx 1.5 & f(2) &= 1.2599210\cdots \\ f(1) &\approx 1.4166667 & f(1) &= 1 \\ f(0) &\approx 1.3333333 & f(0) &= 0 \\ &\vdots & &\vdots \\ f(-8) &\approx 0.6666666 & f(-8) &= -2. \end{aligned}$$

As we can see from the graph in Figure 5.4, the function and the tangent line are quite close when $|x-8|$ is small, but is unreliable as we stray from $x = 8$.

One of the most useful linear approximations in physics is used to approximate $\sin x$ for small $|x|$, i.e., when $|x-0|$ is small, i.e., when x is near zero:

Example 5.3.5 *Find the linear approximation for $f(x) = \sin x$ at $x = 0$.*

Solution: We will use the formula $f(x) \approx f(a) + f'(a)(x-a)$ with $a = 0$ and $f(x) = \sin x$. Now

$$\begin{aligned} f(0) &= \sin 0 = 0, \\ f'(0) &= \cos 0 = 1. \end{aligned}$$

With this data, (5.12) becomes

$$\begin{aligned} &f(x) \approx f(0) + f'(0)(x-0) = 0 + 1(x-0), \text{ i.e.,} \\ &f(x) \approx x. \end{aligned}$$

The graphs of $y = \sin x$ and its linear approximation at $x = 0$, namely $y = x$, are given in Figure 5.5. The approximation is very good for $|x| < 1$.[8]

5.3.3 Linear Approximations and Implicit Functions

Because we can find $\frac{dy}{dx}$ on implicit curves written as equations, we can find tangent lines and therefore linear approximations. In such a case it is not for y as a function of x, but rather for y as a local function on x. Still the method is valid. We offer two examples here.

Example 5.3.6 *Consider the curve $y^2 - x^2 = 9$. Approximate y as a function of x near $(4, -5)$ using a linear approximation.*

[8]Note that $|x| < 1$ in radians corresponds to, approximately $|x| < 57°$, but to use this to approximate $\sin 48°$, for instance, we need to convert back to radians:

$$\sin 48° = \sin \frac{48° \cdot \pi}{180°} = \sin \frac{48\pi}{180} \approx \frac{48\pi}{180} \approx 0.837758041.$$

The actual value of $\sin 48°$ is 0.7431448, when rounded to seven places.

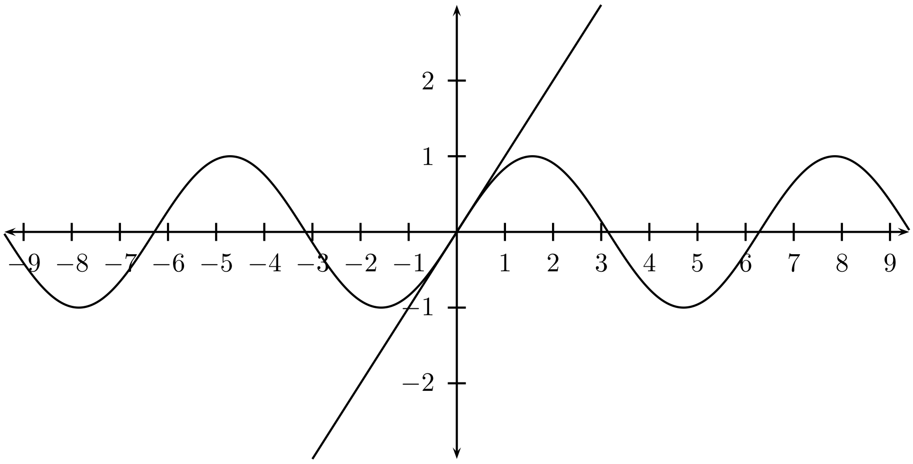


Figure 5.5: Partial graph of $f(x) = \sin x$, and the linear approximation at $x = 0$, which is $y = x$. Though not clear from the printed resolution here, the functions only coincide at $x = 0$. The proof of that fact is left as an exercise.

<u>Solution</u>: We use the usual implicit differentiation technique as in Section 4.5:

$$\begin{aligned} & y^2 - x^2 = 9 \\ \implies \quad & \frac{d}{dx}\left[y^2 - x^2\right] = \frac{d}{dx}[9] \\ \implies \quad & 2y \cdot \frac{dy}{dx} - 2x = 0 \\ \implies \quad & 2y\frac{dy}{dx} = 2x \\ \implies \quad & \frac{dy}{dx} = \frac{2x}{2y} = \frac{x}{y}. \end{aligned}$$

For the point $(4, -5)$—which a quick check shows is on the curve—we have slope

$$\left.\frac{dy}{dx}\right|_{(4,-5)} = \left.\frac{x}{y}\right|_{(4,-5)} = \frac{4}{-5} = -\frac{4}{5}.$$

The tangent line is given by

$$y = -5 - \frac{4}{5}(x - 4).$$

Thus, on the curve near $(4, -5)$, we have $y \approx -5 - \frac{4}{5}(x - 4)$.

This graph and the linear approximation at $(4, -5)$ is given in Figure 5.6.

Example 5.3.7 *Recall that in Example 4.5.5, on page 318 we had the implicit curve given by $5x + x^2 + y^2 + xy = \tan y$, and we found*

$$\frac{dy}{dx} = \frac{5 + 2x + y}{\sec^2 y - 2y - x},$$

and the tangent line to the curve at $(0, 0)$ had slope 5. Thus near $(0, 0)$, we can say that the local function is given by $y \approx 0 + 5(x - 0)$, or $y \approx 5x$, which is a huge advantage over trying to find an actual y for a given x near $x = 0$.

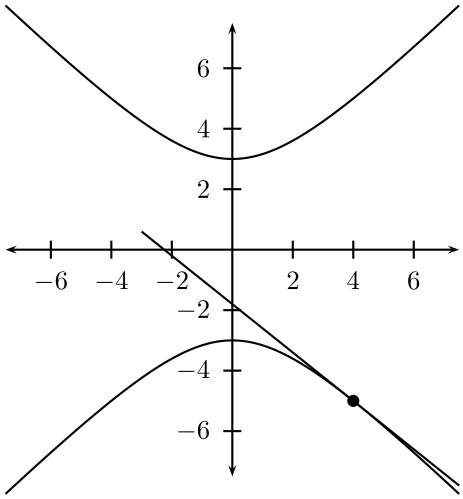


Figure 5.6: Graph for Example 5.3.6, showing a partial graph of the curve $y^2 - x^2 = 9$ and the tangent line at $(4, -5)$, which is also the linear approximation to the local (implicit) function defined near there.

5.3.4 Differentials

Below we define differentials. Eventually we will give numerical and geometric meaning to all of the terms in the definition, but first we define them only formally:

Definition 5.3.2 *Given a function $f(x)$, the* **differential** *of $f(x)$ is given by*

$$df(x) = f'(x)\,dx, \tag{5.13}$$

dx being the **differential of** *x, and where the prime, " $'$ " represents that the derivative is taken with respect to the underlying variable, which here is x.*

This is consistent with our previous use of Leibniz notation:

- $dx = d(x) = (x)'dx = 1 \cdot dx = dx$, as we would hope, and
- $\dfrac{d\,f(x)}{dx} = f'(x) \iff d\,f(x) = f'(x)\,dx$, at least formally, where we (again formally) multiplied both sides by dx.

Now we look at some quick computations which follow from this definition:

- $d\sin x = \cos x\,dx$,
- $dx^2 = 2x\,dx$,
- $d\sqrt{x} = \dfrac{1}{2\sqrt{x}}\,dx = \dfrac{dx}{2\sqrt{x}}$,
- $d\csc x = -\csc x\cot x\,dx$,
- $d\left[\dfrac{x}{x+1}\right] = \dfrac{d}{dx}\left[\dfrac{x}{x+1}\right]dx = \left[\dfrac{(x+1)\frac{d}{dx}(x) - (x)\frac{d}{dx}(x+1)}{(x+1)^2}\right]dx$
  $= \dfrac{(x+1)(1)-(x)(1)}{(x+1)^2}\,dx = \dfrac{1}{(x+1)^2}\,dx = \dfrac{dx}{(x+1)^2}$,

- $d(x\tan x) = \frac{d}{dx}[x\tan x]\,dx = \left[x\cdot\frac{d\,\tan x}{dx} + \tan x\cdot\frac{d(x)}{dx}\right]\,dx = (x\sec^2 x + \tan x)\,dx.$

All of these become old-fashioned derivative problems if we divide these equations by dx. We can fashion differential versions of all of our rules by taking the derivative rules and multiplying both sides by dx. In particular, there are product, quotient and chain rules:

$$d(uv) = u\,dv + v\,du, \tag{5.14}$$

$$d\left[\frac{u}{v}\right] = \frac{v\,du - u\,dv}{v^2}, \tag{5.15}$$

$$df(u(x)) = f'(u(x))u'(x)dx. \tag{5.16}$$

With these we can compute many more differentials directly:

- $d(x\tan x) = x\,d\tan x + \tan x\cdot dx = x\sec^2 x\,dx + \tan x\,dx$, which is exactly the result of our efforts just above to compute this differential.
- $d\sin x^2 = \cos x^2\cdot 2x\,dx = 2x\cos x^2\,dx.$

In fact notice that (5.16) is completely consistent with the idea that $df(u) = f'(u)\,du$:

- By definition: $df(u) = f'(u)\,du$.
- If it happens that u is a function of x, i.e., $u = u(x)$, then

$$df(u) = df(u(x)) = \frac{df(u(x))}{dx}\,dx = \frac{df(u(x))}{du(x)}\cdot\frac{du(x)}{dx}\,dx = \underbrace{f'(u(x))}_{f'(u)}\underbrace{u'(x)\,dx}_{du} = f'(u)\,du.$$

- We now have two methods for computing quantities such as $\frac{d\sin x^2}{dx^2}$:
  1. considering x^2 as a variable in its own right:
  $$\frac{d\sin x^2}{x^2} = \cos x^2;$$
  2. using the definition of differentials (Definition 5.3.2, page 394):
  $$\frac{d\sin x^2}{x^2} = \frac{(\sin x^2)'\,dx}{(x^2)'\,dx} = \frac{\cos x^2\cdot 2x\,dx}{2x\,dx} = \cos x^2.$$

This again shows the robustness of the Leibniz notation, with differentials as well as derivatives. Computations of differentials become ubiquitous as we develop integration techniques in later sections.

So far we have only looked at these differentials formally. Now we will emphasize that these differentials can in fact be interpretted *numerically.*

Recall that $\frac{df(x)}{dx}$ measures how $f(x)$ changes as x changes. More precisely, the fraction $\frac{df(x)}{dx}$ gives us the instantaneous rate of change in $f(x)$ as x changes, at a particular value of x. (This is akin to $\frac{ds(t)}{dt}$ measuring velocity—that is, how position $s(t)$ is changing as t is changing—at a particular value of t.) Note also that $\frac{df(x)}{dx}$ can be interpretted as the slope of the graph of $f(x)$ at the particular value x. Hence $\frac{df(x)}{dx}$ represents an instantaneous "rise/run."

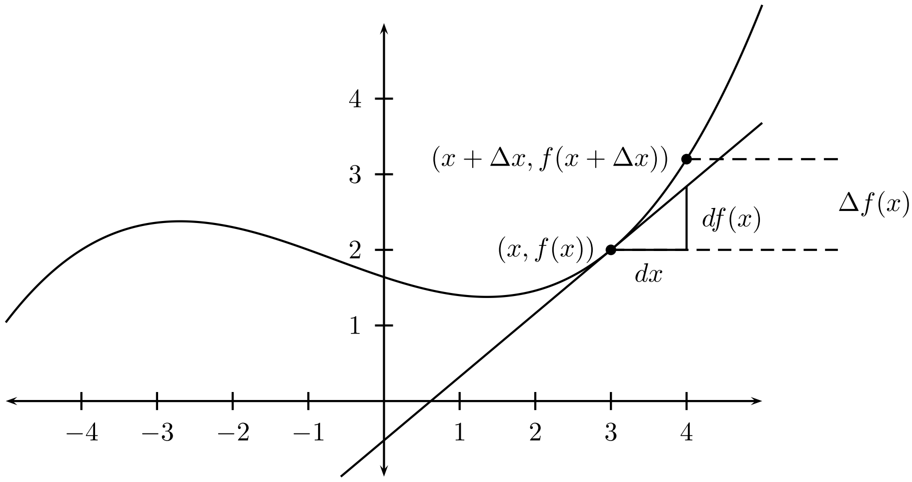


Figure 5.7: Illustration of the geometric meaning of $\frac{df(x)}{dx}$, giving further, geometric meaning to both dx and $df(x)$. For any value x in the domain, if $f'(x)$ is defined, we have a tangent line of slope $f'(x) = \frac{df(x)}{dx}$. Now $df(x) = f'(x)dx$, where dx can represent a "run" and $df(x)$ the resultant "rise" along the tangent line. Note that $df(x)$ is an approximation of the "rise" (not necessarily positive) in the actual function, as the input changes from x to $x + dx$. Moreover that rise, being what is the rise in the tangent line instead of the function itself, is the same as what is given by the linear approximation.

Now we will let dx represent a "run," i.e., a change in x from a fixed value of x, while $df(x)$ will be the resultant "rise," *at the rate of* $\frac{df(x)}{dx}$, i.e., $df(x)$ will represent the "rise" along the tangent line which ran through $(x, f(x))$. This is illustrated in Figure 5.7.

The same justification for using linear approximations for functions allows us to use $df(x)$ to approximate an actual change in the function $f(x)$, as we perturb the x-value by a small quantity dx. In fact, the actual change, the linear approximation, and the differentials are all related. If we call the perturbation in x by both names Δx and dx, we get:

$$\begin{aligned} f(x + \Delta x) &= f(x) + \frac{f(x + \Delta x) - f(x)}{\Delta x} \cdot \Delta x \\ &\approx f(x) + f'(x)\Delta x \\ &= f(x) + \frac{df(x)}{dx} \cdot dx \\ &= f(x) + df(x). \end{aligned} \tag{5.17}$$

This reflects exactly what occurs with the linear approximation:

$$f(x) \approx f(a) + \underbrace{f'(a)(x - a)}_{\left.\frac{df(x)}{dx}\right|_{x=a} \cdot \, dx}, \tag{5.18}$$

where the part of $x + \Delta x$ in (5.17) is played by x in (5.18), the part of x is played by a, and Δx is played by $(x - a)$. Furthermore, $f'(a)(x - a)$ represents $df(x)$ evalutated at $x = a$ with $dx = (x - a)$, so the right-hand side of (5.18) $f(a)$ plus the approximate purturbation of $f(x)$ from $f(a)$ as x strays from a.

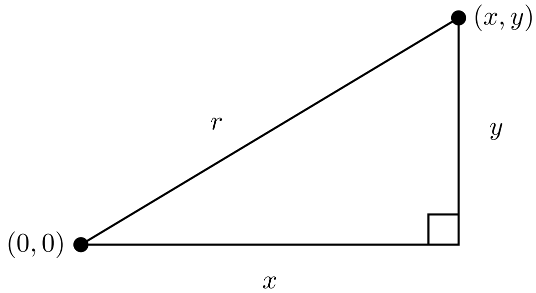


Figure 5.8: If x, y and r are the legs and hypotenuse of a right triangle, then $x^2+y^2=r^2$ according to the Pythagorean Theorem. This implies the $2x\frac{dx}{dt}+2y\frac{dy}{dt}=2r\frac{dr}{dr}$, if x, y and r are functions of t. In fact, if we allow x and y to be respectively horizontal and vertical *displacements* from the origin $(0,0)$, with r the distance from $(0,0)$, the Pythagorean Theorem still holds and so does the equation relating the rates.

5.4 Newton's Approximation Method

5.5 Chain Rule III: Related Rates

In many situation we have variables which are constrained by a natural relationship or equation. For instance, if we have three sides of a right triangle, say x, y and a hypotenuse r, then

$$x^2+y^2=r^2 \tag{5.19}$$

is the well-known Pythagorean Theorem. If x, y, r are all functions of another variable, say t, then as in the previous section we have the left-hand and right-hand sides of Equation 5.19 represent the same function of t, so we can differentiate with respect to t:

$$\begin{aligned} & \frac{d}{dt}\left[x^2+y^2\right]=\frac{d}{dt}\left[r^2\right] \\ \implies \qquad & 2x\frac{dx}{dt}+2y\frac{dy}{dt}=2r\frac{dr}{dt}. \end{aligned} \tag{5.20}$$

This followed from the chain rule.[9] Thus an equation relating variables holding true forces another equation relating the rates of change of these variables to hold true, hence the title of this section.

Example 5.5.1 *Using the illustration in Figure 5.8, suppose at some particular time t_0 we have $x=30\,\text{m}$ and $y=40\,\text{m}$, but x is decreasing at $20\,\text{m/sec}$, and y is increasing at $10\,\text{m/sec}$. How is r changing at that time?*

To answer this we should first point out that

$$\frac{dx}{dt}=-20\,\text{m/sec}, \qquad\qquad \frac{dy}{dt}=10\,\text{m/sec}.$$

What we want is $\frac{dr}{dt}$. What we know are x, y, $\frac{dx}{dt}$, and $\frac{dy}{dt}$. In order to solve for $\frac{dr}{dt}$ in (5.20), we also need r. But that is easy enough, since

$$r^2=x^2+y^2=(30\,m)^2+(40\,m)^2=10,000\,m^2 \implies r=50\,m.$$

[9]Some texts also call this process *implicit differentiation*, mainly because it resembles the work we did in Section 4.5. Related Rates and implicit differentiation are both really just further applications of the chain rule.

(We actually have $r = \pm 50\,\text{m}$ from the algebra, but here we do assume r is a distance and is therefore nonnegative.) Entering these numbers into (5.20) gives

$$2(30\,\text{m})(-20\,\text{m}/\,\text{sec}) + 2(40\,\text{m})(10\,\text{m}/\,\text{sec}) = 2(100\,\text{m})\frac{dr}{dt},$$

so

$$-400\,\text{m}^2/\,\text{sec} = 200\,\text{m}\frac{dr}{dt} \implies \frac{dr}{dt} = -2\,\text{m}/\,\text{sec}\,.$$

Hence the distance to the origin is decreasing at $2\,\text{m}/\,\text{sec}$.

Actually, in the above example we could have avoided the need to compute r if we had solved the algebraic equation for r (recalling that r here is nonnegative):

$$r = \sqrt{x^2 + y^2}.$$

Taking time derivatives now gives us

$$\begin{aligned}\frac{dr}{dt} &= \frac{1}{2\sqrt{x^2+y^2}}\frac{d}{dt}\left(x^2+y^2\right)\\ &= \frac{1}{\sqrt{x^2+y^2}}\left(2x\frac{dx}{dt} + 2y\frac{dy}{dt}\right)\\ &= \frac{1}{\sqrt{x^2+y^2}}\left(x\frac{dx}{dt} + y\frac{dy}{dt}\right).\end{aligned}$$

It should be noted that the first approach had the advantage of easier differentiation, which we paid for by having to solve for r in the algebraic equation before inserting that value into the related rate equation. This second approach puts more effort into preliminary algebraic work and the differentiation step, and thus does not require the value of r because of the simplicity of the left-hand side. This is because whenever a variable occurs only as a single term in the first power, that variable does not appear after differentiation (though its rate, i.e., derivative still does).

Example 5.5.2 *(Ideal Gas)*