ROLLE'S, MEAN-VALUE, INCREASE/DECREASE, EXTREME VALUES

MATH 152, SECTION 55 (VIPUL NAIK)

Corresponding material in the book: Sections 4.1-4.4.

Difficulty level: Moderate to hard. While most of these are ideas you have probably seen at the AP level or equivalent, our treatment of the topics will be somewhat more thorough. Also, this is extemely important as preparation for the process of graphing a function, which in turn is very important as a general tool for understanding all kinds of functions.

What students should definitely get: The statements of Rolle's theorem and the mean value theorem. The relationship between the signs of one-sided derivatives and whether the function value at a point is greater or less than the function value to its immediate left or right. The notions of local maximum, local minimum, point of increase, point of decrease. The definition of critical point. The first derivative test and second derivative test. The procedure for determining absolute maxima and minima.

What students should hopefully get: The distinction between being positive and being nonnegative; similarly, the distinction between being negative and being nonpositive. In particular, the fact that even when difference quotients are strictly positive, the derivative obtained as the limit may be zero. The conceptual distinction between local extreme values (a local condition) and absolute extreme values.

Executive summary

Words...

(1) If a function f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), and $f(a) = f(b) = 0$, then there exists $c \in (a, b)$ such that $f'(c) = 0$. This is called *Rolle's theorem* and is a consequence of the extreme-value theorem.

(2) If a function f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there exists $c \in (a, b)$ such that $f'(c)$ is the difference quotient $(f(b) - f(a))/(b - a)$. This result is called the *mean-value theorem*. Geometrically, it says that for any chord, there is a parallel tangent. Another way of thinking about it is that every difference quotient is equal to a derivative at some intermediate point.

(3) If f is a function and c is a point such that $f(c) \geq f(x)$ for x to the immediate left of c, we say that c is a local maximum from the left. In this case, the left-hand derivative of f at c, if it exists, is greater than or equal to zero. This is because the difference quotient is greater than or equal to zero. Local maximum from the right implies that the right-hand derivative (if it exists) is ≤ 0, local minimum from the left implies that the left-hand derivative (if it exists) is ≤ 0, and local minimum from the right implies that the right-hand derivative (if it exists) is ≥ 0. Even in the case of *strict* local maxima and minima, we still need to retain the equality sign on the derivative because it occurs as a *limit* and a limit of positive numbers can still be zero.

(4) If c is a point where f attains a local maximum (i.e., $f(c) \geq f(x)$ for all x close enough to c on both sides), then $f'(c)$, if it exists, is equal to zero. Similarly for local minimum.

(5) A *critical point* for a function is a point where either the function is not differentiable or the derivative is zero. All local maxima and local minima must occur at critical points.

(6) If $f'(x) > 0$ for all x in the open interval (a, b), f is increasing on (a, b). Further, if f is one-sided continuous at the endpoint a and/or the endpoint b, then f is increasing on the interval including that endpoint. Similarly, $f'(x) < 0$ implies f decreasing.

(7) If $f'(x) > 0$ everywhere except possibly at some isolated points (so that they don't cluster around any point) where f is still continuous, then f is increasing everywhere.

(8) If $f'(x) = 0$ on an open interval, f is constant on that interval, and it takes the same constant value at an endpoint where it's continuous from the appropriate side.

(9) If f and g are two functions that are both continuous on an interval I and have the same derivative on the interior of I, then $f - g$ is a constant function.

(10) There is a *first derivative test* which provides a sufficient (though not necessary) condition for a local extreme value: it says that if the first derivative is nonnegative (respectively positive) on the immediate left of a critical point, that gives a strict local maximum (respectively local maximum) from the left. If the first derivative is negative on the immediate left, we get a strict local minimum from the left. If the first derivative is positive on the immediate right, we get a strict local minimum from the right, and if it is negative on the immediate right, we get a strict local maximum from the right.

The first derivative test is similar to the corresponding "one-sided derivative" test, but is somewhat stronger for a variety of situations because in many cases, one-sided derivatives are zero, which is inconclusive, whereas the first derivative test fails us more rarely.

(11) The second derivative test states that if f has a critical point c where it is twice differentiable, then $f''(c) > 0$ implies that f has a local minimum at c, and $f''(c) < 0$ implies that f has a local maximum at c.

(12) There are also higher derivative tests that work for critical points c where $f'(c) = 0$. These work as follows: we look for the smallest k such that $f^{(k)}(c) \neq 0$. If this k is even, then f has a local extreme value at c, and the nature (max versus min) depends on the sign of $f^{(k)}(c)$ (max if negative, min if positive). If k is odd, then we have what we'll see soon is a point of inflection.

(13) To determine absolute maxima/minima, the candidates are: points of discontinuity, boundary points of domain (whether included in domain or outside the domain; if the latter, then limiting), critical points (derivative zero or undefined), and limiting cases at $\pm\infty$.

(14) To determine absolute maxima and absolute minima, find all candidates (discontinuity, endpoints, limiting cases, boundary points), evaluate at each, and compare. Note that any absolute maximum must arise as a local or endpoint maximum. However, instead of first determining which critical points give local maxima by the derivative tests, we can straightaway compute values everywhere and compare, if our interest is solely in finding the absolute maximum and minimum.

Actions... (think of examples that you've done)

(1) Rolle's theorem, along with the more sophisticated formulations involving increasing/decreasing, tell us that there is an intimate relationship between the zeros of a function and the zeros of its derivative. Specifically, between any two zeros of the function, there is a zero of its derivative. Thus, if a function has r zeros, the derivative has at least $r - 1$ zeros, with at least one zero between any two consecutive zeros of f.

(2) The more sophisticated version tells us that between any two zeros of a differentiable function, the function must attain a local maximum or local minimum. So, if the function is increasing everywhere or decreasing everywhere, there is at most one zero.

(3) The mean-value theorem allows us to use bounds on the derivative of a function to bound the overall variation, or change, in the function. This is because if the derivative cannot exceed some value, then the difference quotient also cannot exceed that value, which means that the function cannot change too quickly on average.

(4) To determine regions where a function is increasing and decreasing, we find the derivative and determine regions where the derivative is positive, zero, and negative.

(5) To determine all the local maxima and local minima of a function, find all the critical points. To find the critical points, solve $f' = 0$ and also consider, as possible candidates, all the points where the function changes definition. *Although a point where the function changes definition need not be a critical point, it is a very likely candidate.*

1. Rolle's theorem and mean-value theorem

1.1. **Rolle's theorem.** Rolle's theorem states that if f is a function defined on a closed interval $[a, b]$ such that the following three conditions hold: (i) f is continuous on $[a, b]$ (ii) f is differentiable on the open interval (a, b) (iii) $f(a) = f(b) = 0$, then there exists $c \in (a, b)$ such that $f'(c) = 0$. (It turns out that the condition that both $f(a)$ and $f(b)$ be equal to *zero* is not necessary – we can weaken it to simply requiring that $f(a)$ equal $f(b)$. The version stated in the book requires them both to be zero).

Now, I'll give you a rough sketch of the proof of Rolle's theorem. One possibility is that f is a constant function, in which case $f'(c) = 0$ for all $c \in (a, b)$. If f is nonconstant, then by the extreme-value theorem, f is either bigger than zero somewhere or smaller than zero somewhere. Assume the former – a similar proof applies for the latter assumption. In this case, f attains a maximum at some point in (a, b). At this point, if we try to calculate the left-hand derivative, we see that the left-hand derivative is greater than or equal to zero. And if we try to calculate the right-hand derivative, we see that the right-hand derivative is less than or equal to zero. Because the function is differentiable at the point, both the left-hand derivative and the right-hand derivative must be equal, which means that they must both be equal to zero.

Now, the crucial point here is understanding *why* the derivative of a function should be zero at a point where it is maximum. And this is very important for some of the stuff we'll be seeing in the near future. So let's understand more clearly what's happening.

At a point c where f attains a maximum, two things are happening. First, $f(c) \geq f(x)$ for $x < c$. This forces that the difference quotient that we form between c and x for any $x < c$ is nonnegative. Hence, the left-hand derivative is the limit of some expression that is nonnegative, so the left-hand derivative itself is nonnegative.

What happens to the right-hand derivative? Well, in this case, $f(x) - f(c)$ is zero or negative, but $x - c$ is positive, so the difference quotient is nonpositive, so the limit, which is the right-hand derivative, is nonpositive. So we have a situation where the left-hand derivative is nonnegative (i.e., positive or zero) and the right-hand derivative is nonpositive (i.e., negative or zero).

Now, for the function to be differentiable, the left-hand derivative and right-hand derivative must be equal, so the derivative must be equal to zero.

Here's another way of thinking about this. Up until the point where the maximum is achieved, the function must be, at least roughly speaking, going up. (This is not correct strictly speaking, but is useful at least in simple cases). And then, immediately after that point, the function must be, at least roughly speaking, going down. So, at that point, it changes from a *going up* to a *going down* function, hence at the point it is oing neither up nor down, so the derivative is zero.

Why is differentiability so important? Well, think of what might happen for a function that has one-sided derivatives but isn't differentiable. In that case, it could have a *sharp peak* – it increases in a straight line, and then takes a turn and starts decreasing in a straight line.

Also, note that differentiability *at the endpoints* is not necessary. So, Rolle's theorem applies for instance to the function $f(x) := \sqrt{1 - x^2}$ on the interval $[-1, 1]$, even though that function is not differentiable *at* the two points -1 and 1.

1.2. **Mean-value theorem.** Here's what the mean-value theorem states. It states that if f is a function that is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there exists $c \in (a, b)$ such that:

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

In other words, given any two points such that the function is continuous on the closed interval between those two points and differentiable on the open interval, then there is a point in the open interval at which the derivative at the interior point equals the difference quotient between the two endpoints.

In other words, there is a point on the graph in between these two points such that the tangent line at the point is *parallel* to the secant line, or chord, joining the points $(a, f(a))$ and $(b, f(b))$.

Note that this result has a somewhat similar flavor to the intermediate-value theorem, but it is a different result.

Please see the book for a description of how the mean-value theorem can be derived from Rolle's theorem.

Aside: The mean-value theorem can be used to prove Darboux's theorem. Recall that the derivative of a differentiable function on an interval need not be continuous on that same interval. However, it comes very close to being continuous if it is defined everywhere on the interval. Specifically, the derivative satisfies the intermediate value property, and hence all its discontinuities must be of the oscillatory kind.

This result is called Darboux's theorem. Although the result is not part of the syllabus, it can be deduced with a little bit of work from the mean-value theorem.

In fact, there are many similar results about derivatives that would become easy if we assumed that the derivative is continuous, but are true even in general, and the proofs of most of these results relies on the mean-value theorem. Since we're not focused on proving theorems, we will not be talking a lot about the mean-value theorem explicitly, but you should keep in mind that it is at the back of a lot of what we do.

2. Local increase and decrease behavior

We will now try to understand, very clearly, the relationship between the *sign of the derivative* and the *behavior of the function near a point.*

2.1. **Larger than stuff on the left.** Suppose c is a point and $a < c$ such that $f(x) \leq f(c)$ for all $x \in (a, c)$. In other words, c is a *local maximum from the left.* What do I mean by that? I mean that $f(c)$ is larger than or equal to f of the stuff on the *immediate* left of it. That doesn't mean that $f(c)$ is a maximum over the entire domain of f – it just means it is greater than or equal to stuff on the immediate left.

Now, we claim that, if the left-hand derivative of f at c exists, then it is greater than or equal to 0. How do we work that out? The left-hand derivative is the limit of the difference quotient:

$$\frac{f(x) - f(c)}{x - c}$$

where $x \to c^-$. Note that for x close enough to c, (i.e., $a < x < c$), the numerator is negative or zero, and the denominator is negative, so the difference quotient is zero or positive. Thus, the limit of this, if it exists, is zero or positive.

There are three other cases. Let's just summarize the four cases:

(1) If c is a point that is a local maximum from the left for f, then the left-hand derivative of f at c, if it exists, is zero or positive.
(2) If c is a point that is a local maximum from the right for f, then the right-hand derivative of f at c, if it exists, is zero or negative.
(3) If c is a point that is a local minimum from the left for f, then the left-hand derivative of f at c, if it exists, is zero or negative.
(4) If c is a point that is a local minimum from the right for f, then the right-hand derivative of f at c, if it exists, is zero or positive.

2.2. **Strict maxima and minima.** We said that for a function f, a point c is a *local maximum from the left* if there exists $a < c$ such that $f(x) \leq f(c)$ for all $x \in (a, c)$. Now, this definition also includes the possibility that the function is constant just before c.

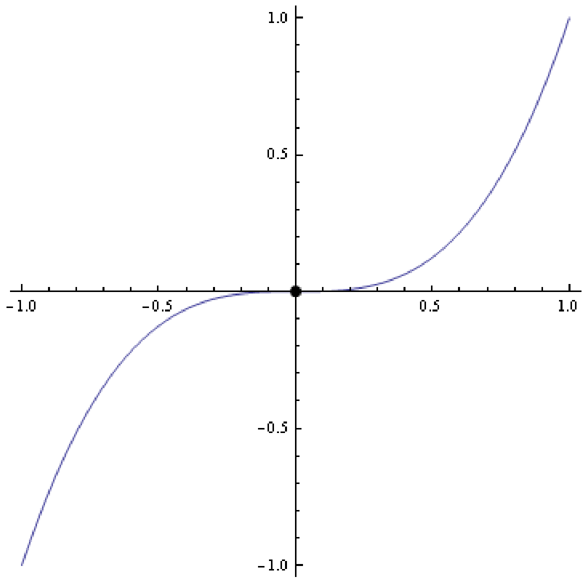
A related notion is that of a *strict local maximum from the left*, which means that there exists $a < c$ such that $f(x) < f(c)$ for all $x \in (a, c)$. In other words, $f(c)$ is *strictly bigger* than $f(x)$ for x to the immediate left of c.

Similarly, we can define the notions of strict local maximum from the right, strict local minimum from the left, and strict local minimum from the right.

2.3. **Does strict maximum/minimum from the left/right tell us more?** Recall that if c is a point that is a local maximum from the left for f, then the left-hand derivative of f at c, if it exists, is greater than or equal to zero. What if c is a point that is a strict local maximum from the left for f? Can we say something more about the left-hand derivative of f at c?

The first thing you might intuitively expect is that that left-hand derivative of f at c should now not just be greater than or equal to zero, it should be strictly greater than zero. But you would be wrong.

It *is* true that if c is a strict local maximum from the left for f, then the difference quotients, as $x \to c^-$, are all positive. However, the *limit* of these difference quotients could still be zero. Another way of thinking about this is that even if the function is increasing up to the point c, it may happen that the rate of increase is leveling off to 0. An example is the function x^3 at the point 0: 0 is a strict local maximum from the left, but the derivative at 0 is 0. Here's a picture:



Later, we will understand this situation more carefully and it will turn out that we are dealing (in this case) with what is called a *point of inflection*.

2.4. **Minimum, maximum from both sides.** So we have some sign information about the derivative closely related to how the function at the point compares with the value of the function at nearby points. Maximum from the left means left-hand derivative is nonnegative, maximum from the right means right-hand derivative is nonpositive, minimum from the left means left-hand derivative is nonpositive, minimum from the right means right-hand derivative is nonnegative.

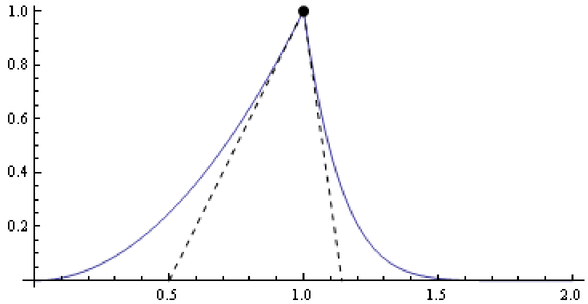
So, let's piece these together:

(1) A *local maximum* for the function f is a point c such that $f(c)$ is the maximum possible value for $f(x)$ in an open interval containing c. Thus, a point of local maximum for f is a point that is both a local maximum from the left and a local maximum from the right. A *strict local maximum* for the function f is a point c such that $f(c)$ is strictly greater than $f(x)$ for all x in some open interval containing c.
(2) A *local minimum* for the function f is a point c such that $f(c)$ is the minimum possible value for $f(x)$ in an open interval containing c. Thus, a point of local minimum for f is a point that is both a local minimum from the left and a local minimum from the right. A *strict local minimum* for the function f is a point c such that $f(c)$ is strictly smaller than $f(x)$ for all x in some open interval containing c.

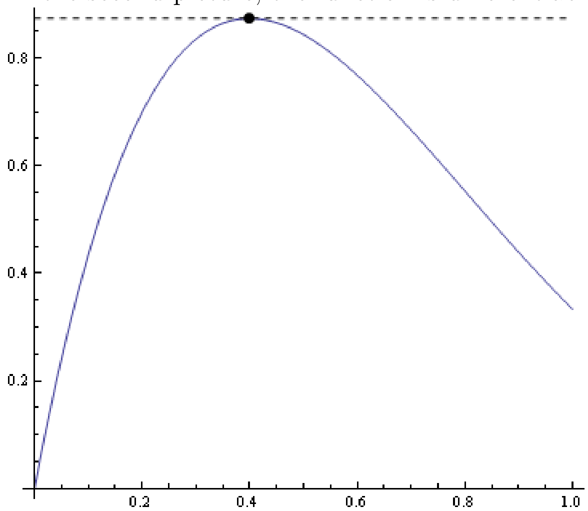
What can we say about local maxima and local minima? We can say the following:

(1) At a local maximum, the left-hand derivative (if it exists) is greater than or equal to zero, and the right-hand derivative (if it exists) is less than or equal to zero. Thus, *if* the derivative exists at a point of local maximum, it *equals zero*. The same applies to strict local maxima.
(2) At a local minimum, the left-hand derivative (if it exists) is less than or equal to zero, and the right-hand derivative (if it exists) is greater than or equal to zero. Thus, *if* the derivative exists at a point of local minimum, it *equals zero*. The same applies to strict local minima.

Below are two pictures depicting points of local maximum. In the first picture, the left-hand derivative is positive, the right-hand derivative is negative, and the function is not differentiable at the point of local maximum.



In the second picture, the function is differentiable, and the derivative is zero.



2.5. **Maximum from the left, minimum from the right.** Suppose c is a point such that it is a local maximum from the left for f and is a local minimum from the right for f. This means that $f(c)$ is greater than or equal to $f(x)$ for x to the immediate left of c, and $f(c)$ is less than or equal to $f(x)$ for x to the immediate right of c. In this case, we say that f is *non-decreasing* at the point c.

In other words, f at c is bigger than or equal to what it is on the left and smaller than or equal to what it is on the right. Well, in this case, the left-hand derivative is greater than or equal to zero and the right-hand derivative is greater than or equal to zero. Thus, if $f'(c)$ exists, we have $f'(c) \geq 0$.

Now consider the case where c is a point that is a local minimum from the left for f and is a local maximum from the right for f. This means that $f(c)$ is less than or equal to $f(x)$ for x to the immediate left of c and greater than or equal to $f(x)$ for x to the immediate right of c. In this case, we say that f is *non-increasing* at the point c.

In other words, f at c is smaller than what it is on the right and larger than what it is on the left. Well, in this case, the left-hand derivative is less than or equal to zero and the right-hand derivative is less than or equal to zero. Thus, if $f'(c)$ exists, we have $f'(c) \leq 0$.

2.6. **Introducing strictness.** We said that f is *non-decreasing* at the point c if $f(c) \geq f(x)$ for x just to the left of c and $f(c) \leq f(x)$ for x just to the right of c. We now consider the *strict* version of this concept. We say that f is *increasing* at the point c if there is an open interval (a, b) containing c such that, for $x \in (a, b)$, $f(x) < f(c)$ if $x < c$ and $f(x) > f(c)$ if $x > c$. In other words, c is a strict local maximum from the left and a strict local minimum from the right.

Well, what can we say about the derivative at a point where the function is increasing, rather than just non-decreasing? We already know that $f'(c)$, if it exists, is greater than or equal to zero, but we might hope to say that the derivative $f'(c)$ is strictly greater than zero. Unfortunately, that is not true.

In other words, a function could be increasing at the point c, in the sense that it is strictly increasing, but still have derivative 0. For instance, consider the function $f(x) := x^3$. This is increasing everywhere, but at the point zero, its derivative is zero.

How can a function be increasing at a point even though its derivative is zero? Well, what happens is that the derivative was positive before the point, is positive just after the point, and becomes zero just momentarily. Alternatively, if you think in terms of the derivative as a limit of difference quotients, all the difference quotients are positive, but the limit is still zero because they get smaller and smaller in magnitude as you come closer and closer to the point. Another way of thinking of this is that you reduce your car's speed to zero for the split second that you cross the STOP line, so as to cimply with the letter of the law without actually stopping for any interval of time.

Similarly, we can define the notion of a function f being *decreasing* at a point c. This means that $f(c) < f(x)$ for x to the immediate left of c and $f(c) > f(x)$ for x to the immediate right of c. As in the previous case, we can deduce that $f'(c)$, if it exists, is less than or equal to zero, but it could very well happen that $f'(c) = 0$. An example is $f(x) := -x^3$, at the point $x = 0$.

2.7. **Increasing functions and sign of derivative.** Here's what we did. We first did separate analyses for what we can conclude about the left-hand derivative and the right-hand derivative of a function based on how the value of the function at the point compares with the value of the function at points to its immediate left. We used this to come to some conclusions about the nature of the derivative of a function (if it exists) at points of local maxima, local minima, and points where the function is nondecreasing and nonincreasing. Let's now discuss a converse result.

So far, we have used information about the nature of changes of the function to deduce information about the sign of the derivative. Now, we want to go the other way around: use information about the sign of the derivative to deduce information about the behavior of the function. And this is particularly useful because now that we have a huge toolkit, we can differentiate practically any function that we can write down. This means that even for functions that we have no idea how to visualize, we can formally differentiate them and work with the derivative. Thus, if we can relate information about the derivative to information about the function, we are in good shape.

Remember what we said: if a function is increasing, it is nondecreasing, and if it is nondecreasing, then the derivative is greater than or equal to zero. Now, a converse for this would mean some condition on the derivative telling us whether the function is increasing.

Unfortunately, the derivative being zero is very inconclusive. The function could be constant, it could be a local maximum, it could be a local minimum, it could be increasing, or it could be decreasing. However, it turns out that if the derivative is *strictly* positive, then we can conclude that the function is increasing.

Specifically, we have the following chain of implications for a function f defined around a point c and differentiable at c:

$f'(c) > 0 \implies f$ is increasing at $c \implies f$ is nondecreasing at $c \implies f'(c) \geq 0$

And each of these implications is strict, in the sense that you cannot proceed backwards with any of them, becausethere are counterexamples to each possible reverse implication.

Similarly, for a function f defined around a point c and differentiable at c:

$f'(c) < 0 \implies f$ is decreasing at $c \implies f$ is nonincreasing at $c \implies f'(c) \leq 0$

2.8. **Increasing and decreasing functions.** A function f is said to be increasing an an interval I (which may be open, closed, half-open, half-closed, or stretching to infinity) if for any $x_1 < x_2$, with both x_1 and x_2 in I, we have $f(x_1) < f(x_2)$. In other words, the larger the input, the larger the output.

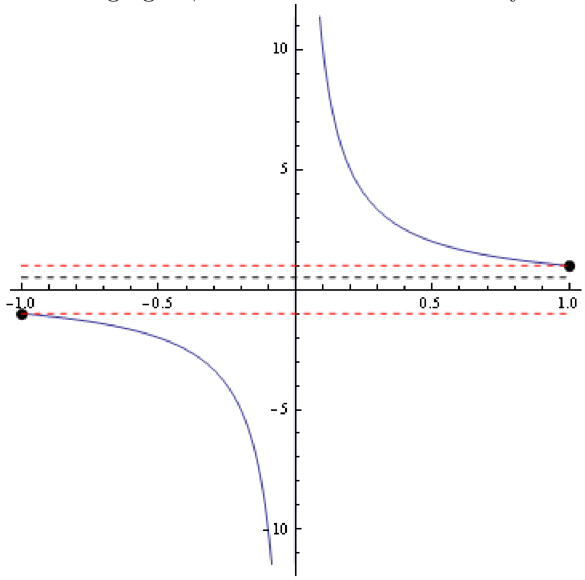
A little while ago, we talked of the notion of a function that is increasing at a point, and that was basically something similar, except that there one of the comparison points was fixed and the other one was restricted to somewhere close by. For a function to be increasing on an interval means that it is increasing at every point in the interior of the interval. If the interval has endpoints, then the function attains a strict local minimum at the left endpoint and a strict local maximum at the right endpoint.

Similarly, we say that f is *decreasing* on an interval I if, for any $x_1, x_2 \in I$, with $x_1 < x_2$, we have $f(x_1) > f(x_2)$. In other words, the larger the input, the smaller the output.

When I do not specify the interval and simply say that a function is increasing (respectively, decreasing), I mean that the function is increasing (respectively, decreasing) over its entire domain. For functions whose domain is the set of all real numbers, this means that the function is increasing (respectively, decreasing) over the set of all real numbers.

An example of an increasing function is a function $f(x) := ax + b$ with $a > 0$. An example of a decreasing function is a function $f(x) := ax + b$ with $a < 0$.

By the way, here's an interesting and weird example. Consider the function $f(x) := 1/x$. This function is not defined at $x = 0$. So, its domain is a union of two disjoint open intervals: the interval $(-\infty, 0)$ and the interval $(0, \infty)$. Now, we see that on each of these intervals, the function is decreasing. In fact, on the interval $(-\infty, 0)$, the function starts out from something close to 0 and then becomes more and more negative, approaching $-\infty$ as x tends to zero from the left. And then, on the interval $(0, \infty)$, the function is decreasing again, down from $+\infty$ all the way to zero.



But, taken together, is the function decreasing? No, and the reason is that at the point 0, where the function is undefined, it is undergoing this *huge* shift – from $-\infty$ to ∞. This fact – that points where the function is undefined can be points where it jumps from $-\infty$ to $+\infty$ or $+\infty$ to $-\infty$ – is a fact that keeps coming up. If you remember, this same fact haunted us when we were trying to apply the intermediate-value theorem to the function $1/x$ on an interval containing 0.

2.9. **The derivative sign condition for increasing/decreasing.** We first state the result for open intervals, where it is fairly straightforward. Suppose f is a function defined on an open interval (a, b). Suppose, further, that f is continuous and differentiable on (a, b), and for every point $x \in (a, b)$, $f'(x) > 0$. Then, f is an increasing function on (a, b).

A similar statement for decreasing: If f is a function defined on an interval (a, b). Suppose, further, that f is continuous and differentiable on (a, b), and for every point $x \in (a, b)$, $f'(x) < 0$. Then, f is a decreasing function on (a, b).

The result also holds for open intervals that stretch to ∞ or $-\infty$.

Note that it is important that f should be defined for all values in the interval (a, b), that it should be continuous on the interval, and that it should be differentiable on the interval. Here are some counterexamples:

(1) Consider the function $f(x) := 1/x$, defined and differentiable for $x \neq 0$. Its derivative is $f'(x) := -1/x^2$, which is negative wherever defined. Hence, f is decreasing on any open interval not containing 0. However, it is *not* decreasing on any open interval containing 0.
(2) Consider the function $f(x) := \tan x$. The derivative of the function is $f'(x) := \sec^2 x$. Note that f is defined for all x that are not odd multiples of $\pi/2$, and the same holds for f'. Also, note that $f'(x) > 0$ wherever defined, because $|\sec x| \geq 1$ wherever defined. Thus, the tan function is increasing on any interval not containing an odd multiple of $\pi/2$. But at each odd multiple of $\pi/2$, it slips from $+\infty$ to $-\infty$.

Let us now look at the version for a closed interval.

Suppose f is a function defined on a closed interval $[a,b]$, which is continuous on $[a,b]$ and differentiable on (a,b). Then, if $f'(x) > 0$ for $x \in (a,b)$, then f is increasing on all of $[a,b]$. Similarly, if $f'(x) < 0$ for $x \in (a,b)$, then f is decreasing on all of $[a,b]$.

In other words, we do *not* need to impose conditions on one-sided derivatives at the endpoints in order to guarantee that the function is increasing on the entire closed interval.

Finally, if $f'(x) = 0$ on the interval (a,b), then f is constant on $[a,b]$.

Some other versions:

(1) The result also applies to half-closed, half-open intervals. So, it may happen that a function f is continuous on $[a,b)$, differentiable on (a,b), and $f'(x) > 0$ for $x \in (a,b)$. In this case, f is increasing on $[a,b)$.
(2) The result also applies to intervals that stretch to infinity in either or both directions.

2.10. **Finding where a function is increasing and decreasing.** Let's consider a function f that, for simplicity, is continuously differentiable on its domain. So, f' is a continuous function. We now note that, in order to find out where f is increasing and decreasing, we need to find out where f' is positive, negative and zero.

Here's an example, Consider the function $f(x) := x^3 - 3x^2 - 9x + 7$. Where is f increasing and where is it decreasing? In order to find out, we need to differentiate f. The function $f'(x)$ is equal to $3x^2 - 6x - 9 = 3(x-3)(x+1)$. By the usual methods, we know that f' is positive on $(-\infty,-1) \cup (3,\infty)$, negative on $(-1,3)$, and zero at -1 and 3. Thus, the function f is increasing on the intervals $(-\infty,-1]$ and $[3,\infty)$ and decreasing on the interval $[-1,3]$.

Note that it is *not* correct to conclude from the above that f is increasing on the set $(-\infty,-1] \cup [3,\infty)$, although it is increasing on each of the intervals $(-\infty,-1]$ and $[3,\infty)$ *separately.* This is because the two pieces $(-\infty,-1]$ and $[3,\infty)$ are in some sense independent of each other. In general, the positive derivative implies increasing conclusions hold on intervals because they are what mathematicians call *connected sets*, and not for disjoint unions of intervals. In the case of this specific function, we note that $f(-1) = 12$ and $f(3) = -20$, so the value of the function at the point 3 is smaller than it is at -1. Thus, it is not correct to think of the function as being increasing on the union of the two intervals.

Similarly, if f is a rational function $x^2/(x^3-1)$, then we get $f'(x) = (-2x - x^4)/(x^3-1)^2$. Now, in order to find out where this is positive and where this is negative, we need to factor the numerator and the denominator. The factorization is:

$$\frac{-x(x+2^{1/3})(x^2 - 2^{1/3}x + 2^{2/3})}{(x-1)^2(x^2+x+1)^2}$$

The zeros of the numerator are 0 and $-2^{1/3}$ and the zero of the denominator is 1. The quadratic factors in both the numerator and the denominator are always positive. Also note that there is a minus sign on the outside.

Hence, f' is negative on $(1,\infty)$, $(0,1)$, and $(-\infty,-2^{1/3})$, positive on $(-2^{1/3},0)$, zero on 0 and $-2^{1/3}$, and undefined at 1. Thus, f is decreasing on $[0,1)$, $(1,\infty)$, and $(-\infty,-2^{1/3}]$, increasing on $[-2^{1/3},0]$.

Now, let's combine this with the information we have about f itself. Note that f is undefined at 1, it is positive on $(1,\infty)$, it is zero at 0, and it is negative on $(-\infty,0) \cup (0,1)$. How do we combine this with information about what's happening with the derivative?

On the interval $(-\infty,-2^{1/3})$, f is negative *and* decreasing. What's happening as $x \to -\infty$? f tends to zero (we'll see why a little later). So, as x goes from $-\infty$ to $-2^{1/3}$, f goes down from 0 to $-2^{2/3}/3$. Then, as x goes from $-2^{1/3}$ to 0, f is still negative but starts going up from $-2^{2/3}/3$ and reaches 0. In the interval from 0 to 1, f goes back in the negative direction, from 0 down to $-\infty$. Then, in the interval $(1,\infty)$, f goes emerges from $+\infty$ and goes down to 0 as $x \to +\infty$.

So we see that information about the sign of the derivative helps us get a better picture of how the function behaves, and allows us to better draw the graph of the function – something that we will try to do more of a short while from now.

Point-value distinction. We use the term *point of local maximum* or *point of local minimum* (or simply *local maximum* or *local minimum*) for the point in the domain, and the term *local maximum value* for the value of the function at the point.

3. Determining local extreme values

3.1. **Local extreme values and critical points.** If f is a function and c is a point in the interior of the domain of f, then f is said to have a *local maximum* at c if $f(x) \leq f(c)$ for all x sufficiently close to c. Here, *sufficiently close* means that there exists $a < c$ and $b > c$ such that the statement holds for all $x \in (a, b)$.

Similarly, we have the concept of *local minimum* at c.

The points in the domain at which local maxima and local minima occur are termed the *points of local extrema* and the values of the function at these points are termed the *local extreme values*.

As we discussed last time, if f is differentiable at a point c of local maximum or local minimum, the derivative of f at c is zero. This suggests that we define a notion.

An interior point c in the domain of a function f is termed a *critical point* if either $f'(c) = 0$ or $f'(c)$ does not exist. Thus, all the local extreme values occur at critical points – because at a local maximum or minimum, either the derivative does not exist, or the derivative equals zero.

Note that not all critical points are points of local maxima and minima. For instance, for the function $f(x) := x^3$, the point $x = 0$ is a critical point, but the function does not attain a local maximum or local minimum at that point. However, critical points give us a small set of points that we need to check against. Once we have this small set, we can use other methods to determine what precisely is happening at these points.

3.2. **First-derivative test.** The first-derivative test basically tries to determine whether something is a local maximum by looking, not just at the value of the derivative *at* the point, but also the value of the derivative *close* to the point.

Basically, we want to combine the idea of *increasing on the left, decreasing on the right* to show that something is a local maximum, and similarly, we combine the idea of *decreasing on the left, increasing on the right* to show that something is a local minimum.

The first-derivative test says that if c is a critical point for f and f is continuous at c (Note that f need not be differentiable at c). if there is a positive number δ such that:

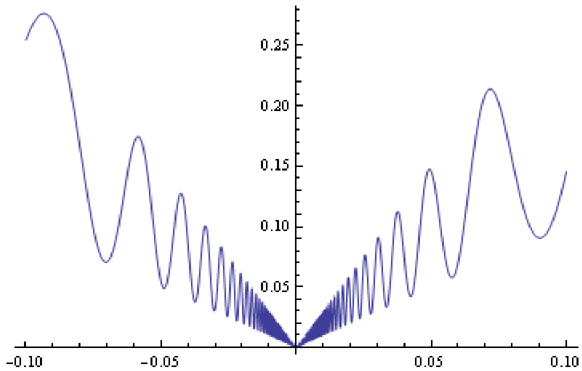
(1) $f'(x) > 0$ for all $x \in (c - \delta, c)$ and $f'(x) < 0$ for all $x \in (c, c + \delta)$, then $f(c)$ is a local maximum, i.e., c is a point of local maximum for f.
(2) $f'(x) < 0$ for all $x \in (c - \delta, c)$ and $f'(x) > 0$ for all $x \in (c, c + \delta)$, then $f(c)$ is a local minimum, i.e., c is a point of local minimum for f.
(3) $f'(x)$ keeps constant sign on $(c - \delta, c) \cup (c, c + \delta)$, then c is not a point of local maximum/minimum for f.

Thus, for the function $f(x) := x^2/(x^3 - 1)$, there is a local minimum at $-2^{1/3}$ and a local maximum at 0.

Recall that for the function $f(x) := x^3$, the derivative at zero is zero, so it is a critical point but it is not a point of local extremum, because the derivative is positive everywhere else.

3.3. **What are we essentially doing with the first-derivative test?** Why does the first-derivative test work? Essentially it is an application of the results on increasing and decreasing functions for closed intervals. What we're doing is using the information about the derivative from the left to conclude that the point is a strict local maximum from the left, because the function is increasing up to the point, and is a strict local maximum from the right, because the function is decreasing down from the point.

3.4. **The first-derivative test is sufficient but not necessary.** For most of the function that you'll see, the first-derivative test will give you a good way of figuring out whether a given critical point is a local maximum or local minimum. There are, however, situations where the first-derivative test fails to work. These are situations where the derivative changes sign infinitely often, close to the critical point, so does not have a constant sign near the critical point. For instance, consider the function $f(x) := |x|(2 + \sin(1/x))$. This attains a local minimum at the point $x = 0$, which is a critical point. However, the derivative of the function oscillates between the positive and negative sign close to zero and doesn't settle into a single sign on either side of zero.



3.5. **Second-derivative test.** One problem with the first-derivative test is that it requires us to make two local sign computations over *intervals*, rather than *at points*. Discussed here is a variant of the first-derivative test, called the second-derivative test, that is sometimes easier to use.

Suppose c is a critical point in the interior of the domain of a function f, and f is twice differentiable at c. Then, if $f''(c) > 0$, c is a point of local minimum, whereas if $f''(c) < 0$, then c is a point of local minimum.

The way this works is as follows: if $f''(c) > 0$, that means that f' is (strictly) increasing at c. This means that f' is negative to the immediate left of c and is positive to the immediate right of c. Thus, f attains a local minimum at c.

Note that the second-derivative test works for critical points where the function is twice-differentiable. In particular, it does not work for the kind of sharp peak points where the function suddenly changes direction. On the other hand, since the second-derivative test involves evaluation of the second derivative at only one point, it may be easier to apply in certain situations than the first-derivative test, which requires reasoning about the sign of a function over an interval.

4. Finding maxima and minima: a global perspective

4.1. **Endpoint maxima and minima.** An *endpoint maximum* is something like a local maximum, except that it occurs at the endpoint of the domain, so the value of the function at the point needs to be compared only with the values of the function at points sufficiently close to it on one side (the side that the domain is in). Similarly, an *endpoint minimum* is like a local minimum, except that it occurs at the endpoint of the domain, so the value of the function at the point needs to be compared only with the values of the function at points sufficiently close to it on one side.

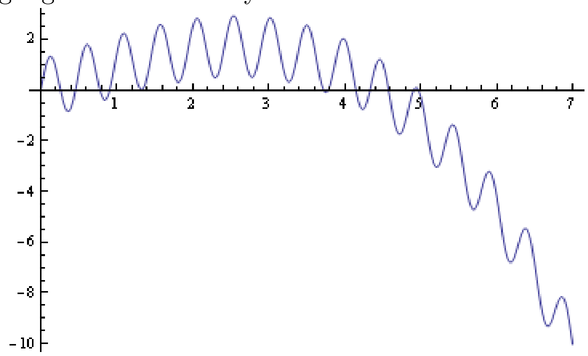
If the endpoint is a left endpoint, then being an endpoint maximum (respectively, minimum) means being a local maximum (respectively, minimum) from the right. If the endpoint is a right endpoint, then being an endpoint maximum (respectively, minimum) means being a local maximum (respectively, minimum) from the left.

4.2. **Absolute maxima and minima.** We say that a function f has an absolute maximum at a point d in the domain if $f(d) \geq f(x)$ for all x in the domain. We say that f has an absolute minimum at a point d in the domain if $f(d) \leq f(x)$ for all x in the domain. The corresponding value $f(d)$ is termed the absolute maximum (respectively, minimum) of f on its domain.

Notice the following very important fact about absolute maxima and minima, which distinguishes them from local maxima and minima. If an absolute maximum value exists, then the value is unique, even though it may be attained at multiple points on the domain. Similarly, if an absolute minimum value exists, then the value is unique, even though it may be attained at multiple points of the domain. Further, assuming the function to be continuous through the domain, and assuming the domain to be connected (i.e., not fragmented into intervals) the range of the function is the interval between the absolute minimum value and the absolute maximum value. This follows from the intermediate value theorem.

For instance, for the cos function, absolute maxima occur at multiples of 2π and absolute minima occur at odd multiples of π. The absolute maximum value is 1 and the absolute minimum value is -1.

Just for fun, here's a picture of a function having lots of local maxima and minima, but all at different levels. Note that some of the local maximum values are less than some of the local minimum values. This highlights the extremely local nature of local maxima/minima.



4.3. **Where and when do absolute maxima/minima exist?** Recall the *extreme value theorem* from some time ago. It said that for a continuous function on a closed interval, the function attains its maximum and minimum. This was basically asserting the existence of absolute maxima and minima for a continuous function on a closed interval.

Notice that any point of absolute maximum (respectively, minimum) is either an endpoint or is a point of local maximum (respectively, minimum). We further know that any point of local maximum or minimum is a critical point. Thus, in order to find all the absolute maxima and minima, a good first step is to find critical points and endpoints.

Another thing needs to be noted. For some funny functions, it turns out that there is no maximum or minimum. This could happen for two reasons: first, the function approaches $+\infty$ or $-\infty$, i.e., it gets arbitrarily large in one direction, somewhere. Second, it might happen that the function approaches some maximum value but does not attain it on the domain. For instance, the function $f(x) = x$ on the interval $(0, 1)$ does not attain a maximum or minimum, since these occur at the endpoints, which by design are not included in the domain.

Thus, the absolute maxima and minima, *if they occur*, occur at critical points and endpoints. But we need to further tackle the question of existence. In order to deal with this issue clearly, we need to face up to something we have avoided so far: limits to infinity.

4.4. **Limits at infinity and to infinity.** We need to tackle two questions: first, what do we mean by trying to evaluate $\lim_{x\to\infty} f(x)$ and $\lim_{x\to-\infty} f(x)$, and second, what do we mean by saying $\lim_{x\to c} f(x) = \infty$ and $\lim_{x\to c} f(x) = -\infty$.

For both, the basic idea is that for something to approach $+\infty$ means that it eventually crosses every arbitrarily large number and does not come back down, while approacing $-\infty$ means that it eventually crosses below every arbitrarily small negative number and does not come back. Formally, we say that:

(1) $\lim_{x\to c} f(x) = +\infty$ if, for every $N > 0$, there exists $\delta > 0$ such that if $0 < |x - c| < \delta$, $f(x) > N$.
(2) $\lim_{x\to c} f(x) = -\infty$ if, for every $N > 0$, there exists $\delta > 0$ such that if $0 < |x - c| < \delta$, $f(x) < -N$.
(3) $\lim_{x\to\infty} f(x) = L$ if, for every $\epsilon > 0$, there exists $N > 0$ such that for $x > N$, $|f(x) - L| < \epsilon$.
(4) $\lim_{x\to-\infty} f(x) = L$ if, for every $\epsilon > 0$, there exists $N > 0$ such that for $x < -N$, $|f(x) - L| < \epsilon$.
(5) $lim_{x\to\infty} f(x) = \infty$ if, for every $N > 0$, there exists $M > 0$ such that if $x > M$, then $f(x) > N$.
(6) $\lim_{x\to\infty} f(x) = -\infty$ if, for every $N > 0$, there exists $M > 0$ such that if $x > M$, then $f(x) < -N$.
(7) $\lim_{x\to-\infty} f(x) = \infty$ if, for every $N > 0$, there exists $M > 0$ such that if $x < -M$, then $f(x) > N$.
(8) $\lim_{x\to-\infty} f(x) = -\infty$ if, for every $N > 0$, there exists $M > 0$ such that if $x < -M$, then $f(x) < -N$.

We will consider limits to infinity in much more detail in 153.

4.5. **Rules of thumb for figuring out limits at infinity.** Here are some rules. Note that each rule also applies to one-sided limits, and often has to be applied in a one-sided sense because the signs of infinity being approached from the two sides may be different:

(1) If the numerator approaches a positive number and the denominator approaches zero from the positive side, then the quotient approaches $+\infty$.

(2) If the numerator approaches a negative number and the denominator approaches zero from the positive side, the quotient approaches $-\infty$.
(3) If the numerator approaches a positive number and the denominator approaches zero from the negative side, the quotient approaches $-\infty$.
(4) If the numerator approaches a negative number and the denominator approaches zero from the negative side, the quotient approaches $+\infty$.

For instance, for the function $f(x) := 1/x^2$, the numerator approaches (in fact, equals) a positive number, and the denominator approaches zero from the positive side, so the limit at 0 is $+\infty$.

For the function $g(x) := 1/x$, as $x \to 0^-$, the numerator is positive and the denominator approaches zero from the negative side, so the quotient approaches $-\infty$. As $x \to 0^+$, the numerator is positive and the denominator approaches zero from the positive side, so the quotient approaches $+\infty$.

Let's use these ideas to revisit our example of the rational function $x^2/(x^3 - 1)$. Recall that here the only point where the function is not defined is $x = 1$. Here, the numerator approaches a positive number (1). As $x \to 1^-$, the denominator approaches 0 from the left, so the quotient approaches $-\infty$, and as $x \to 1^+$, the denominator approaches 0 from the right side, so the quotient approaches $+\infty$. Thus, the left-hand limit is $-\infty$ and the right-hand limit is $+\infty$. Note that, when the limits are $\pm\infty$, we are still allowed to say, and should say, that the limits do not exist. Infinite does not exist.

Next, we look at rules of thumb that guide us when $x \to \infty$. Here are some of these rules:

(1) For a polynomial p of degree one or higher, $p(x) \to \infty$ as $x \to \infty$ if the leading coefficient of p is positive, and $p(x) \to -\infty$ as $x \to \infty$ if the leading coefficient of p is negative.
(2) For a polynomial p of degree one or higher, $p(x) \to \infty$ as $x \to -\infty$ if the leading coefficient of p is positive and the degree of p is even, and $p(x) \to -\infty$ as $x \to -\infty$ if the leading coefficient of p is positive and the degree of p is odd. When the leading coefficient of p is negative, the signs get reversed.
(3) For a rational function, the limits as $x \to \pm\infty$ can be computed by simply looking at the limit of the quotient of the leading terms in the numerator and the denominator.
(4) For a rational function, if the degree of the denominator is greater than the degree of the numerator, the value of the rational function approaches 0 as the input goes to ∞ and as the input goes to $-\infty$. In other words, for such a rational function r, $\lim_{x\to\infty} r(x) = \lim_{x\to-\infty} r(x) = 0$.
(5) For a rational function, if the degree of the denominator is smaller than the degree of the numerator, the limit of the rational function, as $x \to \infty$, is the infinity with the same sign as the quotient of the leading coefficients. As $x \to -\infty$, it is the infinity with the same sign as the product of (the quotient of the leading coefficients) and (-1 to the power of the difference of degrees).
(6) For a rational function, if the numerator and the denominator have equal degrees, then the limit as $x \to \infty$, and the limit as $x \to -\infty$, are both equal to the quotient of the leading coefficients.

4.6. **Strategy for computing absolute maxima and minima.** Here's all the candidates we have to deal with:

(1) All endpoints in the domain, and the function values at those endpoints.
(2) All critical points in the domain, and the function values at those critical points.
(3) For points not in the domain but in the boundary of the domain, as well as at $\pm\infty$ (if the domain stretches to $+\infty$ and/or $-\infty$), we try to compute the limits.

Here's what we get, comparing the values:

(1) If, for any of the points where the function isn't defined, or at $\pm\infty$, the limit is $+\infty$, there isn't any absolute maximum. If, for any of the points where the function isn't defined, or at $\pm\infty$, the limit is $-\infty$, there isn't any absolute minimum.
(2) Consider the values of the function at all the critical points, and the limits at $\pm\infty$ and all other points in the boundary of the domain but not in the domain itself. If the maximum of these is attained by one of the critical points, that is the absolute maximum. If the maximum is attained as one of the limits but not at any of the critical points, there is no absolute maximum. Similar remarks apply for minima.

4.7. **Other subtle issues.** Here are some additional issues:

- When there are only finitely many critical points, endpoints, and limit situations, and we need to find the absolute maximum or absolute minimum, we do *not* need to use the derivative tests to figure out which of them are local maxima, which are local minima, and which are neither. We can simply compute the values and compare.
- However, as the picture shown a little earlier indicates, just looking at the values does not immediately tell us whether we have a local maximum, local minimum, or neither. Some lcoal maximum values may be smaller than some local minimum values.
- If there are infinitely many critical points, endpoints, and limit situations, we may need to think a little more clearly about what is happening. It may be helpful to use derivative tests and facts about even, odd, and periodic functions to eliminate or narrow down possibilities.

5. Important fact critical for integration

We noted a little while back that if f is a continuous function on a closed interval $[a, b]$, and its derivative is zero on the open interval (a, b), then f is constant on $[a, b]$.

This fact has an important corollary, which is critical to the whole setup and process of integration:

> If f and g are continuous functions on a closed interval $[a, b]$ and $f'(x) = g'(x)$ for all $x \in (a, b)$, then $f - g$ is a constant function on $[a, b]$.

We will return to this fact and its implications a little later.

6. Piecewise defined functions

We now consider all the above notions for functions defined piecewise on intervals. As usual, we assume that each of the piece functions is nice enough, which in this case means we assume that it is twice continuously differentiable.

For functions defined piecewise, we need to spearately consider all the points where the definition changes. As far as we are concerned, for each point where the definition changes, we have the following possibilities:

- The function is not continuous at this point: In this case, we need to separately consider the left-hand limit and right-hand limit at the point, and the value at that point, and consider all these as candidates for the local extreme values.
- The function is continuous but not differentiable at this point: Then it is a critical point, and the value there might be a candidate for a local extreme value. Whether it is or not depends on the signs of the one-sided derivatives.
- The function is continuously differentiable at the point, and the derivative is zero: Then again, it is a critical point.
- The function is continuously differentiable at the point, and the derivative is nonzeor: Then, it is not a critical point.

We will consider all these in more detail when we study the graphing of functions.