MAXIMUM AND MINIMUM VALUES: ONE VARIABLE AND TWO

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.7.

What students should definitely get: The definition of critical point in terms of gradient vector being zero, the second derivative test, the use of these to find local and absolute extreme values.

What students should hopefully get: The similarities and differences between the situations for one and two variables.

Executive summary

Words ...

(1) For a directional local minimum, the directional derivative (in the outward direction from the point) is greater than or equal to zero. For a directional local maximum, the directional derivative (in the outward direction from the point) is less than or equal to zero.

Note that even for *strict* directional local maximum or minimum, the possibility of the directional derivative being zero cannot be ruled out.

(2) If a point is a point of directional local minimum from two opposite directions (i.e., it is a local minimum along a line through the point, from both directions on the line) then the directional derivative along the line, if it exists, must equal zero.

(3) If a function of two variables is differentiable at a point of local minimum or local maximum, then the directional derivative of the function is zero at the point in every direction. Equivalently, the gradient vector of the function at the point is the zero vector. Equivalently, both the first partial derivatives at the point are zero.

Points where the gradient vector is zero are termed *critical points*.

(4) If the directional derivatives along some directions are positive and the directional derivatives along other directions are negative, the point is likely to be a *saddle point*. A saddle point is a point for which the tangent plane to the surface that's the graph of the function slides through the graph, i.e., it is not completely on one side.

(5) For a function f of two variables with continuous second partials, and a critical point (a,b) in the domain (so $f_x(a,b) = f_y(a,b) = 0$) we compute the Hessian determinant:

$$D(a,b) = f_{xx}(a,b)f_{yy}(a,b) - [f_{xy}(a,b)]^2$$

If $D(a,b) > 0$ and $f_{xx}(a,b) > 0$, the function has a local *minimum* at the point (a,b). If $D(a,b) > 0$ and $f_{xx}(a,b) < 0$, the function has a local *maximum* at the point (a,b). If $D(a,b) < 0$, we get a saddle point at the point. If $D(a,b) = 0$, the situation is inconclusive, i.e., the test is indecisive.

(6) For a closed bounded subset of $\mathbb{R}^n$ (and specifically $\mathbb{R}^2$) any continuous function with domain that subset attains its absolute maximum and minimum values. These values are attained either at interior points (in which case they are local extreme values and must be attained at critical points) or at boundary points.

(7) *Relation with level curves*: *Typically*, local extreme values correspond to isolated single point level curves. However, this is not always the case, and there are some counterexamples. To be more precise, any *isolated* or *strict* local extreme value corresponds to a (locally) single point level curve.

Actions ...

(1) Strategy for finding local extreme values: First, find all the critical points by solving $f_x(a,b) = 0$ and $f_y(a,b) = 0$ as a pair of simultaneous equations. Next, use the second derivative test for each critical point, and if feasible, try to figure out if this is a point of local maximum, or local minimum, or a saddle point.

(2) To find absolute extreme values of a function on a closed bounded subset of $\mathbb{R}^2$, first find critical points, then find critical points for a parameterization of the boundary, and then compute values at all of these and see which is largest and smallest. *If the list of critical points is finite, and we need to find absolute maximum and minimum, it is not necessary to do the second derivative test to figure out which points give local maximum, local minimum, or neither, we just need to evaluate at all points and find the maximum/minimum.*
(3) When the domain of the function is bounded but not closed, we must consider the possibility of extreme values occurring as we approach boundary points not in the domain. If the domain is not bounded, we must consider directions of approach to infinity.

1. Local increase and decrease behavior: one variable recall

1.1. **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.

1.2. **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.

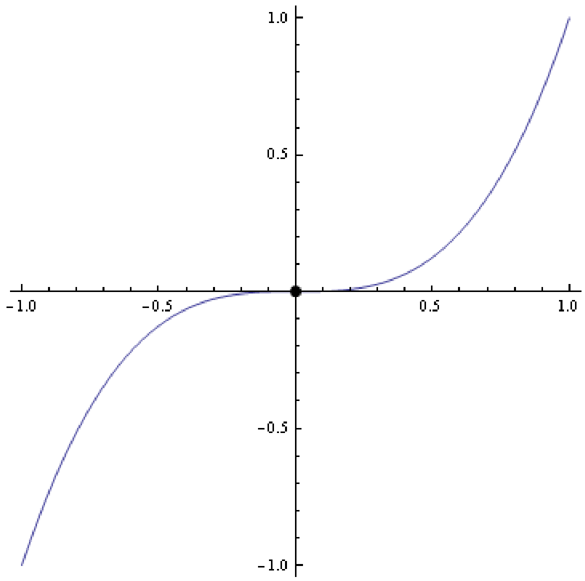
A related notion is that of a *strict local maximum from the left*, which means that there exists $a < c$ such that $f(x) < f(c)$ for all $x \in (a, c)$. In other words, $f(c)$ is *strictly bigger* than $f(x)$ for x to the immediate left of c.

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.

1.3. **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:



As you know by now, the phenomenon we are dealing with in this case is called a *point of inflection*. However, there are many other cases where a similar phenomenon occurs, as will be clear soon.

1.4. **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.

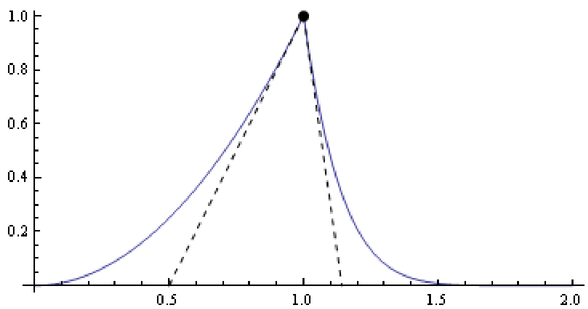
So, let's piece these together:

(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.

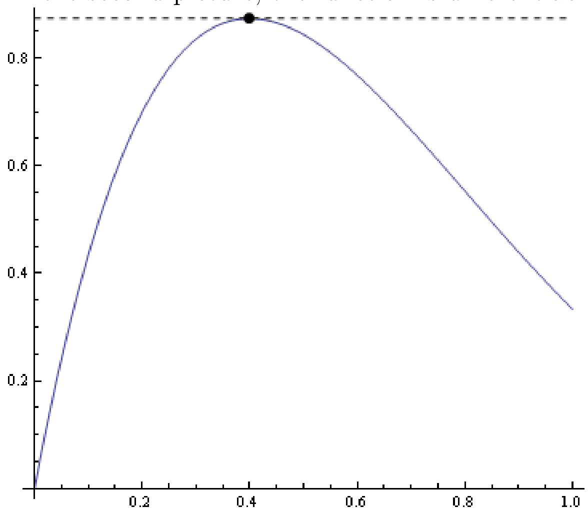
What can we say about local maxima and local minima? We can say the following:

(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.



1.5. **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 $f'(c) \leq 0$.

1.6. **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 comply 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 $x = 0$.

1.7. **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 $c \implies f'(c) \leq 0$

1.8. **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.

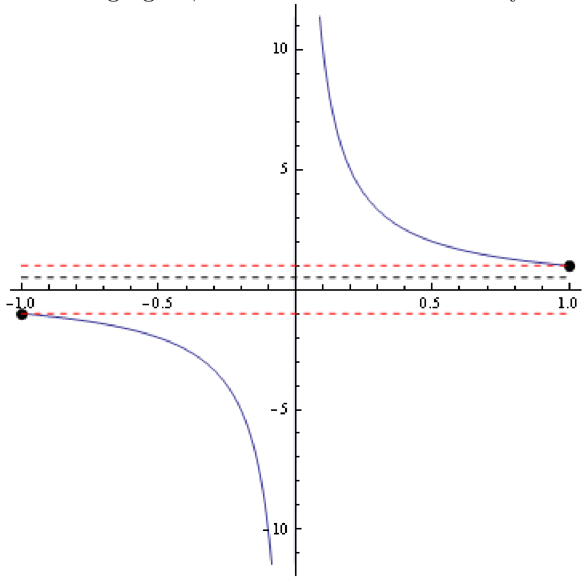
A little while ago, we talked of the notion of a function that is increasing at a point, and that was basically something similar, except that there one of the comparison points was fixed and the other one was restricted to somewhere close by. For a function to be increasing on an interval means that 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.

1.9. **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.

1.10. **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.

2. Strategies for local and absolute maxima/minima

2.1. **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.

2.2. **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:

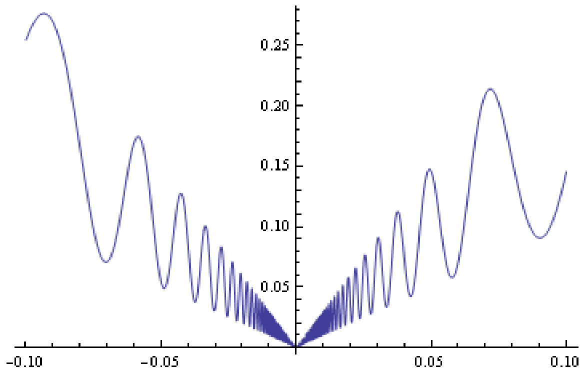
(1) $f'(x) > 0$ for all $x \in (c - \delta, c)$ and $f'(x) < 0$ for all $x \in (c, c + \delta)$, then $f(c)$ is a local maximum, i.e., c is a point of local maximum for f.
(2) $f'(x) < 0$ for all $x \in (c - \delta, c)$ and $f'(x) > 0$ for all $x \in (c, c + \delta)$, then $f(c)$ is a local minimum, 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.

2.3. **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.

2.4. **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.



2.5. **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.

2.6. **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.

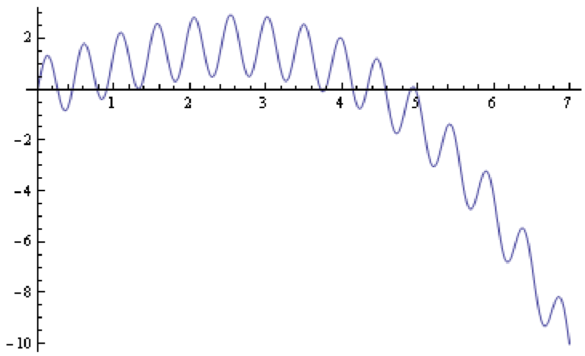
If the endpoint is a left endpoint, then being an endpoint maximum (respectively, minimum) means being a local maximum (respectively, minimum) from the right. If the endpoint is a right endpoint, then being an endpoint maximum (respectively, minimum) means being a local maximum (respectively, minimum)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

2.7. **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.



2.8. **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.

3. Moving to multiple variables

3.1. **Directions and neighborhoods.** The first key difference between functions of one variable and functions of more than one variable is that in the latter case, there are a lot more directions. For a function of 2 variables, there are four directions parallel to the axes: a left and a right approach in each coordinate keeping the other fixed. But there are many other straight line approaches along other directions, which are not parallel to either axes. There could also be curved directions of approach.

Thus, although it is possible to talk of directional maxima and minima (and we'll do this in a moment) we need another approach to defining local maxima and minima. The key idea is that of *neighborhood*, which works by simultaneously covering all directions. We could use as neighborhoods either circular disks or rectangular regions.

3.2. **The technical definition of local maximum and local minimum.** Consider a function f of two variables with domain D. We say that a point (a, b) in the interior of D is a point of *local minimum* for f if there exists an open disk U around (a, b) contained in D such that $f(a, b) \leq f(x, y)$ for all $(x, y) \in U$. Equivalently, there exists a positive number r such that $f(a, b) \leq f(x, y)$ whenever the distance between the points (x, y) and (a, b) is less than r.

Note that with the open disk formulation, we typically use circular disks. However, we could also use square disks centered at (a, b) – these are sometimes easier to work with.

Analogously, we can define *strict local minimum*, *local maximum*, and *strict local maximum*.

3.3. **The technical definition of absolute maximum and absolute minimum.** A point in the domain of a function is termed a point of *absolute maximum* if the value of the function at the point is greater than or equal to the value of the function at all other points in the domain of the function. Similarly, we talk of a point of *absolute minimum* for a function.

3.4. **Directional maximum and directional derivative.** Suppose f is a function of two variables and (a, b) is a point in the domain. For a particular direction of approach to the point (a, b) , e.g., a half-line (or curve) ending at (a, b), we can ask whether (a, b) is a *directional local maximum* or *directional local minimum* from that direction.

We say that (a, b) is a point of directional local maximum from that direction if $f(a, b)$ is greater than or equal to the f-value for all points in the half-line sufficiently close to (a, b). Similarly, we define directional local minimum. We can also define strict directional local maximum and strict directional local minimum.

For functions of one variable, there were only two directions worth pondering: left and right. Ergo, the idea of local maximum/minimum from the left and the right.

In the case of functions of one variable, we related being a local maximum/minimum from the left with the sign of the left-hand derivative (if it exists) and being a local maximum/minimum from the right with the sign of the right-hand derivative (if it exists). We can do the same thing now, and we obtain that:

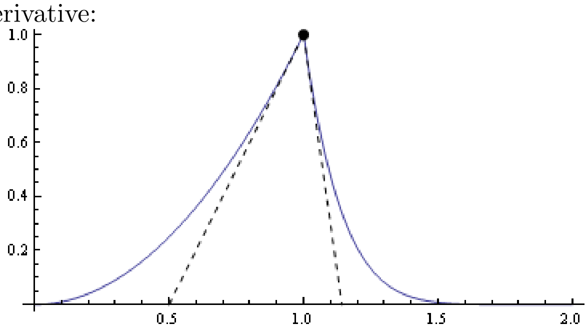
> For a directional local minimum, the directional derivative (in the outward direction from the point) is greater than or equal to zero. For a directional local maximum, the directional derivative (in the outward direction from the point) is less than or equal to zero.

Further, it remains true that *even for strict directional local maxima/minima*, the directional derivative *may well be zero*, i.e., we cannot rule out the possibility of the directional derivative being zero. We saw that for one variable, this proved crucial, because it meant that being a local maximum both from left and right forced the derivative (if it exists) to *equal zero*. The same idea works here:

> If a point is a point of directional local minimum from two opposite directions (i.e., it is a local minimum along a line through the point, from both directions on the line) then the directional derivative along the line, if it exists, must equal zero.

The reasoning is the same as for one variable: being a local minimum from one direction means the directional derivative in that direction is positive or zero. Being a local minimum from the opposite direction means the directional derivative in the opposite direction is positive or zero. *We know that directional derivatives in opposite directions are negatives of each other*, so overall we conclude that both directional derivatives must be zero.

On the other hand, it is possible to have a two variable situation analogous to the function of one variable that has a local maximum with a strictly positive left hand derivative and strictly negative right hand derivative:



For functions of two variables, this would correspond to the graph (which is a surface) not having a well defined tangent plane at the point of local maximum or minimum.

3.5. **Relating directional with the all-directions ideas.** A point of local minimum (respectively strict local minimum, local maximum, strict local maximum) for a function is also a point of local minimum (respectively strict local minimum, local maximum, strict local maximum) if we restrict attention to behavior on any curve/line containing the point. In particular, it is a directional local minimum (respectively, directional local maximum, strict directional local minimum, strict directional local maximum) for all directions.

Thus, we obtain that:

> If a function of two variables is differentiable at a point of local minimum or local maximum, then the directional derivative of the function is zero at the point in every direction. Equivalently, the gradient vector of the function at the point is the zero vector. Equivalently, both the first partial derivatives at the point are zero.

In symbols, if $f(x, y)$ is a function of two variables, and (a, b) is a point of local minimum for f, then, if f is differentiable at (a, b), we have $f_x(a, b) = f_y(a, b) = 0$. *Note that for a differentiable function, both partial derivatives being equal to zero implies that all directional derivatives are equal to zero.*

If we graph the surface $z = f(x, y)$, then pictorially, this means that the tangent plane at a point of local minimum, if it exists, is parallel to the xy-plane, and forms a kind of flat floor for the curve around the point. For a point of local maximum, the tangent plane exists, and forms a kind of flat ceiling around the point.

Points where the gradient vector is zero, i.e., the directional derivative is zero in all directions (which for continuous first partials just means that both partial derivatives are zero) are termed *critical points*. In particular, this means that all candidates for local extreme values occur at critical points.

3.6. **Local maximum, minimum, or neither?** For a function of one variable, recall that the derivative being zero could imply one of many things:

- Point of local maximum, such as x^2 at 0.
- Point of local minimum, such as $-x^2$ at 0.
- Point of increase/constancy – local maximum from left, local minimum from right. The typical examples are points of inflection as with x^3 at 0.
- Point of decrease/constancy – local minimum from left, local maximum from right, such as $-x^3$ at 0.
- None of the above: this includes oscillatory type examples, such as $x^2 \sin(1/x)$, $x \neq 0$ and defined to be 0 at 0. This is oscillatory aroun the point from both sides, though the derivative is still zero.

The same holds with functions of more than one variable – we could very well have the "point of inflection" type situation. However, things become more interesting in two variables, both easier and more complicated. Namely, we have not two, but many, directions, and more importantly, all the directions are connected to each other. By this I mean that I can smoothly rotate from one direction to any other direction, a feat that is impossible in one variable because the left and the right are discrete and opposite directions with no way of bridging across them.[1]

So, suppose we have a point (a, b) in the domain of f such that $f_x(a, b) = 0$ and $f_y(a, b) = 0$. This means that all the directional derivatives are zero.

Now, suppose f is a directional local minimum from some specific direction, e.g., from the positive x-direction. If we now smoothly rotate the direction, either f remains a directional local minimum as we keep rotating, or at some stage, it flips over into not being a directional local minimum at the point, perhaps being a directional local maximum. But *at the direction where the transition occurs, something funny is going on.* For instance, it may be that at the direction where the transition occurs, the function is actually *constant* along the direction. At any rate, the transition from being a local maximum to a local minimum is an interesting and nontrivial phenomenon.

This suggests that the right analogue of point of inflection with increase through – i.e., the right analogue of the type of thinking behind x^3 – is something where there are directions where there is a transition from directional maximum to directional minimum. It will be tricky to go into this in too much detail, but I'll note that the term for a point at which this kind of phenomenon occurs is *saddle point*.

3.7. **In search of a second derivative test.** Obviously, if $f(x, y)$ has a local minimum at (a, b), and f has second derivatives, then it should have a local minimum at (a, b) purely viewed as a function of x (keeping y fixed at b) and also purely viewed as a function of y (keeping x fixed at a). This suggests a second derivative test: we would like to see $f_{xx}(a, b) > 0$ (second derivative test on x) and $f_{yy}(a, b) > 0$ (second derivative test on y). Unfortunately, just having the conditions of both pure second partials greater than 0 at the point is not good enough. There are many other directions.

Luckily, there is an easy fix, and that involves looking at the *mixed partials*. Unfortunately, the logic of this explanation is beyond the scope of the course, so you just need to take it on faith.

Define $D(a, b)$ as the quantity:

$$D(a, b) := f_{xx}(a, b) f_{yy}(a, b) - [f_{xy}(a, b)]^2$$

[1]This is basically the fact that a plane minus a point is still connected, whereas a line minus a point is disconnected.

Then, the second derivative test says the following for a point (a, b) such that f is *twice continuously differentiable* around (a, b) and $f_x(a, b) = f_y(a, b) = 0$:

- If $D(a, b) > 0$ and $f_{xx}(a, b) > 0$ then f attains a local minimum at (a, b). Note that the conditions $D > 0$ and $f_{xx}(a, b) > 0$ together not only imply that $f_{yy}(a, b) > 0$, they also imply that the second partial derivative along *any* direction is positive.
- If $D(a, b) > 0$ and $f_{xx}(a, b) < 0$ then f attains a local maximum at (a, b). Note that the conditions $D > 0$ and $f_{xx}(a, b) < 0$ together not only imply that $f_{yy}(a, b) < 0$, they also imply that the second partial derivative along *any* direction is negative.
- If $D(a, b) < 0$, then f does not attain a local maximum or a local minimum at (a, b). In this case, the point $(a, b, f(a, b))$ is a saddle point of the graph of f: the tangent plane through this point cuts through the surface that's the graph of this function.

  Note that this case definitely occurs if $f_{xx}(a, b)$ and $f_{yy}(a, b)$ have opposite signs, which makes sense because that would mean that the function is a local minimum along one axis and a local maximum along the other. But it could occur even if both of them have the same sign, if the magnitude of $(f_{xy}(a, b))^2$ is bigger than the product of their magnitudes. What that basically means is that there are diagonal directions along which the function's behavior is opposite to what it is along the x and y directions.
- If $D(a, b) = 0$, the second derivative test is inconclusive.

The value $D(a, b)$ can be thought of as the determinant of this 2×2 matrix:

$$\begin{pmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{yx}(a,b) & f_{yy}(a,b) \end{pmatrix}$$

Note that the two off-diagonal entries $f_{xy}(a, b)$ and $f_{yx}(a, b)$ are equal by Clairaut's theorem and the assumption of continuity of partials of f around (a, b). If you ever see multivariable calculus in its proper form, you will learn that this matrix is called a *Hessian* and the conditions of the second derivative test are basically conditions to ensure that the Hessian is a positive definite (respectively, negative definite) matrix.

3.8. **Putting things together: the technique for a function of two variables.** Here now is the overall procedure for finding local and absolute maxima/minima for a twice continuously differentiable function of two variables:

- First, find all the *critical points*. A critical point is a point in the domain at which all directional derivatives are zero, or equivalently, the two first partials are zero. In symbols, (a, b) is a critical point if $f_x(a, b) = 0$ and $f_y(a, b) = 0$.
- Next, for each critical point, use the second derivative test, if feasible, to find out whether it is a point of local maximum, point of local minimum, or a saddle point. If the second derivative test is inconclusive, see if there are other ways of figuring things out.

3.9. **Absolute maxima and minima: boundary issues.** Recall that for functions of one variable, in addition to finding the local maxima/minima, we also needed to consider the endpoint maxima/minima and also the limits to boundary points not in the domain (and to infinity). A similar kind of complication arises for functions of two variables. For simplicity, we restrict attention to the following case: a function f with domain D a closed, bounded subset of $\mathbb{R}^2$ and with the property that f is twice continuously differentiable on the interior of D and its restriction to the boundary of D is "differentiable" under some smooth parameterization of the boundary.

Explanation of terminology:

- The boundary of a subset D of $\mathbb{R}^n$ is the set of points with the property that any open ball centered at the ball intersects D but is not completely contained in D. In other words, it's the points that are in close contact with D and with the complement of D.
- A closed subset of $\mathbb{R}^n$ is a subset that contains all its boundary points. A bounded subset of $\mathbb{R}^n$ is a subset that can be enclosed inside an open unit disk (or equivalently, in a rectangular region). Closed bounded subsets in $\mathbb{R}^n$ are what's called *compact* which makes a lot of facts about them true. (You don't have to worry what this means).

- The interior of a subset D of $\mathbb{R}^n$ is the subset of D comprising those points not in the boundary, i.e., those points for which we can make open balls centered at them lying completely inside D.

In the case of functions of one variable, the subsets live in $\mathbb{R}^1 = \mathbb{R}$ and are one-dimensional, and their boundaries are typically zero-dimensional, i.e., usually, sets of isolated points. In the case of functions of two variables, the subsets live in $\mathbb{R}^2$ and are two-dimensional, and the boundaries are typically one-dimensional, i.e., unions of lines and curves. In general, for n-dimensional subsets of $\mathbb{R}^n$, the boundary sets are expected to be $(n-1)$-dimensional.

The first result of relevance is the *extreme value theorem.* It states that for a closed, bounded subset D of $\mathbb{R}^2$ and a continuous function f on D, f attains its absolute maximum and minimum values at points in D.

The procedure for finding the absolute extreme values, if f is twice continuously differentiable on the interior of D and differentiable under a smooth parameterization of the boundary of D, is as follows:

(1) Find the critical points of f in the interior of D, and hence all the candidates for local extreme values in there.
(2) Find the extreme values of f on the boundary of D.
(3) Compare all these values and use these to find the absolute extremes.

3.10. **Stretching off to infinity.** Things get more complicated when the domain of the function is not a closed bounded region but instead stretches off to infinity in one or more than one direction. In this case, we need to figure out the "limits to infinity" of the function, if any, in order to find the absolute maxima and minima.

Unfortunately, there is more to this than meets the eye, because there are many different directions in which one can go off to infinity, and the limit may be different in each direction.

So, instead of finding these limits, a more useful approach may be to foreclose options, i.e., figure out that, say, if we go off to infinity in any direction, the function value is going to become too large, and thus the absolute minimum will not be attained outside of a certain close dbounded interval.

3.11. **Understanding the relationship with level curves.** It's worth pondering the relationship between local extreme values and level curves.

Level curves denote curves along which the function takes constant values. Local extreme values are typically level curves that are *single point* level curves. For instance, for the function $x^2 + y^2$, the level curves are the circles centered at the origin. The local (and absolute) minimum is the unique single point level curve, which occurs at the origin.

In a subsequent lecture discussion, we will look in detail at various classes of examples of maximum and minimum value computations for functions of two variables.