Matec Notes

Alexey Guzey*

December 23, 2016

I want to thank Elena Kochegarova for invaluable advice, without which these notes would not be written, and Polina Pilyugina, who provided me with her seminar notes, which were the initial inspiration behind these notes.

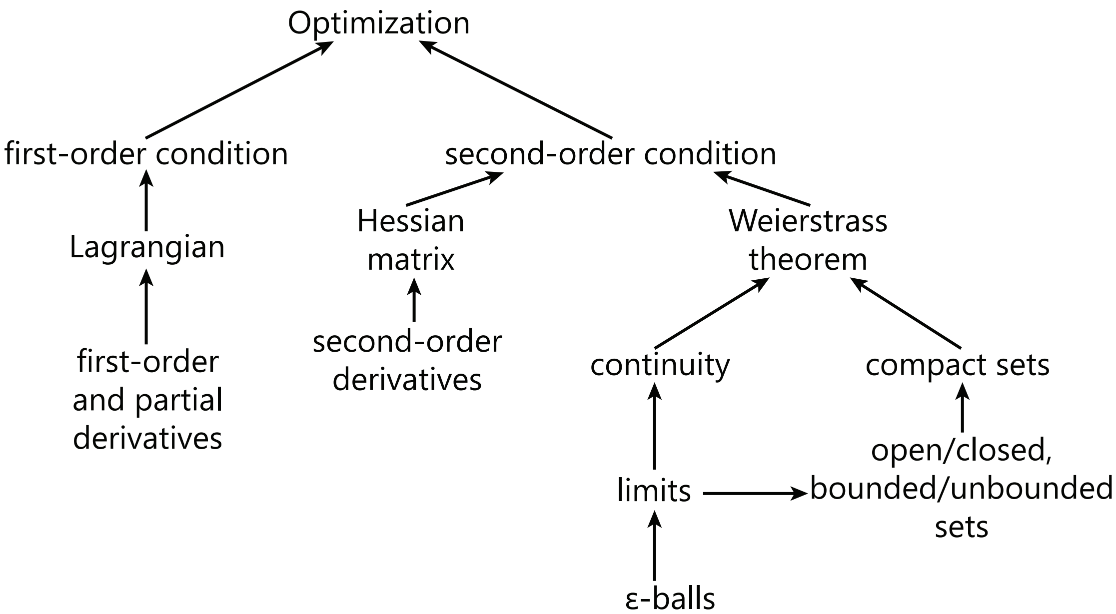
How to use the notes. The principle that guided us during the creation of these notes was to let the student not just memorize the algorithm, but to be able to understand why the algorithm works. Consequently, it focuses much more on the theoretical part of the course, rather than on problem-grinding. To facilitate the process of gaining insight into the concepts your lecturers and seminar teachers want you to learn, explanations were attempted to be made as clear as possible and a lot of effort was made to connect up ideas to each other. To make introduction of new concepts easier, most of the explanations begin with examples in $\mathbb{R}^1$ or $\mathbb{R}^2$ and only then are generalized.

Some chapters feature an appendix in which additional material can be found, such as even deeper explanations or simply something fun and intersting, tangential but outside the scope of the course.

There are two ways to look at the first couple of months of matec:

1. As an extension of first year calculus topics but for several dimensions

2. Getting ready to solve optimization problems. This is totally unobvious, but all the set theory, limits, derivatives of functions of several variables, etc. are needed to be able to fully understand constrained optimization problems, similar to the typical micro utility maximization problems, but more complex. Right now the picture below won't make any sense. But when you start getting the material, you can try to look back at it and you will realize how it's all connected:



*You can contact me regarding typos, suggestions, questions about the material, etc. by email: alex@guzey.com or on VK: vk.com/alexeyguzey

Contents

1 Set Operations and Notation **3**

1.1 Operations on sets . . . 3

1.2 Appendix . . . 4

2 Sequences in $\mathbb{R}^n$ **5**

2.1 Vectors . . . 5

2.2 Balls . . . 6

2.3 Sequences and Their Limits . . . 7

3 Sets **10**

3.1 Open and Closed Sets . . . 10

3.2 Compact Sets . . . 12

3.3 Appendix 1 . . . 14

3.4 Appendix 2 . . . 15

4 Multivariable Functions. Continuity. **16**

4.1 Level curves . . . 16

4.2 Limit of a Function . . . 17

4.3 Continuity . . . 17

4.4 Finding Limits with Polar Coordinates . . . 19

5 Differentiation of Multivariable Functions. Approximation **21**

5.1 Taylor Series Introduction . . . 21

5.2 First-Order (Linear) Approximation for One Variable . . . 22

5.3 First-Order (Linear) Approximation for Two Variables . . . 22

5.4 Tangent Plane . . . 23

5.5 Directional Derivative . . . 23

5.6 Gradient . . . 24

5.7 Chain Rule . . . 25

5.8 Appendix to Chain Rule . . . 28

5.9 Second-order approximation . . . 29

6 Implicit functions **30**

6.1 Implicit Function Theorem 1 . . . 30

6.2 Implicit Function Theorem 2 . . . 32

6.3 Implicit Function Theorem 3 . . . 32

7 Convexity and Concavity. Convex Sets **34**

7.1 $\mathbb{R}^2$ . . . 35

7.2 Appendix (don't read this unless you want to mess with your brain) . . . 38

7.3 What Does Determinant Have To Do With Anything? (don't read this even more) . . . 38

8 Unconstrained Optimization **40**

8.1 Local Optima . . . 40

8.2 Global Optima . . . 41

9 Constrained Optimization **42**

9.1 What the hell is NDCQ? . . . 43

9.2 Lagrange multiplier method . . . 44

9.3 Envelope theorem (unconstrained) . . . 45

9.4 Envelope theorem (constrained) . . . 46

1 Set Operations and Notation

Definition. A set is a collection of distinct objects.
Which means there are no repeats and order doesn't matter.

Example 1.

$$A = \{1, 2, 3\} = \{3, 1, 2\}$$

Example 2.

$$B = \{\underbrace{x}_{\text{all x}} \underbrace{\mid}_{\text{such that}} \underbrace{0 \leq x \leq 1}_{\text{x is bigger or equal than 0 and less or equal than 1}}\}$$

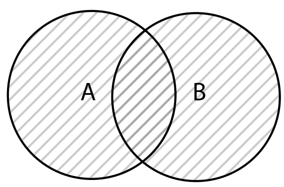
A has three members, while B has an infinite number of members. Thus, sets can either be finite or infinite.

1.1 Operations on sets

First, some notation:

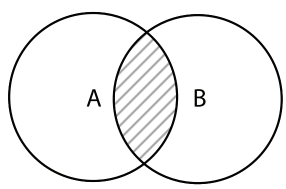
$a \in A$	element a belongs to a set A	$1 \in \{1, 2\}$
$B \subseteq A$	a set B is a subset of a set A	$\{1, 2\} \subseteq \{1, 2, 3, 4\}$, $\{1, 2\} \subseteq \{1, 2\}$
$\emptyset$	empty set	$A = \{\}$
2^A	power set: set of all subsets of a set	e.g. $A = \{1, 2\}$ $2^A = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$

Union of sets.



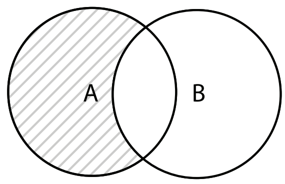
$$A \cup B = \{\underbrace{x}_{\text{all x}} \underbrace{\mid}_{\text{such that}} \underbrace{x \in A}_{\text{x belongs to A}} \underbrace{\text{or } x \in B}_{\text{or x belongs to B}}\}$$

Intersection of sets. $A \cap B$. Venn diagram:



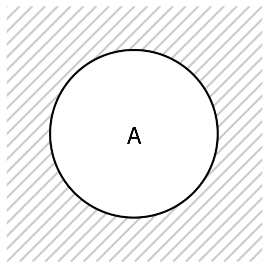
$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.
Verbally: elements that belong to both A and B.

Difference of sets. $A \setminus B$. Venn diagram:



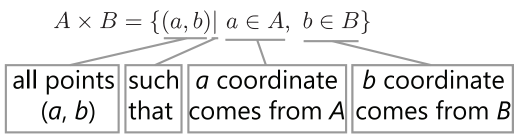
$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$.
Verbally: elements that belong to A and do not belong to B.

Complement or negation of a set. $\bar{A}$ or $\neg A$. Venn diagram:



$\neg A = \{x \mid x \notin A\}$.
Verbally: elements that do not belong to A.

Cartesian product of sets.



Example 1.

$$\{1, 2, 3\} \times \{3, 4\} = \{(1, 3), (1, 4), (2, 3), (2, 4), (3, 3), (3, 4)\}$$

Example 2.

$$\{3, 4\} \times \{1, 2, 3\} = \{(3, 1), (3, 2), (3, 3), (4, 1), (4, 2), (4, 3)\}.$$

1.2 Appendix

Fun fact: Russell's paradox Suppose a barber who shaves all men who do not shave themselves and only men who do not shave themselves. Does the barber shave himself? If he does, then he shouldn't. If he doesn't, then he should. Thus a paradox.

Stated more formally: Let R be the set of all sets that are not members of themselves ($R = \{x | x \notin x\}$). If R is not a member of itself, then its definition dictates that it must contain itself ($R \in R$), and if it contains itself, then it contradicts its own definition as the set of all sets that are not members of themselves ($R \notin R$).

2 Sequences in $\mathbb{R}^n$

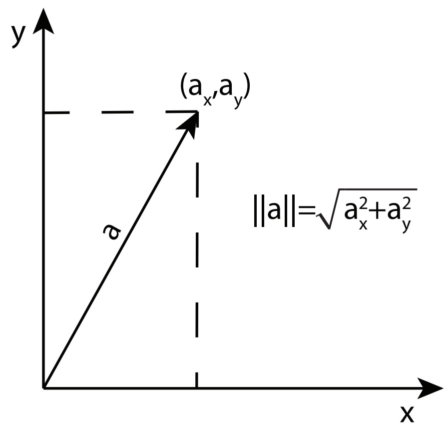
2.1 Vectors

Vectors a in $\mathbb{R}^n$ is given by $a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, where a_n is a nth coordinate of a vector. For example, in $\mathbb{R}^2$ vector $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_x \\ a_y \end{pmatrix}$.

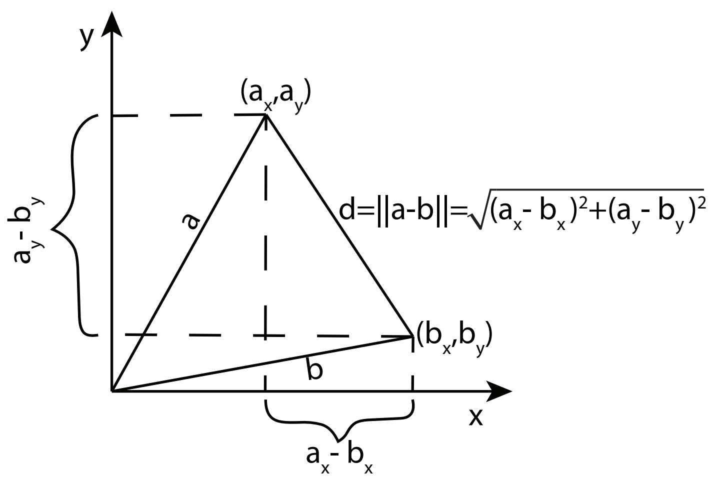
We add vectors together by adding each coordinate: $a + b = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ b_n + b_n \end{pmatrix}$.

2.1.1 Vectors in $\mathbb{R}^2$ (Euclidian space)

Length of a vector[1] a is denoted by $||a||$ and equals $\sqrt{a_x^2 + a_y^2}$ or in sum notation $\sqrt{\sum_{i=1}^{2} a_i^2}$. You already knew this in the form of Pythagoras' theorem:



Distance between vectors a and b is denoted by $||a-b||$ and is equal to $\sqrt{(a_x - b_x)^2 + (a_y - b_y)^2} = \sqrt{\sum_{i=1}^{2} (a_i - b_i)^2}$. As we can see from the picture, it's Pythagoras again:



[1]Technically, it's called *a norm of a vector* but you don't need to think about it.

2.1.2 Vectors in $\mathbb{R}^n$ (Euclidian space)

The way we count lenghts and distances in $\mathbb{R}^n$ is exactly the same as in $\mathbb{R}^2$ but we use more coordinates. Basically, Pythagoras' theorem but for arbitrary number of dimensions.

Definition. Length of a vector a in $\mathbb{R}^n$:

$$||a|| = \sqrt{\sum_{i=1}^{n} a_i^2}$$

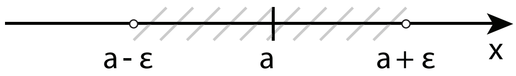
Definition. Distance between two vectors a and b in $\mathbb{R}^n$:

$$||a-b|| = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

2.2 Balls

2.2.1 Balls in $\mathbb{R}^1$

Ball in $\mathbb{R}^1$ is called *an interval.* An interval near the point a, $(a-\varepsilon,\ a+\varepsilon)$



written in math notation as

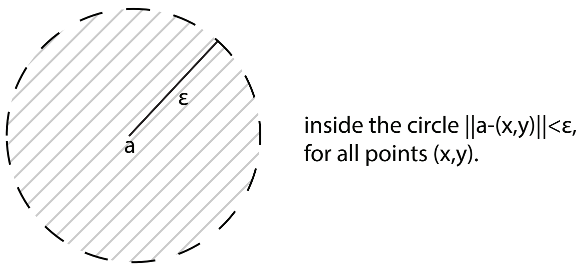
$$\{x \mid a-\varepsilon < x < a+\varepsilon\}$$

is given by

$$B_\varepsilon(a) = \left\{x \in \mathbb{R}^1 \mid ||a-x|| < \varepsilon\right\}$$

2.2.2 Balls in $\mathbb{R}^2$

Ball in $\mathbb{R}^2$ is called *a disk.* A disk around the point a with a radius ε:



written in math notation as

$$\left\{(x,y) \mid \sqrt{(a_x - x)^2 + (a_y - y)^2} < \varepsilon\right\}$$

is given by

$$B_\varepsilon(a) = \left\{x \in \mathbb{R}^2 \mid ||a-x|| < \varepsilon\right\}$$

2.2.3 Balls in $\mathbb{R}^n$

For a ball in $\mathbb{R}^n$ we have exactly the same definition as for disk in terms of $||a-x||<\varepsilon$:

Definition. Ball $B_\varepsilon(a)$ around the point a with a radius ε in $\mathbb{R}^n$ is given byki

$$B_\varepsilon(a)=\{x\in\mathbb{R}^n \mid ||a-x||<\varepsilon\}$$

2.3 Sequences and Their Limits

2.3.1 A sequence of points in $\mathbb{R}^n$

Fibonacci sequence, given by $1,1,2,3,5,8,13,...$ is an example of a sequence in $\mathbb{R}^1$, which is the type of sequences you explored in first year calculus. A sequence in $\mathbb{R}^n$ is an extension of this idea:

Fibonacci sequence	arbitrary sequence in $\mathbb{R}^1$	arbitrary sequence in $\mathbb{R}^n$
$a_1=1$	$a_1=\left(a_1^1\right)$	$a_1=\begin{pmatrix}a_1^1\\ \vdots\\ a_1^m\end{pmatrix}$
$a_2=1$	$a_2=\left(a_2^1\right)$	$a_2=\begin{pmatrix}a_2^1\\ \vdots\\ a_2^m\end{pmatrix}$
$a_3=2$	$a_3=\left(a_3^1\right)$	$a_3=\begin{pmatrix}a_3^1\\ \vdots\\ a_3^m\end{pmatrix}$
$a_n=a_{n-2}+a_{n-1}$	$a_n=\left(a_n^1\right)$	$a_n=\begin{pmatrix}a_n^1\\ \vdots\\ a_n^m\end{pmatrix}$

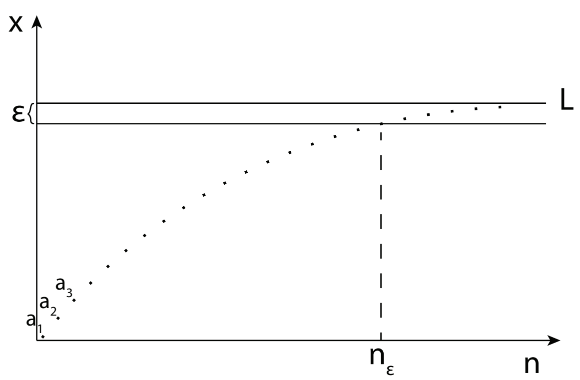
So, $a_n=\begin{pmatrix}a_n^1\\ \vdots\\ a_n^m\end{pmatrix}$ is a nth point of a sequence, a_n^i is an ith coordinate of a point, and the sequence itself is given by $\{a_n\}_{n=1}^{\infty}$.

2.3.2 Limit of a sequence in $\mathbb{R}^1$

In $\mathbb{R}^1$ a sequence has a limit L, if for any arbitrary small number $\varepsilon>0$, there's a point n_ε such that, for all points following n_ε, the sequence lies within the distance between its limit L and ε. Written formally,

$$\lim_{n\to\infty}a_n=L \text{ if } \forall\varepsilon>0\ \exists n_\varepsilon:\ \forall n>n_\varepsilon\ ||a_n-L||<\varepsilon$$

Graphically:



Note, that here the horizontal axis does not carry any meaning by itself: it simply shows the order of the sequence.

2.3.3 Limit of a sequence in $\mathbb{R}^2$

In $\mathbb{R}^2$, we need both coordinates to be close to L, and the following theorem should be fairly obvious:

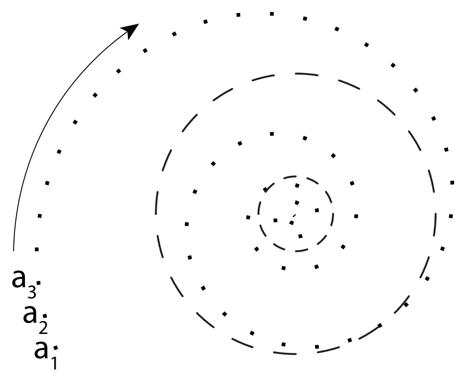
Theorem. A sequence $\{a_n\}_{n=1}^{\infty}$ converges to L if and only if each coordinate of a sequence $\left\{a_n^i\right\}_{n=1}^{\infty}$ converges to the corresponding coordinate L^i.

A natural way to define this, would be to pick an arbitrarily small area – an ε-ball – around L and see whether all points after a certain n_ε lie within this ball. So we modify the definition of a limit by making ε to be radius of a ball. Notice that the definition written in math notation didn't change:

$$\lim_{n\to\infty} a_n = L \text{ if } \forall\varepsilon > 0\ \exists n_\varepsilon :\ \forall n > n_\varepsilon\ ||a_n - L|| < \varepsilon$$

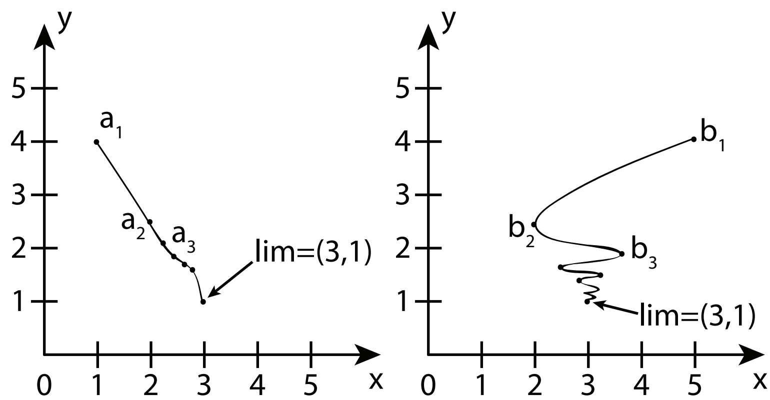
Next, let's see a couple of examples of converging sequences.

First, is a spiral sequence (try to define it explicitly :p). Imagine a ball around its center and start shrinking it. For any arbitrarily small ball, we will find a point on the spiral, such that all points after this point lie within the ball. Thus, spiral's center is its limit.



On the pictures below

left sequence (a)	right sequence (b)
$x_n = 3 - \frac{2}{n}$	$x_n = 3 - \frac{2}{n} \cdot (-1)^n$
$y_n = 1 + \frac{3}{n}$	$y_n = 1 + \frac{3}{n}$



(you can calculate several first values of these sequences to confirm that they are convergent)

2.3.4 Limit of a sequence in $\mathbb{R}^n$

Limit of a sequence in $\mathbb{R}^n$ is very similar to that in $\mathbb{R}^1$ and $\mathbb{R}^2$, except we can't really visualize it (except, perhaps, in $\mathbb{R}^3$), and the definition carries over from $\mathbb{R}^2$ completely.

Definition. The sequence $\{a_n\}_{n=1}^{\infty}$ in $\mathbb{R}^n$ converges to L if

$$\forall \varepsilon > 0 \ \exists n_\varepsilon : \ \forall n > n_\varepsilon \ ||a_n - L|| < \varepsilon$$

2.3.5 Accumulation points of a sequence

Suppose we have a sequence in $\mathbb{R}^1$ given by $a_n = (-1)^n$. Its behavior is pretty straightforward – the sequence jumps back and forth from -1 to 1 *ad infinitum.* We may be tempted to say it has two limits but that would only be half-right: we could apply the definition of a limit either to all odd points, or all even points – but not to both. Such points, where an infinite number, but not necesarrily all of them, lie within a ball $B_\varepsilon(a)$ are called *accumulation points of a sequence.* Note that the limit is always an accumulation point. Also, if the sequence has one accumulation point, then this point is the limit.

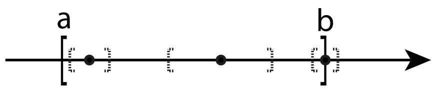
Although the concept of accumulation point of a sequence is rarely used, it will be helpful in understanding the limit of a function.

3 Sets

3.1 Open and Closed Sets

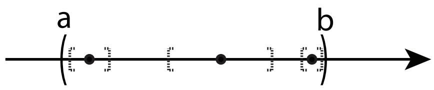
3.1.1 Sets (intervals) in $\mathbb{R}^1$

Back in high school you learned about types of intervals. Interval is called *closed*, if it includes its endpoints. Interval is called *open*, if it doesn't include its endpoints. To understand the difference between open and closed interval in a different way, let's try to draw some $\mathbb{R}^1$ balls (intervals, that is) on it:



We can't find a ball that would be drawn around endpoints a or b and that would lie inside the interval $[a, b]$ completely. For point b, right-hand side of a ball will necessarily be outside the interval; for point a – left-hand side.

Now consider an open interval (a, b):



In contrast to a closed interval, for *any* point in an open interval, we can find a ball which would lie inside the interval completely. We call such a point *internal:*

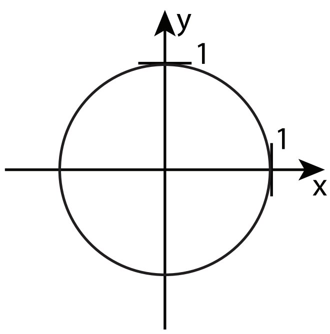
Definition. *Internal point* is a point such that we can draw an ε-Ball arount it, which would contain only points from the set.

Then, an open interval consists only of internal points. In fact, this observation is true for all open sets in $\mathbb{R}^n$. Formally, though, the definition of an open set is:

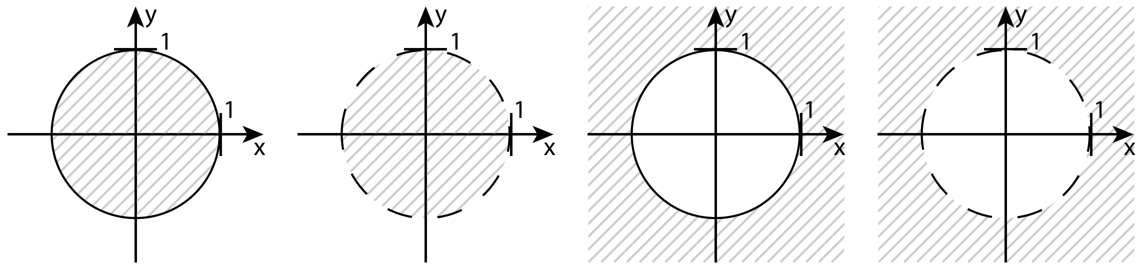
Definition. The set S is called *open,* if for its every point x, there exists an ε-Ball centered at x that lies in S completely.

3.1.2 Sets in $\mathbb{R}^2$

Consider a circle $x^2 + y^2 = 1$ on a plane:



The circle naturally divides the whole plane into 2 parts: inside of it and outside of it. However we also need to decide which part the circle itself belongs to: inside or outside. So there are four arrangements in total:



$x^2 + y^2 \leq 1$	$x^2 + y^2 < 1$	$x^2 + y^2 \geq 1$	$x^2 + y^2 > 1$
closed	open	closed	open
bounded	bounded	unbounded	unbounded

Note that the difference between open and closed sets is whether they contain their *boundary points*. Intuitively, boundary points lie on, duh, bounds, which implies that however small balls around these points we draw, some parts of them will neccesarily be inside the interval and some parts will be outside (check picture for closed interval above, if this is not obvious). Formally:

Definition. *Boundary point* is a point such that every ε-Ball around it contains points from a set and not from a set.

The way to remember the distinction between closed and open set is that closed contains *all* its boundary points, while open doesn't contain *any* of its boundary points (check the picture above with four circles to confirm). If a set contains some but not all its boundary points, then it's neither open nor closed.

Also note that two leftmost circles are bounded, while two rightmost are unbounded. Both visually and intuitively it's obvious: set is bounded if it doesn't extend to infinity in any direction (extending to infinity even along a single line in $\mathbb{R}^2$ is enough to become unbounded), so here's the formal definition:

Definition. Set is called *bounded* if it is contained within some ball.

Note that a bounded set doesn't need to be round itself. It simply needs to be able to be drawn into a ball, and as long as it doesn't include infinity in any direction, it's bounded.

3.1.3 Sets in $\mathbb{R}^n$

In $\mathbb{R}^n$ everything stays basically the same, except we need more axes.

Now, hopefully having acquired intuition, we move on to define closedness of a set formally. To do this let's get back to $\mathbb{R}^1$. Consider an open interval $(0, 1)$ and this sequence on it:

$$\begin{gathered} a_1 = 0.9 \\ a_2 = 0.99 \\ a_3 = 0.999 \\ \ldots \end{gathered}$$

Each member of the sequence is in the interval. However $\lim a_n = 1$, which is outside the interval. Having this sequence in mind, proceed to the definition of a closed set:

Definition. Set is called *closed,* if it contains the limit of any convergent sequence of elements from the set.

Note that $[0, 1]$ containts $\lim a_n$, as well as the limit of any other sequence, which consists of points on the interval. Thus we know that it's closed.

Example 1. Consider the interval $(0, +\infty)$. In $\mathbb{R}^1$ it is given by $\{x \mid x > 0\}$ and it's open and unbounded; in $\mathbb{R}^2$ it is given by $\{(x, y) \mid x > 0,\ y = 0\}$ it's still unbounded but it's no longer open, since line on a plane entirely consists of boundary points (if you try to imagine these sets, this will become obvious).

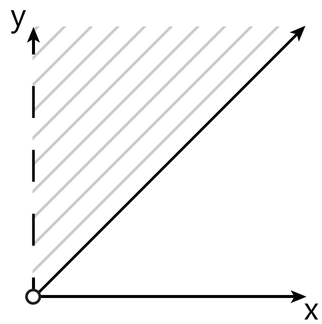
Example 2. (example taken from 31.10.11 mock)

Let D be the domain of the function $f(x, y) = \ln(x) + \sqrt{y - x}$. Find D, the set D^o of internal points of D, and the set ∂D of boundary points of D.

Solution. Domain is the function's all possible inputs, which means we have to ensure that $\ln(x)$ and $\sqrt{y - x}$ receive the correct inputs:

$$\ln(x) \Rightarrow x > 0$$
$$\sqrt{y - x} \Rightarrow y \geq x$$

So D is $\{(x, y) \mid x > 0 \text{ and } y \geq x\}$. Graphically:



From the picture we can see that internal points D^o are $\{(x, y) \mid x > 0 \text{ and } y > x\}$. Recalling the definition of a boundary point, which says that it is a point such that every ε-Ball around it contains points from a set and not from a set, we see that the vertical line $x = 0$ and the diagonal $y = x$ satisfy it. So boundary points ∂D are $\{(x, y) \mid x = 0 \text{ and } y \geq 0 \text{ or } x \geq 0 \text{ and } y = x\}$. Also, we can see that this set is neither open nor closed, as it contains some but not all of its boundary points.

Theorem. Complement of a closed set is an open set; complement of an open set is a closed set.

Example. Complement of a closed interval $[0, 1]$ is open interval $(-\infty, 0) \cup (1, +\infty)$; complement of and open interval $(0, 1)$ is closed interval $(-\infty, 0] \cup [1, +\infty)$.

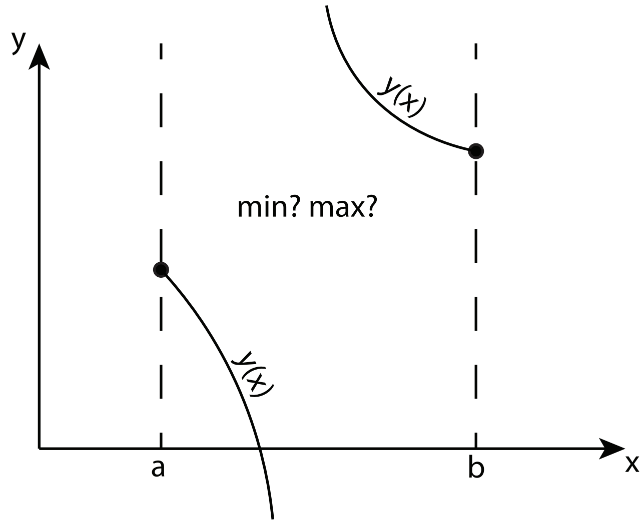
Fun fact. Empty set and its complement (entire line in $\mathbb{R}^1$; entire plane in $\mathbb{R}^2$) are the only sets on $\mathbb{R}^n$ that are simultaneously open and closed.

3.2 Compact Sets

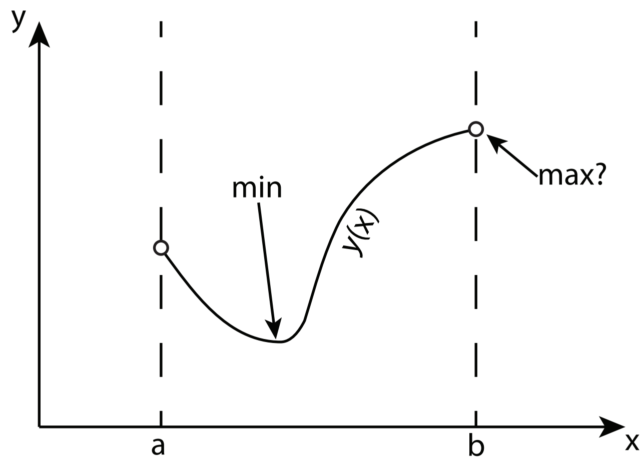
Now, recall the image of a set $\left\{(x, y) \mid x^2 + y^2 \leq 1\right\}$ and proceed to the following definition:

Definition. Set is called *compact* if it is both closed and bounded.

If you want to understand the significance of this definition, let me return to the first year calculus for a moment (you can skip this otherwise). Consider the problem of finding maximum and minimum of a function on an interval. Immediately we start thinking about first and perhaps second derivative. But wait, how do we know that min and max even exist? Consider this function, which is discontinuous on a closed interval $[a, b]$:

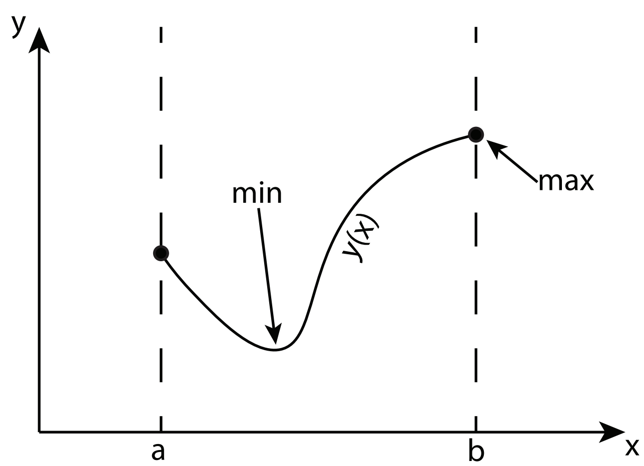


Pretty obvious that it has neither min, nor max on $[a, b]$. Well, maybe requiring function to be continuous would suffice? Then consider this function, which is continous on an open interval (a, b):



Minimum is attained somewhere between a and b, as we can see. But since b is not included, it can't be the point where f attaints its maximum. Where does it then? Somewhere really close to b, obviously. Let's call this point c. Now move to $c + \varepsilon$. We're still inside the interval as $c + \varepsilon < b$ but $f(c+\varepsilon) > f(c)$. So $f(c)$ is not the maximum. It's easy to see that there's in fact no such point, where f attains maximum.

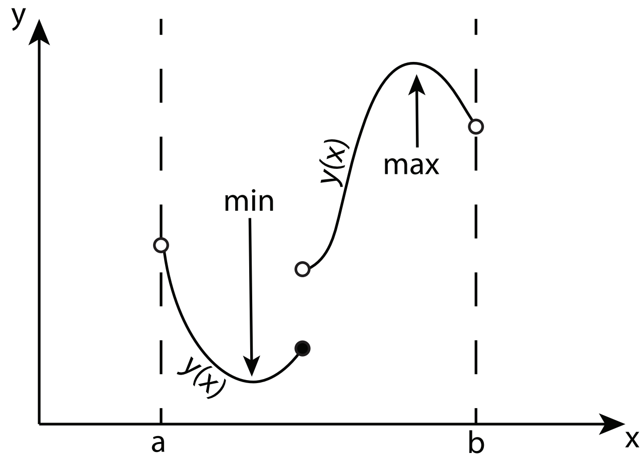
So it turns out, the function has to be both continuous and be defined on a closed interval for us to be confident that it attains both min and max.



This might remind you of the *Extreme Value Theorem*, and it actually is:

Extreme Value Theorem (EVT). If f is continuous on the closed interval $[a, b]$, then f attains its minimum and maximum values on $[a, b]$.

Note, however, that EVT defines *sufficient* conditions i.e. if EVT is satisfied, then min and max are attained for sure. However these conditions are not *necessary* for min and max to be attained. Consider this function, which is not continuous on an open interval:



Both extreme values exist. So the point of EVT is simply to provide a shortcut to us. If it is satisfied, extremes exist. If not – they may or they may not.

Okay, back to matec. Large part of the Math for Economists course is dedicated to the extension of the problem we examined above, except f becomes a function of several variables and constraints are much more complex than $a \leq x \leq b$. *Weierstrass Theorem* that will be introduced further in text, provides a similar shortcut for these cases. The difference is that instead of function needed to be continuous and defined on a closed interval, it will have to be continous and be defined on a *compact* set i.e. a set that is both closed and bounded.

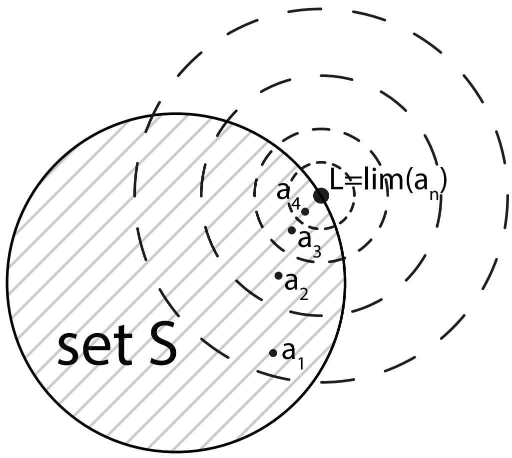
3.3 Appendix 1

Unless you're very comfortable with definitions of open and closed sets, the following is pretty hard to understand. If you don't get it from the first time, try to absorb as much as possible initially and then reread this a couple of days later.

Earlier I wrote that the easy way to remember the distinction between open and closed sets it this:

1. Open set doesn't contain any of its boundary points.
2. Closed set containts all its boundary points.

How is this heuristic connected to the formal definition of open and closed sets? Recall from the definition that a closed set must contain the limit of every convergent sequence, consisting of points from the set. If, for every boundary point, we could find such a sequence that would converge to this boundary point, then, by definition, we would show that a closed set contains all its boundary points.



Note that L is a boundary point, which means there are both points belonging to S and points outside S arbitrarily close to L (in any ε-Ball around L). Next just pick a sequence $\{a_n\}$ from S, such that each next term would lie in an ε-Ball with smaller and smaller radius around L, thus having L as its limit. Now L must belong to S by definition of a closed set. Finally, repeat the process for all boundary points of S.

For open sets it's simpler: definition of an open set basically says that it only contains internal points (since around each point you can draw a ball that would reside inside the set completely), which is pretty much equivalent to saying that it doesn't contain any of its boundary points.

3.4 Appendix 2

Think about the following questions for a few moments, before reading the answers:

1. Consider a finite union of closed sets. Is it open or closed?
2. Consider any (finite or infinite) intersection of closed sets. Is it open or closed?
3. Consider any (finite or infinite) union of open sets. Is it open or closed?
4. Consider a finite intersection of open sets. Is it open or closed?
5. Are all points either boundary or interior?

3.4.1 Answers.

Theorem (questions 1 and 2).

1. A finite union of closed sets is a closed set.

Counterexample for infinite union of closed sets that forms an open set: $\bigcup\limits_{n=1}^{\infty} \left[1 + \frac{1}{n}, 2 - \frac{1}{n}\right] = (1, 2)$.

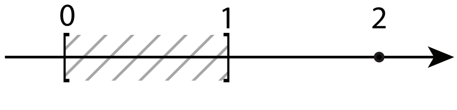
2. Any (finite or infinite) intersection of closed sets is a closed set.

Theorem (questions 3 and 4).

1. Any (finite or infinite) union of open sets is an open set.
2. A finite intersection of open sets is an open set.

Counterexample for infinite intersection of open sets that forms a closed set: $\bigcap\limits_{n=1}^{\infty} \left(1 - \frac{1}{n}, 2 + \frac{1}{n}\right) = [1, 2]$.

Question 5. Consider a set $\{[0, 1] \cup \{2\}\}$ on $\mathbb{R}^1$:



Is point (2) boundary or interior? If you check with the definitions, it's neither. Points like this are called *isolated*. Thus, there are three kinds of points, and boundary and internal points are not complements.

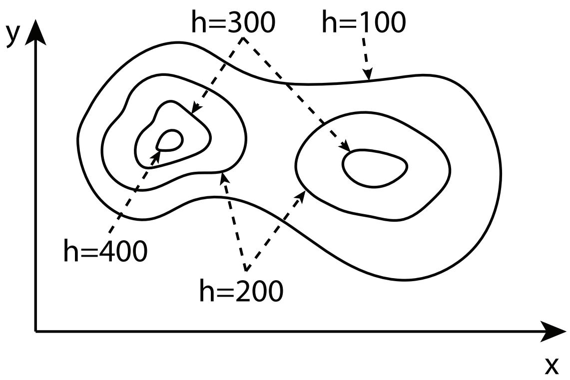
4 Multivariable Functions. Continuity.

Functions and sequences are basically the same things, except sequences are discrete (they're defined on the set of natural numbers and we can number each term of a sequence: $1, 2, 3, ...$) , while functions are defined on the set of real numbers, which are unenumerable. One could say that sequence is a function on $\mathbb{N}$.

4.1 Level curves

Drawing functions of one variable is okay; drawing functions of two variables is hard. Which is why when we have a function of two variables, we frequently try to visualise it on a usual 2-axis graph.

Suppose we have a function which shows height above the sea level of some piece of land $h = f(x, y)$. Then we can draw all points (x, y) for which h is constant, then vary it, here's what happens with the graph:



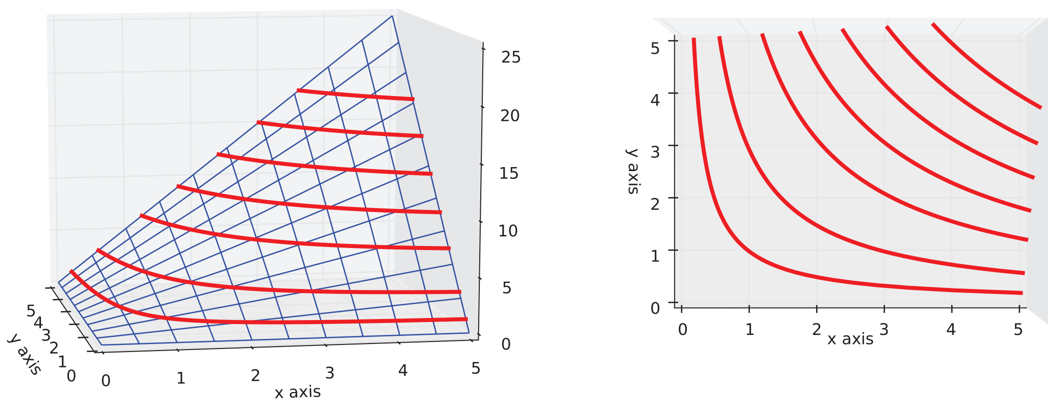
The lines on the graph are *level curves* of the function $h = f(x, y)$. For example, for $h = 100$, a level curve is given by

$$\{(x, y) \mid f(x, y) = 100\}$$

Definition. *C-level curve* (curve of level c) of a function $h = f(x, y)$ for some level c is given by

$$\{(x, y) \mid f(x, y) = c\}$$

Note that the indifference curve from micro is a level curve in this course:



4.2 Limit of a Function

МИЭФ — это такое уникальное место, где три раза рассказывают
о том, что такое предел, и каждый раз берут за это шестьсот тысяч.
Алексей Ахметшин

Limit of a function is pretty much the same thing as limit of a sequence (check section 2.3.2 on page 7 for explanation). In fact, as mentioned earlier, we could view sequences a special kind of functions for which the only values are $f(1)$, $f(2)$, $f(3)$, and so on.

You'll probably never be asked to actually employ the definition of a limit, but you definitely need to understand it conceptually, to be able to prove either existence of a limit or absense of a limit of a function at a point.

While calculating limits of multivariable functions, we can make the same operations as for single-variable functions (lim of a sum, product, quotinent). The problem appears when we try to deal with uncertanties ($\frac{\infty}{\infty}$ and $\frac{0}{0}$): L'Hospital's rule doesn't for multivariable functions. This means that when faced with uncertainty have to find other ways around it. This chapter shows some of the most common techniques.

Example 1: multiplication by the conjugate

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{xy}{3-\sqrt{xy+9}} =$$

Multiply the fraction by $\left(3+\sqrt{xy+9}\right)$ and apply $(a+b)(a-b)=a^2-b^2$ to the denominator:

$$= \lim_{\substack{x\to 0\\ y\to 0}} \frac{xy\left(3+\sqrt{xy+9}\right)}{9-(xy+9)} = \lim_{\substack{x\to 0\\ y\to 0}} \frac{3+\sqrt{xy+9}}{-1} = -6$$

Example 2: change of variables and equivalences

$$\lim_{\substack{x\to 0\\ y\to -3}} \frac{\ln(1+xy)}{x} = \lim_{\substack{x\to 0\\ y\to -3}} \frac{\ln(1+xy)y}{xy} =$$

Substitute $z = xy \to 0$ and apply $\lim(f\cdot g) = \lim(f)\cdot\lim(g)$:

$$\lim_{\substack{z\to 0\\ y\to -3}} \frac{\ln(1+z)}{z}\cdot y =$$

Recall that for $t\to 0$, $\ln(1+t)\sim t$:

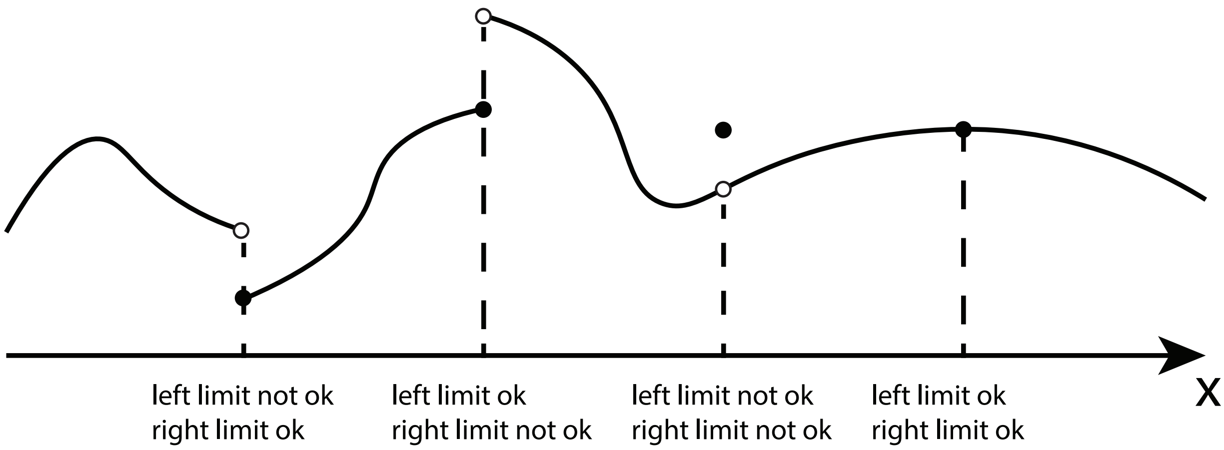
$$\lim_{\substack{z\to 0\\ y\to -3}} \left(\frac{z}{z}\cdot y\right) = -3$$

4.3 Continuity

4.3.1 Continuity on $\mathbb{R}^1$

Definition. $f(x)$ is *continuous* around point x_0, if $\lim\limits_{x\to x_0^-} f(x) = f(x_0)$ and $\lim\limits_{x\to x_0^+} f(x) = f(x_0)$.

Check the picture below to understand why we need these conditions:



4.3.2 Continuity on $\mathbb{R}^2$ and $\mathbb{R}^n$

While there's only two ways to approach a point on the line – from left or right – on a plane (and in higher dimensions) there's an infinite number of directions to do that, and limit's definition extends to accomodate this fact:

Definition. $f(x)$ is *continuous* around point x_0, if $\lim\limits_{x \to x_0} f(x) = f(x_0)$.

Now in order to prove existence of a limit we need to check all possible directions. To prove that the limit doesn't exist, we just need to show that while approaching the point from two different paths (a parallel to sequences: we can prove that the limit of a sequence doesn't exist by showing that it has two accumulation points), function approaches different values. Most often we use the following technique, when we want to show that the limit doesn't exist:

1. Approach the point along the line $y = kx$
2. Approach the point along the parabolic curve $y = kx^2$
3. And so on.

Usually, just checking $y = kx$ is enough. Checking $y = kx^2$ is almost always enough. But nothing hypothetically stops Demeshev or Bukin from coming up with a function, where you need to check $y = kx^{15}$ or something, to prove that the limit doesn't exist.

Exam tip: the main difficulty, when solving such a problem, is to recognize that the limit doesn't exist and not waste time trying to find it.

Example 1. Find the limit of the following function, as $x \to \infty$, $y \to \infty$ or prove that it doesn't exist:

$$f(x, y) = \frac{x^2 + y^4}{x^4 + y^2}$$

Solution: Look at directions $y = kx$:

$$\lim_{x \to \infty} \frac{x^2 + k^4x^4}{x^4 + k^2x^2} = \lim_{x \to \infty} \frac{x^2(1 + k^4x^2)}{x^2(x^2 + k^2)} = \lim_{x \to \infty} \frac{1 + k^4x^2}{x^2 + k^2}$$

Dropping 1 and k^2, as they're bounded and won't matter, we get:

$$\lim_{x \to \infty} \frac{k^4x^2}{x^2} = k^4$$

Which means that the lim of $f(x, y)$ depends on the line on which we approach ∞. Moving along the line $y = 1 \cdot x$, $f(x, y) \to 1$; moving along the line $y = 5 \cdot x$, $f(x, y) \to 625$. Thus $\lim\limits_{x \to \infty} f(x, y)$ doesn't exist.

Example 2. Find the limit of the following function, as $x \to \infty$, $y \to \infty$ or prove that it doesn't exist:

$$f(x, y) = \frac{x^2 y}{x^4 + y^2}$$

Solution: Look at directions $y = kx$:

$$\lim_{x\to\infty} \frac{kx^3}{x^4 + k^2x^2} = \lim_{x\to\infty} \frac{kx^3}{x^2(x^2 + k^2)} = \lim_{x\to\infty} \frac{kx}{x^2 + k^2} = \lim_{x\to\infty} \frac{kx}{x^2} = 0$$

Wait-wait-wait; what if we use parabolas $y = kx^2$?

$$\lim_{x\to\infty} \frac{kx^4}{x^4 + k^2x^4} = \lim_{x\to\infty} \frac{kx^4}{x^4(1 + k^2)} = \frac{k}{1 + k^2}$$

So actually limit doesn't exist! $y = kx$ just couldn't provide us with the right curve.

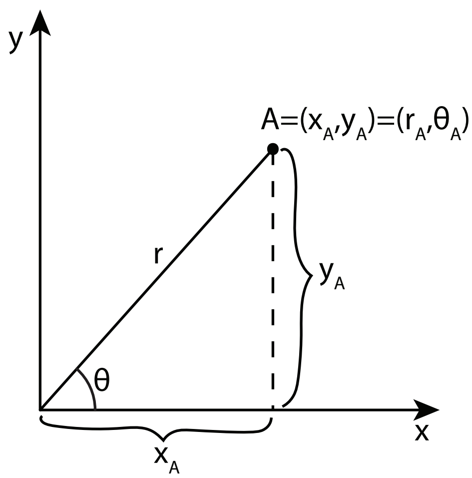
Protip: by checking some, not all directions (such as $y = kx$ in this example) we can only prove that the limit does not exist (if we find different limits along different directions). Finding the "limit" with only some directions doesn't tell us anything about the existence of the actual limit!

Example 3. Find the limit of the following function, as $x \to \infty$, $y \to \infty$ or prove that it doesn't exist:

$$f(x, y) = \frac{x^{15}}{y}$$

4.4 Finding Limits with Polar Coordinates

Usually we define the point on a plane using the Cartesian system with x and y axes. An alternative way to uniquely identify the point is by the distance from the origin and an angle:



To refresh our memory: $\sin\theta = \frac{y}{r}$, $\cos\theta = \frac{x}{r}$, $\tan\theta = \frac{y}{x}$.

So $r = \sqrt{x_A^2 + y_A^2}$, while $\theta = \arctan\left(\frac{y_A}{x_A}\right)$.

The inverse conversion is $x = r \cdot \cos\theta$, $y = r \cdot \sin\theta$.

Protip: Polar coordinates are very useful when we deal with $x^2 + y^2$ in limits.

Example 1. This example just shows explicitly how change of the coordiates to polar works:

$$\lim_{\substack{x\to 3\\ y\to 4}} f(x,y) = \lim_{\substack{r\to 5\\ \theta\to\arctan\left(\frac{4}{3}\right)}} f(r\cos\theta, r\sin\theta) = \lim_{\substack{r\to 5\\ \theta\to\arctan\left(\frac{4}{3}\right)}} g(r,\theta)$$

The question that might pop up is why do we switch g, rather than continue working with f in polar? Consider $f(x,y) = x+y$. Then $f(r,\theta) = r+\theta$. That's hardly what we wanted to achieve, so we introduce $g(r,\theta) = r\cos\theta + r\sin\theta$.

Example 2.

$$\lim_{\substack{x\to 0\\ y\to 0}} f(x,y) = \lim_{\substack{r\to 0\\ \theta\to any}} g(r,\theta)$$

since when $r=0$ we don't have any information about the angle.

Example 3. (example taken from 25.10.12 mock)

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{x^3+y^3}{x^2+y^2} = \lim_{\substack{r\to 0\\ \theta\to any}} \frac{r^3\cos^3\theta + r^3\sin^3\theta}{r^2\cos^2\theta + r^2\sin^2\theta} =$$

Using $\cos^2 x + \sin^2 x = 1$

$$= \lim_{\substack{r\to 0\\ \theta\to any}} r(\cos^3\theta + \sin^3\theta) = 0$$

Since both cos and sin are restricted by -1 and 1, and are therefore bounded, $\cos^3\theta + \sin^3\theta$ won't actually matter.

Notice that we could see from the very beginning that the limit is equal to 0, as, close to the origin, x^3 and y^3 are much smaller than x^2 and y^2.

The result about unimportance of bounded values, when dealing with infinities is mostly obvious but still there's a theorem for it:

Theorem. Limit of a product of infinitely big number and a bounded number is infinity: $+\infty \cdot c = \infty$ and $-\infty \cdot c = \infty$ (here, by ∞, I mean either $+\infty$ or $-\infty$, depending on the sign of c). Limit of a product of an infinitely small number and a bounded number is zero: $0 \cdot c = 0$.

Example 4.

$$\lim_{\substack{x\to 0\\ y\to 0}} \frac{x^2y^2}{(x^2+y^2)^2} = \lim_{\substack{r\to 0\\ \theta\to \text{any}}} \frac{r^2\cos^2\theta\sin^2\theta}{r^2} = \cos^2\theta\sin^2\theta$$

Since both cos and sin change with a change in θ, and we can approach point $(0,0)$ at any angle θ, there is no limit of this function.

5 Differentiation of Multivariable Functions. Approximation

5.1 Taylor Series Introduction

Let's start with "real-life" example. Imagine a car. We know that the car's position at a time t_0 is $s(t_0) = s_0$. However, we don't know its speed. Neither do we know its acceleration. If we are asked about car's position at a time t_1 what do we say? The only value we can use is s_0, so there's no choice but to say that $s(t_1) \approx s_0$. Okay. Now, we suddenly learn about the car's speed at a moment t_0: $v = v_0$. What do we say s_1 is now? Since we don't know anything about car's acceleration, we'll just have to assume its speed is constant. Then $s(t_1) \approx s_0 + v_0(t_1 - t_0)$ i.e. car's initial position and what we assume it drove during the period between t_0 and t_1. Okay, better. But what if we know car's acceleration at t_0: $a = a_0$ as well? Surely we want to use this information, but how do we do it?

In high school physics you were just given a formula:

$$s_1 = s_0 + vt + \frac{at^2}{2}$$

Today you learn that this formula is a special case of *Taylor series*.

Back to math. Car's speed is the rate of change of its position, therefore it's the first derivative $s'(t)$ of $s(t)$. Car's acceleration is the rate of change of its speed, therefore it's the second derivative $s''(t)$ of $s(t)$. What if car's acceleration varies as well? And its acceleration? And so on? Only using $s''(t)$ would be a waste of all the derivatives that follow it.

Okay, here's the formula:

$$s(t_1) \approx s_0 + s'(t_0)(t_1 - t_0) + \frac{s''(t_0)(t_1 - t_0)^2}{2} + \frac{s'''(t_0)(t_1 - t_0)^3}{6} + \ldots$$

In a more general form:

$$f(x) \approx f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{f^{(2)}(x_0)(x - x_0)^2}{2!} + \frac{f^{(3)}(x_0)(x - x_0)^3}{3!} + \ldots$$

And in the most general form possible:

$$f(x) \approx \frac{f^{(0)}(x_0)(x - x_0)^0}{0!} + \frac{f^{(1)}(x_0)(x - x_0)^1}{1!} + \frac{f^{(2)}(x_0)(x - x_0)^2}{2!} + \ldots$$

where $f(x) = f^{(0)}(x)$, $f'(x) = f^{(1)}(x)$ and so on. $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n = \prod_{n=1}^{n} n$ and $0! = 1$.

Definition. *Taylor series* of a function is given by

$$f(x) \approx \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)(x - x_0)^n}{n!}$$

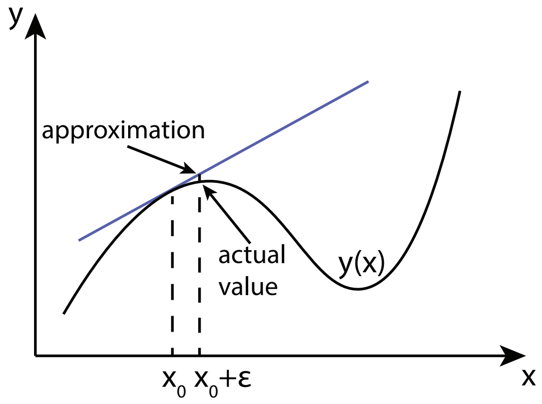
The thing is, it's really hard to understand exactly how we get this formula (I don't really get it myself; blame Akhmetshin :p). If you're interested, Wikipedia has a really great article on Taylor series (link). **But you probably want to memorize the formula and be able to use it when asked.** Taylor series turns up everywhere!

As a rule of thumb, the more derivatives we use, the better approximation is. However this is not universal and even using the infinite number of derivatives does not guarantee the convergence to the true value. Fortunately, within the course the functions are all so nice, we can forget about this and just use Taylor series blindly.

Some terminology: the case when we used car's speed – first derivative – was an instance of *first-order approximation*. The case when we used its acceleration – second derivative – was an instance of *second-order approximation*. These are the only two cases we're going to look in deeply during this course.

5.2 First-Order (Linear) Approximation for One Variable

Back to normal functions. Hopefully, the idea of using derivatives to approximate functions is pretty intuitive now. If we have a function of one variable such that calculating its value at x_0 is trivial, while doing the same thing at $x_0 + \varepsilon$ is nearly impossible, use approximation, usually first-order:



Recall the graphical interpretation of the derivative: it is the slope of the function at a point (or of its tangent line at a point). Equivalently it is shown by the change in f given $dx = 1$. The closer to x_0 we are, the less the slope changes and the more accurate the approximation is. For one variable, generalized form of first-order approximation is:

$$f(x) \approx f(x_0) + f'(x_0)dx = f(x_0) + f'(x_0)(x - x_0)$$

5.3 First-Order (Linear) Approximation for Two Variables

We want to generalize the method of using derivative to approximate a function to the functions of two variables:

$$z = f(x, y)$$

In order to do this, we'll decompose total change of function's value into change due to change in x : dx and change due to change in y : dy.

First with dx. To isolate change of z due to dx we need to fix y i.e. take y as if it were a constant.

$$z = f(x, y_0)$$

Then take this function's derivative, which is the rate of change of the function f along the line $y = y_0$:

$$z'_x = f'(x, y_0)$$

At a specific point (x_0, y_0) it is called *partial derivative*.

Definition. If it exists[2], *partial derivative* of z with respect to x is given by

$$f'_x(x_0, y_0) = f'_x = \frac{\partial z}{\partial x}$$

Note that partial derivatives are denoted with ∂, rather than d. Change in z due to change in x is

$$\frac{\partial z}{\partial x} dx = f'_x dx$$

Now, repeat the same operation for dy.

[2]This is almost always the case within this course, but we can actually come up with a simple looking function that is not differentiable, e.g. $y = (x^2)^{0.5}$, which we usually write as $y = |x|$.

$$z = f(x_0, y)$$

$$f'_y = \frac{\partial z}{\partial y}$$

And change in z due to change in y is

$$\frac{\partial z}{\partial y}dy = f'_y dy$$

Finallly, total change in z equals change due to dx and change due to dy. By combining these we get z's *first total differential.*

Definition. *First total differential* of a function of two variables is given by

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial x}dy = f'_x dx + f'_y dy$$

And we can use this result to approximate z close to (x_0, y_0):

$$z = f(x, y) \approx f(x_0, y_0) + f'_x \cdot (x - x_0) + f'_y \cdot (y - y_0)$$

5.4 Tangent Plane

If in $\mathbb{R}^1$ we find linear approximation of function's value by its tangent line, in $\mathbb{R}^2$ it is tangent plane.

Definition. Tangent plane for a function of two variables is given by

$$z = f(x_0, y_0) + f'_x \cdot (x - x_0) + f'_y \cdot (y - y_0)$$

5.5 Directional Derivative

Directional derivative shows the rate of change of a function in a particular direction we picked. It is a concept which can be thought of as a special case of the first total differential. Key difference between them is that total differential is function of arbitrary dx and dy, while, for directional derivative, length of a vector is always 1, in other words, dx and dy become dependent on each other and $dx^2 + dy^2 = 1$ (check section 2.1.1 on page 5 for explanation). We call such a vector *normalized* or a *unit vector.*

Definition. *Directional derivative* gives the rate of change of $f(x, y)$ at a point (x_0, y_0) in the direction of a unit vector $\vec{u}$ (vector of length 1, where $dx^2 + dy^2 = u_1^2 + u_2^2 = 1$). Its formula is:

$$D_{\vec{u}} f(x, y) = f'_x \cdot u_1 + f'_y \cdot u_2 = \begin{pmatrix} f'_x & f'_y \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

At a point is important because we calculate f'_x and f'_x at a specific point and plug in concrete numbers. Alternatively, if length of a vector is not equal to 1 (such vector is usually denoted as $\vec{l}$), the directional derivative is

$$D_{\vec{l}} f(x, y) = \frac{f'_x \cdot l_1 + f'_y \cdot l_2}{\sqrt{l_1^2 + l_2^2}} = \begin{pmatrix} f'_x & f'_y \end{pmatrix} \cdot \frac{\vec{l}}{||l||}$$

This implies that whether we pick vector $\begin{pmatrix} 5 \\ 3 \end{pmatrix}$ or $\begin{pmatrix} 15 \\ 9 \end{pmatrix}$ directional derivative stays the same. Only change in dy/dx will change it.

5.6 Gradient

Which dx and dy should we pick to achieve the maximum growth of a function (assuming vector length is 1 for simplicity)? The problem is:

$$f'_x \cdot dx + f'_y \cdot dy \to max$$

Note that this is a dot product of vectors $\begin{pmatrix} f'_x \\ f'_y \end{pmatrix}$ and $\begin{pmatrix} dx \\ dy \end{pmatrix}$, since another formulation of dot product is $||\vec{a}|| \cdot ||\vec{b}|| \cdot \cos\alpha$ and cos is maximum ($= 1$) when $\alpha = 0$, it is maximum when vectors are codirected. Thus, to maximize function's growth rate we pick $\begin{pmatrix} dx \\ dy \end{pmatrix}$ such that it is codirected with $\begin{pmatrix} f'_x \\ f'_y \end{pmatrix}$. This, in turn, means that $\begin{pmatrix} f'_x \\ f'_y \end{pmatrix}$ itself points in the direction of maximum growth of the function. This vector is called the *gradient* of the function.

Definition. *Gradient* of a function $f(x, y)$ is given by

$$\vec{\nabla} f(x, y) = \begin{pmatrix} f'_x \\ f'_y \end{pmatrix}$$

Which is simply the vector of partial derivatives of a function. Also, now you can see that we can reformulate directional derivative using the gradient *at a specific point*:

$$D_{\vec{l}} f = \frac{\vec{\nabla} f \cdot \vec{l}}{||\vec{l}\,||}$$

Protip: Remember firmly these three key properties of a gradient, as they're very frequently helpful in the exams. If you are too lazy to memorize all of them, remember property 3.

1. Gradient points in the direction of the most rapid growth of the function (discussed above).
2. Length of the gradient is equal to the maximum rate of growth of the function. Proof:

   $$D_{\vec{l}} f = \frac{\vec{\nabla} f \cdot \vec{l}}{||\vec{l}\,||} = \frac{||\vec{\nabla} f|| \cdot ||\vec{l}\,|| \cdot \cos\alpha}{||\vec{l}\,||} = ||\vec{\nabla} f|| \cdot \cos\alpha = ||\vec{\nabla} f||$$

   $\cos\alpha = 1$ from the derivation of a gradient above.
3. Gradient is orthogonal to the level curve.

Example. (taken from ??.11.2008 mock)
Calculate the directional derivative of the function $f(x, y) = 2x^3 + 2y^2$ at the point $A\,(1, 2)$ in the following directions:

a) $\vec{l} = (1, 3)$

b) $\vec{l}$ which is orthogonal to the curve given by the equation $x^2 + y^2 = 5$

c) Direction of the fastest growth of $f(x, y)$

Solution. In order to learn anything at all about a function, we'll need to know its partial derivatives, so:

$$f'_x = 6x^2 = 6$$
$$f'_y = 4y = 8$$

a) $D = \frac{6 \cdot 1 + 8 \cdot 3}{\sqrt{1^2+3^2}} = \frac{30}{\sqrt{10}}$

b) "Orthogonal" is a synonym to "normal" for xy-plane, and from Jeffrey we remember that the equation of the normal line is $y = f(x_0) + \frac{-1}{f'(x_0)(x-x_0)}$, and it's slope is $\frac{-1}{f'(x_0)}$. Then we need to find f':

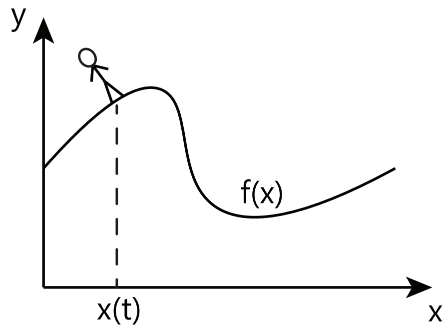
$$y' = -\frac{F'_x}{F'_y} = -\frac{2x}{2y} = -\frac{1}{2}$$

Orthogonal to which is $\frac{1}{-\frac{1}{2}} = 2$. Now just pick any vector with dy twice of dx, e.g. $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and calculate directional derivative in its direction: $D = \frac{6 \cdot 1 + 8 \cdot 2}{\sqrt{1^2+2^2}} = \frac{22}{\sqrt{5}}$

c) Direction of the fastest growth is the direction of the gradient i.e. $\begin{pmatrix} 6 \\ 8 \end{pmatrix}$: $D = \frac{6 \cdot 6 + 8 \cdot 8}{\sqrt{6^2+8^2}} = 10$

5.7 Chain Rule

Suppose we have a one-dimensional mountain, height of which at every point x is given by $f(x)$, and a hiker walking on it, whose coordinate at a time t is given by $x(t)$:



Then, if we want to learn the height of the hiker at some time t, the function we work with is $f(x(t))$. This was sort of a justification for the existence of the chain rule but this is where the real world example ends. If you'd like to learn more, Math Insight has a great page about it: http://mathinsight.org/chain_rule_multivariable_introduction.

5.7.1 $f(x(t))$

Suppose we have a function $f(x(t))$ and we want to find its first total differential df. Change in f is equal to derivative of f multiplied by the change in the argument:

$$df = \frac{df}{dx} dx$$

But since x depends on t, we can't just leave dx be, and dx becomes

$$dx = \frac{dx}{dt} dt$$

Then

$$df = \frac{df}{dx} \left(\frac{dx}{dt} dt \right)$$

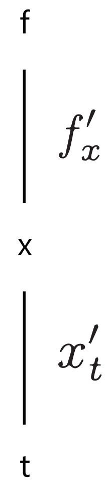
Transfering dt to the other side, we find the derivative $\frac{df}{dt}$:

$$\frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}$$

Alternatively:

$$f_t' = f_x' \cdot x_t'$$

We can draw a diagram[3] to help us understand this process. It seems complicated and unnecessary right now, but it will help a lot with more complex functions:



Next, just multiply both terms by flowing downwards, to get the exact same result, as written above.

5.7.2 $f(x(t), y(t))$

Recall that once we start to deal with functions of several variables, e.g. $f(x, y)$, df needs to be decomposed into change due to dx and change due to dy. Note that we use ∂'s here, because $\frac{\partial f}{dx}$ and $\frac{\partial f}{dy}$ are partial derivatives:

$$df = \frac{\partial f}{dx}dx + \frac{\partial f}{dy}dy$$

But when x and y are dependent variables, e.g. $f(x(t), y(t))$, we need to count that in, and df becomes:

$$df = \frac{\partial f}{dx}\left(\frac{dx}{dt}dt\right) + \frac{\partial f}{dx}\left(\frac{dy}{dt}dt\right)$$

Transferring dt to the other side we get:

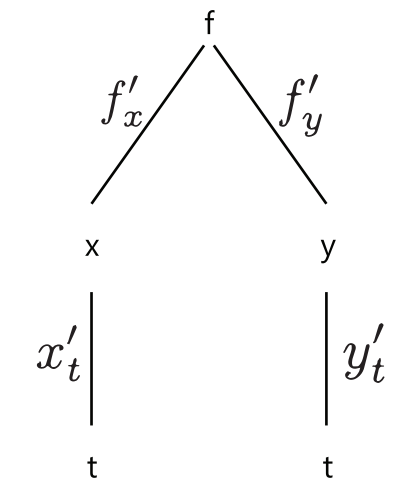
$$\frac{df}{dt} = \frac{\partial f}{dx}\frac{dx}{dt} + \frac{\partial f}{dy}\frac{dy}{dt}$$

Alternatively:

$$f_t' = f_x' \cdot x_t' + f_y' \cdot y_t'$$

However, we could get exactly the same result by drawing a diagram:

[3] Idea by Paul Dawkins: http://tutorial.math.lamar.edu/Classes/CalcIII/ChainRule.aspx

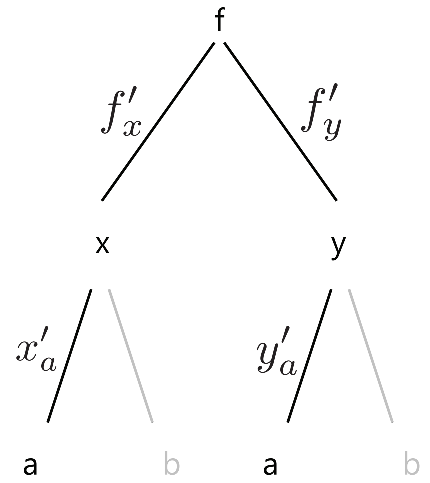


Now, we add up both branches to each other and get precisely:

$$f'_x \cdot x'_t + f'_y \cdot y'_t$$

5.7.3 $f\left(x(a,b), y(a,b)\right)$

Let's find a partial derivative f'_a of the function $f\left(x(a,b), y(a,b)\right)$, using a diagram. Here, we're only interested in the branches that end with a (I greyed out branches we don't need):



After multiplying each element of black branches and then adding up branches to each other, the result is:

$$f'_a = f'_x \cdot x'_a + f'_y \cdot y'_a$$

By analogy we can get f'_b. Also, by analogy we can work with more complex functions with the help of diagrams.

Example. (taken from 29.12.2011 mock)

The function $f(x,y)$ is given by $f(x,y) = u^2(x,y) + v^3(x,y)$. The value of u and v and their respective gradients at the point $(x,y) = (1,1,)$ are also known, $u(1,1) = 3$, $v(1,1) = -2$, $\nabla u(1,1) = (1,4)$, $\nabla v(1,1) = (-1,1)$. Find $\nabla f(1,1)$ if $u, v \in C^1$.

Solution.

$$\nabla f = (\nabla u)^2 + (\nabla v)^3 = \left(1^2 + (-1)^2\,,\ 4^3 + 1^3\right) = (2,65)$$

If you found yourself nodding along and the equation above did not raise any red flags you should stop immediately and try to understand why the thing I just did was completely wrong. Correct solution is after the following appendix.

5.8 Appendix to Chain Rule

If Bukin or Demeshev feel particularly sadistic when composing the exam, they might come up with something like this:

Example. (taken from 21.01.2009 mock)

Calculate all partial derivatives of the first and second order of u with respect to x and y if $u = f(a,b)$ and $a = x + xy,\ b = x/y$.

The first thing to do here is to rewrite $u = f(a,b)$ as $u = f\left(a(x,y), b(x,y)\right)$ to better understand the task and not bother with calculations for now. Let's focus on u'_x:

$$u'_x = f'_a \cdot a'_x + f'_b \cdot b'_x$$

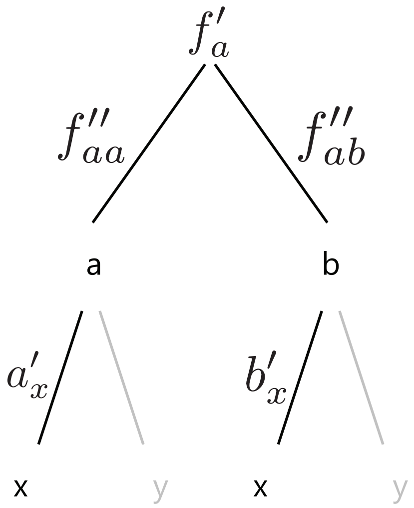
Now, we move on to the second-order derivatives:

$$u''_{xx} = \left(f''_{aa} \cdot a'_x + f''_{ab} \cdot b'_x\right) a'_x + f'_a \cdot a''_{xx} + \left(f''_{ba} \cdot a'_x + f''_{bb} \cdot b'_x\right) b'_x + f'_b \cdot b''_{xx}$$

Simple! Right now your face probably looks a lot like this:

😱😱😱

What the hell happened to f'_a and f'_b? The first thing to realize is that f'_a is actually $f'_a\left(a\left(x,y\right), b\left(x,y\right)\right)$and f'_b is actually $f'_y\left(a\left(x,y\right), b\left(x,y\right)\right)$ and we can differentiate them further *just as if they were ordinary functions.* So let's slow down a little and get back *drawing*:



Notice that this diagram gives $f''_{aa} \cdot a'_x + f''_{ab} \cdot b'_x$, which is exactly what you can see in u''_{xx} above. $a'_x + f'_a \cdot a''_{xx}$ part is the result of applying $(f \cdot g)' = f' \cdot g + f \cdot g'$. Second half of u''_{xx} is derived in exactly the same fashion. Same for u''_{xy}, u''_{yx}, and u''_{yy}. **It takes some effort to understand the process, but once you draw a few diagrams, it becomes rather straightforward**. But back to u''_{xx}:

$$u''_{xx} = \left(f''_{aa} \cdot a'_x + f''_{ab} \cdot b'_x\right) a'_x + f'_a \cdot a''_{xx} + \left(f''_{ba} \cdot a'_x + f''_{bb} \cdot b'_x\right) b'_x + f'_b \cdot b''_{xx}$$

We could try to simplify this, but really it's simpler to just plug in the numbers. We don't know f so all f' and f'' just stay as they are. $a'_x = 1 + y$; $a''_{xx} = 0$; $b'_x = \frac{1}{y}$; $b''_{xx} = 0$. Then:

$$u''_{xx} = \left(f''_{aa} \cdot (1+y) + f''_{ab} \cdot \frac{1}{y}\right)(1+y) + f'_a \cdot 0 + \left(f''_{ba} \cdot (1+y) + f''_{bb} \cdot \frac{1}{y}\right)\frac{1}{y} + f'_b \cdot 0 =$$
$$= f''_{aa}\left(1+y\right)^2 + \left(f''_{ab} + f''_{ba}\right)\frac{1+y}{y} + f''_{bb}\frac{1}{y^2}$$

I very strongly encourage you to try to calculate at least u''_{xy} by yourself and compare the results:

$$u''_{xy} = \left(f''_{aa} \cdot a'_y + f''_{ab} \cdot b'_y\right) a'_x + f'_a \cdot a''_{xy} + \left(f''_{ba} \cdot a'_y + f''_{bb} \cdot b'_y\right) b'_x + f'_b \cdot b''_{xy} =$$
$$= f''_{aa} \cdot x\,(1+y) + f''_{ab} \frac{(1+y)\,x}{-y^2} + f'_a + f''_{ba}\frac{x}{y} + f''_{bb}\frac{x}{-y^3} + f'_b \frac{1}{-y^2}$$

Correct solution to the example in the previous section. To recap: $f(x,y) = u^2(x,y) + v^3(x,y)$, $u(1,1) = 3$, $v(1,1) = -2$, $\nabla u(1,1) = (1,4)$, $\nabla v(1,1) = (-1,1)$. It's not a coincidence that this example is given in the chain rule section:

$$\nabla f = \nabla\left(u^2\right) + \nabla\left(v^3\right) = 2u \cdot \nabla u + 3v^2 \cdot \nabla v$$

5.9 Second-order approximation

Young's Theorem. If the function is $\in C^2$ (twice continuously differentiable), then $f''_{xy} = f''_{yx}$[4].

Protip: Although you could always just rewrite f''_{yx} as f''_{xy}, it's a good idea to calculate them both independently to confirm that they are the same and that you did everything correctly.

If a function of two variables $f(x,y)$ is twice continuously differentiable ($f \in C^2$), its second-order total differential is:

$$d^2f = d(df) = d(f'_x dx + f'_y y dy) = d(f'_x)dx + d(f'_y)dy = (f''_{xx}dx + f''_{xy}dy)dx + (f''_{yx}dx + f''_{yy}dy)dy =$$

$$= f''_{xx}dx^2 + 2f''_{xy}dxdy + f''_{yy}dy^2$$

Subsequently, the second-order approximation of a function of two variables $f(x,y)$ is its Taylor polynomial up to second-order derivative:

$$f(x,y) \approx f(x_0,y_0) + f'_x(x_0,y_0)(x-x_0) + f'_y(x_0,y_0)(y-y_0) +$$
$$+ \frac{1}{2}(f''_{xx}(x-x_0)^2 + 2f''_{xy}(x-x_0)(y-y_0) + f''_{yy}(y-y_0)^2)$$

Example. Use second-order approximation to approximate the function $f(x,y) = x^3y^5 + x^2 - y^3 + xy$ at a point $(1,1)$

Solution.

$$df = (3x^2y^5 + 2x + y)dx + (5y^4x^3 - 3y^2 + x)dy$$

$$d^2f = (6xy^5 + 2)dx^2 + 2(15x^2y^4 + 1)dxdy + (20y^3x^3 - 6y)dy^2$$

$$f(x,y) \approx f(1,1) + f'_x(1,1)(x-1) + f'_y(1,1,)(y-1) +$$
$$+ \frac{1}{2}(f''_{xx}(x-1)^2 + 2f''_{xy}(x-1)(y-1) + f''_{yy}(y-1)^2) =$$
$$= 2 + 6(x-1) + 3(y-1) + \frac{1}{2}(8(x-1)^2 + 2 \cdot 16(x-1)(y-1) + 14(y-1)^2)$$

Note that since we only use first two differentials, the approximation is only accurate around the point $(1,1)$.

[4]You could try to picture a surface in xyz-space in your head, then imagine how we first take f'_x and then f''_{xy} or f'_y and then f''_{yx} on it, and, with a considerable effort, might see, why this theorem true. There's no short and clear explanation.

6 Implicit functions

Suppose we want to find the derivative $\frac{dy}{dx}$ of an *implicit* function $xy = 1$. Well, simple enough, just write it *explicitly* as $y = \frac{1}{x}$, and differentiate it:

$$\frac{dy}{dx} = y' = \left(\frac{1}{x}\right)' = -\frac{1}{x^2}$$

But now suppose the function is $x + \sin y + xy = 0$. Whatever we try to do, there's no way to place all ys on the one side and all xs on the other side. We are forced to differentiate the implicit function. The way we do it is using *Implicit Function Theorem.*

6.1 Implicit Function Theorem 1

Let's continue with $x + \sin y + xy = 0$. The important thing to realize here is that, although we can't disentangle x from y, the function itself still exists, and there's nothing fundamentally different between an *implicit* function $F(x, y) = 0$ and an *explicit* function $y = f(x)$. One of the implications is that we can still view y as a function of x:

$$F(x, y) = x + \sin y + xy = 0 \Rightarrow F(x, y(x)) = x + \sin y(x) + xy(x) = 0$$

This also means it's possible to find the derivative of the implicit function $\frac{dy}{dx}$ at a point, just as if it were explicit. Since $F(x, y(x)) = 0$, $F'(x, y(x)) = 0$.

Now, remembering the chain rule, we differentiate with respect to x:

$$\frac{\partial F}{\partial x} + \frac{\partial F}{dy} \cdot \frac{dy}{dx} = 0$$

then

$$\frac{\partial F}{dy} \cdot \frac{dy}{dx} = -\frac{\partial F}{\partial x}$$

and finally

$$\frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}$$

which is usually written as

$$y' = -\frac{F'_x}{F'_y}$$

The result of this derivation should have been familiar to you from the previous year Calculus as an *Implicit Function Theorem* (IFT). Since in this course we'll study more than one IFT, we are going to call it IFT1.

IFT1. If we have equation $F(x, y) = 0$ and such point (x_0, y_0) that:

1. point (x_0, y_0) satisfies equation $F(x_0, y_0) = 0$
2. the function F is continuosly differentiable[5] ($F \in C^1$)
3. $F'_y(x_0, y_0) \neq 0$

[5] Actually, we only need it to be differentiable around the point, but you don't need to think about it.

Then explicit function $y = f(x)$ is defined *near the point* (x_0, y_0) and its derivative $y'$is equal to

$$y' = -\frac{F'_x}{F'_y}$$

Condition 1 is needed because we need to make sure the point we're trying to find the derivative at actually belongs to the graph of the function.

Condition 2 is needed because, well, unless the function is differentiable at a point, we can't really take its derivative (**Exam tip**: usually implicit functions given are polynomials, which are always continuously differentiable; you can simply state this fact to show that the condition holds).

Condition 3 is needed since we divide by F'_y when calculating the derivative, and the function would not be defined if F'_y was 0 (like with $x^2 + y^2 = 1$ at $(1, 0)$ and $(-1, 0)$, as $F'_y = 2y = 0$ at these points).

Example. (taken from 25.03.2015 mock)

Consider the equation $y^3 + xy + 3x^2 + 2x^3 = 7$.

(a) Does this equation define the implicit function $y(x)$ at a point $(x = 1,\ y = 1)$?

(b) If the function $y(x)$ is defined, find its second-order Taylor expansion.

Solution.

(a) Let's check the three conditions:

1. $y^3 + xy + 3x^2 + 2x^3$ at $(1, 1)$ is $1 + 1 + 3 + 2 = 7$ – correct.
2. Polynomial, thus C^1.
3. $F'_y = 3y^2 + x = 3 + 1 = 4 \neq 0$ – satisfied.

Thus, we can conclude that this equation does indeed define the implicit function $y(x)$ at a point $(x = 1,\ y = 1)$.

(b) Second-order Taylor expansion of any function is given by:

$$y(x_0) + y'(x_0)(x - x_0) + \frac{y''(x_0)(x - x_0)^2}{2}$$

Check Taylor Series Introduction if you forgot this formula. And for our case it would look the following way:

$$y(1) + y'(1)(x - 1) + \frac{y''(1)(x - 1)^2}{2}$$

Then we can find $y'(1)$ by using the formula $y' = -\frac{F'_x}{F'_y}$:

$$y' = -\frac{y + 6x + 6x^2}{3y^2 + x} == \frac{1 + 6 + 6}{3 + 1} = -\frac{13}{4}$$

Now, remember that y is a function of x: $y(x)$, so both $F'_x = y + 6x + 6x^2$ and $F'_y = 3y^2 + x$ are functions of x, not of y, and when we write y we actually mean $y(x)$, so differentiate accordingly:

$$y'' = \left(-\frac{F'_x}{F'_y}\right)' = -\frac{F''_x F'_y - F'_x F''_y}{\left(F'_y\right)^2} = -\frac{(y' + 6 + 12x)\left(3y^2 + x\right) - (6y \cdot y' + 1)\left(y + 6x + 6x^2\right)}{\left(3y^2 + x\right)^2}$$

$$y'' = \left(-\frac{F'_x}{F'_y}\right)' = -\frac{F''_x F'_y - F'_x F''_y}{\left(F'_y\right)^2} =$$

$$= -\frac{(y' + 6 + 12x)\left(3y^2 + x\right) - (6y \cdot y' + 1)\left(y + 6x + 6x^2\right)}{(3y^2 + x)^2} =$$

$$= -\frac{\left(\frac{-13}{4} + 6 + 12\right)(3 + 1) - \left(6 \cdot \frac{-13}{4}\right)(1 + 6 + 6)}{(3 + 1)^2} = \frac{4115}{16}$$

Finally, the answer is

$$y \approx 1 - \frac{13}{4}(x - 1) + \frac{\frac{4115}{16}(x - 1)^2}{2}$$

6.2 Implicit Function Theorem 2

A slight generalization of IFT1 is the case when we have one dependent variable y and several independent variables $x_1, \dots, x_n$, so the equation becomes

$$F(x_1, \dots, x_n, y) = F(x_1, \dots, x_n, y(x_1, \dots, x_n)) = 0$$

Fortunately, we're actually only interested in the partial derivative $\frac{dy}{dx_i}$ of this function, which means that all the derivatives not involving x_i are 0 (as $c' = 0$). So our new expression is

$$F'_{x_i} + F'_y \cdot y'_{x_i} = 0$$

IFT2. If we have equation $F(x_1, \dots, x_n, y) = 0$ and such point $\left(x_0^1, \dots, x_{0,}^n y_0\right)$ that:

1. point $\left(x_0^1, \dots, x_{0,}^n y_0\right)$ satisfies equation $F(x_1^0, \dots, x_n^0, y) = 0$
2. the function is continuosly differentiable ($F \in C^1$)
3. $F'_y\left(x_0^1, \dots, x_{0,}^n y_0\right) \neq 0$

Then explicit function $y = f(x_1, \dots, x_n)$ is defined *near the point* $\left(x_0^1, \dots, x_{0,}^n y_0\right)$ and its partial derivatives $y'_{x_i}$are equal to

$$y'_{x_i} = -\frac{F'_{x_i}}{F'_y}, \text{ for any } i = 1, \dots, n$$

6.3 Implicit Function Theorem 3

The final generalization happens when there are n independent variables and m simultaneous equations. We will actually only work with the case of one independent variable and two functions, as, going beyond, everything gets too complicated. In equations below x is an independent variable, while $y(x)$ and $z(x)$ are dependent, i.e. they're functions of x:

$$\begin{cases} F(x, y, z) = 0 \\ G(x, y, z) = 0 \end{cases}$$

Differentiating each function with respect to x by using chain rule (check IFT1 if you forgot) to each function:

$$\begin{cases} F'_x + F'_y \cdot y'_x + F'_z \cdot z'_x = 0 \\ G'_x + G'_y \cdot y'_x + G'_z \cdot z'_x = 0 \end{cases}$$

Alternatively:

$$\begin{cases} F'_y \cdot y'_x + F'_z \cdot z'_x = -F'_x \\ G'_y \cdot y'_x + G'_z \cdot z'_x = -G'_x \end{cases}$$

Thus, we have a system of two equation with two unknowns: y'_x and z'_x, which we can solve by Cramer's rule.

IFT 3. If we have equations $\begin{cases} F(x,y,z) = 0 \\ G(x,y,z) = 0 \end{cases}$ and such point (x_0, y_0, z_0) that:

1. point (x_0, y_0, z_0) satisfies equations $\begin{cases} F(x,y,z) = 0 \\ G(x,y,z) = 0 \end{cases}$
2. the functions are continuosly differentiable ($F,\ G \in C^1$)
3. *Jacobian* (matrix of partial derivatives) given by

$$J = \begin{pmatrix} F'_y & F'_z \\ G'_y & G'_z \end{pmatrix}$$

4. is such that$|J| \neq 0$ at a point (x_0, y_0, z_0)

then, by application of *Cramer's rule*[6] the derivatives we're interested in are given by

$$y'_x = -\frac{\begin{vmatrix} F'_x & F'_z \\ G'_x & G'_z \end{vmatrix}}{|J|} = -\frac{\begin{vmatrix} F'_x & F'_z \\ G'_x & G'_z \end{vmatrix}}{\begin{vmatrix} F'_y & F'_z \\ G'_y & G'_z \end{vmatrix}}$$

$$z'_x = -\frac{\begin{vmatrix} F'_y & F'_x \\ G'_y & G'_x \end{vmatrix}}{|J|} = -\frac{\begin{vmatrix} F'_y & F'_x \\ G'_y & G'_x \end{vmatrix}}{\begin{vmatrix} F'_y & F'_z \\ G'_y & G'_z \end{vmatrix}}$$

To remind you, determinant of a 2×2 matix

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Exam tip: You will *absolutely certainly* be asked to employ IFT1, or IFT2, or IFT3, or any combination of these on the exam, so even if the explanations of these are unclear, just memorize the results of each: y' for IFT1; y'_{x_i} for IFT2; and y'_x, z'_x for IFT3; and make sure you can plug in the right numbers in formulas when asked.

[6] You could just memorize the formulas below, but Wikipedia actually has a wonderful (still rather difficult to understand, though) geometric explanation of this formula. Do check it out, if you're interested: https://en.wikipedia.org/wiki/Cramer%27s_rule#Geometric_interpretation

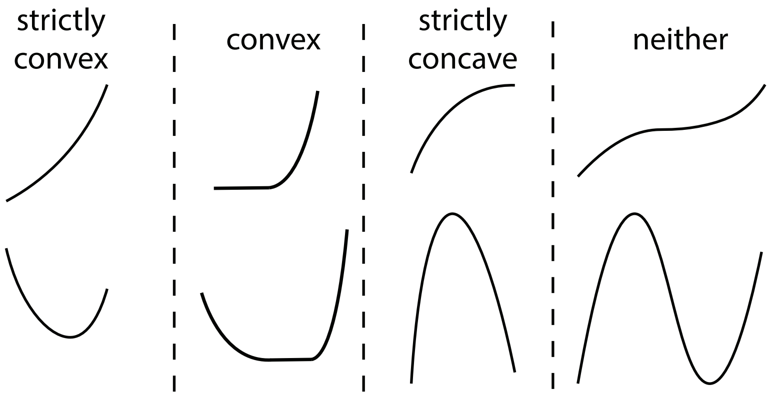
7 Convexity and Concavity. Convex Sets

Remark. In contrast to the course, the topics "Convexity and Concavity" and "Unconstrained Optimization" are presented in a different order here, because it feels more natural to me this way.

First derivative shows the slope of the function: $f'(x) > 0 \Rightarrow$ slope positive $\Rightarrow$ function increases; $f'(x) < 0 \Rightarrow$ slope negative $\Rightarrow$ function decreases. Recall an example from Taylor Series Introduction chapter. Slope of a function is analogous to its speed: speed is positive $\Rightarrow$ function increases; speed is negative $\Rightarrow$ function decreases.

Second derivative then is the "acceleration" of a function: function is speeding up $\Rightarrow f'' > 0$; function is slowing down $\Rightarrow f'' < 0$.

We call functions that are speeding up *convex* and functions that are slowing down *concave*[7].



Protip: an easy way to remember which one is convex and which is concave is to note that $y = -x^2$ looks a lot like a cave. Coincidentally, it is also con*cave*.

Okay, this was the basic intuition, but it is waaaaay too imprecise, even for me. Actually, if the function is always speeding up, i.e. $f'' > 0$, then it's called *strictly convex*. Simply *convex* means that it does not slow down, i.e. $f'' \geq 0$. Same for concave. So lines like $y = 2x$ are both convex and concave.

Furthermore, you could say that concave function like $y = -x^2$ is first slowing down and then speeding up, pointing to the absolute value of its first derivative. Well, technically, by "speeding up" I mean "speeding up upwards or slowing down downwards". Same for "slowing down".

The technical formulation for convex function is the following:

Definition. A function is *convex* on (a, b) if the inequality

$$f(\alpha x + (1 - \alpha) y) \leq \alpha f(x) + (1 - \alpha) f(y)$$

is satisfied for any two points x and y from (a, b) and any α in $[0, 1]$.

Protip: Although you are rarely asked for this definition, sometimes, remembering it and understanding its geometrical meaning (it is explained wonderfully in Jeffrey on page 49) is extremely helful in the exams (see Example at the end of this chapter).

7.0.1 Convex sets

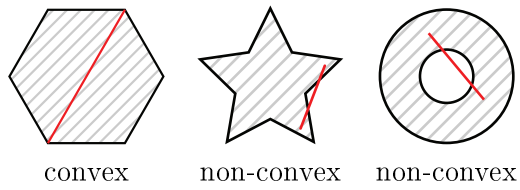
It's rather obvious that $f(x) = x^2$ is convex. However, what if we define the domain (all inputs) to be $(1, 4) \cup (8, 14)$, rather than $(-\infty, +\infty)$. Is $f(x)$ still convex on its domain?

The first thing to notice here is that the definition above only describes the sitiation of (a, b) – a single interval, while here we have two intervals. But let's ignore this for a moment and proceed anyway. Then, by taking two points in the domain, say $x = 2$ and $y = 10$, and taking their middle i.e. $\alpha = 0.5$, we get $f(0.5 \cdot 2 + 0.5 \cdot 10) = f(6)$, which is not defined!

What we found is that the initial question does not make any sense – the function can't be either convex or concave on a set like this. In $\mathbb{R}^1$ the set (domain) needs to be "connected". In $\mathbb{R}^n$ the

[7] Sometimes *convex* is called *concave up* and *concave* is called *concave down*.

situation is more complicated: here, the set (domain) needs to be *convex*, i.e. all of its points have to be connected by a straight line segment, for us to be able to determine convexity of a function. Some examples:



Definition. A set is called *convex* if given any two points a, b in that set, the straight line segment ab joining them lies entirely within that set.

Formally, a set V is called *convex* if

$$\forall a, b \in V \text{ point } \alpha a + (1-\alpha)b \in V,\ 0 \leq \alpha \leq 1$$

Notice, that for a concave function, e.g. $y = ln(x)$, the area below it – caled subgraph – looks like a convex set; and for a convex function, e.g. $y = x^2$ the area above it – called epigraph – looks like a convex set. Thus, a theorem:

Theorem. If f is concave, then its subgraph is a convex set. If f is convex, then its epigraph is a convex set.

Example 1. Determine whether the following set is convex: $\{(x, y) \mid y = x^2\}$

Answer: If you skipped the relevant seminar, you've probably thought "of course it is, since $y = x^2$ is convex!". But if you reread the problem, it does not actually say anything about the epigraph of $y = x^2$. The points in this example all lie on the parabola itself. Since when we connect any two of these points, we get off the parabola, the set in question is not convex.

This was an intuitive explanation, but to prove it formally we'll need to make use of the definition of a convex set written just above. Let $a = (-2, f(-2)) = (-2, 4)$, $y = (1, f(1)) = (1, 1)$. We can pick any $\alpha \in (0, 1)$ but let's take $\alpha = 0.6$ here, as an example. Then $\alpha a + (1-\alpha)b = 0.6(-2, 4) + 0.4(1, 1) = (-0.8, 2.8)$. Since $f(-0.8) = 0.64 \neq 2.8$ this point does not lie in the set. Thus we get a contradiction with the definition and a proof that the set is not convex.

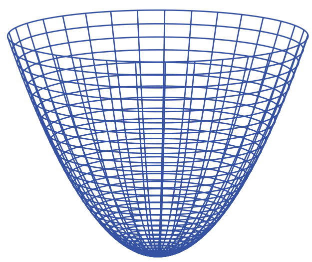
Example 2. Determine whether the following set is convex: $\{(x, y) \mid y \geq x^2\}$

Answer: This set is convex, since it describes the area above the parabola $y = x^2$ and the theorem is applicable.

7.1 $\mathbb{R}^2$

But what do we do with a function of two variables? Rather than simply checking f'', we now have four partial derivatives: $f''_{xx}, f''_{xy}, f''_{yx} f''_{yy}$. Let's start with a simple example:

$$f(x, y) = x^2 + y^2$$



It's visually obvious that this function is convex, so let's see what happens to second-order partial derivatives in this case:

$$\begin{matrix} f'_x = 2x \\ f'_y = 2y \end{matrix} \Rightarrow \begin{matrix} f''_{xx} = 2 \\ f''_{xy} = 0 \\ f''_{yx} = 0 \\ f''_{yy} = 2 \end{matrix}$$

Note that cross derivatives (f_{xy} and f_{yx}) are 0 and we can forget about them for now. Seeing that $f''_{xx} > 0$ **at the entire domain of the function**, we may say that f is **always** speeding up along the x-axis; And since the same could be said about y-axis, we may conclude that the function is convex as a whole.

Usually we arrange partial derivatives in the form of the *Hessian matrix*:

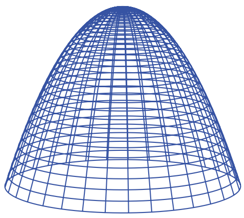
$$H = \begin{pmatrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Definition. *Hessian* matrix is given by

$$H = \begin{pmatrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{pmatrix}$$

Switching signs of the function we get

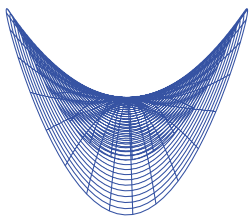
$$f(x, y) = -x^2 - y^2$$



$$H = \begin{pmatrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

Which is obviously concave. Finally, for function

$$f(x, y) = x^2 - y^2$$



$$H = \begin{pmatrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

As $f''_{xx} > 0$, the function is speeding up along its x-axis; $f''_{yy} < 0$, so function is slowing along the y-axis, which means that it's neither concave nor convex.

Using these three functions for intuition, we can proceed to a more formal treatment. If we define $H_1 = |f''_{xx}|$ and $H_2 = |H| = \begin{vmatrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{vmatrix}$, we can create the following table:

$f(x)$	H	H_1	H_2	convexity/concavity	definiteness
x^2+y^2	$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$	>0	>0	strictly convex	positive definite
$-x^2-y^2$	$\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$	<0	>0	strictly concave	negative definite
x^2-y^2	$\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$	something else		neither	neither

"Positive definite" and "negative definite" is what matrices, which satisfy the given conditions are called. You should remember them because sometimes these terms are used in the exams.

H_1 and H_2 in the table above are called *leading principal minors* of a matrix. Formally:

Definition. Let A be an $n \times n$ matrix. The kth order *leading principal minor* of A is the determinant of a matrix obtained by deleting the last $n-k$ rows and columns of A.

So 1st order leading principal minor of A : H_1, is obtained by deleting all but the first row and column. H_2 is obtained by deleting all but first two rows and columns. And so on. Consequently,$|H_n|$ is the determinant of a $n \times n$ matrix.

A general rule for finding whether a function f is strictly convex or strictly concave is:

1. f is strictly convex if and only if all its leading principal minors are strictly positive (>0).
2. f is strictly concave if and only if all its leading principal minors alternate signs as follows:

$$H_1 < 0,\ H_2 > 0,\ H_3 < 0,\ \text{and so on}$$

But what if the general pattern above holds, but some leading principal minor H_m is 0? This is where intuition about speeding up and slowing down along axes ends, and where we'll need to do a *lot* more calculations. In this case we'll unfortunately need to check all *principal minors* to determine whether the function is convex or concave:

Definition. Let A by an $n \times n$ matrix. A *principal minor* of A is the determinant of a matrix obtained by deleting $n-k$ rows of A, and the same $n-k$ columns of A.

So, for a 2×2 matrix there are two 1st order principal minors: $D_{11} = |f''_{xx}|$ and $D_{12} = \left|f''_{yy}\right|$, and one 2nd order principal minor: $D_2 = \begin{vmatrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{vmatrix}$.

For a 3×3 matrix there are three 1st order principal minors: $D_{11} = |f''_{xx}|$ (remove 2nd and 3rd rows and columns), $D_{12} = \left|f''_{yy}\right|$, (remove 1st and 3rd rows and columns) and $D_{13} = |f''_{zz}|$ (remove 2nd and 3rd rows and columns); three 2nd order principal minors: $D_{21} = \begin{vmatrix} f''_{xx} & f''_{xy} \\ f''_{yx} & f''_{yy} \end{vmatrix}$ (remove 3rd row and column), $D_{22} = \begin{vmatrix} f''_{xx} & f''_{xz} \\ f''_{zx} & f''_{zz} \end{vmatrix}$ (remove 2nd row and column), and $D_{23} = \begin{vmatrix} f''_{yy} & f''_{yz} \\ f''_{zy} & f''_{zz} \end{vmatrix}$ (remove 1st row and column); and one 3rd order principal minor: $D_3 = \begin{vmatrix} f''_{xx} & f''_{xy} & f''_{xz} \\ f''_{yx} & f''_{yy} & f''_{yz} \\ f''_{zx} & f''_{zy} & f''_{zz} \end{vmatrix}$.

A general rule for finding whether a function f is convex or concave is:

1. f is convex if and only if all its principal minors are non-negative (≥ 0).
2. f is concave if and only if all its principal minors alternate signs as follows:

$$D_1 \leq 0,\ D_2 \geq 0,\ D_3 \leq 0,\ \text{and so on}$$

The table for convexity/concavity for functions of two variables is:

D_{1m}	D_{2m}	convexity/concavity	definiteness
≥ 0	≥ 0	convex	positive semidefinite
≤ 0	≥ 0	concave	negative semidefinite
something else		neither	neither

Example. If you understand the formal definition of concavity/convexity, you might find yourself quite happy upon seeing a problem like this on the exam (this one was taken from 25.03.2015 mock):

Let $f(x)$ be a concave function defined on $[0; \infty)$ and $f(0) = 0$. Is it true that for $k \geq 1$ the following inequality holds: $kf(x) \geq f(kx)$?

I strongly suggest you try to solve this problem yourself before reading the solution.

Solution. Recalling that concave means that

$$f(\alpha x_1 + (1-\alpha)x_2) \geq \alpha f(x_1) + (1-\alpha)f(x_2)$$

We need to figure out a way to turn this into $kf(x) \geq f(kx)$. The first thing to notice is that this looks a lot like the definition, except for this pesky $1-\alpha$ term. Recalling that $f(0) = 0$ and setting $x_2 = 0$ (setting $x_2 < x_1$ is counterintuitive, but the definition doesn't actually say that x_2 most be greater than x_1), we get the following:

$$f(\alpha x_1) \geq \alpha f(x_1)$$

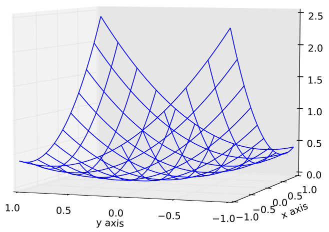
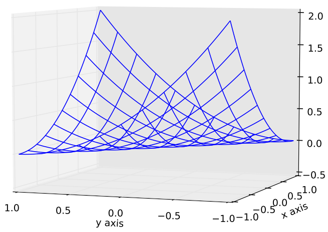
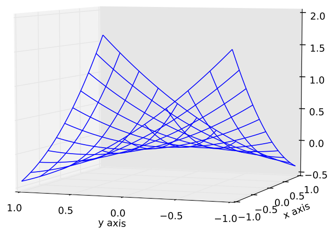
But in this formulation $\alpha(f(x_1)$ is to the right of $\geq$, while in the problem formulation it's to the left of $\geq$. Then we may notice that taking $\alpha = \frac{1}{k}$ solves this problem:

$$kf(\frac{x_1}{k}) \geq f(x_1)$$

Now it's pretty obvious that to get from this to $kf(x) \geq f(kx)$ we just need to take $x_1 = kx$.

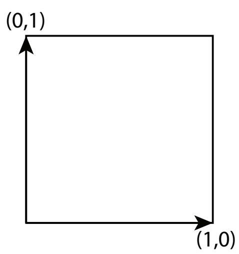
7.2 Appendix (don't read this unless you want to mess with your brain)

You can think of every leading/principal leading minor as of cross-section of a function:

$0.7x^2 + xy + 0.7y^2$	$0.5x^2 + xy + 0.5y$	$0.3x^2 + xy + 0.3y$
$\begin{pmatrix} 1.4 & 1 \\ 1 & 1.4 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.6 & 1 \\ 1 & 0.6 \end{pmatrix}$
$H_1 > 0,\ H_2 > 0$	$H_1 > 0,\ H_2 = 0$	$H_1 > 0,\ H_2 < 0$
strictly convex	convex	neither
		

7.3 What Does Determinant Have To Do With Anything? (don't read this even more)

Suppose we have a 1 by 1 square, which we can get from vectors $(1,0)$ and $(0,1)$:

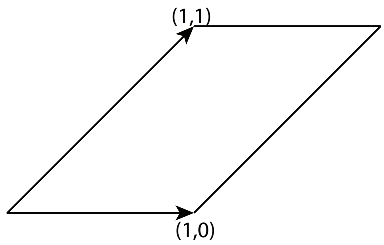


This square can be written in a matrix form as

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

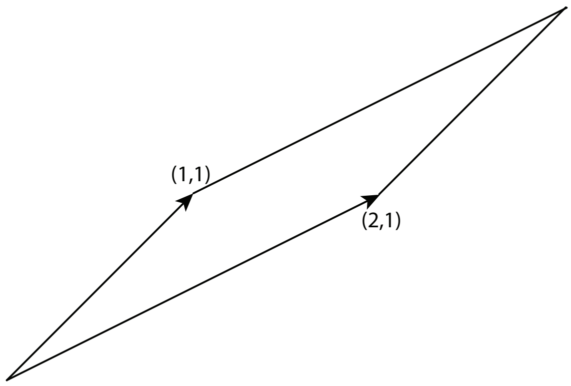
where the first row denotes vector $(1, 0)$ and second row denotes vector $(0, 1)$. Area of the square is 1 and determinant of this matrix is 1. Now let's add the first row of the matrix to the second row and get the following rhombus:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$



Area stayed the same and determinant stayed the same. Now let's add the second row to the first row:

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$



And again, both area and determinant stayed the same. This should give you an intuitive understanding why determinant gives area of the figure[8] and why determinant is 0, when vectors are dependent. What else does this argument show is that **however twisted the initial figure is, we can always reduce it to the "fundamental" form of the diagonal matrix**[9] (all figure's angles are 90 degree). The determinant will stay the same. "Fundamental" does not mean unique. This has something to do with Hessian but I'm not sure what exactly. Sorry.

[8] Also, multiplication of a row multiplies area and determinant by a constant.

[9] Numbers on the diagonal are eigenvalues of the matrix.

8 Unconstrained Optimization

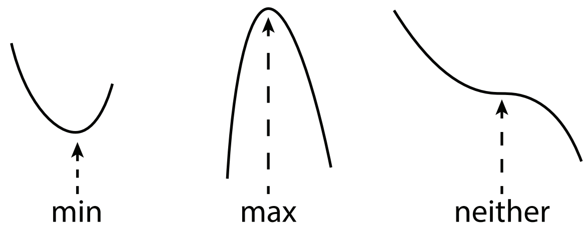
8.1 Local Optima

Suppose we have a function $y = f(x)$ and we want to find its minimum and maximum. How do we do it? Start looking for *stationary points*, i.e. points where the function is flat i.e. $y' = 0$.

Definition. Point is called *stationary* (or *critical*), if all partial derivatives of the function are equal to 0 at this point, i.e. $\nabla f = 0$.

This was the *first-order condition* (since it's based on the first derivative) for the min or max, also called *necessary condition*. Usually we just say "*FOC*", though.

However, finding such a point is not sufficient, since we don't know if it is a minimum, maximum, or neither of these.



To ascertain which point it is, we need to find the second derivative of a function. If the second derivative is positive, then the first derivative (slope) is increasing, the function is speeding up, and, looking at the picture above, we have the case of min. If the second derivative is negative, then the slope is decreasing, the function is slowing down, and we have the case of max. If the second derivative is zero, then this is an inflection point.

$$\begin{aligned} y' = 0, \quad y'' > 0 \Rightarrow & \quad \text{speeding up} \Rightarrow \text{min} \\ y' = 0, \quad y'' < 0 \Rightarrow & \quad \text{slowing down} \Rightarrow \text{max} \end{aligned}$$

This was the *second-order condition* (since it's based on the second derivative) for the min or max, also called *sufficient condition*. Usually we just say "*SOC*", though.

All of this sounds suspiciously similar to our discussion of convexity and concavity in the previous chapter. And it is in fact the same discussion. Finding that the function attains a local minimum at a point is exactly the same as finding that a function is convex **in this point's vicinity** (look at the picture above if it's not clear why!). Finding that the function attains a local maximum at a point is exactly the same as finding that a function is concave **in this point's vicinity**. If this is not clear, try to imagine a differentiable function that would be convex around a point, where it attains maximum.

Therefore, the rule for convexity becomes the rule for local min and the rule for concavity becomes the rule for local max, with the difference that all **Hessian matrices are calculated at stationary points**:

H_1	H_2	H_3	definiteness	min/max
> 0	> 0	> 0	positive definite	local min
≥ 0	≥ 0	≥ 0	positive semidefinite	inconclusive
< 0	> 0	< 0	negative definite	local max
≤ 0	≥ 0	≤ 0	negative semidefinite	inconclusive
something else			indefinite	saddle point

Exam tip: If you get an inconclusive result, generally you don't need to look any further and can just write "inconclusive" in answer.

You can check the pictures in section 7.2 on page 38 for geometric intuition regarding these rules. The middle picture there: $0.5x^2 + xy + 0.5y$, sheds some light on the inconclusive case.

The reason we're only talking about local optima right now is that we're calculating Hessian matrices at specific points and therefore cannot know what happens with the function on its whole domain.

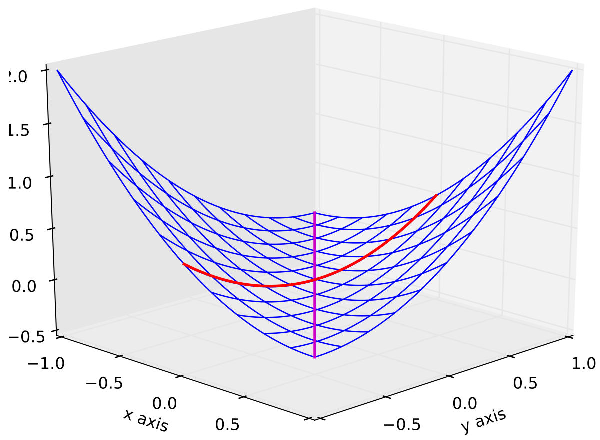
Protip: Recall that the Young's Theorem says that if the function is $\in C^2$ (twice continuously differentiable), then $f''_{xy} = f''_{yx}$. Since functions given for such exercices almost always satisfy this theorem, Hessian matrices are almost always symmetric.

8.2 Global Optima

If the function is either concave or convex, then its only critical point is its global maximum or minimum, respectively. In all other cases, finding global minima and maxima of a function is much less straightforward, as there's no universal rule for this problem. There are two general ways to proceed further:

1. Try to prove that there's no global min/max
2. Prove that a local extremum is also a global one.

Let's start with trying to prove there's no global min/max. The usual way to do this would be to show that the function goes to infinity in some direction. For example, let $f(x, y) = 0.5x^2 + xy + 0.5y^2$:



So, suppose we want to prove that it has no global maximum. Let's try the direction $x = y$, so $f(x, y) = 0.5y^2 + y^2 + 0.5y^2 = 2y^2$. Now check that $\lim\limits_{y\to+\infty} 2y^2 = +\infty$, so indeed this function has no global maximum. In fact – picture makes it clear – for this function, we could check literally any direction other than $x = -y$ (purple line) and the result would stay the same. For example, let $x = 0$ (red line), then $f(x, y) = 0.5y^2$ and $\lim\limits_{y\to-\infty} 0.5y^2 = +\infty$; or let $y = x^2$! Then $f(x, y) = 0.5x^2 + x^3 + 0.5x^4$ and $\lim\limits_{x\to+\infty} 0.5x^2 + x^3 + 0.5x^4 = +\infty$;

Proving that a local extremum is also the global one is harder. There's basically three options:

1. The function is convex or kinda convex and everything works out
2. Transformation to polar coordinates works and everything's easy as well (just find lim of a function as $r \to \infty$ to prove that the local limit is a global one)
3. The two above don't work and you're fuc..uh... you have to come up with something on the spot.

9 Constrained Optimization

Note: You can skip this explanation of the lagrangian if you wish and move right to the method itself.

Definition of the economic good is that it's something that is both scarce and desirable. It's pretty obvious that the lack of constraints when optimizing, does not go hand-in-hand with the scarcity condition. Almost all optimization problems encountered in real life do have some constraints placed on them.

Here, I'll use a utility maximization problem faced by a consumer as an example. The constraint is their income $I = 5$. Available goods are x and y. Utility function $U(x, y) = x \cdot y$. For simplicity, we'll assume price of both goods to be equal to 1, so the generic income constraint $P_x \cdot x + P_y \cdot y = I$ transforms into $x + y = 5$ Formally, the task is:

$$\begin{cases} U(x,y) = x \cdot y \to \underset{x,y}{max} \\ x + y = 5 \end{cases}$$

There are several ways to solve this problem.

The most obvious one is to express $y = 5 - x$ and substitute this into the original equation, making the problem

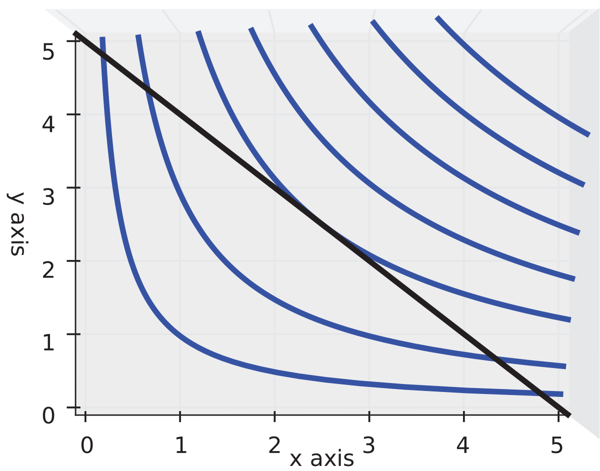
$$x \cdot (5 - x) = 5x - x^2 \to \underset{x}{max}$$

Then we set derivative to 0

$$5 - 2x = 0$$
$$x = 2.5$$

We know that $5x - x^2$ is a parabola with branches downwards, which means that $x = 2.5$ is its maximum. Pretty simple.

In microeconomics we would probably use graphs to solve it. So let's draw the indifference curves (which are called level curves in our course; check section 4.1 on page 16) and income constraint:



The solution $x = 2.5$ is immediately obvious. The important thing to notice here is that the income constaint is tangent to the optimal level-curve. As we can see, any level-curve that crosses the income constraint, but is not tangent to it, is not optimal.

Now let's change the income constraint $x+y = 5$ to the general form of $g(x, y) = x+y$. Again, from the picture above, we can see that the level-curves of f and g are tangent. Recall that the gradient's third property says that it is orthogonal to the level-curve. Since f and g share the level-curve at a point, their gradients are orthogonal to the same line, and they must be codirected.

The fact that ∇f and ∇g are codirected means that they are coefficients of each other and we can get one from the other by multiplying it by some number. Usually λ is used here:

$$\nabla f = \lambda \nabla g \iff \begin{cases} f'_x = \lambda g'_x \\ f'_y = \lambda g'_y \end{cases}$$

By including the original income constraint $g(x, y) = c$, we get three equations with three unknowns and the problem becomes:

$$\begin{cases} f'_x = \lambda g'_x \\ f'_y = \lambda g'_y \\ g(x, y) = c \end{cases}$$

Solving this system of equations will get us first-order conditions. This approach might seem somewhat unwieldy, especially when we can just subsitute $y = 5 - x$, but when constraints become more complex, e.g. $y^2 + x^2 = 1$, it's the most convenient way to solve an optimization problem.

The way to remember those three equations is to introduce the *Lagrangian function:*

$$L(x, y, \lambda) = f(x, y) + \lambda \left(c - g(x, y)\right)$$

By taking its partial derivatives L'_x, L'_y, and L'_λ and equating them to zero we get exactly the original system:

$$\begin{cases} L'_x = 0 \\ L'_y = 0 \\ L'_\lambda = 0 \end{cases} \Rightarrow \begin{cases} f'_x - \lambda g'_x = 0 \\ f'_y - \lambda g'_y = 0 \\ c - g(x, y) = 0 \end{cases}$$

Weierstrass Theorem. Function continuous on a compact (closed and bounded) set attains its minimum and maximum.

I suggest you check back on the discussion of the significance of this theorem in section 3.2 on page 12.

9.1 What the hell is NDCQ?

"Always check NDCQ" your seminar teacher tells you. What the hell does that even mean? Well its actual formulation is:

If $\exists\,(x, y) : \begin{cases} \nabla g\,(x, y) = (0, 0) \\ g\,(x, y) = c \end{cases}$, then remember $(x*, y*)$ as a candidate for extremum. If $\nexists$, then NDCQ holds.

Basically it says that when the gradient of the constraint is equal to 0, while satisfying the constraint, then the Lagrangian won't detect this point while looking for extremum (since we can't solve $\nabla f = \lambda \nabla g$). This means that we need to check this point separately later.

Example 1. Suppose the constraint $x^2 + 3y^2 = 4$. Then its gradient is $(2x, 6y)$. $2x = 0 \to x = 0$, $6y = 0 \to y = 0$. Since $0^2 + 3 \cdot 0^2 = 0 \neq 4$, NDCQ is satisfied.

Example 2. Suppose the constraint is $x^2 + 3y^2 = 0$. Then its gradient is $(2x, 6y)$. $2x = 0 \to x = 0$, $6y = 0 \to y = 0$. Since $0^2 + 3 \cdot 0^2 = 0$, NDCQ is violated. Then we'll need to calculate the value of the function at a point $(0, 0)$ later and check if it is an extremum.

9.2 Lagrange multiplier method

Steps (solution of an example is below):

1. **Check NDCQ (non-degenerate constraint qualification).**

If $\exists\,(x*, y*) : \begin{cases} \nabla g\,(x*, y*) = (0,0) \\ g\,(x,y) = c \end{cases}$, then remember $(x*, y*)$ as a candidate for extremum. If $\nexists$, then NDCQ holds.

2. **Introduce Lagrangian function.**

$L(x, y, \lambda) = f(x, y) + \lambda\,(c - g(x, y))$
Check FOC (necessary condition):

$$\begin{cases} L'_x = 0 \\ L'_y = 0 \\ L'_\lambda = 0 \end{cases} \Rightarrow \begin{cases} f'_x - \lambda g'_x = 0 \\ f'_y - \lambda g'_y = 0 \\ c - g(x, y) = 0 \end{cases}$$

Find critical points $(x*, y*, \lambda*)$

3. Check SOC (sufficient condition).

Bordered Hessian in our case is:

$$\bar{\mathrm{H}} = \begin{pmatrix} 0 & g'_x & g'_y \\ g'_x & L''_{xx} & L''_{xy} \\ g'_y & L''_{yx} & L''_{yy} \end{pmatrix}$$

Note that $L''_{xy} = L''_{yx}$, so you only need to calculate one of them. In our case, $n = 2$, $m = 1$, so if $\bar{\mathrm{H}} > 0$, then $(x*, y*, \lambda*)$ is maximum. If $\bar{\mathrm{H}} < 0$, then $(x*, y*, \lambda*)$ is minumum.

9.2.1 Bordered Hessian

Exam tip: If you can only memorize one thing from the entire course for the exam, memorize this!!
Let n be the number of variables and m be the number of constraints. The general rule for the Bordered Hessian when finding max is:

1. Calculate the determinant of the Hessian (recall that the Hessian is the last principal leading minor)
2. If its sign is $(-1)^n$, then start to calculate the determinants of the previous principal leading minors i.e. remove rightmost column and the bottom row one by one. **The signs must alternate**.
3. Calculate the determinants of the last $n - m$ leading principal minors or until the pattern breaks down (so you know this is not max).

The general rule for the Bordered Hessian when finding min is:

1. Calculate the determinant of the Hessian (recall that the Hessian is the last principal leading minor)
2. If its sign is $(-1)^m$, then start to calculate the determinants of the previous principal leading minors i.e. remove rightmost column and the bottom row one by one. **The signs must all equal to** $(-1)^m$.
3. Calculate the determinants of the last $n - m$ leading principal minors or until the pattern breaks down (so you know this is not min).

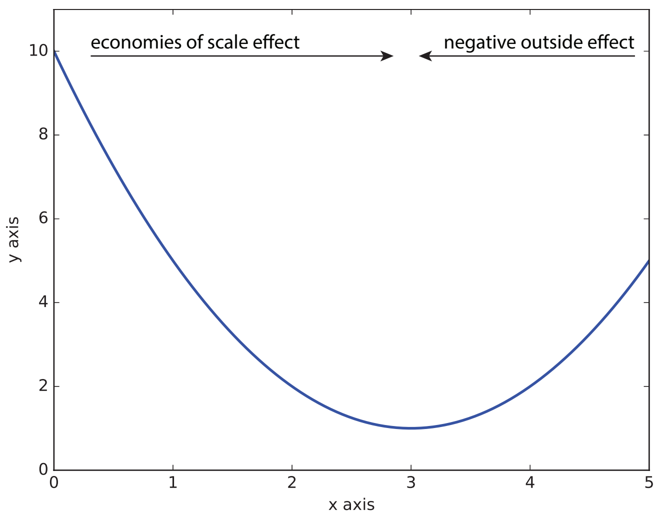
The rule when signs must alternate and when they stay the same is hopefully familiar to you from the discussion of unconstrained optimization. If it's not, you can use mnemonics to remember it (remember the *cave*?) The metaphor that came to my mind is that in order to stay at the top (max) you always need to fight different enemies (so need alternate strategies and stuff); and when you're just trying to hold on (min) you're digging into the trenches and just do one thing (signs stay the same). If this didn't help, try come up with your own mnemonic! Anyway, here's another one: Note that n is always bigger than m. So when we are finding max (big number) we care about $(-1)^n$ and when we are finding min (small number) we care about $(-1)^m$.

Examples. There are great examples and a deeper explanation of the Lagrange Multipliers method in the OptimizationHOWTO by A.Kalchenko. (if the link doesn't work, go to Mathematics for Economists page in icef-info and scroll to the bottom of the page). It should also help if you found my explanation of the Lagrangian and/or Bordered Hessian convoluted and unintelligible.

9.3 Envelope theorem (unconstrained)

Imagine yourself several of years from now: a successful ICEF graduate, you are in a very competitive business of growing marijuana. You learned well from the microeconomics courses that the only way to survive in competitive markets is to minimize *Average Total Cost*. Your ATC is affected by the economies of scale, which increase your production efficiency, and by the fact that if you, um, produce too much of the good, law enforcement agencies will spend much more resources trying to bust you, thus increasing your costs. This model suggests the following quadratic function, which you need to minimize:

$$ATC = y(x) = x^2 - 6x + 14 \to \min_x$$
$$y' = 2x - 6 = 0$$
$$x = 3, \ y = 5$$



So you find that the optimal production is 3 units of your top-notch product. However, the police has suddenly become much more active, which affects the coefficient a, changing your ATC function to

$$y(x) = x^2 - 4x + 14 \to \min_x$$
$$y' = 2x - 4 = 0$$
$$x = 2, \ y = 10$$

As expected, the optimal quantity has fallen from 3 to 2, while ATC has risen from 5 to 10. However, instead of calculating the optimal production every time the activity level of the police changes, we could solve the equation once for arbitragy a: $y(x, a) = x^2 - ax + 14$ and then just substitute the appropriate a into the solution to find the answer. To see how it works, let's do this procedure:

$$y(x,a) = x^2 - ax + 14 \to \min_x$$
$$y' = 2x - a = 0$$
$$x = \tfrac{a}{2},\ y = -\tfrac{a^2}{4} + 14$$

Substituting $a = 4$ we get $x = 2,\ y = 10$, exactly as before.

The final expression for ATC is $y(x,a) = -\frac{a^2}{4} + 14$. It tells us the optimal value of our function, depending on some parameter a, and it is called *the value function*, usually denoted $V(a)$. In our case $V(a) = -\frac{a^2}{4} + 14$.

Note that to find the effect of a marginal change in police activity on ATC we would need to take the derivative of $y(x,a) = x^2 - ax + 14$ by a:

$$y'_a(x,a) = -x$$

And since the optimal $x = \frac{a}{2}$, substituting it,

$$y'_a(x,a) = -\frac{a}{2}$$

On the other hand, we could find the effect of a marginal change in police activity on ATC by taking the derivative of $V(a) = -\frac{a^2}{4} + 14$ directly, as it shows ATC for all a:

$$V'(a) = -\frac{a}{2}$$

What we just saw is exactly the statemetent of the Envelope theorem. Mathematically, it is stated as

Theorem 1 (unconstrained optimization). Let $f(x,a) \in C^1$ and $f(x,a) \to \max_x = f(x^*(a),a) = V(a)$ i.e. we rename the result of the maximization as $V(a)$. Then $V'(a) = \frac{df(x^*(a),a)}{da} = \left.\frac{\partial f(x,a)}{\partial a}\right|_{x^*(a)}$

9.4 Envelope theorem (constrained)

Most often the Envelope theorem is used in the constrained case, e.g. with the Lagrangian. In this case the mathematical formulation gets very clunky but the result is that $L = f(x,y) + \lambda g(x,,y)$ becomes $V(a)$. This means that all you need to do is to take the Lagrangian derivative with respect to the parameter, at the optimal point you have found. The intuition behind this is that L kinda incorporates $f(x)$ and $g(x)$ together, which means that we can work (i.e. take derivative) with it directly.

Example. It is known that the point $(1,0)$ is the constrained local maximum of the function $f(x,y) = 5x - ky - 3x^2 + 2xy - 5y^2$ subject to $x + y = 1$.

(a) Find the value of k and the maximum value of the function f

(b) Using Envelope theorem find the new value of maximum if k will increase by 0.1

Solution.

(a) First, set up the Lagrangian

$$L = 5x - ky - 3x^2 + 2xy - 5y^2 + \lambda(1 - x - y)$$

Then solve it and find k

$$\begin{cases} L'_x = 5 - 6x + 2y - \lambda = 0 & (1) \\ L'_y = -k + 2x - 10y - \lambda = 0 & (2) \\ L'_\lambda = 1 - x - y = 0 \Rightarrow y = 1 - x & (3) \end{cases}$$
$$5 - 6 + 2 - 2 - \lambda = 0 \qquad (1)$$
$$\lambda = -1 \qquad (1)$$
$$-k + 2 + 1 = 0 \qquad (2)$$
$$k = 3 \qquad (2)$$

So $f(x, y) = 2$ at $x = 1,\ y = 0$ and $k = 3$
(b)

$$df = L'_k dk = -0.1y = 0$$