
Lagrange-Hamilton-series part 2 - Multi-dimensional extrema, partial derivatives and the gradient

An article posted on bathtubphysics.de

Jan Bergner

2023-03-18



Introduction

In the last part of this series, we re-iterated the basics for finding extrema of one-dimensional functions.

The fundamental insight was that an extremum (that is, a local one) must have a tangent of slope 0.

This idea can be extended into three dimensions and beyond. We will focus primarily on the 3D-case, i. e. functions like $z(x, y)$, as this can still be visualized.

For example, consider the function

$$z(x, y) = 4 - x^2 - y^2.$$

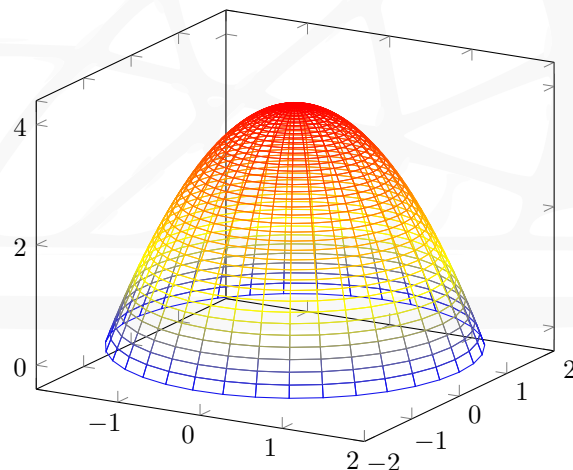


Figure 1: The function $z(x, y) = 4 - x^2 - y^2$. A paraboloid. (Which is just a rotated parabola.)

If we forget the y -term for a moment, this is just a flipped and slightly shifted parabola. We would instantly recognise that there is a maximum at $x = 0$ and $z = 4$.

Now, we can just rotate that parabola around the z -axis to find the function we are really interested in - the function $z = 4 - x^2 - y^2$.

Obviously, this function still has a maximum at $x = 0$ and $z = 4$, but with an additional constraint of y also being equal to 0.

We can picture the rotation of the simple parabola together with the rotation of its tangent at the extreme point to realize that the rotated tangent will form a plane.

As the tangent in the one-dimensional case must have a slope of 0 and we rotate around the z -axis which is orthogonal to this tangent, the plane also won't ascent or descent into any direction.

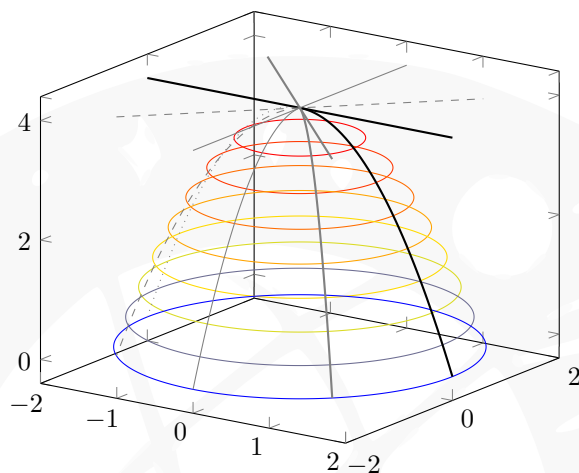


Figure 2: The function $z(x, y) = 4 - x^2 - y^2$ as a rotation. The tangents through $z = 4$ form a plane.

The reasoning with a plane that does not ascend nor descend into any direction will work for any extremum in a three-dimensional situation.

Think of it this way:

If we have a function with a maximum that is three-dimensional we should be able to balance a wooden plate right onto this maximum. And this plate cannot have a non-zero slope anywhere.

With this insight, let us find a “plane-equation” that allows us to describe a “tangent plane” to a 3-dimensional function and will let us judge whether it has zero slope into either direction. To do this, we will start with the general plane equation and after that, we will need to introduce the so-called partial derivatives.

Preliminary thoughts

The plane equation

Fortunately, it is particularly easy to obtain a vector equation for planes, starting with the vector form of the straight line equation.

Let us recall this equation for a **straight line**:

$$\vec{x}_l = \vec{x}_0 + \lambda \vec{v}$$

$\vec{x}_0$ was any given point on the straight line, $\vec{v}$ was a direction vector and λ is the number of $\vec{v}$ -steps to go from $\vec{x}_0$ to reach a certain point $\vec{x}_l$ on the straight line.

Now, we can construct a plane by simply “sliding” a straight line into any direction $\vec{w}$, that is not exactly parallel to $\vec{v}$.

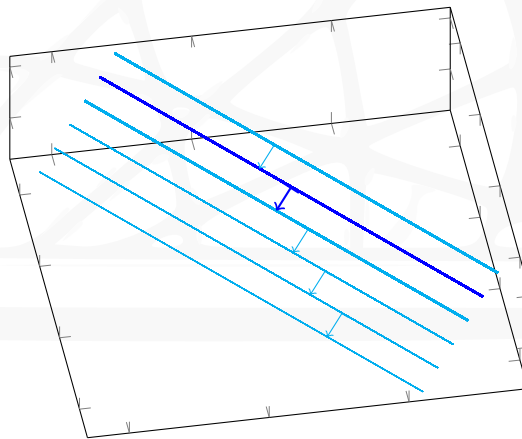


Figure 3: The (thick and dark blue) straight line slides down (and up) in (opposite) direction of the vector $\vec{w}$

The thick and dark blue straight line “slides” in $\vec{w}$ -direction to create a new straight line (light blue) that is parallel to the old one. This is repeated with the newly created straight line and the now newly-created straight line is again parallel to all the existing ones.

These “representative” straight lines form a plane, as illustrated in the following plot:

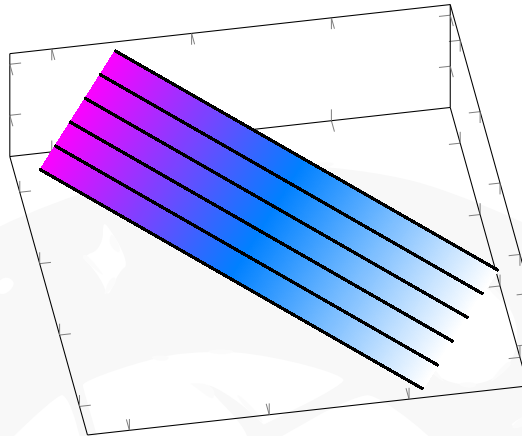


Figure 4: “Representative” straight lines form a plane. The colour indicates the z -value at a given point of the plane.

Now, how can we put actual numbers on that?

We have seen, that we can use “representative” straight lines to form a plane and that these representatives are parallel to each other. In terms of the straight-line-equation, this means, they all need to have the same direction vector $\vec{v}$ and must differ in the given “initial” point $\vec{x}_0$.

If we denote the original straight line by 0 and the i th representative by an index i , the set of straight line equations can be written like this:

$$\vec{x}_{l,i} = \vec{x}_{0,i} + \lambda \vec{v}$$

Considering the fact that any given “initial” point arises from a number of “slides” in $\vec{w}$ -direction, we can write:

$$\vec{x}_{0,i} = \vec{x}_{0,0} + i\vec{w}$$

Actually, no-one can stop us from using non-integer values for i . ($i = \frac{1}{2}$ once again, simply corresponds to a slide of $\frac{1}{2}\vec{w}$.)

Thus, we can replace $i \rightarrow \mu$, where μ is any real number, to get to any possible straight line inside the plane that is parallel to the original one.

As an equation, this reads

$$\vec{x}_{0,\mu} = \vec{x}_0 + \mu\vec{w}.$$

If we insert this, we get our full plane equation:

$$\begin{aligned}x_{l,\mu}^{\vec{}} &= x_{0,\mu}^{\vec{}} + \lambda \vec{v} \\x_{l,\mu}^{\vec{}} &= \vec{x}_0 + \mu \vec{w} + \lambda \vec{v}\end{aligned}$$

We rename $x_{l,\mu}^{\vec{}} \rightarrow x_p^{\vec{}}$ and re-order the terms on the right-hand-side, as is convention:

$$\boxed{x_p^{\vec{}} = \vec{x}_0 + \lambda \vec{v} + \mu \vec{w}}$$

This is our final plane equation.

Notice, that there are infinitely many possible choices for $\vec{w}$ that would give us the same plane, as we can slide our initial point $\vec{x}_0$ to any point on the next representative straight line.

This would correspond to choosing

$$\begin{aligned}x_{0,\mu}'^{\vec{}} &= x_{0,\mu}^{\vec{}} + \lambda' \vec{v} \\&= \vec{x}_0 + \mu \vec{w} + \lambda' \vec{v} \\&= \vec{x}_0 + \mu \underbrace{\left(\vec{w} + \frac{\lambda'}{\mu} \vec{v} \right)}_{=: \vec{w}'}\end{aligned}$$

instead of $x_{0,\mu}^{\vec{}}$ as initial point for the μ -representative.

This last fact will be important for us.

Indeed, it will allow for an easy construction with somewhat basic direction vectors.

Assume, that we want to construct a plane that does **not** have an infinite slope, i. e. there is no straight line inside it that is just vertical. Especially, we can exclude $x_p = \text{const.}$ or $y_p = \text{const.}$ (Remember, we will ultimately be on the look-out for planes that are tangent to 3-dimensional functions. If x_p or y_p were constant, this would correspond to an infinite slope.)

Let us construct vectors

$$\vec{u}_x = \begin{pmatrix} 1 \\ 0 \\ \zeta_x \end{pmatrix}, \quad \vec{u}_y = \begin{pmatrix} 0 \\ 1 \\ \zeta_y \end{pmatrix}.$$

ζ_x and ζ_y shall be arbitrary at this point and we want to show, that any plane (that satisfies the aforementioned condition) can be constructed with those two vectors.

This can be achieved, if

$$\lambda \vec{v} + \mu \vec{w} = \rho_x \vec{u}_x + \rho_y \vec{u}_y$$

is possible.

Let's work this through:

$$\begin{aligned} \lambda \vec{v} + \mu \vec{w} &= \rho_x \vec{u}_x + \rho_y \vec{u}_y \\ \begin{pmatrix} \lambda v_x + \mu w_x \\ \lambda v_y + \mu w_y \\ \lambda v_z + \mu w_z \end{pmatrix} &= \begin{pmatrix} \rho_x \\ \rho_y \\ \rho_x \zeta_x + \rho_y \zeta_y \end{pmatrix} \\ \lambda v_z + \mu w_z &= \zeta_x (\lambda v_x + \mu w_x) + \zeta_y (\lambda v_y + \mu w_y) \end{aligned}$$

In the last step, we replaced ρ_x and ρ_y in the last vector component by the first two components.

Next, consider, that λ and μ are arbitrary at this point. Thus, we can divide the whole equation by μ and redefine $\nu = \frac{\lambda}{\mu}$.

$$\begin{aligned} \lambda v_z + \mu w_z &= \zeta_x (\lambda v_x + \mu w_x) + \zeta_y (\lambda v_y + \mu w_y) \\ \nu v_z + w_z &= \zeta_x (\nu v_x + w_x) + \zeta_y (\nu v_y + w_y) \\ 0 &= \nu v_z + w_z - [\zeta_x (\nu v_x + w_x) + \zeta_y (\nu v_y + w_y)] \\ 0 &= \nu v_z + w_z - \nu \zeta_x v_x - \zeta_x w_x - \nu \zeta_y v_y - \zeta_y w_y \\ 0 &= \begin{pmatrix} -\zeta_x \\ -\zeta_y \\ 1 \end{pmatrix} \cdot [\nu \vec{v} + \vec{w}] \end{aligned}$$

At this point, the term inside the square brackets can basically be anything. (Indeed, it will be a vector inside the plane.) So, we can redefine

$$\vec{t} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} := \nu \vec{v} + \vec{w}$$

With this new vector $\vec{t}$, the question becomes, whether we can always find ζ_x and ζ_y in such a way, that the following equation is satisfied:

$$\begin{aligned} 0 &= \begin{pmatrix} -\zeta_x \\ -\zeta_y \\ 1 \end{pmatrix} \cdot [\nu \vec{v} + \vec{w}] \\ 0 &= \begin{pmatrix} -\zeta_x \\ -\zeta_y \\ 1 \end{pmatrix} \cdot \vec{t} \\ 0 &= \begin{pmatrix} -\zeta_x \\ -\zeta_y \\ 1 \end{pmatrix} \cdot \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \\ 0 &= t_z - \zeta_x t_x - \zeta_y t_y \\ \zeta_x &= \frac{1}{t_x} (t_z - \zeta_y t_y) \end{aligned}$$

Indeed, we should be able to find ζ_x and ζ_y in a way, that the equation is satisfied except for some edge cases.

The first edge case is $t_x = 0, t_y \neq 0$. In this case, ζ_x becomes irrelevant and the equation reduces to

$$t_z - \zeta_y t_y = 0,$$

- i. e. we can still find a valid combination of ζ_x and ζ_y . (If $t_y = 0$ while $t_x \neq 0$, we get the symmetric case with x and y exchanged.)

Now, what if $t_x = t_y = 0$?

Recall the definition

$$\begin{aligned}\vec{t} &= \nu \vec{v} + \vec{w} \\ \vec{t} &= \frac{\lambda}{\mu} \vec{v} + \vec{w} \\ \mu \vec{t} &= \lambda \vec{v} + \mu \vec{w}\end{aligned}$$

As we can see (and pointed out before), $\vec{t}$ is just any vector residing inside the plane. (With an arbitrary length.) Thus, its x - and y -component cannot be 0 at the same time, because otherwise, the vector would correspond to a vertical straight line, i. e. to an infinite slope. This is, what we excluded from the beginning.

So in fact, we can always describe a plane that does not include vertical straight lines by direction vectors like this:

$$\vec{u}_x = \begin{pmatrix} 1 \\ 0 \\ \zeta_x \end{pmatrix}, \quad \vec{u}_y = \begin{pmatrix} 0 \\ 1 \\ \zeta_y \end{pmatrix}$$

Partial derivatives

So far, we have dealt with one-dimensional functions like $f(x)$ and their derivative

$$\left. \frac{df}{dx} \right|_{x=x_0} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Or, if we do not look for a specific point but rather a derivative function,

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

Let us rewrite that a bit:

$$df = f'(x)dx$$

The **total differential** df on the left-hand-side can be interpreted as a tiny change to the function f , evaluated at x , which is per the right-hand-side proportional to the tiny change dx in x , where the “local proportionality factor” is the first derivative $f'(x)$.

Of course, this only works for infinitely tiny dx and df , because we approximate the function $f(x)$ around x as a straight line. (Its tangent.)

Now, we want to step it up to a function $f(x, y)$ of two variables.

Let us begin with a simple plane to get a feeling for everything. A plane can be written as a function

$$z(x, y) = m_x x + m_y y.$$

One example is given in the plot below.

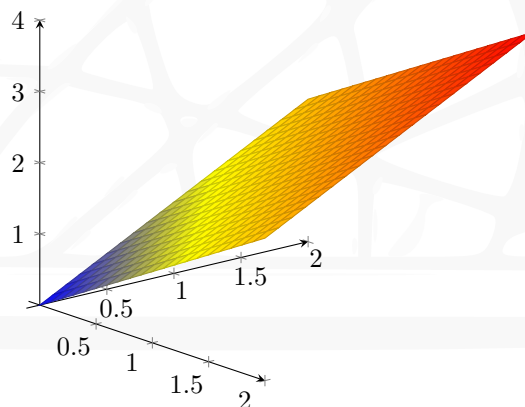


Figure 5: The simple plane $z = x + y$.

The plot shows the case of $m_x = m_y = 1$.

Now, we can define

$$z(x + \Delta x, y + \Delta y) =: z(x, y) + \Delta z$$

to write:

$$z(x, y) + \Delta z = m_x (x + \Delta x) + m_y (y + \Delta y)$$

$$z(x, y) + \Delta z = m_x x + m_x \Delta x + m_y y + m_y \Delta y$$

$$\Delta z = m_x \Delta x + m_y \Delta y$$

This one was easy. We can just replace $\Delta \rightarrow d$, to get an expression for the total differential in z :

$$\boxed{dz = m_x dx + m_y dy}$$

Now, consider a (potentially squished) paraboloid:

$$z(x, y) = m_x x^2 + m_y y^2$$

Again, defining

$$z(x + \Delta x, y + \Delta y) =: z(x, y) + \Delta z,$$

we can write:

$$\begin{aligned} z(x, y) + \Delta z &= m_x (x + \Delta x)^2 + m_y (y + \Delta y)^2 \\ &= m_x (x^2 + 2x \cdot \Delta x + \Delta x^2) + m_y (y^2 + 2y \cdot \Delta y + \Delta y^2) \\ \Delta z &= m_x (2x \cdot \Delta x + \Delta x^2) + m_y (2y \cdot \Delta y + \Delta y^2) \\ \Delta z &= m_x (2x + \Delta x) \Delta x + m_y (2y + \Delta y) \Delta y \end{aligned}$$

Now, if we replace $\Delta \rightarrow d$, i. e. we make the deltas infinitesimal, they do not really contribute to the sums, anymore. (For any finite $2x, 2y$. I. e. strictly speaking, we need to exclude $(x, y) = (0, 0)$.)

Therefore,

$$\begin{aligned} 2x + dx &\approx 2x \\ 2y + dy &\approx 2y. \end{aligned}$$

Notice, we don't get rid of the factors dx and dy in this manner!

We obtain:

$$\boxed{dz = 2m_x x dx + 2m_y y dy}$$

Let's do one more:

$$z(x, y) = mx \cdot y$$

As usual, we start with our deltas:

$$\begin{aligned} z(x, y) + \Delta z &= m(x + \Delta x) \cdot (y + \Delta y) \\ &= m(xy + y\Delta x + x\Delta y + \Delta x\Delta y) \\ \Delta z &= m(y\Delta x + x\Delta y + \Delta x\Delta y) \end{aligned}$$

As $\Delta \rightarrow d$, the term $\Delta x\Delta y$ is small, even compared to the other infinitesimal summands. (Remember, we are doing a limiting process. For any finite but small $\Delta x, \Delta y$, this last term will be approximately negligible.)

Thus:

$$\boxed{dz = mydx + mxdy}$$

We can notice a pattern. We always end up with something of the form

$$dz = g_x(x, y)dx + g_y(x, y)dy,$$

where the g -functions have some dependency on x and y .

Is this a general thing?

After all, we have seen, that this is *exactly* possible for a plane. And using tangent planes was our whole objective.

Now, how can we find those g -functions?

Consider the function

$$z(x, y) = x^2,$$

that only depend on x , but is also defined for y , i. e. for each y , we get the same $z_y(x)$ -relation, which in this case, happens to be a parabola.

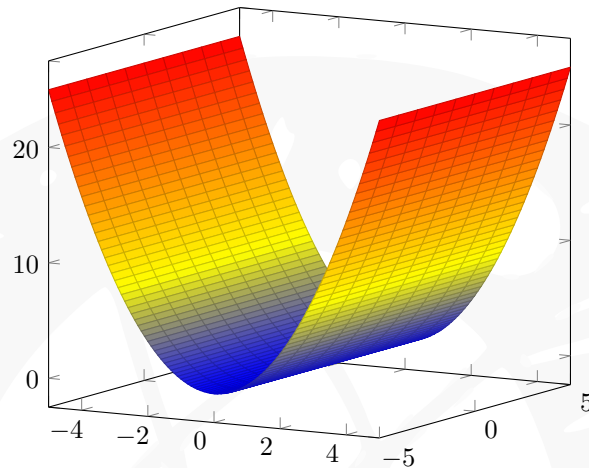


Figure 6: A parabola “sliding” in y -direction.

At any given point of this function, going a tiny step dy into the y -direction does not change the z -coordinate at all, i. e. $dz = 0$. In order to satisfy this, g_y has to be equal to 0:

$$g_y(x, y) = 0$$

Effectively, this reduces our differential equation to a 1-dimensional problem:

$$dz = g_x(x, y)dx$$

If we had a function z , that would only depend on x , we would be done at this point, since this equation really is just the definition of a first derivative.

But wait! This is exactly the case, here. $z = z(x)$ actually is a function that only depend on x , albeit being also defined for another variable y , but without formal dependency.

Therefore, we can deduce:

$$\begin{aligned} dz &= z'(x)dx \\ z'(x) &= g_x(x, y) = g_x(x) \end{aligned}$$

But what happens to a function that actually depend on y ?

Well, let us consider the function

$$z(x, y) = x \cdot y$$

again:

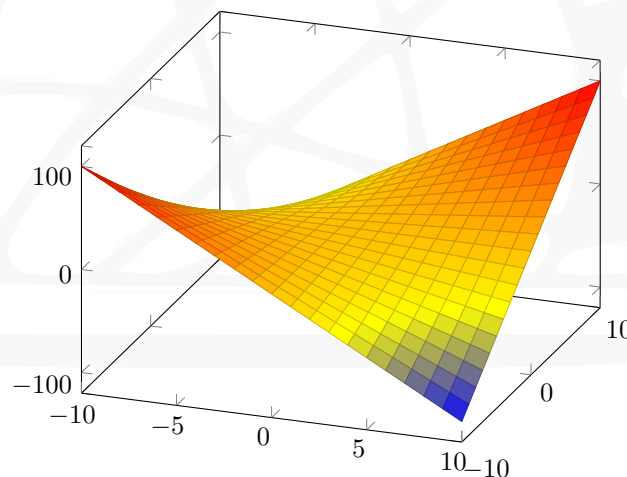


Figure 7: The function $z(x, y) = x \cdot y$.

This time, let us picture x as full variable but y as a parameter that will only assume discrete values. This is, again, the idea of choosing some “representatives” for the whole function. In this case, these are representative cross-sections at constant values of y , making our function a group of 1-dimensional functions

$$z_y(x) = y \cdot x :$$

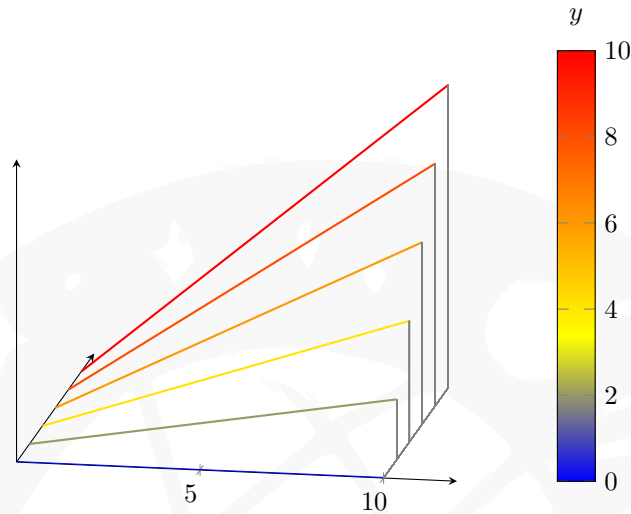


Figure 8: The group of “representative” functions $z_y(x) = x \cdot y$.

This last plot shows z versus x for selected values of the so-considered “parameter” y . The skewed gray line is a parallel line to the y -axis (i. e. $x = 10$ and $z = 0$ are kept constant), while the vertical gray lines represent a constant x and y from $z = 0$ to the $z = z_y(x = 10)$, i. e. the line straight down from the function to the x - y -plane.

As we see, for each given y , we get a straight line representing the group $z_y(x)$, whose slope is just y :

$$z'_y(x) = y$$

Alright, for every function of the group, we get a constant slope y , which is, indeed, what we expect for a straight line.

Now, we can certainly evaluate the group of functions z_y for any y we want and therefore, we can define a derivative in x for any given y . All that is left to do is the step from discrete groups of 1-dimensional functions to a continuous 2-dimensional function.

In order to do that, let us think about what a discrete, constant y actually *means* for our problem.

Recall, that for $z = z(x, y)$,

$$dz = g_x(x, y)dx + g_y(x, y)dy.$$

Dealing with a discrete group of functions means, that we are limited to one value of y and that one alone. No small shifts in y are possible, not even infinitesimal ones.

In other words:

$$dy = 0$$

However, this is something we cannot only achieve with a group of functions but also with the 2-dimensional function by requiring y to be kept constant and let the whole shift dz be due to a shift dx .

Therefore, we can find g_x for any given y_0 simply by:

$$g_x(x, y_0) = \frac{dz(x, y_0)}{dx}$$

y_0 just indicates that $y = y_0$ shall be kept constant. However, we might as well just write y instead of y_0 - as long as we keep in mind that y is meant to be kept constant.

With this, we can define the **partial derivative** g_x of z with respect to x as the total derivative (i. e. the “regular” one that we already know) when y is kept and considered constant.

The partial derivative is denoted as a differential quotient with a “rounded d” instead of a regular one:

$$\frac{\partial z(x, y)}{\partial x} = \left. \frac{dz(x, y)}{dx} \right|_{y=\text{const.}}$$

The “rounded d” ∂ is sometimes called “del” when reading out loud a formula. However, this might be confusing, because the nabla operator ∇ (we will get to that guy, later) is also sometimes referred to as “del”.

Sadly, I am not aware of any unambiguous convention on how to read the ∂ .

Notice, that the definition can seamlessly be generalized to higher dimensions.

Suppose, we have a function f that depends on N variables $\{x_1, x_2, \dots, x_N\}$:

$$f(x_1, x_2, \dots, x_N)$$

The partial derivative with respect to any variable x_i can then be defined as:

$$\frac{\partial f}{\partial x_i} := \left. \frac{df}{dx_i} \right|_{x_j = \text{const.}, j \neq i}$$

Sometimes, the partial derivative is also written using this shorthand notation:

$$\partial_x f := \frac{\partial f}{\partial x}$$

Summary

First of all, if you want to read more on partial derivatives, you can find multiple sources like textbooks, online-courses or wikipedia.

Then, the really important take-away is, that you can just take **partial derivatives** of a function f with respect to any of its arguments $\{x_1, x_2, \dots, x_N\}$ by keeping all the other arguments constant:

$$\partial_{x_i} f := \frac{\partial f}{\partial x_i} := \left. \frac{df}{dx_i} \right|_{x_j = \text{const.}, j \neq i}$$

The total differential of f becomes:

$$df = \sum_{i=1}^N \frac{\partial f}{\partial x_i} dx_i$$

Examples for using partial derivatives

In theory, we now have mastered partial derivatives, but in practice, there is usually a bit of confusion as to where partial derivatives can be used and where we need total (i. e. “regular”) derivatives.

Let us discuss some examples, so we get a better feeling for all this.

Chain rule

Consider the function

$$f(x) = (\ln x)^2 = \ln^2 x.$$

We can find the derivative as

$$f'(x) = 2 \ln x \cdot \frac{1}{x} = 2 \frac{\ln x}{x}.$$

To find this derivative, we used the hopefully well-known **chain rule**.

So far, this is nothing new. The chain rule formally states:

$$\frac{d}{dx} [f(y(x))] = f'(y(x)) \cdot y'(x)$$

Here, it is worth being careful about what depends on what. The function f depends on another function y (the logarithm in the example above), which depends on a variable x .

Thus, the whole thing ultimately depends only on x , since y depends on x and cannot be varied independent from x .

However, we completely ignore this fact when taking the derivative f' . We do that “pretending” that y is our only independent variable.

Now, what do we actually calculate here?

Remember the function

$$z(x, y) = x^2$$

from the last section.

Let us exchange the variables x and y :

$$z(x, y) = y^2$$

We can make the function $z(x, y)$ into the function $f(x)$ by setting

$$y(x) = \ln x,$$

which is like cutting through the function z , like we did before. But instead of keeping y constant, i. e. cut along a straight line, we cut along the function $y(x)$.

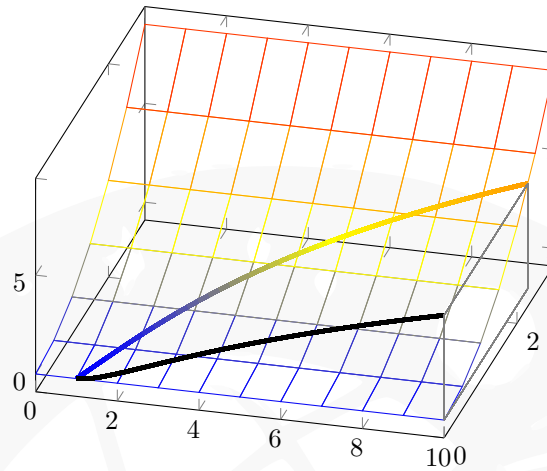


Figure 9: The function $z(x) = \ln^2 x$ (black), constructed from $z = y^2$ (mesh) and “cutting” along $y = \ln x$ (rainbow curve) and projecting into x - z -plane.

Here, we see the function $z(x, y) = y^2$ represented as the rainbow-colored mesh. The rainbow-colored curve is what we get, if we cut along $y(x) = \ln x$.

Finally, we can “project” the rainbow-curve onto the x - z -plane to get the function $z(x) = \ln^2 x$.

(We used the z -axis to represent the z -function, which coincides with $f(x)$.)

The fact that we could “project” the rainbow curve onto a plane means, that we effectively reduce the function to depend on just one instead of two variables.

In the equations, this is reflected by the fact that there *actually* is a function $y(x)$, that relates the two variables.

Recall the relation between the differentials for a function $z(x, y)$:

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy$$

In our case, there is a relation between y and x and thus, we can rewrite dy in terms of dx like

$$dy = y'(x) dx.$$

Inserting yields:

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} \cdot y'(x) dx$$

Now, we can rewrite

$$y'(x) = \frac{\partial y}{\partial x},$$

where it does not matter whether we choose a partial or total derivative, since there are no other variables that must be kept constant.

Combining these equations gets us to:

$$\begin{aligned} dz &= \left[\frac{\partial z}{\partial x} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x} \right] dx \\ \Rightarrow \frac{dz}{dx} &= \frac{\partial z}{\partial x} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x} \end{aligned}$$

Let us now consider a function of the form $f(g(x))$, i. e. a function that we would differentiate using the chain rule. Relating this to the above equation means replacing $z \rightarrow f$ and $y \rightarrow g$:

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial x}$$

As stated before,

$$\frac{\partial g}{\partial x} = g'(x)$$

and since f only directly depends on g ,

$$\frac{\partial f}{\partial x} = 0.$$

Using those insight, we get:

$$\frac{df}{dx} = \frac{\partial f}{\partial g} \cdot g'(x)$$

Now, what is up with $\frac{\partial f}{\partial g}$? Well, this is a partial derivative of f with respect to g , which means that we can replace it by the regular derivative if the function f only depends on the “variable” g , i. e.:

$$\frac{\partial f}{\partial g} = f'(g) = f'(g(x))$$

Inserting gets us back to our original chain rule:

$$\frac{df}{dx} = f'(g(x)) \cdot g'(x)$$

Amazing! Partial derivatives as they appear in the equation relating the differentials ($dz = \partial_x z dx + \partial_y z dy$) and the regular old definition of a variable change ($dy = y'(x)dx$) directly gets us to the regular old chain rule!

If you want to get a somewhat more visual explanation on the chain rule, I have to point you to the respective video on the great channel **3blue1brown**.

However, I also tried to visualize this. Unfortunately, in one single image, everything gets quite cluttered.

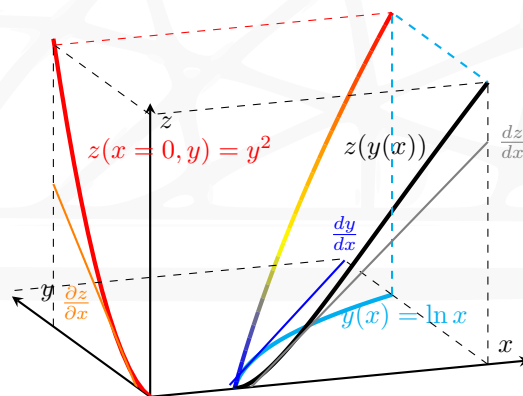


Figure 10: Where we can find partial and total derivatives.

The graphic shows our function $z(y(x)) = (\ln x)^2$ (black curve), where $y(x) = \ln x$ (blue curve) and $z(y) = z(x=0, y) = y^2$ (red curve).

The rainbow-colored curve is the function $y(x)$ “projected” onto the function $z(x, y) = y^2$, which is not completely visualized due to the lack of space. (The projection for the rightmost shown point of the function goes along the *vertical* dashed blue line.)

Once again, projecting the rainbow-curve into the x - z -plane gives us the function $z(y(x))$. (Again, the projection goes along the dashed blue line, but of course, this time, it is the other one.)

The other dashed lines are for (hopefully) better orientation. In particular, the dashed black lines don’t really have any meaning beyond showing where particular points of our functions fall inside a primary plane or onto any of the axes.

The red dashed line should remind you, that this is also the projection for the rainbow-curve into the y - z -plane.

As for the derivatives - they are not shown directly but only by one respective representative tangent.

This means, that the gray, dark blue and orange straight lines are the tangents to the black, cyan and red functions, respectively. The slope of these tangents is given by the total or partial derivatives of the functions in the tangent points.

(Notice, that we have chosen the special case of $\partial z / \partial x = 0$ to prevent even more confusion.)

Multi-dimensional chain rule

We can now generalize the chain rule to multiple dimensions. Suppose, we have two functions $x(t)$ and $y(t)$, which both depend on the same variable t .

Furthermore, consider a function $z(x(t), y(t))$ and let us construct the derivative, starting with the differentials of x and y . (Yes, we have seen this before...)

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy$$

Since x and y ultimately depend on t , let us do the corresponding variable changes:

$$dx = \frac{\partial x}{\partial t} dt$$
$$dy = \frac{\partial y}{\partial t} dt$$

Let us insert, collect like terms and re-arrange:

$$\begin{aligned} dz &= \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy \\ &= \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial t} dt + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t} dt \\ &= \left[\frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t} \right] dt \\ \Rightarrow \frac{dz}{dt} &= \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial t} \end{aligned}$$

That wasn't too bad...

We can now step it up.

Consider a scenario, where we have N independent variables $\{x_1, x_2, \dots, x_N\}$ and a set of M functions that depend on those variables:

$$\{y_1(x_1, x_2, \dots, x_N), y_2(x_1, x_2, \dots, x_N), \dots, y_M(x_1, x_2, \dots, x_N)\}$$

A useful shorthand is defining the vector

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

and say, that every function $y_i = y_i(\vec{x})$ depends on this very vector.

Now, let $z = z(y_1(\vec{x}), y_2(\vec{x}), \dots, y_M(\vec{x}))$ be a M -dimensional function that depends on those functions y_i .

We can easily construct the differential:

$$dz = \sum_{i=1}^M \frac{\partial z}{\partial y_i} dy_i$$

And now, as we can do the same for the dy_i :

$$dy_i = \sum_{j=1}^N \frac{\partial y_i}{\partial x_j} dx_j$$

All together, we get the generalized multi-dimensional chain rule:

$$\begin{aligned} dz &= \sum_{i=1}^M \frac{\partial z}{\partial y_i} \cdot \left[\sum_{j=1}^N \frac{\partial y_i}{\partial x_j} dx_j \right] \\ &= \sum_{i,j} \frac{\partial z}{\partial y_i} \cdot \frac{\partial y_i}{\partial x_j} dx_j \end{aligned}$$

Unfortunately, we cannot rewrite this as one total derivative, since the function ultimately depends on N x_j instead of just one.

You probably heard about this before, but we need to be careful about notation.

In particular:

$$\frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x} \neq \frac{\partial z}{\partial x}$$

Partial derivatives written as a “fraction” are *not really* fractions! Most importantly, we cannot “cancel factors”, but instead, we must evaluate partial derivatives individually.

Contour line equations

Once again, let us consider a function $z(x, y)$ whose differential is

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy.$$

We could also rewrite this as a vector:

$$\vec{r}(x, y) = \begin{pmatrix} x \\ y \\ z(x, y) \end{pmatrix}$$

The vector perspective will be helpful in the future. That’s why it is mentioned.

Now, let us have a look at a *contour line* in z , i. e. all the points on the function, where $z = z_0$ has a constant value.

You can picture the function $z(x, y)$ as a hilly landscape.

Then think of a hiking trail that follows the slopes of the hills, always staying precisely at the same altitude.

The altitude would correspond to the z -value of our function.

For example, the following plot shows the function

$$z(x, y) = 4 - (x - 2)^2 - (y - 2)^2$$

rainbow-shaded and the contour line

$$z = 3.8$$

in black:

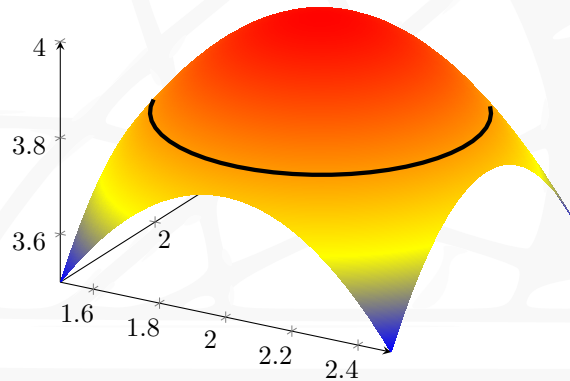


Figure 11: The function $z(x, y) = 4 - (x - 2)^2 - (y - 2)^2$ (rainbow-shaded) and the contour line $z = 3.8$ (black).

Now, let us find a general formula for a contour line.

Since $z = z_0$ will be constant, there must be a relation $y(x)$ that links x and y .

In the easiest case, we can just solve the equation

$$z_0 = z(x, y)$$

for y (or x , if we prefer) to get that relation as a function $y(x)$ (or $x(y)$, respectively).

However, it is possible that we *cannot* express the relation between x and y as a function, because we might get multiple possible y s for a given x . (The contour line in the above example is reflecting just that, since with a circle, we get two values of y for almost all values of x .)

To deal with this fact, we can define an additional parameter t and let $x = x(t)$ and $y = y(t)$.

In practice, think of yourself following that hiking path and let t indicate the time, you reach any given point on it.

Indeed, you could, for example, also let t be the total distance you have covered. But I think, there is benefit in using that rather physically motivated example.

In any case, this illustrates why it should always be possible to find such a parameterization. A physical path *can* indeed be followed - in time.

However, we need to limit ourselves to continuous contour lines for now.

To understand, why this might be an issue, picture two mountains next to each other which might both have a contour line of the same height going around them in a closed loop.

To find a general formula for any contour line to a function $z(x, y)$, let's start again with our favourite differential. (Remember, $dz = 0$ is the very definition of a contour line.)

$$dz \stackrel{!}{=} 0 = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy$$

If it is actually possible to find a *function* $y(x)$, we can re-arrange:

$$\begin{aligned} 0 &= \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy \\ \frac{\partial z}{\partial y} dy &= - \frac{\partial z}{\partial x} dx \\ \frac{dy}{dx} = y'(x) &= - \frac{\partial z / \partial x}{\partial z / \partial y} \end{aligned}$$

Notice, that the partial derivatives on the right-hand-side may themselves be functions of x and y . This means, that we cannot simply integrate the equation to get our contour-line-function $y(x)$. Calculating the partial derivatives will let you end up with a differential equation you need to solve. Depending on the original function $z(x, y)$, this might be an easy task or downright impossible.

Next, we get to the equation for the parameterized case, using

$$\begin{aligned} dx &= \frac{\partial x}{\partial t} dt =: \dot{x} dt \\ dy &= \frac{\partial y}{\partial t} dt =: \dot{y} dt. \end{aligned}$$

Be aware, that we use a dot instead of a dash to indicate the derivative with respect to time t . (Again,

this is something very common in physics. Derivatives with respect to time are so common, that it makes sense to introduce this extra symbol.)

Now, let's wrap this up:

$$\begin{aligned} 0 &= \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy \\ 0 &= \frac{\partial z}{\partial y} \dot{y} dt + \frac{\partial z}{\partial x} \dot{x} dt \\ 0 &= \frac{\partial z}{\partial y} \dot{y} + \frac{\partial z}{\partial x} \dot{x} \end{aligned}$$

This is not that easy, anymore, since we need to solve for both, $x(t)$ and $y(t)$, which is rather difficult when there is only one equation.

However, this was to be expected, as there are infinitely many possible solutions on how to parameterize.

Think of the analogy with t being the time your actual hike took. This says nothing about your *velocity* at a given point or time, respectively.

If you ran the first kilometer and after that, took a break before continuing at normal walking speed, you would get a different parameterization than if you had walked with constant speed. But still, both would have described the same trail.

The velocity (without proof at this point) is the vector

$$\vec{v}(t) = \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix},$$

- i. e. you could, for example, enforce your speed to be constant by requiring

$$(\vec{v}(t))^2 = \dot{x}^2 + \dot{y}^2 = v_0^2.$$

This would, indeed, give you a second equation to work with. (Albeit not a linear one.)

In summary, the contour line equations (depending on the actual situation) are:

$$\begin{aligned} y'(x) &= - \frac{\partial z / \partial x}{\partial z / \partial y} \\ 0 &= \frac{\partial z}{\partial y} \dot{y} + \frac{\partial z}{\partial x} \dot{x} \end{aligned}$$

Let us now get into some **examples**, again.

Consider $z(x, y) = x \cdot y$ for a start.

Here, we can simply solve for y :

$$\begin{aligned} z_0 &= x \cdot y \\ y(x) &= \frac{z_0}{x} \end{aligned}$$

Let us now use the derivative equations:

$$\begin{aligned} y'(x) &= - \frac{\partial_x z}{\partial_y z} \\ \frac{dy}{dx} &= - \frac{y}{x} \end{aligned}$$

This is a separable differential equation:

$$\begin{aligned} \frac{dy}{y} &= - \frac{dx}{x} \\ \ln |y| &= - \ln |x| + c \\ |y| &= \frac{e^c}{|x|} \end{aligned}$$

Notice, the term e^c in the right-hand-side denominator is a positive constant that would represent positive z_0 . You need to consider the possible cases for x s and y s to work out the absolute values, properly, to eventually give you the minus signs you need.

Now finally, let us use the parameterized form:

$$\begin{aligned} 0 &= \partial_y z \dot{y} + \partial_x z \dot{x} \\ &= x \dot{y} + y \dot{x} \\ &= \frac{d}{dt} (x \cdot y) \end{aligned}$$

Of course, another integration gets us back to our original equation.

Okay. Next one.

Now, let $z(x, y) = xy \cdot e^{xy}$. This one is not easily solved for either x or y .

The partial derivatives are:

$$\begin{aligned} \partial_x z &= y \cdot e^{xy} + xy^2 \cdot e^{xy} \\ &= y \cdot e^{xy} \cdot (1 + xy) \\ \partial_y z &= x \cdot e^{xy} + x^2 y \cdot e^{xy} \\ &= x \cdot e^{xy} \cdot (1 + xy) \end{aligned}$$

By inserting, most of the stuff cancels:

$$\begin{aligned} y'(x) &= - \frac{\partial_x z}{\partial_y z} \\ &= - \frac{y}{x} \end{aligned}$$

We have seen this one before. The contour lines for this function are the same as in the last example.

However - we cannot expect the z -values to also be identical since now, the integration constant that was identified with z_0 in the last example does not have the same meaning.

Tangent plane equation

We can close the loose ends, now.

Recall the vectors

$$\vec{u}_x = \begin{pmatrix} 1 \\ 0 \\ \zeta_x \end{pmatrix}, \quad \vec{u}_y = \begin{pmatrix} 0 \\ 1 \\ \zeta_y \end{pmatrix}.$$

We stated that these vectors can be used to construct the tangent plane at a point $\vec{x}_0 = (x_0, y_0, z(x_0, y_0))^T$ on a function $z(x, y)$, whose equation is given by:

$$\begin{aligned} \vec{x}_T &= \vec{x}_0 + \lambda_x \vec{u}_x + \lambda_y \vec{u}_y \\ \begin{pmatrix} x_T \\ y_T \\ z_T \end{pmatrix} &= \begin{pmatrix} x_0 \\ y_0 \\ z(x_0, y_0) \end{pmatrix} + \begin{pmatrix} \lambda_x \\ 0 \\ \lambda_x \zeta_x \end{pmatrix} + \begin{pmatrix} 0 \\ \lambda_y \\ \lambda_y \zeta_y \end{pmatrix} \end{aligned}$$

From the first two components, we can derive

$$\begin{aligned} \lambda_x &= x_T - x_0 \\ \lambda_y &= y_T - y_0 \end{aligned}$$

This can be inserted into the third component:

$$z_{\{T; x_0, y_0\}}(x_T, y_T) = z(x_0, y_0) + \zeta_x \cdot (x_T - x_0) + \zeta_y \cdot (y_T - y_0)$$

Since we are talking about a *tangent* plane, it should have the same partial derivatives as the original function around $\vec{x}_0$. (This really is the summary of everything we have done so far.)

Let us calculate these partial derivatives:

$$\begin{aligned} \frac{\partial z_{\{T; x_0, y_0\}}}{\partial x_T} &= \left. \frac{\partial z}{\partial x} \right|_{\vec{x}=\vec{x}_0} = \zeta_x \\ \frac{\partial z_{\{T; x_0, y_0\}}}{\partial y_T} &= \left. \frac{\partial z}{\partial y} \right|_{\vec{x}=\vec{x}_0} = \zeta_y \end{aligned}$$

This is nice. We have found our ζ_x and ζ_y :

$$\zeta_x = \left. \frac{\partial z}{\partial x} \right|_{\vec{x}=\vec{x}_0}$$

$$\zeta_y = \left. \frac{\partial z}{\partial y} \right|_{\vec{x}=\vec{x}_0}$$

Therefore, we can now construct the tangent plane direction vectors for each point:

$$\vec{u}_x(\vec{x}) = \begin{pmatrix} 1 \\ 0 \\ \frac{\partial z}{\partial x} \end{pmatrix}, \quad \vec{u}_y(\vec{x}) = \begin{pmatrix} 0 \\ 1 \\ \frac{\partial z}{\partial y} \end{pmatrix}$$

Notice that again, our whole reasoning can be extended into higher dimensions.

Therefore, we can also consider a function $f(x_1, x_2, \dots, x_N)$ of N variables and define N vectors $\vec{u}_{x_1}, \vec{u}_{x_2}, \dots, \vec{u}_{x_N}$, that form the N -dimensional tangent shape to the function f . (In higher dimensions, we call this a hyperplane.)

$$\vec{u}_{x_i} = \frac{\partial}{\partial x_i} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \\ f(x_1, x_2, \dots, x_N) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ \frac{\partial f}{\partial x_i} \end{pmatrix}$$

The *one* in the right-hand-side-vector is, of course in the i th component.

What's more, we just introduced the the “operator notation” for a partial derivative:

$$\frac{\partial}{\partial x} \bullet := \frac{\partial \bullet}{\partial x}$$

When applied to a vector $\vec{u} = (u_1, u_2, \dots, u_M)^T$, the partial derivative operator acts component-wise:

$$\frac{\partial}{\partial x} \vec{u} = \begin{pmatrix} \frac{\partial}{\partial x} u_1 \\ \frac{\partial}{\partial x} u_2 \\ \vdots \\ \frac{\partial}{\partial x} u_M \end{pmatrix}$$

The gradient

It is now useful to define the so-called **gradient vector** for a “test function” $f(x, y)$:

$$\nabla f(x, y) := \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

Of course, we can extend into higher dimensions.

A function

$$g = g(x_1, x_2, \dots, x_N)$$

that depends on N variables would have a gradient vector operator like

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_N} \end{pmatrix},$$

meaning

$$\nabla g = \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \\ \vdots \\ \frac{\partial g}{\partial x_N} \end{pmatrix}.$$

The operator ∇ is called the “nabla-operator” which is sometimes also pronounced as “del”.

In general, the nabla-operator acts on a scalar function that depends on N variables by returning the N -component-vector of the N partial derivatives of the function to either of the variables.

Dimensions

Notice, that a function depending on N variables generally lives in a $N + 1$ -dimensional space.

For example, a function $f(x, y) = 4 - x^2 - y^2$ (a flipped and shifted paraboloid) depends on *two* variables but lives in a 3-dimensional space and could be represented as a vector

$$\vec{f} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ f(x, y) \end{pmatrix}.$$

In contrast, the gradient vector

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

is only two-dimensional, i. e. living in the x - y -plane.

We can visualize this in two dimensions using a colour map for the function and representative vectors for the gradient.

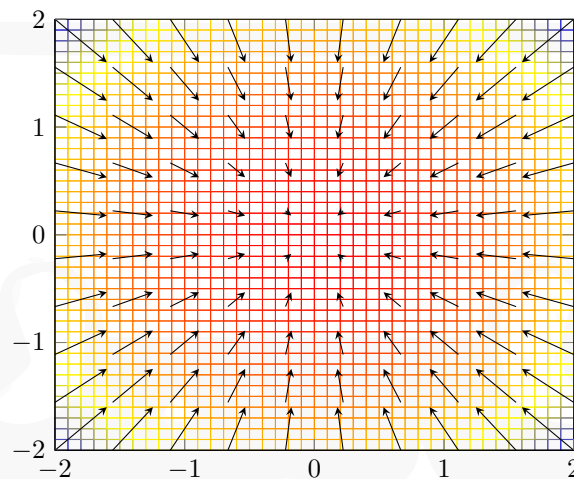


Figure 12: The function $f(x, y) = 4 - (x - 2)^2 - (y - 2)^2$ (rainbow-shaded) and the gradient vector field (black). Red means high values for f and blue means low ones.

The more the colour is in the red area, the higher the value of the function f is. The more it is blue, the lower is the function's value.

As for the gradient vector field, the length of the arrows is $\frac{1}{10}$ of the proper length for better visualization.

Interestingly, the vectors all seem to point to the local maximum...

Steepest ascent

Indeed, the gradient always points into the direction of locally steepest ascent.

Be aware, that the gradient vector has one dimension less than the function it is applied to. Therefore, it is not a tangent vector to the function.

In the case of a function of two variables, the gradient is a direction vector in these two dimensions, i. e. in the x - y -plane, while in contrast, “ascent” would seem to imply going upwards, i. e. increasing your z -coordinate.

Let us prove this considering a function

$$f(x_1, x_2, \dots, x_N) =: f(\vec{x})$$

that has a gradient

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{pmatrix}.$$

Now consider a **unit vector**

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix},$$

meaning it must have a length of 1. (Indeed, it does not have to be a unit vector, as long as we use a general vector of fixed length.)

Now, we can define any point in the function's parameter space that is reachable by going from $\vec{x}$ into $\vec{u}$ -direction as

$$\vec{y}(t) = \vec{x} + t \cdot \vec{u},$$

where t is telling us, how many "steps" we have to go into $\vec{u}$ -direction.

Therefore, we can calculate the rate of change in f as derivative $\frac{df}{dt}$, where f is now considered to be a function of $\vec{y}$:

$$\frac{d}{dt}f(\vec{y}) = \nabla_{\vec{y}}f \cdot \frac{d}{dt}\vec{y} = \nabla_{\vec{y}}f \cdot \vec{u}$$

Here, we gave the nabla operator an index $\vec{y}$ to indicate, that the partial derivatives are with respect to $\vec{y}$'s components.

Now, the rate of change is just the dot product of the gradient times the direction vector $\vec{u}$. This is, of course, largest, when both point into the same direction.

q.e.d.

In turn $-\nabla z$ is the direction of steepest *descent*.

In any case, we can visualize this for a plane. (Remember, that for an infinitely small area around a given point on a function of two variables, the local behaviour can be approximated by the tangent plane to this point.)

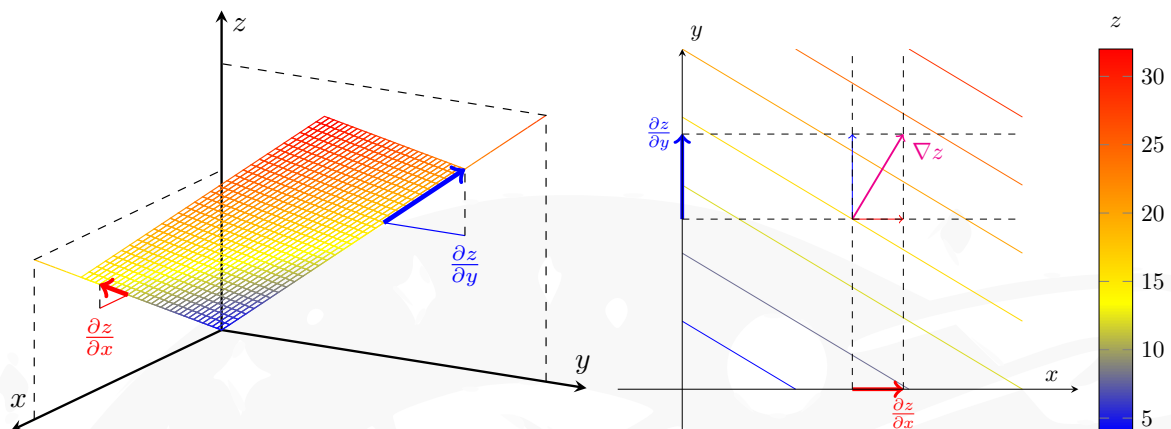


Figure 13: A plane with ascend vectors (left) whose horizontal components are equal to the respective partial derivative. On the right, this is the same plane represented as contour lines. The ascend vectors combined form the gradient, which is pointing into the direction of steepest ascent.

The left plot shows a plane with “ascend vectors”. Notice, that their horizontal component must be equal to the partial derivative if our reasoning should work.

The right plot shows the same plane, represented as contour lines. Combining the “ascend vectors” (or rather, the components that are not in z -direction), we get the gradient. (magenta)

We can visually see that it is orthogonal to the contour lines, which, for a plane, is the obvious direction of steepest ascent.

Notice, that for the right plot, the axes are equally scaled.

Additionally, we notice, that going along the y -axis, the next contour line is reached faster than by going along the x -axis. Of course, this means, that the plane is steeper in y -direction than in x -direction. This is also, why the “ y -ascend-vector” (blue) is longer than the “ x -ascend-vector”. (Which also goes for their horizontal components.)

En passant, we realized another important fact.

The direction of steepest ascent, i. e. the direction of the gradient is orthogonal to a plane’s contour line.

In the next chapter of this series, we will show that this holds for other functions, as well.

(Finally) Finding extrema

Now, we can finally turn to finding actual (local) extrema. Admittedly, there is no longer much to it.

Recall our tangent plane equation

$$\vec{x} = \vec{x}_0 + \lambda \vec{u}_x + \mu \vec{u}_y,$$

where

$$\vec{u}_x = \begin{pmatrix} 1 \\ 0 \\ \zeta_x \end{pmatrix} \quad \text{and} \quad \vec{u}_y = \begin{pmatrix} 0 \\ 1 \\ \zeta_y \end{pmatrix}.$$

Using

$$\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \text{and} \quad \vec{x}_0 = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix},$$

we can rewrite the plane equation to:

$$z = z_0 + \zeta_x (x - x_0) + \zeta_y (y - y_0)$$

The gradient of this plane “function” is

$$\nabla z = \begin{pmatrix} \frac{\partial z}{\partial x} \\ \frac{\partial z}{\partial y} \end{pmatrix} = \begin{pmatrix} \zeta_x \\ \zeta_y \end{pmatrix}.$$

Now, a **horizontal plane** must not change its z -value, meaning the z -components of the direction vectors $\vec{u}_x$ and $\vec{u}_y$ must be 0:

$$\zeta_x = \zeta_y = 0$$

In turn, this means, the gradient has to be 0, as well.

This makes sense, as the gradient being the zero-vector (which is the only one that has no orientation) implies, that there *is actually* no direction of steepest ascent. (Remember, this is, where the gradient points to, generally speaking.)

After all, the only way that there is *no* steepest ascent direction in a plane is, if there is no ascent at all, which is indeed what we want.

I might be repeating myself, but remember that the gradient of a function at a given point is identical to its tangent plane's gradient.

Therefore, we obtain the **required condition for a local extremum** of a function at a given point to be the gradient being zero:

$$f(\vec{x}_E) \text{ being an extremum} \Rightarrow \nabla f|_{\vec{x}=\vec{x}_E} \stackrel{!}{=} 0$$

Notice, this equation, again, holds for higher dimensions. The steepest-ascent-property had been derived for a general N -dimensional case.

The condition for an extremum is always a tangent hyperplane (the N -dimensional pendant to the regular plane, as mentioned) having no ascent, i. e. there is no direction of steepest ascent which in turn, means, that the gradient must become the N -dimensional zero-vector.

This is equivalent to all partial derivatives being zero.

HEADS UP!

We have only derived a **required** condition for a local extremum. A **sufficient** condition would also require to check, whether the function actually in- or decreases into either direction.

Examples

Let us do some example calculations.

$$z = 4 - x^2 - y^2$$

We had started this whole article with

$$z(x, y) = 4 - x^2 - y^2.$$

The gradient is:

$$\nabla z = \begin{pmatrix} -2x \\ -2y \end{pmatrix} = -2 \begin{pmatrix} x \\ y \end{pmatrix}$$

There is one way to get a zero-vector:

$$\nabla z|_{\vec{x}=\vec{x}_E} = -2 \begin{pmatrix} x_E \\ y_E \end{pmatrix} \stackrel{!}{=} \vec{0}$$

$$\Leftrightarrow x_E = 0, \quad y_E = 0$$

This is the extremum we expected. (In fact, this is a maximum.)

$$z = (x^2 + y^2)^2 - x^2 - y^2 + \frac{1}{4}$$

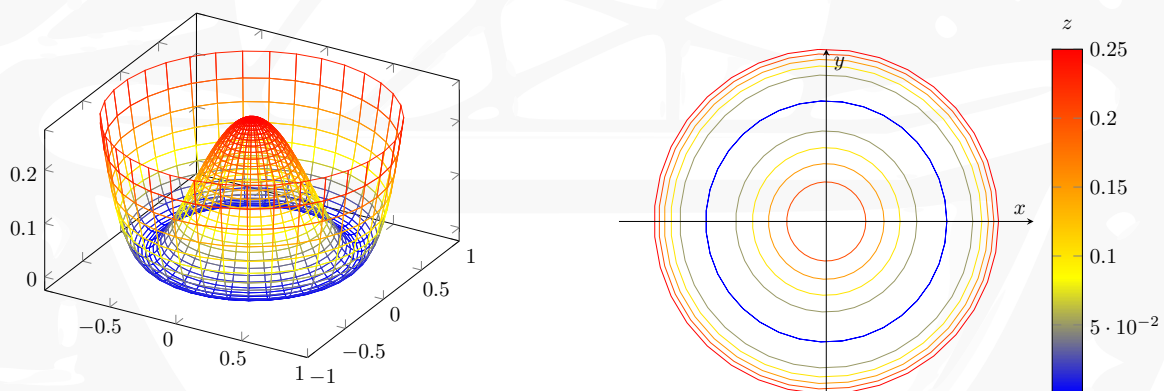


Figure 14: The “mexican-hat”-function.

This function, which is also sometimes known as a mexican hat function, is a bit more intricate:

$$\begin{aligned} z(x, y) &= (x^2 + y^2)^2 - x^2 - y^2 + \frac{1}{4} \\ &= x^4 + y^4 + 2x^2y^2 - x^2 - y^2 + \frac{1}{4} \end{aligned}$$

Next, let us calculate the gradient:

$$\nabla z = \begin{pmatrix} 4x^3 + 4xy^2 - 2x \\ 4y^3 + 4x^2y - 2y \end{pmatrix} = \begin{pmatrix} 2x \cdot (2x^2 + 2y^2 - 1) \\ 2y \cdot (2y^2 + 2x^2 - 1) \end{pmatrix}$$

In the interest of thoroughness, here is an example gradient plot.

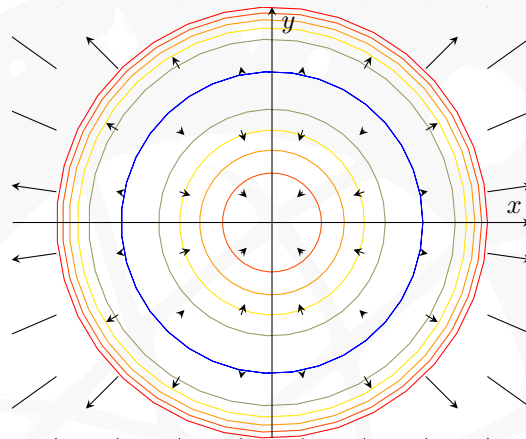


Figure 15: The “mexican-hat”-function and its gradient.

For the extrema, we get the rather obvious case of

$$(x_E, y_E) = (0, 0),$$

which actually is a local maximum.

But now, look at the term inside the inner brackets:

$$2x_E^2 + 2y_E^2 - 1$$

This one is identical in both components and can also become zero to give us a set of critical points:

$$\begin{aligned} 2x_E^2 + 2y_E^2 - 1 &\stackrel{!}{=} 0 \\ x_E^2 + y_E^2 &= \frac{1}{2} \end{aligned}$$

This is just the equation of a circle. In particular, it is the blue line in the contour plot.

However, let us insert that into the original function:

$$\begin{aligned}
 z(x_E, y_E) &= \left(\overbrace{x_E^2 + y_E^2}^{\frac{1}{2}} \right)^2 - \left(\overbrace{x_E^2 + y_E^2}^{\frac{1}{2}} \right) + \frac{1}{4} \\
 &= \left(\frac{1}{2} \right)^2 - \left(\frac{1}{2} \right) + \frac{1}{4} \\
 &= \frac{1}{4} - \frac{1}{2} + \frac{1}{4} \\
 &= 0
 \end{aligned}$$

As we can see, all the points on the circle get us to the same value for the function. ($z(x_E, y_E) = 0$)

Thus, none of these points truly is a minimum, as there are always two directions to go (along the circle, that is), where the function does not increase but stays constant.

On the other hand, all points on the circle lead to the lowest possible value for the function. So, one could argue, that our definition of a minimum (or an extremum, in general) is too narrow.

After all, nature has that tendency to minimize things. If the mexican-hat-function were a landscape and right at the top of the central hill we had a fountain, the water would flow down the hill and hit one of the points on the circle before finding a resting position.

Of course, we cannot just fill that valley with water. May be we have a fountain that vertically shoots droplets into the air.

These droplets might be ejected with a low rate, so one that reaches the bottom of the valley can stay there and evaporate, before the next one emerges.

If you really want to simulate the mathematically optimal analogy by a physical setup, you'd have to make sure, that the droplet's friction with the hill is large enough to basically drain all its kinetic energy, while simultaneously being small enough, so that the droplet can still follow gravity to the bottom, creating a trajectory that always follows the steepest descent, i. e. the negative gradient.

$$z = \sin x \cdot \cos y$$

Let's talk about another one:

$$z(x, y) = \sin x \cdot \cos y$$

The gradient is:

$$\nabla z = \begin{pmatrix} \cos x \cdot \cos y \\ -\sin x \cdot \sin y \end{pmatrix}$$

To find an extremum, both components must be zero.

Since the cosine of a number and the sine of that number can never be 0 simultaneously, we can only get a zero vector as the gradient if

$$\sin x = 0, \cos y = 0 \text{ or } \cos x = 0, \sin y = 0.$$

Thus, we could have the following sets of solutions:

$$(A) \quad x = n\pi, \quad y = \frac{\pi}{2} + m\pi, \quad n, m \in \mathbb{Z}$$

$$(B) \quad x = \frac{\pi}{2} + n\pi, \quad y = m\pi, \quad n, m \in \mathbb{Z}$$

Let us consider these sets individually. Set A implies $\sin x = 0$ and thus:

$$z(x_A, y_A) = 0.$$

But since this would hold for every value of y , we once again cannot have a true extremum.

Here, these are saddle points. (Or mountain passes, if we once again use the analogy of a hilly landscape.)

Now, let us look at situation B .

$$\begin{aligned} z(x_B, y_B) &= \sin(x_B) \cdot \cos(y_B) \\ &= \sin\left(\frac{\pi}{2} + n\pi\right) \cdot \cos(m\pi) \end{aligned}$$

Both, the sine and the cosine can give us values of $+1$ or -1 . Thus, these are the values, the function can assume:

$$z(x_B, y_B) = \pm 1$$

Obviously, $z = -1$ are the minima and $z = +1$ are the maxima.

$$f = (x - x_0)^2 - (y - y_0)^2 + (z - z_0)^2$$

Let us step it up one dimension.

$$f(x, y, z) = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2$$

We obtain the gradient to be

$$\nabla f = 2 \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix}.$$

The critical point for an extremum is, therefore:

$$\begin{pmatrix} x_E \\ y_E \\ z_E \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$$

Now, let us check, if this is indeed an extremum.

To do that, let us shift away from the candidate point by a small step of length h (i. e. h is positive) into either direction.

The direction vector can simply be the unit vector in r -direction in spherical coordinates:

$$\vec{e}_r = \begin{pmatrix} \sin \theta \cdot \cos \phi \\ \sin \theta \cdot \sin \phi \\ \cos \theta \end{pmatrix}$$

We can do that shift by setting

$$\vec{x} = \vec{x}_E + h\vec{e}_r$$

and let h become infinitely small.

That gets us to:

$$\begin{aligned} z(\vec{x}_E + h\vec{e}_r) &= (x_E + h \sin \theta \cdot \cos \phi - x_0)^2 \\ &\quad + (y_E + h \sin \theta \cdot \sin \phi - y_0)^2 \\ &\quad + (z_E + h \cos \theta - z_0)^2 \\ &= (h \sin \theta \cdot \cos \phi)^2 + (h \sin \theta \cdot \sin \phi)^2 + (h \cos \theta)^2 \\ &= h^2 \left[(\sin \theta \cdot \cos \phi)^2 + (\sin \theta \cdot \sin \phi)^2 + (\cos \theta)^2 \right] \\ &= h^2 \cdot \vec{e}_r^2 \\ &= h^2 \end{aligned}$$

Indeed, we don't get any dependency on the direction. Our function's value around our candidate point just depends on the length of the step we do into either direction.

Therefore, going away from our candidate point will always increase the function's value and we really have a true minimum.

Summary

This article was quite a feat.

Indeed, it is the longest one I have written so far (that also took me the largest amount of time) and we had to introduce a huge amount of concepts and ideas that will accompany us going forward in this series of articles and quite possibly, also in others.

Thus, I want to summarize the mayor points for you.

- We discussed how to construct an equation for a **(hyper)plane**.
- We reasoned, that we can always find a hyperplane that is **tangent** to a function at a given point.
- To find extrema, we could use the **required condition**, that moving inside the tangent hyperplane must not change the "function's coordinate", i. e. there can be no ascend or descent in either direction.
- To characterize the rate of change of a function into either coordinate's direction, we introduced **partial derivatives**. The partial derivative with respect to any coordinate is what we would get as a **total derivative** if all the other variables were just constants.
- We highlighted the fact, that one should be careful about what really is a depending variable and what is an independent variable when evaluating partial versus total derivatives.

Furthermore, we have discussed the relation between **total differentials**, partial derivatives and the **chain rule**.

- Using partial derivatives, we could construct a tangent hyperplane. This came with the insight, that a partial derivative of a function, evaluated at a given point, is equal to the partial derivative of the tangent hyperplane at this point.
- We defined the **gradient** vector of a function using partial derivatives.
- We could prove, that the gradient will always point into the direction of **steepest ascent**.
- With the steepest-ascent-property of the gradient, we could reason, why at any local extremum, the gradient vector must be the zero-vector.
- Finally, we managed to find actual extrema of some example functions using the gradient.

Now, let that sink in for a moment and make sure, you have a decent understanding of everything inside this article - especially the distinction between partial and total derivatives, as this is fundamental for almost everything we will be doing going forward.