
Lagrange-Hamilton-series part 1 - Extrema of one-dimensional functions

An article posted on bathtubphysics.de

Jan Bergner

2022-11-02



Motivational prelude

I like discussing problems - usually problems in theoretical classical mechanics - that require the Lagrange or Hamilton formalism for finding a solution.

This is why I decided that it might be useful to give a more thorough introduction into the topic.

Admittedly, this will be for my benefit as well, since I will have to brush up on some details.

In any case, my goal is that someone with a high-school-level math education should be able to follow my articles and get the basic gist of everything up to the point of applying this stuff.

However, I am a physicist by education and by temperament and thus, I don't promise final mathematical rigor. The result will probably be something, that you could use as a university crash course for physicists.

Minima and maxima of one-dimensional functions

We start gently with a recap of stuff that should be familiar from high school.

Suppose, we are given a function $y(x)$ and want to find the maxima and minima of this function.

There are two possible scenarios to consider:

1. The function is minimal or maximal when $x \rightarrow \pm\infty$.
2. At some (or multiple) point $x = x_E$, the function has a local minimum or maximum.

Local extrema and extrema "at" $\pm\infty$

Let us have a look at the following graph.

We are shown three example functions with their (local) minima and maxima. In particular:

$$y_1(x) = \frac{1}{10} \cdot (x + 1)^2 - 2$$

$$y_2(x) = 4 \cdot e^{-\frac{(x-1)^2}{4}}$$

$$y_3(x) = \sin x - 1$$

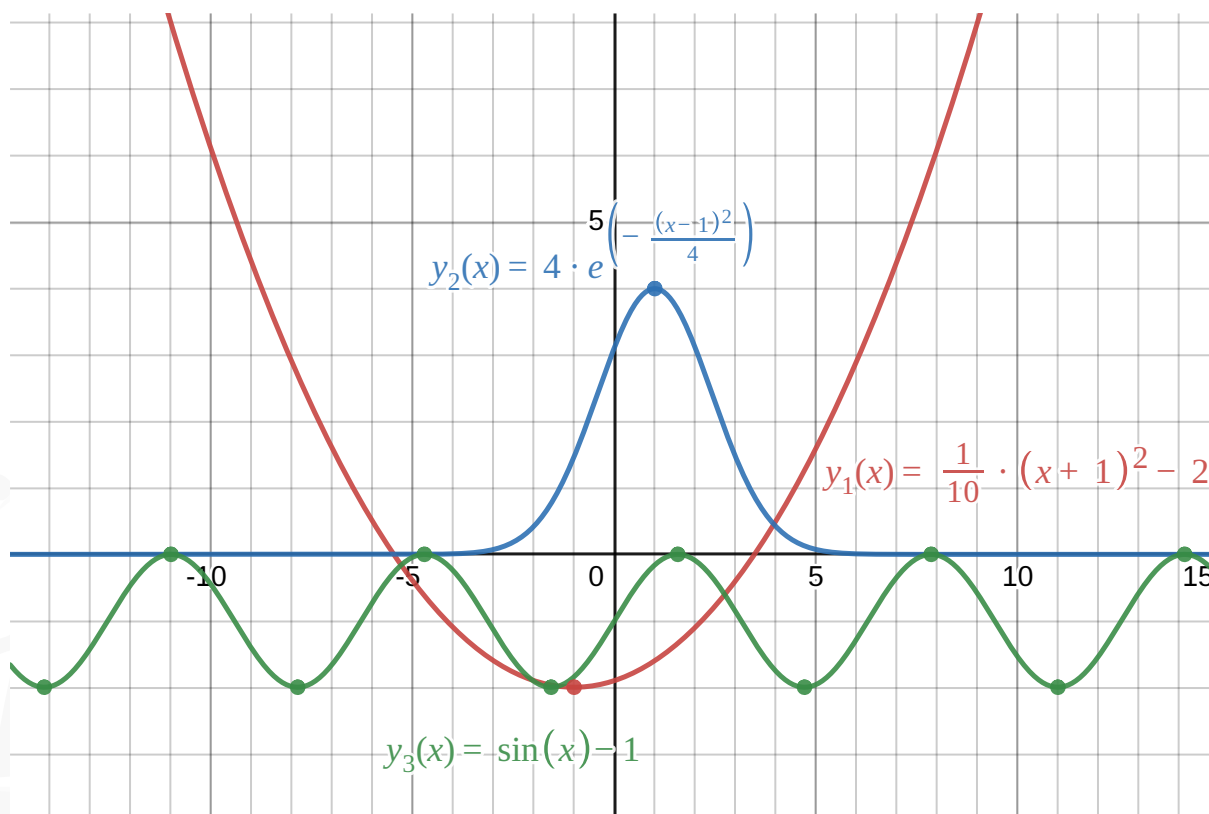


Figure 1: Some functions with their minima and maxima.

- y_1 (red)

Our first function is a shifted and squished parabola. There is a local (and in this case also a global) minimum at $(-1, -2)$. A local maximum does not exist, but the function blows off to $+\infty$ when $x \rightarrow \pm\infty$.

- y_2 (blue)

This one is a slightly scaled and shifted bell curve. Here, we see a local (and also global) maximum at $(1, 4)$, while there is no local minimum. However, as $x \rightarrow \pm\infty$, the curve approaches 0, which is, indeed an upper boundary for the function's lowest value. (A so-called infimum.)

- y_3 (green)

Here, we have a sine function that is shifted "down" one unit.

In this case, it is not really clear, what happens "at" $\pm\infty$, but instead there is an infinite number of local (and again, global) maxima and minima. In particular,

Maxima: $(\frac{\pi}{2} + 2\pi n, 0)$, $n \in \mathbb{Z}$

Minima: $(-\frac{\pi}{2} + 2\pi n, -2)$, $n \in \mathbb{Z}$

To summarize, we have seen three types of behaviours for $x \rightarrow \pm\infty$. A function might approach a finite value, it might “blow off” or it might not approach anything. (This is just by the provided examples, of course. One might discuss additional “classes” of functions.)

Unfortunately, there is not a lot we can do about this except for actually checking these limits. That is, why we won't really consider extreme behaviour for $x \rightarrow \pm\infty$ from here on.

Local extrema and their tangents

For the local extrema, we can make a crucial observation.

There is one thing that all local extrema have in common. Tangents to the function $y(x)$ through an extreme point (x_E, y_E) are always horizontal lines, i. e. they can be described as function

$$y_T(x) = y_E,$$

themselves.

Alternatively, we can use the derivatives at x_E , which must vanish:

$$\left. \frac{dy}{dx} \right|_{x_E} = \frac{dy_T}{dx} = 0$$

The following plot illustrates this for the example functions $y_1(x)$ and $y_2(x)$.

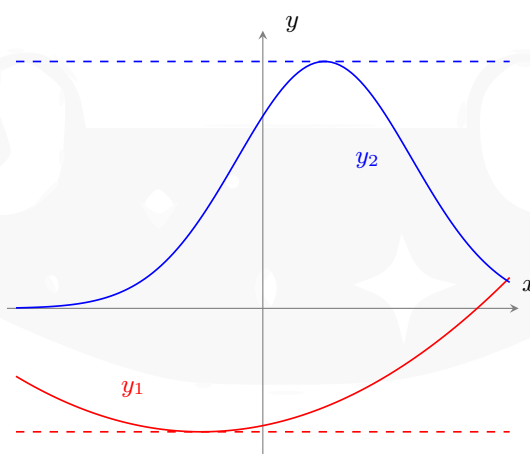


Figure 2: Tangents to the local extrema of y_1 and y_2 .

Let us re-iterate, how to find a tangent to a function in general.

Tangents to functions

General straight lines

A tangent is, of course, a straight line (l), which in general, has the following equation:

$$y_l = m_l \cdot x_l + n_l$$

m_l is the slope of the straight line and n_l is the y -axis-intersection coordinate. (This does not work for vertical lines, which are described by $x_l = x_C$, where x_C is the constant x -value, where the straight line resides.)

Alternative form of the straight line equation

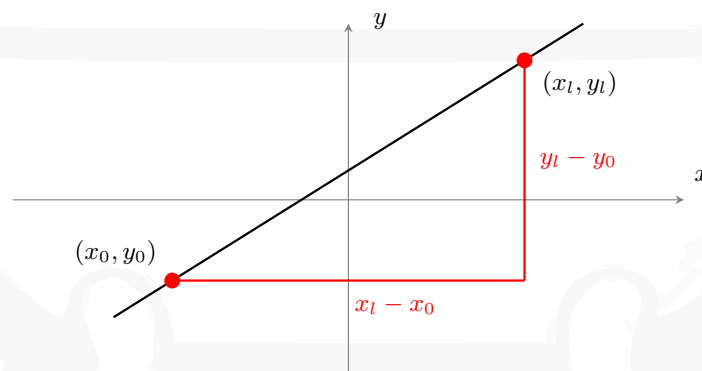


Figure 3: The slope of a straight line with one given and one variable point.

If one point (x_0, y_0) on the line is known, we can use the alternative form

$$\frac{y_l - y_0}{x_l - x_0} = m_l$$

to describe the straight line, which can be re-arranged to get back to the original form:

$$\begin{aligned}\frac{y_l - y_0}{x_l - x_0} &= m_l \\ y_l - y_0 &= m_l (x_l - x_0) \\ y_l &= m_l \cdot x_l - \underbrace{m_l \cdot x_0 + y_0}_{+n_l}\end{aligned}$$

As we can see, by identifying

$$n_l = y_0 - m_l \cdot x_0,$$

we get the usual straight-line-equation from the alternate form.

However, we could also replace

$$\begin{aligned}x_l &\rightarrow x \\ y_l &\rightarrow y(x),\end{aligned}$$

where $y(x)$ is some function, we want to find a tangent to.

If we let $x \rightarrow x_0$, we get the slope of the tangent through (x_0, y_0) , which is, in turn, the definition of $y(x)$'s first derivative $y'(x_0) = \left. \frac{dy}{dx} \right|_{x_0}$, evaluated at x_0 :

$$y'(x_0) = \left. \frac{dy}{dx} \right|_{x_0} = \lim_{x \rightarrow x_0} \frac{y(x) - \overbrace{y(x_0)}^{y_0}}{x - x_0} = m_T$$

(m_T merely indicates, that now, we are no longer talking about any straight line's slope, but about the slope of a tangent.)

Straight line equation with vectors

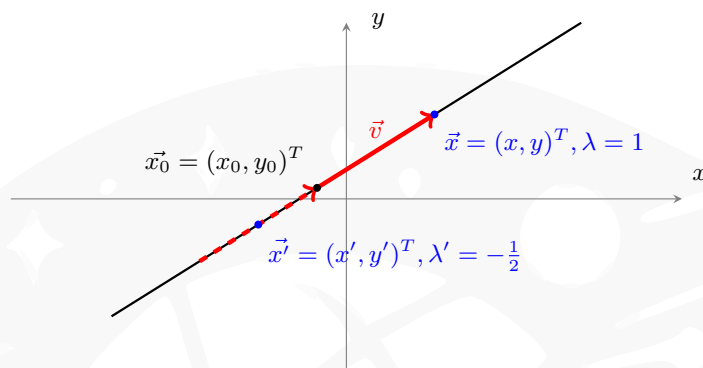


Figure 4: The straight line using one fixed point (x_0, y_0) and a direction vector \vec{v} .

The straight line can also be written using vectors

$$\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\vec{x}_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

like so:

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

where \vec{v} is a direction vector and λ is a real number.

In fact, \vec{v} is not one vector, but there are infinitely many possible vectors \vec{v} , as we only care for the direction of the vector, not its length. You can think of it as a mere direction on a compass.

This means, that λ is a real number that will tell us, how many “steps” of \vec{v} we need to go, starting from \vec{x}_0 , to reach a certain point on the tangent.

(If we make \vec{v} shorter, we can still reach any point on the straight line by making λ 's absolute value larger, in turn.)

Next, let us find that vector

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}.$$

In fact, the vector equation can be re-written as two “normal” equations:

$$x = x_0 + \lambda v_x$$

$$y = y_0 + \lambda v_y$$

Dividing the first equation by v_x , we get

$$\lambda = \frac{1}{v_x} (x - x_0),$$

which can be inserted into the second equation:

$$\begin{aligned} y &= y_0 + \lambda v_y \\ &= y_0 + v_y \left[\frac{1}{v_x} (x - x_0) \right] \\ &= y_0 + \underbrace{\frac{v_y}{v_x}}_{=m_T} (x - x_0) \end{aligned}$$

There, we have it.

If we choose v_x and v_y in a way, that

$$\frac{v_y}{v_x} = m_T,$$

the vector equation will always reproduce the straight line equation.

For the sake of simplicity, we can choose:

$$v_x = 1$$

$$v_y = m_T$$

With this choice, the tangent is described by

$$\vec{x} = \vec{x}_0 + \lambda \begin{pmatrix} 1 \\ m_T \end{pmatrix}.$$

Notice, there are other conventions about how to choose the components of the direction vector \vec{v} .

For example, the vector might be normalized to a length of 1, which has the advantage of making the value of λ more meaningful as a physical distance.

In our case, however, we only want an easy way to construct any valid direction vector.

Using tangents to find local extrema

As pointed out before, every local extreme point (x_E, y_E) on a function $y(x)$ will have a horizontal tangent, i. e. a tangent that can be described by a function

$$y_T(x) = y_E,$$

whose slope - i. e. the first derivative of the function $y(x)$ - is 0:

$$y'(x_E) = \left. \frac{dy}{dx} \right|_{x_E} = m_T = 0$$

This usually can be used to find extrema by calculating the derivative, set it to zero and solve for x_E .

But be aware! This is only a *required* condition, not a sufficient one.

This means, that we cannot conclude to have found an extremum, just because the derivative is zero.

Let us look at some example plots to understand this.

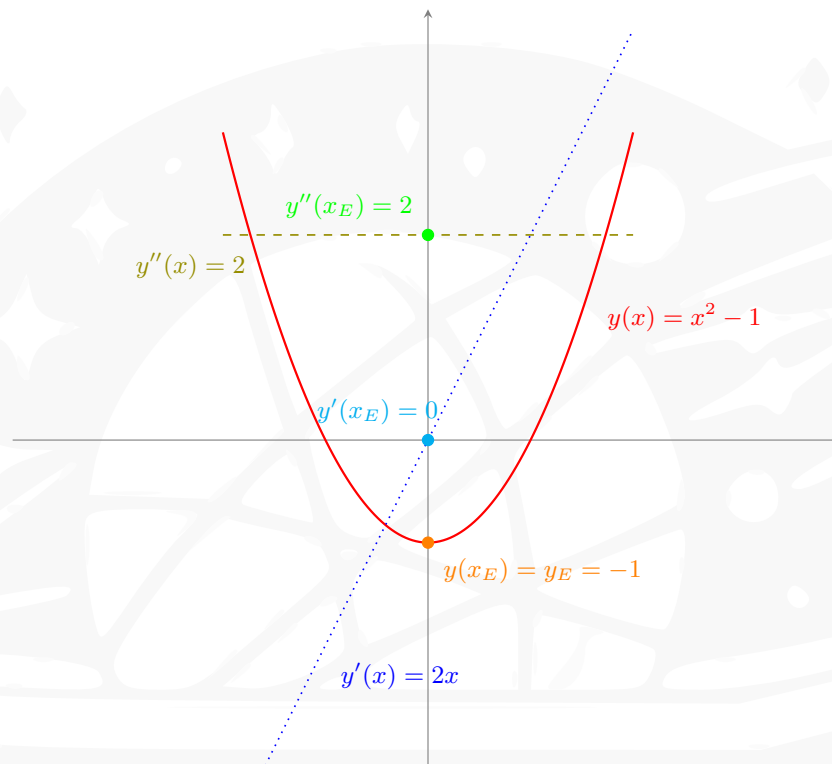


Figure 5: A (shifted) parabola with its first two derivatives.

This plot shows a regular parabola that is shifted down by 1 unit (red curve). The dotted blue curve is the first derivative of that parabola, while the dashed green curve is the second derivative.

Notice the local (and also global) minimum at $x_E = 0$.

The fact that at the minimum, the tangent's slope is 0 corresponds to the first derivative at $x = x_E$ being 0.

As for the second derivative, in this case, it's just the constant number 2.

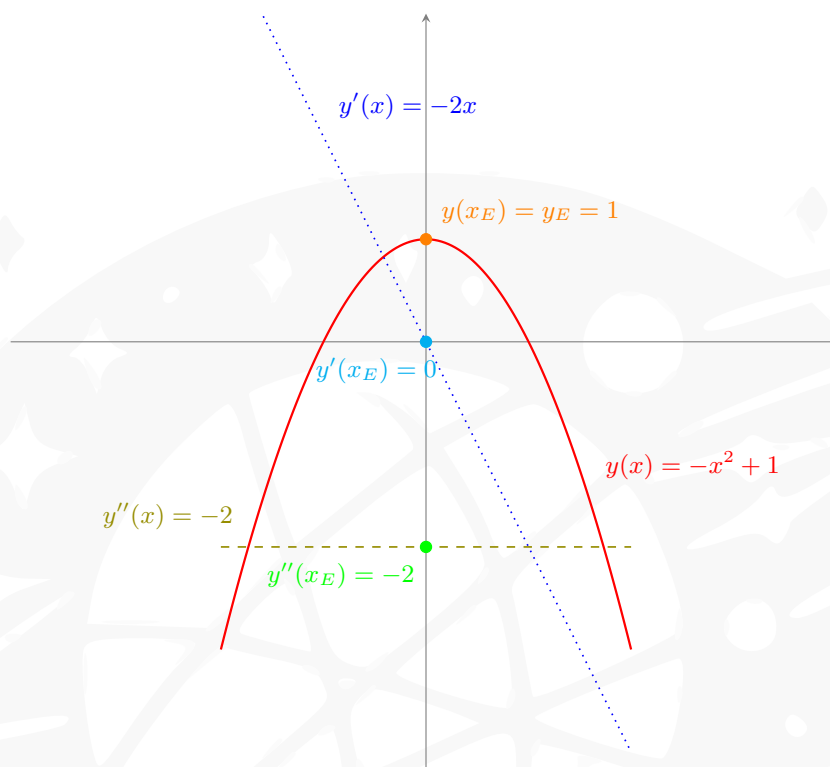


Figure 6: A (shifted and flipped) parabola with its first two derivatives.

The function in this graph is the same parabola as before, except that it's flipped over the x -axis.

Once again, we have an extremum - which is a maximum in this case - at $x_E = 0$.

Yet again, the first derivative at $x_E = 0$ is 0 as well.

However, the second derivative, while again being just a constant number, now has a value of -2 .

This is reflective of the fact that now, the first derivative has a constant negative slope of -2 while before, it had constant positive slope of $+2$.

Indeed, this is a more general point.

If the **second derivative** at the extremum is **positive**, so is the **slope** of the **first derivative** around the extreme point.

This implies that we must have a **negative slope to the left of the extremum** (i. e. for smaller x -values) in the **original function**, because a positive slope in the first derivative, which will reach the value of 0 at the extremum means that to its left, the first derivative will have a negative value.

By similar logic, we get a **negative slope to the right in the original function**. Thus, the extremum is a local **minimum**. In turn, a negative second derivative at the extremum implies a maximum.

A local extremum at x_E requires a first derivative of 0 at this point. Furthermore, the extremum is a minimum, if the second derivative at the extremum is positive, while it is a maximum if the second derivative has a negative value.

As a matter of fact, the first derivative being equal to 0 and the second one being *not* equal to 0 is a sufficient condition for the existence of a local extremum.

But what if the second derivative is equal to 0?

Let us have a look at two more graphs.

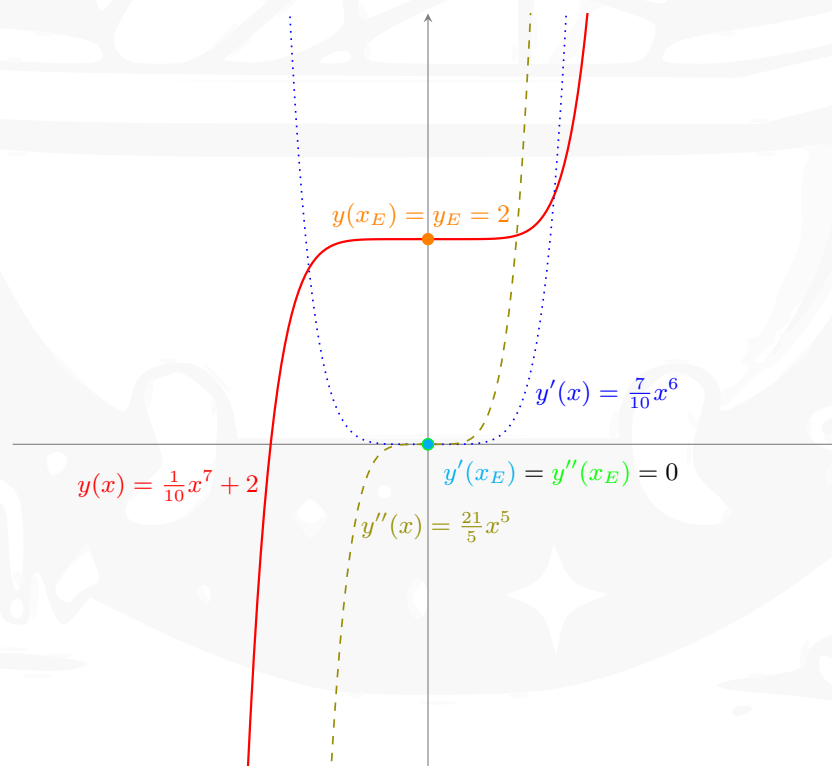


Figure 7: A (shifted and scaled) x^7 -curve with its first two derivatives.

This one is a (squished and shifted) x^7 -function.

At $x_C = 0$, the first and second derivatives are zero. (Notice, we used an index C instead of E , which shall remind us, that unlike before, not every “candidate point” will actually be an extremum.)

Indeed the tangent to the function at $x_C = 0$ has a slope of 0 as is evidenced by the first derivative.

But we can visually see that this is **not** an extremum. It is the case of an increasing function whose (positive) slope gets smaller and smaller until it reaches the value of 0 at the critical point $x_C = 0$ and from there, the slope is increasing again. The extremum candidate point at $(x_C, y(x_C))$ is what is called a *saddle point*.

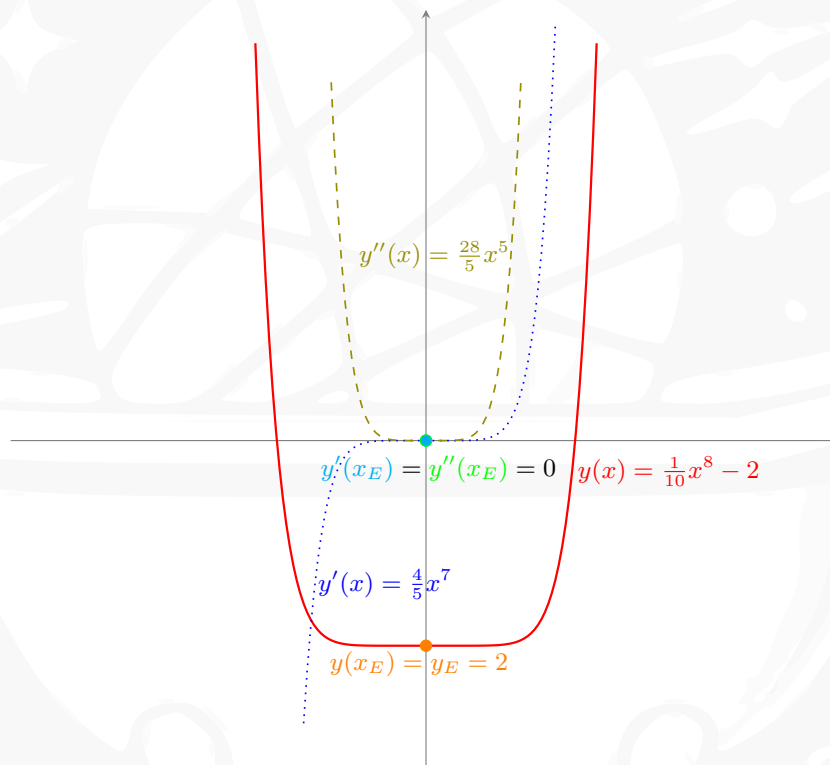


Figure 8: A (shifted and scaled) x^8 -curve with its first two derivatives.

Here, the situation is different.

We see the squished and shifted function x^8 .

Its first and second derivative at $x_C = 0$ are 0 and the tangent to the original function at $x_C = 0$ is a horizontal line, i. e. has a slope of 0. However, in contrast to the last case, here we really have a proper minimum.

But how can we prove this?

As it happens, the first derivative at $x_C = 0$ is 0. (Remember, this is a *required* condition.) Furthermore, the function that describes the first derivative has a saddle point there.

Therefore, the tangent to the first derivative is also a horizontal line, i. e. the second derivative still is 0.

But recall what we've discussed before.

An extremum is characterised by the fact that the function's slope left to the critical point and right to the critical point have opposite signs.

Here, this is the case as we can see with the first derivative. It is negative to the left, i. e. for smaller x than the critical point and is positive to the right of the critical point i. e. for larger x . Thus, in case of a second derivative being 0, we have to check whether the first derivative does a sign change when approaching the point from either side and going through it.

Formally, we need to check, if

$$\operatorname{sgn}(y'(x_0 - \delta)) = -\operatorname{sgn}(y'(x_0 + \delta)),$$

for $\delta \rightarrow 0$, where sgn is the sign-function, which returns 1 for a positive input, -1 for a negative input and 0 for an input of 0.

Thus, we can check, if

$$\lim_{\delta \rightarrow 0} \operatorname{sgn}(y'(x_0 - \delta)) = -\lim_{\delta \rightarrow 0} \operatorname{sgn}(y'(x_0 + \delta))$$

to prove the existence of an extremum, if the second derivative vanishes, as well.

To summarize:

-
- If we want to find a local extremum of a function $y(x)$, we can calculate the **first derivative** $y'(x)$ and find all so-called critical points, where the first derivative is 0. This is a **required condition** for an extremum of the original function $y(x)$.
 - To check, whether a critical point that was found like this really is an extremum of the function, we can first check if the **second derivative** $y''(x)$'s value at the critical point is non-zero. This is **sufficient** for an extremum.
 - If the second derivative has a positive value, the extremum is a local minimum while for a negative second derivative a local maximum was found.

- In the case of the second derivative also being zero, we need to check whether the **sign of the first derivative**'s value changes when going through the critical point. This is also a **sufficient condition** for an extremum, given that the first derivative was 0.

Now, our memories should have been sufficiently refreshed to go beyond one dimension in the next part.

