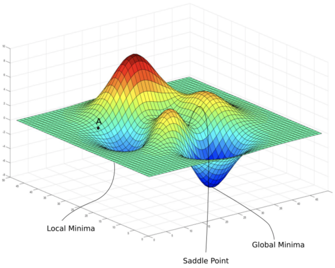
Optimization

Conditions for optimality



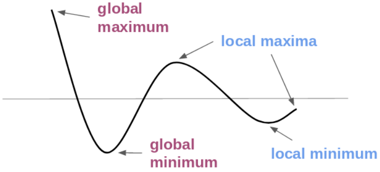
Learning goals

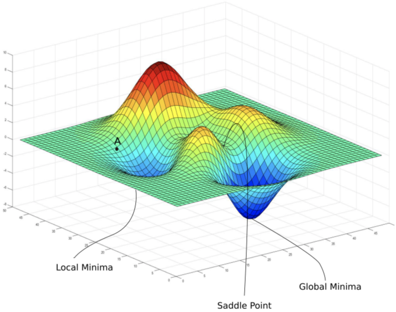
- Optima
- First & Second Order Conditions

DEFINITION LOCAL AND GLOBAL MINIMUM

Given $\mathcal{S} \subseteq \mathbb{R}^d$, $f : \mathcal{S} \to \mathbb{R}$:

- f has a **global minimum** in $\mathbf{x}^* \in \mathcal{S}$, if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{S}$
- f has a **local minimum** in $\mathbf{x}^*$, if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in B_\epsilon(\mathbf{x}^*)$, with $B_\epsilon(\mathbf{x}^*) := \{\mathbf{x} \in \mathcal{S} \mid \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ denoting all points that have a maximum distance of ϵ to $\mathbf{x}^*$ ("ϵ-neighboorhood").





Source (left): `https://en.wikipedia.org/wiki/Maxima_and_minima`.

Source (right): `https://wngaw.github.io/linear-regression/`.

EXISTENCE OF OPTIMA

$$f: \mathcal{S} \subseteq \mathbb{R}^d \to \mathbb{R}$$

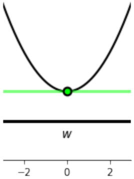
- *f* continous:
  - A real-valued function f, which is defined on a **compact set**, must attain a minimum and a maximum according to the Extreme Value Theorem.
- *f* not continous:
  - In general no statement possible about existence of maximum/minimum.

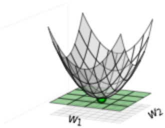
Conditions for Optimality

FIRST ORDER CONDITION

There is one specific property that holds for both the univariate and multivariate case at a (local) optimum: The Taylor series approximation (i.e., the tangent hyperplane) at a local minimum is perfectly flat.

This means that at every local minima the first order derivatives are 0. This condition is also called the **first-order condition**.





Examples for a univariate (left) and multivariate (right) function. Both functions are strictly convex, and have a unique local optimum, which is also the global one. The tangent (hyperplane) is perfectly flat at the optimum.

Source: Watt, 2020, Machine Learning Refined.

FIRST ORDER CONDITION

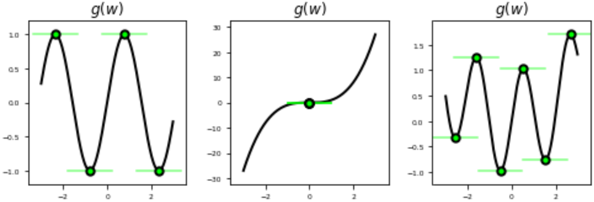
Because at every local minimum $\mathbf{x}^*$ the first order derivative is necessarily always zero, we call this the **first-order** or **necessary** condition.

- **First-order condition (univariate):** Let $\mathbf{x}^* \in \mathbb{R}$ be a local minimum of f. Then:
$$f'(\mathbf{x}^*) = 0$$
- **First-order condition (multivariate):** Let $\mathbf{x}^* \in \mathbb{R}^d$ be a local minimum of f. Then:
$$\nabla f(\mathbf{x}^*) = (0, 0, ..., 0)^\top$$

The points at which the first order derivative is zero are called **stationary points**.

FIRST ORDER CONDITION

However, this is only a **necessary** condition. It is not valid the other way around, i.e. not every point at which the first order derivative is 0 (stationary point) is also a local minimum.

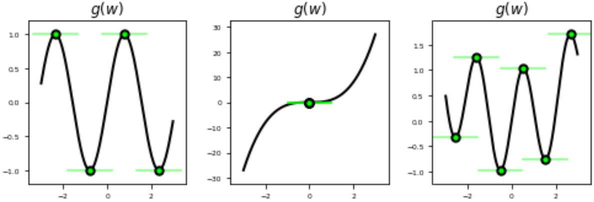


Left: There are for points which fulfill the necessary condition, two of them are in fact local (and also global) minima, and two are local (and also global) maxima (which is not what we are looking for!). Middle: The function has one point that fulfills the necessary condition, but it is neither a local minimum nor a local maximum; such points are called "saddle points". Right: A function that has multiple local minima and maxima.

Source: Watt, 2020, Machine Learning Refined.

SECOND ORDER CONDITION

However, the function's second order derivative helps us identifying whether a stationary point $\mathbf{x}^*$ is a local optimum or not. In cases where the second derivative $f''(\mathbf{x}^*) > 0$ (i.e. the function is convex in the neighborhood of $\mathbf{x}^*$), $\mathbf{x}^*$ is a local minimum.



Left: Around the two minima, the function is convex; around the two maxima the function is concave. Middle: The function is neither convex nor concave around the stationary point. Right: Again, concave behavior around the maxima, and convex behavior around the minima of the function.

Source: Watt, 2020, Machine Learning Refined.

SECOND ORDER CONDITION

We introduce the second order conditions.

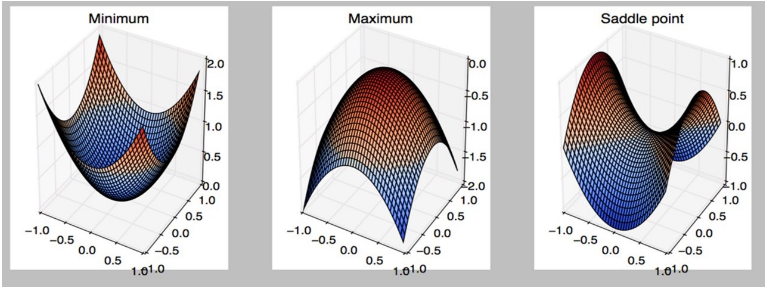
- **Second-order condition (univariate)**: A **stationary** point $x^* \in \mathcal{S} \subseteq \mathbb{R}$ fulfills
$$f''(x^*) > 0.$$
- **Second-order condition (multivariate):** A **stationary** point $\mathbf{x}^* \in \mathcal{S} \subseteq \mathbb{R}^d$ fulfills
$$\nabla^2 f(\mathbf{x}^*) \text{ is positive semi-definitie}$$
(all eigenvalues are positive). This means the curvature is positive in all directions.

If the second-order conditions are fulfilled, then the point is a **local** minimum. The second-order condition is **sufficient** to prove that a point is a local minimum.

SECOND ORDER CONDITION

Example:

1. Local min: If [f convex $\Leftrightarrow$ All eigenvalues positive], then global min
2. Local max: If [f concave $\Leftrightarrow$ All eigenvalues negative], then global max
3. Some eigenvalues positive and some negative $\Leftrightarrow$ saddle point



Let $f : \mathcal{S} \to \mathbb{R}$ be convex on the convex set $\mathcal{S}$. Then the following applies:

- Any local minimum of f is also a global minimum (see chapter *Conditions for Optimality*).
- If f is strictly convex, f has exactly one local minimum on $\mathcal{S}$ and it is also the unique global minimum of f on $\mathcal{S}$ (see chapter *Conditions for Optimality*).
- Sublevel sets $S_1 = \{\mathbf{x} \mid f(\mathbf{x}) < a\}$ and $S_2 = \{\mathbf{x} \mid f(\mathbf{x}) \leq a\}$, $a \in \mathbb{R}$, form convex sets.