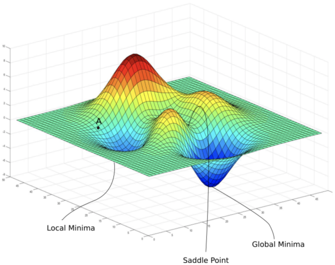
Optimization

Conditions for optimality



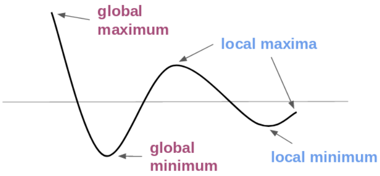
Learning goals

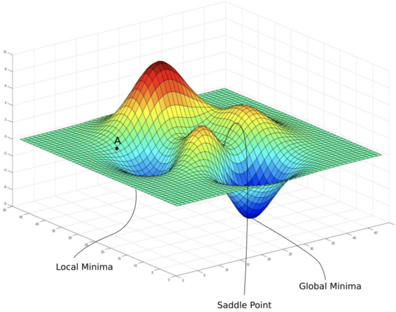
- Local and global
- First & second order conditions

DEFINITION LOCAL AND GLOBAL MINIMUM

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

- f has **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 $B_\epsilon(\mathbf{x}^*) := \{\mathbf{x} \in \mathcal{S} \mid \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ ("ϵ"-ball round $\mathbf{x}^*$).





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

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

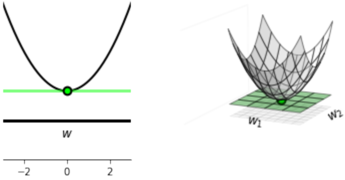
EXISTENCE OF OPTIMA

$$f : \mathcal{S} \to \mathbb{R}$$

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

FIRST ORDER CONDITION FOR OPTIMALITY

Let $f \in \mathcal{C}^1$. **Observation:** At a local minimum 1st order Taylor series approximation is perfectly flat; 1st order derivatives are 0.



(Strictly) convex functions (left: univariate; right: multivariate) with unique local minimum, which is the global one. Tangent (hyperplane) is perfectly flat at the optimum.

Source: Watt, 2020, Machine Learning Refined.

FIRST ORDER CONDITION FOR OPTIMALITY

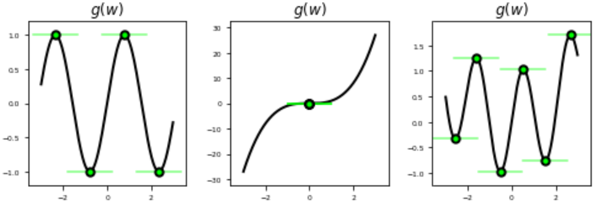
At every local minimum $\mathbf{x}^*$ the first derivative is necessarily always zero; it is therefore called **first-order** or **necessary** condition.

- **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 FOR OPTIMALITY

The condition is **not sufficient**: Not every stationary point ($\nabla f(\mathbf{x}) = 0$) is a local minimum.

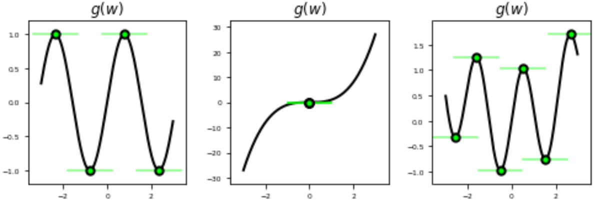


Left: Four points fulfill the necessary conditions; but two of the points are local maxima (not minima). Middle: One point fulfills the necessary condition, but is not a local optimum. Right: Multiple local minima and maxima.

Source: Watt, 2020, Machine Learning Refined.

SECOND ORDER CONDITION FOR OPTIMALITY

Let $f \in \mathcal{C}^2$. A stationary point $\mathbf{x}$ (i.e., $f(\mathbf{x}) = 0$) is a local minimum if $f''(\mathbf{x}) > 0$ (i.e., the function is locally convex).



Left / Right: Function has positive curvature in all directions at the minima, and negative curvature around the maxima. Middle: Curvature is positive in one, and negative in the other direction.

Source: Watt, 2020, Machine Learning Refined.

SECOND ORDER CONDITION FOR OPTIMALITY

Let $f \in \mathcal{C}^2$.

- **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.

Second-order condition is **sufficient** to prove a local minimum.

Note: For a convex function, $\nabla^2 f(\mathbf{x})$ is always p.s.d.; therefore, any stationary point is the local (also global) minimum.

CONDITIONS FOR OPTIMALITY AND CONVEXITY

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

- Any local minimum is also global minimum
- If f strictly convex, f has exactly one local minimum which is also unique global minimum on $\mathcal{S}$