



NTNU

Norwegian University of
Science and Technology

TTK4135 – Lecture 2

Optimality Conditions for Constrained Optimization: KKT Conditions

Lecturer: Lars Imsland

Purpose of Lecture

- Optimization problems and Convexity
- Motivating Examples for KKT Conditions
- KKT Conditions

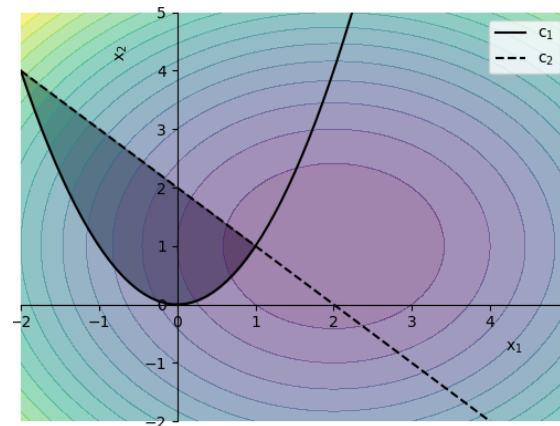
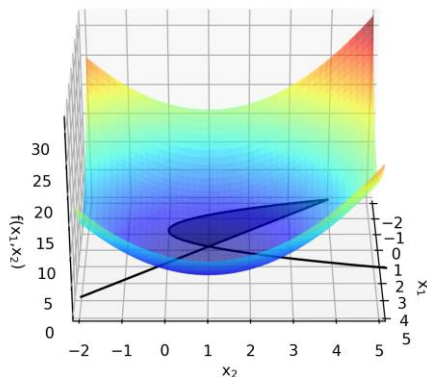
Reference: Chapter 12.1, 12.2 in N&W

General Optimization Problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{aligned} c_i(x) &= 0, & i \in \mathcal{E}, \\ c_i(x) &\geq 0, & i \in \mathcal{I}. \end{aligned}$$

- Example:

$$\min (x_1 - 2)^2 + (x_2 - 1)^2 \quad \text{subject to} \quad \begin{aligned} x_1^2 - x_2 &\leq 0, \\ x_1 + x_2 &\leq 2. \end{aligned}$$



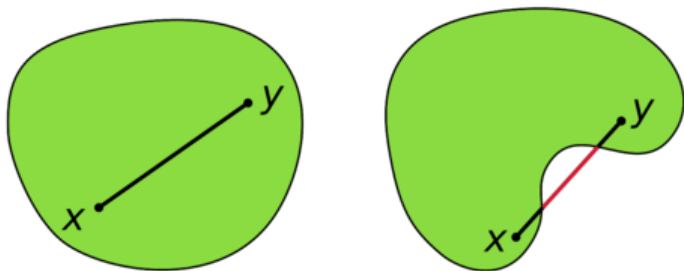
- What if we add equality-constraint $x_1 = 0$?

Definitions: Feasible Set and Solutions

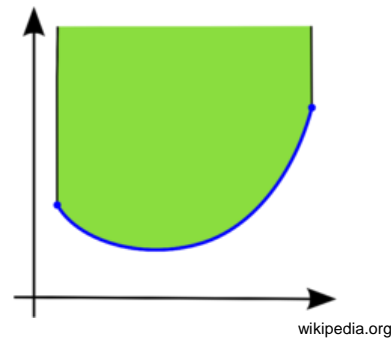
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases} \quad (\text{P})$$

- Feasible set: $\Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}$
- A vector x^* is a *global solution* to (P) if $x^* \in \Omega$ and $f(x) \geq f(x^*)$ for $x \in \Omega$.
- A vector x^* is a *local solution* to (P) if $x^* \in \Omega$ and there is a neighborhood \mathcal{N} of x^* such that $f(x) \geq f(x^*)$ for $x \in \mathcal{N} \cap \Omega$.
- A vector x^* is a *strict local solution* to (P) if $x^* \in \Omega$ and there is a neighborhood \mathcal{N} of x^* such that $f(x) > f(x^*)$ for $x \in \mathcal{N} \cap \Omega$ with $x \neq x^*$.

Convexity: An important property



If the line segment between any two points within a **set** is inside the set, the set is **convex**.



A **function** is **convex** if the epigraph is a convex set.

- A convex optimization problem: Both $f(x)$ and the feasible set convex
- Convex optimization problems are preferable!
 - For convex optimization problems, **every local minimum is also a global minimum**. **Sufficient to search for a local minimum!** Which is much easier than searching for global minimum.
 - For many convex optimization problems, it is easy to find derivatives, exploit structure, etc. making them efficient to solve.
 - They typically have “guaranteed complexity”.

Convexity: Conditions

- When is an optimization problem convex?

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}$$

- Conditions for a convex optimization problem:

- $f(x)$ is a convex function:

$$\forall x, y \in \Omega, \forall \alpha \in [0, 1] : \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

- The feasible set $\Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0, i \in \mathcal{E}, c_i(x) \geq 0, i \in \mathcal{I}\}$ is convex:

$$\forall x, y \in \Omega, \forall \alpha \in [0, 1] : \quad \alpha x + (1 - \alpha)y \in \Omega$$

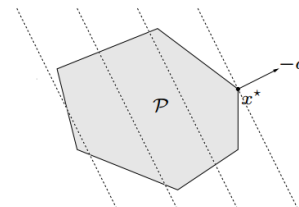
- When is the feasible set convex?
 - $c_i(x), i \in \mathcal{E}$ are linear
 - $c_i(x), i \in \mathcal{I}$ are concave

Convex problems: Any local solution is global

Types of Constrained Optimization Problems

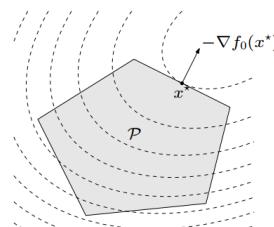
- Linear programming
 - Convex problem
 - Feasible set polyhedron

$$\begin{aligned} \min \quad & c^\top x \\ \text{subject to} \quad & Ax \leq b \\ & Cx = d \end{aligned}$$



- Quadratic programming
 - Convex problem if $P \geq 0$
 - Feasible set polyhedron

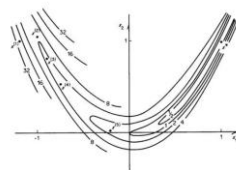
$$\begin{aligned} \min \quad & \frac{1}{2}x^\top Px + q^\top x \\ \text{subject to} \quad & Ax \leq b \\ & Cx = d \end{aligned}$$



- Nonlinear programming
 - In general non-convex!

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & g(x) = 0 \\ & h(x) \geq 0 \end{aligned}$$

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



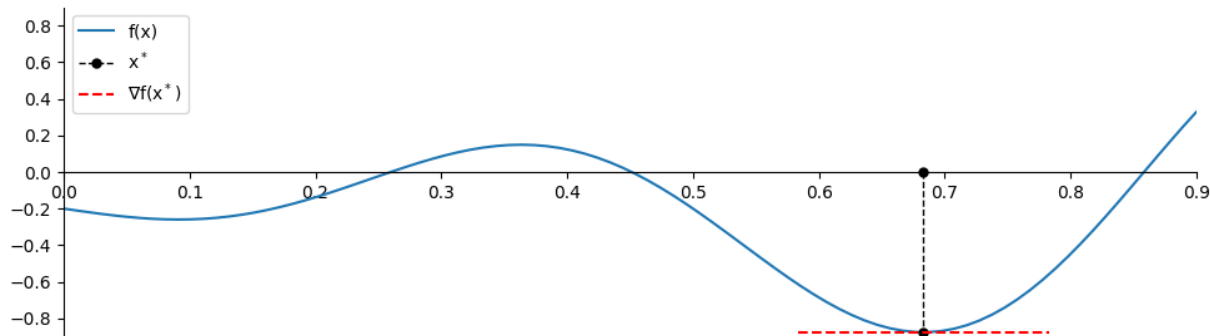
$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & c_i(x) = 0, \quad i \in \mathcal{E}, \\ & c_i(x) \geq 0, \quad i \in \mathcal{I}. \end{aligned}$$

Necessary Conditions for Unconstrained Optimization

$$\min_{x \in \mathbb{R}^n} f(x)$$

Theorem 2.2 (First-Order Necessary Conditions).

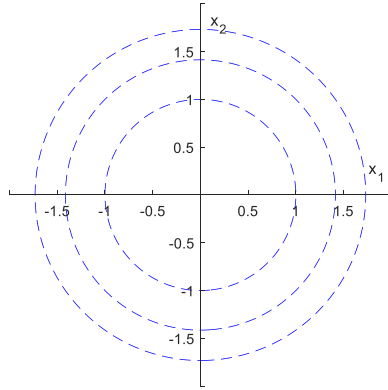
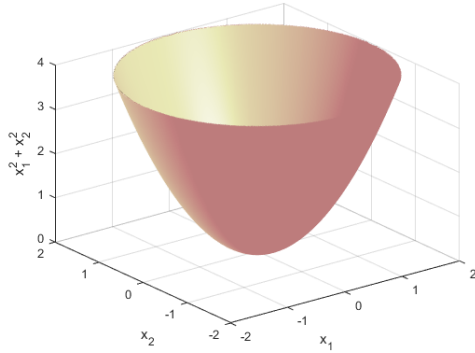
If x^ is a local minimizer and f is continuously differentiable in an open neighborhood of x^* , then $\nabla f(x^*) = 0$.*



- What about constrained problems?

Contours/level curves, gradients and directions

$$f(x_1, x_2) = x_1^2 + x_2^2$$



Contours/level curves, gradients and directions

KKT Conditions

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{aligned} c_i(x) &= 0, & i \in \mathcal{E}, \\ c_i(x) &\geq 0, & i \in \mathcal{I}. \end{aligned}$$

- Lagrangian

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

- Note: One Lagrangian multiplier λ_i for each constraint
- Necessary conditions for x^* to be a solution (under some mild regularity conditions):

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\ c_i(x^*) &= 0, & \forall i \in \mathcal{E}, \\ c_i(x^*) &\geq 0, & \forall i \in \mathcal{I}, \\ \lambda_i^* &\geq 0, & \forall i \in \mathcal{I}, \\ \lambda_i^* c_i(x^*) &= 0, & \forall i \in \mathcal{E} \cup \mathcal{I}. \end{aligned}$$

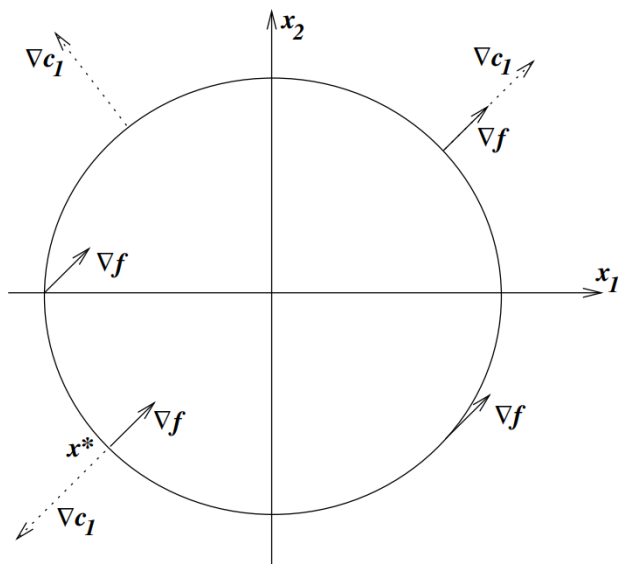
- These are called the *KKT conditions*

We will not prove KKT, but study 3 motivating cases (Ex. 12.1-12.3 in N&W)

Looking for points where there are no descent directions...
...as these are potential local solutions

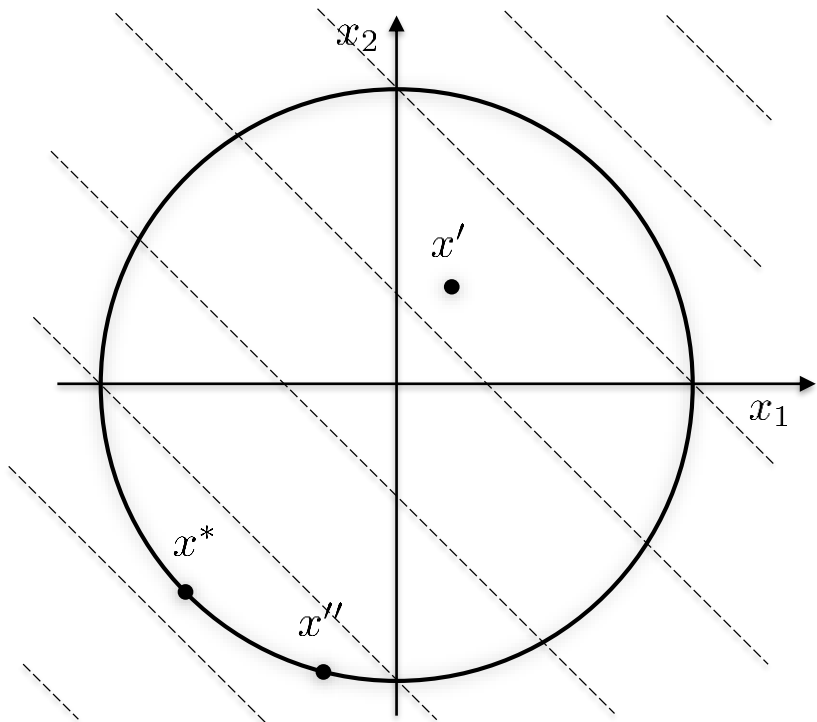
Case I: Equality constraint (Example 12.1)

$$\min x_1 + x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0$$



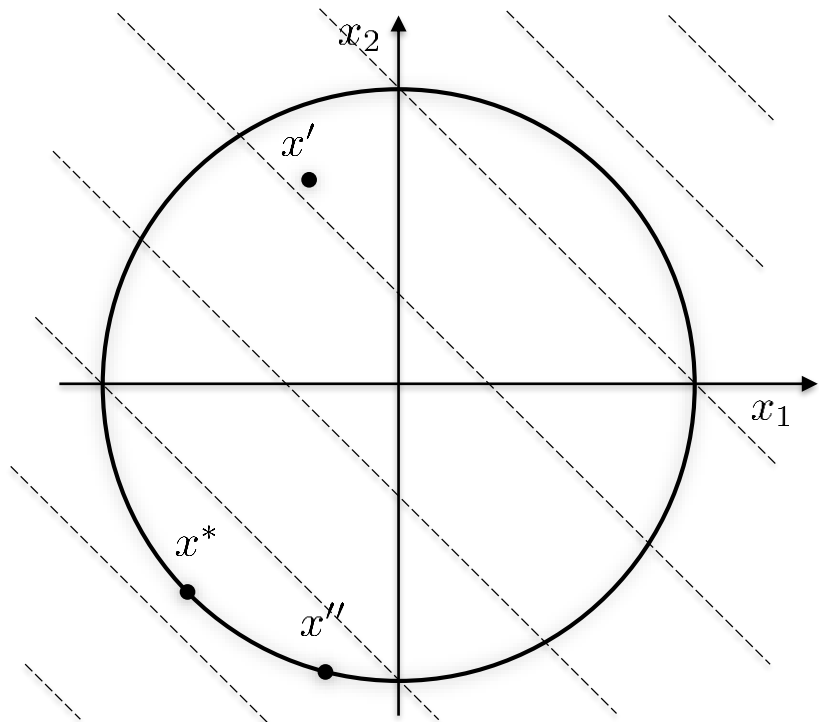
Case II: Inequality constraint (Example 12.2)

$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0$$



Case II: Inequality constraint (Example 12.2)

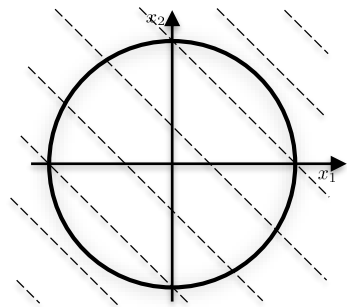
$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0$$



Active Set

The active set $\mathcal{A}(x)$ at any feasible point x consists of the equality constraint indices from \mathcal{E} together with the indices of the inequality constraints i for which $c_i(x) = 0$. That is,

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(x) = 0\}$$



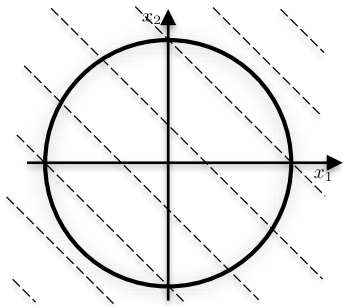
$$\min x_1 + x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0$$

$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0$$

Set of Feasible Directions

Given a feasible point x and the active constraint set $\mathcal{A}(x)$, the set of linearized feasible directions $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d \mid \begin{array}{ll} d^\top \nabla c_i(x) = 0, & \text{for all } i \in \mathcal{E}, \\ d^\top \nabla c_i(x) \geq 0, & \text{for all } i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}$$

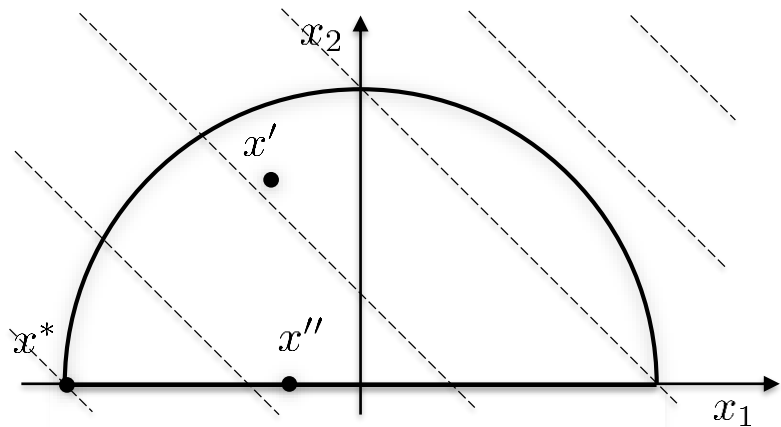


$$\min x_1 + x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 2 = 0$$

$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0$$

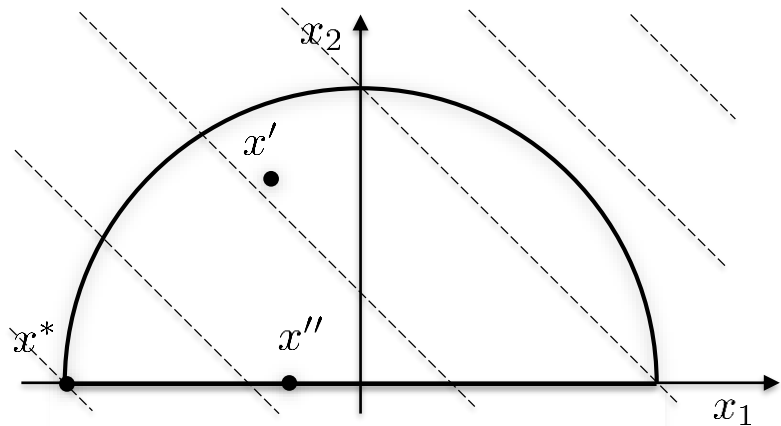
Case III: Two inequality constraints (Example 12.3)

$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0, \quad x_2 \geq 0$$

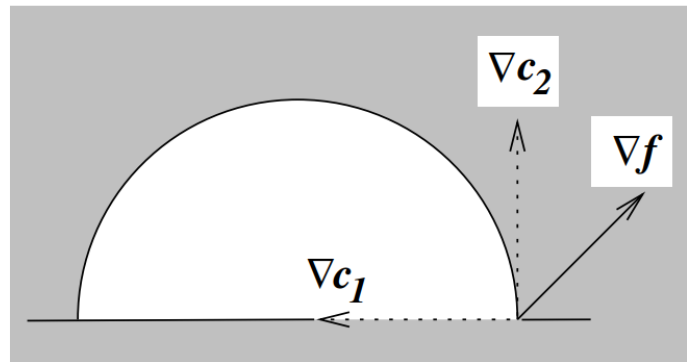
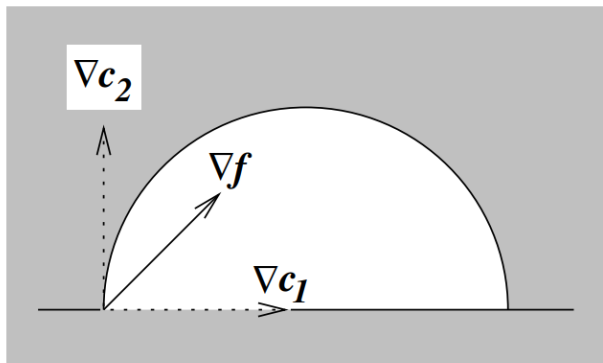


Case III: Two inequality constraints (Example 12.3)

$$\min x_1 + x_2 \quad \text{s.t.} \quad 2 - x_1^2 - x_2^2 \geq 0, \quad x_2 \geq 0$$



Case III: Two inequality constraints (Example 12.3)



$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E}, \\ c_i(x) \geq 0, & i \in \mathcal{I}, \end{cases} \quad (12.1)$$

Theorem 12.1 (First-Order Necessary Conditions).

Suppose that x^ is a local solution of (12.1), that the functions f and c_i in (12.1) are continuously differentiable, and that the LICQ holds at x^* . Then there is a Lagrange multiplier vector λ^* , with components λ_i^* , $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)*

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \quad (12.34a)$$

$$c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E}, \quad (12.34b)$$

$$c_i(x^*) \geq 0, \quad \text{for all } i \in \mathcal{I}, \quad (12.34c)$$

$$\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I}, \quad (12.34d)$$

$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}. \quad (12.34e)$$

Linear Independence Constraint Qualification (LICQ)

Given the point x and the active set $\mathcal{A}(x)$, we say that the linear independence constraint qualification (LICQ) holds if the set of active constraint gradients $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ is linearly independent.

Why are KKT-conditions so important?

- KKT conditions can be used to solve nonlinear programming problems, but only for *very simple* problems
- But: **Most algorithms for constrained optimization look for candidate solutions that fulfill KKT conditions**
 - These are iterative algorithms that stop when KKT conditions fulfilled
- And also:
 - When faced with an optimization problem that you don't know how to handle, write down the optimality conditions
 - Often you can learn about a problem by examining the properties of its optimal solutions
- And finally:
 - The Lagrange multipliers tell you the 'hidden cost' of constraints

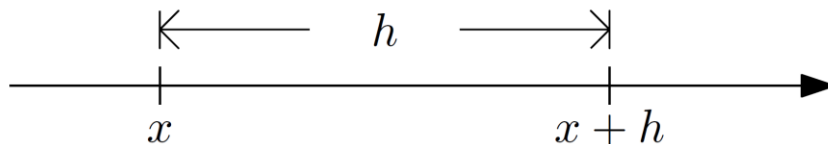
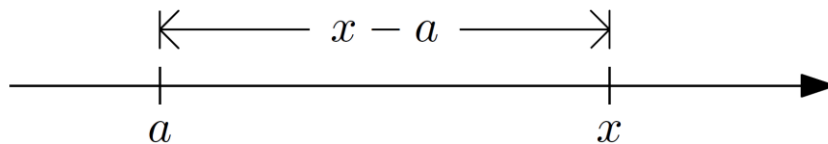
Taylor Expansions

- From calculus?

$$f(x) = f(a) + (x - a)f'(a) + \frac{(x - a)^2}{2}f''(a) + O(\|x - a\|^3)$$

- This course:

$$f(x + h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + O(\|h\|^3)$$



$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}$$

Lagrangian:
$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

KKT-conditions (First-order necessary conditions):
 If x^* is a local solution and LICQ holds, then there exist λ^* such that

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\ c_i(x^*) &= 0, \quad \forall i \in \mathcal{E}, \\ c_i(x^*) &\geq 0, \quad \forall i \in \mathcal{I}, \\ \lambda_i^* &\geq 0, \quad \forall i \in \mathcal{I}, \\ \lambda_i^* c_i(x^*) &= 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}. \end{aligned}$$

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, \\
c_i(x^*) &= 0, \quad \forall i \in \mathcal{E}, \\
c_i(x^*) &\geq 0, \quad \forall i \in \mathcal{I}, \\
\lambda_i^* &\geq 0, \quad \forall i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) &= 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}.
\end{aligned}$$