NTNU | Norwegian University of Science and Technology

TTK4135 – Lecture 2
Optimality Conditions for Constrained Optimization: KKT Conditions

Lecturer: Lars Imsland

Purpose of Lecture

- Recap: Optimization problems and Convexity

- Necessary conditions for constrained optimization:
  - KKT conditions
  - Motivating examples

Reference: Chapter 12.1, 12.2 in N&W

General Optimization Problem

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

$\mathcal{E} = \emptyset$

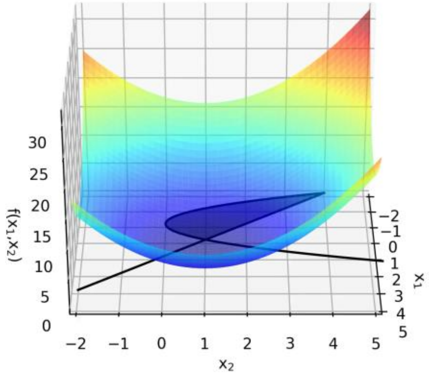
$\mathcal{I} = \{1, 2\}$

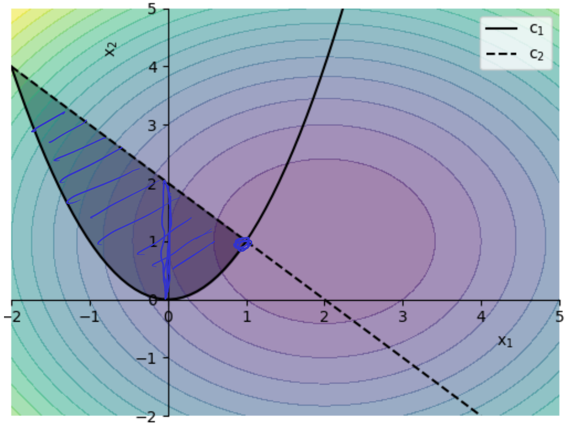
$- c_1(x)$

- Example:

$$\min_{x \in \mathbb{R}^2} (x_1 - 2)^2 + (x_2 - 1)^2 \quad \text{subject to} \quad \begin{array}{l} x_1^2 - x_2 \leq 0, \\ x_1 + x_2 \leq 2. \end{array}$$

$c_2(x) = 2 - x_1 - x_2$





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

$c_3 = x_1$

$\mathcal{E} = \{3\}$

Definitions: Feasible Set and Solutions

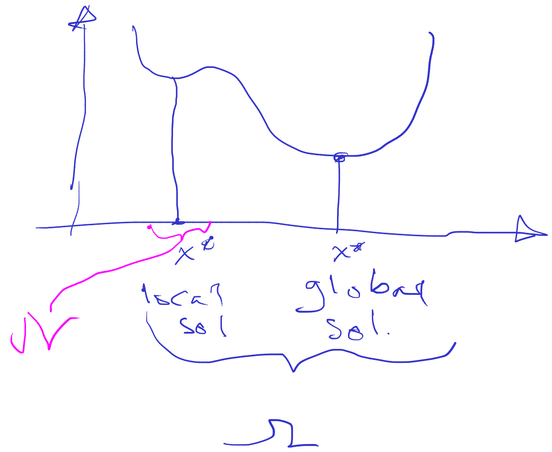
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases} \qquad \text{(P)}$$

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

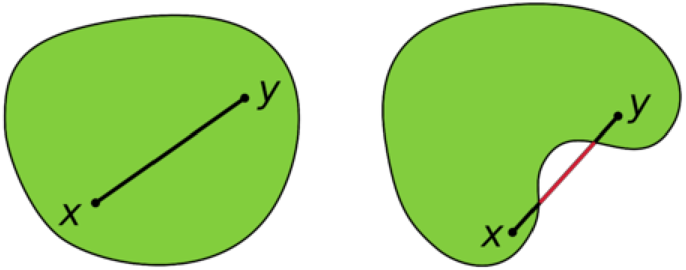
- A vector x^* is a *global solution* to (P) if $x^* \in \Omega$ and $f(x) \geq f(x^*)$ for $x \in \Omega$.

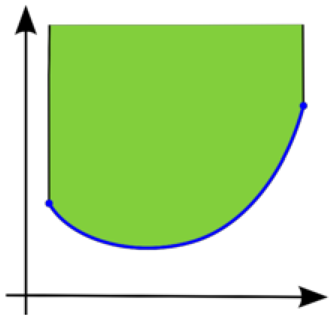
- A vector x^* is a *local solution* to (P) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x) \geq f(x^*)$ for $x \in \mathcal{N} \cap \Omega$.

- 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 $\Omega = \{x \in \mathbb{R}^n | c_i(x) = 0,\ i \in \mathcal{E},\ c_i(x) \geq 0,\ i \in \mathcal{I}\}$ 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

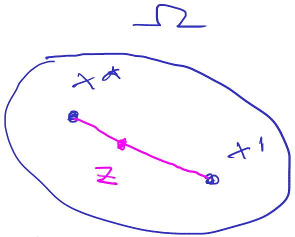
Convex problems: Any local solution is global

Proof by contradiction:

Assume convex opt. prob. and x^* local solution

That is: There exist $x' \in \Omega$, $x' \neq x^*$ s.t. $f(x') < f(x^*)$

Define $z = \alpha x^* + (1-\alpha) x'$, $\alpha \in [0,1]$



$$
\begin{aligned}
f(z) &= f(\alpha x^* + (1-\alpha) x') \\
&\le \alpha f(x^*) + (1-\alpha) f(x') \\
&= f(x^*) - (1-\alpha) f(x^*) + (1-\alpha) f(x') \\
&= f(x^*) + (1-\alpha) \underbrace{\left(f(x') - f(x^*) \right)}_{<0} \\
&\le f(x^*)
\end{aligned}
$$

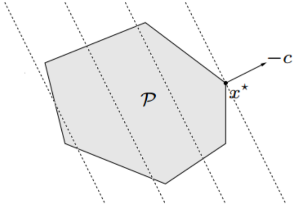
$\Rightarrow$ x^* cannot be local sol.

Contradiction!

Types of Constrained Optimization Problems

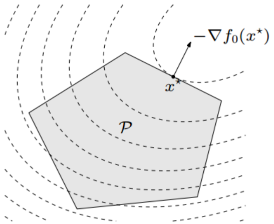
- Linear programming
  - Convex problem
  - Feasible set polyhedron

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



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

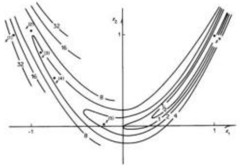
$$\begin{array}{ll} \min & \frac{1}{2} x^\top P x + q^\top x \\ \text{subject to} & Ax \leq b \\ & Cx = d \end{array}$$



- Nonlinear programming
  - In general non-convex!

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

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



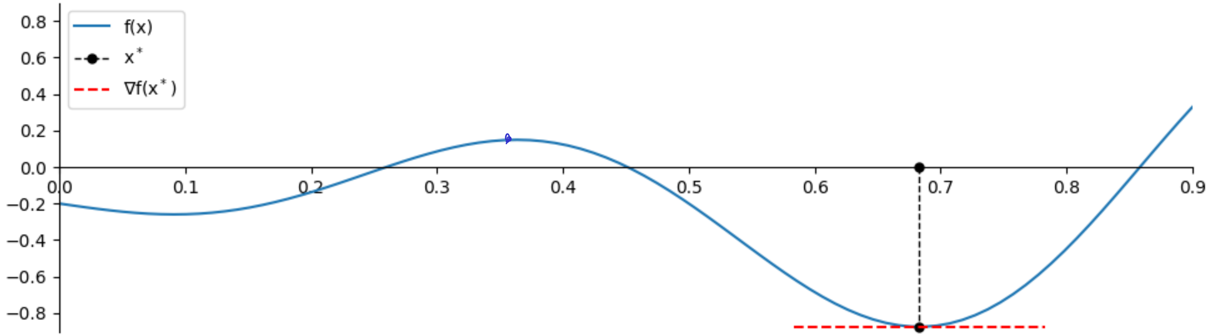
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}$$

Necessary Conditions for Unconstrained Optimization

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

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 $\nabla f(x^*) = 0$.

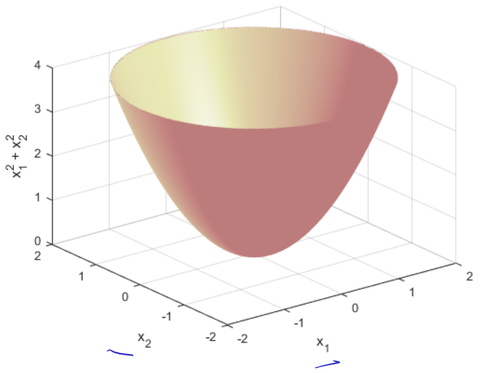


When are there no descent/downhill directions?

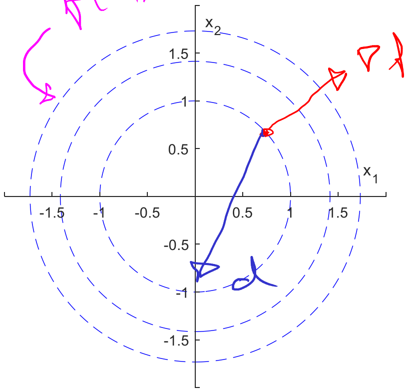
- What about constrained problems?

Contours/level curves, gradients and directions

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



$f(x_1, x_2) = k, \; k = 1, 2, 3, \ldots$



$$x' = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

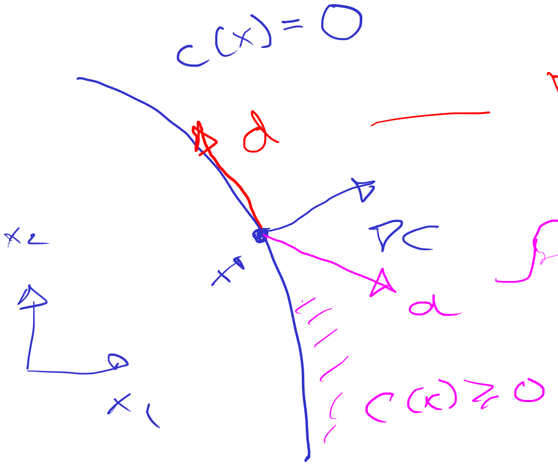
$$\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$$\nabla f(x') = \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}$$

Given direction d: If $\nabla f^T d < 0$, then

d is a descent direction (for $f(x)$)

Contours/level curves, gradients and directions



$\nabla c^T d = 0$: d feasible direction for $c(x)=0$

$\nabla c^T d \geq 0$: d feasible direction for $c(x)\geq 0$

Observation: In a local solution, there cannot be feasible descent direction

Necessary conditions for optimality
KKT Conditions

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

- Lagrangian

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

- Note: One Lagrangian multiplier λ_i for each constraint

- Necessary conditions for x^* to be a solution (under some mild regularity conditions):

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

- These are called the *KKT conditions*

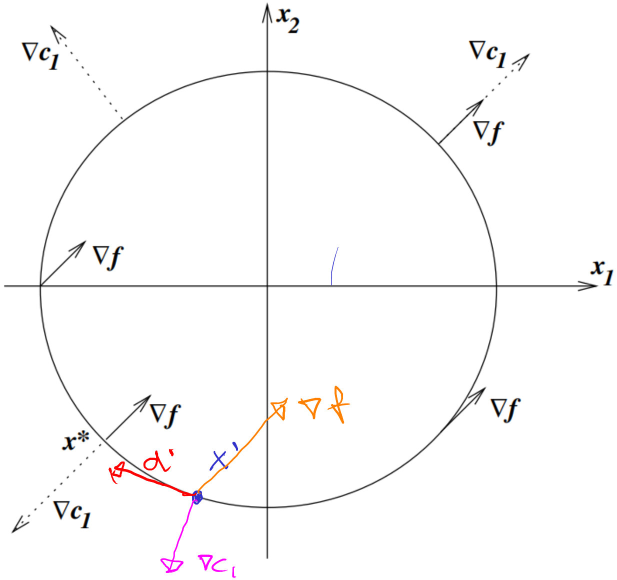
We will not prove KKT, but study 3 motivating cases (Ex. 12.1-12.3 in N&W)

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

Case I: Equality constraint (Example 12.1)

$$\min \underbrace{x_1 + x_2}_{f(x)} \qquad \text{s.t.} \qquad \underbrace{x_1^2 + x_2^2 - 2 = 0}_{c_1(x)}$$

$$\nabla f = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \nabla c_1 = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$



x': Exist feasible descent direction

$d'^T \nabla c_1(x') = 0$ and $d'^T \nabla f(x') < 0$

x^*: No feasible descent direction.

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$$

$$\mathcal{L}(x, \lambda) = f(x) - \lambda_1 c_1(x)$$

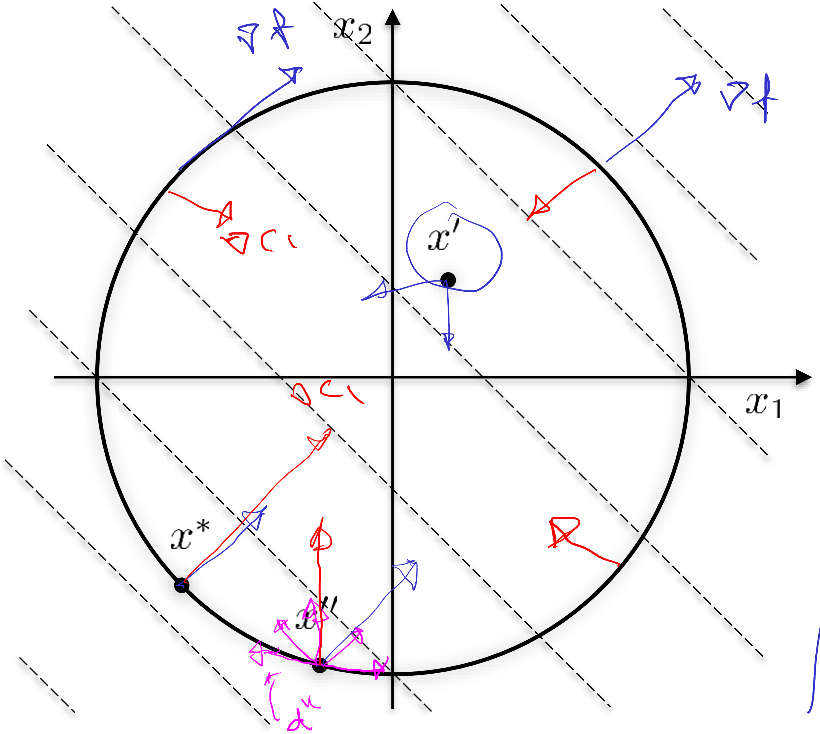
KKT: $\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \lambda_1^* \nabla c_1(x^*) = 0$

$$c_1(x^*) = 0$$

Case II: Inequality constraint (Example 12.2)

$\min\ x_1 + x_2 \qquad \text{s.t.} \qquad \underbrace{2 - x_1^2 - x_2^2}_{c_1(x)} \geq 0$

$\nabla c_1(x) = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}$



x': exist feasible descent directions $d'^T \nabla f(x') < 0$

x'': exist feasible descent dir.

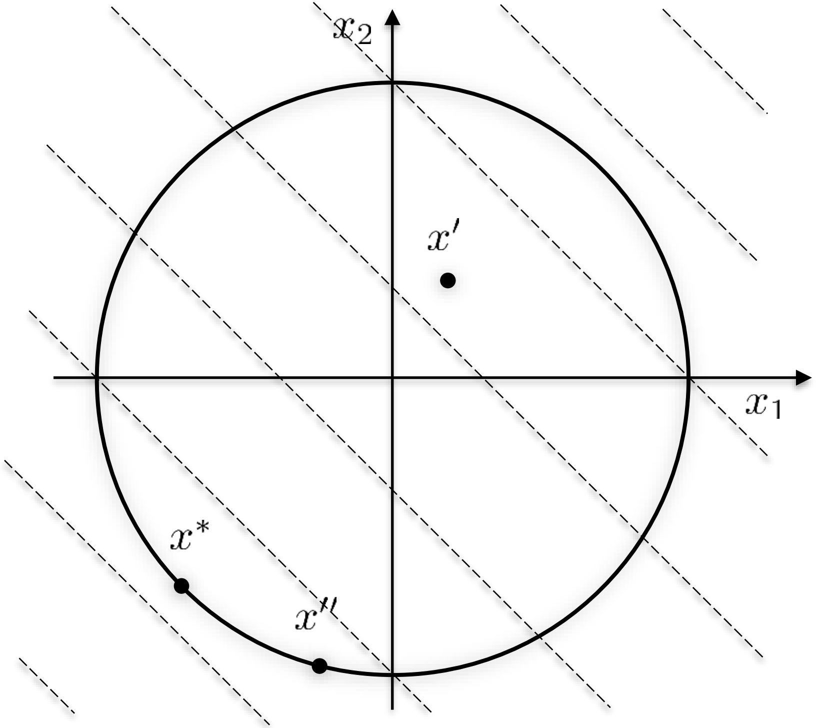
$d''^T \nabla f \leq 0$ and $d''^T \nabla c_1(x) \geq 0$

x^*: No feasible descent direction

$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*),\ \lambda_1^* > 0$

Case II: Inequality constraint (Example 12.2)

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



$\mathcal{L}(x,\lambda) = f(x) - \lambda_1 c(x)$

KKT: $\nabla f(x^*) - \lambda \nabla c_1(x^*) = 0$

$c_1(x^*) \geq 0$

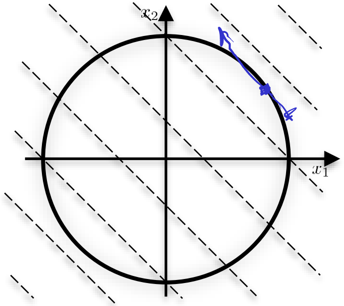
$\lambda_1^* > 0$

$\lambda_1^* \cdot \underbrace{c_1(x^*)}_{=0} = 0$

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,

$$\mathcal{A}(x) = \mathcal{E} \cup \left\{ i \in \mathcal{I} \middle| c_i(x) = 0 \right\}$$



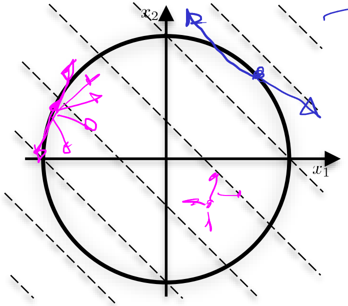
$$\min\ x_1 + x_2 \qquad \text{s.t.} \qquad x_1^2 + x_2^2 - 2 = 0$$

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

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

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

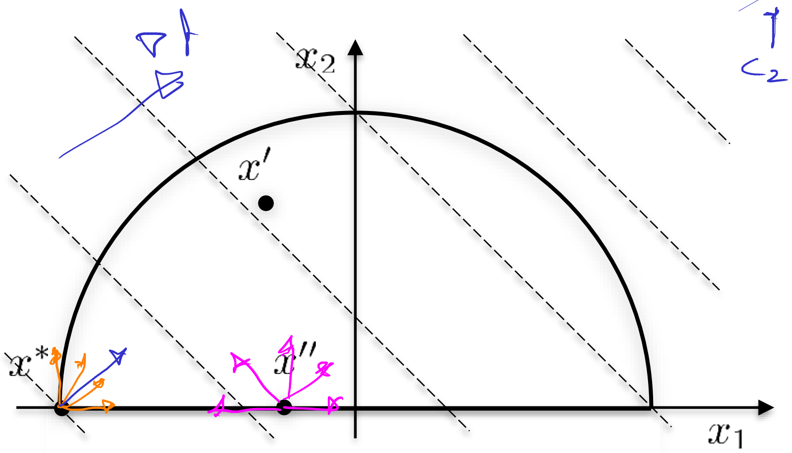


$$\min\ x_1 + x_2 \qquad \text{s.t.} \qquad x_1^2 + x_2^2 - 2 = 0$$

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

Case III: Two inequality constraints (Example 12.3)

$$\min \underbrace{x_1 + x_2}_{f(x)} \qquad \text{s.t.} \qquad \underbrace{2 - x_1^2 - x_2^2 \geq 0}_{c_1}, \quad \underbrace{x_2 \geq 0}_{c_2}$$



x': $\mathcal{A}(x') = \emptyset$

$\mathcal{F}(x') = \mathbb{R}^2$

x'': $\mathcal{A}(x'') = \{2\}$

$\mathcal{F}(x'') = \{d \mid d^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} \geq 0\}$

x^*: $\mathcal{A}(x^*) = \{1, 2\}$

$\mathcal{F}(x^*) = \{d \mid d^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} \geq 0,$
$d^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} \geq 0\}$

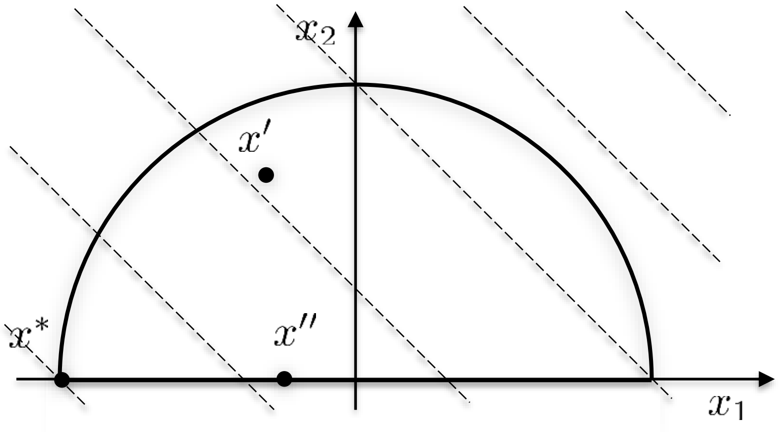
$\nabla f(x^*) \in \mathcal{F}(x^*)$

$\Rightarrow \nabla f(x^*) = \lambda_1^* \nabla c_1(x^*) + \lambda_2^* \nabla c_2(x^*)$

$\lambda_1^* \geq 0, \lambda_2^* \geq 0$

Case III: Two inequality constraints (Example 12.3)

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



$$\mathcal{L}(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x)$$

KKT:

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \lambda_1^* \nabla c_1(x^*) - \lambda_2^* \nabla c_2(x^*) = 0$$

$$c_1(x^*) \geq 0$$

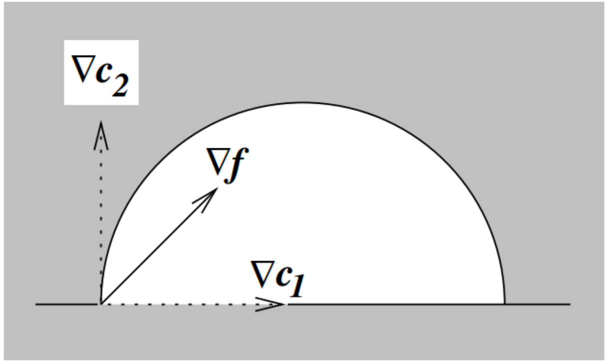
$$c_2(x^*) \geq 0$$

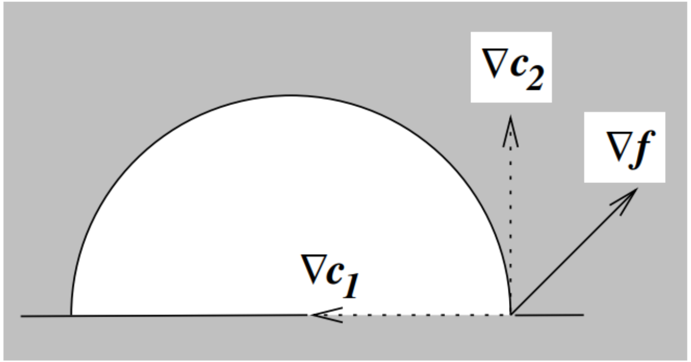
$$\lambda_1^* \geq 0$$

$$\lambda_2^* \geq 0$$

$$\lambda_1^* c_1(x^*) = 0 \qquad \lambda_2^* c_2(x^*) = 0$$

Case III: Two inequality constraints (Example 12.3)





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

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

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^*)

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

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

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

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

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

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 search 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