NTNU | Norwegian University of Science and Technology

TTK4135 – Lecture 3
Optimality Conditions for Constrained Optimization (KKT & 2nd order)

Lecturer: Lars Imsland

Purpose of Lecture

- Repetition of definitions:
  - Gradient, Hessian
  - Feasible Set
  - Local vs Global Optima

- Conditions for optimality
  - **KKT conditions** (1st order, necessary conditions)
    - Example
    - Constraint qualifications
  - 2nd order conditions (necessary and sufficient)

- Reference: Chapter 12.3, 12.5 (12.8, 12.9) in N&W

Administrative

- We need more members in the reference group
  - Send me an e-mail: lars.imsland@ntnu.no

- The first Matlab assessment is now active
  - Do not be intimidated by the amount of text – the task is probably simpler than you think
  - You have unlimited attempts
  - You can discuss the problem and get help from your classmates, but everyone must complete on their own
  - It is not obligatory, but will count towards 20% part grade
    - Finish all 6: A, finish 5: B, finish 4: C, finish 3: D, finish 1 or 2: E

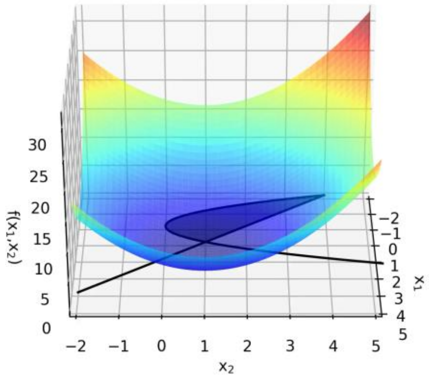
  - Schedule: New problem ca. every 2^{nd} week, deadline 3 weeks after

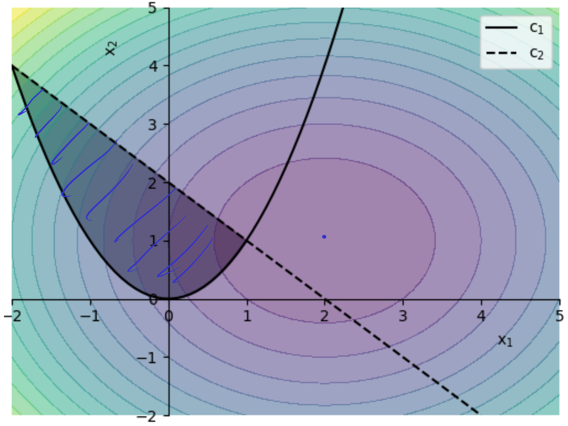
General Optimization Problem

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

- Example:

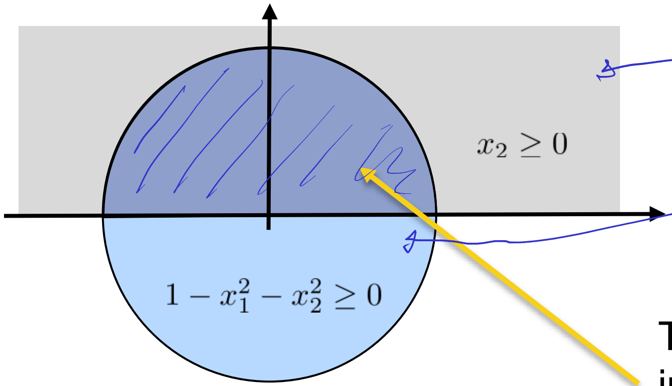
$$\min\ (x_1 - 2)^2 + (x_2 - 1)^2 \quad \text{subject to} \quad \begin{aligned} x_1^2 - x_2 &\leq 0, \\ x_1 + x_2 &\leq 2. \end{aligned}$$





Feasible Set

Feasible set: Collection of all points that satisfy all constraints:



$c_1(x) = x_2 \geq 0$

$c_2(x) = 1 - x_1^2 - x_2^2 \geq 0$

The feasible set is the intersection of the grey and blue area

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

Gradient and Hessian

- The *gradient* (or first derivative) of a function $f(x)$ of several variables is defined as

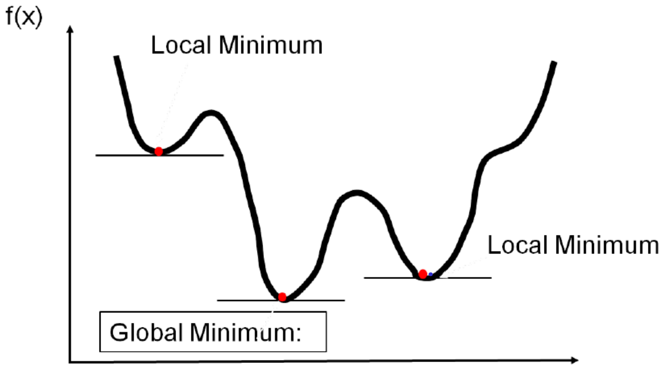
$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T$$

- The matrix of second partial derivatives of $f(x)$is known as the *Hessian*, and is defined as

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- We will frequently use $\nabla^2_{xx} \mathcal{L}(x^*, \lambda^*)$, the *Hessian of the Lagrangian*

Local and Global Optima



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

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

A point 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 $x \in \mathcal{N} \cap \Omega$.

Convex optimization problems: local solutions are global.

Unconstrained Optimality Conditions

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

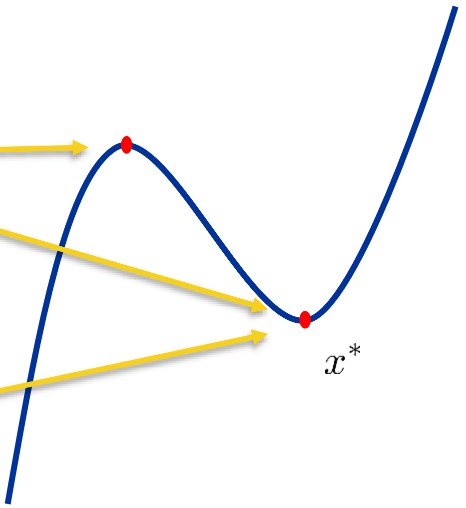
We want to test a point x^* for local optimality:

Necessary condition:

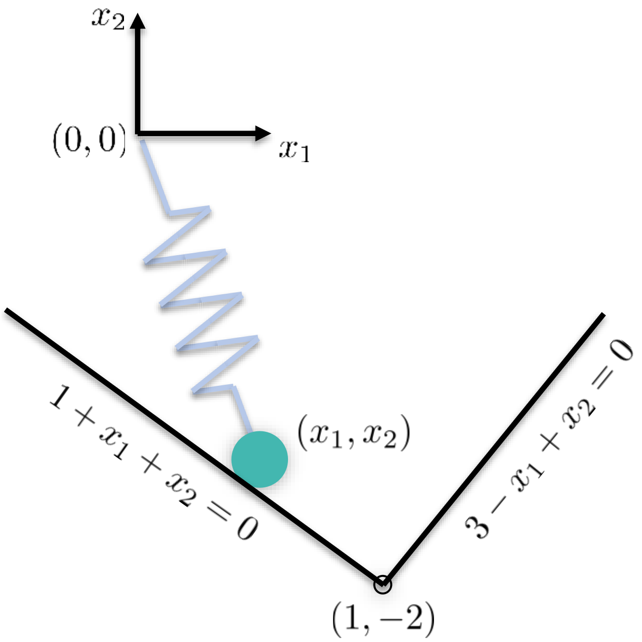
$\nabla f(x^*) = 0$ (stationarity)

Sufficient condition:

x^* stationary and $\nabla^2 f(x^*) > 0$



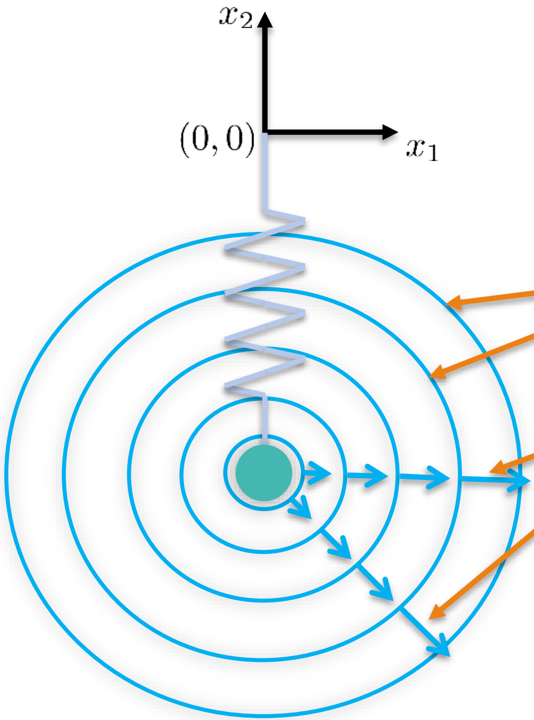
Simple example: Ball and Spring



To find position at rest, minimize potential energy!

$$
\begin{aligned}
\min_{x \in \mathbb{R}^2} \quad & \underbrace{x_1^2 + x_2^2}_{\text{spring}} + \underbrace{m x_2}_{\text{gravity}} \\
\text{subject to} \quad & c_1(x) = 1 + x_1 + x_2 \geq 0 \\
& c_2(x) = 3 - x_1 + x_2 \geq 0
\end{aligned}
$$

Ball and Spring: No Constraints



$$\min_{x \in \mathbb{R}^2} \quad x_1^2 + x_2^2 + m x_2$$

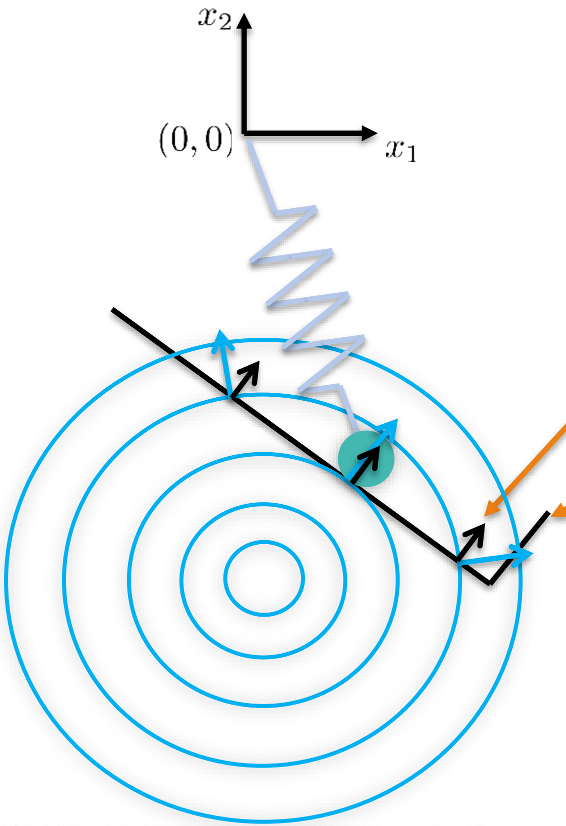
Contour lines of $f(x)$

Gradient of $f(x)$ $\quad \nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 + m \end{pmatrix}$

Unconstrained minimum:

$$\nabla f(x^*) = 0 \Rightarrow \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{m}{2} \end{pmatrix}$$

Ball and Spring: With one (active) constraint



$$\min_{x \in \mathbb{R}^2} \quad x_1^2 + x_2^2 + m x_2$$
$$\text{subject to} \quad c_1(x) = 1 + x_1 + x_2 \geq 0$$
$$c_2(x) = 3 - x_1 + x_2 \geq 0$$

Gradient $\nabla c_1(x)$ of active constraint

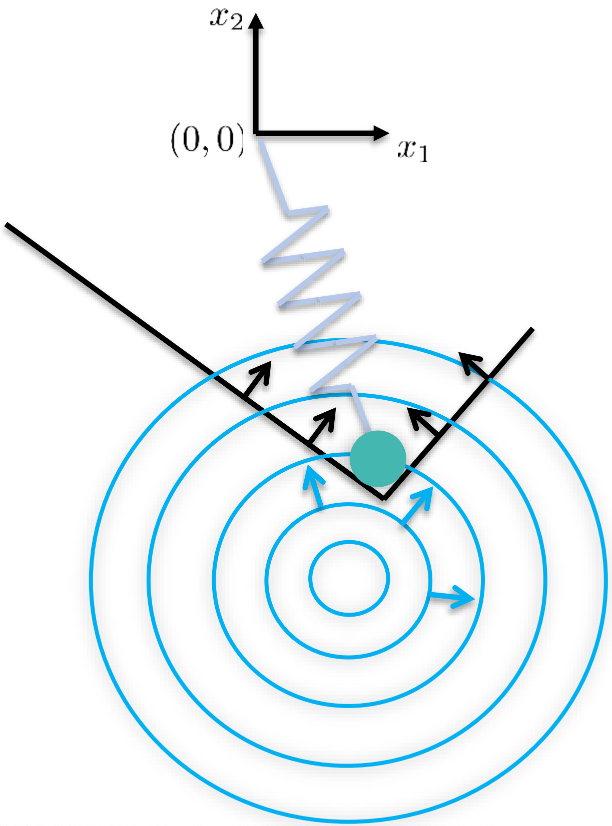
Inactive constraint $c_2(x)$

Constrained minimum:

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

Lagrange multiplier

Ball and Spring: With two active constraints



$$\min_{x \in \mathbb{R}^2} \quad x_1^2 + x_2^2 + m x_2$$

$$\text{subject to} \quad c_1(x) = 1 + x_1 + x_2 \geq 0$$

$$c_2(x) = 3 - x_1 + x_2 \geq 0$$

Constrained minimum at "equilibrium of forces":

$$\nabla f(x^*) = \lambda_1 \nabla c_1(x^*) + \lambda_2 \nabla c_2(x^*), \quad \lambda_1,\ \lambda_2 \geq 0$$

"Constraint forces"

The Lagrangian

For constrained optimization problems, introduce modification of objective function:

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

- Multipliers for *equality* constrains may have both signs in a solution
- Multipliers for *inequality* constraints cannot be negative (cf. shadow prices)
- For (inequality) constraints that are *inactive*, multipliers are zero

KKT Conditions (Theorem 12.1)

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

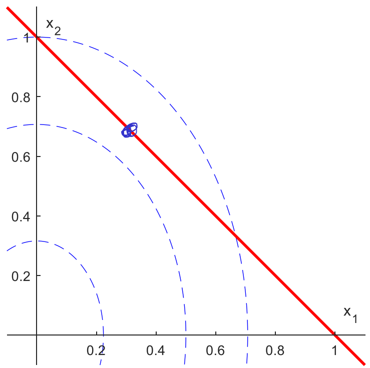
$$
\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, && && \text{(stationarity)} \\
c_i(x^*) &= 0, && \forall i \in \mathcal{E}, && \\
c_i(x^*) &\geq 0, && \forall i \in \mathcal{I}, && \text{(primal feasibility)} \\
\lambda_i^* &\geq 0, && \forall i \in \mathcal{I}, && \text{(dual feasibility)} \\
\lambda_i^* c_i(x^*) &= 0, && \forall i \in \mathcal{E} \cup \mathcal{I}. && \text{(complementarity condition/ complementary slackness)}
\end{aligned}
$$

Either $\lambda_i^* = 0$ *or* $c_i(x^*) = 0$

(*strict* complimentarity: Only one of them is zero)

KKT Ex. 1

$$\min_{x \in \mathbb{R}^2} \quad 2x_1^2 + x_2^2$$
$$\text{s.t.} \quad c_1(x) = x_1 + x_2 - 1 = 0$$



$$\mathcal{L}(x,\lambda) = 2x_1^2 + x_2^2 - \lambda_1(x_1 + x_2 - 1)$$

KKT: $\nabla_x \mathcal{L}(x^*,\lambda^*) = 0$

$c_1(x^*) = 0$

$$\frac{\partial \mathcal{L}}{\partial x_1} = 4x_1 - \lambda_1 = 0 \rightarrow x_1 = \frac{\lambda_1}{4}$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 2x_2 - \lambda_1 = 0 \rightarrow x_2 = \frac{\lambda_1}{2}$$

$$x_1 + x_2 = 1$$

$$\frac{\lambda_1}{4} + \frac{\lambda_1}{2} = 1 \longrightarrow \lambda_1 = \frac{4}{3}$$

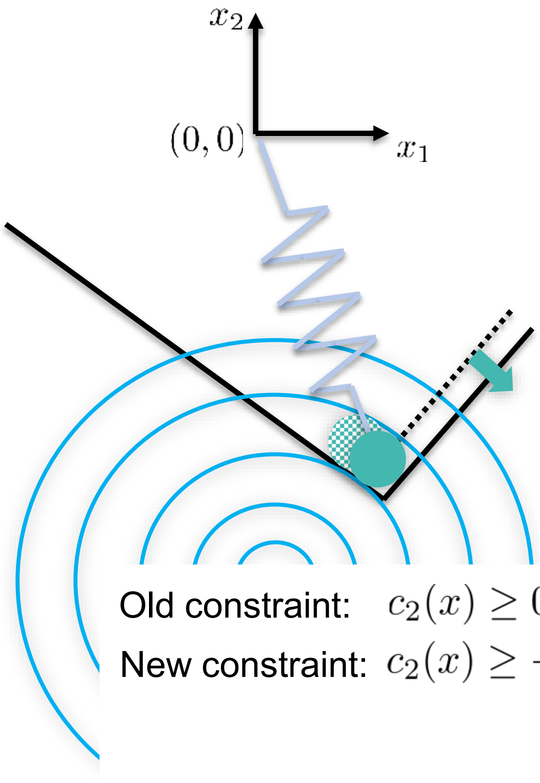
$$\Rightarrow x_1^* = \frac{1}{3}, \quad x_2^* = \frac{2}{3}, \quad \lambda_1^* = \frac{4}{3}$$

Solvability of KKT conditions

- KKT conditions can only be solved for very simple problems
  - The main complexity is the complementarity conditions – that is, deciding which inequality constraints are active or not

- What is then the use of the KKT conditions?
  - Algorithms for LP and QP are constructed by searching for points that fulfill the KKT conditions
    - LPs and (some) QPs are convex – KKT are necessary *and* sufficient
  - For nonlinear programming, we use KKT to check whether a certain iterate is a *candidate* solution
    - In general KKT are *necessary* conditions!

Multipliers: “Shadow prices”



Old constraint: $c_2(x) \geq 0$

New constraint: $c_2(x) \geq -\varepsilon$

What happens if we relax a constraint?

Feasible set becomes larger, so new minimum $f(x^*_\varepsilon)$ becomes smaller.

How much would we gain?

$$f(x^*_\varepsilon) \approx f(x^*) - \lambda\varepsilon$$

That is: The Lagrangian multipliers are the “hidden cost” (aka “shadow prices”) of constraints

KKT Conditions (Theorem 12.1)

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

$$
\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, && && \text{(stationarity)}\\
c_i(x^*) &= 0, && \forall i \in \mathcal{E}, && \\
c_i(x^*) &\geq 0, && \forall i \in \mathcal{I}, && \text{(primal feasibility)}\\
\lambda_i^* &\geq 0, && \forall i \in \mathcal{I}, && \text{(dual feasibility)}\\
\lambda_i^* c_i(x^*) &= 0, && \forall i \in \mathcal{E} \cup \mathcal{I}. && \text{(complementarity condition/ complementary slackness)}
\end{aligned}
$$

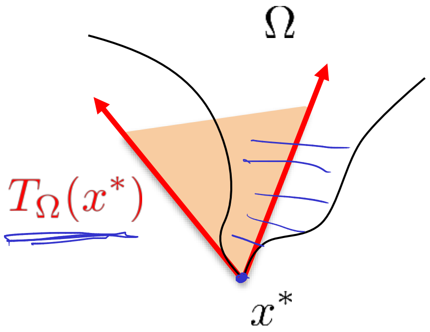
Either $\lambda_i^* = 0$ *or* $c_i(x^*) = 0$

(*strict* complimentarity: Only one of them is zero)

Geometric description of feasible directions

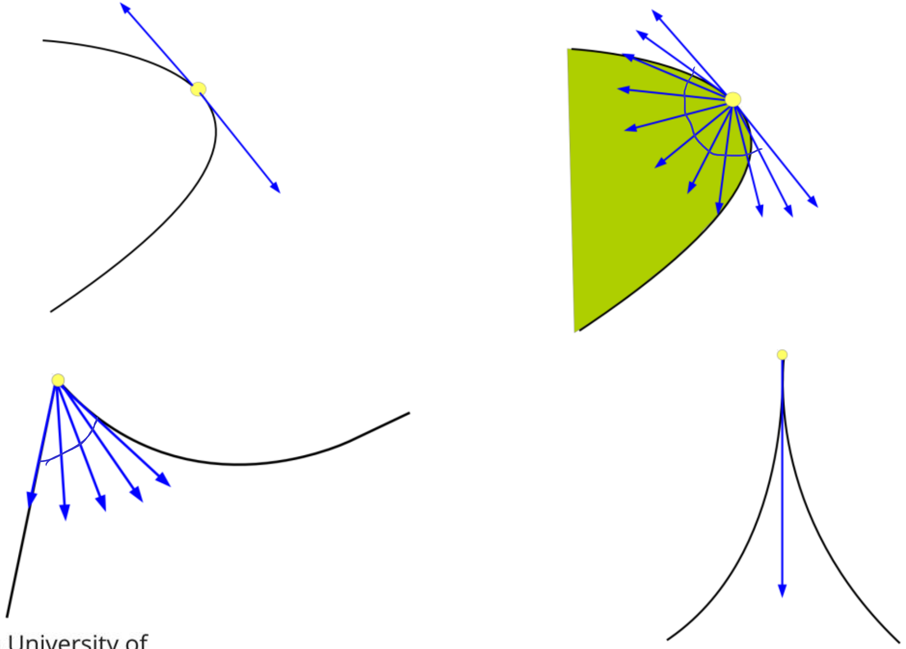
- Recall: A possible solution is a point where there are no directions that are **both** feasible and descent directions
  - Directions should be interpreted in a geometric sense

- A Tangent Cone: A geometric description of the set of feasible directions



Tangent Cone

The tangent cone to a set Ω at a point $x \in \Omega$, denoted by $T_\Omega(x)$, consists of the limits of all (secant) rays which originate at x and pass through a sequence of points $p_i \in \Omega - \{x\}$ which converges to x.



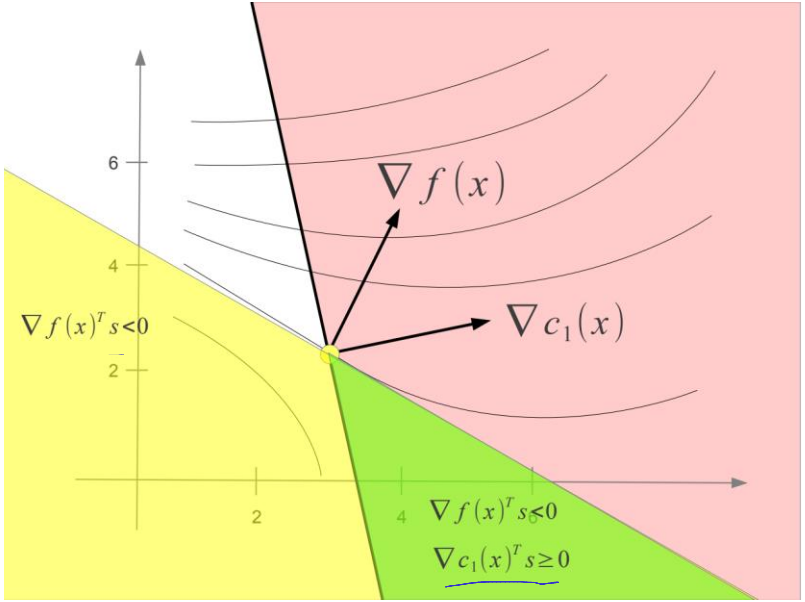
Geometric and algebraic descriptions

- A Tangent Cone: A geometric description of the feasible directions

- However, for KKT conditions we need an algebraic description using constraint gradients
  - Feasible directions for constraint *i:*

$$d^\top \nabla c_i(x) \geq 0$$

- Constraint qualifications are needed to ensure that geometric and algebraic descriptions are equivalent!



Active Set

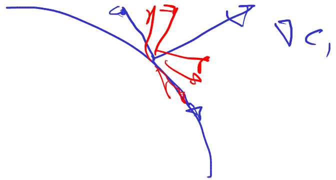
The active set $\mathcal{A}(x)$ at any feasible point x consists of the equality constraint indices from $\mathcal{E}$ together with the indices of the inequality constraints i for which $c_i(x) = 0$. That is,

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

Set of (linearized) 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

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



- Note 1: The definition of $T_\Omega(x)$ depends on the geometry of the feasible set Ω.
- Note 2: The definition of $\mathcal{F}(x)$ depends on the algebraic definition of the constraints.

Constraint Qualifications

- Constraint Qualifications are needed to rule out special cases where optimal solutions does not fulfill the KKT conditions

  A constraint qualification is an assumption that ensures similarity of the constraint set Ω and its linearized approximation, in a neighborhood of a point x^*.

  – In other words: Constraint qualifications ensure that the linearized feasible set $\mathcal{F}(x^*)$ and the tangent cone $T_\Omega(x^*)$ are the same

- The most used Constraint Qualification is LICQ:

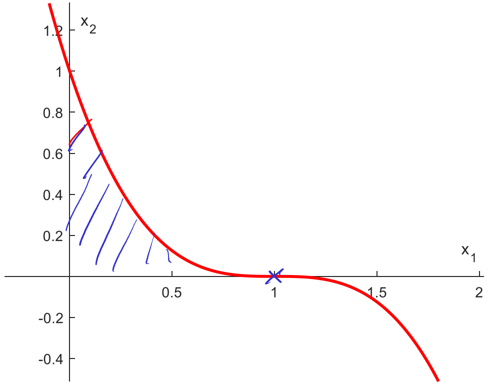
  Given the point x and the active set $\mathcal{A}(x)$, we say that the *linear independence constraint qualification* (LICQ) holds if the set of active constraint gradients $\{\nabla c_i(x),\ i \in \mathcal{A}(x)\}$ is linearly independent.

  – Other constraint qualifications exists (N&W 12.6, not syllabus)

- Note: LICQ implies uniqueness of Lagrange multipliers

LICQ Ex.

$$
\begin{aligned}
\min_{x \in \mathbb{R}^2} \quad & -x_1 \\
\text{s.t.} \quad & c_1(x) = (1-x_1)^3 - x_2 \geq 0 \\
& c_2(x) = x_1 \geq 0 \\
& c_3(x) = x_2 \geq 0
\end{aligned}
$$



Clearly, $x^* = \begin{pmatrix}1\\0\end{pmatrix}$, $\mathcal{A}(x^*) = \{1,3\}$

$\nabla c_1(x^*) = \begin{pmatrix}0\\-1\end{pmatrix}$, $\nabla c_3(x^*) = \begin{pmatrix}0\\1\end{pmatrix}$

Not lin. indep. $\Rightarrow$ LICQ does not hold.

KKT:

$$\nabla f(x^*) = \lambda_1 \nabla c_1(x^*) + \lambda_3 \nabla c_3(x^*)$$

$$\begin{bmatrix}-1\\0\end{bmatrix} = \lambda_1 \begin{bmatrix}0\\-1\end{bmatrix} + \lambda_3 \begin{bmatrix}0\\1\end{bmatrix}$$

That is: KKT cannot hold in x^*.

KKT necessary&sufficient for convex problems

- Important result not given directly in the book:

> For a **convex** constrained optimization problems where Slater's condition is fulfilled, the KKT conditions are **necessary and sufficient** for optimality

- Slater's condition is a **constraint qualification**, similar to LICQ
  - Basically: The feasible region must have an interior point
  - LICQ implies Slater's condition
  - Slater's condition implies that strong duality hold (more on this later for LP/QPs)

- **Slater's condition is fulfilled for LPs and (convex) QPs**

- Book: Only shows this result for LPs and (convex) QPs

SECOND ORDER CONDITIONS

2^{nd} Order Conditions: Critical Cone

- We have found a point x^* that fulfills KKT conditions
- Say there are directions $w \in \mathcal{F}(x^*)$ that does not lead to an increase in the objective function, that is $w^\top \nabla f(x^*) = 0,\ w \neq 0$. How do we decide whether x^* is actually a minimum?
- Second-order conditions answer this by looking at the curvature (2^{nd} derivative) in these directions

- Define the *critical cone*:

$$w \in \mathcal{C}(x^*, \lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^\top w = 0, & \forall i \in \mathcal{E}, \\ \nabla c_i(x^*)^\top w = 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \underline{\lambda_i^* > 0}, \\ \nabla c_i(x^*)^\top w \geq 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{cases}$$

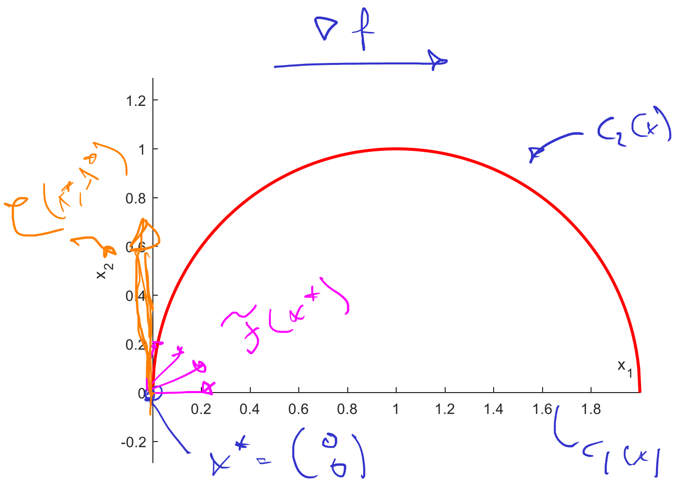
- Note: $\mathcal{C}(x^*, \lambda^*) \subseteq \mathcal{F}(x^*)$. Difference: Inequalities with positive Lagrange multiplier treated as equalities
- $\mathcal{C}(x^*, \lambda^*)$ contains the "undecided" directions from $\mathcal{F}(x^*)$, the directions where decrease/increase cannot be decided from $\nabla f(x^*)$ alone

Critical cone Ex.

$$\min_{x\in\mathbb{R}^2} \quad x_1$$

$$\text{s.t.} \quad c_1(x) = x_2 \geq 0$$

$$c_2(x) = -(x_1-1)^2 - x_2^2 + 1 \geq 0$$



$x^* = \begin{pmatrix}0\\0\end{pmatrix}, \quad \mathcal{A}(x^*) = \{1, 2\}$

KKT fulfilled with $\lambda_1^* = 0$, $\lambda_2^* = 0.5$

$\nabla c_1(x^*) = \begin{bmatrix}0\\1\end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix}2\\0\end{bmatrix}$

$\mathcal{F}(x^*) = \{d \mid d \geq 0\}$

$w \in \mathcal{C}(x^*, \lambda^*) \Leftrightarrow \begin{cases}\nabla c_2^T w = 0\\ \nabla c_1^T w \geq 0\end{cases}$

$\Rightarrow \mathcal{C}(x^*, \lambda^*) = \underline{\left\{ w = \begin{pmatrix}0\\w_2\end{pmatrix} \;\middle|\; w_2 \geq 0 \right\}}$

2nd Order Conditions: Necessary & Sufficient

- Second-order necessary conditions (Theorem 12.5):

Suppose that x^* is a local solution and that the LICQ condition is satisfied. Let λ^* be the Lagrange multiplier vector for which the KKT conditions are satisfied. Then

$$w^\top \nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w \geq 0, \qquad \text{for all } w \in \mathcal{C}(x^*, \lambda^*)$$

- Second-order sufficient conditions (Theorem 12.6):

Suppose that for some feasible point $x^* \in \mathbb{R}^n$ there is a Lagrange multiplier vector λ^* such that the KKT conditions are satisfied. Suppose also that

$$w^\top \nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w > 0, \qquad \text{for all } w \in \mathcal{C}(x^*, \lambda^*), w \neq 0$$

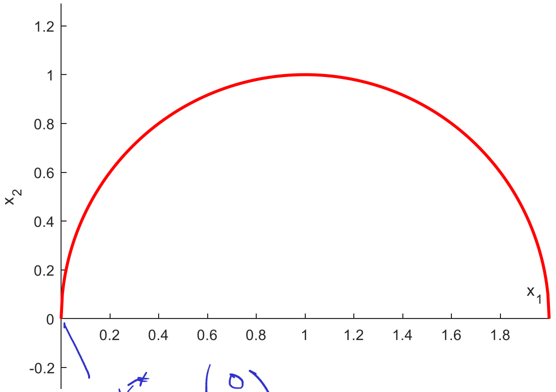
Then x^* is a strict local solution.

2nd order cond., Ex.

$$\min_{x\in\mathbb{R}^2} \quad x_1$$

$$\text{s.t.} \quad c_1(x) = x_2 \geq 0$$

$$c_2(x) = -(x_1-1)^2 - x_2^2 + 1 \geq 0$$



$x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\mathcal{L}(x,\lambda) = x_1 - \lambda_1 x_2 - \lambda_2(-(x_1-1)^2 - x_2^2 + 1)$$

$$\nabla_x \mathcal{L}(x,\lambda) = \begin{pmatrix} 1 + 2\lambda_2(x_1-1) \\ -\lambda_1 + 2\lambda_2 x_2 \end{pmatrix}$$

$$\nabla^2_{xx} \mathcal{L}(x,\lambda) = \begin{bmatrix} 2\lambda_2 & 0 \\ 0 & 2\lambda_2 \end{bmatrix}$$

$$\nabla^2_{xx} \mathcal{L}(x^*,\lambda^*) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad [\lambda_2^* = 0.5]$$

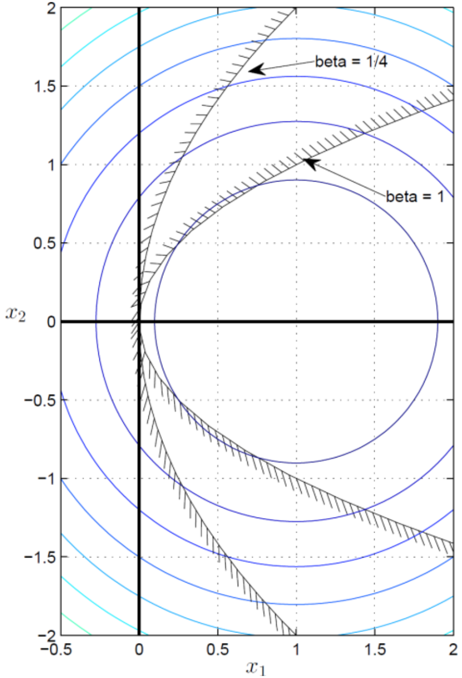
$$\begin{bmatrix} 0 \\ w_2 \end{bmatrix}$$

$$w \in \mathcal{C}(x^*,\lambda^*): \quad \underline{w^T \nabla^2_{xx} \mathcal{L}(x^*,\lambda^*) w}$$

$$= w_2^2 > 0$$

That is: Suff. cond. holds

Example:



$$\min_{x \in \mathbb{R}^2} \quad f(x) = \frac{1}{2}\left((x_1 - 1)^2 + x_2^2\right)$$

$$\text{s.t.} \quad c_1(x) = -x_1 + \beta x_2^2 = 0$$

$x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, KKT : $\lambda_1^* = 1$ | $\nabla c_1(x^*) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

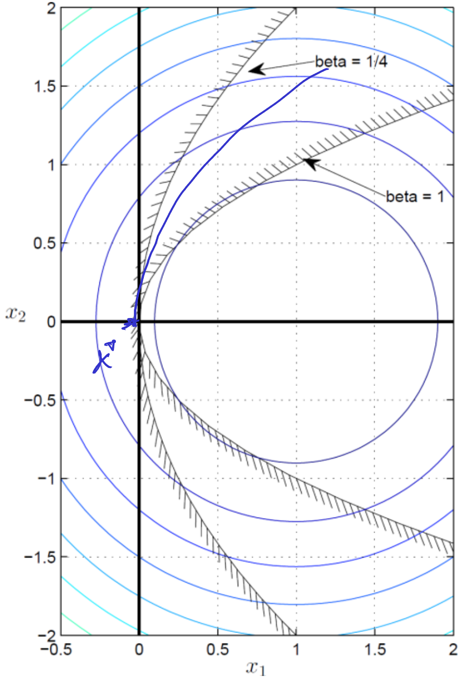
$$\mathcal{F}(x^*) = \mathcal{C}(x^*, \lambda^*) = \{ w \mid [-1 \ \ 0] w = 0 \}$$
$$= \left\{ \begin{bmatrix} 0 \\ w_2 \end{bmatrix} \right\}$$

$$\mathcal{L}(x, \lambda) = \frac{1}{2}\left[(x_1 - 1)^2 + x_2^2\right] - \lambda_1 \left(-x_1 + \beta x_2^2\right)$$

$$\nabla_x \mathcal{L}(x, \lambda) = \begin{bmatrix} x_1 - 1 + \lambda_1 \\ x_2 - 2\lambda_1 \beta x_2 \end{bmatrix}$$

$$\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 1 & 0 \\ 0 & 1 - 2\lambda_1^* \beta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 - 2\beta \end{bmatrix}$$

Example:



$$\min_{x\in\mathbb{R}^2} \quad f(x) = \frac{1}{2}\left((x_1-1)^2 + x_2^2\right)$$

$$\text{s.t.} \quad c_1(x) = -x_1 + \beta x_2^2 = 0$$

$$w \in \mathcal{C}(x^*, \lambda^*) : w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w$$

$$= \begin{bmatrix} 0 \\ w_2 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 1-2\beta \end{bmatrix} \begin{bmatrix} 0 \\ w_2 \end{bmatrix}$$

$$= (1-2\beta) w_2^2 > 0$$

when $\beta < \frac{1}{2}$

---

x^* strict local solution

when $\beta < \frac{1}{2}$

Positive Definiteness

A square, symmetric matrix A is *positive definite* if the following equivalent conditions hold:

- There is a positive scalar α such that

$$x^\top A x \geq \alpha x^\top x, \qquad \text{for all } x \in \mathbb{R}^n.$$

- $x^\top A x > 0, \qquad \text{for all } x \neq 0.$
- If all *eigenvalues* $\lambda_i > 0$.

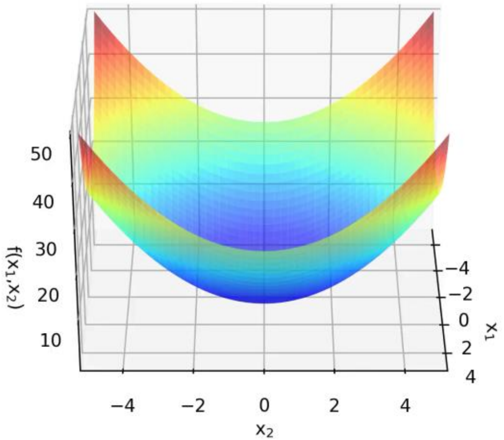
We also write $A > 0$ when A is PD.

A square matrix A is *positive semidefinite* if

$$x^\top A x \geq 0, \qquad \text{for all } x \in \mathbb{R}^n$$

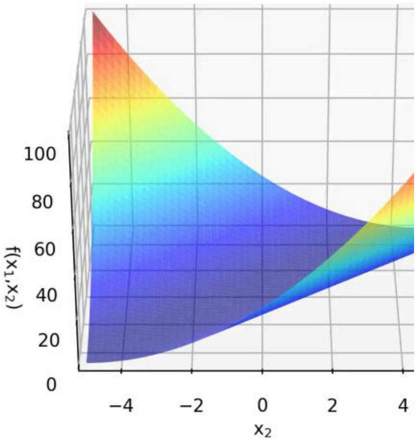
We also write $A \geq 0$ when A is PSD.

Visualizations



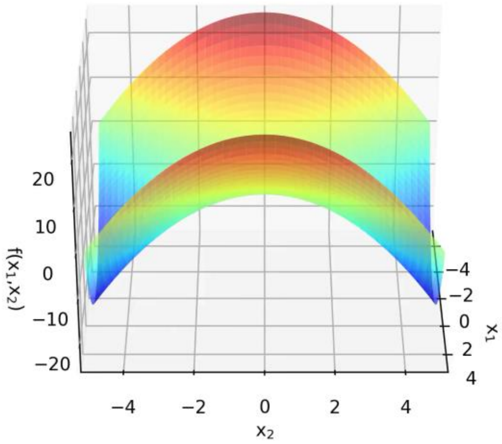
Positive Definite

$x^T Px = x_1^2 + x_2^2$



Positive Semi-definite

$x^T Px = x_1^2 + 2x_1x_2 + x_2^2$



Indefinite

$x^T Px = x_1^2 - x_2^2$