Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

What is optimization and how do we use it?

Aachener Verfahrenstechnik | RWTH AACHEN UNIVERSITY

Definition: Numerical (Mathematical) Optimization

Optimization (in everyday language):

Improvement of a good solution by intuitive, brute-force or heuristics-based decision-making

Numerical (Mathematical) Optimization:
Finding the best possible solution using a mathematical problem formulation and a rigorous/ heuristic numerical solution method

Often the term **mathematical programming** is used as an alternative to numerical optimization. This term dates back to the times before computers. The term *programming* referred to the solution of *planning problems*.

For those interested in the history of optimization: *Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, Extra Volume - Optimization Stories, 21st International Symposium on Mathematical Programming, Berlin, August 19–24, 2012*

Formulation of Optimization Problems (1)

The general formulation of an optimization problem consists of:

- The **variables** (also called decision variables, degrees of freedom, parameters, …)
- An **objective function**
- A **mathematical model** for the description of the system to be optimized
- **Additional restrictions** on the optimal solution, including bounds of the variables.

The mathematical model of the system under consideration and the additional restrictions are also referred to as **constraints**.

The objective function can either be **minimized** or **maximized**.

Formulation of Optimization Problems (2)

- The **objective function** describes an economical measure (operating costs, investment costs, profit, etc.), or technological, or ...

- The mathematical modeling of the system results in models to be added to the optimization problem as **equality constraints**.

- The **additional constraints** (mostly linear inequalities) result, for instance, from:
  - plant- or equipment-specific limitations (capacity, pressure, etc.)
  - material limitations (explosion limit, boiling point, corrosivity, etc.)
  - product requirements (quality, etc.)
  - resources (availability, quality, etc.)

[1] Edgar T. F., Himmelblau D. M., Optimization of Chemical Processes, 2nd Edition, McGraw-Hill, 2001.

Solution of Optimization Problems

What defines the solution of an optimization problem?

- Those **values** of the **influencing variables** (decision variables or degrees of freedom) are sought, which maximize or minimize the objective function.
- The values of the degrees of freedom must **satisfy** the **mathematical model** and **all additional constraints** like, for instance, physical or resource limitations at the optimum.
- The solution is, typically, a **compromise** between **opposing effects**. In process design, for instance, the investment costs can be reduced while increasing the operating costs (and vice versa).

Applications of Optimization

Optimization is widely used in science and engineering, and in particular in process and energy systems engineering, e.g.,

- **Business decisions** (determination of product portfolio, choice of location of production sites, analysis of competing investments, etc.)
- **Design decisions: Process, plant and equipment** (structure of a process or energy conversion plant, favorable operating point, selection and dimensions of major equipment, modes of process operation, etc.)
- **Operational decisions** (adjustment of the operating point to changing environmental conditions, production planning, control for disturbance mitigation and set-point tracking, etc.)
- **Model identification** (parameter estimation, design of experiments, model structure discrimination, etc.)

Short Examples

- Engineering: **design** and **operation**

- Operations research, e.g., **airlines**
  - How to **schedule routes**: results in **huge linear programs (LPs)**
  - How to price airline tickets?
    - Should the airline aim at always having full airplanes?
  - Must consider uncertainty, typically as **stochastic formulation**

- **Navigation systems**: how to go from A to B in shortest time (or shortest distance, lowest fuel consumption or ...)

- LaTeX varies **spacing and arrangement of figures** to maximize visual appeal of documents

- Successful **natural processes** not using numerical methods
  - Evolution of species
  - Behavior of animals
  - Equilibrium processes in nature maximize entropy generation

Check Yourself

- What constitutes an optimization problem?
- What types of problems are typically found?
- Why do we typically seek a compromise in optimization?
- What is the difference between a nonlinear program, an optimal control problem and a stochastic program?

Bilevel Optimization in Grad School

max great papers

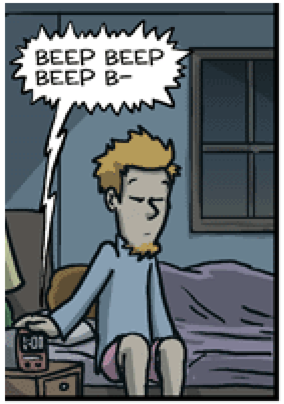
Constraints:

- # nervous breakdowns < OSHA limit
- sponsors happy

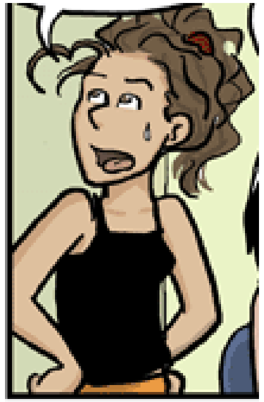
Variables:

- Pressure
  Where are my paperz?
- Encouragement
  Occasional free beer and food

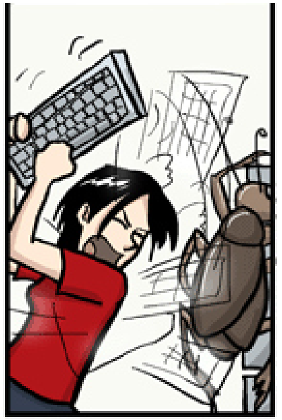
max slack



max social impact



min graduation time



max papers

Constraints:

- sleep > 4hrs
- pay rent
- keep funding

Variables:

- work load
- free lunch schemes
- seem busy schemes

Applied Numerical Optimization

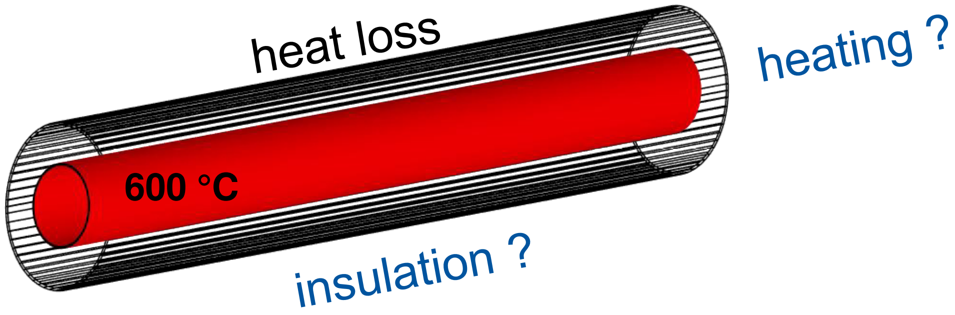
Prof. Alexander Mitsos, Ph.D.

Examples of optimization problems – basic examples

Aachener Verfahrenstechnik | RWTH AACHEN UNIVERSITY

Example: Design of a Pipeline (1)

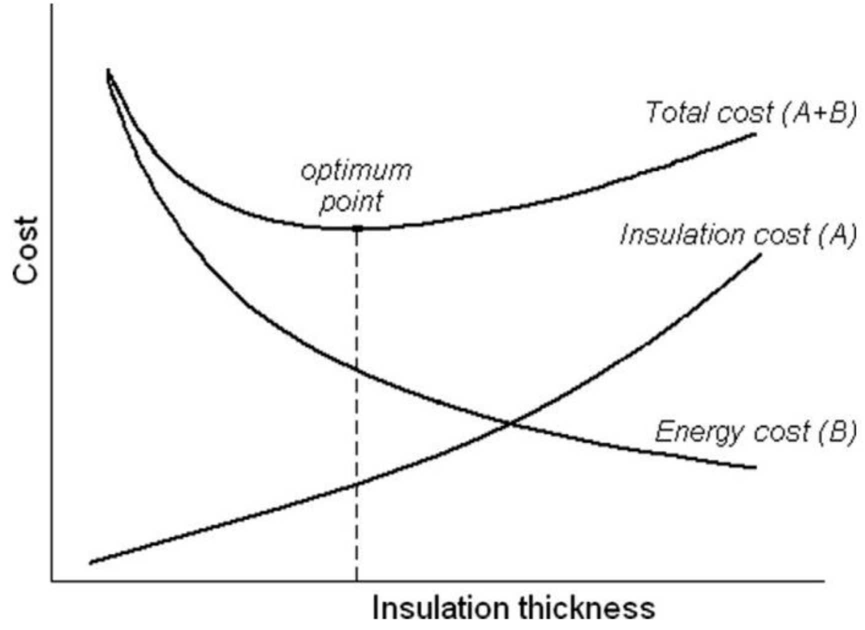
- A fluid at temperature 600°C flows through a pipeline.
- Surface heat losses must be balanced by additional heating.
- The heating costs (**operational costs**) are proportional to the heat loss, which can be reduced by the installation of an insulation (**investment costs**).



Example: Design of a Pipeline (2)

- The aim is to find the best compromise between the cost of additional heating and cost of additional insulation. The **objective function** corresponds thus to the **total (annualized) cost**.

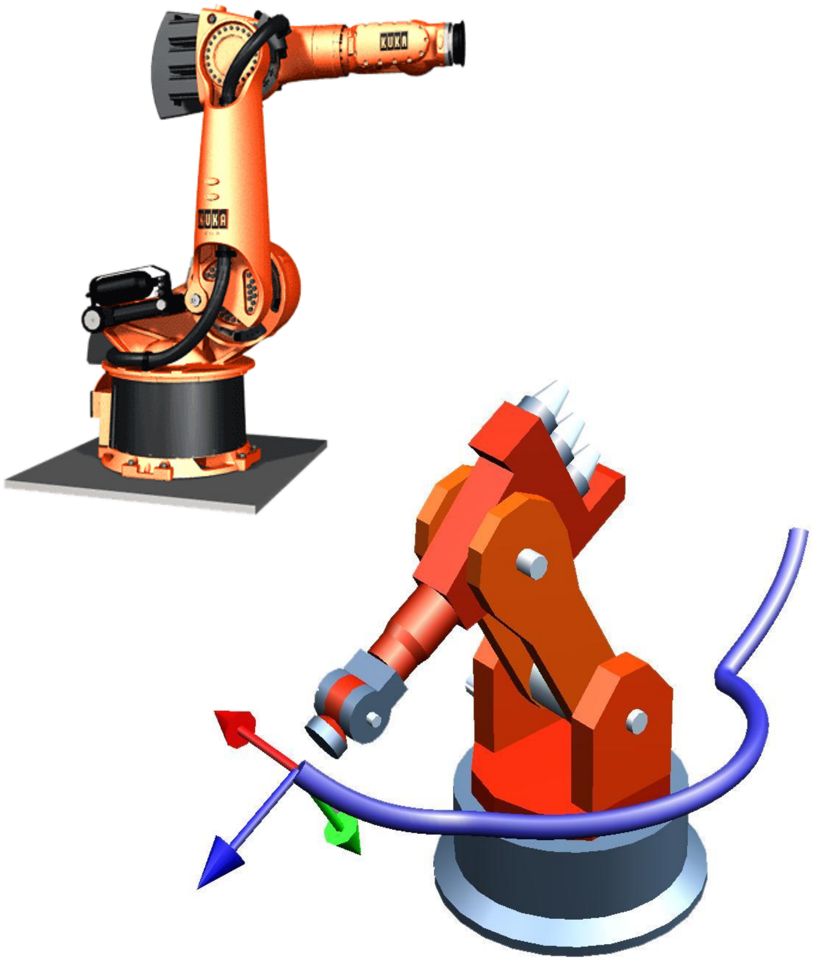
- The **degree of freedom** is the **insulation thickness**.



[1]

[1] Kaynakli, Economic thermal insulation thickness for pipes and ducts: A review study, 2014

Example: Optimal Motion Planning of Robots



Task:

- Transportation and accurate positioning of a part, e.g., during the assembly of an automobile windscreen.

Aims:

- Short cycle time for production, e.g., minimization of transportation time through optimal motion planning
- Correct positioning of the part during assembly
- No collisions during movement

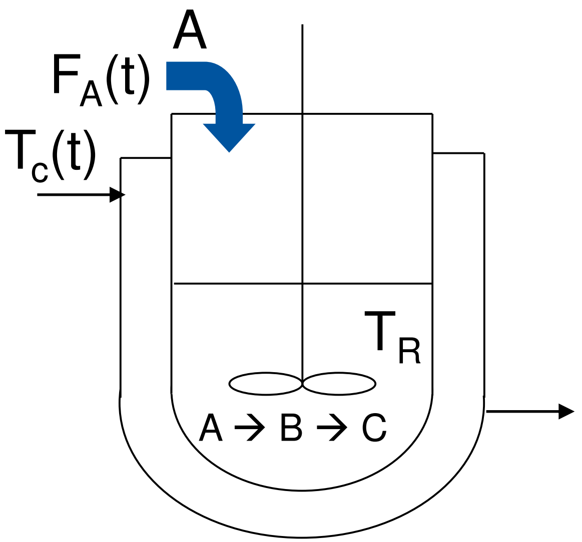
Source: FG Simulation und Systemoptimierung, TU Darmstadt; Kuka Roboter GmbH

Example: Optimization of Semi-batch Reactor Operation

In a semi-batch reactor, a product B should be manufactured according to the reaction scheme

$$A \rightarrow B \rightarrow C$$

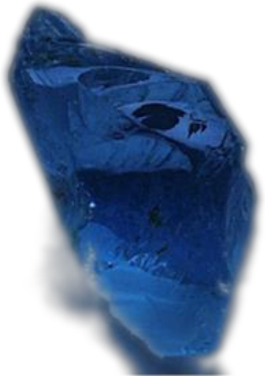
where C is an undesirable by-product.

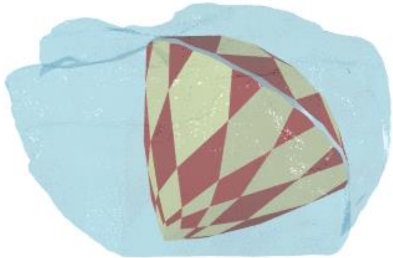


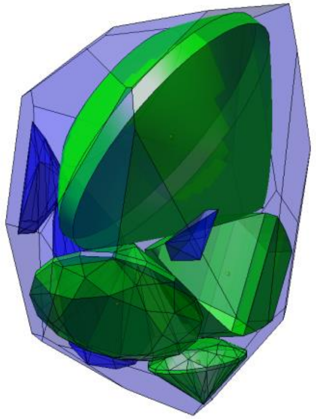
Optimization problem:

- The selectivity of the reaction can be maximized over the batch by manipulating the **dosage of reactant A** and the **reaction temperature**.
- The degrees of freedom are **functions of time.**
- Like in robot motion planning, this problem is an **optimal trajectory planning** problem.

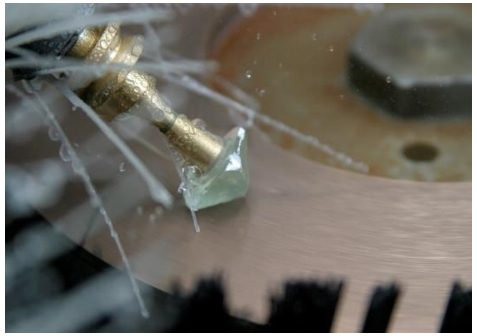
Example: Gemstone Cutting as (Multi-Body) Design Centering Problem







Optimal Cut?



maximize gemstone volume
minimize waste

Pictures from Fraunhofer ITWM
Algorithm by Oliver Stein, solver by ITWM

Check Yourself

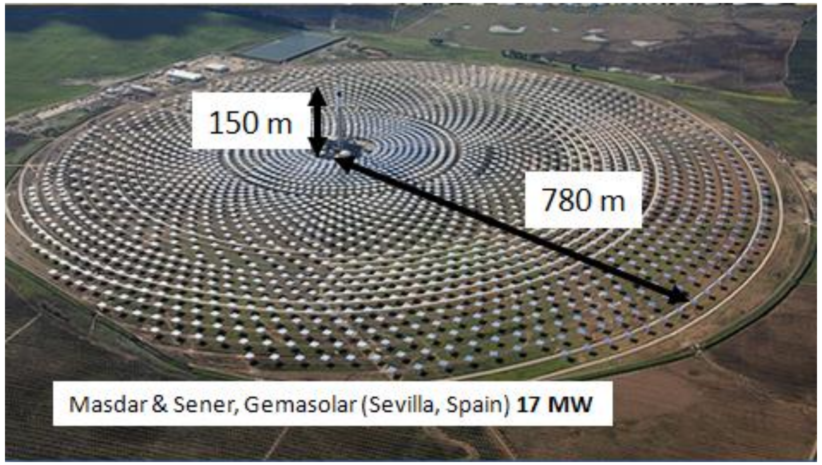
- For each of the considered examples, state: variables, objective function, model, additional constraints
- Formulate the application of your interest as an optimization problem
- What are some limits of optimization?

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Examples of optimization problems – solar thermal

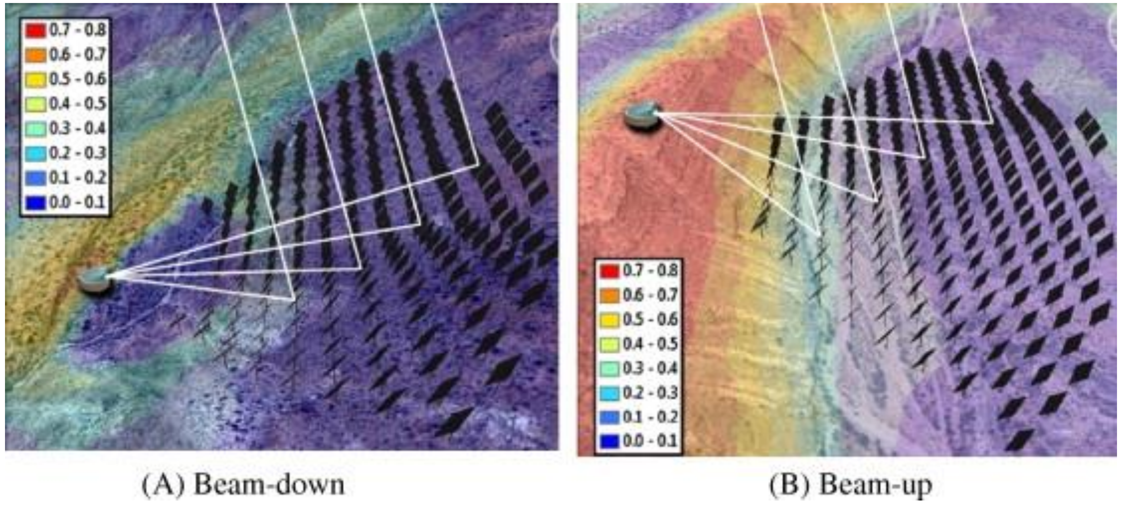
Example: Heliostat Fields – Construct on Plane or Hill?



Masdar & Sener, Collage: D. Codd

eSolar, Collage: D. Codd

- **Renewable energy** requires huge land areas and is expensive
- **Central receiver plants** - a promising scalable technology
- Can use hills in beam-down (CSPonD) or beam-up (“natural-tower")



Noone, et int. Mitsos*, Solar Energy

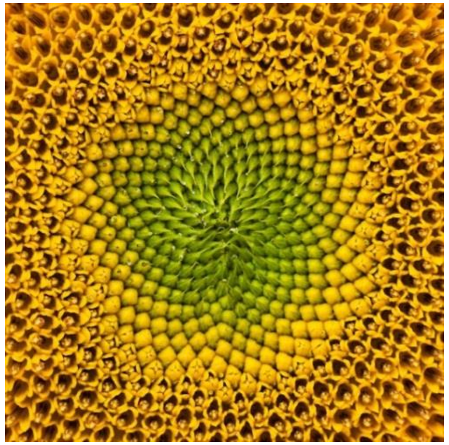
CSPonD: Slocum*, Codd, et int. Mitsos, Solar Energy

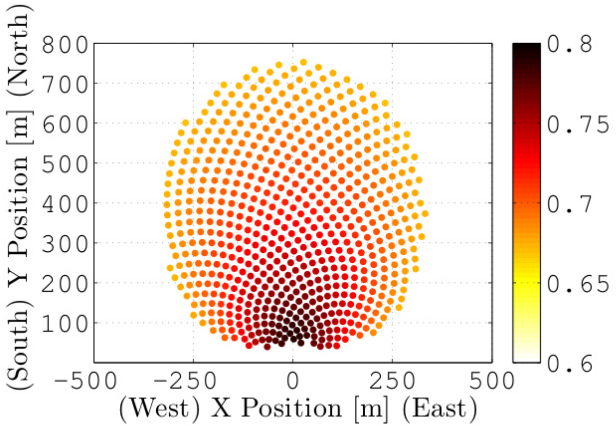
Example: Heliostat Fields – Optimization Applicable?

- Objective: Maximize field efficiency and minimize land area usage
  – Minimize economical & ecological costs
  – Factorial number of local minima
- Noone (with guidance by Mitsos) developed and validated a model suitable for optimization (fast yet accurate, compatible with reverse mode algorithmic differentiation)
- Heuristic global methods (genetic algorithm, multistart) prohibitive for realistic number of heliostats
- Local optimization from arbitrary initial guess not suitable as results are very sensitive to initial guess
- Heuristic solution tried: Start with existing designs and optimize locally
- Result obtained: Spiral pattern recognized by Prof. Manuel Torrilhon
- Long-term goal: Deterministic global optimization using Relaxation of Algorithms [1]

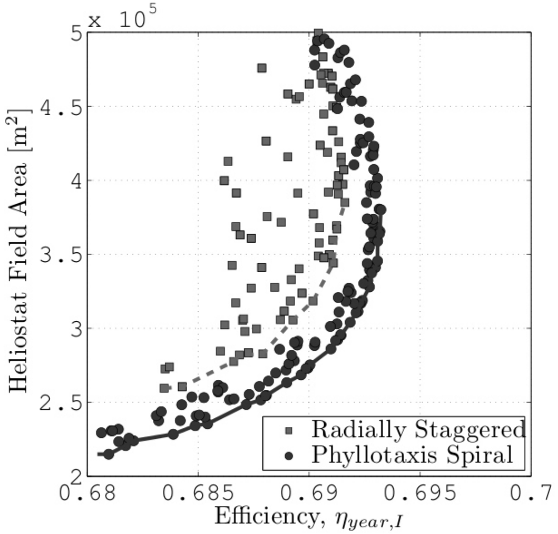
[1] Mitsos A., Chachuat B. and Barton P., McCormick-based relaxations of algorithms. SIAM Journal on Optimization, 2009

Example: Heliostat Field Optimization – Some Results





- Identified **spiral pattern** from **local optimization** of radially staggered pattern [1]
  - Abengoa concurrently proposed spiral
- Optimized biomimetic spiral → appreciable improvement in efficiency, substantial savings in land area.



http://www.bbc.co.uk/mundo/noticias/2012/01/120123_girasol_energia_solar_am.shtml, https://www.popsci.com/technology/article/2012-01/sunflower-design-inspires-more-efficient-solar-power-plants/ and picked up by many more…

[1] Noone C. J., Torrilhon M., Mitsos A., Heliostat field optimization: A new computationally efficient model and biomimetic layout, Solar Energy, 2012

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

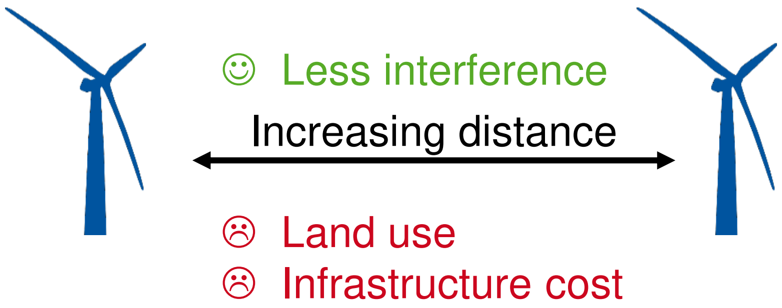
Examples of optimization problems - wind

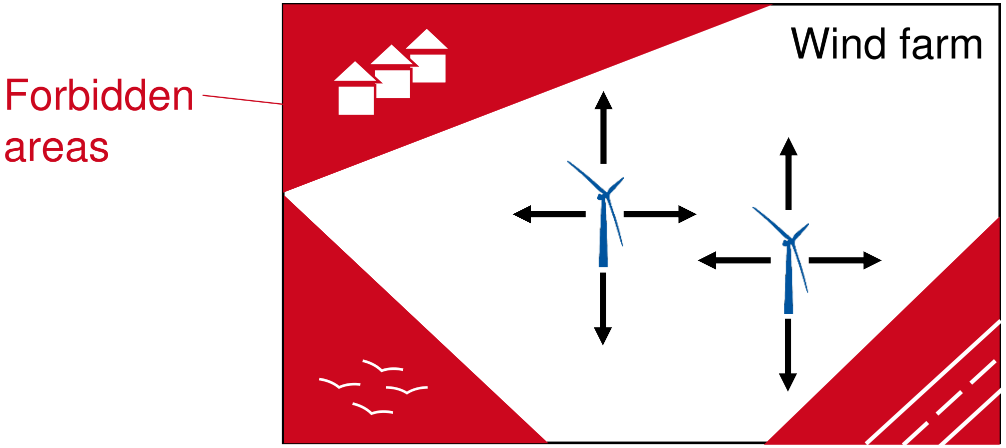
Aachener Verfahrenstechnik | RWTH AACHEN UNIVERSITY

Example: Wind Farm Layout Optimization

- **Wind turbines** are built in groups (=wind farms) to produce more electricity in a given limited area
- **Wind farm layout:** Where to position turbines within farm limits? Potentially also: How many turbines?
- **Typical objectives:**
  - Maximize annual electricity production
  - Minimize levelized cost of electricity (=cost per unit of electrical energy)

- **Key Factor:** Distance between turbines

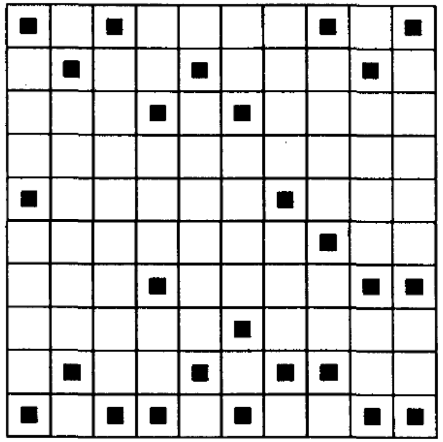


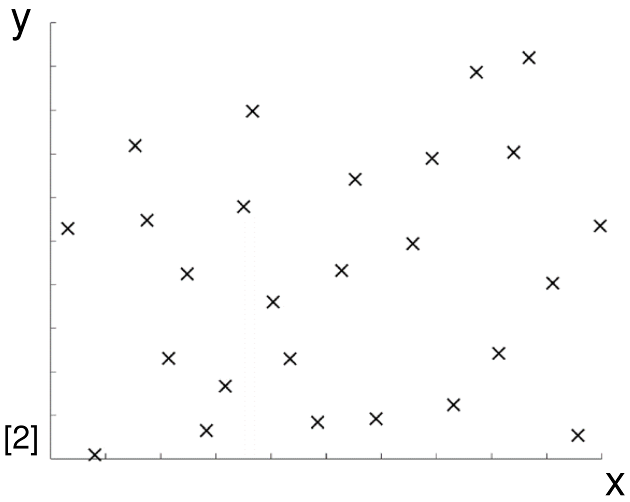


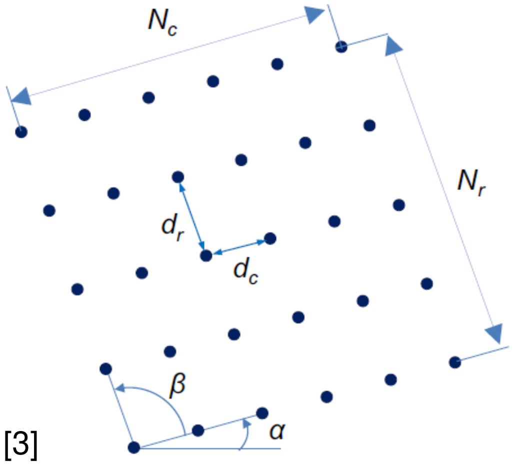
[1] https://commons.wikimedia.org/wiki/File:Wind_farm_near_North_Sea_coast.jpg (CC BY-SA 4.0)

Example: Wind Farm Layout Optimization – How to Describe Layout?

Fixed cells	Continuous positions	Patterns

[1] 

[2] 

[3] 

Fixed cells
- ☺ easy to optimize #turbines
- ☹ less freedom
- ☹ many discrete variables

Continuous positions
- ☺ most freedom
- ☹ difficult to optimize #turbines
- ☹ many continuous variables

Patterns
- ☺ few variables
- ☹ less freedom
- ☹ complex areas difficult

→ Has implications on applicable optimization algorithms & quality of solution

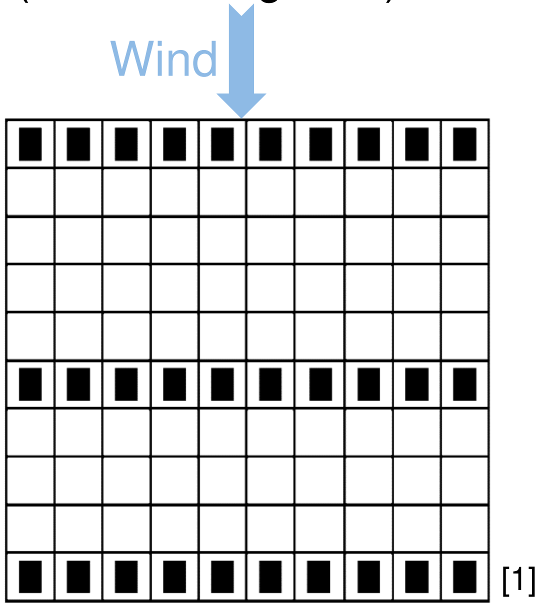
[1] Mosetti et al. (1994). J. Wind Eng. Ind. Aerod., 51, 105.
[2] DuPont & Cagan (2012). J. Mech. Des., 134, 081002. Mod.
[3] González et al. (2017). Appl. Energ., 200, 28.

Example: Wind Farm Layout Optimization – Global Optimization

- **Most basic case:** constant wind from one direction, minimize levelized cost of electricity

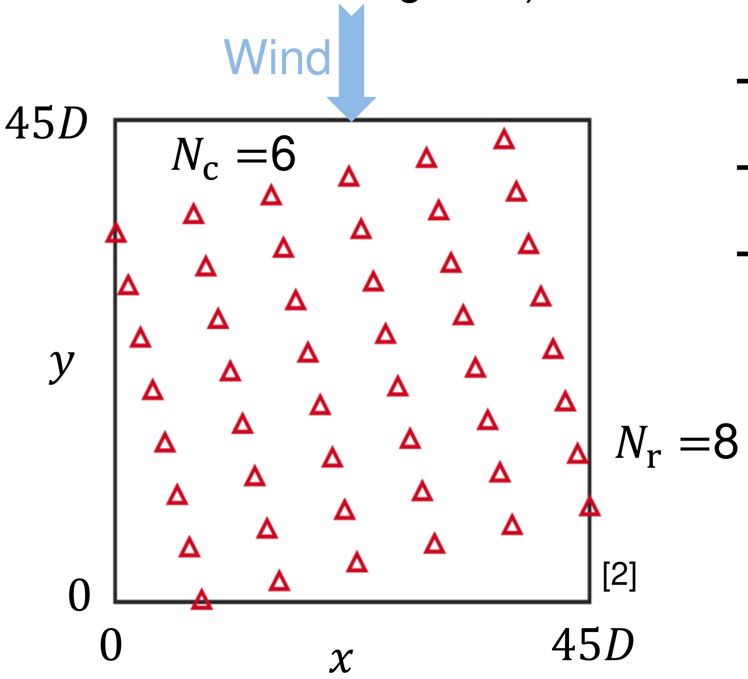
Benchmark solution

- Fixed cells approach
- Genetic algorithm (stochastic global)



Improved solution

- Pattern approach
- MAiNGO (open source, deterministic global)



→ Levelized cost of electricity - 13 %

→ Annual electricity production +68 %

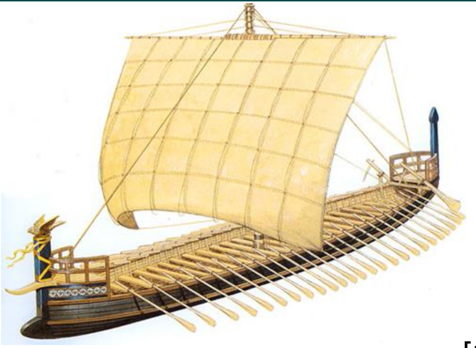
→ Efficiency +4.4 %-pt

→ Both are optimized layouts

→ Problem formulation and algorithm make a difference

[1] Grady et al. (2005), Renew. Energ., 30, 259.
[2] Bongartz (2020), in preparation.

Example: Sailing – Technology Choice



[1]

Ancient sailing: Fixed mast; no boom → mostly downwind

[2]

[3]

Classic sailing: Fixed mast; boom → can go upwind

[4]

Novel hulls (catamaran)
Novel sails (wing, …)

Typically, inventions by human creativity, not by mathematical optimization.

[5]

Windsurfing: Mast moves

[6]

Kite-surfing: No mast

[7]

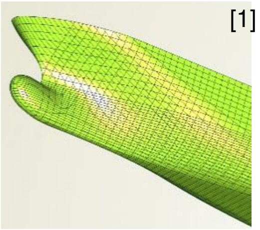
Hydrofoiling: wing in water.
From planning to flying!

[8]

Wing instead of sail or kite.
No mast, no boom, no ropes

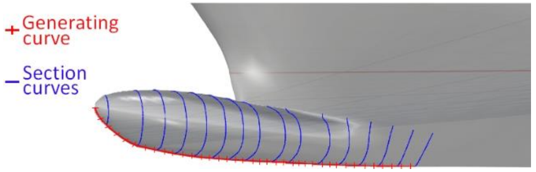
[1] via http://www.salimbeti.com/micenei/ships.htm [2] Photo credit Eva Lambidoni
[3] Photo credit Jörn Viell [4] Alexander Mitsos & Daniel Fouquet (skipper), (photo credit Eva Lambidoni)
[5] Alexander Mitsos (photo credit Stephanie Mitsos) [6] Verena Niem & Dimitris Chatzigeorgiou
[7] Sotiris Kontolios [8] Robby Naish

Example: Sailing – Optimization



CAD model of a ship's hull

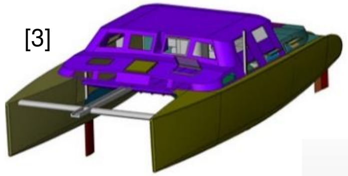
Infinite degrees of freedom



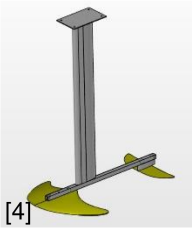
Geometric parametrization of bulbous bow [2]

Complex physics! Coupled hydro-, aero- and structural-dynamics problem

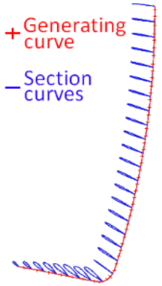
CAD model → Parametric modeler → (Shape parameters) → Mathematical modeling → (Objective function, Constraints) → Optimization algorithm → Optimal shape



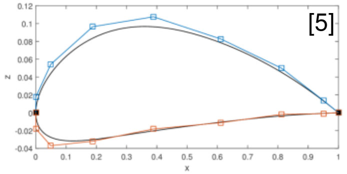
CAD model of a Catamaran



CAD model of a hydrofoil



Geometric parametrization of hydrofoil [2]



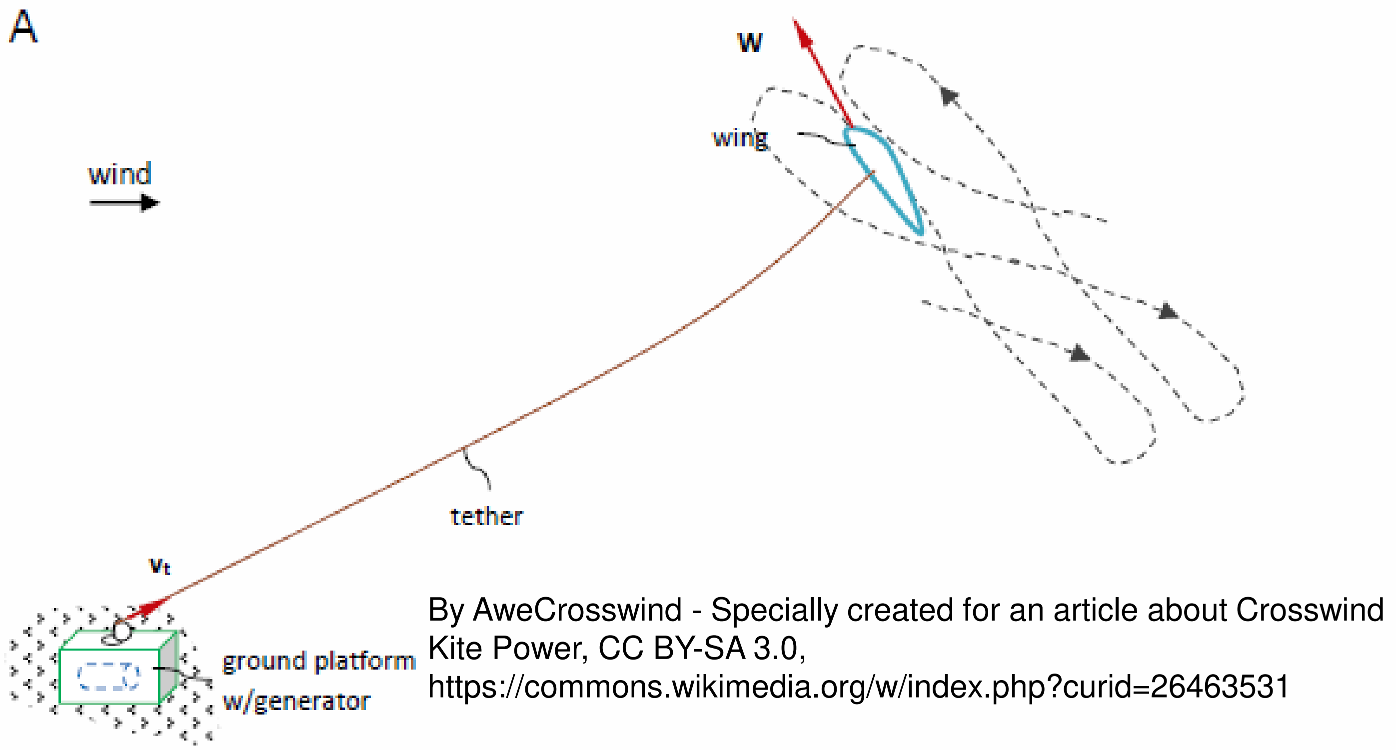
Exemplary geometry approximation using B-splines

Choice of parametrization decides the degrees of freedom

[1] via https://www.friendship-systems.com/solutions/for-ship-design (copyright free)
[2] Berrini et al., Geometric Modelling and Deformation for Shape Optimization of Ship Hulls and Appendages
[3] via http://www.wavescalpel.com/development-first-solidworks-cad-model-for-wavescalpel/ (copyright free)
[4] via https://grabcad.com/ (copyright free)
[5] Masters et al., A Geometric Comparison of Aerofoil Shape Parameterisation Methods, 2016

Renewable Electricity Generation by Kite: Optimization of Operation

- Wind power generation by kite



By AweCrosswind - Specially created for an article about Crosswind Kite Power, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=26463531

- Kite has to be moved to generate power, $\int F(t)v(t)dt > 0$
- Hard optimal control under uncertainty problem

- Optimization over finite control

$$\begin{aligned} &\underset{u(t)}{\text{maximize}} \quad \bar{T}(t_f) := \frac{1}{t_f}\int_0^{t_f} T(t)dt, \\ &\text{subject to} \quad |u(t)| \le u_{\max}, \\ &r\sin(\theta(t))\cos(\phi(t)) \ge z_{\min}, \\ &|\psi(t)| \le 2\pi. \end{aligned}$$

- Noisy data, uncertain wind prediction
- Inaccurate control model
- Path found by ad-hoc schemes or based on nonlinear model-predictive control

Costello, Francois & Bonvin European Journal of Control 2017

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Classification and issues of optimization

Aachener Verfahrenstechnik
RWTH AACHEN UNIVERSITY

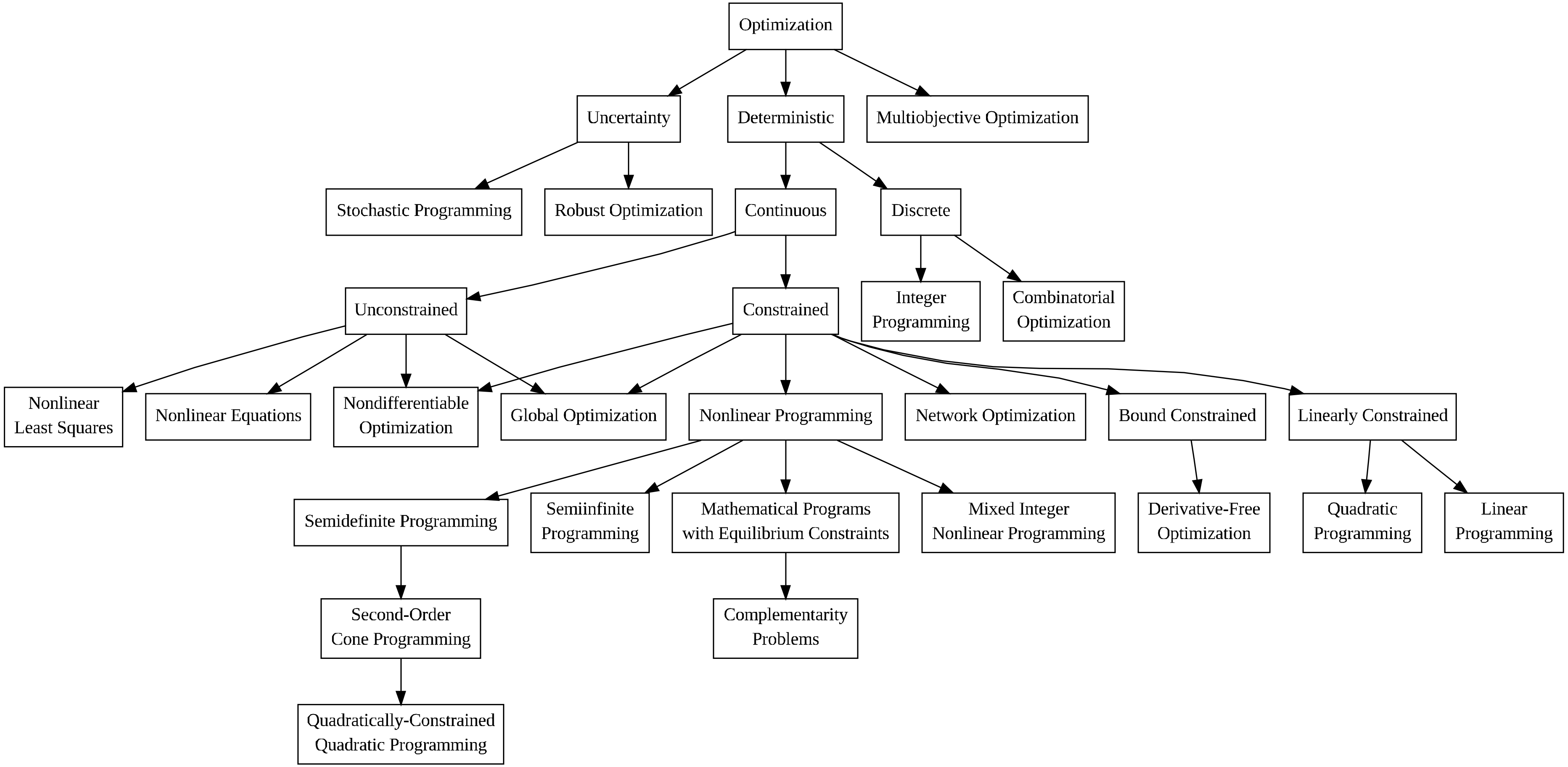
Classification of Optimization Problems

Optimization problems are classified with respect to the type of the objective function, constraints and variables, in particular

- **Linearity of objective function and constraints:**
  - Linear (LP) versus nonlinear programs (NLP)
  - NLPs can be convex or nonconvex, smooth or nonsmooth
- **Discrete and/or continuous variables:**
  - Integer programs (IP) and mixed-integer programs (MIP or MILP and MINLP, respectively)
- **Time-dependence:**
  - Dynamic optimization or optimal control programs (DO or OCP)
- **Stochastic or deterministic models and variables:**
  - Stochastic programs, semi-infinite optimization, …
- **Single objective vs multi-objective, single-level vs multi-level, ...**

NEOS Classification of Stationary Optimization Problems

http://neos-guide.org/content/optimization-taxonomy



Common Terminology Used in Numerical Optimization

- An **optimization problem:** mathematical formulation to find the best possible solution out of all feasible solutions. Typically comprising one or multiple objective function(s), decision variables, equality constraints and/or inequality constraints.

- An **algorithm** is a procedure for solving a problem based on conducting a sequence of specified actions. The terms ‘algorithm’ and ‘solution method’ are commonly used interchangeably.

- A **solver** is the implementation of an algorithm in a computer using a programming language. Often, the terms ‘solver’ and ‘software’ are used interchangeably.

Formulation and Solution of Optimization Problems

1. Determine variables and phenomena of interest through **systems analysis**
2. Define optimality criteria: **objective function(s)** and (additional) **constraints**
3. Formulate a **mathematical model** of the system and determination of **degrees of freedom** (number and nature)
4. Identify of the **problem class** (LP, QP, NLP, MINLP, OCP etc.)
5. Select (or develop) a suitable **algorithm**
6. Solve the problem using a numerical **solve**r
7. **Verify the solution** through sensitivity analysis, understand results, …

Some Issues with Optimization

- Not a button-press technology
  - Need expertise for model formulation, algorithm selection and tuning, checking results, …

- "Optimizer's curse": solution using good algorithm and bad model will look better than what it is
  - Random error: if the model has a random error and we optimize, the true objective value of the solution found will be worse than the calculated one
  - If model allows for nonphysical solution with good objective value, good optimizer will pick such
    - On the other hand, model has to just lead in correct direction, not be correct

- Many engineering (design) problems are nonconvex, but global algorithms are inherently very expensive

- Often optimal solution at constraint, thus tradeoff good vs. robust solution

Check Yourself

- What is the difference between a nonlinear program, an optimal control problem and a stochastic program?
- What are the steps in formulating and solving an optimization problem?
- What are some issues in optimization?
- Formulate the application of your interest as an optimization problem

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Formal definition of optimization

Aachener Verfahrenstechnik

RWTH AACHEN UNIVERSITY

Some Simple Optimization Problems and Their Solutions

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$ objective function

$$\text{s.t.}\quad c_i(\boldsymbol{x}) = 0, \forall i \in E$$ equality constraints (EC)

$$c_i(\boldsymbol{x}) \leq 0, \forall i \in I$$ inequality constraints (IC)

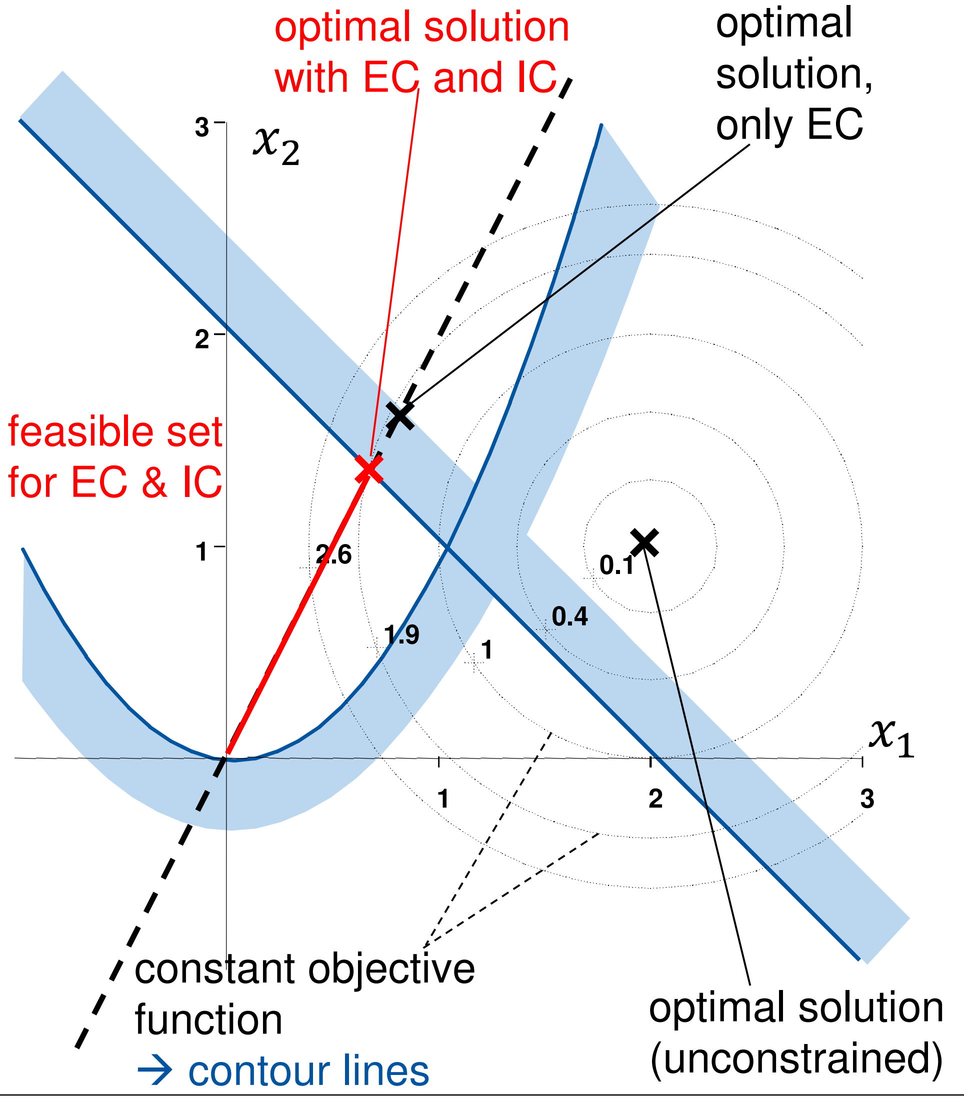
Example:

$$\min_{\boldsymbol{x}} (x_1 - 2)^2 + (x_2 - 1)^2$$

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

$$x_1^2 - x_2 \leq 0$$

$$x_1 + x_2 \leq 2$$



Nonlinear Optimization Problem (Nonlinear Program, NLP)

General formulation:

$$\min_{\boldsymbol{x} \in D} f(\boldsymbol{x})$$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, i \in E$$

$$c_i(\boldsymbol{x}) \leq 0, i \in I$$

$\boldsymbol{x} = [x_1, x_2, \dots, x_n]^T \in D \subseteq R^n$ a vector (point in n-dimensional space)

D **host set**

$f : D \to R$ **objective function**

$c_i : D \to R$ constraint functions $\forall\, i \in E \cup I$

E the index set of **equality constraints**

I the index sets of **inequality constraints**

The **constraints** and the host set define **the feasible set**, i.e., the set of all feasible solutions:

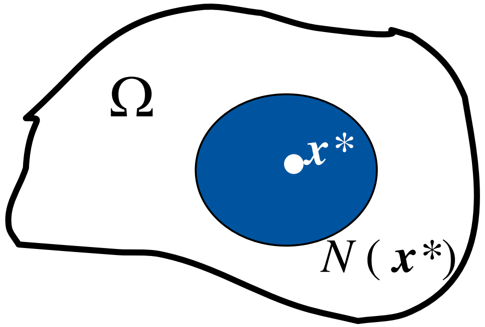
$$\Omega = \{\boldsymbol{x} \in D \mid c_i(\boldsymbol{x}) \leq 0\ \forall\, i \in I, c_i(\boldsymbol{x}) = 0\ \forall\, i \in E\}$$

Equivalent formulation: $\boxed{\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x})}$

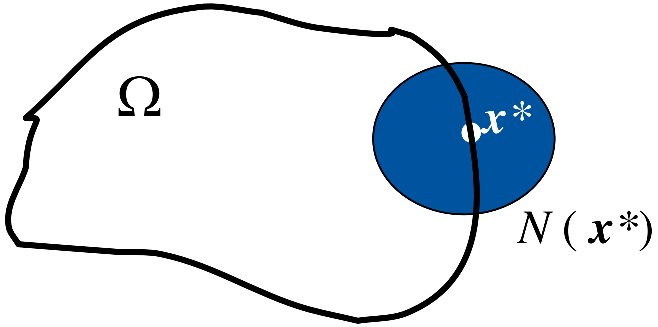
What Is an Optimal Solution ?

Definition (optimal solution, minimum):

$$\boxed{\min_{x\in\Omega} f(\boldsymbol{x})}$$



Solution in interior of feasible set



Solution on boundary of feasible set

a) $\boldsymbol{x}^*$ is a local solution if $\boldsymbol{x}^* \in \Omega$ and a neighborhood $N(\boldsymbol{x}^*)$ of $\boldsymbol{x}^*$ exists: $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})\ \forall \boldsymbol{x} \in N(\boldsymbol{x}^*) \cap \Omega$

b) $\boldsymbol{x}^*$ is a strict local solution if $\boldsymbol{x}^* \in \Omega$ and a neighborhood $N(\boldsymbol{x}^*)$ of $\boldsymbol{x}^*$ exists: $f(\boldsymbol{x}^*) < f(\boldsymbol{x})\ \forall \boldsymbol{x} \in N(\boldsymbol{x}^*) \cap \Omega$, $\boldsymbol{x} \neq \boldsymbol{x}^*$

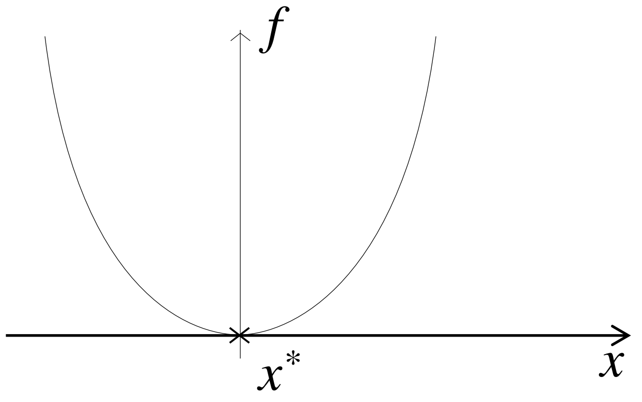
c) $\boldsymbol{x}^*$ is a global solution if $\boldsymbol{x}^* \in \Omega$ and $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})\ \forall \boldsymbol{x} \in \Omega$

More formally, these are solution points

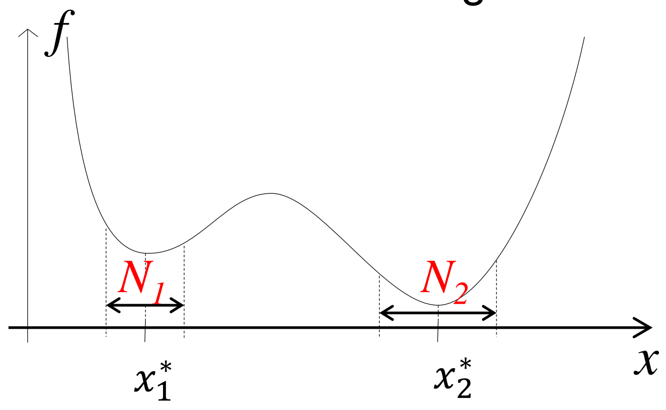
Optimal Solution – Some Examples

$$\min_{x \in R} f(x)$$

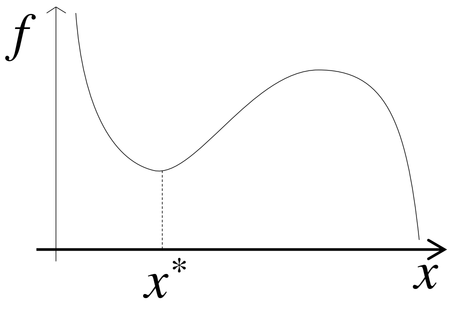
a) strict global minimum



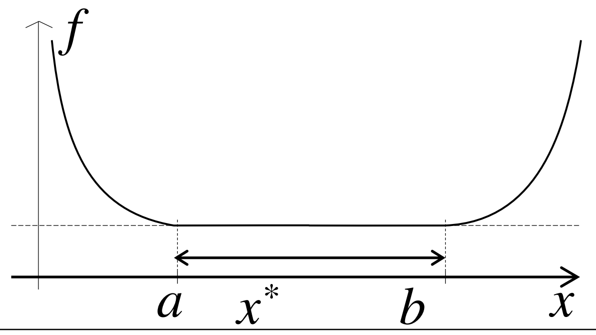
b) Two strict local minima,
out of which one is strict global minimum



c) a strict local minimum,
no global minimum



d) each $x^* \in [a, b]$ is a local and global minimum
no strict minima



Check Yourself

- Write down the general definition of optimization problem
- Definition of local and global solution of an optimization problem?
- Is every local solution also a global solution? Is every global solution also a local solution?
- What is the feasible set of an optimization problem?
- Can a solution be in the interior of the feasible set? On its boundary? Outside the feasible set?
  - Draw the corresponding picture
- For given problem recognize the (local or global) optimal solution points

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Mathematical background

Nonlinear Optimization Problem (Nonlinear Program, NLP)

General formulation:

$$\min_{\boldsymbol{x} \in D} f(\boldsymbol{x})$$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) = 0, i \in E$$

$$c_i(\boldsymbol{x}) \leq 0, i \in I$$

$\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T \in D \subseteq R^n$ a vector (point in *n*-dimensional space)

D **host set**

$f : D \rightarrow R$ **objective function**

$c_i : D \rightarrow R$ constraint functions $\forall\, i \in E \cup I$

E the index set of **equality constraints**

I the index sets of **inequality constraints**

The **constraints** and the host set define **the feasible set**, i.e., the set of all feasible solutions:

$$\Omega = \{\boldsymbol{x} \in D \mid c_i(\boldsymbol{x}) \leq 0\ \forall\, i \in I,\ c_i(\boldsymbol{x}) = 0\ \forall\, i \in E\}$$

Equivalent formulation:

$$\boxed{\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x})}$$

Directional Derivative

Definition:

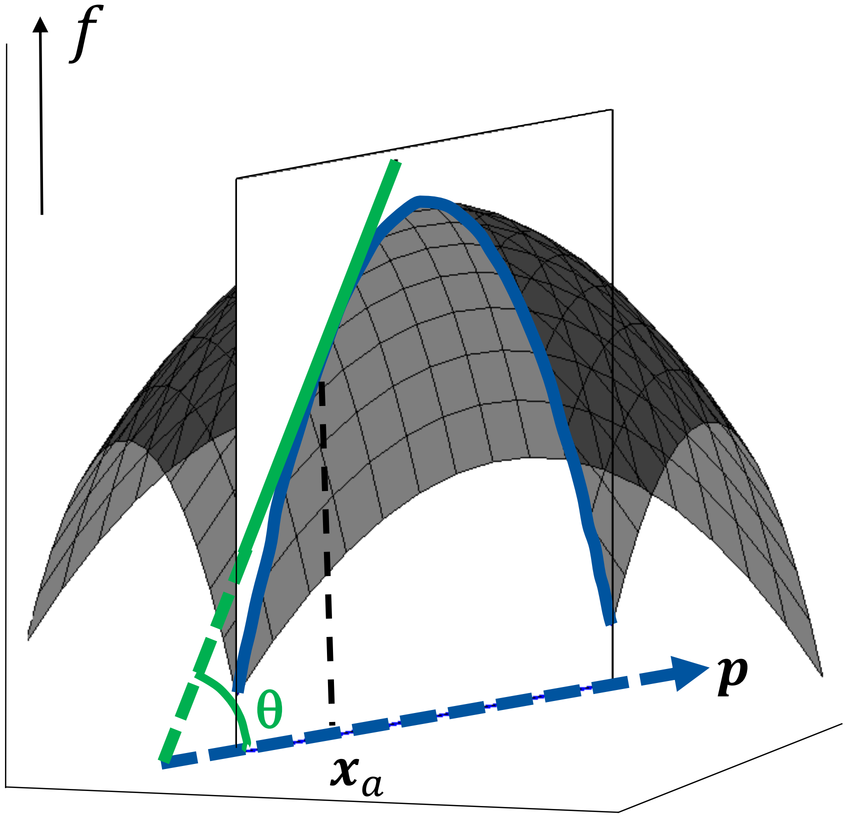
Let $f: D \to R$, $D \subseteq R^n$, $\boldsymbol{x} \in D$ and $\boldsymbol{p} \in R^n$ with $\|\boldsymbol{p}\| = 1$.

f is differentiable at the point $\boldsymbol{x} = \boldsymbol{x}_a$ in the direction $\boldsymbol{p}$ if the limit,

$$D(f, \boldsymbol{p})|_{\boldsymbol{x}=\boldsymbol{x}_a} = \lim_{\varepsilon \to 0} \frac{f(\boldsymbol{x}_a + \varepsilon \boldsymbol{p}) - f(\boldsymbol{x}_a)}{\varepsilon} =: \boldsymbol{\nabla}_{\boldsymbol{p}} f(\boldsymbol{x}_a)$$

exists and is finite.

$D(f, \boldsymbol{p})$ is called the directional derivative of f in the direction $\boldsymbol{p}$.



Gradient

Definition:

- The first derivative of a scalar, continuous function f is called the gradient of f at point $\boldsymbol{x}$:

$$\boldsymbol{\nabla} f(\boldsymbol{x}) = \begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{\boldsymbol{x}} \\ \vdots \\ \left.\frac{\partial f}{\partial x_n}\right|_{\boldsymbol{x}} \end{bmatrix}.$$

Remarks:

- If $\boldsymbol{x}$ is a function of time t, the chain rule applies:

$$\left.\frac{df}{dt}\right|_{\boldsymbol{x}(t)} = \boldsymbol{\nabla} f(\boldsymbol{x})^T \left.\frac{d\boldsymbol{x}}{dt}\right|_t = \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x_i}\right|_{\boldsymbol{x}(t)} \left.\frac{\partial x_i}{\partial t}\right|_t .$$

- The directional derivative is related to the gradient:

$$D(f(\boldsymbol{x}), \boldsymbol{p}) = \boldsymbol{\nabla}_{\boldsymbol{p}} f(\boldsymbol{x}) = \lim_{\varepsilon \to 0} \frac{f(\boldsymbol{x} + \varepsilon \boldsymbol{p}) - f(\boldsymbol{x})}{\varepsilon} = \boldsymbol{\nabla} f(\boldsymbol{x})^T \boldsymbol{p}$$

Hessian (matrix)

Definition:

- The second derivative of a scalar, twice continuously differentiable function f is the symmetric Hessian (matrix) $\boldsymbol{H}(\boldsymbol{x})$ of the function f

$$\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{\nabla}^2 f(\boldsymbol{x}) = \begin{bmatrix} \left.\frac{\partial^2 f}{\partial x_1^2}\right|_{\boldsymbol{x}} & \cdots & \left.\frac{\partial^2 f}{\partial x_1 x_n}\right|_{\boldsymbol{x}} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial^2 f}{\partial x_1 x_n}\right|_{\boldsymbol{x}} & \cdots & \left.\frac{\partial^2 f}{\partial x_n^2}\right|_{\boldsymbol{x}} \end{bmatrix}.$$

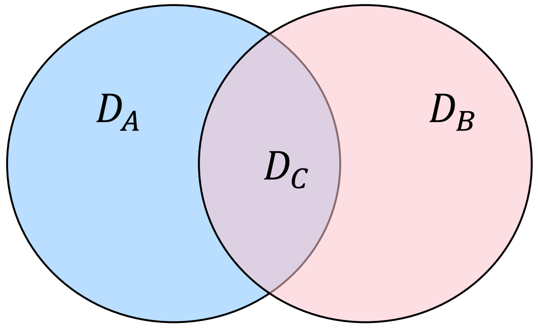
Necessary and Sufficient Conditions: Definitions and Properties

- Necessary condition: Statement *A* is a *necessary condition* for statement *B* if (and only if) the falsity of *A* guarantees the falsity of *B.* In math notation: $\text{not}A \Rightarrow \text{not}B$
- Sufficient condition: Statement *A* is a *sufficient condition* for statement *B*, if (and only if) the truth of *A* guarantees the truth of *B.* In math notation: $A \Rightarrow B$
- If statement *A* is necessary condition for statement *B,* then *B* is sufficient condition for statement *A.*
  - $\text{not}A \Rightarrow \text{not}B$ implies $B \Rightarrow A$
- If statement *A* is sufficient condition for statement *B,* then *B* is necessary condition for statement *A.*
  - $A \Rightarrow B$ implies $\text{not}B \Rightarrow \text{not}A$
- In optimization we would like to have easy to check conditions that tell us if a candidate point
  - is a local optimum (sufficient condition for optimality is sufficient)
  - is not an optimal condition (necessary condition is violated)

  Ideally we want conditions that are necessary and sufficient for local optimum (or even better for global)

Necessary and Sufficient Conditions: Examples

- Simple example: let $x \in R$ and $y = x^2$. Statement *A* "x *is positive*" and statement *B* "y *is positive*"
  - *A* is sufficient for *B*. Proof: *A* true $\Leftrightarrow\ x > 0 \Rightarrow x^2 > 0 \Rightarrow y > 0 \Leftrightarrow$ *B* true
  - *A* is not necessary for *B*. Proof by counter-example: $x = -1 \Rightarrow y = x^2 = 1$, so *B* is true and *A* is false
- Example for sets. Let $D_A, D_B \subset R^n$ and $D_C = D_A \cap D_B$
  - Statement *A*: $\boldsymbol{x} \in D_A$
  - Statement B: $\boldsymbol{x} \in D_B$
  - Statement *C*: $\boldsymbol{x} \in D_C$
  - *A* is necessary for *C*, *B* is necessary for *C*
  - *C* is sufficient for *A*, *C* is sufficient for *B*
  - (*A* and *B*) is both necessary and sufficient for *C*



Check Yourself

- Which functions are continuous, differentiable, continuous and differentiable?
- How is the directional derivative of a function defined? How is the partial derivative related to the directional derivative?
- What is the definition of the gradient and the Hessian of a function?

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Optimality conditions in smooth unconstrained problems

Aachener Verfahrenstechnik

RWTH AACHEN UNIVERSITY

Unconstrained Optimization

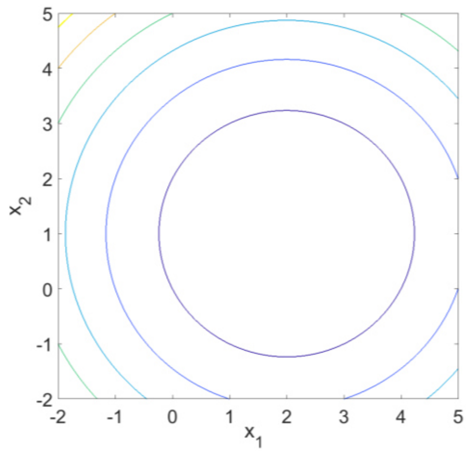
Unconstrained optimization problem:

Special case for which the feasible set $\Omega = R^n$

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

$\boldsymbol{x}^*$ is a local solution if $\boldsymbol{x}^* \in R^n$ and a neighborhood $N(\boldsymbol{x}^*)$ of $\boldsymbol{x}^*$ exists: $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \forall \boldsymbol{x} \in N(\boldsymbol{x}^*)$

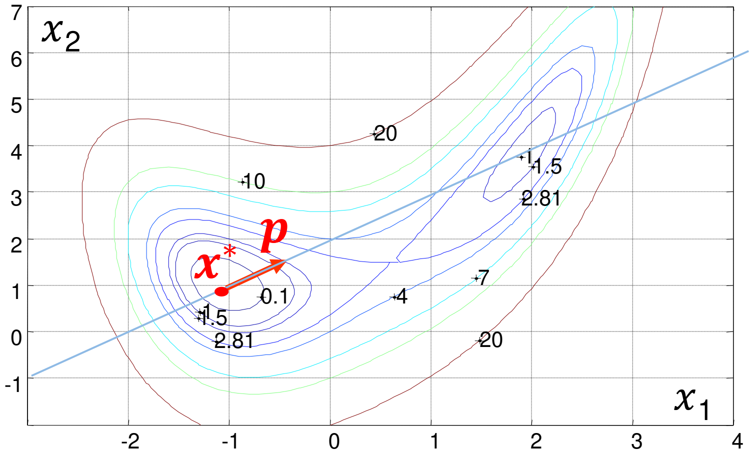
We want easy to check conditions
Necessary: if $\boldsymbol{x}^*$ is optimal then conditions is satisfied
Sufficient: if condition is satisfied then $\boldsymbol{x}^*$ is optimal
Ideally both necessary and sufficient!



First-Order Necessary Conditions

Theorem (First-Order Necessary Conditions):

Let f be continuously differentiable and let $\boldsymbol{x}^* \in R^n$ be a local minimizer of f, then $\boxed{\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}}$



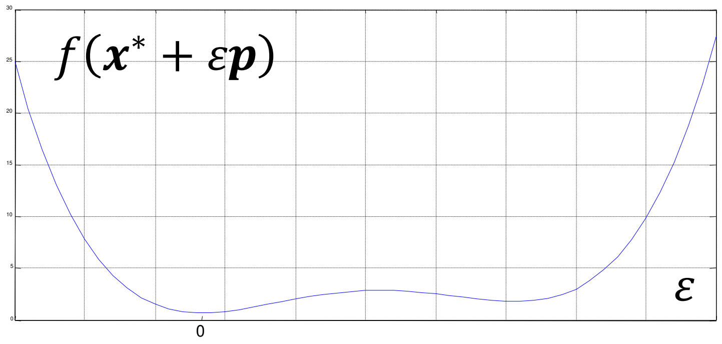
Proof:

As $\boldsymbol{x}^*$ is a local minimizer of f, for each $\boldsymbol{p} \in R^n$, there exists $\tau > 0$, such

that $f(\boldsymbol{x}^* + \varepsilon \boldsymbol{p}) \geq f(\boldsymbol{x}^*)\ \forall\ \varepsilon \in [0, \tau]$.

By the definition of the directional derivative:

$$\boldsymbol{\nabla}_{\boldsymbol{p}} f(\boldsymbol{x}^*) = \lim_{\varepsilon \to 0} \frac{f(\boldsymbol{x}^* + \varepsilon \boldsymbol{p}) - f(\boldsymbol{x}^*)}{\varepsilon} = \boldsymbol{\nabla} f(\boldsymbol{x}^*)^T \boldsymbol{p} \geq 0 \qquad (1)$$



The special choice, $\boldsymbol{p} = -\boldsymbol{\nabla} f(\boldsymbol{x}^*)$, leads to

$$\boldsymbol{\nabla} f(\boldsymbol{x}^*)^T \boldsymbol{p} = -\boldsymbol{\nabla} f(\boldsymbol{x}^*)^T \boldsymbol{\nabla} f(\boldsymbol{x}^*) = -\|\boldsymbol{\nabla} f(\boldsymbol{x}^*)\|^2 \leq 0 \ \ \text{(norm property)} \qquad (2)$$

(1) and (2) $\Rightarrow \boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$.

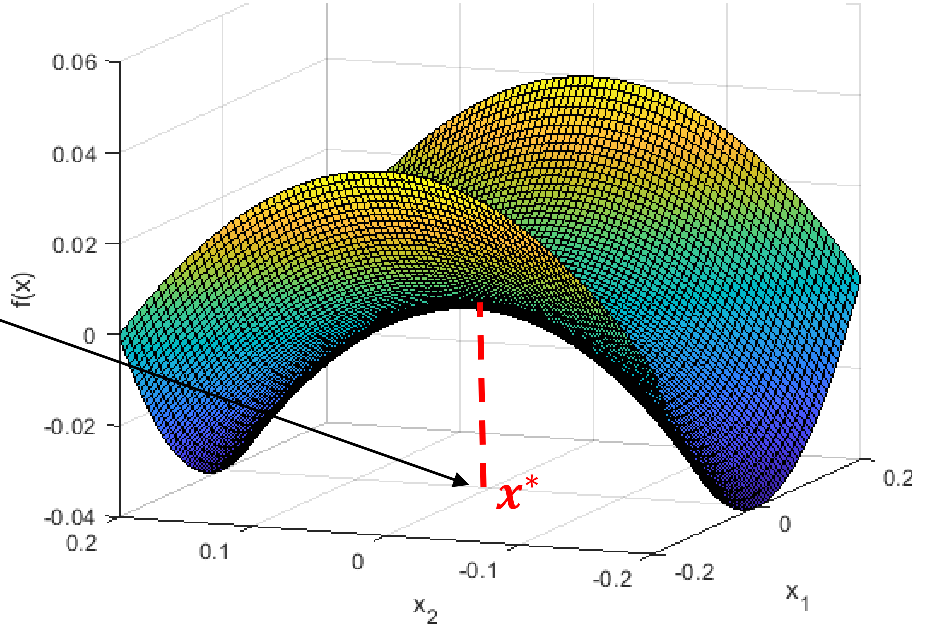
Stationary Points

- Let f be continuously differentiable and $\boldsymbol{x}^* \in R^n$. If $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$ holds, then $\boldsymbol{x}^*$ is called a stationary point of f.
- This condition is a **necessary**, but **not a sufficient** condition for a local minimum.
- Example: $f(x) = -x^2$ possesses its only stationary point at $x^* = 0$, as $\nabla f(x^*) = -2x^* = 0$. This point is, however, not a minimum but rather the unique global maximum.

Saddle Points

- A stationary point does not have to be a minimum or a maximum. Such a stationary point is called a saddle point.
- Example: the gradient of $f(\boldsymbol{x}) = x_1^2 - x_2^2$ is $\boldsymbol{\nabla} f(\boldsymbol{x}) = [2x_1, -2x_2]^T$. Thus, $\boldsymbol{x}^* = \boldsymbol{0}$ is its only stationary point. As f is positively curved in x_1-direction and negatively curved in x_2-direction, $\boldsymbol{x}^*$ is a saddle point.

saddle point



Second-Order Necessary Conditions

Theorem (Second-order necessary conditions):

Let f be twice continuously differentiable and let $\boldsymbol{x}^* \in R^n$ be a local minimizer of f, then

$$\boxed{\begin{array}{ll} 1. & \boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}, \\ 2. & \boldsymbol{\nabla}^2 f(\boldsymbol{x}^*) \text{ is positive semidefinite.} \end{array}}$$

These conditions are **only necessary and not sufficient**

- The only stationary point of $f(x) = x^3$ is $x^* = 0$, with $\nabla f(0) = 0, \nabla^2 f(0) = 0$. The above conditions are fulfilled. $x^* = 0$ is not a local minimum but rather a saddle point.
- The only stationary point of $f(x) = -x^4$ is $x^* = 0$, with $\nabla f(0) = 0, \nabla^2 f(0) = 0$. The above conditions are fulfilled. $x^* = 0$ is not a local minimum but rather a local maximum.

Second-Order Necessary Conditions: Informal Proof by Contradiction

Let f be twice continuously differentiable and let $\boldsymbol{x}^* \in R^n$ be a local minimizer of f.

Assume $\boldsymbol{\nabla}^2 f(\boldsymbol{x}^*)$ is not positive semidefinite.

Thus, $\exists \boldsymbol{p} \in R^n$: $\boldsymbol{p}^{\boldsymbol{T}} \boldsymbol{\nabla}^2 f(\boldsymbol{x}^*) \boldsymbol{p} < 0$

Taylor expansion at $\boldsymbol{x}^*$ gives

$$f(\boldsymbol{x}^* + \epsilon \boldsymbol{p}) = f(\boldsymbol{x}^*) + \epsilon \boldsymbol{\nabla} f(\boldsymbol{x}^*)^T \boldsymbol{p} + \frac{1}{2} \epsilon^2 \boldsymbol{p}^T \boldsymbol{\nabla}^2 f(\boldsymbol{x}^*) \boldsymbol{p} + O(\epsilon^3).$$

$\boldsymbol{x}^*$ is a local minimum and thus by first-order necessary condition

$$\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$$

For sufficiently small ϵ, $O(\epsilon^2)$ dominates over $O(\epsilon^3)$. Since $\boldsymbol{p}^{\boldsymbol{T}} \boldsymbol{\nabla}^2 f(\boldsymbol{x}^*) \boldsymbol{p} < 0$

$$f(\boldsymbol{x}^* + \epsilon \boldsymbol{p}) < f(\boldsymbol{x}^*)$$

$\boldsymbol{x}^*$ is not a local minimum

Sufficient Optimality Conditions

Theorem (sufficient optimality conditions):

Let f be twice continuously differentiable and let $\boldsymbol{x}^* \in R^n$, if

1. $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$,
2. $\boldsymbol{\nabla}^2 f(\boldsymbol{x}^*)$ is positive definite.

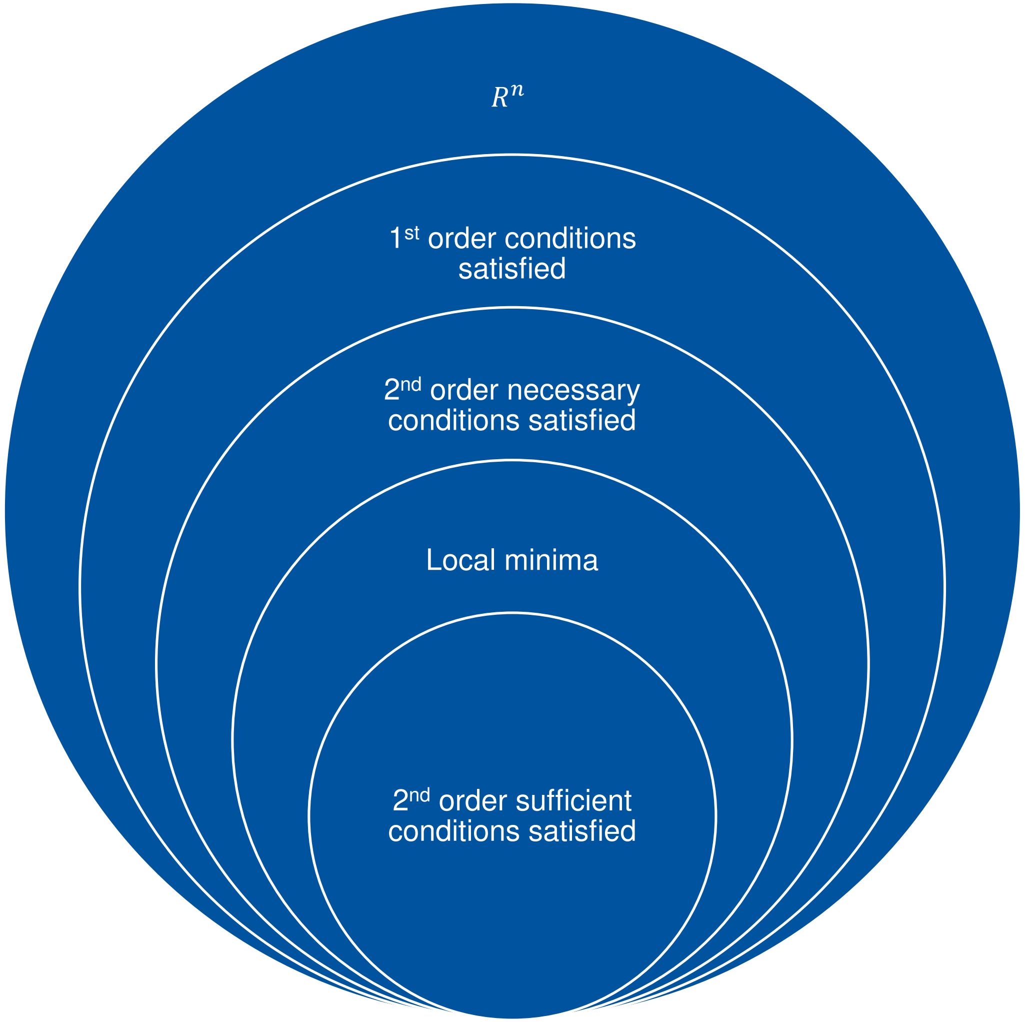
then $\boldsymbol{x}^*$ is a strict local minimizer of f.

Proof similar to second order necessary

Remark

- $f(x) = x^4$ attains at $x^* = 0$ its (unique) strict global minimum. Further, $\nabla f(0) = 0$ and $\nabla^2 f(0) = 0$ hold, thus the 2nd condition in the above theorem is violated.
- Hence, the conditions mentioned in the theorem are **sufficient but not necessary**

Optimality Conditions for Smooth Problems



- Optimality conditions are at a point, not for the whole R^n
- All the sets shown are true subsets
- The first-order necessary conditions exclude non-stationary points
- The second-order necessary conditions exclude some saddle points and some local maxima, but not all

Check Yourself

- What is a stationary point? Are there different kinds of stationary points?
- What are the first-order necessary conditions of optimality for smooth unconstrained problems?
- What are the second-order necessary conditions of optimality for smooth unconstrained problems?
- What are there the second-order sufficient conditions for smooth unconstrained problems?
- Are there any necessary and sufficient optimality conditions? In general vs for specific classes

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Examples for optimality conditions

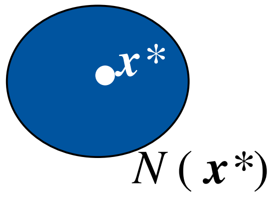
Smooth Unconstrained Optimization (Recap)

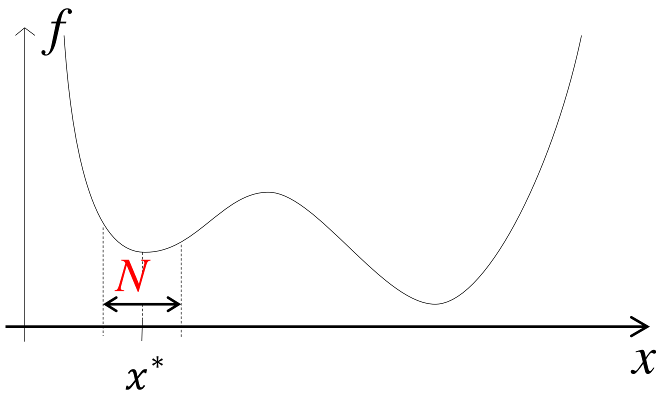
Unconstrained optimization problem: $\min_{\boldsymbol{x} \in R^n} f(\boldsymbol{x})$

Special case for which the feasible set $\Omega = R^n$

Definition:

$\boldsymbol{x}^*$ is a local solution if $\exists N(\boldsymbol{x}^*)$: $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \forall \boldsymbol{x} \in N(\boldsymbol{x}^*)$





Optimality Conditions:

1st-Order Necessary : *If* $\boldsymbol{x}^*$ *is a local minimum then* $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$

2nd-Order Necessary: *If* $\boldsymbol{x}^*$ *is a local minimum then* $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$ *and* $\boldsymbol{H}(\boldsymbol{x}^*)$ *is positive semi definite*

2nd-Order Sufficient : *If* $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$ *and* $\boldsymbol{H}(\boldsymbol{x}^*)$ *is positive definite then* $\boldsymbol{x}^*$ *is a local minimum*

Example for Application of Necessary Condition (1)

Problem

Find all stationary points of the function

$$f(\boldsymbol{x}) = x_1^4 + x_1^2(1 - 2x_2) + 2x_2^2 - 2x_1x_2 + 4.5x_1 - 4x_2 + 4$$

and use these to determine all minima.

Solution

The stationary points $\boldsymbol{x}^*$ are defined by the condition, $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$

$$\boldsymbol{\nabla} f(\boldsymbol{x}) = \begin{bmatrix} \left.\dfrac{\partial f}{\partial x_1}\right|_{\boldsymbol{x}} \\ \left.\dfrac{\partial f}{\partial x_2}\right|_{\boldsymbol{x}} \end{bmatrix} = \begin{bmatrix} 4x_1^3 + 2x_1(1 - 2x_2) - 2x_2 + 4.5 \\ -2x_1^2 + 4x_2 - 2x_1 - 4 \end{bmatrix} = \boldsymbol{0}$$

$$\Rightarrow \begin{cases} 4x_1^3 + 2x_1(1 - 2x_2) - 2x_2 + 4.5 = 0 \\ -2x_1^2 + 4x_2 - 2x_1 - 4 = 0 \end{cases}$$

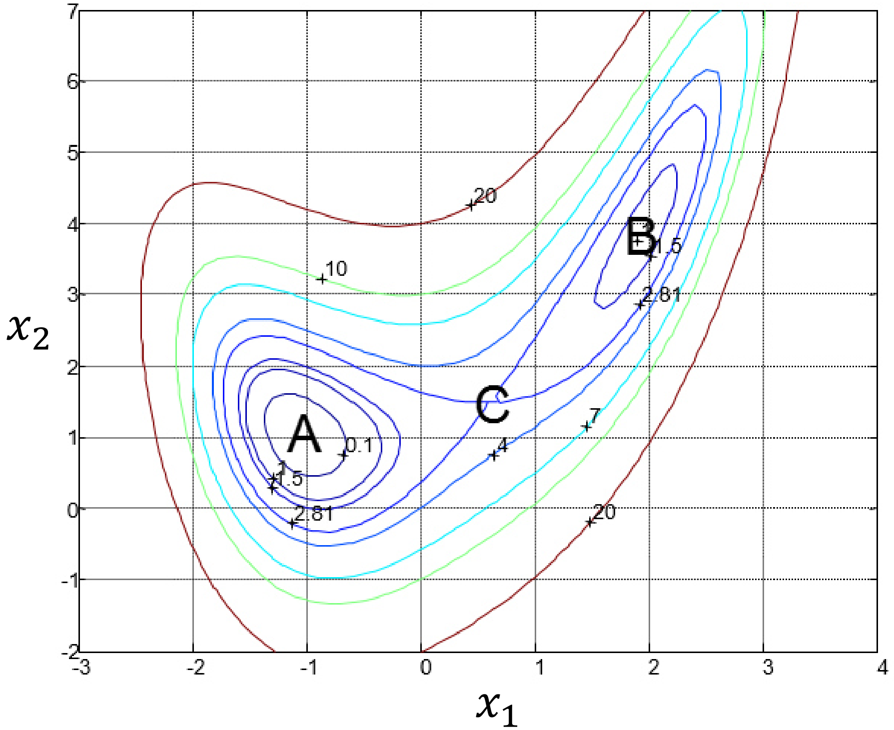
Example for Application of Necessary Condition (2)

- Solving the system of equations results in the stationary points

$$\boldsymbol{A}(-1.053, 0.9855),$$

$$\mathbf{B}(1.941, 3.854),$$

$$\boldsymbol{C}(0.6117, 1.4929).$$



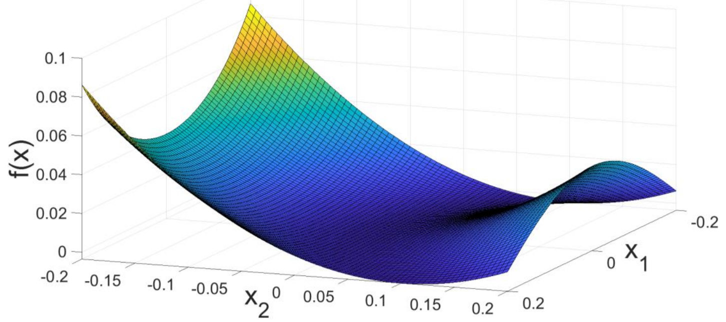
- To classify the stationary points, we investigate the definiteness of the Hessian $\boldsymbol{H}(\boldsymbol{x})$

$$\boldsymbol{H}(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 + 2(1 - 2x_2) & -4x_1 - 2 \\ -4x_1 - 2 & 4 \end{bmatrix}$$

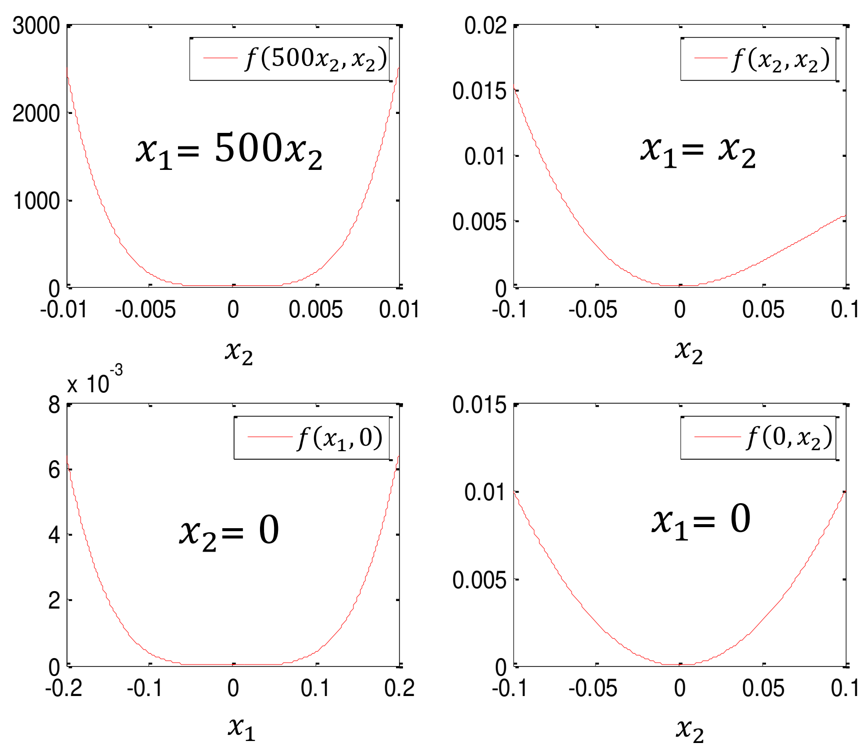
- At $\boldsymbol{A}$ and $\mathbf{B}$ all eigenvalues are positive. By 2nd order sufficient conditions $\boldsymbol{A}$ and $\mathbf{B}$ are local minima. (A is indeed the unique global minimizer)
  At $\boldsymbol{C}$, the Hessian has one positive and one negative eigenvalue. The 2nd order necessary conditions are violated and thus $\boldsymbol{C}$ is not a local minimum ($\boldsymbol{C}$ is indeed a saddle point)

A Funky Function [1] (1)

$$f(\boldsymbol{x}) = (x_2 - x_1^2)(x_2 - 4x_1^2)$$



- $\boldsymbol{\nabla} f(\boldsymbol{x}) = \begin{bmatrix} -10x_1x_2 + 16x_1^3 \\ 2x_2 - 5x_1^2 \end{bmatrix}$, $\boldsymbol{\nabla} f(\boldsymbol{0}) = \boldsymbol{0}$
- $\boldsymbol{H}(\boldsymbol{x}) = \begin{bmatrix} -10x_2 + 48x_1^2 & -10x_1 \\ -10x_1 & 2 \end{bmatrix}$, $\boldsymbol{H}(\boldsymbol{0}) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$
- Necessary conditions (1st and 2nd) satisfied
- Sufficient conditions not satisfied

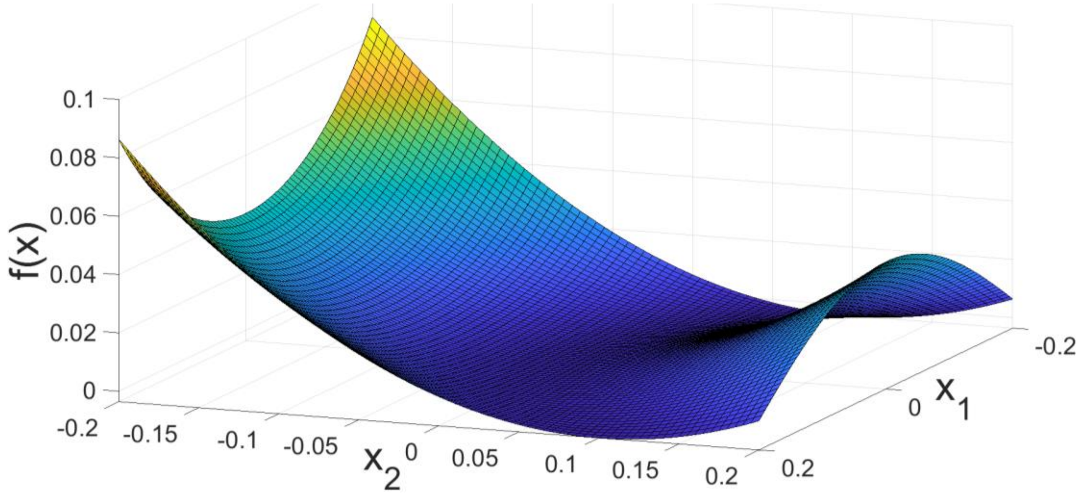


- **0** is local minimum w.r.t. every line through it
- **0** is not a local minimum of f
- How can that be?

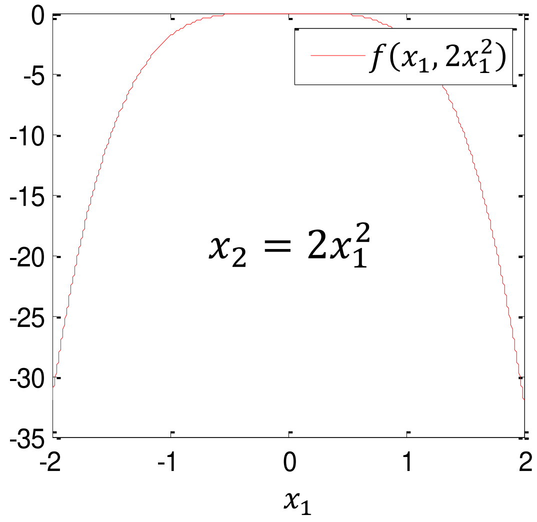
[1] Bertsekas D.P., Nonlinear Programming – Third Edition, Athena Scientific, 2016.

A Funky Function [1] (2)

$$f(\boldsymbol{x}) = (x_2 - x_1^2)(x_2 - 4x_1^2)$$



- Take $x_2 = 2x_1^2$. We have $f(\boldsymbol{x}) = f(x_1, 2x_1^2) = -2x_1^4$ and clearly (0, 0) is not a minimum along that curve.



- Trick: it is not a minimum for a curve that passes through it

[1] Bertsekas D.P., Nonlinear Programming – Third Edition, Athena Scientific, 2016.

Applied Numerical Optimization

Prof. Alexander Mitsos, Ph.D.

Convexity in optimization

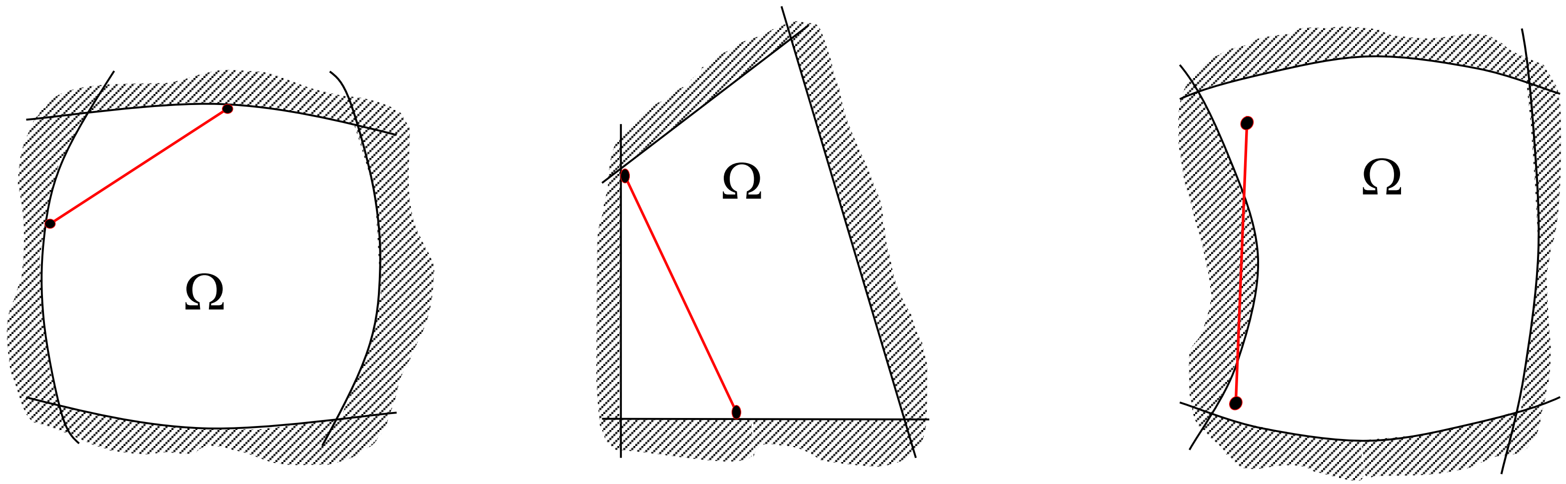
Aachener Verfahrenstechnik

RWTH AACHEN UNIVERSITY

Convexity of a Set

Definition (convex set):

- A set $\Omega \subseteq R^n$ is convex, if $\forall\, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega$ and $\forall\, \alpha \in [0,1]$, $\alpha \boldsymbol{x}_1 + (1\text{-}\alpha)\boldsymbol{x}_2 \in \Omega$



convex

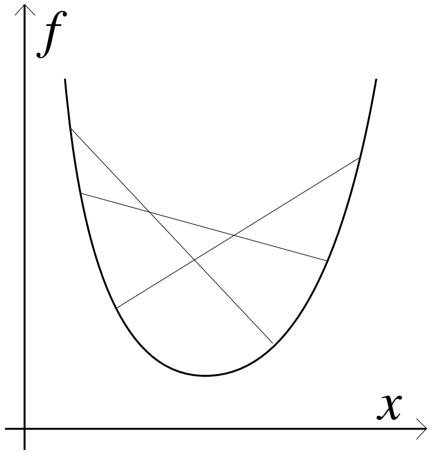
nonconvex

- The constraints define the feasible set $\Omega = \{\boldsymbol{x} \in D \mid c_i(\boldsymbol{x}) \leq 0\ \forall\, i \in I,\, c_i(\boldsymbol{x}) = 0\ \forall\, i \in E\}$
- Convexity of Ω makes a big difference in theoretical properties and in numerical solution
- A set is either convex or nonconvex (no "concave sets" please)

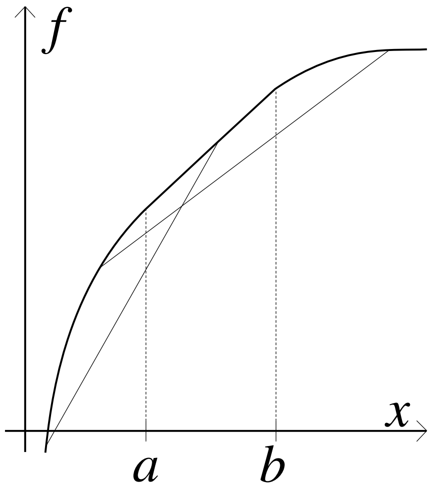
Convexity of a Function

Definition (convex function): assume D is convex

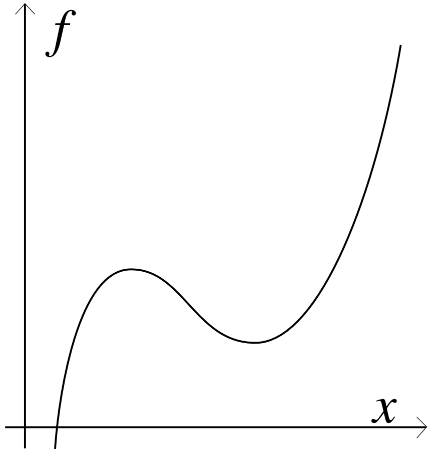
- A function f is convex on D, if $\forall\ \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathrm{D}$, $\forall\ \alpha \in [0,1]$: $f(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) \leq \alpha f(\boldsymbol{x}_1) + (1\text{-}\alpha) f(\boldsymbol{x}_2)$
- f is strictly convex on D, if $\forall\ \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathrm{D}$, $\forall\ \alpha \in (0,1)$: $f(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) < \alpha f(\boldsymbol{x}_1) + (1\text{-}\alpha) f(\boldsymbol{x}_2)$
- f is (strictly) concave on D, if $(-f)$ is (strictly) convex
- Affine linear functions are both convex and concave, but not strict



strictly convex



concave, but not strictly



neither convex, nor concave

Criteria for Convexity (1)

Definition (positive definite):

- A symmetric ($n \times n$)-matrix $\boldsymbol{A}$ is called positive definite, if $\boldsymbol{p}^T\boldsymbol{A}\boldsymbol{p} > 0 \ \forall\ \boldsymbol{p} \in R^n, \boldsymbol{p} \neq \boldsymbol{0}$.
- A symmetric ($n \times n$)-matrix $\boldsymbol{A}$ is called positive semi-definite, if $\boldsymbol{p}^T\boldsymbol{A}\boldsymbol{p} \geq 0 \ \forall\ \boldsymbol{p} \in R^n$.
- If ($-\boldsymbol{A}$) is positive (semi-)definite, then $\boldsymbol{A}$ is called negative (semi-)definite.

Theorem:

A symmetric ($n \times n$)-matrix $\boldsymbol{A}$ is positive definite, if $\lambda_k > 0, \forall k \in \{1, \dots, n\}$ where λ_k represent the eigenvalues of $\boldsymbol{A}$, i.e., the solutions of $\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$.

Similarly positive semi-definite if $\lambda_k \geq 0, \forall k \in \{1, \dots, n\}$

Theorem (convexity under differentiability):

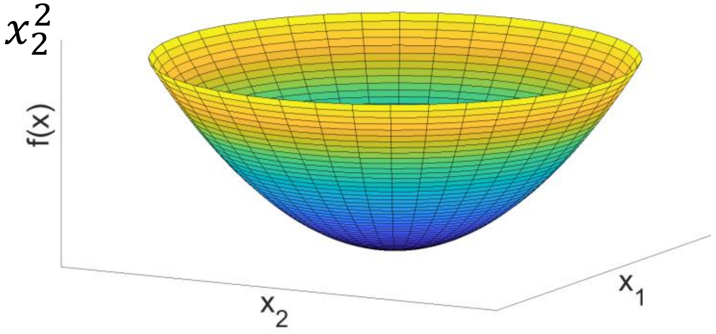
- f is convex, iff the Hessian $\boldsymbol{H}(\boldsymbol{x})$ is positive semi-definite $\forall \boldsymbol{x} \in D$.
- If $\boldsymbol{H}(\boldsymbol{x})$ is positive definite $\forall \boldsymbol{x} \in D$, then f is strictly convex.

Criteria for Convexity (2)

	if $\forall\, \boldsymbol{x} \in D$		
f is	$\boldsymbol{H}(\boldsymbol{x})$ is	all λ_k are	$\forall\, \boldsymbol{p} \in R^n$, $\boldsymbol{p}^T \boldsymbol{H}(\boldsymbol{x}) \boldsymbol{p}$ is
strictly convex	positive definite	> 0	> 0
convex	positive semi-definite	≥ 0	≥ 0
strictly concave	negative definite	< 0	< 0
concave	negative semi-definite	≤ 0	≤ 0
neither convex, nor concave	-	≥ 0 or ≤ 0	≥ 0 or ≤ 0
	Definiteness of the Hessian	Sign of the eigenvalues	Sign of the quadratic form

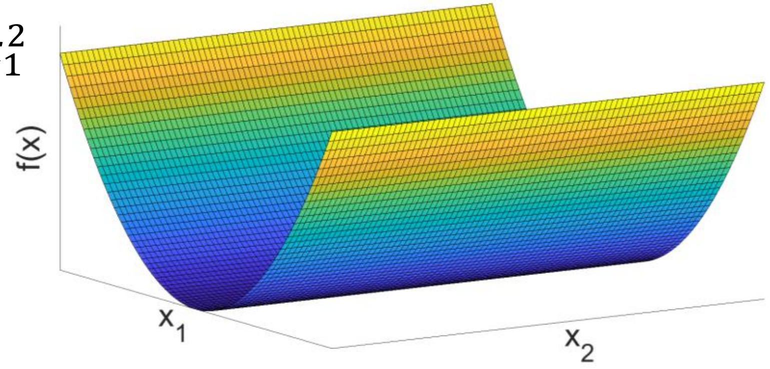
Geometric Illustration of Convexity for $x \in R^2$

$f(\boldsymbol{x}) = x_1^2 + x_2^2$



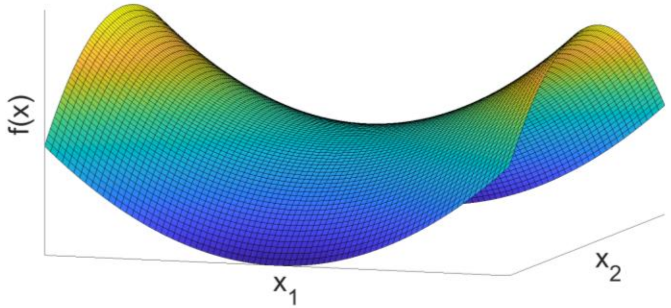
$\lambda_{1,2} > 0$

$f(\boldsymbol{x}) = x_1^2$



$\lambda_1 > 0, \lambda_2 = 0$

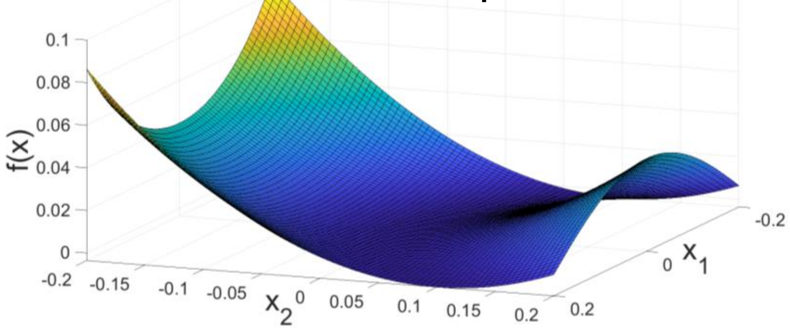
$f(\boldsymbol{x}) = x_1^2 - x_2^2$



$\lambda_1 > 0, \lambda_2 < 0$

$f(\boldsymbol{x}) = (x_2 - x_1^2)(x_2 - 4x_1^2)$

Sign of eigenvalues depend on $\boldsymbol{x}$



Aachener Verfahrenstechnik

RWTH AACHEN UNIVERSITY

Convex Optimization Problems

Definition (convex optimization problem):

- The optimization problem $\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$ is convex, if the objective function f is convex and if the feasible set Ω is convex.
- If D is a convex set, $c_i \forall\ i \in I$ are convex on D and $c_i \forall\ i \in E$ are linear then $\Omega = \{\boldsymbol{x} \in \mathrm{D} \mid c_i(\boldsymbol{x}) \leq 0\ \forall\ i \in I, c_i(\boldsymbol{x}) = 0\ \forall\ i \in E\}$ is convex.
- Apart from trivial exceptions:
  - Ω is nonconvex, if any $c_i \forall\ i \in E$ is a nonlinear function.
  - Ω is nonconvex, if any $c_i \forall\ i \in I$ is nonconvex on D.
  - Extra challenge: find such exceptions
- "… in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity." R. Tyrrell Rockafellar in SIAM Review, 1993

Optimality Conditions for Smooth Convex Problems

- Let f be twice continuously differentiable and convex.

- Since f is smooth and convex, $\boldsymbol{\nabla}^2 f(\boldsymbol{x})$ is positive semi-definite for all $\boldsymbol{x}$.

- If $\boldsymbol{x}^* \in R^n$ is a local solution point, then it also is a global solution point.
  - Proof argument: Convexity implies that first derivative does not decrease when we move away from $\boldsymbol{x}^*$. So we cannot find other distinct local minimum.

- A point $\boldsymbol{x}^* \in R^n$ is a **global solution** point if and only if $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$
  - Convexity implies $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \left(\boldsymbol{\nabla} f(\boldsymbol{x}^*)\right)^T (\boldsymbol{x} - \boldsymbol{x}^*)$
  - With stationarity $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$

- Simply said:
  - The first order optimality condition is both necessary and sufficient.
  - A stationary point is equivalent to a local solution point and a global solution point.
  - In constrained problems a similar property exists: convexity implies that the first-order optimality conditions are both necessary and sufficient

Check Yourself

- When is an optimization problem convex?
- How can we check the convexity of a smooth unconstrained optimization problem? (at least in principle)
- Are there any necessary and sufficient optimality conditions? In general vs for specific classes