# Chapter 0
# Organization and Introduction

*Convex Optimization for Computer Vision*
SS 2018

Michael Moeller
Visual Scene Analysis
Department of Computer Science
and Electrical Engineering
University of Siegen

# Organizational Stuff

# Requirements, or "is this something for me?"

### Necessary

- Interest in mathematical theory
- Solid background in analysis and linear algebra
- Numerics (Matlab)

### Nice to know

- Image processing and computer vision
- Optimization or inverse problems lectures offered by math department

# Formalities

## Exercises

- Exercise sheets covering the content of the lecture will go online every Friday
- Exercises contain theoretical as well as programming problems
- You have one week for the exercise sheets and will turn in your solutions on Friday in the mailbox outside of H-A 7106.
- You may work on the exercises in groups of two
- The solutions will be discussed in the exercises on Monday
- **Reaching at least** 50% **of the total exercise points is required for being admitted to the final exam**
- If solutions have obviously been copied, both groups will get 0 points

# Formalities

### Questions within the lecture

The more we discuss in the lecture, the more interesting the course will be! Please don't be shy to say something!

### Examination

- Depending on the number of attendees, the final exam will be either oral or written.
- This lecture is worth 10 credits.

# Formalities

## Miscellaneous

- Hartmut's office: H-A 7116
- My office: H-A 7106
- Office hours: Please write an email.
- Lecture: Starts at quarter past. Short break in between.
- Course website: `http://www.vsa.informatik.uni-siegen.de/en/convex-optimization-computer-vision-0`
- To access the course material: username: "student", password "100%brain"

# What is it good for?

## General setup

An optimization problem aims at finding some $\hat{u}$ such that

$$\hat{u} \in \arg\min_u E(u) \qquad \text{s.t. } \; u \in C,$$

for an *energy function* $E : S \to \mathbb{R}$. In this course we will limit ourselves to $S = \mathbb{R}^n$.

One often distinguishes *continuous* and *discrete* optimization based on whether $u$ can take a whole range of values, e.g. $u_i \in [0, 1]$ or merely discrete values, e.g. $u_i \in \{0, 1\}$.

In this lecture we will consider continuous optimization methods only. More specifically, we will require $C$ and $E$ to be *convex* - a terminology to be discussed in the first chapter.

Very nice big picture by S.J. Wright: *Continuous Optimization (Nonlinear and Linear Programming)* from whom I will steal some examples.

# Example 1: Profit maximization

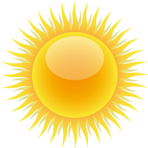
A company wants to maximize its profit under certain contraints given by the availability of resources.

- A company has two products.
- Producing the amount x of product 1 requires
  - using machine A for $5x$ units of time,
  - using machine B for $2x$ units of time.
- Producing the amount x of product 2 requires
  - using machine A for $1x$ units of time,
  - using machine B for $4x$ units of time.
- Product 1 sells for twice as much as product 2.
- Machine A is available for 160 units of time.
- Machine B is available for 240 units of time.

How much of product 1 and 2 should the company produce in order to maximize its profit?

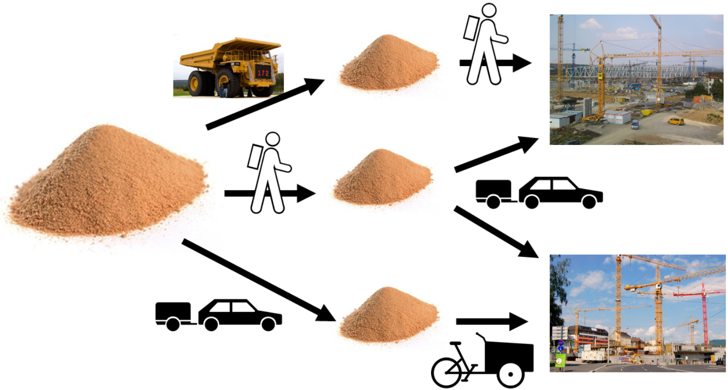
# Example 2: Weather forecast





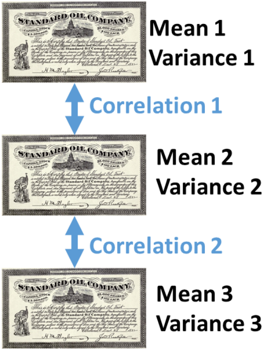
- The ”weather” is described by a set of differential equations.
- Predicting the weather means evolving these equations.
- One needs to determine the state the weather is currently in.
- The latter requires fitting measurements of the atmospheric state to the model, and includes additional information (e.g. the previous model that was fitted).
- The above fitting can be phrased as an optimization problem.

# Example 3: Transport problems




- Given: Some demand for certain goods.
- Given: Sources of these goods.
- Given: Possible ways of transportation between notes in the graph, their respective costs, and possible constraints.
- Desired: The cheapest way of satisfying the demand.

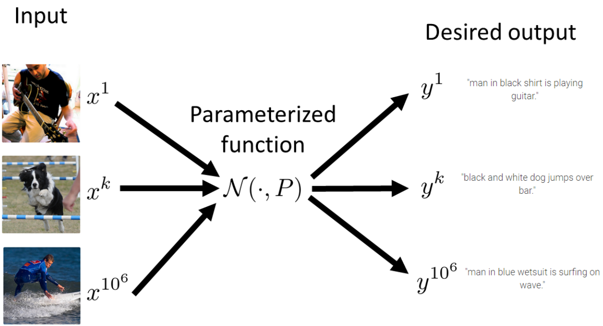
# Example 4: Portfolio optimization




Balance expected mean and risk

Find an investment that maximizes the expected profit for given suitable variance/risk.

# Example 5: Machine Learning




Example taken from `http://cs.stanford.edu/people/karpathy/deepimagesent/`.

Optimization problem: $\min_P \sum_k d(\mathcal{N}(x^k, P), y^k)$.

The inputs and outputs can be anything!

`https://www.youtube.com/watch?v=0FW99AQmMc8`
`https://www.youtube.com/watch?v=fa5QGremQf8`
`https://www.youtube.com/watch?v=5aogzAUPilE`

# Optimization vs. Convex Optimization

# Optimization algorithms

We have seen several examples for continuous optimization problems of the form

$$\hat{u} \in \arg\min_u E(u) \qquad \text{s.t.} \quad u \in C \tag{1}$$

In general it will be impossible to determine $\hat{u}$ directly. Instead we will construct sequences of iterates $u^k$ such that $u^k \to \hat{u}$.

There is no general efficient way to construct such iterates for a generic problem like (1).

Most often, the more structure of a specific optimization problem we can exploit the faster our algorithm we be!

# Classification into particular areas

The problem

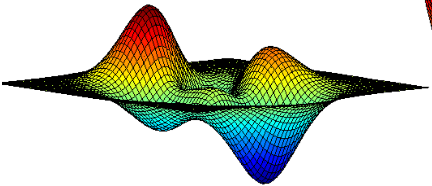
$$\hat{u} = \arg\min_u E(u) \qquad \text{s.t.} \quad u \in C$$

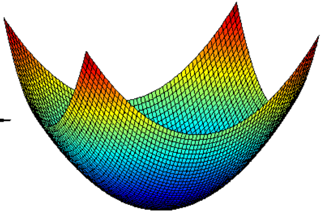
is particularly structured if $E$ is linear in $u$ and the set $C$ can be described by affine linear inequality constraints. In this case we speak of *linear programming*. In this lecture we will consider a more general class of problems.

In *quadratic programming* the set $C$ can still be described by affine linear inequality constraints, but $E$ is allowed to contain additional terms like $\sum_{i,j} A_{i,j} u_i u_j$. For a *positive semi-definite matrix* $A$, quadratic programs are convex.

A particularly important classification of problems in computer vision are *nonsmooth problems* for which $E$ is not *continuously differentiable*. Moreover, computer vision problems are often *constrained problems* in which $C \subsetneq \mathbb{R}^n$.

# Why dealing with convex optimization only?





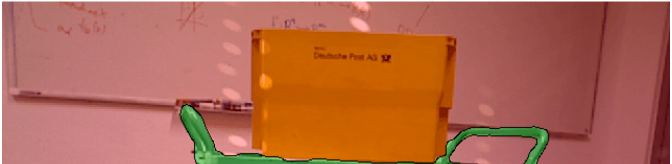
**Practical aspects:**
→ Almost any attempt to minimize a nonconvex function can get stuck in a local minimum.
→ Ambiguity: Modeling problem or bad local minimum?

**A basis to start from:**
→ Several techniques from convex optimization can also be applied to nonconvex optimization
→ Several techniques from nonconvex optimization iteratively solve convex problems (e.g. by linearizations or majorizations).

# Convex Optimization in CV

# Denoising

# Segmentation

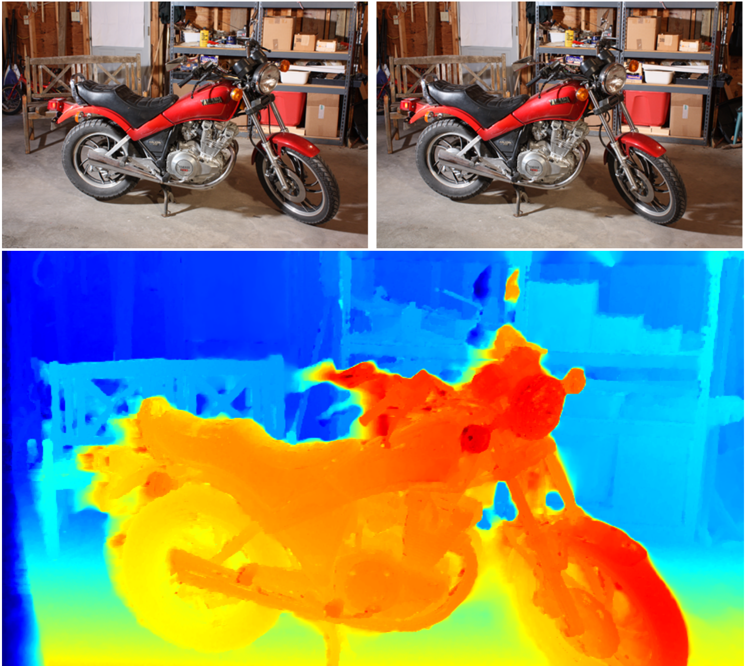
# Inpainting


[1]

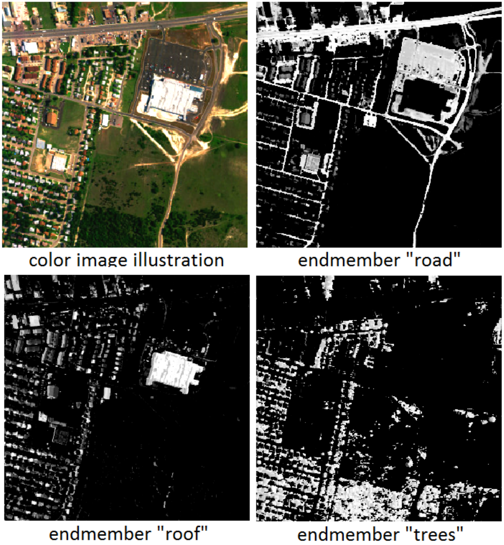
$$\min_{u, u|_{\partial\Omega} = f|_{\partial\Omega}} \int_\Omega |\nabla u(x)| \, dx$$

[1] https://www.theguardian.com/science/gallery/2014/mar/31/national-science-photography-competition-in-pictures

# Stereo Matching



# Unmixing




$$\min_{u} \|Au - f\|_2^2 + \alpha \|u\|_1$$

# Image fusion

+

=

$$u^{k+1} \in \arg\min_u \|Bu - f\|_2^2 + \alpha \int_\Omega |\nabla u(x)| - \langle u(x), p^k(x)\rangle \, dx$$

# Image fusion

Based on $u^{k+1} \in \arg\min_{u} \|u-f\|_2^2 + \alpha \int_\Omega |\nabla u(x)| - \langle u(x), p^k(x) \rangle \, dx$

# Image deblurring

$$\arg\min_u \|k * u - f\|_2^2 + \alpha \int_\Omega |\nabla u(x)| \, dx$$

# Discrete setting - motivating a mathematical understanding

All previous examples were formulated in a spatially continuous setting, representing the unknown as a function, and the energy being a functional on a suitable function space.

In this lecture we will only discuss finite dimensional optimization, i.e. how can we minimize and energy function

$$E : C \subset \mathbb{R}^d \to \mathbb{R}$$

for a convex set $C$.

To be able to develop and understand general optimization techniques, we first have to learn/remind ourselves how to describe some properties of sets, functions, and matrices mathematically. This will be chapter 1!

# Overview

# Lecture overview

## Chapter 1: Mathematical basics and convex analysis

Basics of multivariable calculus and linear algebra:

- Open, closed, bounded and compact sets
- Continuity of functions $f : \mathbb{R}^n \to \mathbb{R}^m$
- Differentiability of functions $f : \mathbb{R}^n \to \mathbb{R}^m$, chain rule
- Eigenvectors, semi-definiteness, induced norms
- Representation of arbitrary linear operators as matrices

Basics of convex analysis:

- Convex sets
- Convex extended real valued functions in $\mathbb{R}^n$
- Existence of minimizers
- Optimality conditions and subdifferential calculus

Goal: Everyone knows all necessary tools to follow the lecture!

# Lecture overview

**Chapter 2: Gradient based methods**

Optimization algorithms based on (generalized) gradient methods

- Gradient descent
- Gradient projection
- Proximal gradient method
- Convergence analysis using average operators

Goal: Establish basic minimization strategies based on energy descent methods most suitable for (partly) smooth energy functions.

# Lecture overview

## Chapter 3: Convex conjugation and duality

- Primal and dual formulation of a problem
- Convex conjugate
- Saddle point problems
- Optimality conditions

Goal: Increase the number of tools to reformulate and analyze more complex convex minimization problems.

# Lecture overview

**Chapter 4: Primal-dual optimization schemes**

- Primal-dual hybrid gradient method
- Alternating directions method of multipliers
- Convergence analysis based a reduction to the proximal point algorithm
- Stopping criteria + step size selection

Goal: Learn about state-of-the-art first order non-smooth optimization methods.

These are the algorithms used in most publications on variational method in imaging and computer vision!