Chapter 0
Organization and Introduction

Convex Optimization for Computer Vision and Machine Learning
WS 2017/2018

Virginia Estellers, Emanuel Laude
Computer Vision Group
Faculty of Informatics
Technical University of Munich

Organization

Organization

Weekly: 2-hour lecture (Virginia), 2-hour tutorial (Emanuel)

Lecture: Starts at quarter past. Short break in between

Course material
`https://vision.in.tum.de/teaching/ws2017/in2330`
Lectures based on the course created by M. Moeller in 2016

Office hours: please write us an email

- Virginia's office 02.09.037
- Emanuel's office 02.09.039

Assessment: weekly exercise + written/oral final exam

Exercises

Exercise sheets posted every week

Theoretical and programming problems

One week for each sheet, solutions discussed in class

Exercises in groups of 1 or 2. Copied solutions get 0 points

Exercise points +0.5(up to) to your final exam (above 5.0)

Interactions

Please don't be shy to ask questions

- They make the course more interesting
- They adapt the content to your background and interests

Please don't be shy to email us

- with suggestions or questions about blurry topics
- we will clarify topics that you found confusing
- we will adapt the exercises to help you understand them

Course Requirements

Necessary

- Basic background in Analysis
- Background in linear Algebra
- Basic Numerical Programming (Matlab)

Useful

- Image processing, computer vision, machine learning
  Computer Vision I and II: Variational Methods, Multiple View Geometry, Machine Learning for Robotics and Computer Vision
- Numerical Optimization
  Numerisches Programmieren

Why Convex Optimization

Optimization Problems

Given $E : S \subset \mathbb{R}^n \to \mathbb{R}$, we want to find $\hat{u}$ such that

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

where

u is the optimization variable

C is the constraint set

E is the objective or energy function

We can only solve[1] (1) for a subset of problems

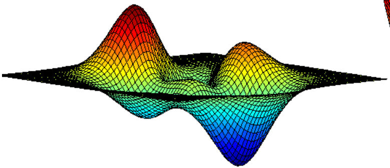
- least-squares problem $\min_u \|Au - b\|^2 \to u = (A^T A)^{-1} A^T b$
- linear program: no analytic solution, simplex algorithm
- convex problem: guarantees for existence/uniqueness solution

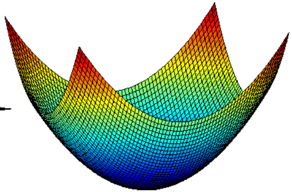
[1] no restriction to local minima or exhaustive search strategies

Why Convex Optimization

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

where C is a *convex set* and E is a *convex function*





- converge to local minima of nonconvex functions raises the question whether the model or the minimum are wrong
- sequential convex optimization of nonconvex problems (linearization or majorization)

Optimization algorithms

To construct iterates efficiently, we exploit the structure of E/C

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

In this course, the energy function and the constraints define a computer vision or machine learning model

- robust to noise and outliers in the data
- regular: generalize well, do not overfit the training data
- sparse: explain the data with as few variables as possible

This usually results in large nonsmooth optimization problems

Denoising

$$\min_u \|u - f\|_1 + \alpha \int_\Omega |\nabla u(x)|\, dx$$

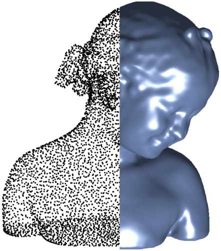
Denoising

$$\min_u \|u - f\|_1 + \alpha \int_\Omega |\nabla u(x)| \, dx$$

Image deblurring

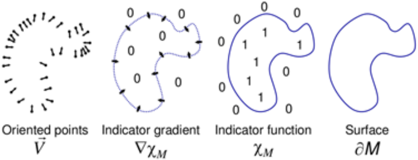
$$\arg\min_u \|k * u - f\|_2^2 + \alpha \int_\Omega |\nabla u(x)|\, dx$$

Surface Reconstruction



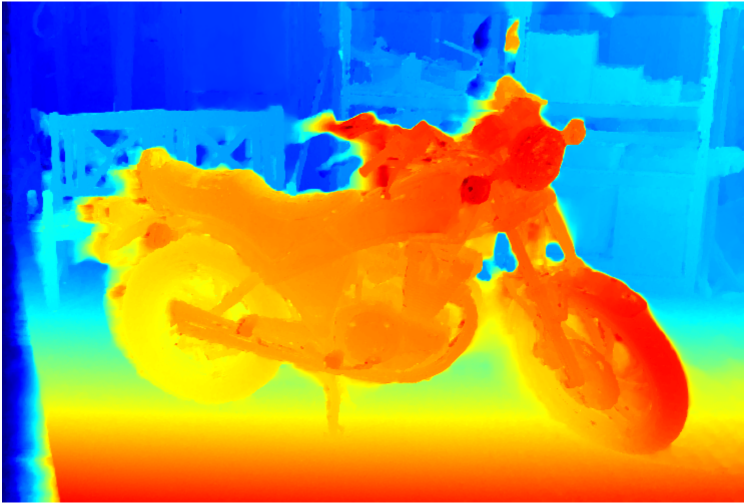
Reconstruct an implicit surface $\partial M = \{x \colon \chi(x) = 0\}$ from an oriented point cloud $\{(x_i, V_i)\}_{i=0..N}$

$$\min_u \sum_{i=0}^{N} \alpha \chi^2(x_i) + \beta \| |V_i - \nabla \chi(x_i)\|^2 + \int_\Omega \|Hu(x)\|^2 \, dx$$



Stereo Matching

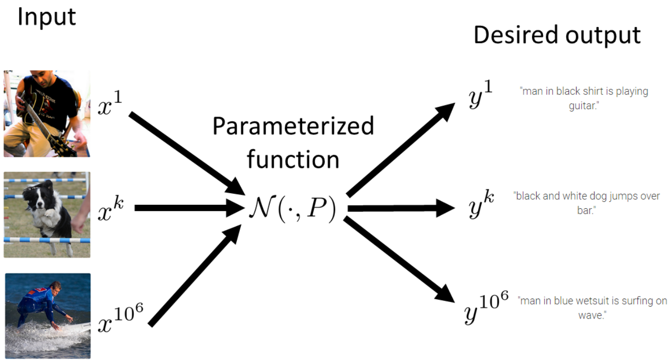
Stereo Matching



Convexification of $\min_{v} \int_{\Omega} |f^1(x + v(x)) - f^2(x)| + \alpha |\nabla v(x)| \, dx$

Machine Learning Framework

Find the parameters (weights) P of the model $\mathcal{N}(\cdot, P)$ that explains the training examples $\{(x_i, y_i)\}_{i=0}^{N}$



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

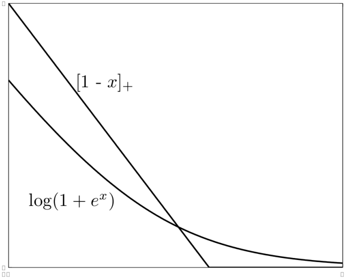
Optimization problem: $\min_P \sum_i \mathcal{L}(\mathcal{N}(x_i, P), y_i)$.

Common Machine Learning Loss Functions

Linear regression $\min_w \|Xw - y\|^2$

Binary labels

- SVM loss: $\min_w \sum_k [1 - y_k x_k^T w]_+$
- Binary logistic loss: $\min_w \sum_k \log(1 + \exp(-y_k x_k^T w))$



Example taken from

`https://people.eecs.berkeley.edu/~jordan/courses/294-fall09/lectures/optimization/slides.pdf`

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
- Linear operators in matrix form, eigenvectors, semi-definiteness

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
- Subgradient descent
- Convergence analysis

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

- Concept: Averaged operators
- Primal-dual hybrid gradient method
- Proximal point algorithm
- Douglas-Rachford splitting
- Alternating directions method of multipliers
- Convergence analysis based on maximally monotone operators
- Primal and dual residuals. Choice of primal and dual stepsizes

Goal: Learn about state-of-the-art first order optimization methods and their relations.

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

Lecture overview

Chapter 5: To be defined by your interests

- Majorize-Minimize algorithm
- Other splitting algorithms: Peaceman-Rachford, etc.

Example of Linear Program: 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.
- The cost of a time unit of machine A is half the cost of a time unit of machine B.
- 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?