Chapter 0
Organization and Overview

Convex Optimization for Machine Learning & Computer Vision
WS 2019/20

Tao Wu
Zhenzhang Ye

Computer Vision Group
Department of Informatics
TU Munich

Organization

Whether this lecture fits you?

Prerequisites

- Background in Mathematical Analysis and Linear Algebra.
- Implementation in Python (or Matlab).
- Interest in Mathematical Theory (why algorithms work).

Whether this lecture fits you?

Prerequisites

- Background in Mathematical Analysis and Linear Algebra.
- Implementation in Python (or Matlab).
- Interest in Mathematical Theory (why algorithms work).

Nice plus (but not necessary)

- Experience in Machine Learning and Computer Vision
  e.g., CV I & II, ML for CV, Probab. Graphical Models in CV.
- Knowledge and experience in Continuous Optimization
- Knowledge in Functional Analysis (which provides generalization of concepts and theorems)

Course overview

Exercise session (organized by Zhenzhang Ye)

- Exercise sheets covering the content of the lecture will be passed out every Wednesday.
- Exercises contain theoretical as well as programming questions.
- Should submitted solutions be obviously copied, both groups would get 0 points.
- You may work on the exercises in groups of two.
- You are encouraged to present your solution on board at exercise class.
- To get a 0.3 grade bonus, you need to fulfill 75% of the total exercise points and present solution at least once.

Course overview

Lectures

1. Essential theory from convex analysis.
2. Formulation and analysis of optimization algorithms.
3. Implementation of algorithms on selected applications.
4. Extended topic (tentative): Stochastic optimization.

Tao Wu
Zhenzhang Ye

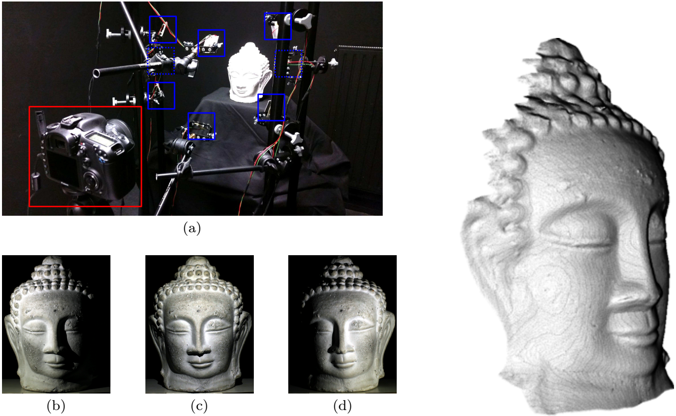
Contact us

Miscellaneous info

- Tao's office: 02.09.061
- Zhenzhang's office: 02.09.060
- Office hours: Please write an email.
- Lecture: Starts at quarter past; Short break in between.
- Course website (where you check out announcements):
  `https://vision.in.tum.de/teaching/ws2019/cvx4cv`
- Submit your programming exercises per email to:
  `yez@in.tum.de`
- Passcode for accessing course materials:
  `legendre`

Variational Methods in Computer Vision

Photometric stereo for 3D reconstruction

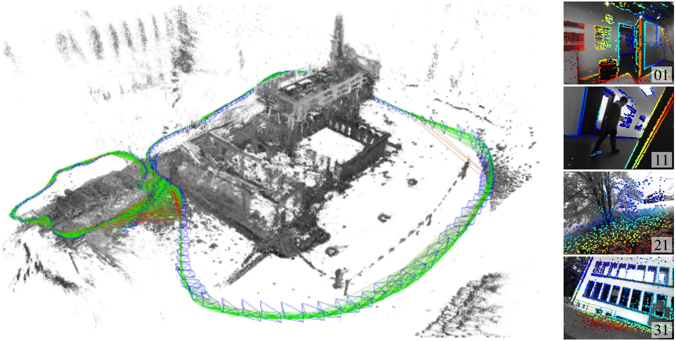


LED photometric stereo [Quéau et al '18]

Minimize photometric error from a Lambertian shading model:

$$\min_{\rho, z \in \mathbb{R}^{\Omega}} \sum_{i=1}^{n} \sum_{j \in \Omega} \psi\left(\rho_j \left\{\mathbf{l}_j^i(z) \cdot \mathbf{n}_j(z)\right\}_+ - I_j^i\right).$$

Visual odometry



Direct sparse odometry (DSO) [Engel et al '18]

Minimize photometric error of reprojected features:

$$\min_{\{c_i\},\{u_i\},\{d_{\mathbf{p}}\}} \sum_{i \in \mathcal{F}} \sum_{\mathbf{p} \in \mathcal{P}_i} \sum_{j \in \mathcal{Q}_{\mathbf{p}}} f_{i,\mathbf{p},j}(c_i, u_i, d_{\mathbf{p}}, u_j) + \lambda \sum_{i \in \mathcal{F}} g_i(c_i, u_i).$$

Foreground-background separation by matrix decomposition

raw video frames Z $\approx$ background A + foreground B

[Candès et al '11]

Low-rank and sparse matrix decomposition:

$$\min_{A,B\in\mathbb{R}^{n\times m}} \|A\|_{\text{nuclear}} + \lambda\|B\|_{\ell_1} + \delta\{\|A+B-Z\|_2 \leq \epsilon\}.$$

Tao Wu
Zhenzhang Ye

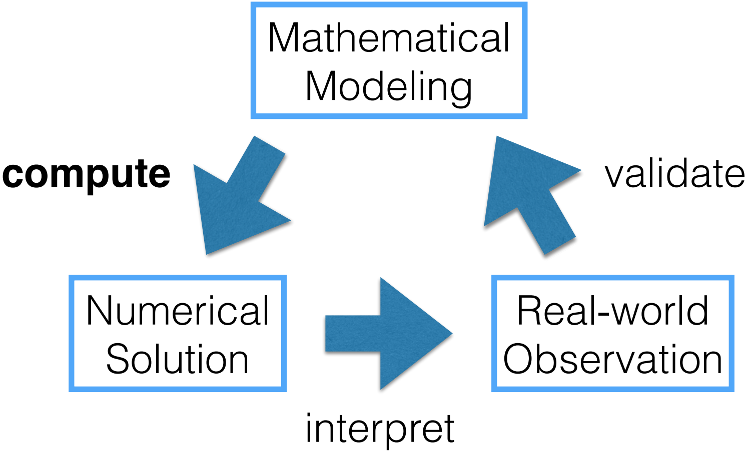
Image classification by logistic regression

MNIST handwritten digits.

Minimize negative log-likelihood:

$$\min_{W,b} \; -\frac{1}{N}\sum_{n=1}^{N} \log\left(\frac{\exp(\langle W_{Y_n,\cdot}, X_n\rangle + b_{Y_n})}{\sum_{k=1}^{10}\exp(\langle W_{k,\cdot}, X_n\rangle + b_k)}\right) + R(W,b).$$

Driving cycle



Appetizer: Image segmentation

- Image segmentation / clustering:

image

segmentation ($L = 4$)

Appetizer: Image segmentation

- Image segmentation / clustering:

image

segmentation ($L = 4$)

- Variational method for finding label function $u : \Omega \to \Delta^L$

$$\min_u \sum_{j\in\Omega} \left(\delta\{u_j \in \Delta^L\} + \langle u_j, f_j\rangle\right) + \alpha \sum_{l=1}^{L} \sum_i \omega_i \|(\nabla u^l)_i\|,$$

where

- Pointwise constraint: Δ^L is the unit simplex in $\mathbb{R}^L$.
- Unary term: $f : \Omega \to \mathbb{R}^L$ is pre-computed.
- Pairwise term: $\sum_i \omega_i \cdot (\nabla u^l)_i$ is the weighted total-variation.

An instance of convex optimization

- The variational model

$$\min_u \sum_{j\in\Omega} \left(\delta\{u_j \in \Delta^{L-1}\} + \langle u_j, f_j\rangle\right) + \alpha \sum_{l=1}^{L}\sum_i \omega_i \|(\nabla u^l)_i\|,$$

  is a special case of **convex optimization**

$$\text{minimize } J(u) + \delta\{u \in C\},$$

  with **convex objective** J and **convex constraint** C.

- This course is about **theory** and **practice** for solving convex optimization problem that arise from computer vision and machine learning.

Tao Wu
Zhenzhang Ye

Apply a solver

- Put into canonical form:

$$\min_{u \in \mathbb{R}^n} F(Ku) + G(u), \qquad \text{(primal)}$$

where $F : \mathbb{R}^m \to \mathbb{R},\ G : \mathbb{R}^n \to \mathbb{R}$ are *convex functions*, $K \in \mathbb{R}^{n \times m}$ is a matrix.

Tao Wu
Zhenzhang Ye

Apply a solver

- Put into canonical form:

$$\min_{u \in \mathbb{R}^n} F(Ku) + G(u), \qquad \text{(primal)}$$

where $F : \mathbb{R}^m \to \mathbb{R},\ G : \mathbb{R}^n \to \mathbb{R}$ are *convex functions*, $K \in \mathbb{R}^{n \times m}$ is a matrix.

- Reformulate the problem (by introducing *dual variable* p):

$$\max_{p \in \mathbb{R}^m} -F^*(p) - G^*(-K^\top p), \qquad \text{(dual)}$$

$$\max_{p \in \mathbb{R}^m} \min_{u \in \mathbb{R}^n} \langle Ku, p \rangle - F^*(p) + G(u), \qquad \text{(saddle-point)}$$

where F^* is the *convex conjugate* of F.

Tao Wu
Zhenzhang Ye

Apply a solver

- Put into canonical form:

$$\min_{u\in\mathbb{R}^n} F(Ku) + G(u), \qquad \text{(primal)}$$

where $F:\mathbb{R}^m \to \mathbb{R},\ G:\mathbb{R}^n \to \mathbb{R}$ are *convex functions*, $K \in \mathbb{R}^{n\times m}$ is a matrix.

- Reformulate the problem (by introducing *dual variable p*):

$$\max_{p\in\mathbb{R}^m} -F^*(p) - G^*(-K^\top p), \qquad \text{(dual)}$$

$$\max_{p\in\mathbb{R}^m} \min_{u\in\mathbb{R}^n} \langle Ku, p\rangle - F^*(p) + G(u), \qquad \text{(saddle-point)}$$

where F^* is the *convex conjugate* of F.

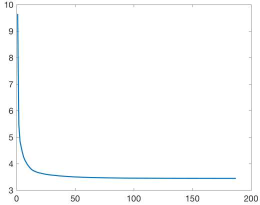
- Apply *PDHG* on the saddle-point formulation:

$$u^{k+1} = \arg\min_u \left\langle u, K^\top p^k \right\rangle + G(u) + \frac{s}{2}\|u - u^k\|^2,$$

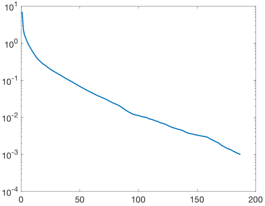
$$p^{k+1} = \arg\min_p -\left\langle K(2u^{k+1} - u^k), p \right\rangle + F^*(p) + \frac{t}{2}\|p - p^k\|^2.$$

What you are expected to learn from this course

energy value



primal-dual gap



- Does a minimizer always exist?
- How to characterize a minimizer via optimality condition?
- How to derive an (efficient) optimization algorithm?
- How to analyze and observe the convergence?
- Implementation in Python (with numpy).

Ready to start?