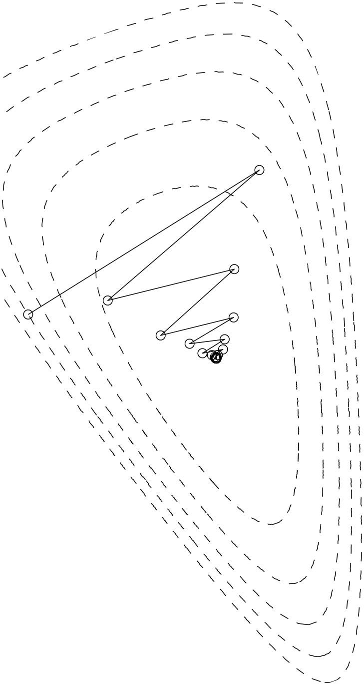


Introduction to Optimization

Introduction

Marc Toussaint
University of Stuttgart
Summer 2015

Why Optimization is interesting!

- In an otherwise unfortunate interview I've been asked why "we guys" (AI, ML, optimal control people) always talk about optimality. "People are by no means optimal", the interviewer said. I think that statement pinpoints the whole misunderstanding of the role and concept of optimality principles.
  - *Optimality principles are a means of scientific (or engineering) description.*
  - It is often easier to describe a thing (natural or artifical) via an optimality priciple than directly

- Which science does *not* use optimality principles to describe nature & artifacts?
  - Physics, Chemistry, Biology, Mechanics, ...
  - Operations research, scheduling, ...
  - Computer Vision, Speach Recognition, Machine Learning, Robotics, ...

- Endless applications

Teaching optimization

- Standard: *Convex Optimization, Numerical Optimization*
- Discrete Optimization (Stefan Funke)
- Exotics: Evolutionary Algorithms, Swarm optimization, etc

Teaching optimization

- Standard: *Convex Optimization, Numerical Optimization*
- Discrete Optimization (Stefan Funke)
- Exotics: Evolutionary Algorithms, Swarm optimization, etc

- In this lecture I try to cover the standard topics, but include as well work on stochastic search & global optimization

Rough Types of Optimization Problems

- Generic optimization problem:
  Let $x \in \mathbb{R}^n,\ f:\ \mathbb{R}^n \to \mathbb{R},\ g:\ \mathbb{R}^n \to \mathbb{R}^m,\ h: \mathbb{R}^n \to \mathbb{R}^l$. Find

$$\min_x \quad f(x)$$
$$\text{s.t.} \quad g(x) \le 0\ , \quad h(x) = 0$$

- **Blackbox**: only $f(x)$ can be evaluated
- **Gradient**: $\nabla f(x)$ can be evaluated
- Gauss-Newton type: $f(x) = \phi(x)^\top \phi(x)$ and $\nabla \phi(x)$ can be evaluated
- **2nd order**: $\nabla^2 f(x)$ can be evaluated

- "Approximate upgrade":
  – Use samples of $f(x)$ to approximate $\nabla f(x)$ locally
  – Use samples of $\nabla f(x)$ to approximate $\nabla^2 f(x)$ locally

Optimization in Machine Learning: SVMs

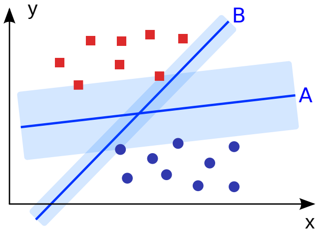
- optimization problem

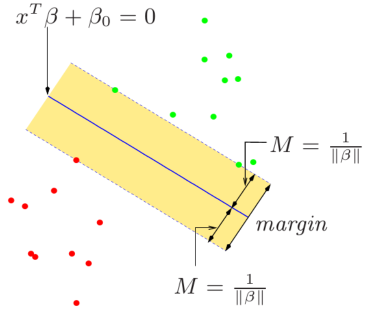
  $\max_{\beta, \|\beta\|=1} M$ subject to $y_i(\phi(x_i)^\top \beta) \geq M, \quad i = 1, \ldots, n$

- can be rephrased as

  $\min_\beta \|\beta\|$ subject to $y_i(\phi(x_i)^\top \beta) \geq 1, \quad i = 1, \ldots, n$

  Ridge regularization like ridge regression, but different loss





Optimization in Robotics

- Trajectories:
  Let $x_t \in \mathbb{R}^n$ be a joint configuration and $x = x_{1:T} = (x_1, \dots, x_T)$ a trajectory of length T. Find

$$\begin{aligned} \min_x \quad & \sum_{t=0}^T f_t(x_{t-k:t})^\top f_t(x_{t-k:t}) \\ \text{s.t.} \quad & \forall_t : \; g_t(x_t) \le 0 \,, \quad h_t(x_t) = 0 \end{aligned} \tag{1}$$

- Control:

$$\min_{u,\ddot q,\lambda} \quad \|u - a\|^2_H \tag{2}$$

$$\text{s.t.} \quad u = M\ddot q + h + J_g^\top \lambda \tag{3}$$

$$J_\phi \ddot q = c \tag{4}$$

$$\lambda = \lambda^* \tag{5}$$

$$J_g \ddot q = b \tag{6}$$

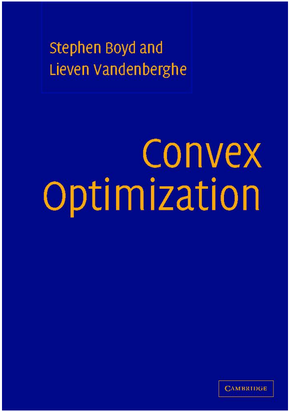
Optimization in Computer Vision

- Andres Bruhn's lectures
- Flow estimation, (relaxed) min-cut problems, segmentation, ...

Planned Outline

- Unconstrained Optimization: Gradient- and 2nd order methods
  - stepsize & direction, plain gradient descent, steepest descent, line search & trust region methods, conjugate gradient
  - Newton, Gauss-Newton, Quasi-Newton, (L)BFGS
- Constrained Optimization
  - log barrier, squared penalties, augmented Lagrangian
  - Lagrangian, KKT conditions, Lagrange dual, log barrier $\leftrightarrow$ approx. KKT
- Special convex cases
  - Linear Programming, (sequential) Quadratic Programming
  - Simplex algorithm
  - Relaxation of integer linear programs
- Global Optimization
  - infinite bandits, probabilistic modelling, exploration vs. exploitation, GP-UCB
- Stochastic search
  - Blackbox optimization (0th order methods), MCMC, downhill simplex

Books

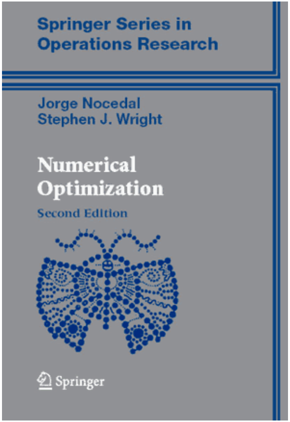


Boyd and Vandenberghe: *Convex Optimization.*

http://www.stanford.edu/~boyd/cvxbook/

(this course will not go to the full depth in math of Boyd et al.)

Books



Nocedal & Wright: *Numerical Optimization*

`www.bioinfo.org.cn/~wangchao/maa/Numerical_Optimization.pdf`

Organisation

- Webpage:

http://ipvs.informatik.uni-stuttgart.de/mlr/marc/teaching/15-Optimization/

  - Slides, Exercises & Software (C++)
  - Links to books and other resources

- Admin things, please first ask:
  Carola Stahl, Carola.Stahl@ipvs.uni-stuttgart.de, Raum 2.217

- Rules for the tutorials:
  - Doing the exercises is crucial!
  - At the beginning of each tutorial:
    – sign into a list
    – mark which exercises you have (successfully) worked on
  - Students are randomly selected to present their solutions
  - You need 50% of completed exercises to be allowed to the exam
  - Please check 2 weeks before the end of the term, if you can take the exam