Convex Optimization with Engineering Applications

Ulf T. Jönsson

Royal Institute of Technology (KTH)

Outline

Outline

Introduction to the course

This course gives you an introduction to

1. the theory of convex sets and convex functions
2. several important classes of convex optimization problems
3. algorithms for unconstrained and constrained optimization
4. applications of convex optimization in engineering

This years teachers

- Anders Forsgren (Fundamental theory and algorithms)
- Ulf Jönsson (Theory topics and applications)

Course literature

S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004, ISBN: 0521833787.

Supporting material can also be found at Stephen Boyd's homepage

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

and at Lieven Vandenberghe's homepage

http://www.ee.ucla.edu/ vandenbe/

Lecture slides can be found at the course homepage

http://www.math.kth.se/optsyst/forskning/forskarutbildning/SF3849/info.html

Note: This set of slides are to a very large extent adopted from previous course slides by M. Johansson. They also borrow from the lectures by S. Boyd (Stanford) and L. Vandenberghe (UCLA).

Outline

1. Introduction to the course
2. Classes of optimization problems
3. Applications
4. Administrative issues

General mathematical optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- $x = (x_1, \ldots, x_n)$: optimization variables
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: objective function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$: inequality constraint functions
- $h_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, p$: equality constraint functions

$x^* \in \mathcal{F} = \{x \in \mathbf{R}^n : f_i(x) \leq 0; h_i(x) = 0,\ \forall i\}$ is a (global) optimal solution if

$$f_0(x^*) \leq f_0(x), \quad \forall x \in \mathcal{F}.$$

Solving optimization problems

The general mathematical optimization problem

- includes a large number of relevant problems
- is very difficult to solve in general
  - long computation times
  - cannot ensure that the global optimum has been found

Structure of the problem needs to be explored in order to

- develop efficient algorithms
- verify that an optimal solution has been found.

Some optimization problems with nice structure

- least squares problems
- linear programming problems
- convex optimization problems
  - geometric programming
  - second-order cone programming
  - semidefinite programming

Least-squares optimization

$$\text{minimize} \quad \|Ax - b\|_2^2 = \quad \text{minimize} \quad (Ax - b)^T (Ax - b)$$

- If A has full column rank then the analytical solution is

$$x^* = (A^T A)^{-1} A^T b$$

- There exists reliable and efficient algorithms and software
- least squares problems are easy to recognize and useful in approximation problems

Linear programming (LP)

Linear program on *standard form*

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Linear program on *inequality form*

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \end{aligned}$$

They are equivalent because adding slack variables to the inequality form gives an equivalent problem on standard form

$$\begin{aligned} \text{minimize} \quad & \begin{bmatrix} c^T & -c^T & 0 \end{bmatrix} z \\ \text{subject to} \quad & \begin{bmatrix} A & -A & I \end{bmatrix} z = b \\ & z = \begin{bmatrix} x_+^T & x_-^T & s^T \end{bmatrix} \geq 0 \end{aligned}$$

Properties of LP

- No analytical formula
- Reliable and efficient software
  - The simplex algorithm
  - Interior point algorithms
- Many practical problems can be formulated as LP

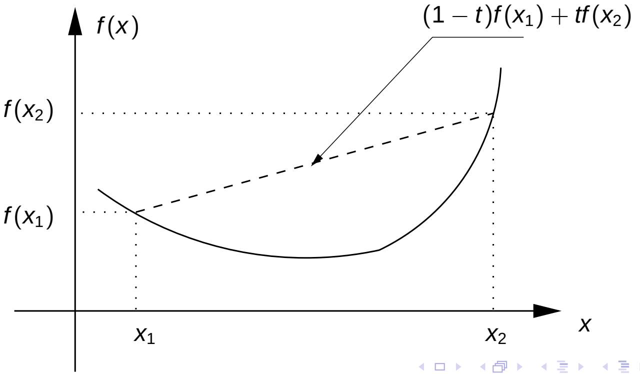
Convex optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & a_i^T x = b_i, \quad i = 1, \dots, p \end{array}$$

where the objective and the constraint functions are convex

$$f_i((1-t)x_1 + tx_2) \leq (1-t)f_i(x_1) + tf_i(x_2), \qquad t \in [0,1]$$

for $i = 0, 1, \dots, m$.

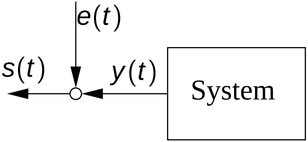


Properties of convex programs

- No analytic solution
- Reliable and efficient algorithms
- Requires ingenuity to recognize or transform to convex form
- Many problems in control, communication and signal processing can be formulated as convex optimization problems.

Outline

Linear regression (Model fitting)



Problem: Fit a linear regression model to measurement data

$$\text{Regression model}: \quad y(t) = \sum_{j=1}^{n} \alpha_j \psi_j(t)$$

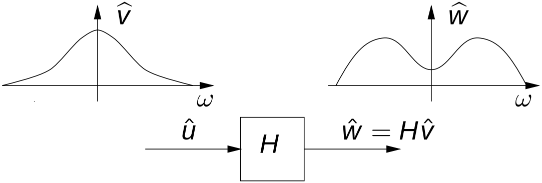
- $\psi_k(t)$, $k = 1, \ldots, n$: Regressors (known functions)
- $\alpha_k = k = 1, \ldots, n$: Model parameters (to be determined)
- $e(t)$ measurement noise (not known)
- $s(t)$ observations.

Idea for model fitting: Least mean square of the prediction errors

$$
\begin{aligned}
&\text{minimize}\frac{1}{2}\sum_{i=1}^{m}(\sum_{j=1}^{n}\alpha_j\psi_j(t_i)-s(t_i))^2\\
=&\text{minimize}\frac{1}{2}\sum_{i=1}^{m}(\psi(t_i)^T x - s(t_i))^2\\
=&\text{minimize}\frac{1}{2}(Ax-b)^T(Ax-b) \qquad (1)
\end{aligned}
$$

$$
\psi(t)=\begin{bmatrix}\psi_1(t)\\ \vdots\\ \psi_n(t)\end{bmatrix},\quad x=\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{bmatrix},\quad A=\begin{bmatrix}\psi(t_1)^T\\ \vdots\\ \psi(t_n)^T\end{bmatrix},\quad b=\begin{bmatrix}s(t_1)\\ \vdots\\ s(t_n)\end{bmatrix}
$$

Optimal filter design



Finite impulse response (FIR) filter

$$y(t) = \sum_{\tau=0}^{n-1} h_\tau u(t-\tau) = h_0 u(t) + h_1 u(t-1) + \cdots + h_{n-1} u(t-n+1).$$

Frequency response of the FIR filter

$$\begin{aligned} H(\omega) &= h_0 + h_1 e^{-j\omega} + \ldots + h_{n-1} e^{-i(n-1)\omega} \\ &= \sum_{t=0}^{n-1} h_t \cos(t\omega) + i \sum_{t=0}^{n-1} h_t \sin(t\omega) \end{aligned}$$

Chebyshev design via convex optimization

Chebyshev design: minimize deviation from desired frequency response

$$\text{minimize} \quad \max_{\omega\in[0,2\pi]} |H(\omega) - H_{\text{des}}(\omega)|$$

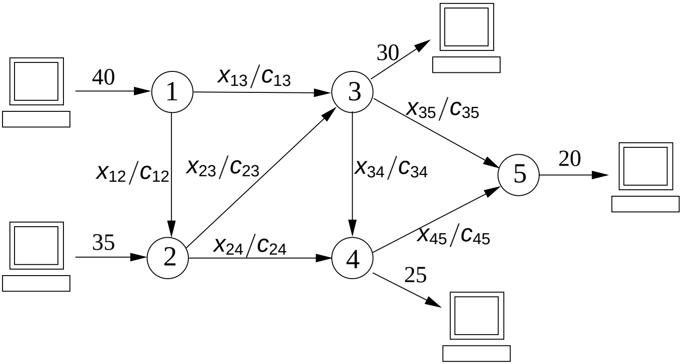
After discretizing frequency we obtain

$$\begin{array}{ll} \text{minimize} & t \\ \text{subj.to} & |H(\omega_k) - H_{\text{des}}(\omega_k)| \leq t, \quad k = 1, \ldots, K \end{array}$$

which results in the *second order cone program*

$$\begin{array}{ll} \text{minimize} & t \\ \text{subj.to} & \left\| \begin{bmatrix} 1 & \cos(\omega_k) & \ldots & \cos((n-1)\omega_k) \\ 0 & -\sin(\omega_k) & \ldots & -\sin((n-1)\omega_k) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{n-1} \end{bmatrix} - \begin{bmatrix} \mathrm{Re}\, H_{\text{des}}(\omega_k) \\ \mathrm{Im}\, H_{\text{des}}(\omega_k) \end{bmatrix} \right\| \leq t, \\ & \text{for all } k \end{array}$$

Network flow problem



The problem is to send data from servers in nod 1 and 2 to terminals in node 3, 4, 5.

The cost of traffic in the link between node i and j is c_{ij} Kr/Kbyte.

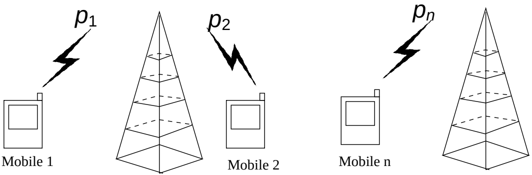
We want to minimize the cost for the total data traffic.

Resulting optimization problem

$$
\begin{array}{ll}
\text{minimize} & \sum\limits_{\text{all arcs}} c_{ij} x_{ij} \\
\text{subject to} & x_{12} + x_{13} = 40 \\
& -x_{12} + x_{23} + x_{24} = 35 \\
& -x_{13} - x_{23} + x_{34} + x_{35} = -30 \\
& -x_{24} - x_{34} + x_{45} = -25 \\
& -x_{35} - x_{45} = -20 \\
& x_{ij} \geq 0, \ \text{for all arcs in the graph}
\end{array}
$$

- This is an LP.

Power allocation in wireless networks



Several mobiles communicates in uplink with the base stations. The capacity depends on the signal-to-interference ratio

$$\gamma_i(p) = \frac{g_{ii}p_i}{\sum_{j \neq i} g_{ij}p_j + \sigma_i}$$

where g_{ij} are channel gains and σ is the thermal noise.

Case 1: Minimal power subject to target SIR constraints

$$
\begin{array}{ll}
\text{minimize} & \sum_{i=1}^{n} p_i \\
\text{subject to} & \gamma_i(p) \geq \gamma_i^T \\
& p_i \geq 0
\end{array}
$$

which can be re-written as an LP.

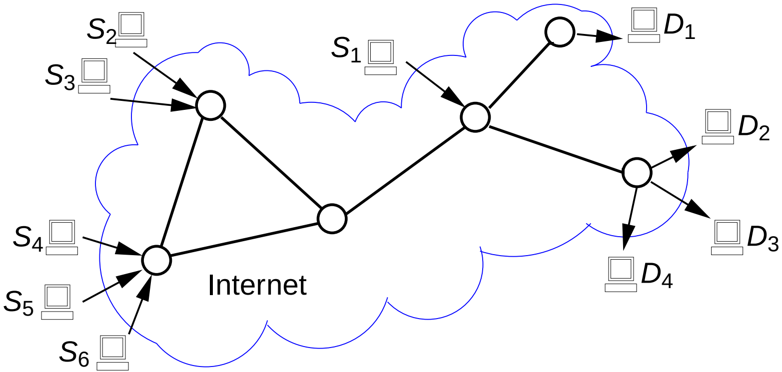
Case 2: A weighted throughput problem

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{n} w_i \log(\gamma_i(p)) \\ \text{subject to} & \sum_{i=1}^{n} p_i \leq p_{tot} \end{array}$$

which can be re-formulated as the *geometric program*

$$\begin{array}{ll} \text{minimize} & \prod_{i=1}^{n} t_i^{w_i} \\ \text{subject to} & g_{ii} p_i^{-1} t_i^{-1} (\sigma_i + \sum_{j \neq i} g_{ij} p_j) \leq 1, \; i = 1, \ldots, n \\ & \sum_{i=1}^{n} p_i \leq p_{tot} \end{array}$$

Internet Congestion Control



- Several users compete for limited network resourses.
- Congestion control helps to share the resources fairly and efficiently.

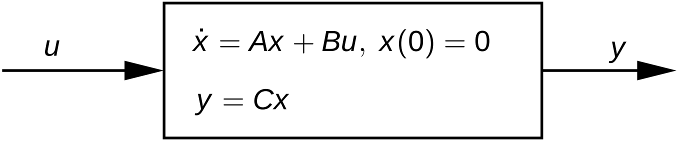
Network Utility Maximization (NUM)

$$
\begin{array}{rll}
\text{maximize} & \sum_{p=1}^{P} u_p(s_p) & \\
\text{subject to} & \sum_{p \in \mathcal{P}} s_p \leq c_l, \quad l = 1, \ldots, L & \qquad (NUM) \\
 & s_p \geq 0, \quad p = 1, \ldots, P &
\end{array}
$$

- s_p data rate from user $p = 1, \ldots, P$
- c_l capacity constraint on link $l = 1, \ldots, L$
- u_p, $p = 1, \ldots, P$, strictly concave *utility function*

- The solution to NUM gives the optimal rates.
- Iterative algorithms based on dual decompostion techniques provides a decentralized control scheme:
  1. Source controllers update the data rates based on congestion information from the links
  2. Link controllers update the congestion signal (queue length in buffer) based on the total data rate in the link.

$\mathbf{L}_2$ gain computation



The $\mathbf{L}_2$ gain of the system is the smallest γ such that

$$\int_0^T y(t)^T y(t) dt \leq \gamma^2 \int_0^T u(t)^T u(t) dt$$

for all $T > 0$ and all square integrable inputs u.

Fact: The $\mathbf{L}_2$ gain is γ if and only if there exists $P > 0$ such that $V(x) = x^T P x$ satisfies the quadratic inequality

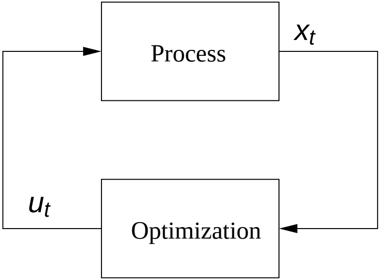
$$\frac{\partial V}{\partial x}(Ax + Bu) \leq \gamma^2 u^T u - y^T y, \ \forall (x, u, y = Cx).$$

It follows that the $\mathbf{L}_2$ gain can be computed as

$$\begin{array}{ll} \text{minimize} & \gamma^2 \\ \text{subject to} & \begin{bmatrix} A^T P + PA + C^T C & PB \\ B^T P & -\gamma^2 I \end{bmatrix} \preceq 0 \\ & P \succ 0 \end{array}$$

- *Linear matrix inequality* that can be solved using interior point algorithms.

Model predictive control (MPC)



- The idea behind MPC is to use an optimization algorithm as controller.
- The optimization is done based on predicted state variables.
- The prediction is done based on a model which is the reason for the term “model predictive control”

The algorithm

1. Measure $x_{t|t} := x_t$.
2. Determine $U_t^* = (u_{t|t}^*, u_{t+1|t}^*, \dots, u_{t+N-1|t}^*)$ by solving

$$\min \sum_{k=0}^{N-1} f_0(x_{t+k|t}, u_{t+k|t}) \quad \text{subj. to} \quad \begin{cases} x_{t+k+1|t} = f(x_{t+k|t}, u_{t+k|t}), \\ x_{t+k|t} \in \mathbf{X},\ u_{t+k|t} \in \mathbf{U} \\ x_{t|t} = x_t \end{cases}$$

3. Apply $u_t := u_{t|t}^*$
4. Let $t := t+1$ and go to 1.

$x_{t+k+1|t} = f(x_{t+k|t}, u_{t+k|t})$ is the predicted state given $x_{t|t}$.

Online Optimization Versus Explicit MPC

1. Online optimization: Solve on-line in run-time the optimization $\min_{U_t} J(x_{t|t}, U_t)$, where

$$U_t = (u_{t|t}, u_{t+1|t}, \dots, u_{t+N-1|t})$$

$$J(x_{t|t}, U_t) = \sum_{k=0}^{N-1} f_0(x_{t+k|t}, u_{t+k|t})$$

$$\text{subj. to} \quad \begin{cases} x_{t+k+1|t} = f(x_{t+k|t}, u_{t+k|t}), \\ x_{t+k|t} \in \mathbf{X}, \; u_{t+k|t} \in \mathbf{U} \end{cases}$$

2. Find explicit solution

$$U_t^* = (\mu(0, x_{t|t}), \dots, \mu(N-1, x_{t+N-1|t})) = \operatorname{argmin}_{U_t} J(x_{t|t}, U_t)$$

and use $u_{t|t}^* = \mu(x_{t|t}) := \mu(0, x_{t|t})$ as state feedback function.

Linear quadratic case

Let $x_k := x_{t+k|t}$ and $u_k = u_{t+k|t}$. The linear quadratic MPC optimization becomes

$$
\begin{aligned}
\text{minimize} \quad & \sum_{k=1}^{K} x_k^T Q x_k + u_k^T R u_k \\
\text{subject to} \quad & \begin{cases} x_{k+1} = A x_k + B u_k\,, & x_0 \text{ given} \\ C x_k + D u_k \leq \eta, & k = 1, \ldots, K \end{cases}
\end{aligned}
$$

This is a convex optimization problem if $R > 0$ and $Q \geq 0$.

The following closed formula for the system response

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\mathcal{X}} = \underbrace{\begin{bmatrix} A \\ A^2 \\ \vdots \\ A^n \end{bmatrix}}_{\mathcal{A}} x_0 + \underbrace{\begin{bmatrix} B & 0 & \dots & 0 \\ AB & B & \dots & 0 \\ A^{n-1}B & A^{n-2}B & \dots & B \end{bmatrix}}_{\mathcal{B}} \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}}_{\mathcal{U}}$$

gives a compact formulation of the cost

$$\begin{aligned} \sum_{k=1}^{K} x_k^T Q x_k + u_k^T R u_k &= \mathcal{X}^T (Q \otimes I) \mathcal{X} + \mathcal{U}^T (R \otimes I) \mathcal{U} \\ &= x_0^T \mathcal{Q} x_0 + 2 x_0^T \mathcal{S}^T \mathcal{U} + \mathcal{U}^T \mathcal{R} \mathcal{U} \end{aligned}$$

where

$$\begin{bmatrix} \mathcal{Q} & \mathcal{S}^T \\ \mathcal{S} & \mathcal{R} \end{bmatrix} = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ 0 & I \end{bmatrix}^T \begin{bmatrix} Q \otimes I & 0 \\ 0 & R \otimes I \end{bmatrix} \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ 0 & I \end{bmatrix}$$

and for the constraint

$$C x_k + D u_k \le \eta \;\Leftrightarrow\; \underbrace{(C \otimes I)\mathcal{A}}_{\mathcal{C}} x_0 + \underbrace{((C \otimes I)\mathcal{B} + D \otimes I)}_{\mathcal{D}} \mathcal{U} \le \underbrace{\eta \otimes \mathbf{1}}_{b}$$

We have arrived at a quadratic program with inequality constraints

$$\begin{array}{ll} \text{minimize} & x_0^T \mathcal{Q} x_0 + 2x_0^T \mathcal{S}^T \mathcal{U} + \mathcal{U}^T \mathcal{R} \mathcal{U} \\ \text{subject to} & \left\{ \mathcal{C} x_0 + \mathcal{D} \mathcal{U} \leq b \right. \end{array}$$

The Karush-Kuhn-Tucker (KKT) optimality conditions become

$$\begin{aligned} \mathcal{R} u + \mathcal{S} x_0 + \mathcal{D}^T \lambda &= 0 \\ \lambda \geq 0 \\ \mathcal{C} x_0 + \mathcal{D} \mathcal{U} &\leq b \\ \lambda_i (\mathcal{C}_i x_0 + \mathcal{D}_i u - b_i) = 0 \end{aligned}$$

Explicit MPC

By solving the KKT conditions explicitly as a funcion of x_0 we get

$$u = -\mathcal{R}^{-1}(\mathcal{S}x_0 + \mathcal{D}^T\lambda)$$

within a polyhedron defined by the optimality constraints.

Application of optimization to engineering systems

There are many ways in which optimization can be used in design and analysis of engineering systems

- formulate design as an optimization problem
- interpret solution as algorithm for an underlying optimization problem
- extend fundamental theory using optimization-theoretic techniques

It gives a powerful, versatile and widely applicable viewpoint.

Course goals

You will learn to

1. Recognize/formulate problems as convex optimization problems.
2. Develop numerical codes for problems of moderate size
3. Characterize optimal solution, give limits of performance,...

Course contents

Theoretical tools

- convex sets, convex functions, convex optimization, duality

Important classes of convex optimization problems

- linear programming, quadratic programming, second-order cone programming, geometric programming, semidefinite programming

Numerical algorithms

- simple algorithms, basic ideas behind interior point algorithms, hands-on-experience with some algorithms

Applications

- filter design, wireless power control, network flow problems, model predictive control, linear matrix inequalities in analysis and design of control systems, ...

- Only small samples from a vast research field will be covered
  - The course gives you a guided tour
- The emphasis will be on core ideas and concepts not on detailed proofs
- Motivation for the theory comes from the need to understand
  - Ideas behind proofs are often useful/crucial for understanding
- No prior knowledge of optimization required but it helps

Outline

Course content and schedule

1. Introduction
2. Convexity
3. Linear programming and the simplex method.
4. Linear programming, Lagrangian relaxation and duality
5. Convex programming and semidefinite programming.
6. Geometric programming and second-order cone programming
7. Sensitivity and multiobjective optimization
8. Decomposition and large-scale optimization
9. Smooth convex unconstrained and equality-constrained minimization
10. Interior methods
11. Applications in communications and control
12. Applications in communications and control

Course requirements

There is one version of the course given this time, the 6-credit version

1. The 6-credit version requires successful completion of homework assignments and the presentation of a short lecture on a special topic.

Homework sets

There will be a total of four mandatory hand-ins distributed during the course.

Topic	Deadline
Convexity	November 16
Convex optimization problems	November 29
Duality	December 9
Numerical algorithms	December 21

Late homework solutions are not accepted.

Research paper presentations

Read and present a research paper, individually or in groups

In 10-15 minutes, try to address the following:

1. What engineering problem is being solved? Why is it important?
2. How can the problem be formulated as an optimization problem
3. Is the problem convex? Use what you learned in course to explain!
4. How is the problem solved? Explain any new techniques!
5. What type of results are obtained?
6. Are there any obvious flaws, or extensions? Try to teach us what you have learned!

Preliminary presentation dates: Tuesday December 21, 2010 and Monday January 10, 2011.