Lecture 1 : Introduction to Convex Optimization CS709

Instructor: Prof. Ganesh Ramakrishnan

Introduction: Mathematical optimization

- **Motivating Example**
- **Applications**
- **Convex optimization**
- **Least-squares(LS) and linear programming(LP) - Very common place**

- **Course goals and topics**
- **Nonlinear optimization**
- **Brief history of convex optimization**

Mathematical optimization

$x^* = \text{argmin } f_0(x)$ s.t $f_i(x) \leq b_i$

(Mathematical) Optimization problem:-

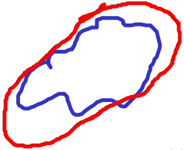
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq b_i, \; i = 1, \ldots, m. \end{aligned}$$

$f_i(x) - b_i \leq 0$

x = $(x_1, \ldots, x_n)$: optimization variables
$f_i : R^n \rightarrow R$, i = 1,...,m : constraint functions
optimal solution x^* has smallest value of f_0 among all vectors that satisfy the constraints

Almost Every Problem can be posed as an Optimization Problem

- Given a set $\mathcal{C} \subseteq \Re^n$ find the ellipsoid $\mathcal{E} \subseteq \Re^n$ that is of smallest volume such that $\mathcal{C} \subseteq \mathcal{E}$.
  Hint: First work out the problem in lower dimensions
- Sphere $\mathcal{S}_r \subseteq \Re^n$ centered at $\mathbf{0}$ is expressed as: $\{x \mid \|x\|_2 \leq r\}$



$$\|x\|_2 = \sqrt{\sum_i |x_i|^2}$$

Ax

Ellipsoid: Symmetrically scaled & rotated & possibly translated sphere

$Ax + b$

$\{x \mid \|Ax+b\|_2 \leq 1\}$

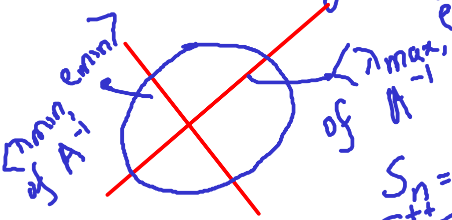
scale to radius = 1

Almost Every Problem can be posed as an Optimization Problem

- Given a set $\mathcal{C} \subseteq \Re^n$ find the ellipsoid $\mathcal{E} \subseteq \Re^n$ that is of smallest volume such that $\mathcal{C} \subseteq \mathcal{E}$. *Hint: First work out the problem in lower dimensions.*
- Sphere $\mathcal{S}_r \subseteq \Re^n$ centered at $\mathbf{0}$ is expressed as: $\mathcal{S} = \{\mathbf{u} \in \Re^n \mid \|\mathbf{u}\|_2 \leq r\}$
- Ellipsoid $\mathcal{E} \subseteq \Re^n$ is expressed as: $\{v \mid \|Av + b\|_2 \leq 1\}$ $A \in S_n^{++}$

Volume of ellipsoid $\propto$ prod of eigenvalues of A^{-1}

$= \det(A^{-1})$



S_n = family of symmetric $n \times n$ matrices

$S_n^{++} \subseteq S_n$ = family of positive def in S_n

Almost Every Problem can be posed as an Optimization Problem

- Given a set $\mathcal{C} \subseteq \Re^n$ find the ellipsoid $\mathcal{E} \subseteq \Re^n$ that is of smallest volume such that $\mathcal{C} \subseteq \mathcal{E}$. *Hint: First work out the problem in lower dimensions.*
- Sphere $\mathcal{S}_r \subseteq \Re^n$ centered at $\mathbf{0}$ is expressed as: $\mathcal{S} = \left\{\mathbf{u} \in \Re^n \,|\, \|\mathbf{u}\|_2 \leq r\right\}$
- Ellipsoid $\mathcal{E} \subseteq \Re^n$ is expressed as:
  $\mathcal{E} = \left\{\mathbf{v} \in \Re^n | A\mathbf{v} + \mathbf{b} \in \mathcal{S}_1\right\} = \left\{\mathbf{v} \in \Re^n \,|\, \|A\mathbf{v} + \mathbf{b}\|_2 \leq 1\right\}$. Here, $A \in S^n_{++}$, that is, A is an $n \times n$ (strictly) positive definite matrix.
- The optimization problem will be:

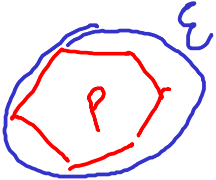
$$\begin{aligned} \underset{[a_{11}, a_{12} \ldots, a_{nn}, b_1, \ldots b_n]}{\text{minimize}} \quad & det(A^{-1}) \\ \text{subject to} \quad & \mathbf{v}^T A\mathbf{v} > 0, \ \forall\ \mathbf{v} \neq 0 \\ & \|A\mathbf{v} + \mathbf{b}\|_2 \leq 1, \ \forall \mathbf{v} \in \mathcal{C} \end{aligned}$$

$f_0(x)$; $A \in S^n_{++}$; $x = [A, b]$; $A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & & & \\ a_{n1} & \ldots & \ldots & a_{nn} \end{bmatrix}$; $b = [b_1 \ldots b_n]^T$; $\lambda_1(A), \lambda_2(A) \ldots \lambda_n(A) > 0$; Simplify when C is a polygon; $C \subseteq \mathcal{E}$

Every Problem can be posed as an Optimization Problem (contd.)

- Given a polygon $\mathcal{P}$ find the ellipsoid $\mathcal{E}$ that is of smallest volume such that $\mathcal{P} \subseteq \mathcal{E}$.
- Let $\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_p$ be the corners of the polygon $\mathcal{P}$
- The optimization problem will be:

$$\begin{aligned} \underset{[a_{11}, a_{12} \ldots, a_{nn}, b_1, ..b_n]}{\text{minimize}} \quad & det(A^{-1}) \\ \text{subject to} \quad & -\mathbf{v}^T A \mathbf{v} > 0, \ \forall \ \mathbf{v} \neq 0 \\ & \|A\mathbf{v}_i + \mathbf{b}\|_2 \leq 1, \ i \in \{1..p\} \end{aligned}$$



V_i's are vertices of the polygon

Natural questions to address:

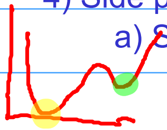
1) Abstract from this experience to help formulate an optimization problem in a new situation: Are there known/manageable families of optimization problems to which I could reduce my new problem?

* Linear Programs
* Quadratic Programs
* Positive Semi-definite Programs
* Conic Programs

2) Analysis: Does the problem have a unique solution or a solution at all?

3) Algorithms: How do I compute the best or nearly best solutions if they exist?

4) Side points:



a) Study how the solutions change with change in constraints? For example, if the ellipsoid was to be centred at origin or was to be axis-aligned, the optimal solution could be very different

Every Problem can be posed as an Optimization Problem (contd.)

- Given a polygon $\mathcal{P}$ find the ellipsoid $\mathcal{E}$ that is of smallest volume such that $\mathcal{P} \subseteq \mathcal{E}$.
- Let $\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_p$ be the corners of the polygon $\mathcal{P}$
- The optimization problem will be:

$$\begin{aligned} \underset{[a_{11}, a_{12} ..., a_{nn}, b_1, ..b_n]}{\text{minimize}} \quad & det(A^{-1}) \\ \text{subject to} \quad & -\mathbf{v}^T A \mathbf{v} > 0, \ \forall \ \mathbf{v} \neq 0 \\ & \|A\mathbf{v}_i + \mathbf{b}\|_2 \leq 1, \ i \in \{1..p\} \end{aligned}$$

- How would you pose an optimization problem to find the ellipsoid of largest volume that fits inside $\mathcal{C}$?

So Again: Mathematical optimization

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq b_i, \; i = 1, \ldots, m. \end{aligned}$$

$x = (x_1,...,x_n)$: optimization variables

$f_i : R^n \rightarrow R$, $i = 1,...,m$: constraint functions

optimal solution x^* has smallest value of f_0 among all vectors that satisfy the constraints

Examples

portfolio optimization

- variables: amounts invested in different assets
- constraints: budget, max./min. investment per asset, minimum return
- objective: overall risk or return variance

when constraints are violated, you have scams

Examples

device sizing in electronic circuits

- variables: device widths and lengths
- constraints: manufacturing limits, timing requirements, maximum area
- objective: power consumption

Examples

$\mathcal{D}$ for eg = set of pairs of (audio file, transcription)

x^i y^i

data fitting - **machine learning**

- variables: model parameters $= w$
- constraints: prior information, parameter limits $= C, \Omega$ → Stability of opt algos
- objective: measure of misfit or prediction error $= L(w)$

Bayesian prior

Generalization

$$\min_{w} \sum_{(x^i, y^i) \in \mathcal{D}} L(w, x^i, y^i)$$

s.t. $w \in C$

$\Omega(w) \leq \lambda$

More Generally..

- x represents some action such as
  - portfolio decisions to be made
  - resources to be allocated
  - schedule to be created
  - vehicle/airline deflections
- Constraints impose conditions on outcome based on
  - performance requirements
  - manufacturing process
- Objective $f_0(x)$ should be desirably small
  - total cost
  - risk
  - negative profit

Solving optimization problems

general optimization problems

- very difficult to solve
- methods involve some compromise, e.g., very long computation time, or not always finding the solution

exceptions: certain problem classes can be solved efficiently and reliably

- least-squares problems
- linear programming problems
- convex optimization problems

Least-squares

$(Ax^* - b) \perp C(A)$

$(Ax^* - b)^T Ax = 0 \quad \forall x$

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2^2$$

$x \in \mathbb{R}^n$

solving least-squares problems

- analytical solution: $x^* = (A^T A)^{-1} A^T b$
- reliable and efficient algorithms and software
- computation time proportional to $n^2 k$ ($A \in R^{k \times n}$); less if structured
- a mature technology

using least-squares

- least-squares problems are easy to recognize
- a few standard techniques increase flexibility (e.g., including weights, adding regularization terms)

$(A^T A + \lambda I)^{-1} A^T b = \arg\min_x \|Ax - b\|_2^2 + \lambda \|x\|_2^2$

Linear programming

c_i = cost of i^{th} vegetable
x_i = amt of " "

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & a_i^T x \geq b_i, \; i = 1, \ldots, m. \end{array}$$

Constraints on vitamins, minerals, proteins, carbs

solving linear programs

- no analytical formula for solution
- reliable and efficient algorithms and software
- computation time proportional to n^2m if $m \geq n$; less with structure
- a mature technology

using linear programs

- not as easy to recognize as least-squares problems
- a few standard tricks used to convert problems into linear programs (e.g., problems involving l_1- or l_∞-norms, piecewise-linear functions)

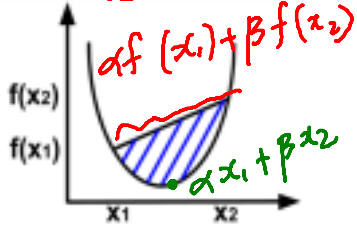
Convex optimization problem

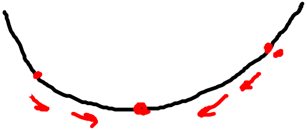
$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \ \ i = 1, \ldots, m. \end{array}$$

- objective and constraint functions are convex:

$$f_i(\alpha x_1 + \beta x_2) \leq \alpha f_i(x_1) + \beta f_i(x_2)$$

if $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$





- includes least-squares problems and linear programs as special cases

Convex optimization problem

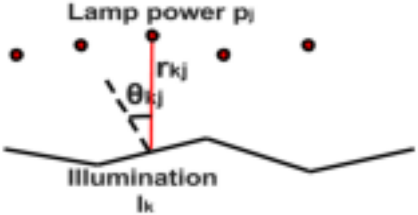
solving convex optimization problems

- no analytical solution
- reliable and efficient algorithms
- computation time (roughly) proportional to $\{n^3, n^2m, F\}$, where F is cost of evaluating f_i's and their first and second derivative
- almost a technology

using convex optimization

- often difficult to recognize
- many tricks for transforming problems into convex form
- surprisingly many problems can be solved via convex optimization

Example: m lamps illuminating n(small, flat) patches



intensity I_k at patch k depends linearly on lamp powers p_j:

$$I_k = \sum_{j=1}^{n} a_{kj} p_j,\, a_{kj} = {r_{kj}}^{-2} max\{cos\theta_{kj}, 0\}$$

problem: Provided the fixed locations(a_{kj}'s), achieve desired illumination I_{des} with bounded lamp powers

$$\begin{array}{ll} \underset{p_j}{\text{minimize}} & max_{k=1,..,n} \mid log(I_k) - log(I_{des}) \mid \\ \text{subject to} & 0 \leq p_j \leq p_{max},\ j = 1, \ldots, m. \end{array}$$

Example: m lamps illuminating n(small, flat) patches

How to solve? Some approximate(suboptimal) 'solutions':-

1. use uniform power: $p_j = p$, vary p
2. use least-squares:

$$\underset{p_j}{\text{minimize}} \quad \sum_{k=1}^{n} \| I_k - I_{des} \|_2^2$$

   round p_j if $p_j > p_{max}$ or $p_j < 0$
3. use weighted least-squares:

$$\underset{p_j}{\text{minimize}} \quad \sum_{k=1}^{n} \| I_k - I_{des} \|_2^2 + \sum_{j=1}^{m} w_j \| p_j - p_{max}/2 \|_2^2$$

   iteratively adjust weights w_j until $0 \le p_j \le p_{max}$
4. use linear programming:

$$\begin{aligned} \text{minimize} \quad & max_{k=1,..,n} \mid I_k - I_{des} \mid \\ \text{subject to} \quad & 0 \le p_j \le p_{max},\ j = 1, \ldots, m. \end{aligned}$$

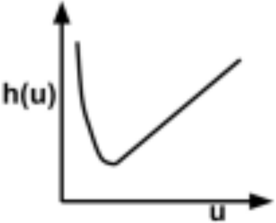
Example: m lamps illuminating n(small, flat) patches

- Use convex optimization: problem is equivalent to

$$\underset{p_j}{\text{minimize}} \quad f_0(p) = max_{k=1,..,n} h(I_k/I_{des})$$
$$\text{subject to} \quad 0 \le p_j \le p_{max}, \; j = 1, \dots, m.$$

will revisit

with $h(u) = \max\{u, 1/u\}$



- f_0 is convex because maximum of convex functions is convex
- **exact** solution obtained with effort ≈ modest factor × least-squares effort

Example: m lamps illuminating n(small, flat) patches

Additional constraints does adding 1 or 2 below complicate the problem?

1. no more than half of total power is in any 10 lamps.
2. no more than half of the lamps are on ($p_j > 0$).

Example: m lamps illuminating n(small, flat) patches

Additional constraints does adding 1 or 2 below complicate the problem?

1. no more than half of total power is in any 10 lamps.
2. no more than half of the lamps are on ($p_j > 0$).

- **answer:** with (1), still easy to solve; with (2), extremely difficult.

Example: m lamps illuminating n(small, flat) patches

Additional constraints does adding 1 or 2 below complicate the problem?

1. no more than half of total power is in any 10 lamps.
2. no more than half of the lamps are on ($p_j > 0$).

- **answer:** with (1), still easy to solve; with (2), extremely difficult.
- **moral:** (untrained) intuition doesn't always work; without the proper background very easy problems can appear quite similar to very difficult problems.

Course goals and topics

Goals

- recognize/formulate problems (such as the illumination problem) as convex optimization problem
- develop code for problems of moderate size (1000 lamps, 5000 patches)
- characterize optimal solution (optimal power distribution), give limits of performance, etc

Topics

- Convex sets, (Univariate) Functions, Optimization problem
- Unconstrained Optimization: Analysis and Algorithms
- Constrained Optimization: Analysis and Algorithms
- Optimization Algorithms for Machine Learning
- Discrete Optimization and Convexity (Eg: Submodular Minimization)
- Other Examples and applications (MAP Inference on Graphical Models, Majorization-Minimization for Non-convex problems)

Grading and Audit

Grading

- Quizzes and Assignments: 15%
- Midsem: 25%
- Endsem: 45%
- Project: 15%

Audit requirement

- Quizzes and Assignments and Project

Nonlinear optimization

traditional techniques for general nonconvex problems involve comprom **local optimization methods** (nonlinear programming)

- find a point that minimizes f_0 among feasible points near it
- fast, can handle large problems
- require initial guess
- provide no information about distance to (global) optimum

global optimization methods

- find the (global) solution
- worst-case complexity grows exponentially with problem size

these algorithms are often based on solving convex subproblems

Brief history of convex optimization

theory (convex analysis): ca1900–1970

algorithms

- 1947: simplex algorithm for linear programming (Dantzig)
- 1960s: early interior-point methods (Fiacco & McCormick, Dikin, . . .)
- 1970s: ellipsoid method and other subgradient methods
- 1980s: polynomial-time interior-point methods for linear programming (Karmarkar 1984)
- late 1980s–now: polynomial-time interior-point methods for nonlinear convex optimization (Nesterov & Nemirovski 1994)

applications

- before 1990: mostly in operations research; few in engineering
- since 1990: many new applications in engineering (control, signal processing, communications, circuit design, . . .); new problem classes (semidefinite and second-order cone programming, robust optimization)