1. Introduction

- mathematical optimization
- least-squares and linear programming
- convex optimization
- example
- course goals and topics
- nonlinear optimization
- brief history of convex optimization

Mathematical optimization

(mathematical) optimization problem

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

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

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

EVERY PROBLEM CAN BE POSED AS AN OPTIMIZATION PROBLEM!

① Given a set C (or polygon C), find the ellipsoid $\mathcal{E}$ that is of smallest volume s.t. $C \subseteq \mathcal{E}$

Think in 2 dimensions if that helps

Expression for $\mathcal{E}$?

Expression for sphere S in n-dimensions centered at O

$$S = \{ u \in \mathbb{R}^n \mid \|u\|_2 \leq r \} \qquad \|u\|_2 = \left(\sum_{i=1}^{n} u_i^2\right)^{1/2}$$

$$\mathcal{E} = \{ v \in \mathbb{R}^n \mid Av + b \in S \} = \{ v \in \mathbb{R}^n \mid \|Av+b\|_2 \leq 1 \}$$

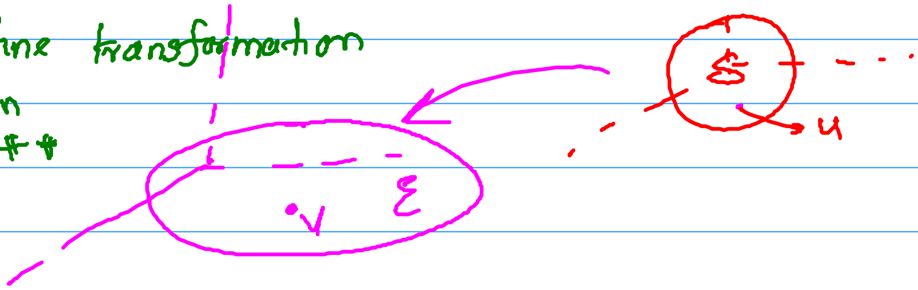
b is nx1 vector

A is nxn matrix

affine transformation

$A \in S^n_{++}$

i.e all eigenvalues of A are strictly positive



$x = [A, b] = [a_{11}\ a_{12} \ldots a_{1n}, a_{21}, a_{22} \ldots a_{2n} \ldots a_{n1}\ a_{n2} \ldots a_{nn}, b_1, b_2 \ldots b_n]$

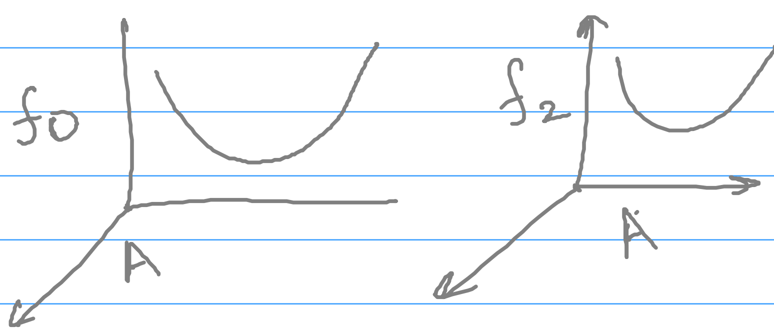
$f_0(x) \propto \det(A^{-1})$

$f_1(x) = A \in S^n_{++}$ i.e $r^T A r > 0 \quad \forall r \neq 0$

$f_2(x) = \forall v \in C \quad \|Av+b\|_2 \leq 1$

If C is a polygon, consider corners $v_1\, v_2 \ldots v_k$ of C

& $f_i(x) = \|Av_i + b\|_2 - 1 \leq 0$



H/W: How would you express optimization problem to find ellipsoid $\mathcal{E}$ of largest volume that fits inside C?

Examples

portfolio optimization

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

device sizing in electronic circuits

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

data fitting → Machine learning

- variables: model parameters
- constraints: prior information, parameter limits
- objective: measure of misfit or prediction error

$\min_{w} L_D(w) + C_D(w)$

L_D = loss component

D = data set

C_D = complexity of $\{w\}$ = prior belief . . .

5 coin tosses . . all heads

$\|p_H, p_T\|_2$

$\|[p_H, p_T]\| = \|[1, 0]\| = 1$

$\|[\hat{p}_H, \hat{p}_T]\|_2 = \|[1/2, 1/2]\|_2 = \frac{1}{\sqrt{2}}$

Solving optimization problems

Another formulation

$\min_{w} L_D(w)$

s.t. $C_D(w) \le \theta$

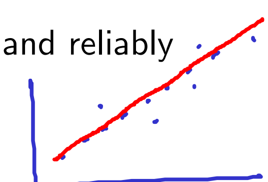
general optimization problem

- 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

Not prerequisite.

- least-squares problems → $D = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_n & y_n \end{bmatrix}$
- linear programming problems
- convex optimization problems



Find line s.t. you minimize sum of squares of distances of pts from line

Least-squares

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

s.t $\|x\|_2 \leq \theta$ → soln is slight variant

solving least-squares problems

- analytical solution: $x^\star = (A^TA)^{-1}A^Tb$ → H/w: Geometric interpretation
- reliable and efficient algorithms and software
- computation time proportional to n^2k ($A \in \mathbf{R}^{k\times n}$); 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)

Linear programming

c_i = cost of ith ingredient

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & a_i^Tx \geq b_i, \quad i = 1,\ldots,m \end{array}$$

$[x_1 \ldots x_n]$ = amnt of each of n ingredients to be purchased

each constraint corresponds to some mineral/protien/carb/vitamin constraint

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 programming

- 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

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

- objective and constraint functions are convex:

$$f_i(\alpha x + \beta y) \le \alpha f_i(x) + \beta f_i(y)$$

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

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

solving convex optimization problems

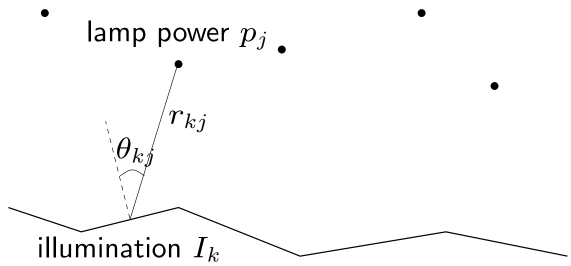
- no analytical solution
- reliable and efficient algorithms
- computation time (roughly) proportional to $\max\{n^3, n^2 m, F\}$, where F is cost of evaluating f_i's and their first and second derivatives
- almost a technology

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:

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

problem: achieve desired illumination I_{des} with bounded lamp powers

$$\begin{array}{ll} \text{minimize} & \max_{k=1,\ldots,n} |\log I_k - \log I_{\mathsf{des}}| \\ \text{subject to} & 0 \le p_j \le p_{\mathsf{max}}, \quad j = 1, \ldots, m \end{array}$$

how to solve?

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

   $$\text{minimize} \quad \sum_{k=1}^{n} (I_k - I_{\mathsf{des}})^2$$

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

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

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

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

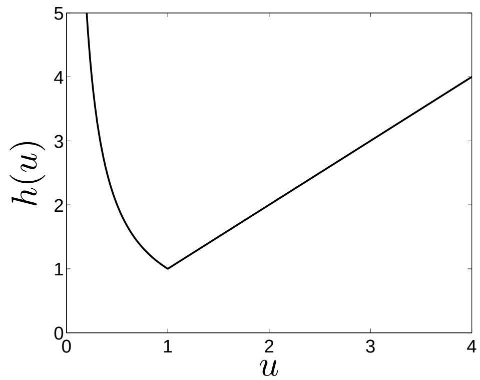
   which can be solved via linear programming

of course these are approximate (suboptimal) ‘solutions’

5. use convex optimization: problem is equivalent to

$$\begin{array}{ll} \text{minimize} & f_0(p) = \max_{k=1,\ldots,n} h(I_k/I_{\text{des}}) \\ \text{subject to} & 0 \le p_j \le p_{\text{max}}, \quad j = 1,\ldots,m \end{array}$$

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



f_0 is convex because maximum of convex functions is convex

exact solution obtained with effort $\approx$ modest factor $\times$ least-squares effort

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

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

topics

1. convex sets, functions, optimization problems
2. examples and applications
3. algorithms

Nonlinear optimization

traditional techniques for general nonconvex problems involve compromises

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)