Convex Optimization

TTIC 31070 / CMSC 35470 / BUSF 36903 / STAT 31015

Prof. Nati Srebro

Lecture 1:

Optimization Problems

Optimization Problems

$$
\begin{array}{llll}
(P) & \min\limits_{x \in \mathbb{R}^n} & f_0(x) & \\
 & s.t. & f_i(x) \leq b_i & i = 1 \ldots m
\end{array}
$$

$$f_0, f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$$

Examples:

- Minimize cost (maximize profit) while achieving goals
- Find maximum likelihood parameters
- Minimize error of model on data
- Find minimum energy configuration

Optimization Problems

$$(P) \qquad \min_{x \in \mathbb{R}^n} \quad f_0(x)$$
$$s.t. \quad f_i(x) \le b_i \qquad i = 1 \dots m$$

$$f_0, f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$$

- Def: $x \in \mathbb{R}^n$ is feasible for (P) iff it satisfies $f_i(x) \le b_i \; \forall_{i=1\dots m}$ and $\underbrace{x \in dom(f_0)}_{f_0(x) < \infty}$
- Def: The optimal value of (P) is:
$$p^* = \inf \{f_0(x) \mid f_i(x) \le b_i \; \forall_{i=1\dots m}\}$$
- Def: $x^* \in \mathbb{R}^n$ is an optimum (aka optimal point) if it is feasible and $f_0(x^*) = p^*$

$\min \quad x \log(x)$	$\min \quad -\log(x)$ $s.t. \quad x \le 2$	$\min \quad \log(x^2+1)$	$\min \quad [\lvert x \rvert - 1]_+$
$(0, \infty)$ feasible	$(0,2]$ feasible	$\mathbb{R}$ feasible	$\mathbb{R}$ feasible
$p^* = -1/e$	$p^* = \log(2)$	$p^* = 0$	$p^* = 0$
$x^* = 1/e$	$x^* = 2$	$x^* = 0$	$x^* \in [-1, +1]$

Optimization Problems

$$(P) \qquad \min_{x\in\mathbb{R}^n} \quad f_0(x)$$
$$s.t. \quad f_i(x) \le b_i \qquad i = 1 \dots m$$

$$f_0, f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$$

- Def: $x \in \mathbb{R}^n$ is feasible for (P) iff it satisfies $f_i(x) \le b_i \ \forall_{i=1\dots m}$ and $\underbrace{x \in dom(f_0)}_{f_0(x) < \infty}$
- Def: The optimal value of (P) is:
$$p^* = \inf\{f_0(x) \mid f_i(x) \le b_i \ \forall_{i=1\dots m}\}$$
- Def: $x^* \in \mathbb{R}^n$ is an optimum (aka optimal point) if it is feasible and $f_0(x^*) = p^*$
- Def: We say (P) is infeasible, and $p^* = \infty$, if no point $x \in \mathbb{R}^n$ is feasible
- Def: We say (P) is unbounded from below if $p^* = -\infty$

$$\boxed{\begin{array}{ll} \min & \log(x-5) \\ s.t. & x \le 2 \end{array}}$$

infeasible
$p^* = \infty$

$$\boxed{\begin{array}{ll} \min & x \\ s.t. & x \le -1 \\ & -x \le -1 \end{array}}$$

infeasible, $p^* = \infty$

$$\boxed{\begin{array}{ll} \min & 5 - x^2 \\ s.t. & x \ge 0 \end{array}}$$

$[0, \infty)$ feasible
$p^* = -\infty$

$$\boxed{\begin{array}{ll} \min & 1/x \\ s.t. & x \ge 3 \end{array}}$$

$(3, \infty)$ feasible
$p^* = 0$
no x^*

Example: Lemonade Stand

$$\text{profit}(x) = (x-1)100e^{-5x}$$

$$\min_x f(x) \qquad\qquad f(x) = -(x-1)100e^{-5x}$$

$$0 = f'(x^*) = -100\left(e^{-5x} - 5e^{-5x}(x-1)\right) = -100(6-5x^*)e^{-5x}$$

➔ $x^* = 1.2, p^* = -0.0496$

Example: Least Squares

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} (\langle a_i, x \rangle - b_i)^2 = \|A^\top x - b\|^2$$

- Data: $a_i \in \mathbb{R}^n, b_i \in \mathbb{R}$ ($A = [a_1, \dots, a_m] \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^m$)
- Optimization variable: x

$0 = \nabla f(x^*) = 2A(A^\top x^* - b)$ ➔ $AA^\top x^* = Ab$ ➔ $x^* = (AA^\top)^{-1} Ab$

Example: ℓ_1 Regression

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{m} |\langle a_i, x\rangle - b_i| = \|A^\top x - b\|_1$$

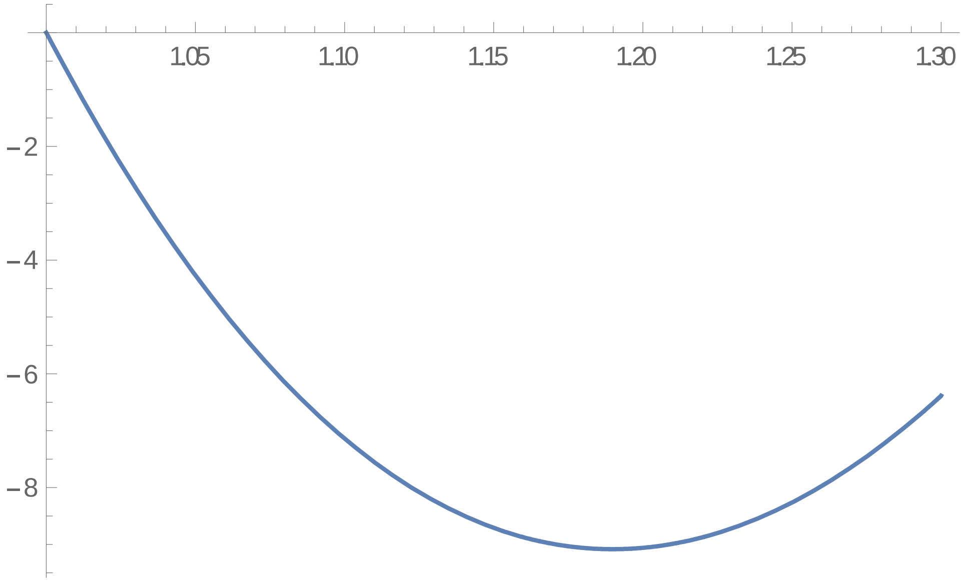
- Rewrite as a linear program:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} & \sum_{i=1}^m z_i \\ s.t. & -z_i \leq \langle a_i, x\rangle - b_i \leq z_i \quad i = 1..m \end{array}$$

Example

$$
\begin{aligned}
\min \quad & 100(1-x)(1-3\log x) \\
s.t. \quad & x \geq 1 \\
& x \leq 1.3
\end{aligned}
$$

$$x^* = 1.18098 \ldots$$



Optimization Problems

$$(P) \qquad \min_{x \in \mathbb{R}^n} \quad f_0(x)$$
$$s.t. \qquad f_i(x) \leq b_i \qquad i = 1..m$$

$$f_0, f_1, \dots, f_m: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$$

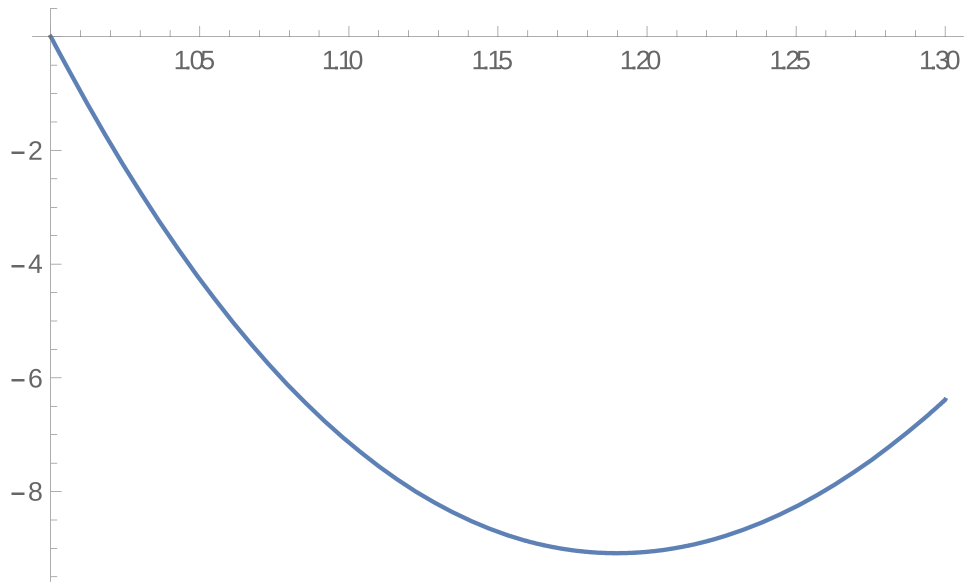
- Def: $x \in \mathbb{R}^n$ is feasible for (P) iff it satisfies $f_i(x) \leq b_i \; \forall_{i=1\dots m}$ and $x \in dom(f_0)$
- Def: The optimal value of (P) is:
$$p^* = \inf \{f_0(x) \mid f_i(x) \leq b_i \; \forall_{i=1\dots m}\}$$
or $p^* = \infty$ if no feasible point exists
- Def: $x^* \in \mathbb{R}^n$ is an optimum (aka optimal point) if it is feasible and $f_0(x^*) = p^*$

- **Def: $x \in \mathbb{R}$ is ϵ-suboptimal if it is feasible and $f_0(x) \leq p^* + \epsilon$, i.e.**
$$\forall_{\text{feasible } x'} f_0(x) \leq f_o(x') + \epsilon$$

How did I find an ϵ-suboptimum?

$$\begin{array}{ll} \min & 100(1-x)(1-3\log x) \\ s.t. & x \geq 1 \\ & x \leq 1.3 \end{array}$$

$$x^* = 1.18098\ldots$$



Grid Search

$$\min_{x\in\mathbb{R}} \quad f_0(x)$$
$$s.t. \quad MIN \leq x \leq MAX$$

- Parameter: $\delta > 0$
- Method: Evaluate $f_0(x)$ at
$$x \in \{MIN, MIN+\delta, MIN+2\delta, \dots, MIN + \left\lfloor \frac{MAX-MIN}{\delta} \right\rfloor \delta\}$$
returning minimum

- Analysis: We will always have $\tilde{x}$ in the grid with $|\tilde{x} - x^*| \leq \delta$, and so:
$$|f_0(\tilde{x}) - f_0(x^*)| \leq |x - x^*| \cdot |f'(\overline{x})| \leq \delta \cdot D$$
- For $\hat{x}$ return by Grid Search we have:
$$f_0(\hat{x}) \leq f_0(\tilde{x}) \leq f_0(x^*) + \delta D$$
- Conclusion: If $\forall_{MIN \leq x \leq MAX} |f_0'(x)| \leq D$, and we use $\delta = \frac{\epsilon}{D}$, we can find an ϵ-suboptimal solution using $\frac{(MAX-MIN)D}{\epsilon}$ evaluations.

Grid Search

- Only depends on specific forms of access (*oracles*) to f, not on the form of the function
  - In this case: evaluation oracle $x \mapsto f(x)$
  - Later on, also $x \mapsto \nabla f(x)$, $x \mapsto \nabla^2 f(x)$, others
- Runtime guarantee (on #access and operations) in terms of specific assumptions / quantities, and desired ϵ
  - In this case: $|f'| \leq D$ (Lipschitz assumption)

- But, disappointing runtime:
  - $O\left(\frac{1}{\epsilon}\right)$ means exponential in #digits of precision
  - In higher dimension, grid of size $\left(\frac{MAX-MIN}{\delta}\right)^n$ ensures $\|x - x^*\| \leq \delta\sqrt{n}$ ➔ runtime is $\left(\frac{MAX-MIN}{\epsilon}\sqrt{n}D\right)^n$
- Can't do any better without more assumptions:

  Theorem: for any ϵ and any algorithm making $< {}^{1}/_{3\epsilon}$ evaluation queries, there exists a function $f: [0,1] \to \mathbb{R}$, with $|f'| \leq 1$ for which the algorithm fails to find a ϵ-suboptimal solution.

Bisection Search

$$\min_{x \in \mathbb{R}} \quad f(x)$$
$$s.t. \quad MIN \leq x \leq MAX$$

- Assume $|f'| \leq D$ and f is convex
- Access to $f(x)$, $f'(x)$

Init: $x_L^{(0)} = MIN, x_H^{(0)} = MAX$

Iter: $x^{(k)} = \frac{x_L^{(k)} + x_H^{(k)}}{2}$

If $f'\left(x^{(k)}\right) = 0$, stop

If $f'\left(x^{(k)}\right) < 0$: $\quad x_L^{(k+1)} \leftarrow x^{(k)}$, $\quad x_H^{(k+1)} \leftarrow x_H^{(k)}$

If $f'\left(x^{(k)}\right) > 0$: $\quad x_L^{(k+1)} \leftarrow x_L^{(k)}$, $\quad x_H^{(k+1)} \leftarrow x^{(k)}$

Bisection Search

- Claim: If $f(x)$ is convex and $\forall_{MIN \le x \le MAX} f'(x) \le D$, then
$$f\left(x^{(k)}\right) \le p^* + D(MAX - MIN)2^{-k}$$

- Conclusion: #iterations, and therefor #evals and runtime, to find ϵ-suboptimal solution:
$$O\left(\log\left(\frac{MAX - MIN}{\epsilon} D\right)\right)$$

Bisection Search

$$\min_{x \in \mathbb{R}} \quad f(x)$$
$$s.t. \quad MIN \le x \le MAX$$

Init: $x_L^{(0)} = MIN, x_H^{(0)} = MAX$

Iter: $x^{(k)} = \frac{x_L^{(0)} + x_H^{(0)}}{2}$

If $\left\| x_L^{(k)} - x_H^{(k)} \right\| \le \frac{\epsilon}{D}$, stop

If $f'\left(x^{(k)}\right) = 0$, stop

If $f'\left(x^{(k)}\right) < 0$: $x_L^{(k+1)} \leftarrow x^{(k)}$
$x_H^{(k+1)} \leftarrow x_H^{(k)}$

If $f'\left(x^{(k)}\right) > 0$: $x_L^{(k+1)} \leftarrow x_L^{(k)}$
$x_H^{(k+1)} \leftarrow x^{(k)}$

Convex Optimization Problems

$$
\begin{array}{llll}
(P) & \min\limits_{x \in \mathbb{R}^n} & f_0(x) & \\
 & s.t. & f_i(x) \leq b_i & i = 1 \ldots m
\end{array}
$$

- Def: (P) is a <u>convex optimization problem</u> if $f_0, f_1, \ldots, f_m$ are convex functions

- **In this course:** methods for solving convex optimization problems of form (P), based on oracle access to $f_0, f_1, \ldots, f_m$, with guarantees based on their properties

Optimization vs Solving Equations

$$(P) \qquad \min_{x \in \mathbb{R}^n} f(x)$$

- Claim (Optimality Condition for Unconstrained Optimization):
  If $f(x)$ is convex and differentiable in its domain, then

  $$x^* \text{ is optimal iff } \nabla f(x^*) = 0$$

- Conclusion:

  $$\text{Minimizing } f(x) \Leftrightarrow \text{solving } \nabla f(x) = 0$$

- As we shall see later—also for constrained optimization

About the Course

- Methods for solving convex optimization problems, based on oracle access, with guarantees based on their properties
  - And also a few more specific methods…
- Understanding different optimization methods
  - Understanding their derivation
  - When are they appropriate
  - Guarantees (a few proofs, not a core component)
- Working and reasoning about optimization problems
  - Optimality conditions
  - Duality
  - Standard forms: LP, QP, SDP, etc

- Prerequisites:
  - Linear Algebra (vector fields, linear transformations, matrices, eigenvalues)
  - Multi-dimensional Calculus (gradients, Hessians, partial derivatives, directional derivatives)
  - Some background in Algorithms (runtime analysis, proving correctness of an algorithm), and programming

Course Structure

TAs: Blake Woodworth (head), Haoyang Liu, Greg Naitzat, Angela Wu

Contact us: `convex-optimization-2018-staff@ttic.edu`

Communication, homework, lecture slides, help forum via course page on canvas

If not registered, please complete webform

- Lectures Mondays and Wednesdays 12:05PM
- Recitations (choose one): Monday 4-5PM TTIC 530
OR Tuesday 4:30-5:30 Ryerson 276
- Homeworks due every Friday (50% of the grade)

  Some Python programming, mostly completing provided code (can use other languages, eg MATLAB, R, Julia, etc, if you prefer—no code or support provided)
- TA office hours: Tuesday, Wednesday, Thursday and Friday (see website)
- 7-8 homeworks (50% of grade), final (50% of grade)
- Books:
  - Boyd and Vandenberghe "Convex Optimization" (about 70% of the class)
  - Nocedal and Wright "Numerical Optimization"
  - Nemirovski "Efficient Methods in Convex Programming"

We are looking for additional graders

Homework #1

- Available now, due next Friday 1/12, back 1/17
- Material covered: this lecture + Monday's lecture (convexity)
- Used also to evaluate control of prerequisites and readiness to take class
  - A,B: satisfactory performance; prepared to take the class
  - C: borderline; may take class but should consider difficulty and/or preparation
  - Below C: advised to not take class this quarter
- For this homework only: no collaboration on required questions (OK to collaborate on future homework)