Convex Optimization
TTIC 31070 / CMSC 35470 / BUSF 36903 / STAT 31015
Prof. Nati Srebro

Lecture 1: Optimization Problems
Optimization Problems

\[(P) \quad \min_{x \in \mathbb{R}^n} f_0(x) \]

s.t.

\[f_i(x) \leq b_i \quad i = 1 \ldots m\]

\[f_0, f_1, \ldots, f_m : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}\]

Examples:

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

\[ (P) \quad \min_{x \in \mathbb{R}^n} f_0(x) \]

\[ \text{s.t.} \quad f_i(x) \leq b_i \quad i = 1 \ldots m \]

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

- Def: \( x \in \mathbb{R}^n \) is **feasible** for \((P)\) iff it satisfies \( f_i(x) \leq b_i \ \forall i = 1 \ldots m \) and \( x \in \text{dom}(f_0) \)

- Def: The **optimal value** of \((P)\) is:

\[ p^* = \inf \{ f_0(x) \mid f_i(x) \leq b_i \ \forall i = 1 \ldots m \} \]

- Def: We say \((P)\) is **infeasible**, and \( p^* = \infty \), if no point \( x \in \mathbb{R}^n \) is feasible

- 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 **unbounded from below** if \( p^* = -\infty \)
Examples

\[ \begin{align*}
\text{min} & \quad x \\
\text{s. t.} & \quad x \leq -1 \\
& \quad -x \leq -1
\end{align*} \]

Infeasible, \( p^* = \infty \)

\[ \begin{align*}
\text{min} & \quad 5 - x^2 \\
\text{s. t.} & \quad x \geq 0
\end{align*} \]

Unbounded, \( p^* = -\infty \)

\[ \begin{align*}
\text{min} & \quad x \log(x) \\
\text{finite} & \quad p^* = -1/e \\
\text{unique} & \quad x^* = 1/e
\end{align*} \]
Example: Lemonade Stand

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

\[ \min_x f(x) \quad f(x) = -(x - 1)100e^{-5x} \]

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

\[ \Rightarrow x^* = 1.2, \quad p^* = -0.0496 \]
Example: Least Squares

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

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

\[
0 = \nabla f(x^*) = 2A^T(Ax^* - b) \Rightarrow A^T Ax^* = A^T b \Rightarrow x^* = (A^T A)^{-1} A^T b
\]
Example: $\ell_1$ Regression

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

- Rewrite as a linear program:

$$\begin{align*}
\min & \quad \sum_{i=1}^{m} z_i \\
\text{s.t.} & \quad -z_i \leq \langle a_i, x \rangle - b_i \leq z_i \quad i = 1..m
\end{align*}$$
Example

\[
\begin{align*}
\min & \quad 100(1 - x)(1 - 3 \log x) \\
\text{s.t.} & \quad x \geq 1 \\
& \quad x \leq 1.3
\end{align*}
\]

\[x^* = 1.18098 \ldots\]
Optimization Problems

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

\[ f_0, f_1, ..., f_m : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \]

- **Def:** \( x \in \mathbb{R}^n \) is **feasible** for (P) iff it satisfies \( f_i(x) \leq b_i \ \forall i=1...m \) and \( x \in \text{dom}(f_0) \)
- **Def:** The **optimal value** of (P) is:
  \[ p^* = \inf \{ f_0(x) \mid f_i(x) \leq b_i \ \forall i=1...m \} \]
  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 **\( \epsilon \)-suboptimal** if it is feasible and \( f_0(x) \leq p^* + \epsilon \), i.e.
  \[ \forall \text{feasible } x', f_0(x) \leq f_0(x') + \epsilon \]
How did I find a $10^{-5}$-suboptimum?

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

$x^* = 1.18098 \ldots$
Grid Search

\[
\min_{x \in \mathbb{R}} f(x) \\
\text{s.t. } \ MIN \leq x \leq MAX
\]

• Parameter: \( \delta > 0 \)
• Method: Evaluate \( f(x) \) at
  \[x \in \{MIN, MIN + \delta, MIN + 2\delta, \ldots, MIN + \left\lfloor \frac{MAX - MIN}{\delta} \right\rfloor \delta\}\]
  returning minimum

• Analysis: We will always have \( x \) in the grid with \( |x - x^*| \leq \delta \), and so:
  \[|f(x) - f(x^*)| \leq |x - x^*| \cdot |f'(\tilde{x})| \leq \delta \cdot D\]
• Conclusion: If \( \forall_{MIN \leq x \leq MAX} |f'(x)| \leq D \), and we use \( \delta = \frac{\epsilon}{D} \), we can find an \( \epsilon \)-suboptimal solution using at most \( \frac{(MAX-MIN)D}{\epsilon} \) 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 $\epsilon$
  • 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 $\left(\frac{\text{MAX} - \text{MIN}}{\delta}\right)^n$ ensures $\|x - x^*\| \leq \delta \sqrt{n}$ \implies runtime is $\left(\frac{\text{MAX} - \text{MIN}}{\epsilon} \sqrt{nD}\right)^n$

• Can’t do any better without more assumptions
Bisection Search

\[
\begin{align*}
\min_{x \in \mathbb{R}} & \quad f(x) \\
\text{s.t.} & \quad \text{MIN} \leq x \leq \text{MAX}
\end{align*}
\]

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

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

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

- If \(f'(x^{(k)}) = 0\), stop
- If \(f'(x^{(k)}) < 0\): \(x_L^{(k+1)} \leftarrow x^{(k)}\)
  \(x_H^{(k+1)} \leftarrow x_H^{(k)}\)
- If \(f'(x^{(k)}) > 0\): \(x_L^{(k+1)} \leftarrow x_L^{(k)}\)
  \(x_H^{(k+1)} \leftarrow x^{(k)}\)
Bisection Search

• Claim: If $f(x)$ is convex and $\forall_{MIN \leq x \leq MAX} f'(x) \leq D$, then
  $$f(x^{(k)}) \leq p^* + D2^{-k(MAX-MIN)}$$

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

$$\min_{x \in \mathbb{R}} f(x)$$

s.t. $MIN \leq x \leq MAX$

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

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

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

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

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

$x_H^{(k+1)} \leftarrow x_H^{(k)}$

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

$x_H^{(k+1)} \leftarrow x^{(k)}$
Convex Optimization Problems

\[
\begin{align*}
(P) \quad \min_{x \in \mathbb{R}^n} & \quad f_0(x) \\
\text{s.t.} & \quad f_i(x) \leq b_i \quad i = 1 \ldots m
\end{align*}
\]

• Def: \((P)\) is a **convex optimization problem** if \(f_0, f_1, \ldots, f_m\) are convex functions

• Claim: \((P)\) is convex \(\Rightarrow\) feasible set is convex
  
  Proof sketch: sublevel sets \(\{x | f_i(x) \leq b_i\}\) are convex, and intersections of convex sets is convex

• Converse is not true!

• **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
Optimality Condition: Unconstrained Optimization

\[(P) \quad \min_{x \in \mathbb{R}^n} f(x)\]

- Claim: If \(f(x)\) is convex and differentiable in its domain, then:
  \[x^* \text{ is optimal iff } \nabla f(x^*) = 0\]

Sketch of \(\Leftarrow\): \[f(x) \geq f(x^*) + \langle x - x^*, \nabla f(x^*) \rangle\]

Minimizing \(f(x)\) \(\iff\) solving \(\nabla f(x) = 0\)

- In fact, even if not convex...

\textit{next lecture}
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 opt 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

Lecturers: Prof. Nati Srebro + Prof. Ofer Meshi
TA: Behnam Neyshabur

• Lectures Tuesdays and Thursdays
• Recitations Wednesday(?)
• 7-8 homeworks (50% of grade)
  • Due Fridays
  • No late homeworks accepted without prior arrangement
  • Some Python programming, mostly completing provided code (can use other languages, eg MATLAB, R, Julia, etc, if you prefer—no code or support provided)

• Final (50% of grade)
• Books