Introduction to Optimization Theory

Lecture #11 - 10/19/20
MS&E 213 / CS 2690

Aaron Sidford
sidford@stanford.edu
Plan for Today

Motivation
- Recap where we are
- Motivate next unit

Convex sets
- Another perspective on convex functions

Oracles
- Structure of convex sets

Hyperplanes and Subgradients
## Recap

<table>
<thead>
<tr>
<th>Regularity</th>
<th>Oracle</th>
<th>Goal</th>
<th>Algorithm</th>
<th>Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 1, f(x) \in [0,1], x_\ast \in [0,1]$</td>
<td>value</td>
<td>$\frac{1}{2}$-optimal</td>
<td>anything</td>
<td>$\infty$</td>
</tr>
<tr>
<td>$n = 1, x_\ast \in [0,1], L$-Lipschitz</td>
<td>value</td>
<td>$\epsilon$-optimal</td>
<td>$\epsilon$-net</td>
<td>$\Theta(L/\epsilon)$</td>
</tr>
<tr>
<td>$x_\ast \in [0,1], L$-Lipschitz $| \cdot |_\infty$</td>
<td>value</td>
<td>$\epsilon$-optimal</td>
<td>$\epsilon$-net</td>
<td>$(\Theta(L/\epsilon))^n$</td>
</tr>
<tr>
<td>$L$-smooth and bounded</td>
<td>value, gradient</td>
<td>$\epsilon$-optimal</td>
<td>$\epsilon$-net</td>
<td>exponential</td>
</tr>
<tr>
<td>$L$-smooth</td>
<td>gradient</td>
<td>$\epsilon$-critical</td>
<td>gradient descent</td>
<td>$O(L(f(x_0) - f_\ast)\epsilon^{-2})$</td>
</tr>
<tr>
<td>$L$-smooth $\mu$-strongly convex</td>
<td>gradient</td>
<td>$\epsilon$-optimal</td>
<td>gradient descent</td>
<td>$O((L/\mu) \log((f(x_0) - f_\ast)/\epsilon))$</td>
</tr>
<tr>
<td>$L$-smooth convex</td>
<td>gradient</td>
<td>$\epsilon$-optimal</td>
<td>gradient descent</td>
<td>$O(L|x_0 - x_\ast|_2^2/\epsilon)$</td>
</tr>
<tr>
<td>$L$-smooth $\mu$-strongly convex</td>
<td>gradient</td>
<td>$\epsilon$-optimal</td>
<td>gradient descent</td>
<td>$O(L\sqrt{L/\mu} \log((f(x_0) - f_\ast)/\epsilon))$</td>
</tr>
<tr>
<td>$L$-smooth $\mu$-strongly convex</td>
<td>gradient</td>
<td>$\epsilon$-optimal</td>
<td>gradient descent</td>
<td>$O\left(L|x_0 - x_\ast|_2^2/\epsilon\right)$</td>
</tr>
</tbody>
</table>
How?

**ε-net**
- Check enough points to cover optimal points
- Check random points

**Local Greedy**
- Iteratively, locally decrease function value

\[ x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k) \]

\[ x_{k+1} = \text{argmin}_{x \in U_k(x)} \]

\[ U_k(x_k) = f(x_k) \quad \text{and} \quad U_k(x) \geq f(x) \quad \text{for all} \quad x. \]

**Acceleration**
- Combine upper and lower bounds
- Is there a more general lower bound phenomena?
Next Few Weeks

• What if function is non-differentiable?
• What if function is very non-smooth?
• What if cannot make sufficient local progress?

Idea
Develop new potential functions!
Develop new notions of progress!
Develop new methods!
Many Examples

**Max Functions**
- \( \min_{x \in \mathbb{R}^n} \max_{i \in [m]} f_i(x) \)
- Can solve if \( f_i \) are smooth and convex.
- *What if many of them? (m large)*

**Ill Conditioned Problem**
- \( \min_{x \in \mathbb{R}^n} \frac{1}{2} \| Ax - b \|_2^2 + \lambda \| x \|_1 \)
- Can solve if \( L \)-smooth and \( \mu \)-strongly convex
- *What if \( L/\mu \gg n^c \)?*
Linear Programming

Input
• $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$

Goal
• $\min_{x \in \mathbb{R}^n} c^T x$ for $P \overset{\text{def}}{=} \{x : Ax \geq b\}$
• $= \min_{x \in \mathbb{R}^n} c^T x + \psi_P(x)$ for $\psi_P(x) \overset{\text{def}}{=} \begin{cases} 0 & Ax \geq b \\ \infty & \text{otherwise} \end{cases}$
The Picture

Polytope

\[ A \mathbf{x} \geq b \]

\[
\begin{pmatrix}
- & a_1 & - \\
- & a_2 & - \\
\vdots & \vdots & \vdots \\
- & a_k & - \\
- & a_m & - \\
\end{pmatrix} \mathbf{x} \geq 
\begin{pmatrix}
b_1 \\
b_2 \\
\vdots \\
b_k \\
b_m \\
\end{pmatrix}
\]

\[ c^T \mathbf{x} \]

(Closed) Half-space

half(\(a_i, b_i\)) \(\overset{\text{def}}{=} H_x (a_i, b_i) \overset{\text{def}}{=} \{ x \in \mathbb{R}^n : a_i^T x \geq b_i \} \]

\[ \min_{x \in \mathbb{R}^n : Ax \geq b} c^T x \]

Solution: \(x_\star\)

What to do here?
Our Approach

**Step #1**
- Obtain a better understanding of convex sets
- Connect convex set structure to convex function structure

**Step #2**
- Consider different oracles for convex functions
  - Subgradient oracle and subgradient methods
  - Separation oracle and cutting plane methods
  - Barrier oracle and interior point methods

**Step #3**
- Have a good winter break!
- Along the way we will learn
  - Online learning, SGD, Newton’s method, and more!
Plan for Today

Motivation
- Recap where we are
- Motivate next unit

Convex sets
- Another perspective on convex functions

Oracles
- Structure of convex sets

Hyperplanes and Subgradients
Convex Set

Definition: a set $S \subseteq \mathbb{R}^n$ is convex if and only if for all $x, y \in S$ and $t \in [0,1]$ we have $tx + (1 - t)y \in S$.

- “contains the line segment between every pair of points”
- “closed under convex combinations”
Convexity Examples and Properties

**Lemma**: if $C$ is a set (possibly infinite) of convex sets in $\mathbb{R}^n$ then $\bigcap_{S \in C} S$ is convex

**Proof**: $x, y \in \bigcap_{S \in C} S$ implies that $tx + (1 - t)y \in S$ for all $S \in C$ and $t \in [0,1]$

**Lemma**: if $S$ is convex, its closure (union of limit points) is convex

**Lemma**: for all $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ the half-space, $\text{half}(a, b) \overset{\text{def}}{=} H_{\geq}(a, b) \overset{\text{def}}{=} \{x \in \mathbb{R}^n \mid a^\top x \geq b\}$, is convex

**Corollary**: Polytopes, i.e. $\{x \in \mathbb{R}^n \mid Ax \geq b\}$, are convex

**Theorem**: all closed convex sets are intersections of (a possibly infinite) set of halfspaces.
Convex function minimization?

(sub)level set: $\text{level}_\leq(f, v) = \{x \in \mathbb{R}^n \mid f(x) \leq v\}$

strict (sub)level set: $\text{level}_<(f, v) = \{x \in \mathbb{R}^n \mid f(x) < v\}$

Note: $x$ is $\epsilon$-optimal $\iff x \in \text{level}_\leq(f, f_* + \epsilon)$

Lemma: If $f: \mathbb{R}$ convex then $\text{level}_\leq$ and $\text{level}_<$ are always convex.

Proof: if $f(x) \leq v$ and $f(y) \leq v$ then
\[
    f(t \cdot x + (1 - t) \cdot y) \leq t \cdot f(x) + (1 - t) \cdot f(y) \leq t \cdot v
\]

Optimizing a convex function $\iff$ finding a point in a convex set

Use convexity structure!

What to do here?

$x_*$

$x_k$
Convex function minimization?

(sub)level set: \( \text{level}_\leq(f, v) = \{ x \in \mathbb{R}^n \mid f(x) \leq v \} \)

strict (sub)level set: \( \text{level}_<(f, v) = \{ x \in \mathbb{R}^n \mid f(x) < v \} \)

Note: \( x \) is \( \epsilon \)-optimal \( \iff x \in \text{level}_\leq(f, f^* + \epsilon) \)

Lemma: If \( f: \mathbb{R} \) is convex then \( \text{level}_\leq \) and \( \text{level}_< \) are always convex.

Is the converse true?

No!

Quasiconvex: function with convex level sets

Problem

\[
\min_{x \in \mathbb{R}^n} f(x)
\]
Convexity and Convex Functions?

**Definition:** for $f: \mathbb{R}^n \rightarrow \mathbb{R}$ its epigraph is

$$\text{epi}(f) = \{(x, t) | x \in \mathbb{R}^n, t \in \mathbb{R}, f(x) \leq t\}$$

**Theorem:** $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function $\iff$ $\text{epi}(f)$ is a convex set

**Proof $\Rightarrow$:** Let $(x, v_x), (y, v_y) \in \text{epi}(f)$.

Convexity: $f(t \cdot x + (1 - t) \cdot y) \leq t \cdot f(x) + (1 - t) \cdot f(y)$

Definition of epigraph: $f(t \cdot x + (1 - t) \cdot y) \leq t \cdot v_x + (1 - t) \cdot v_y$

Same as: $t(x, v_x) + (1 - t)(y, v_y) \in \text{epi}(f)$
Convexity and Convex Functions?

**Definition:** for $f: \mathbb{R}^n \to \mathbb{R}$ its epigraph is
\[
epi(f) = \{(x, t) \mid x \in \mathbb{R}^n, t \in \mathbb{R}, f(x) \leq t\}
\]

**Theorem:** $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function $\iff$ $\text{epi}(f)$ is a convex set

**Proof $\Leftarrow$:** $(x, f(x)), (y, f(y)) \in \text{epi}(f)$ for all $x, y \in \mathbb{R}^n$

Convexity: $t(x, f(x)) + (1-t)(y, f(y)) \in \text{epi}(f)$

Definition of epigraph: $f(t \cdot x + (1-t) \cdot y) \leq t \cdot f(x) + (1-t) \cdot f(y)$
Plan for Today

Motivation
- Recap where we are
- Motivate next unit

Convex sets
- Another perspective on convex functions

Oracles
- Structure of convex sets

Hyperplanes and Subgradients
How obtain information about level sets?

**Idea:** Differentiable Case

- \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) convex
- \( \iff f(y) \geq f(x) + \nabla f(x)^\top (y - x) \)
- \( \Rightarrow \text{level}_{\leq}(f, f(x)) \subseteq \{ y : \nabla f(x)^\top (y - x) \leq 0 \} \)
- \( \iff \text{level}_{\leq}(f, f(x)) \subseteq H_{\geq}(-\nabla f(x), -\nabla f(x)^\top x) \)
- Is this information enough?

**Cutting Plane Methods**

- Will cover in a few weeks
- **This week:** just prove the oracle exists for quasi-convex functions
Another Idea

Idea: Differentiable Case
• \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) convex
• \( \Leftrightarrow f(y) \geq f(x) + \nabla f(x)^T (y - x) \)
• **Subgradient:** \( g \) is subgradient of \( f \) at \( x \) if \( f(y) \geq f(x) + g^T (y - x) \) for all \( y \in \mathbb{R}^n \)
• \( \partial f(x) = \{ \text{set subgradients of } f \text{ at } x \} \)

- (sub)level set: \( \text{level}_\leq(f, v) = \{ x \in \mathbb{R}^n | f(x) \leq v \} \)
- strict (sub)level set: \( \text{level}_<(f, v) = \{ x \in \mathbb{R}^n | f(x) < v \} \)

**Subgradient descent**

- **Will cover this week / next week**
- **This week:** just prove existence and relate to convexity

query
\[ x \in \mathbb{R}^n \]
\[ \text{subgradient oracle} \]
\[ g \in \partial f(x) \]
Plan for Today

Motivation
- Recap where we are
- Motivate next unit

Convex sets
- Another perspective on convex functions

Oracles
- Structure of convex sets

Hyperplanes and Subgradients