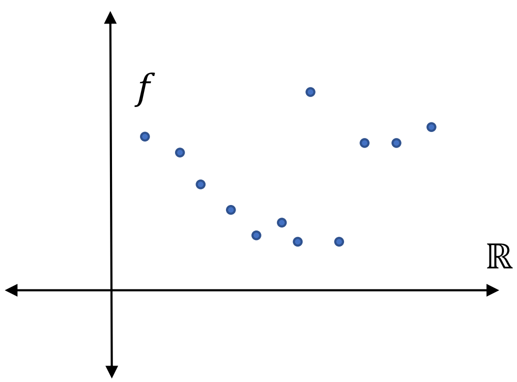
Introduction to Optimization Theory

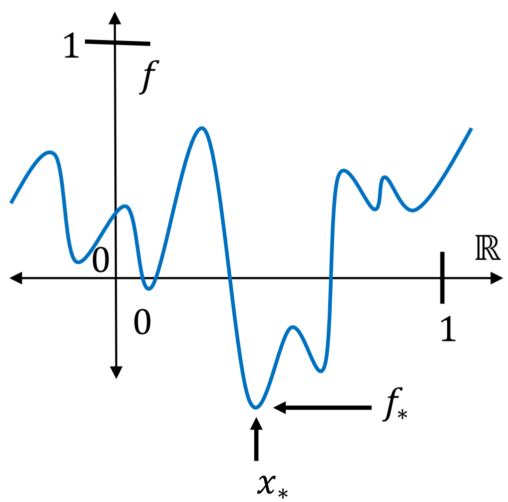
Lecture #11 - 10/19/20

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

Problem

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

Regularity	Oracle	Goal	Algorithm	Iterations
$n = 1$, $f(x) \in [0,1]$, $x_* \in [0,1]$	value	½-optimal	anything	∞
$n = 1$, $x_* \in [0,1]$, L-Lipschitz	value	ϵ-optimal	ϵ-net	$\Theta(L/\epsilon)$
$x_* \in [0,1]$, L-Lipschitz in $\|\cdot\|_\infty$	value	ϵ-optimal	ϵ-net	$\left(\Theta(L/\epsilon)\right)^n$
L-smooth and bounded	value, gradient	ϵ-optimal	ϵ-net	exponential
L-smooth	gradient	ϵ-critical	gradient descent	$O(L(f(x_0) - f_*)\epsilon^{-2})$
L-smooth μ-strongly convex	gradient	ϵ-optimal	gradient descent	$O\left((L/\mu)\log([f(x_0) - f_*]/\epsilon)\right)$
L-smooth convex	gradient	ϵ-optimal	gradient descent	$O(L\|x_0 - x_*\|_2^2/\epsilon)$
L-smooth μ-strongly convex	gradient	ϵ-optimal	gradient descent	$O\left(\sqrt{L/\mu}\log([f(x_0) - f_*]/\epsilon)\right)$
L-smooth μ-strongly convex	gradient	ϵ-optimal	gradient descent	$O\left(\sqrt{L\|x_0 - x_*\|_2^2/\epsilon}\right)$

How?

ϵ-net

- Check enough points to cover optimal points
- Check random points

Acceleration

- Combine upper and lower bounds
- Is there a more general lower bound phenomena?

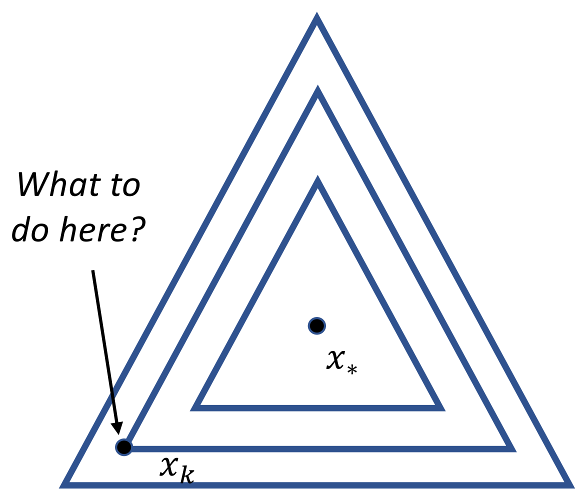
Local Greedy

- Iteratively, locally decrease function vaue
- $x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k)$
- $x_{k+1} = \underset{x}{\operatorname{argmin}}\, U_k(x)$ for where $U_k(x_k) = f(x_k)$ and $U_k(x) \geq f(x)$ for all x.

Next Few Weeks

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

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



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 μ-strongly convex
- *What if $L/\mu \gg n^c$?*

A Canonical Example

Linear Programming

Input

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

Goal

- $\min_{x \in P} c^\top x$ for $P \stackrel{\text{def}}{=} \{x : Ax \geq b\}$
- $= \min_{x \in \mathbb{R}^n} c^\top x + \psi_P(x)$ for $\psi_P(x) \stackrel{\text{def}}{=} \begin{cases} 0 & Ax \geq b \\ \infty & \text{otherwise} \end{cases}$

The Picture

(Closed) Half-space

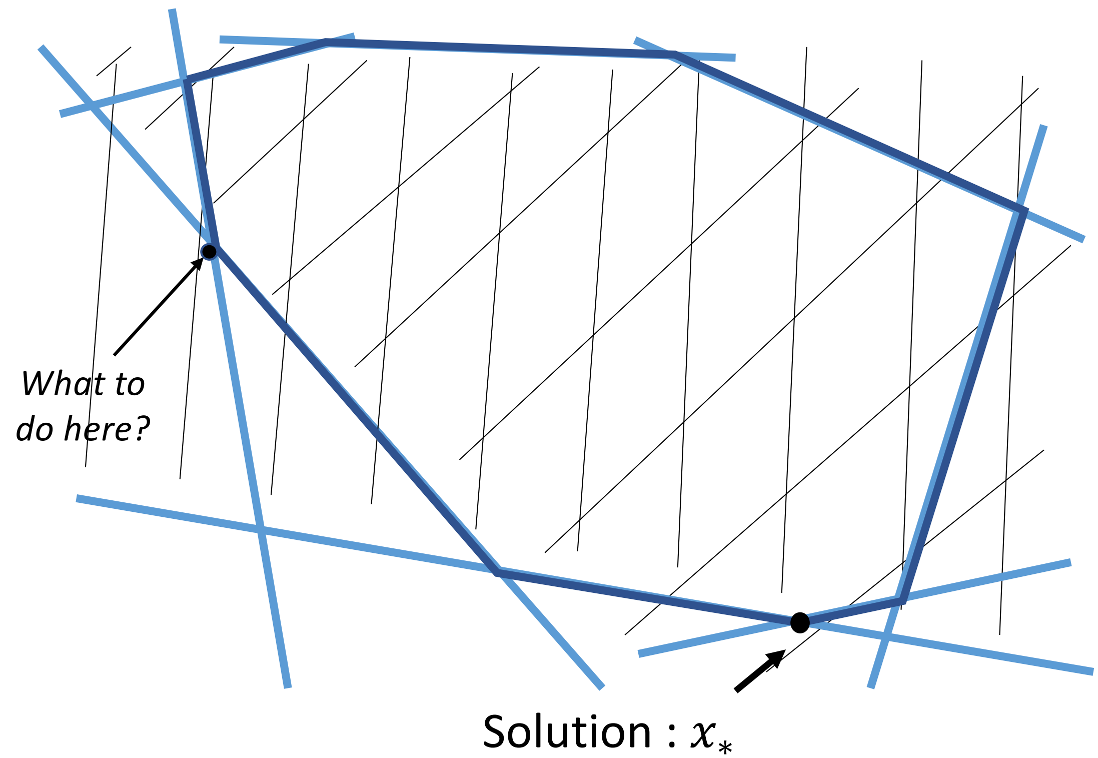
$$\mathrm{half}(a_i, b_i) \stackrel{\text{def}}{=} H_{\geq}(a_i, b_i) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : a_i^\top x \geq b_i\}$$

$$\min_{x \in \mathbb{R}^n:\, Ax \geq b} c^\top x$$

Polytope $Ax \geq b$

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

$c^\top x$



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

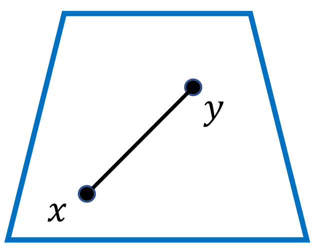
Hyperplanes and Subgradients

Convex Set

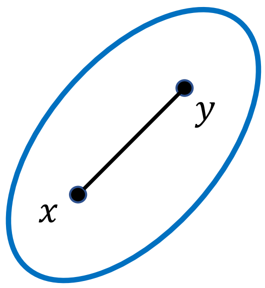
$tx + (1 - t)y$ for $t \in [0,1]$ is a "convex combination" of x and y"

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 $tx + (1 - t)y \in S$.

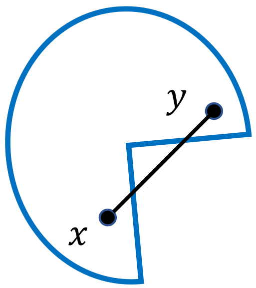
- *"contains the line segment between every pair of points"*
- *"closed under convex combinations"*



Convex



Convex



Non-convex

Convexity Examples and Properties

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

Proof: $x, y \in \cap_{S \in C} S$ 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, $\text{half}(a, b) \stackrel{\text{def}}{=} H_{\geq}(a, b) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid a^\top x \geq b\}$, 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?

Problem

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

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

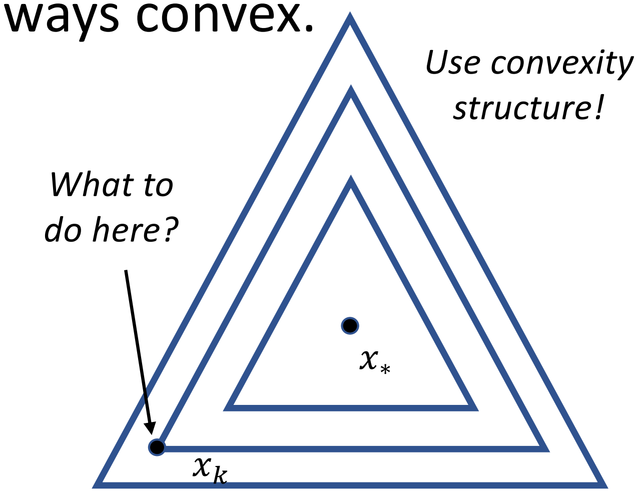
strict (sub)level set: $\text{level}_{<}(f, v) = \{x \in \mathbb{R}^n \mid f(x) < v\}$

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

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

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

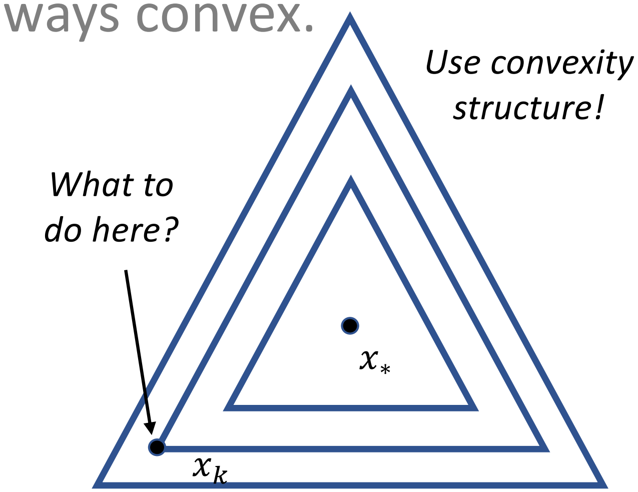


Convex function minimization?

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

(sub)level set: $\mathrm{level}_{\leq}(f, v) = \{x \in \mathbb{R}^n \mid f(x) \leq v\}$
strict (sub)level set: $\mathrm{level}_{<}(f, v) = \{x \in \mathbb{R}^n \mid f(x) < v\}$
Note: x is ϵ-optimal $\Leftrightarrow x \in \mathrm{level}_{\leq}(f, f_* + \epsilon)$

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



Is the converse true?

No!

Quasiconvex: function with convex level sets

Convexity and Convex Functions?

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

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

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

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

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: $t(x, v_x) + (1-t)(y, v_y) \in \mathrm{epi}(f)$

Convexity and Convex Functions?

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

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

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

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

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

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

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

How obtain information about level sets?

Idea: Differentiable Case

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

Cutting Plane Methods

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

query $x \in \mathbb{R}^n$ → **separation oracle** → output $g \in \mathbb{R}^n$ such that $\text{level}_{\leq}\big(f, f(x)\big) \subseteq H_{\geq}(g, g^\top x)$

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

Another Idea

<u>**Idea**</u>: Differentiable Case

- $f: \mathbb{R}^n \to \mathbb{R}$ convex
- $\Leftrightarrow f(y) \geq f(x) + \nabla f(x)^\top (y - x)$
- **Subgradient**: g is subgradient of f at x if $f(y) \geq f(x) + g^\top (y - x)$ for all $y \in \mathbb{R}^n$
- $\partial f(x) = \{\text{set subgradients of } f \text{ at } x\}$

Subgradient descent

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

<u>query</u>
$x \in \mathbb{R}^n$ → **subgradient oracle** → <u>output</u>
$g \in \partial f(x)$

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