AI-3001: Advanced Topics in Machine Learning **Aug-Nov 2022**

Lecture 4: Convex Optimization Review

Lecturer: Ganesh Ghalme *Scribes: Ganesh Ghalme*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

4.1 Convex Optimization Review

Definition 4.1 (affine combination). *Given two points $x_1, x_2 \in \mathbb{R}^d$ we define a line passing through these points by the set of points $\{x \in \mathbb{R}^d : x = \alpha x_1 + (1-\alpha)x_2, \alpha \in \mathbb{R}\}$*

Definition 4.2 (Convex combination). *Given two points $x_1, x_2 \in \mathbb{R}^d$ we define a convex combination is the set $\{x : x = \alpha x_1 + (1-\alpha)x_2, \alpha \in [0,1]\}$.*

Definition 4.3 (Convex Set). *We call a set $X \subseteq \mathbb{R}^d$convex if for all pairs $x_1, x_2 \in X$ we have $\alpha x_1 + (1-\alpha)x_2 \in X$.*

Definition 4.4 (Convex Hull). *A set of all convex combinations of points in a given set is the convex hull of that set. Given $X \subseteq \mathbb{R}^d$ we call $conv(X) = \{\alpha x_1 + (1 - \alpha x_2) : x_1, x_2 \in X, \alpha \in [0,1]\}$.*

- Convex hull of a set is the 'smallest' set that contains that set.

- Equivalent definition: Let $h(x) \leq f(x) \forall x \in X$ be a convex function. Then h is a conex hull of f if there does not exist a function $g \neq h$ such that 1. g is convex 2. $g(x) \leq f(x)$ for all $x \in X$, and 3. $g(x) > h(x)$ for some $x \in X$.

Examples of convex sets

- Empty set, point, line
- Norm ball $\{x \ : \ ||x|| \leq r\}$
- Probability simplex $\{x : \sum_{i=1}^d x_i = 1, x_i \geq 0 \forall i \in [d]\}$
- Set of all polynomials of degree atmost d
- Set of all polygons $\{x : Ax \leq b\}$ where $\leq$ is interpreted componentwise
- Hyperplanes ($\{x : a^T x = b\}$) and halfspaces $\{x : a^T x \leq b\}$

Properties of Convex Sets

- **Separating Hyperplane Theorem** Let $X, Y \subseteq \mathbb{R}^d$ be convex sets. Then there exists a vector $\mathbf{a}$ and a constant b such that for all $\mathbf{x}$ and $\mathbf{y}$ we have

$$\mathbf{x}^T\mathbf{a} \geq b \text{ for all } \mathbf{x} \in X$$
$$\mathbf{y}^T\mathbf{a} \leq b \text{ for all } \mathbf{y} \in Y$$

- **Following operations preserve convexity**
  - Intersection
  - Image (and pre-image) of affine function: Let $S \subseteq R^n$ be a convex set and $f : \mathbb{R}^n \to \mathbb{R}^m$ be an affine function (is of the form $f(x) = Ax+b$) then the image of S defined as $f(S) = \{f(x) \in \mathbb{R}^m : x \in S\}$ is convex.

Definition 4.5 (convex function)**.** *A function $f : \mathbb{R}^d \to \mathbb{R}$ is called convex if for any $x_1, x_2 \in \mathbb{R}^d$ and $\alpha \in [0,1]$ we have*

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha \cdot f(x_1) + (1-\alpha) \cdot f(x_2).$$

Definition 4.6 (Strictly convex function)**.** *The inequality is strict in the above equation.*

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha \cdot f(x_1) + (1-\alpha) \cdot f(x_2).$$

Theorem 4.7. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. The following conditions are equivalent.*

- *f is a convex function i.e.$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha \cdot f(x_1) + (1-\alpha) \cdot f(x_2)$.*
- *$f(y) \geq f(x) + \nabla f(x)^T (y-x)$ for all $x, y \in \mathbb{R}^n$.*
- *$(\nabla f(x) - \nabla f(y))^T (x-y) \geq 0$ for all $x, y \in \mathbb{R}^n$.*
- *If f is twice differentiable then $\nabla^2 f(x)$ is positive semi-definite for all $x \in \mathbb{R}^n$ that is $y^T \nabla^2 f(x) y \geq 0$ for all $y \in \mathbb{R}^n$.*

Definition 4.8 (α-strongly convex function)**.** *Let f be a differential function. We call f an α-strongly convex function if any one of the below equivalent conditions is satisfied*

- *$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{\alpha}{2}||x-y||^2$*
- *$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\alpha}{2}||y-x||^2$*
- *$(\nabla f(y) - \nabla f(x))^T (y-x) \geq \alpha I$*
- *If f is twice differentiable, $\nabla^2 f(x) \geq \alpha$ for all $x \in \mathbb{R}^n$*
- *$f(x) - \frac{\alpha}{2}||x||^2$ is convex.*

Definition 4.9 (β-smooth functions)**.** *Let f be a differential function. We call f a β-smooth function if*

$$f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{\beta}{2}||y-x||^2.$$

Definition 4.10 (Exp-concave functions)**.** *A function f is called α-exp concave if*

$$h(x) = e^{-\alpha f(x)} \tag{4.1}$$

is concave.

Observation 1. *exp-concavity $\implies$ convexity.*

Examples:

- e^{ax} is convex for any value $a \in \mathbb{R}$.
- x^a is convex for $a \leq 0$ and $a \geq 1$ and concave for $a \in [0,1]$.

- $\log(x)$ is concave over $\mathbb{R}_{++}$
- $A^T x + b$ is both convex and concave.
- $x^T Q x + b^T x + c$ is convex over $\mathbb{R}$ if and only if Q is positive semi-definite.
- e^x is strictly convex but not strongly convex.
- square loss is 2-strongly convex
- KL-divergence defined as
$$D(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$
is strictly convex but not strongly convex. Here, P and Q are probability distributions defined on the same probability space.
- $f(x) = \sum_i x_i \log(x_i)$ is 1-strongly convex.
- Square loss is $\frac{1}{2}$-exp-concave (exercise)

4.2 Probability Bounds

We begin with basic definitions and cover (a very small) subset of fundamental results relevant for this course.

1. **Markov's Inequality** Let X be a non-negative random variable and $\varepsilon > 0$ then we have
$$\mathbb{P}[X \geq \varepsilon] \leq \frac{\mathbb{E}[X]}{\varepsilon}$$
2. **Chebyshev's Inequality** Let X be a real-valued random variable with finite second moment then we have
$$\mathbb{P}[|X - \mathbb{E}[X]| \geq k\sigma] \leq \frac{1}{k^2}$$
Here, $\sigma = \texttt{Var}(X)$.
3. **Upper Bound on Variance** Let X be a random variable taking values in $[a, b]$ then
$$\texttt{Var}[X] \leq \frac{(b-a)^2}{4}. \tag{4.2}$$
Proof. Let $\mathbb{E}[X] = \mu$.
$$\begin{aligned} \texttt{Var}[X] &\leq \mathbb{E}[(X-\mu)^2 + (b-X)(X-a)] && \text{(since } X \in [a,b]\text{)} \\ &= \mathbb{E}[\mu^2 - 2X\mu + (a+b)X - ab] \\ &= -\mu^2 + (a+b)\mu - ab \\ &= (\mu - a)(b - \mu) \end{aligned}$$
The above inequality holds for any value of μ. Hence, it also holds for $\mu = \frac{a+b}{2}$. This gives the desired result. $\square$
4. **Hoeffdings Inequality)** Let X be a random variable taking values in $[a, b]$ and $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ where each random variable X_i is an independent copy of X, then we have
$$\mathbb{P}[|\overline{X} - \mathbb{E}[X]| \geq \varepsilon] \leq 2 \cdot \exp\Big(-\frac{2n\varepsilon^2}{(b-a)^2}\Big).$$

4.3 Other Important Inequalities/Theorems

1. AM-GM inequality: Given n reals $a_1, a_2, \cdots, a_n$ we have

$$\frac{1}{n}\sum_{i=1}^{n} a_i \geq \left(\Pi_{i=1}^{n} a_i\right)^{1/n}$$

2. (Taylor's Series Expansion) Let f be $n+1$ times continuously differentiable function at every point between x and y. Then we have

$$f(x) = \sum_{m=0}^{n} \frac{(x-y)^m}{m} f^{(m)}(y) + \frac{(x-z)^{(n+1)}}{(n+1)!} f^{(m+1)}(z)$$

   for some z between x and y.

3. Cauchy-Shwartz Inequality:

$$|\langle x, y\rangle| \leq \ ||x||||y||$$

4. Holder's Inequality:

$$\sum_{i=1}^{n} |x_i y_i| \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}$$

   where $p, q \in (1, \infty)$ are such that $\frac{1}{p} + \frac{1}{q} = 1$.

5. $e^x \geq 1 + x \quad \forall x \in \mathbb{R}$.
6. $e^{-x} \leq 1 - x + \frac{x^2}{2} \quad \forall x \in \mathbb{R}$.
7. $e^{-x-x^2} \leq 1 - x \quad \forall x \leq \frac{1}{2}$.
8. $\log(1+x) \geq x - \frac{x^2}{2} \quad \forall x \geq 0$.
9. $\log(1+x) \leq x \quad \forall x \geq 0$.
10. **Jensen's Inequality**
   Let ϕ be a convex function then,

$$\phi\left(\frac{\sum_i a_i x_i}{\sum_i a_i}\right) \leq \frac{\sum_i a_i \phi(x_i)}{\sum_i a_i}.$$

   Equivalently, for any random variable X, we have

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}[\phi(X)].$$

Proof. $\phi(\frac{\sum_i a_i x_i}{\sum_i a_i}) = \phi(\sum_i (\frac{a_i}{\sum_i a_i}) x_i) \leq \sum_i \frac{a_i}{\sum_i a_i} \phi(x_i) = \frac{\sum_i a_i x_i}{\sum_i a_i}$. $\square$