CSE6243: Advanced Machine Learning **Fall 2024**

Lecture 4: Optimization: Convex set and Function II

Lecturer: Bo Dai *Scribes: Arjun Bansal, Atticus Rex*

Note: *LaTeX template courtesy of UC Berkeley EECS Department.*

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 Recap

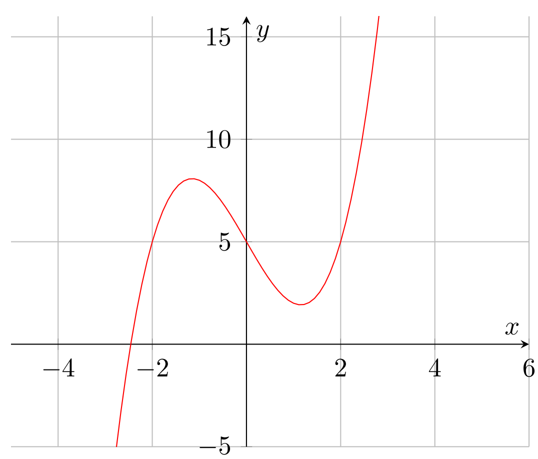
In previous class we discussed the following topics:

1. Definition of a Convex Set
2. Examples of the convex set
3. Operations that preserve convexity
4. 0th, 1st, and 2nd order conditions for convex functions and proofs for why these conditions hold

4.1.1 Conditions for Convexity

Definition 4.1 *0th Order Condition for Convexity*
A function f is convex if and only if for all $x, y \in \Omega$ and $t \in [0, 1]$, we have:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \tag{4.1}$$



For example the function $f(x) = x^3 - 4x + 5$ is convex for $x > -1$ because it satisfies the 0th order condition for convexity. The function is not convex for all x as there are more than 2 local minima's. This also mean positive quadratic functions are convex in general.

Definition 4.2 *1st Order Condition for Convexity*
A function f is convex if and only if for all $x, y \in \Omega$ and $t \in [0,1]$, we have:

$$f(y) \geq f(x) + \nabla f(y)^T (y - x) \tag{4.2}$$

This condition is equivalent to saying that the function must lie above its tangent line at any point in the domain.

Definition 4.3 *2nd Order Condition for Convexity*
A function f is convex if and only if for all $x \in \Omega$, we have:

$$\nabla^2 f(x) \geq 0 \tag{4.3}$$

This condition is equivalent to saying that the function must have a non-negative Hessian at any point in the domain.
Proof: Proof of 2nd order condition for convexity:

$$f \in \Omega \iff \nabla^2 f(x) \geq 0 \tag{4.4}$$

Proof of $\rightarrow$: Use the Taylor Expansion of f around x:

$$f(x+h) = f(x) + h^T \nabla f(x) + \frac{1}{2} h^T \nabla^2 f(z) h$$

We can rearrearange the terms to get:

$$\frac{1}{2} h^T \nabla^2 f(z) h = f(x+h) - (f(x) + h^T \nabla f(x))$$

We know from the 1st order condition that $f(x+h) \geq f(x) + \nabla f(x)^T h$. Therefore:

$$0 \leq f(x+h) - \big(f(x) + h^\top \nabla f(x)\big) = \frac{1}{2} h^T \nabla^2 f(z) h \geq 0$$

Therefore the only if condition holds.
Proof of $\leftarrow$: We can use the same Taylor Expansion as above:

$$f(x+h) = f(x) + h^T \nabla f(x) + \frac{1}{2} h^T \nabla^2 f(z) h$$

Since we know that $\nabla^2 f(z) \geq 0$, we can say that:

$$f(x+h) \geq f(x) + h^T \nabla f(x)$$

This is the same as the first order condition for convexity and therefore the if condition holds and the Hessian is a semi-positive definite matrix.

■

4.2 New Material

4.2.1 Convex Function Examples

1. Affine Functions: $f(x) = a^T x + b$. Can easily be shown to be convex using the 2nd order condition the Hessian is $0 \geq 0$.

2. Quadratic Functions: $f(x) = x^T Qx + b^T + c$ where $Q \in \mathbb{R}^{n\times n}$ and is a SPSD, $x, b \in \mathbb{R}^n$, $c \in \mathbb{R}$ Can prove that this is convex using the 2nd order condition the Hessian is $Q \geq 0$ as Q is a SPSD.
3. Vector Norms $||x||_p = \left(\sum_{i=1}^{d} |x_i|^p\right)^{1/p}$ where $p \geq 1$. Proof of convexity is given below.
4. Trace of a Matrix $tr(Ax)$ where $A \in \mathbb{R}^{p\times d}$ and $x \in \mathbb{R}^{d\times p}$. This is convex as the trace is a linear function.
5. Norm of a Matrix $||z||_{op} = \lambda_{max}(x)$
6. Frobenius Norm $||x||_F = \sqrt{tr(A^T A)}$
7. Logarithmic Sum Exponents: $f(x) = \log(\sum_{i=0}^{n} \exp(x_i))$. The proof of this is given below.

4.2.1.1 Vector Norms Proof of Convexity

Proof: We can prove that the p-norm is convex by using the 0th order condition for convexity. We can use the triangle inequality to prove this.

$$||tx + (1-t)y||_p \leq ||tx||_p + ||(1-t)y||_p = t||x||_p + (1-t)||y||_p$$
$$||tx + (1-t)y||_p \leq t||x||_p + (1-t)||y||_p$$

This is equivalent to the 0th order condition for convexity and therefore the p-norm is convex. ■

The infinity norm is also convex as it is the maximum of the absolute values of the elements of the vector.

$$||x||_\infty = \max_i(x_i)$$

4.2.1.2 Log-Sum-Exp Proof of Convexity

Easiest to prove using the 1st order condition for convexity.

$$\nabla f(x) = \left[\frac{\exp(x_i)}{\sum_{i=1}^{n} \exp(x_i)}\right]_i^d$$

The above gradient is a probability distribution and therefore the function is convex.

4.2.2 Operations that preserve Convex functions

1. Non-negative weighted sum of convex functions: $f(x) = \sum_{i=1}^{n} \alpha_i f_i(x)$ where f_i are convex and $\alpha_i \geq 0$.
2. A special case of the above is $n \to \infty$ and $\sum_{i=1}^{n} \alpha_i = 1$.This becomes the Expectation function $g(x) = \mathbb{E}_\alpha[f(x_i)] = \int p(z)f(x,z)dz$ where $p(z)$ is a probability distribution.
3. The Max operation $g(x) = \max_{i=1}^{n} f_i(x)$ If f is convex then g is also convex.
4. A special case of the above is $\tilde{g}(x) = \max_{i=1}^{n} f_i(x,y)$ is also convex is f is convex in x but not necessarily in y.
5. Spectral Norm of a Matrix $||A||_{op} = \lambda(x) = \max_{y^T y=1} y^T Xy$ is convex.

6. Partial Min $g(x) = \min_y f(x,y)$ f needs to be convex in both x and y for g to be convex
7. Inner Affine function $g(x) = f(Ax+b)$, f needs to be convex for g to be convex.
8. Composite functions:

$$g(x) = f(h(x))$$
$$x \in \mathbb{R}^d, h(\cdot) : \mathbb{R}^d \to \mathbb{R}$$
$$f : \mathbb{R} \to \mathbb{R}$$
$$g(\cdot) : \mathbb{R}^d \to \mathbb{R}$$

$$g(\cdot)\text{convex} \begin{cases} f(\cdot) \text{ is convex and strictly increasing and } h \text{ is convex} \\ f(\cdot) \text{ is convex and strictly decreasing and } h \text{ is concave} \end{cases}$$

4.2.3 Convex Function Applications

Say we have a probability distribution function

$$p(x,z) = p(x|z)p(z)$$

with $D = x_1, x_2, ..., x_n$. We use the maximum likelihood function to estimate the parameters of the distribution, that is, we want to maximize $\log p(x)$

$$\begin{aligned} \mathcal{L}(z) &= -\sum_{i=1}^{n} \log p(x) && (4.5) \\ &= -\sum_{i=1}^{n} \log \int_z p(x,y)dz && (4.6) \\ &= -\sum_{i=1}^{n} \log \int_z \frac{p(x_i,z)}{q(z|x_i)} q(z|x_i)dz && (4.7) \end{aligned}$$

where q is an auxillary distribution.

The shape of a log function is concave and therefore by making it negative the maximum likelihood function is convex.

By the Jensen inequality we know that: $f(\mathbb{E}|x|) \leq \mathbb{E}[f(x)]$. Therefore we can say that:

$$\mathcal{L}(z) \leq \min_q \sum_{i=1}^{n} -\mathbb{E}_{q(z|x)} \log(p(x_i,z)) - \log(q(z|x_i)) \quad (4.8)$$

This is the variational lower bound of the maximum likelihood function.