Recitation 22

Thursday November 30, 2022

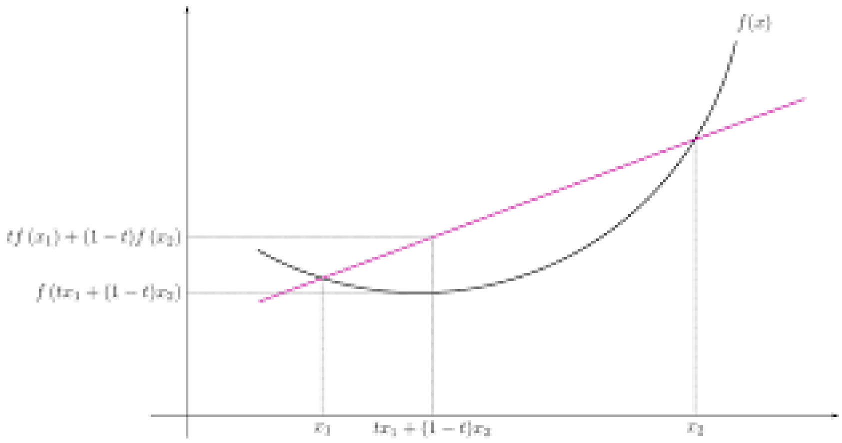
1 Recap

1.0.1 Convex function

Convex Function: Let $X \subseteq \mathbb{R}^n$ be a convex set. A function $f : X \to \mathbb{R}$ is *convex* if

$$\forall x_1, x_2 \in X, \forall t \in [0,1] : f(tx_1 + (1-t)x_2) \le tf(x_1) + (1-t)f(x_2),$$

i.e., the "graph is below the chord".



1. Sum: If $f_1(x)$ and $f_2(x)$ are convex functions, then for $c_1, c_2 \ge 0$, $c_1 f_1(x) + c_2 f_2(x)$ is also a convex function.

2. Pointwise maximum: If $f_1, ..., f_n$ are convex functions, then $f(x) = \max\{f_1(x), ..., f_n(x)\}$ is also convex.

3. Linear change of coordinates: If f is a convex function, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, then $g(x) = f(Ax + b)$ is also convex.

4. Sublevel sets:
$$S_\gamma = \{x \in \mathbb{R}^n : f(x) \le \gamma\}$$
If $f(x)$ is convex, then S_γ is a convex set.

5. Epigraph of a function:
$$\text{epi}f = \{(x,y) : x \in \mathbb{R}^n, y \in \mathbb{R} : \quad f(x) \le y\}.$$
If f is a convex function, then epi$f$ is a convex set.

6. Convex Optimization Problem:

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ &\text{s.t.}\;\; g_i(x) \le 0 \qquad i = 1,\dots,m \end{aligned}$$

   where f and $g_1,\dots,g_m$ are convex functions.

2 Hessian Matrix

- A Hessian matrix of a function f is $H(x) := \nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{ij}$.
- $H(x)$ is always symmetric.
- A function f is convex if and only if $H(x)$ is psd for any x.
- Warning: oftentimes $H(x)$ depends on the values of x. In turns, its eigenvalues depends on the values of x.
- Practice: among the two functions $f(x,y) = 2x^2 - 4xy + y^2$ and $g(x,y) = e^{-2x-y}$, what are their Hessian matrices? Which function has Hessian matrix that depends on x, y and which one does not? Which function is convex?

2.1 Gradient Descent

Gradient descent is used to find the minimum of a function $f(x)$ given access to its gradients $\nabla f(x)$ and a stepsize γ.

Algorithm 1 Gradient Descent

Input: initial guess x_0, step size $\gamma > 0$

while $\nabla f(x_k) \neq 0$ **do** $x_{k+1} = x_k - \gamma \nabla f(x_k)$ **return** x_k;

3 Exercises

1. For the following functions, what are their Hessian matrix? Are they convex?

   (a) $f(x_1, x_2, x_3) = x_1 x_2 x_3$ for $x_1, x_2, x_3 \in \mathbb{R}$.

   (b) $f(x_1, x_2) = e^{-2x_1 - x_2}$ for $x_1, x_2 \in \mathbb{R}^+$.

   (c) $f(x_1, x_2) = x_1^2 + 5x_2^2 - 4x_1 x_2$ for $x_1, x_2 \in \mathbb{R}$.

   (d) $f(x_1, x_2) = 4x_1^2 + x_2^2 + 5x_1 x_2$ for $x_1, x_2 \in \mathbb{R}$.

   (e) $f(x_1, x_2) = x_1 + x_2 + \frac{1}{x_1 x_2}$ for $x_1, x_2 \in \mathbb{R}^+$.

2. Denote $f(x) = \frac{1}{2} \cdot x^2 - 3x$ has global minimum at $x^* = 3$.

   (a) What is $\nabla f(x)$?

   (b) Suppose that we apply gradient descent with step size $\gamma > 0$. How do we express x_{n+1} in terms of x_n and γ?

3. More practice. Are the following functions convex? Explain your reasoning.

   (a) $f(x, y) = e^{2x-y} + (3x^2 + 2y^2 - xy)$

   (b) $f(x, y, z) = xyz$

   (c) $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \log x_i$

4 Solutions

1. (a) $H(x_1,x_2,x_3) = \begin{bmatrix} 0 & x_3 & x_2 \\ x_3 & 0 & x_1 \\ x_2 & x_1 & 0 \end{bmatrix}$ which is not psd when we substitute (x_1,x_2,x_3) with $(1,1,1)$. $H(1,1,1) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$, thus, f is not convex.

   (b) $H(x) = e^{-2x_1-x_2}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$ with eigenvalues 0 and $5e^{-2x_1-x_2} > 0$. Thus, f is convex. In addition, $\lambda_1 = 5e^{-2x_1-x_2} < 5$ meaning that f is smooth with $L = 5$.

   (c) $H(x) = \begin{bmatrix} 2 & -4 \\ -4 & 10 \end{bmatrix}$ with eigenvalues $6 \pm 4\sqrt{2} > 0$. f is smooth with $L = 6+4\sqrt{2}$.

   (d) $H(x) = \begin{bmatrix} 8 & 5 \\ 5 & 2 \end{bmatrix}$ with eigenvalues $5 \pm \sqrt{34}$. Thus, f is not convex.

   (e) $H(x) = \begin{bmatrix} \frac{2}{x_1^3 x_2} & \frac{1}{x_1^2 x_2^2} \\ \frac{1}{x_1^2 x_2^2} & \frac{2}{x_1 x_2^3} \end{bmatrix}$. The two eigenvalues λ_1, λ_2 satisfies $\lambda_1 + \lambda_2 = tr(A) > 0$ and $\lambda_1\lambda_2 = \det A < 0$ for x_1, x_2 large.

2. (a) $\nabla f(x) = \frac{\partial f}{\partial x} = x - 3$.

   (b) $x_{n+1} = x_n - \gamma \cdot \nabla f(x_n) = x_n - \gamma(x_n - 3) = (1-\gamma)x_n + 3\gamma$.

3. (a) Convex. e^{2x-y} is convex because $f(t) = e^t$ is convex and $(x,y) \to 2x - y$ is a change of coordinates. $3x^2 + 2y^2 - xy$ is also convex because $\begin{bmatrix} 3 & -1/2 \\ -1/2 & 2 \end{bmatrix}$ is psd.

   (b) Not convex. Consider the points $a, b = (1,-1,1), (-1,1,1)$. The we have, $f(\frac{1}{2}a + \frac{1}{2}b) = f(0,0,1) = 0 > \frac{1}{2}f(a) + \frac{1}{2}f(b) = -1$. Another proof is to restrict $z = 1$, which is $f(x,y,1) = xy$. This function is not convex (try to plot it).

   (c) Convex. Each $x_i \log x_i$ term is convex. The sum of convex functions remains convex.