Recitation #9

Irina Espejo (iem244@nyu.edu)

Center for Data Science

DS-GA 1014: Optimization and Computational Linear Algebra for Data Science

recall: Optimization

Hessian and Gradient

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. It's gradient is the column vector

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f(x)}{\partial x_n}(x) \end{pmatrix}$$

The gradient indicates the direction of maximum growth of $f(x)$. The Hessian $H_f(x)$ is

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial^2 x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial^2 x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial^2 x_n} \end{pmatrix}$$

$H_f(x)$ is symmetric.

recall: Convexity

Convex sets and functions

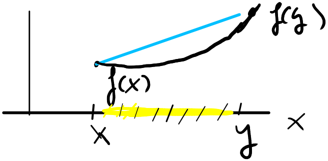
A set $C \subset \mathbb{R}^n$ is convex if for all $x, y \in C$ and all $\alpha \in [0,1]$,

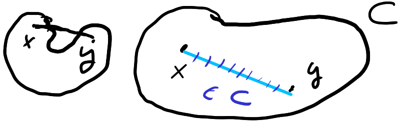
$$\alpha x + (1-\alpha) y \in C$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for all $x, y \in \mathbb{R}^n$ and all $\alpha \in [0,1]$ it holds that

global property

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$





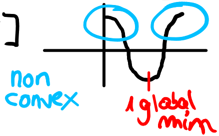
practice: Gradient and Hessian

Exercise 1

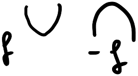
1. If f only has 1 global minima and no local minima, then f is convex ✓True <u>✗False</u>
2. Linear combination of two convex functions is convex ✓True <u>✗False</u>
3. Convex functions are differentiable at all points ✓True <u>✗False</u>
4. Norms are convex functions <u>✓True</u> ✗False
5. If f is convex, then $g(x) = f(Ax + b)$ is also convex <u>✓True</u> ✗False
6. Sum of a non-convex function with another function is never convex ✓True <u>✗False</u>
7. Union of convex sets is convex ✓True <u>✗False</u>
8. Intersection of convex sets is convex <u>✓True</u> ✗False

9. Maximum of two convex functions is convex ✓True ✗False
10. Every subspace is a convex set ✓True ✗False
11. Every convex set is a subspace ✓True ✗False

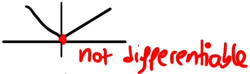
① Counterexample $f(x)=\cos(x)$ $x\in[0,\pi]$



② f convex, $-f$ is non-convex



③ Counterexample $f(x)=|x|$



④ To prove it: let $x,y\in\mathbb{R}^n$ $\alpha\in[0,1]$

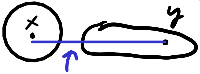
$$\|\alpha x+(1-\alpha)y\| \underset{\substack{\uparrow\\ \text{triangular}\\ \text{inequality}}}{\le} \|\alpha x\|+\|(1-\alpha)y\| = |\alpha|\,\|x\|+|1-\alpha|\,\|y\|$$

(5) convexity is a global property $x' = Ax + b$ is just another $x' \in \mathbb{R}^n$

(6) Counter example let f be nonconvex then $f - f = 0$ is convex

(7) Take two convex sets

$C_1 \cup C_2 =$



there are points in the segment outside the union

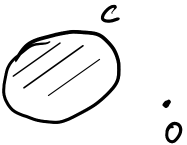
(8) for any $x, y \in C_1 \cap C_2$ x — y $\in C_1 \cap C_2$ in particular $x, y \in C_1$ which is convex

(9) See Ex 2

Exercise 1

10 True by definition of subspace.

11



convex sets might not contain the zero

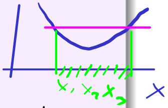
Exercise 2

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $t \in \mathbb{R}$. We define the epigraph $\text{epi}(f) \subset \mathbb{R}^{n+1}$ to be the set of all points above the graph of f

$$\text{epi}(f) = \{(x, t) \in \mathbb{R}^{n+1} \mid \underline{f(x) \leq t}\}$$

$t = 2$

$(x_1, 2)\ (x_2, 2)\ (x_3, 2)$



Show that

- Prove that f is convex if and only if $\text{epi}(f)$ is convex set
- Prove that if f, g are convex functions then $h(x) = \max(f(x), g(x))$ is convex.

Exercise 2

2.1. $\Longrightarrow$

Hypo: f is a convex function $\overset{\text{Goal}}{\Longrightarrow}$ $\boxed{\text{epi}(f) \text{ is a convex set}}$

for all $x, y \in \mathbb{R}^n$, $\alpha \in [0,1]$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

Goal

I want to $(x,t), (x',t') \in \text{epi}(f)$ $\alpha \in [0,1]$

show $\alpha(x,t) + (1-\alpha)(x',t') \in \text{epi}(f)$

$$f(\alpha x + (1-\alpha) x') \leq \alpha t + (1-\alpha) t'$$

Exercise 2

$$f(\alpha x+(1-\alpha)\cancel{y}\,x') \leq \alpha f(x)+(1-\alpha)f(\cancel{y}\,x')$$

$(x,t),(x',t') \in epi(f)$

$f(x) \leq t$

$f(x') \leq t'$

$$f(\alpha x+(1-\alpha)\cancel{y}\,x') \leq \alpha t+(1-\alpha)t'$$

this means that $(\alpha x+(1-\alpha)x', \alpha t+(1-\alpha)t') \in epi(f)$ $\square$

Exercise 2

$\Leftarrow$ epi(f) is convex $\Rightarrow$ f is convex

Let $(x,t), (x',t') \in \text{epi}(f)$ and $\alpha \in [0,1]$

Because epi(f) is convex $\alpha(x,t)+(1-\alpha)(x',t') \in \text{epi}(f)$

So by definition of epi(f): $f(\alpha x+(1-\alpha)x') \leq \alpha t+(1-\alpha)t'$

and f is convex $\square$

2.2

Not necessary to do it

recall: Taylor n-variables

Taylor's approximation

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function, Taylor's formula for all $x \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$ "small"

- Order 1: $f(x+h) \approx f(x) + \langle \nabla f(x), h \rangle$
- Order 2: $f(x+h) \approx f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} h^T H_f(x) h$

practice: Taylor's approximation

Exercise 3

Apply Taylor's formula to second order to the following functions with $h \in \mathbb{R}^2$ small

1. $f(x, y) = e^x \cos(y)$ at $(x, y) = (0, 0)$
2. $f(x) = \|x\|^2$ with $x \in \mathbb{R}^2$ at $x = 0$

2.1 $\nabla f(x,y) = \left(e^x \cos(y), -e^x \sin(y)\right)$

at $(0,0)$ $(1, 0)$

$$H_f(x,y) = \begin{pmatrix} e^x \cos^2(y) & -e^x \cos y \sin y \\ -e^x \cos y \sin y & -e^x \cos^2(y) \end{pmatrix}$$

2.2 $\nabla f(x) = 2x$ $\quad H = 2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

practice: useful derivatives

Exercise 4

Calculate the gradient and Hessian of the following functions. Assume $X \in \mathbb{R}^n$

- $f(x) = \|x\|^2$ $\rightarrow$ $\nabla f = 2x$
- $f(x) = \|Ax\|^2$ with $A \in \mathbb{R}^{n\times n}$
- $f(x) = x^T Ax$

- $f(x) = x_1^2 + \dots + x_n^2 = \|x\|^2$

$$\frac{\partial f(x)}{\partial x_i} = 2x_i \qquad \nabla f(x) = 2x$$

Exercise 4

- $f(x) = \|Ax\|^2 = \langle Ax, Ax \rangle$

$$\nabla f(x) = \langle \partial_x (Ax), Ax \rangle + \langle Ax, \partial_x (Ax) \rangle$$

$$= \langle A, Ax \rangle + \langle Ax, A \rangle$$

$$= A^T A x + (Ax)^T A = 2AA^T x$$

$$* \; \partial_x (Ax) = \partial_x \begin{pmatrix} \sum_j a_{1j} x_j \\ \vdots \\ \sum_j a_{nj} x_j \end{pmatrix} = A$$

- Apply same rule as before $\nabla f(x) = (A + A^T)x$