Lab 10: Gradient Descent and Convexity

EECS 245, Fall 2025 at the University of Michigan

due by the end of your lab section on Wednesday, November 5th, 2025

Name: ____

uniqname: ____

Each lab worksheet will contain several activities, some of which will involve writing code and others that will involve writing math on paper. To receive credit for a lab, you must complete all activities and show your lab TA by the end of the lab section.

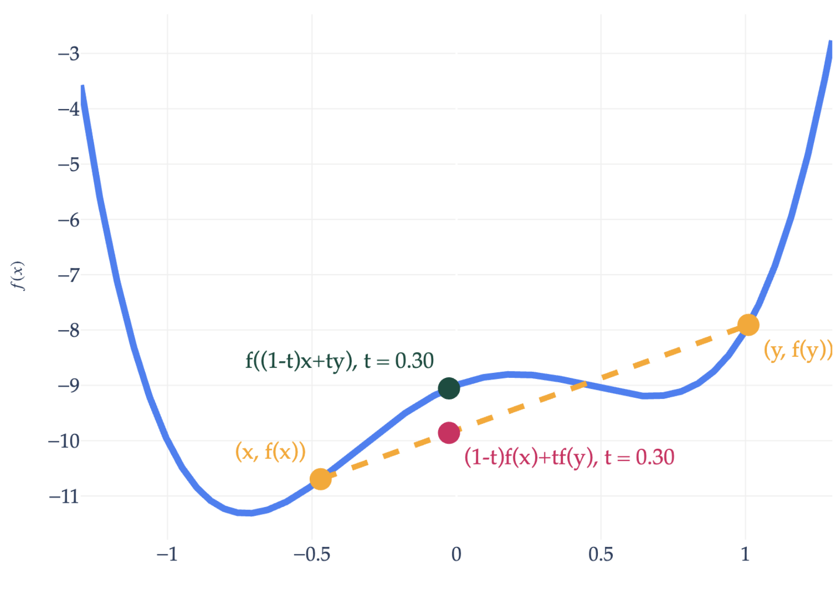
While you must get checked off by your lab TA **individually**, we encourage you to form groups with 1-2 other students to complete the activities together.

Recap: Convexity

A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if for all $\vec{x}$ and $\vec{y}$ in its domain, and for any $t \in [0, 1]$,

$$f((1-t)\vec{x} + t\vec{y}) \leq (1-t)f(\vec{x}) + tf(\vec{y})$$

The English interpretation of this definition is that **the line connecting any two points on the graph of f always lies on or above the graph of f**. Intuitively, a convex function is a function that curves upward, like a bowl.



A non-convex function

Activity 1: Using Convexity to Prove Inequalities

a) Suppose $f : \mathbb{R} \to \mathbb{R}$ is a convex function such that $f(0) = 0$. Prove that for all $y \in \mathbb{R}$ and $t \in [0, 1]$,

$$f(ty) \leq tf(y)$$

b) Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Prove that $2f(5) \leq f(3) + f(7)$.

Activity 2: Understanding Complex Proofs

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. It turns out that the function $g(\vec{x})$, defined by

$$g(\vec{x}) = f(A\vec{x} + \vec{b})$$

for some $n \times n$ matrix A and vector $\vec{b} \in \mathbb{R}^n$, is also convex, no matter what A and $\vec{b}$ are. We're not going to ask you to prove this on your own: instead, we'll give you a proof and ask you questions to ensure you understand it.

Our **goal** is to show that $g((1-t)\vec{x} + t\vec{y}) \leq (1-t)g(\vec{x}) + tg(\vec{y})$, for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $t \in [0,1]$. We'll start with the "left-hand side" of the definition, and try and leverage f's convexity.

$$\begin{aligned}
g((1-t)\vec{x} + t\vec{y}) &= f\left(A\left((1-t)\vec{x} + t\vec{y}\right) + \vec{b}\right) && (1)\\
&= f\left((1-t)A\vec{x} + tA\vec{y} + \vec{b}\right) && (2)\\
&= f\left((1-t)(A\vec{x} + \vec{b}) + t(A\vec{y} + \vec{b})\right) && (3)\\
&\leq (1-t)f(A\vec{x} + \vec{b}) + tf(A\vec{y} + \vec{b}) && (4)\\
&= \boxed{(1-t)g(\vec{x}) + tg(\vec{y})} && (5)
\end{aligned}$$

---

a) In which line did we use the fact that f is convex?

b) How did we move from line (1) to line (2), i.e. $f\left(A\left((1-t)\vec{x} + t\vec{y}\right) + \vec{b}\right) = f\left((1-t)A\vec{x} + tA\vec{y} + \vec{b}\right)$?

c) How did we move from line (2) to line (3), i.e. $f\left((1-t)A\vec{x} + tA\vec{y} + \vec{b}\right) = f\left((1-t)(A\vec{x} + \vec{b}) + t(A\vec{y} + \vec{b})\right)$?

Recall, $g(\vec{x}) = f(A\vec{x} + \vec{b})$, where A is an $n \times n$ matrix and $\vec{x}, \vec{b} \in \mathbb{R}^n$. On the last page, we showed that if f is convex, then g is convex.

Now, let's explore what happens if f is **strictly** convex. Recall, this means that for all (non-equal) $\vec{x}$ and $\vec{y}$ in its domain, and for any $t \in (0, 1)$,

$$f((1-t)\vec{x} + t\vec{y}) < (1-t)f(\vec{x}) + tf(\vec{y})$$

d) Suppose $\text{rank}(A) = n$. Explain why it's impossible for $A\vec{x} + \vec{b} = A\vec{y} + \vec{b}$ for two different vectors $\vec{x}$ and $\vec{y}$.

e) Suppose $\text{rank}(A) < n$. Explain why it's possible for $g(\vec{x}) = g(\vec{y})$ for two different vectors $\vec{x}$ and $\vec{y}$. *Hint: Think about nullsp*(A).

f) Using the above reasoning, explain why if f is strictly convex, then g is strictly convex if $\text{rank}(A) = n$, and is (not strictly) convex if $\text{rank}(A) < n$.

g) What were your thoughts on this type of activity, where we give you a proof and ask you questions about it?

◯ Hated it ◯ Didn't like it ◯ Neutral ◯ Liked it ◯ Loved it

Activity 3: Gradient Descent Gone Wrong

Suppose $\vec{x} \in \mathbb{R}^2$. Let

$$f(\vec{x}) = x_1^3 + \|\vec{x}\|^2 = x_1^3 + x_1^2 + x_2^2$$

To minimize $f(\vec{x})$, we use gradient descent, with a learning rate of $\alpha = \frac{1}{4}$.

a) Open Desmos and plot the related function $g(x) = x^3 + x^2$. Even though this is a scalar-to-scalar function, and f is vector-to-scalar, they are related. What do you notice about the shape of the graph?

b) Find $\nabla f(\vec{x})$, the gradient of $f(\vec{x})$.

c) Recall, $\vec{x}^{(t)}$ is the guess for $\vec{x}^*$ at timestep t. Let $\vec{x}^{(t)} = \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \end{bmatrix}$.

Show that

$$x_1^{(t+1)} = \frac{1}{2}x_1^{(t)} - \frac{3}{4}(x_1^{(t)})^2, \qquad x_2^{(t+1)} = \frac{1}{2}x_2^{(t)}$$

d) For any initial guess $\vec{x}^{(0)}$, what does $x_2^{(t)}$ converge to as $t \to \infty$?

e) Suppose $\vec{x}^{(0)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$.

i) Find $\vec{x}^{(1)}$.

ii) Will gradient descent eventually converge, given this initial guess and learning rate?

f) Suppose $\vec{x}^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

i) Find $\vec{x}^{(1)}$.

ii) Will gradient descent eventually converge, given this initial guess and learning rate?