# Lab 10: Gradient Descent and Convexity Solutions

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.

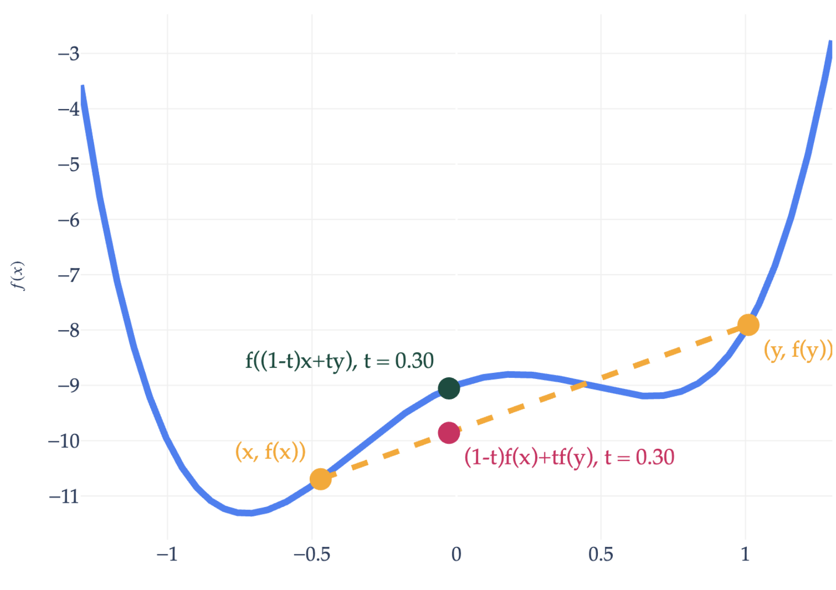
While you must get checked off by your lab TA **individually**, we encourage you to form groups with 1-2 other students to complete the activities together.

## 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 $t \in [0, 1]$,

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

**Solution:**

$$\begin{aligned} f((1-t)\cdot 0 + ty) &\le (1-t)f(0) + tf(y) \\ f(ty) &\le tf(y) \end{aligned}$$

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

**Solution:** We'll start by solving for $t$ using $x = 3$ and $y = 7$ with the expression for the input on the left side of the inequality.

$$\begin{aligned} 5 &= (1-t)x + ty \\ &= (1-t)\cdot 3 + 7t \\ &= 3 - 3t + 7t \\ &= 3 + 4t,\ \ t = \frac{1}{2} \end{aligned}$$

Then, plug the variables into the inequality and simplify.

$$\begin{aligned} f((1-\frac{1}{2})\cdot 3 + \frac{1}{2}\cdot 7) &\le (1-\frac{1}{2})\cdot f(3) + \frac{1}{2}\cdot f(7) \\ f(\frac{3}{2} + \frac{7}{2}) &\le \frac{f(3)}{2} + \frac{f(7)}{2} \\ f(5) &\le \frac{f(3)}{2} + \frac{f(7)}{2} \\ 2f(5) &\le f(3) + f(7) \end{aligned}$$

### 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?

> **Solution:** Line 4. We simplified the original expression to the form of the formal definition's left side in line 3, and line 4 is where we connect it back to the right side.

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

> **Solution:** Distributing $A$ by left multplying it to the terms in the parentheses.

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

> **Solution:** Add $t\vec{b} - t\vec{b}$ to the input expression, that way we can increase the number of terms without changing the value of the expression.

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 $\vec{y}$.

> **Solution:** If $\text{rank}(A) = n$, then the columns of $A$ are linearly independent, so $A\vec{x}$ and $A\vec{x}$ must be different for any $\vec{x} \neq \vec{y}$.

**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$(A)$.*

> **Solution:** If $\text{rank}(A) < n$, then $A$'s columns are linearly dependent, so $A\vec{x}$ and $A\vec{y}$ can be the same vector. In that case, $f(A\vec{x} + \vec{b}) = f(A\vec{y} + \vec{b}) \to g(\vec{x}) = g(\vec{y})$.

**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 $\text{rank}(A) < n$.

> **Solution:** We can show this with a proof by cases. In both cases, we'll start from line 4 of the proof on the previous page, but with $f$ being strictly convex.
> Case 1: $\text{rank}(A) = n$
>
> $$\begin{aligned} g((1-t)\vec{x} + t\vec{y}) &< (1-t)f(A\vec{x} + \vec{b}) + tf(A\vec{y} + \vec{b}) \\ &< (1-t)g(\vec{x}) + tg(\vec{y}) \end{aligned}$$
>
> Case 2: $\text{rank}(A) < n$
> We know from part **e)** that it's possible for $g(\vec{x}) = g(\vec{y})$. Using proof by contradiction, assume that $g$ is strictly convex.
>
> $$\begin{aligned} g((1-t)\vec{x} + t\vec{y}) &< (1-t)f(A\vec{x} + \vec{b}) + tf(A\vec{y} + \vec{b}) \\ &< (1-t)g(\vec{x}) + tg(\vec{y}) \\ &< (1-t)g(\vec{x}) + tg(\vec{x}) \\ &< g(\vec{x}) \end{aligned}$$
>
> This is a contradiction, because if $t = 0$, then the left side of the inequality is $g(\vec{x})$, leaving us with $g(\vec{x}) < g(\vec{x})$.

**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?

> **Solution:** $g(x)$ is a cubic function, with a local minimum and local maximum. It's not convex.

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

> **Solution:**
>
> $$\nabla f(\vec{x}) = \begin{bmatrix} 3x_1^2 + 2x_1 \\ 2x_2 \end{bmatrix}$$

**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

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

**Solution:** Use the formula for gradient descent, $\vec{x}^{(t+1)} = \vec{x}^{(t)} - \alpha \nabla f(\vec{x}^{(t)})$

$$\begin{aligned} x_1^{(t+1)} &= x_1^{(t)} - \alpha \nabla f(x_1^{(t)}) \\ &= x_1^{(t)} - \frac{1}{4}(3(x_1^{(t)})^2 + 2x_1) \\ &= x_1^{(t)} - \frac{3}{4}(x_1^{(t)})^2 - \frac{2}{4}x_1^{(t)} \\ &= \frac{1}{2}x_1^{(t)} - \frac{3}{4}(x_1^{(t)})^2 \end{aligned}$$

$$\begin{aligned} x_2^{(t+1)} &= x_2^{(t)} - \alpha \nabla f(x_2^{(t)}) \\ &= x_2^{(t)} - \frac{1}{4}(2x_2^{(t)}) \\ &= x_2^{(t)} - \frac{1}{2}x_2^{(t)} \\ &= \frac{1}{2}x_2^{(t)} \end{aligned}$$

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

**Solution:** $x_2^{(t)}$ converges to 0 as $t \to \infty$ because it's being divided at each iteration.

**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?

**Solution:**
i. $x_2$ updates by being scaled, and since it's already 0 we can just focus on finding $x_1^{(1)}$.

$$\begin{aligned} x_1^{(1)} &= \frac{1}{2}x_1^{(0)} - \frac{3}{4}(x_1^{(0)})^2 \\ &= \frac{1}{2}(-1) - \frac{3}{4}(-1)^2 \\ &= -\frac{1}{2} - \frac{3}{4} \\ &= -\frac{5}{4} \end{aligned}$$

ii. Gradient descent will not converge, it will continue to decrease until $-\infty$. We're not moving in the direction of the local minimum but still decreasing,.

**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?

**Solution:** i. Similar to part **e)**, we only have to find $x_1^{(1)}$

$$\begin{aligned} x_1^{(1)} &= \frac{1}{2}x_1^{(0)} - \frac{3}{4}(x_1^{(0)})^2 \\ &= \frac{1}{2}(1) - \frac{3}{4}(1)^2 \\ &= -\frac{1}{4} \end{aligned}$$

ii. Gradient descent will converge. If we run more iterations, we'll see that $x_1^{(t)}$ is approcing zero because the absolute value keeps decreasing. $x_1^{(2)}$, for instance, is $-\frac{11}{64}$, and $|-\frac{11}{64}| < |-\frac{1}{4}|$