# Lab 10: Gradient Descent and Convexity

EECS 245, Winter 2026 at the University of Michigan

**due** by the end of your lab section

**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: Gradient Descent (Chapter 8.3)

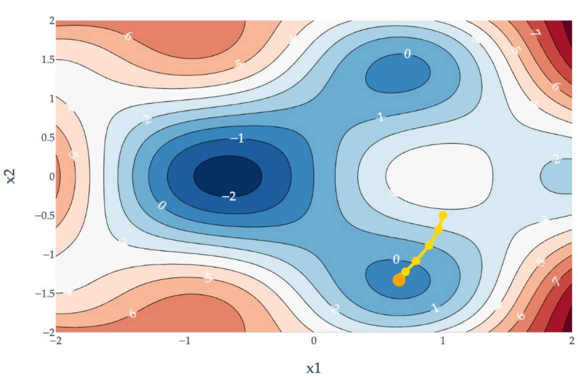
Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is a **differentiable** vector-to-scalar function, meaning that all of its partial derivatives are defined everywhere. To find $\vec{x}^*$, the **minimizer** of $f$:

1. Choose a positive number, **ff**. This number is called the **learning rate**, or **step size**.
2. Choose an **initial guess** for the minimizer, $\vec{x}^{(0)}$.
3. Then, repeatedly update the guess using the **update rule**:

$$\vec{x}^{(t+1)} = \vec{x}^{(t)} - \alpha \nabla f(\vec{x}^{(t)})$$

4. Terminate once the algorithm converges, which happens when the norm of the gradient, $\|\nabla f(\vec{x}^{(t)})\|$, is below some small **tolerance** level, e.g. $0.001$ (since this must mean we're very close to a minimum).

**Intuition**: The gradient vector $\nabla f(\vec{x}^{(t)})$ tells us the direction of greatest increase in $f$ at the current guess $\vec{x}^{(t)}$, so $-\nabla f(\vec{x}^{(t)})$ is the direction that will **decrease** our function the most. The distance moved in that direction is determined by the step size $\alpha$, which scales $-\nabla f(\vec{x}^{(t)})$. To update our guess, we add $-\alpha \nabla f(\vec{x}^{(t)})$ to the old guess, $\vec{x}^{(t)}$. Then, we repeat this process.




Example gradient descent path for a function $f : \mathbb{R}^2 \to \mathbb{R}$.

## Activity 1: 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?

### Activity 2: Gradient Descent for Empirical Risk Minimization

Suppose we have a dataset of 3 points, $(1, 2), (3, 5), (6, -1)$. We'd like to fit a simple linear regression model, $h(x_i) = w_0 + w_1 x_i$, to this dataset by minimizing average squared loss.

While we already know a closed-form solution for the optimal parameters — we've seen multiple equivalent versions of these formulas throughout the semester — let's use gradient descent to find them.

Let $\vec{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$. Using an initial guess of $\vec{w}^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and a step size of $\alpha = 0.1$, perform one iteration of gradient descent. What is $\vec{w}^{(1)}$?

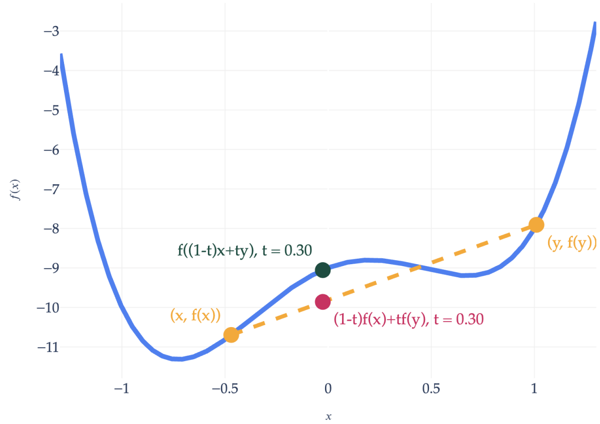
*Hint: You can proceed either by using the gradient of* $R_{sq}(\vec{w}) = \frac{1}{n}\|\vec{y} - X\vec{w}\|^2$ *from the notes or by computing the partial derivatives of* $R_{sq}(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (w_0 + w_1 x_i))^2$ *with respect to* $w_0$ *and* $w_1$.

### Activity 3: Convexity and Gradient Descent

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.

**a)** If a function is convex, it **must** have a global minimum.
◯ True ◯ False

**b)** If a function is convex, then gradient descent **must** converge on it given any initial guess and step size.
◯ True ◯ False

**c)** If a function is convex, then any local minimum is also a global minimum.
◯ True ◯ False

**d)** If a function is convex and has a global minimum, then its minimizer **must** be unique.
◯ True ◯ False

## Activity 4: Linear Approximation

Suppose $f : \mathbb{R} \to \mathbb{R}$, i.e. $f$ is a scalar-to-scalar function. In general, the **tangent line** to $f(x)$ at $x = a$ is given by the equation

$$f(x) \approx \underbrace{f(a) + \left(\frac{\mathrm{d}f}{\mathrm{d}x}(a)\right)(x - a)}_{\text{tangent line}}$$

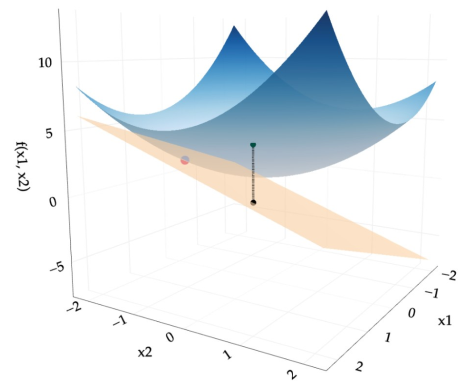
The $\approx$ symbol means that the tangent line is an approximation of $f(x)$ near $x = a$; in **Appendix 2**, we defined it as the **best linear approximation** of $f(x)$ near $x = a$. The expression on the right is a line with slope $\frac{\mathrm{d}f}{\mathrm{d}x}(a)$ that passes through the point $(a, f(a))$.

**a)** Draw $f(x) = (x - 2)^2 + 5$ and its tangent line at $x = 3$.

**b)** For vector-to-scalar functions, the best linear approximation at $\vec{x} = \vec{a}$ is given by

$$f(\vec{x}) \approx f(\vec{a}) + (\nabla f(\vec{a})) \cdot (\vec{x} - \vec{a})$$

If $\vec{x} \in \mathbb{R}^2$, this is called the **tangent plane**; if $\vec{x} \in \mathbb{R}^3$ or higher, it is called the **tangent hyperplane**.




Tangent plane for a function $f : \mathbb{R}^2 \to \mathbb{R}$.

Find the tangent plane to $f(\vec{x}) = \|\vec{x}\|^2 - 3$ at $\vec{x} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$.

## Activity 5: Using Convexity to Prove Inequalities

Proofs like this will not appear on Midterm 2 but will appear on the Final Exam.

**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 $2f(5) \leq f(3) + f(7)$.