Homework 1

Due: 11:59 pm, 3 October

1. **Gradient Descent for a Strongly Convex Function (Unconstrained Case)**

   Consider the 2-D quadratic function

   $$f(x) \;=\; \tfrac{1}{2}\, x^\top Q x + b^\top x,$$

   where

   $$Q = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}, \qquad b = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

   Take the initial point $x_0 = (3,\, 3)^\top$.

   (a) Show that f is α-strongly convex and β-smooth with specified α and β.

   (b) Set $\eta = 0.1$. Perform gradient descent $x_{k+1} = x_k - \eta \nabla f(x_k)$ for 10 steps and visualize the trajectory..

   (c) Repeat with step size 2.0 and visualize the trajectory.

   (d) Use the optimal constant step size. Run 10 steps of GD and plot the trajectory.

2. **Projected Gradient Descent for a Strongly Convex Function(Constrained Case)**

   Use the same quadratic objective

   $$f(x) = \tfrac{1}{2} x^\top Q x + b^\top x,$$

   but constrain the feasible set to

   $$\mathcal{X} = \{\, x \in \mathbb{R}^2 : a^\top x = 1 \,\}, \qquad a = (1,\, -1)^\top.$$

   (a) Derive the closed-form Euclidean projection of an arbitrary $z \in \mathbb{R}^2$ onto $\mathcal{X}$:

   $$\Pi_{\mathcal{X}}(z) := \arg\min_{x \in \mathcal{X}} \|x - z\|_2.$$

(b) Perform projected gradient descent. Run

$$y_{k+1} = x_k - \eta \nabla f(x_k), \qquad x_{k+1} = \Pi_{\mathcal{X}}(y_{k+1})$$

with step size $\eta = 0.1$ for 10 iterations, starting from $x_0 = (3, 3)^\top$. Plot the pre-projection iterates y_k and the projected iterates x_k.

3. **Gradient Descent for a Nonconvex Function**

Consider the nonconvex function

$$f(x) \;=\; x^4 + \tfrac{1}{3}x^3 - 2x^2 - x + 1.$$

This landscape has a *sharp* minimum near $x = 1$ and a *flat* minimum near $x = -1$.

(a) Start at $x_0 = -2$ and run gradient descent

$$x_{k+1} = x_k - \eta\, \nabla f(x_k),$$

with $\eta = 0.01$ for 20 iterations. Plot the trajectory of x_k.

(b) Start at $x_0 = 2$. Run gradient descent with $\eta = 0.01$ for iterations and plot the trajectory of x_k.

(c) Discuss how the different starting points lead GD to converge to different minima in this nonconvex landscape.

* Flatness of minima is closely linked to the generalization ability of deep learning models. We will later study flatness-aware optimization.

Submission:

- Jupyter Notebook (code and plots)
- A PDF report summarizing results and visualizations