Continuous Optimization **Spring 2025**

Exercise 2: Gradient Descent

Lecturer: Aurelien Lucchi

Problem 1 (Quadratic function):

Consider a quadratic function $f : \mathbb{R}^d \to \mathbb{R}$ of the form $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{R}^{d\times d}$ is symmetric invertible and $\mathbf{b} \in \mathbb{R}^d, c \in \mathbb{R}$.

1. Prove that f is smooth with constant $2\|\mathbf{A}\|$, where we recall that $\|\mathbf{A}\| := \sup_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}$.
2. What's the minimum value of f?

Problem 2 (Biased gradients):

Consider the gradient descent update with a bias:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta\nabla f(\mathbf{x}_k) + \epsilon_k, \tag{1}$$

where $\eta > 0$ is the step size and $\epsilon_k > 0$ is a bias. We assume that $\eta \leq \frac{1}{L}$.

1. Show that
$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + \frac{\eta}{2}\left(-\|\nabla f(\mathbf{x}_k)\|^2 + \|\epsilon_k\|^2\right).$$
2. Conclude that
$$\min_{k=1\ldots K} \|\nabla f(\mathbf{x}_k)\|^2 \leq \frac{\eta}{2K}(f(\mathbf{x}_1) - f(\mathbf{x}^*)) + \frac{1}{K}\sum_{k=1}^{K} \|\epsilon_k\|^2,$$

Problem 3 (Normalized Gradient Descent):

In this exercise, we consider a variant of gradient descent known as normalized gradient descent. At each iteration, it normalizes the gradient by its norm, which yields the following update step:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \eta\nabla f(\mathbf{x}_k), \tag{2}$$

where $\eta > 0$ is a chosen step size.

We assume that f is convex and L-smooth. Prove that

1. $$\|\nabla f(\mathbf{x}_k)\| \leq \frac{f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})}{\eta} + \frac{L\eta}{2}.$$
2. If we choose $\eta = \frac{2\epsilon}{L}$, how many iterations do we need to obtain $\frac{1}{k}\sum_{i=0}^{k-1} \|\nabla f(\mathbf{x}_k)\| \leq \epsilon$?

Problem 4 (Programming):

Complete TODOs in the Jupyter Notebook provided by implementing the Gradient Descent optimizer for a Linear Regression task. Then, study the behavior of the optimizer for different step sizes, initialization, and maximum number of iterations.