Homework - October 1, 2024

Student: Ayrton Chilibeck, achilibe@ualberta.ca
Lecturer: Lili Mou, UoA.F24.466566@gmail.com

Problem 1: The Epsilon-Neighbourhood

Prove that, given

$$\lambda = 1 - \frac{\epsilon}{2||y - x||}$$

and

$$z = \lambda x + (1 - \lambda)y$$

That z is in the ϵ-neighbourhood of x.

In order to prove that z is in the ϵ-neighbourhood of λ, it suffices to prove that the distance $z - x$ is less than ϵ. First we can rearrange the formulae to get a clear picture of the inequality:

$$\begin{aligned} z &= \lambda x + (1 - \lambda)y \\ z - x &= (\lambda x + (1 - \lambda)y) - x \\ &= (1 - \lambda)(y - x) \end{aligned}$$

$$\begin{aligned} \lambda =& 1 - \frac{\epsilon}{2||y - x||} \\ \epsilon =& (1 - \lambda)2\sqrt{y^2 - x^2} \end{aligned}$$

We can now prove that $\epsilon > z - x$ by contradiction:

Suppose that $\epsilon < z - x$. Then

$$\begin{aligned} \epsilon =& z - x \\ (1 - \lambda)2\sqrt{y^2 - x^2} =& (1 - \lambda)(y - x) \\ 2\sqrt{y^2 - x^2} =& (y - x) \end{aligned}$$

This is false, since we can take some $y = 3, x = 2$ and arrive at $10 < 1$. Thus it holds that z is in the epsilon-neighbourhood of x. □

Problem 2: Global optimum of a Convex Function

Prove that the global optimum of a convex function f is where $\nabla f = 0$.

Recall the definition of convexity with respect to the first-order principle from last week:

$$f(y) \geq f(x) + \nabla f(x)^{\mathsf{T}}(y - x)$$

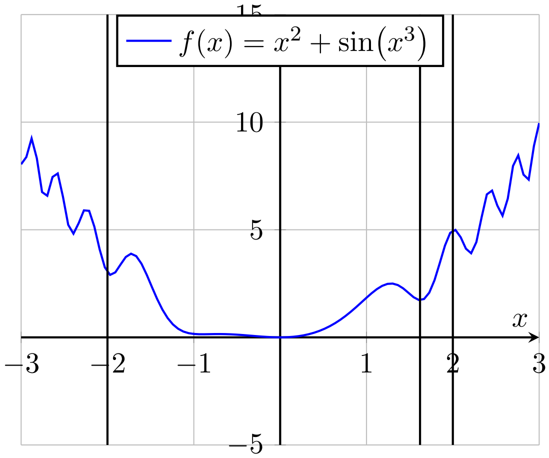

Figure 1: Graph of $x^2 + sin(x^3)$ depicting an annealing schedule where gradient descent will converge to a local optimum instead of a global one.

where $f : \mathbb{R}^n \to \mathbb{R}^n$. Note that this is the extension from the scalar equation from last week to n-dimensions.

Since we are given $\nabla f(x) = 0$ at our point x, we can substitute into the convexity function:

$$\begin{aligned} f(y) \geq & f(x) + \nabla f(x)^{\mathsf{T}}(y - x) \\ f(y) \geq & f(x) + 0 \\ f(y) \geq & f(x) \end{aligned}$$

So $\forall x, y \in \mathbb{R}^n$, $f(y) \geq f(x)$, which is the required condition to find a global optimum for a function. □

Problem 3: Learning-Rate Annealing Schedule

In the gradient descent algorithm, $\alpha > 0$ is the learning rate. If is small enough, then the function value guarantees to decrease. In practice, we may anneal α, meaning that we start from a relatively large α, but decrease it gradually.
Show that α cannot be decreased too fast. If α is decreased too fast, even if it is strictly positive, the gradient descent algorithm may not converge to the optimum of a convex function.

Consider the loss contour of $x^2 + sin(x^3)$ as shown in figure 1. If we take an annealing schedule of $\alpha = \{4, 0.6\}$, then we fail to find the global optimum for this function, rather settling at a local optimum.