Problem set: Gradient descent and Fixed points

Exercise (i) (1P) Let $f_1, f_2, \dots, f_m$ be smooth with parameters $L_1, L_2, \dots L_m$. Show that the function $f := \sum_{i=1}^m f_i$f is smooth with parameter $\sum_{i=1}^m L_i$.

Exercise (ii) (1P) Let f be smooth with parameter L and A a matrix. Show that $f \circ A$ is smooth with parameter $L\|A\|^2$.

Computing Fixed Points

Gradient descent turns up in a surprising number of situations which apriori have nothing to do with optimization. In this exercise we will see how computing the fixed point of functions can be seen as a form of gradient descent. Suppose that we have a 1-Lipschitz continuous function $g : \mathbb{R} \to \mathbb{R}$ such that we want to solve for

$$g(x) = x.$$

A simple strategy for finding such a fixed point is to run the following algorithm: starting from an arbitary x_0, we iteratively set

$$x_{k+1} = g(x_k). \tag{1}$$

Exercise (iii) (3P) Enter the missing code snippets in the jupyter notebook. Partial credit will be awarded.

We will try solve for x starting from $x_0 = 1$ in the following two equations:

$$x = \log(1 + x), \quad \text{and} \quad x = \log(2 + x). \tag{2}$$

What difference do you observe in the rate of convergence between the two problems? Let's understand why this happens:

Exercise (iv) (3P) Theoretical fixed point questions.

- We want to re-write the update (1) as a step of gradient descent. To do this, we need to find a function f such that the gradient descent update is identical to (1):

  $$x_{k+1} = x_k - \alpha f(x_k) = g(x_k).$$

  Derive such a function f.

- Give sufficient conditions on g to ensure convergence of procedure (2). What α would you need to pick? Hint: We know that gradient descent on f with fixed step-size converges if f is convex and smooth. What does this mean in terms of g?

- What condition does g need to satisfy to ensure linear convergence? Are these satisfied for the problems in (2)