Labs
Optimization for Machine Learning
Spring 2025

USC
Computer Science
Sai Praneeth Karimireddy
spkreddy.org/optmlspring2025

Problem Set 3, Feb 5, 2024
(Gradient Descent Continued.)

Solve exercise 18 from lecture notes.

Gradient descent on a quadratic function. Consider the quadratic function $f(x) = \frac{1}{2}x^\top Ax + \langle b, x\rangle + c$, where A is a $d \times d$ symmetric matrix, $b \in \mathbb{R}^d$ and c in $\mathbb{R}$.

1. What are the minimal conditions on A, b and c that ensure that f is strictly convex ? For the rest of the exercise we assume that these conditions are fulfilled.
2. Is f strongly convex ?
3. Prove that f has a unique minimum x^* and give its closed form expression.
4. Show that f can be rewritten as $f(x) = \frac{1}{2}(x - x^*)^\top A(x - x^*) + f(x^*)$
5. From an initial point $x_0 \in \mathbb{R}^d$, assume we run gradient descent with step-size $\gamma > 0$ on the function f. Show that the n^{th} iterate x_n satisfies $x_n = x^* + (I_d - \gamma A)^n\ (x_0 - x^*)$, where I_d is the $d \times d$ identity matrix.
6. In which range must the step-size γ be so that the iterates converge towards x^* ?

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

$$x_{t+1} = g(x_t)\,. \tag{1}$$

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

$$x = \log(1 + x)\,, \text{ and} \tag{2}$$
$$x = \log(2 + x)\,. \tag{3}$$

Follow the Python notebook provided here:

Google Drive Link (click here)

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

Theoretical questions.

1. 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):

$$x_{t+1} = x_t - \gamma f'(x_t) = g(x_t)\,.$$

   Derive such a function f.

2. Give sufficient conditions on g to ensure convergence of procedure (1). What γ 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?*

3. What condition does g need to satisfy to ensure *linear* convergence? Are these satisfied for problems (2) and (3) in the exercise?