

Assignment II - Exercise 4 (solution)

Machine Learning II
2021-2022 - UMONS
Souhaib Ben Taieb

In this exercise, we ask you to prove that gradient descent (GD) converges to the global optimum for **L -Lipschitz convex** functions, which are a rich class of functions that cover many problems in machine learning.

Definition 1. $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is **convex** if for all $x, y \in \mathbb{R}^d$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Definition 2. A smooth function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is **L -Lipschitz** if for all $x, y \in \mathbb{R}^d$,

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2.$$

Property 1. If $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is **convex**, then for all x, y ,

$$f(x) + \langle \nabla f(x), y - x \rangle \leq f(y), \quad (1)$$

where $\langle \cdot, \cdot \rangle$ is the dot product.

Property 2. If $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is **L -Lipschitz convex**, then for all x, y ,

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|_2^2. \quad (2)$$

In the following, gradient descent (GD) will refer to the following algorithm:

1. Choose $x_0 \in \mathbb{R}^d$ and step-size $\eta > 0$
2. For $i = 0, 1, 2, \dots$, compute

$$x_{i+1} = x_i - \eta \nabla f(x_i)$$

Theorem. If $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is **L -Lipschitz convex**, and $x^* = \operatorname{argmin}_x f(x)$, then GD with step-size $\eta \leq \frac{1}{L}$ satisfies:

$$f(x_k) \leq f(x^*) + \frac{\|x_0 - x^*\|^2}{2\eta k}. \quad (3)$$

First, show that

$$f(x_{i+1}) \leq f(x_i) - \frac{\eta}{2}\|\nabla f(x_i)\|_2^2 \quad (4)$$

for $i = 0, 1, \dots, k$. To do so, apply the following steps:

1. Use the fact that f is **L -Lipschitz convex**, i.e. use (2) with $y = x_{i+1}$ and $x = x_i$.

$$f(x_{i+1}) \leq f(x_i) + \langle \nabla f(x_i), x_{i+1} - x_i \rangle + \frac{L}{2}\|x_{i+1} - x_i\|_2^2$$

2. Replace x_{i+1} by $x_i - \eta \nabla f(x_i)$, and use the fact that $\langle a, a \rangle = \|a\|_2^2$.

$$f(x_{i+1}) \leq f(x_i) - \eta\|\nabla f(x_i)\|_2^2 + \frac{L\eta^2}{2}\|\nabla f(x_i)\|_2^2 = f(x_i) - \eta\left(1 - \frac{L\eta}{2}\right)\|\nabla f(x_i)\|_2^2$$

3. Use the fact that $L\eta \leq 1$.

$$f(x_{i+1}) \leq f(x_i) - \frac{\eta}{2} \|\nabla f(x_i)\|_2^2.$$

Secondly, show that

$$f(x_{i+1}) \leq f(x^*) + \frac{1}{2\eta} \left(\|x_i - x^*\|_2^2 - \|x_{i+1} - x^*\|_2^2 \right) \quad (5)$$

To do so, apply the following steps:

1. Combine (4) with the fact that f is convex, i.e. by (1), we have

$$f(x_i) \leq f(x^*) + \langle \nabla f(x_i), x_i - x^* \rangle.$$

$$f(x_{i+1}) \leq f(x^*) + \langle \nabla f(x_i), x_i - x^* \rangle - \frac{\eta}{2} \|\nabla f(x_i)\|_2^2$$

2. Use the fact that $\nabla f(x_i) = \frac{1}{\eta}(x_i - x_{i+1})$.

$$f(x_{i+1}) \leq f(x^*) + \frac{1}{\eta} \langle x_i - x_{i+1}, x_i - x^* \rangle - \frac{1}{2\eta} \|x_i - x_{i+1}\|_2^2$$

3. Show that

$$f(x_{i+1}) \leq f(x^*) + \frac{1}{2\eta} \|x_i - x^*\|_2^2 - \frac{1}{2\eta} \|x_i - x^* - \eta \nabla f(x_i)\|_2^2.$$

Note that

$$\frac{1}{\eta} \langle a, b \rangle - \frac{1}{2\eta} \|a\|_2^2 = -\frac{1}{2\eta} \left(-2\langle a, b \rangle + \|a\|_2^2 \right)$$

can be rewritten, by completing the square, as follows

$$\frac{1}{2\eta} \|b\|_2^2 - \frac{1}{2\eta} \left(\|b\|_2^2 - 2\langle a, b \rangle + \|a\|_2^2 \right) = \frac{1}{2\eta} \|b\|_2^2 - \frac{1}{2\eta} \|b - a\|_2^2.$$

Finally, by summing (5) for $i = 0, \dots, k-1$, show that

$$\sum_{i=0}^{k-1} (f(x_{i+1}) - f(x^*)) \leq \frac{1}{2\eta} \|x_0 - x^*\|_2^2, \quad (6)$$

$$\sum_{i=0}^{k-1} (f(x_{i+1}) - f(x^*)) \leq \frac{1}{2\eta} (\|x_0 - x^*\|_2^2 - \|x_k - x^*\|_2^2) \leq \frac{1}{2\eta} \|x_0 - x^*\|_2^2,$$

Furthermore, since by (4), $f(x_0), f(x_1), \dots, f(x_k)$ is non-increasing, we have $f(x_k) - f(x^*) \leq f(x_i) - f(x^*)$ for all $i < k$. Thus, we have

$$k(f(x_k) - f(x^*)) \leq \frac{1}{2\eta} \|x_0 - x^*\|_2^2,$$

which proves (3).