1 Part I: Choose the correct answer

1.1 Gradient Descent and smoothness

Question 1: Define $f(x) = ax^2 + b$, $x \in \mathbb{R}$. Consider running gradient descent with a constant-step size. For which one of the following statements, is it **NOT** possible to find a combination of starting point, step size and positive real numbers a and b where the statement happens at some step t?

1. $x_{t+1} < 0 < x_{t+2} < x_t$
2. $x_{t+1} < x_{t+2} < 0 < x_t$
3. $x_{t+1} < 0 < x_t < x_{t+2}$
4. $x_{t+1} \neq x_t$, $x_t = x_{t+2}$

Question 2: Define $f(x) = x^4$ with domain $D_f = [-2, 2]$. Assume we want to find a point x_T with $f(x_T) \leq \epsilon$ starting from $x_0 \in D_f$.Among the following statements, which is true and provides the tightest bound?

1. Using Nesterov acceleration and an appropriate step size, we have $T \in \mathcal{O}(\frac{1}{\epsilon})$ since f is smooth and convex over D_f.
2. Using Nesterov acceleration and an appropriate step size, we have $T \in \mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ f is smooth and convex over D_f.
3. Using Gradient Descent and an appropriate step size, we have $T \in \mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ since f is smooth and convex over D_f.
4. Using Gradient Descent and an appropriate step size, we have $T \in \mathcal{O}(\log(1/\epsilon))$ since f is smooth and strongly convex over D_f.

1.2 Subgradient Descent

Question 3: Consider the function $f(x) = 4x - x^2$ for $x \in \mathbb{R}$. The gradient and the subgradient at $x = 2$ are, respectively,

1. $0, [-1, 1]$
2. $0, \{0\}$
3. 0, does not exist
4. $2, \{2\}$

1.3 Stochastic Gradient Descent

Question 4: Assume we perform constant step-size stochastic gradient descent on $f(x) = \frac{1}{2}(f_1(x) + f_2(x))$ where $f_1(x) = (x-1)^2$ and $f_2(x) = (x+1)^2$, i.e., $x_{t+1} = x_t - a\nabla f_{i_t}(x_t)$ at each iteration i_t is chosen uniformly random in {1,2}. Which of the following statement(s) is (are) false:

1. For $a = 1$, we can guarantee that the iterates stay in a bounded set.
2. $x = 0$ is the global minimum of f.
3. Whatever the choice of constant step-size $a > 0$, the iterates cannot converge as t goes to infinity.
4. For $a = 2$, for any starting point x_0 and after the first iteration, the iterates will belong to $[-1, +1]$.

2 Part II: True or False

Question 5: If a function is strictly convex, then it is also strongly convex.

1. True
2. False

Question 6: For a convex, L-smooth function f with a global minimum and running gradient descent with step size $a = 1/L$ we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\|\nabla f(x_t)\|^2 \leq \mathcal{O}\left(\frac{1}{T}\right), \quad T > 0.$$

1. True
2. False

3 Part III: Open Questions

Question 1: Assume that f has L-Lipschitz gradients, i.e., $\forall(x,y)$, $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2$. The gradient descent update: $x_{t+1} = x_t - a\nabla f(x_t)$, with step size $a = \frac{1}{L}$ Show the following inequality

$$f(x_{t+1}) \leq f(x_t) - \frac{1}{2L}\|\nabla f(x_t)\|_2^2$$

How do you interpret the above inequality?

Question 2: Assume the same as in Question 1, show that

$$f(x_{t+1}) - f(x_t) \leq \frac{L}{2}(\|x_t - x^*\|_2^2 + \|x_{t+1} - x^*\|_2^2)$$