ELE 539 / COS 512: Homework 2

due on Mar. 16, 2021 (11:59 PM Blackboard)

1 AGD Monotonically Decreases The Hamiltonian

In class, we have showed that GD satisfies the descent lemma, i.e. $f(x_{t+1}) \leq f(x_t) - (\eta/2) \cdot \|\nabla f(x_t)\|^2$, where η is the learning rate. It can be also shown that, per iteration, AGD does not necessarily decrease the function value of the iterates. In this question, we will prove that AGD monotonically decreases "the Hamiltonian" instead. Formally, recall the update equations of Nesterov's accelerated gradient descent.

$$\begin{aligned} y_t =& x_t + \gamma(x_t - x_{t-1}), \\ x_{t+1} =& y_t - \eta \nabla f(y_t). \end{aligned}$$

[3 points] Let the Hamiltonian at t^{th} iterate be defined as $E_t := f(x_t) + \|x_t - x_{t-1}\|^2/(2\eta)$. Prove that when f is convex and ℓ-smooth, if we choose $\eta \leq 1/\ell$, and $\gamma \in [0,1]$, then we have:

$$E_{t+1} \leq E_t - \frac{1-\gamma^2}{2\eta}\|x_t - x_{t-1}\|^2.$$

2 Lower Bounds for Smooth and Strongly Convex Functions

In this question, we prove the following lower bound on the iteration complexity for optimizing smooth and strongly convex functions using "span algorithms". Recall that "span algorithms" is a class of algorithms satisfying $x_1 = 0$, and $x_{t+1} \in \text{span}(g_1, \ldots, g_t)$. Here $\{g_i\}_{i=1}^t$ are the gradients returned by the oracle in past iterates. Concretely, we aim to prove

Theorem 1. *For any $t \in \mathbb{N}$, any $\ell, \alpha \in \mathbb{R}_+$ and $\alpha \leq \ell$, for any span algorithm, there exists a α-strongly convex and ℓ-smooth function f, such that:*

$$\min_{1 \leq s \leq t} f(x_s) - f(x^\star) \geq \Omega\left(\alpha \cdot \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2t} \cdot \|x_1 - x^\star\|^2\right)$$

where $\kappa = \ell/\alpha$ is the condition number. We consider the following hard instance:

$$f(x) = \frac{\ell - \alpha}{8}\left[(e_1^\top x - 1)^2 + \sum_{i=1}^{2t-1}(e_i^\top x - e_{i+1}^\top x)^2 + \zeta(e_{2t}^\top x)^2\right] + \frac{\alpha}{2}\|x\|^2$$

where $\zeta \in [0,1]$ is a scalar that will be determined later.

(a) [1 point] Prove that for any $\zeta \in [0, 1]$, the hard function f is ℓ-smooth, and α-strongly convex.

(b) [2 points] Show that there exists a $\zeta \in [0, 1]$, such that the minimum $x^\star$ of function f satisfies:

$$e_i^\top x^\star = \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^i, \qquad \text{for any } i \in [2t].$$

(c) [2 points] Use above results to prove Theorem 1.

3 Mirror Descent for Smooth Convex Functions

In this question, we will bound the iteration complexity of mirror descent for smooth convex function. For simplicity, we consider the unconstrained problem. Formally, we prove

Theorem 2. *Let Φ be a mirror map that is ρ-strongly convex on $\mathbb{R}^d$ w.r.t $\|\cdot\|$, function f be convex and ℓ-smooth w.r.t $\|\cdot\|$. Then mirror descent with learning rate $\eta = \rho/\ell$ satisfies:*

$$f(x_t) - f(x^\star) \le \frac{\ell \cdot D_\Phi(x^\star, x_1)}{\rho(t-1)}$$

Note that $\|\cdot\|$ can be any norm, which is not necessarily 2-norm. By ℓ-smooth w.r.t $\|\cdot\|$, we mean

$$\forall x, y, \qquad \|\nabla f(x) - \nabla f(y)\|_* \le \ell \|x - y\| \tag{1}$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. This is also equivalent to

$$\forall x, y, \qquad f(y) \le f(x) + \langle \nabla f(x), y - x\rangle + \frac{\ell}{2}\|y - x\|^2. \tag{2}$$

You can directly use (1) and (2) for ℓ-smooth condition in this question.

(a) [2 points] Prove the following descent lemma:

$$f(x_{s+1}) - f(x_s) \le -\frac{1}{\eta} D_\Phi(x_s, x_{s+1}).$$

(b) [1 point] Prove the following inequality:

$$f(x_s) - f(x^\star) \le \frac{1}{\eta}\left[D_\Phi(x_s, x_{s+1}) + D_\Phi(x^\star, x_s) - D_\Phi(x^\star, x_{s+1})\right].$$

(c) [1 point] Use above results to show that the average value of the iterates is upper bounded as:

$$\frac{1}{t-1}\sum_{s=2}^{t}\left[f(x_s) - f(x^\star)\right] \le \frac{\ell \cdot D_\Phi(x^\star, x_1)}{\rho(t-1)}$$

(d) [1 point] Prove the last iterate guarantee as in Theorem 2.