Oracle complexity of Lipschitz convex functions

Huanle Xu *

May 9, 2018

1 Basic Elements of Iterative Algorithms

In this lecture we consider the unconstrained minimization of a function f that satisfy the following requirements:

- (i) f admits a minimizer x^* on $\mathbb{R}^n$ such that $||x^*|| \leq R$.
- (ii) f is convex on $\mathbb{R}^n$.
- (iii) f is L-Lipschitz on the l_2-ball of radius R, that is for any $x \in \mathbb{R}^n$ such that $||x|| \leq R$ and any $||\nabla f(x)|| \leq L$.

1.1 Gradient Descent

The simplest strategy to minimize a differentiable function is probably the Gradient Descent scheme. It is an iterative algorithm that starts at some initial point $x_1 \in \mathbb{R}^n$, and then iterates the following equation:

$$x_{t+1} = x_t - \eta \nabla f(x_t),$$

where $\eta > 0$ is the step-size fixed beforehand. The idea of this scheme is simply to do small steps that minimize the local first order Taylor approximation of the function f. When the function is convex but not differentiable it seems rather natural to replace gradients by gradients. This gives the gradient Descent strategy.

Here we need to be a bit careful if we want to analyze this scheme under assumptions (i)-(ii)-(iii). Indeed we have a control on the size of the gradients only in a ball of radius R. Furthermore we also know that the minimizer of the function lies in this ball. Thus it makes sense to enforce that if we leave the ball, then we first project back the point to the ball before taking another gradient step. This gives the (Projected) gradient Descent:

$$\begin{aligned}
&y_{t+1} = x_t - \eta g_t, \text{ where } g_t = \nabla f(x_t) \\
&\text{if } \|y_{t+1}\| \leq R \text{ then } x_{t+1} = y_{t+1}, \text{ otherwise } x_{t+1} = \frac{R}{\|y_{t+1}\|} y_{t+1}.
\end{aligned}$$

The following elementary result gives a rate of convergence for the gradient method.

Theorem 1. *Assume that f satisfies (i)-(ii)-(iii), and that $x_1 = 0$. Then the (Projected) gradient Descent with $\eta = \frac{R}{L\sqrt{t}}$ satisfies for $\bar{x}_t \in \left\{\frac{1}{t}\sum_{s=1}^{t} x_s; \text{argmin}_{1 \leq s \leq t} f(x_s)\right\}$,*

$$f(\bar{x}_t) - f^* \leq \frac{RL}{\sqrt{t}}.$$

*Huanle Xu is with the College of Computer Science and Technology, Dongguan University of Technology. E-mail: {xuhl}@dgut.edu.cn.

Proof. Using the definition of gradients, the definition of the method, and an elementary identity, one obtains

$$\begin{aligned} f(x_s) - f(x^*) &\leq g_s^\top (x_s - x^*) \\ &= \frac{1}{\eta}(x_s - y_{s+1})^\top (x_s - x^*) \\ &= \frac{1}{2\eta}\left(\|x_s - x^*\|^2 + \|x_s - y_{s+1}\|^2 - \|y_{s+1} - x^*\|^2\right) \\ &= \frac{1}{2\eta}\left(\|x_s - x^*\|^2 - \|y_{s+1} - x^*\|^2\right) + \frac{\eta}{2}\|g_s\|^2. \end{aligned}$$

Now note that $\|g_s\| \leq L$, and $\|y_{s+1} - x^*\| \geq \|x_{s+1} - x^*\|$ (just do a picture to convince yourself). Summing the resulting inequality over s, and using that $\|x_1 - x^*\| \leq R$ yield

$$\sum_{s=1}^{t} \left(f(x_s) - f(x^*)\right) \leq \frac{R^2}{2\eta} + \frac{\eta L^2 t}{2}.$$

Plugging in the value of η directly gives the statement (recall that by convexity $f(\bar{x}_t) \leq \frac{1}{t}\sum_{s=1}^{t} f(x_s)$). □