Accelerated Mirror descent methods for overparametrized networks

Kartik Srinivas*
Department of Computer Science
Indian Institute of Technology, Hyderabad
Sangareddy, Telangana
arkartik@student.ubc.ca

Abstract

An analysis of Stochastic Mirror descent, Proximal Mirror descent and their respective accelerated versions on overparametrized networks

Theoretical Results

The function $\Phi(x) = \frac{1}{p}\|x\|_p^p$ is strongly convex with a parameter.

The map used in the problem takes the gradients to a different space namely the dual space of the system. The question is, what exactly is the norm that is being used on both of the sides.

The update equation is based on the fenchel conjugate of the function and is as follows

$$\nabla w^*(y) = \text{argmax}_{x \in X} \langle x, y \rangle - w(x) \tag{1}$$

Using this we get

$$\nabla w^*(y) = \text{argmax}_{x \in X} \langle x, y \rangle - \frac{1}{p}\|x\|_p^p$$

$$y = \nabla_x \frac{\sum |x_j^*|^p}{p}$$

$$y_j = |x_j^*|^{p-1}\text{sgn}(x_j)$$

$$x_j^* = |y_j|^{1/(p-1)}\text{sgn}(y_j)$$

Therefore the update step is

$$x_{t+1} = |\nabla w(x_t) - \eta \nabla f(x_t)|^{\frac{1}{p-1}}\text{sgn}(\nabla w(x_t) - \eta \nabla f(x_t)) \tag{2}$$

*Also affiliated with the University of British Columbia, Vancouver at the time of this work

Preprint. Under review.