

# CSC 576: Accelerated Gradient Descent Algorithm

Ji Liu

Department of Computer Sciences, University of Rochester

November 18, 2014

## 1 Introduction

The accelerated gradient descent (AGD) algorithm is an advanced version of the GD algorithm. The AGD algorithm was first proposed by Nesterov in 1983. AGD is proven to be the optimal method among all gradient based (or first order) algorithms. However, the original AGD algorithm can only handle smooth convex optimization problem. Recent progresses extend AGD to a more general type of convex problem:

$$\min_x f(x) + g(x),$$

where  $f(x)$  is convex smooth and  $g(x)$  is convex and closed. Similar optimal convergence rate can be obtained.

## 2 Algorithm

The AGD algorithm can be summarized in 1: There are two ways to decide the steplength  $\gamma$ .

---

**Algorithm 1** AGD

---

**Require:**  $x_1, K$

**Ensure:**  $x_K$ ;

Initialize  $t_1 = 1, y_1 = x_1$ ;

**for**  $k = 1 : K$  **do**

$$x_{k+1} = \text{prox}_{\gamma g(\cdot)}(y_k - \gamma \nabla f(y_k));$$

$$t_{k+1} = \frac{1}{2} + \frac{1}{2} \sqrt{1 + 4t_k^2};$$

$$y_{k+1} = x_{k+1} + \left(\frac{t_k - 1}{t_{k+1}}\right) (x_{k+1} - x_k);$$

**end for**

---

First, similar to the gradient descent method, to ensure convergence we can safely set  $\gamma$  to be small enough, in particular,  $\gamma \leq (\|\nabla^2 f(x)\|^{-1}) \forall x$ . Second, we can apply the line search scheme to adaptively decide the steplength such that

$$f(x_{k+1}) \leq m_{y_k, \gamma}(x_{k+1}).$$

(see early notes for the line search scheme).

Note that three properties of the sequence  $\{t_k\}$ :

- $\{t_k\}$  is positive and increasing;
- $t_{k+1} \geq t_k + \frac{1}{2}$ ;
- $\frac{t_0-1}{t_1} = 0$  and  $\lim_{t \rightarrow \infty} \frac{t_k-1}{t_{k+1}} = 1$ .

### 3 Convergence Rate

AGD is the optimal gradient based approach, because it provides the optimal convergence rate. Assume the Lipschitz condition (as our early note) in the following.

**Assumption 1.** *Assume that the convex smooth function  $f(x)$  has a Lipschitz gradient, that is, there exists a constant  $L$  such that*

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

Under this assumption, if the steplength is chosen to be small enough or decided by the line search, we have the convergence rate

$$F(x_k) - F^* \leq O\left(\frac{1}{k^2}\right).$$

Recall that the convergence rate for GD is just  $O(1/k)$ .