CS 270: Combinatorial Algorithms

Professor: Prasad Raghavendra, Spring 2021
Scribe: Vishal Raman

Contents

§1 January 20th, 2021

§1.1 Intro to Gradient Descent

Problem: Given a function f, we wish to minimize f. Gradient Descent takes the natural approach of going in the direction of locally steepest decrease.

§1.2 Convexity and Convex Functions

Definition 1.1. Convexity A set $K \subset \mathbb{R}^n$ is convex if and only if for all $x, y \in K$, the line segment joining x, y is also in K. In other words, for $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda)y \in K.$$

Definition 1.2 (Convex Function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for all x, y, $lambda \in [0, 1]$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

> **Remark 1.3.** Visually, if we draw the line segment between two points on the graph of the function, it should lie above the function.

Definition 1.4. A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

Recall for $u, v \in \mathbb{R}^n$.

$$\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i,$$

and

$$\|u\|_2 = \sqrt{\sum u_i^2}.$$

Definition 1.5. A twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if

$$H_f(x) \succcurlyeq 0,$$

in other words, the Hessian is positive-semidefinite.

§1.3 Unconstrained Optimization

We have the following problem: for $f : \mathbb{R}^n \to \mathbb{R}$ convex, we wish to find $\min_{x \in \mathbb{R}^n} f(x)$.

Algorithm: for x_0, the initial point, $x_{t+1} \leftarrow x_t - \eta \nabla f(x_t)$, for a parameter η. We output the average of the points $\frac{1}{T} \sum_i x_i$.

If we let $x^* = \text{argmin}_x f$, then

$$f(x^*) \geq f(x) + \langle \nabla f(x), x^* - x \rangle.$$

It follows that

$$\langle -\nabla f(x), x^* - x \rangle \geq f(x) - f(x^*) \geq 0.$$

Theorem 1

After t steps(for appropriate η),

$$\frac{1}{t}\sum_{i=1}^{t} f(x_i) \leq f(x^*) + O(\frac{RL}{\sqrt{t}})$$

where $R = \|x_0 - x^*\|$ and f is L-Lipschitz: for all x, y $|f(x) - f(y)| \leq L\|x - y\|$.

§1.4 Constrained Optimization

Given a convex set $K \subset \mathbb{R}^n$, we minimize a convex function f.

Algorithm: We have an initial point x_0, $y_{t+1} = x_t - \eta\nabla f(x_t)$. Then $x_{t+1} = \pi_K(y_{t+1})$, where π_K is a projection onto K, defined by

$$\pi_K(y) = \operatorname*{argmin}_{z\in K} \|z - y\|.$$

For convex sets, the same theorem holds, since

$$\|y_{t+1} - x_*\| \geq \|\pi_K(y_{t+1}) - x^*\|.$$

Definition 1.6. A function is α-strongly convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \alpha(\lambda(1-\lambda))\|x - y\|^2.$$

Definition 1.7. A function is β smooth if

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y) - \frac{\beta}{2}(\lambda(1-\lambda))\|x - y\|^2.$$

Theorem 2

After t steps, we can get convergence(for appropriate η)

$$\frac{1}{t}\sum_{i} f(x_i) - f(x^*) \leq e^{-\alpha t/\beta}.$$

We call β/α the condition number.

§2 January 25th, 2021

§2.1 Gradient Descent, Continued

Theorem 3

After t steps(for appropriate η),

$$\frac{1}{t}\sum_{i=1}^{t} f(x_i) \le f(x^*) + O(\frac{RL}{\sqrt{t}})$$

where $R = \|x_0 - x^*\|$ and f is L-Lipschitz: for all x, y $|f(x) - f(y)| \le L\|x - y\|$.

Proof. We will use the fact

$$2\langle a - c, b - c\rangle = \|a - c\|^2 + \|b - c\|^2 - \|a - b\|^2.$$

Note that

$$\begin{aligned} f(x_t) - f(x^*) &\le \langle \nabla f(x_t), x_t - x^*\rangle \\ &\le \left\langle \frac{x_t - x_{t+1}}{\eta}, x_t - x^*\right\rangle \\ &\le \frac{1}{2\eta}\left(\|x_t - x_{t+1}\|^2 + \|x_t - x^*\|_2 - \|x_{t+1} - x^*\|^2\right) \end{aligned}$$

It follows that

$$\sum_{i=1}^{t}(f(x_i) - f(x^*)) \le \frac{1}{2\eta}\sum_{i=1}^{t}\|x_{i+1} - x_i\|^2 + \frac{1}{2\eta}\|x_1 - x^*\|^2.$$

Then, $\|x_{i+1} - x_i\|^2 = \|\eta\nabla f(x_i)\|^2 = \eta^2 L^2$ so our expression

$$\sum_{i=1}^{t}(f(x_i) - f(x^*)) \le \frac{\eta R^2 t}{2} + \frac{R^2}{2\eta}.$$

Choosing $\eta = \frac{R}{L\sqrt{t}}$ gives the desired bound. $\square$

§2.2 Projected Gradient Descent