Course Notes for EE227C (Spring 2018): Convex Optimization and Approximation

Instructor: Moritz Hardt

Email: hardt+ee227c@berkeley.edu

Graduate Instructor: Max Simchowitz

Email: msimchow+ee227c@berkeley.edu

January 23, 2018

2 Lecture 2: Gradient Descent

2.1 Gradient Descent

The procedure of gradient descent is defined by the recursion:

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

where η is the step size. This works to solve the problem

$$\min_{x \in \Omega} f(x)$$

for f convex, differentiable, and L-Lipschitz

Definition 2.1 (L-Lipschitz)**.** A function is said to be *L-Lipschitz* if its gradient is bounded,

$$\|\nabla f(x)\| \leqslant L$$

Fact 2.2. *f(x) is L-Lipschitz implies that the difference between two points in the range is bounded,*

$$|f(x) - f(y)| \leqslant L\|x - y\|$$

Question 2.1. How do we ensure that $x_{t+1} \in \Omega$?

Solution: Project onto Ω

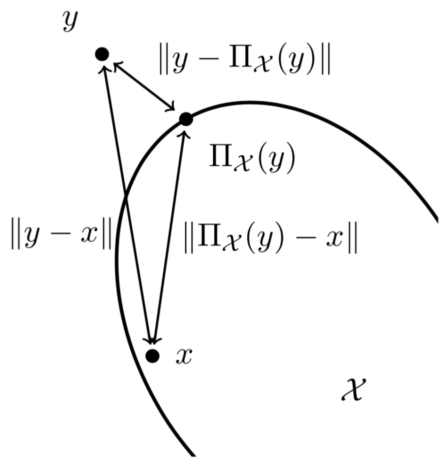

Figure 1: Projection of y onto set $\mathcal{X}$.

Definition 2.3 (Projection)**.** The *projection* of a point y onto a set Ω is defined as

$$\Pi_\Omega(x) = \operatorname*{argmin}_{y \in \Omega} \|x - y\|$$

Example 2.4. A projection onto the Euclidean ball B_2 is just normalization:

$$\Pi_{B_2}(x) = \frac{x}{\|x\|}$$

The crucial property of projections is that they satisfy the following condition:

$$\|\Pi_\Omega(y) - x\|^2 \leqslant \|y - x\|^2$$

i.e. the projection of y onto a convex set containing x is closer to x. See Figure **??** for a geometric picture.

Lemma 2.5.

$$\|\Pi_\Omega(y) - x\|^2 \leqslant \|y - x\|^2 - \|y - \Pi_\Omega(y)\|^2$$

Which follows from the Pythagorean theorem. Note that this lemma implies the above property.

2.1.1 Modifying Gradient Descent with Projections

So now we can modify our original procedure to use two steps.

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

$$x_{t+1} = \Pi_\Omega(y_{t+1})$$

And we are guaranteed that $x_{t+1} \in \Omega$. Note that computing the projection may be the hardest part of your problem, as you are computing an argmin. However, there are convex sets for which we know explicitly how to compute the projection (see Example 2.4).

Theorem 2.6 (Projected Gradient Descent for Lipchitz Functions)**.** *Assume that function f is convex, differentiable, and closed with bounded gradients. Let L be the Lipchitz constant of f over the convex domain Ω. Let R be the upper bound on the distance from the initial point x_1 to the optimal point $x^* = \arg\min_{x \in \Omega} f(x)$ (i.e. $\|x_1 - x^*\|_2$). Let t be the number of iterations of project gradient descent.*

If the learning rate η is set to $\eta = \frac{R}{L\sqrt{(t)}}$, then $f\left(\frac{1}{t}\sum_{s=1}^{t} x_s\right) - f\left(x^*\right) \leqslant \frac{RL}{\sqrt{t}}$.

This means that the difference between the functional value of the average point during the optimization process from the optimal value is bounded above by a constant proportional to $\frac{1}{\sqrt{t}}$.

Before proving the theorem, recall that

- First order characterization of convexity: $f(y) = f(x) + \nabla f(x)^\top (y - x)$
- "Fundamental Theorem of Optimization": An inner product can be written as a sum of norms: $u^\top v = \frac{1}{2}(\|u\|^2 + \|v\|^2 - \|u - v\|^2$. This property can be seen by writing $\|u - v\|$ as $\|u - v\| = \|u\|^2 + \|v\|^2 - 2u^\top v$.
- L-Lipchitz: For all x, $\|\nabla f(x)\| \leqslant L$.
- $\|\Pi_\Omega(y) - x\|^2 \leqslant \|y - x\|^2 - \|y - \Pi_\Omega(x)\|^2$

Proof of Theorem 2.6 for compact sets. The proof begins by first bounding the difference in function values $f(x_s) - f(x^*)$.

$$f(x_s) - f(x^*) \leqslant \tag{1}$$

■