CS 4/527, Gradient Descent

Jared Saia
University of New Mexico

The Problem

Given:

- Convex function f

Goal: Find x that minimizes $f(x)$

Variables

- $D = max_{x,y}|x - y|$
- G is an upperbound on $|\nabla f(x)|$ for any x considered

Note: all norms are 2-norms. D is the *diameter* of the search space

Convexity - Another View

A convex function is **above the tangent plane at any point.** In algebra, we can say

$$f(x+z) \geq f(x) + \nabla f(x) \cdot z, \text{ for all } x, z$$

(Recall that $\nabla f(x)$ is the vector whose i-th coordinate is $\partial f / \partial x_i$)

This is equivalent to:

$$f(x) - f(y) \leq \nabla f(x) \cdot (x - y) \text{ for all } x, y$$

Gradient Descent Algorithm

$\alpha \leftarrow \frac{D}{G\sqrt{T}}$

Repeat for $i = 0$ to T:

1. $x_{i+1} \leftarrow x_i - \alpha \nabla f(x_i)$

Output $z = \frac{1}{T}\sum_i x_i$

Theorem 1

Theorem 1 *Let x^* be the value that minimizes f. Then, for any $\epsilon > 0$, if we set $T = \frac{4D^2G^2}{\epsilon^2}$, gradient descent ensures:*

$$f(z) \leq f(x^*) + \epsilon$$

Proof (I)

$$
\begin{aligned}
|x_{i+1} - x^*|^2 &= |x_i - x^* - \alpha \nabla f(x_i)|^2 \\
&= |x_i - x^*|^2 + \alpha^2 |\nabla f(x_i)|^2 - 2\alpha \nabla f(x_i) \cdot (x_i - x^*)
\end{aligned}
$$

First step holds by definition of x_{i+1}. Last step holds by noting that $|v|^2 = v \cdot v$ and using linearity.

Proof (II)

From last slide, we have:

$$|x_{i+1} - x^*|^2 \;\leq\; |x_i - x^*|^2 + \alpha^2 |\nabla f(x_i)|^2 - 2\alpha \nabla f(x_i) \cdot (x_i - x^*)$$

Reorganizing, and using definition of G, we get:

$$\nabla f(x_i) \cdot (x_i - x^*) \;\leq\; \frac{1}{2\alpha}(|x_i - x^*|^2 - |x_{i+1} - x^*|^2) + \frac{\alpha}{2} G^2$$

Using Slide 3, we then get:

$$f(x_i) - f(x^*) \;\leq\; \frac{1}{2\alpha}(|x_i - x^*|^2 - |x_{i+1} - x^*|^2) + \frac{\alpha}{2} G^2$$

Proof (III)

Now sum last inequality for $i = 1$ to T. After cancellations, we get.

$$\sum_{i=1}^{T} (f(x_i) - f(x^*)) \;\leq\; \frac{1}{2\alpha}(|x_1 - x^*|^2 - |x_T - x^*|^2) + \frac{T\alpha}{2}G^2$$

Divide the above inequality by T. By convexity, $f(\frac{1}{T}(\sum_i x_i)) \leq \frac{1}{T}\sum_i f(x_i)$. Since $z = \frac{1}{T}\sum_i x_i$, we now get

$$f(z) - f(x^*) \leq \frac{D^2}{2\alpha T} + \frac{\alpha}{2}G^2.$$

Since $\alpha = \frac{D}{G\sqrt{T}}$, the right hand side is at most $2\frac{DG}{\sqrt{T}}$. Then since $T = \frac{4D^2G^2}{\epsilon^2}$, we see that $f(z) \leq f(x^*) + \epsilon$