15-451/651 Algorithms, Spring 2019
Recitation #10B (Optional) Worksheet

Convex Functions

Recall that a function f over $\mathbf{R}^n$ is *convex* if for any two inputs $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ and any $\lambda \in [0, 1]$ we have

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

In other words, the line segment from $(\mathbf{x}, f(\mathbf{x}))$ to $(\mathbf{y}, f(\mathbf{y}))$ stays "above" the function. Alternatively, if the function is differentiable then it is convex iff for all $\mathbf{x}, \mathbf{y}$, we have

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

Moreover, recall that a set $K \subseteq \mathbf{R}^n$ is *convex* if for any two points $\mathbf{x}, \mathbf{y} \in K$ and any $\lambda \in [0, 1]$ we have that the point $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \in K$. In other words, the line segment from $\mathbf{x}$ to $\mathbf{y}$ stays inside the set.

Gradient Descent

In the lecture notes (Theorem 7) we show that starting with the point $\mathbf{x}_0$ and using the gradient descent rule $\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - \eta \nabla f_t(\mathbf{x}_t)$ at each step, we get that for any fixed point $\mathbf{x}^*$, we have the following bound.

$$\sum_{t=0}^{T-1} f_t(\mathbf{x}_t) \leq \sum_{t=0}^{T-1} f_t(\mathbf{x}^*) + \underbrace{\frac{\eta}{2} G^2 T + \frac{1}{2\eta} \|\mathbf{x}_0 - \mathbf{x}^*\|^2}_{\text{regret}(T)}, \tag{1}$$

where G is an upper bound on the norm of the gradient $\|\nabla f(\mathbf{x})\|$. Let's see how to use this to find a point $\hat{\mathbf{x}}$ at which the function value is very close to the minimum value.

1. Suppose we get a fixed function $f_t = f$ at each step. From the above expression, show that if we set $\hat{\mathbf{x}} = \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{x}_t$, then

$$f(\hat{\mathbf{x}}) \leq f(x^*) + \frac{\text{regret}(T)}{T}.$$

Solution: Use the first definition of convexity to show that

$$f(\hat{\mathbf{x}}) = f\left(\frac{1}{T} \sum_{t=0}^{T-1} \mathbf{x}_t\right) \leq \frac{1}{T} \sum_{t=0}^{T-1} f(\mathbf{x}_t).$$

This is $1/T$ times the LHS of (1). Which is at most $1/T$ times the RHS of (1), which is $f(\mathbf{x}^*) + \text{regret}(T)/T$.

2. Suppose we set the "learning rate" $\eta = \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|}{G\sqrt{T}}$. Show that regret$(T) \leq G\|\mathbf{x}_0 - \mathbf{x}^*\|\sqrt{T}$.

   Solution: Substituting and doing some simple algebra.

3. Combining the above two parts, show that after $T = \left(\frac{G\|\mathbf{x}^* - \mathbf{x}_0\|}{\varepsilon}\right)^2$ steps, the function value $f(\hat{\mathbf{x}}) \leq f(\mathbf{x}^*) + \varepsilon$.

   Solution: Substituting and doing some simple algebra.

Now let's see what we can get in a setting like that for the experts algorithm. Suppose you know the function $f(\mathbf{x}) = \sum_i c_i x_i$ for some $\mathbf{c} = (c_1, \ldots, c_n) \in [0, M]^n$ (i.e., f is linear) and suppose we have some convex body K contained within the unit cube: i.e., $K \subseteq \{\mathbf{x} \mid 0 \leq x_i \leq 1 \ \forall i \in \{1, 2, \ldots, n\}\}$.

4. What is the diameter of K? (The diameter is the maximum Euclidean distance between two points in K.)

   Solution: The maximum distance is bounded by the max-distance between $(0, 0, \ldots, 0)$ and $(1, 1, \ldots, 1)$, which is $\sqrt{1^2 + 1^2 + \ldots + 1^2} = \sqrt{n}$.

5. If you start with some $\mathbf{x}_0 \in K$, give an upper bound on $\|\mathbf{x}_0 - \mathbf{x}^*\|$.

   Solution: $\|\mathbf{x}_0 - \mathbf{x}^*\|$ is at most the diameter of the cube, so setting $D := \sqrt{n}$ suffices.

6. What is the maximum value of $\|\nabla f(x)\|$ at any point $\mathbf{x} \in K$?

   Solution: $\nabla f(x) = \nabla(c_1 x_1 + \ldots + c_n x_n) = \mathbf{c}$, so $\|\nabla f(x)\| = \|\mathbf{c}\| \leq M\sqrt{n}$. Hence you can set $G = M\sqrt{n}$.

7. Plugging these values in, what expressions do you get for T, η?

   Solution: Recall $T = (\frac{\|\mathbf{x}_0 - \mathbf{x}^*\| G}{\varepsilon})^2 = (\frac{\sqrt{n} \cdot M\sqrt{n}}{\varepsilon})^2 = (\frac{Mn}{\varepsilon})^2$. Substituting, $\eta = \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|}{G\sqrt{T}} = \frac{\varepsilon}{M^2 n}$.