COS-511: Learning Theory **Spring 2017**

Lecture 23: Bandit Convex Optimization

Lecturer: Roi Livni

23.0.1 Online Routing

One could model the online routing problem as a multi-armed bandit problem. Each of the N "arms" of the bandit is a path throughout the network; the loss function measures the time it takes a packet to travel along that path. This approach would work, but the number of paths throughout the network scales exponentially with the number of nodes. Can we do better?

Recall that the dimension of the flow polytope scales *polynomially* with the number of nodes in the network. If we could optimize over the flow polytope directly, we might obtain a better regret bound.

This motivates a more general setting, called **bandit convex optimization**, which is OCO minus the gradient.

23.1 Bandit Convex Optimization

In bandit convex optimization (BCO), as in online convex optimization, the player's goal is to play some $\mathbf{x}_t \in \mathcal{K}$ in the t-th round so as to minimize regret:

$$\text{Regret}_T = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \min_{\mathbf{x}\in\mathcal{K}} \sum_{t=1}^{T} f_t(\mathbf{x})$$

In BCO, however, the player is given far more limited feedback than in OCO: only a single number, $f_t(\mathbf{x}_t) \in \mathbb{R}$, instead of the whole loss function $f_t : \mathcal{K} \to \mathbb{R}$. Can we still do online learning without a gradient?

23.1.1 FKM Algorithm

The idea behind the FKM (Flaxman, Kalai, and McMahan) algorithm is to follow an unbiased estimator of the gradient.

For a function in one dimension $f : \mathbb{R} \to \mathbb{R}$, an unbiased estimator of the derivative is given by:

$$\tilde{f}'(x) = \begin{cases} \frac{1}{\delta} f(x+\delta) & \text{w.p. } \frac{1}{2} \\ -\frac{1}{\delta} f(x-\delta) & \text{w.p. } \frac{1}{2} \end{cases}$$

As $\delta \to 0$, the estimator $\tilde{f}'(x)$ tends toward $f'(x)$ in expectation:

$$\lim_{\delta\to 0} \mathbf{E}\left[\tilde{f}'(x)\right] = \lim_{\delta\to 0} \frac{f(x+\delta) - f(x-\delta)}{2\delta} = f'(x)$$

It turns out that this approach also works in higher dimension, and we will take as an (approximate) unbiased estimator for the gradient, the following one point sample:

$$\tilde{\nabla} f_\delta(\mathbf{x}) = \frac{d}{\delta} f(\mathbf{x} + \delta \mathbf{v})\mathbf{v} \quad \text{where} \quad \mathbf{v} \text{ drawn uniformly over } d\text{-dimensional unit sphere } \mathbb{S}_{d-1}$$

This estimate can in fact be interpreted as an estimate of a smoothed version of f around $\mathbf{x}$, specifically by Stoke's theorem we have the following:

If we let $\hat{f}^\delta(\mathbf{x}) = \mathbb{E}_{\|\mathbf{u}\| \le 1}\left[f(\mathbf{x} + \delta \mathbf{u})\right]$ then

$$\nabla \hat{f}^\delta(\mathbf{x}) = \frac{d}{\delta} \mathop{\mathbb{E}}_{\|\mathbf{v}\| \sim S_{d-1}} \left[f(\mathbf{x} + \delta \mathbf{v})\mathbf{v}\right]$$

How can we cheaply sample a vector $\mathbf{v}$ uniformly over the unit sphere? (That is, $||\mathbf{v}|| = 1$ and all directions should be equally likely.) One way is to sample each element of $\mathbf{v}$ independently from the standard normal distribution, and then scale $\mathbf{v}$ to have unit norm.

Assume for simplicity that $\mathbf{0} \in \mathcal{K}$.

where $\mathcal{K}_\delta$ is defined as:

$$\mathcal{K}_\delta = \left\{ \mathbf{x} \in \mathbb{R}^d : \frac{\mathbf{x}}{1-\delta} \in \mathcal{K} \right\}$$

We project onto this "shrunken" decision set $\mathcal{K}_\delta$ instead of the original decision set $\mathcal{K}$ in order to ensure that $\mathbf{x}_t + \delta \mathbf{v}$ lies within the domain of $f_t : \mathcal{K} \to \mathbb{R}$.

Algorithm 1 FKM

Set $\mathbf{x}_1 \in \mathcal{K}$ arbitrary
for $t = 1, 2, ...$ to T **do**
 $\mathbf{y}_t = \mathbf{x}_t + \delta \mathbf{v}$, $\mathbf{v} \sim S_n$ the sphere uniformly
 Play $\mathbf{y}_t$, suffer loss $f_t(\mathbf{y}_t)$
 Update $\mathbf{x}_{t+1} = \Pi_{K_\delta}[\mathbf{x}_t - \tilde{\nabla}_t]$, $\tilde{\nabla}_t = \frac{n}{\delta} f_t(\mathbf{y}_t) \cdot \mathbf{v}$
end for

and the shrunken set is defined as $\mathcal{K}_\delta = \{x | \frac{x}{1-\delta} \in \mathcal{K}\}$ to avoid moving outside of $\mathcal{K}$ when we add the sampling from the sphere. FKM achieves a regret bound according to the following theorem.

Theorem 23.1. *The FKM algorithm attains a regret of* $O(dT^{\frac{3}{4}})$.

One can use the FKM algorithm for, say, the online routing problem, and the regret will scale polynomially, rather than exponentially, with the number of nodes in the network. To prove this, we begin with the following two lemmas.

Lemma 23.2. $\forall \mathbf{x} \in \mathcal{K}_\delta, B_\delta(\mathbf{x}) = \{\mathbf{y} | \mathbf{y} = \mathbf{x} + \delta \mathbf{u}\} \subseteq \mathcal{K}$

Lemma 23.3. $\forall \mathbf{x}^* \in \mathcal{K}, \exists \mathbf{x}^*_\delta \in \mathcal{K}_\delta$ *s.t.* $|\mathbf{x}^* - \mathbf{x}^*_\delta| = O(\delta)$

Proof. First, note that

$$\mathbb{E} \sum_{t=1}^{T} [f_t(\mathbf{y}_t) - f_t(\mathbf{x}^*)] \le \mathbb{E}[\sum_{t=1}^{T} f_t(\mathbf{y}_t)] - f_t(\mathbf{x}^*_\delta) + \delta T G$$

Next, note that $|\hat{f}^\delta(\mathbf{x}) - f(\mathbf{x})| \leq \mathbb{E}_{\|\mathbf{u}\|\leq 1} |f(\mathbf{x} + \delta \mathbf{u}) - f(\mathbf{x})| \leq \delta G$. Thus we have

$$\begin{aligned}
\mathbb{E}[\text{Regret}_T] &\leq \mathbb{E} \sum_{t=1}^{T} f_t(\mathbf{y}_t) - f_t(\mathbf{x}_\delta^*) + \delta T G \\
&\leq \mathbb{E} \sum_{t=1}^{T} [\hat{f}_t^\delta(\mathbf{x}_t) - \hat{f}_t^\delta(\mathbf{x}_\delta^*)] + 3\delta T G \\
&\leq \text{Regret}_T(OGD)(\tilde{\nabla}_1, \ldots, \tilde{\nabla}_T) + 3\delta T G \\
&\leq \frac{D^2}{\eta} + \eta \sum_{t=1}^{T} |\tilde{\nabla}_t|^2 + 3\delta T G && \text{by OGD regret bound} \\
&\leq \frac{D^2}{\eta} + \eta T \frac{n^2}{\delta^2} + 3\delta T G && \text{by definition of } \tilde{\nabla}_t \\
&= O(\sqrt{nGD} T^{\frac{3}{4}}) && \text{taking } \eta = \frac{\delta}{n} D \cdot \sqrt{T}, \delta = \frac{\sqrt{nD}}{\sqrt{G} T^{1/4}}
\end{aligned}$$

□