THEORETICAL MACHINE LEARNING COS 511

LECTURE #19

APRIL 12, 2016

LECTURER: ELAD HAZAN SCRIBE: EDDIE ZHOU

Last lecture, we introduced the multi-armed bandit problem, or learning with partial information. We are motivated now to examine problems exponential in the number of decisions.

1 FKM Algorithm for BCO

Below we outline the FKM algorithm introduced last lecture, also known as "Gradient Descent without the Gradient".

Algorithm 1 FKM

```
procedure
    Set x_1 ∈ K arbitrary
    for t = 1, 2, ... to T do
        y_t = x_t + δu, u ~ S_n the sphere uniformly
        Play y_t, suffer loss f_t(y_t)
        Update x_{t+1} = ∏_{K_δ}[x_t − ∇̃_t], ∇̃_t = (n/δ) f_t(y_t) · u
    end for
end procedure
```

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

end procedure

To see the intuition behind FKM, recall that $\mathrm{E}[\tilde{\nabla}_t] = \nabla \hat{f}_t^\delta(\mathbf{x}_t)$ as explained in Lecture 18 using Stoke's Theorem. We have

$$\lim_{\delta \to 0} \frac{f(x+\delta) - f(x-\delta)}{2\delta} = f'(x)$$

$$\frac{f(x+\delta) - f(x-\delta)}{2\delta} = \hat{f}^\delta(x) \quad \text{if } \delta \nrightarrow 0$$

$$\hat{f}^\delta(x) = \mathop{\mathrm{E}}_{u \sim Ball}[f(x+\delta u)]$$

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 1.1. $E\,Regret(FKM) = O(T^{\frac{3}{4}})$

To prove this, we begin with the following two lemmas.

Lemma 1.2. $\forall \mathbf{x} \in \mathcal{K}_\delta, B_\delta(\mathbf{x}) = \{\mathbf{y} | \mathbf{y} = \mathbf{x} + \delta\mathbf{u}\} \subseteq \mathcal{K}$

Lemma 1.3. $\forall \mathbf{x}^* \in \mathcal{K}, \exists \mathbf{x}_\delta^* \in \mathcal{K}_\delta$ *s.t.* $|\mathbf{x}^* - \mathbf{x}_\delta^*| = O(\delta)$

Proof. First, note that

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

Thus we have

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

□

2 BLO (Bandit Linear Optimization)

In these some cases of Bandit Convex Optimization, the loss functions are linear. Many interesting examples can be reduced to or already have such loss functions (e.g. online routing). In this case, there exists an efficient algorithm that achieves a tight regret bound of $O(\sqrt{T})$. In ordinary BCO, we lost tightness at three different places:

- nonlinearity of loss functions: normally, $\hat{f}_\delta \neq f$
- the shrunken version of $\mathcal{K}$, $\mathcal{K}_\delta$, is needed for the exploration space
- $|\tilde{\nabla}_t| \sim \frac{1}{\delta^2}$, too large

The first issue is solved by the nature of the loss functions: for linear functions, we have $\hat{f}_\delta = f$. To solve the second and third issues, we make use of **self-concordant barrier functions**.

Definition 2.1. (self-concordant barrier function): $\forall \mathcal{K} \subseteq \mathbb{R}^n, \exists$ function $R_\mathcal{K}$ which is a self-concordant barrier function with "nice" differential properties (including smoothness and convexity). First, $\nabla^2 R_K(\mathbf{x}) \forall \mathbf{x} \in \mathcal{K}$ defines an ellipsoid $\subseteq \mathcal{K}$, the Dikin ellipsoid. Then, we have $\varepsilon_\mathbf{x} = \{\mathbf{y} | (\mathbf{y}-\mathbf{x})^T \nabla^2 R_\mathcal{K}(\mathbf{x})(\mathbf{y}-\mathbf{x}) \le 1\}$. If $R_\mathcal{K}$ is a self-concordant barrier function, we have $\varepsilon_\mathbf{x} \subseteq \mathcal{K}$.

Example 2.2. $\forall$ polytope $\{Ax \le b\}$, $R(\mathbf{x}) = -\frac{1}{m}\sum_{i=1}^{m} \log(A_i \mathbf{x} - b_i)$ is a self-concordant barrier function.

Example 2.3. For the ball, $R(\mathbf{x}) = -\log(1 - \|\mathbf{x}\|^2)$ is a self-concordant barrier function.

Next, we introduce an algorithm that makes use of self-concordant barrier functions to address the loss of tightness detailed earlier.

Algorithm 2 Scrible Algorithm [Abernethy, Hazan, and Rakhlin]

```
procedure
    Set x_1 ∈ K arbitrary
    for t = 1, 2, ... to T do
        A_t = [∇^2 R(x_t)]^{-1/2}]
        y_t = x_t + A_t u, u ~ S_n the sphere uniformly
        ∇̃_t = n f_t(y_t) A_t^{-1} u
        x_{t+1} = argmin_{x∈K} {∑_{i=1}^t ∇_i · x + (1/η) R(x)}
    end for
end procedure
```

Note that because of how we defined A_t, we have $\mathbf{u}^T A_t^T \nabla^2 A_t \mathbf{u} \leq 1$, which solves our third issue. The use of the Dikin ellipsoid from the self-concordant barrier function removes the need to shrink our $\mathcal{K}$, while still maintaining the exploration space. The Scrible algorithm therefore provides a tight regret bound:

Theorem 2.4. $E[Regret(Scrible)] \leq O(\sqrt{T}\log T)$

Proof.

$$\begin{aligned}
\mathrm{E}[\sum_{t=1}^{T} f_t(\mathbf{y}_t) - f_t(\mathbf{x}^*)] &= \mathrm{E}[\sum_{t=1}^{T} \nabla_t(\mathbf{y}_t - \mathbf{x}^*)] && \text{by linearity of loss function} \\
&= \mathrm{E}[\sum_{t=1}^{T} \tilde{\nabla}_t(\mathbf{x}_t - \mathbf{x}^*)] && \text{by } \mathrm{E}[\mathbf{y}_t] = \mathbf{x}_t, \mathrm{E}[\tilde{\nabla}_t] = \nabla_t \\
&\leq \mathrm{Regret}(RFTL) \text{ on } \{\tilde{\nabla}_t\} && \text{recognizing RFTL} \\
&\leq \frac{1}{\eta} \cdot D_R + \eta \sum_{t=1}^{T} (\|\tilde{\nabla}_t\|_t)^2 && \text{by RFTL Theorem} \\
&= \frac{1}{\eta} \cdot D_R + \eta \left[(n f_t(\mathbf{y}_t))^2 \mathbf{u}^T A_t^{-1} \nabla^2 R(\mathbf{x}_t) A_t^{-1} \mathbf{u}\right] \\
&\leq \frac{1}{\eta} \cdot D_R + \eta n T
\end{aligned}$$

Now, using the fact that

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

we have

$$\begin{aligned}
&\sum_{t=1}^{T} \mathrm{E}[f_t(\mathbf{x}_t)] - \sum_{t=1}^{T} f_t(\mathbf{x}^*) \\
&\leq \frac{1}{\eta} \cdot D_R + \eta n T + \delta T G D \\
&= \frac{\mathcal{R}(\mathbf{x}_\delta^*) - \mathcal{R}(\mathbf{y}_1)}{\eta} + \eta n T + \delta T G D \\
&\leq \frac{\nu \log \frac{1}{\delta}}{\eta} + \eta n T + \delta T G D
\end{aligned}$$

After choosing $\eta = O(\frac{1}{\sqrt{T}})$ and $\delta = O(\frac{1}{T})$, we arrive at the desired regret bound. □

Example 2.5. Online Shortest Path

$$
\begin{aligned}
f_t(path) &= \sum_{e \in path} \mathbf{w}_e^t \\
\mathbf{x}_t \in \mathcal{K} &= \mathbb{R}^{|E|} \text{ flow polytope} \\
f_t(\mathbf{x}_t) &= \mathbf{w}_t \cdot \mathbf{x}_t \\
\sum_{e \in \nu} x_e &= 0, -1 \leq x_e \leq 1 \text{ flow conservation and capacity constraint}
\end{aligned}
$$