Announcements

- HW3 is due Nov 27 (next Wednesday)
- Work on projects!! (presentation after three weeks)

21.1 Bandit problem in stochastic shortest path (continue on the last lecture)

21.1.1 FTRL in the bandit setting

In every round,

\[ x_t = \arg \min_{x \in K} \sum_{s=1}^{t-1} f_s \cdot x + \lambda R(x) \]  \hfill (21.1)

From the last lecture, we can use estimated loss function. Therefore,

\[ x_t = \arg \min_{x \in K} \sum_{s=1}^{t-1} \tilde{f}_s \cdot x + \lambda R(x) \]  \hfill (21.2)

The regret bound becomes,

\[
\text{Regret}_T \leq \sum_{t=1}^{T} \lambda D_R(x_t, x_t + 1) + \lambda R(x^*) \\
\leq \sum_{t=1}^{T} \frac{||\tilde{f}_t||^2}{\lambda} + \lambda R(x^*) \\
\leq \frac{T \cdot G}{\lambda} + \lambda D \\
\leq 2\sqrt{T \cdot G \cdot D} \hfill (21.3)
\]

Problem: As \( x_t \) approaches the boundary, \( ||\tilde{f}_t||_x \) grows very large. So above inequality breaks.

Solution: Use regularization with Self Concordance Function

21.1.2 Self Concordance Function

Classical Newton's method \( \text{Let our objective function be } g(x). \text{ we want to minimize} \)

\[
\min_{x \in D} g(x) \hfill (21.4)
\]
By adding self-concordance regularization term $R$, 

$$\min_{x \in D} g(x) + \lambda R(x)$$  \hspace{1cm} (21.5)

The Newton’s update rule becomes 

$$x_{t+1} \leftarrow x_t + (\nabla^2_x R)^{-1} \nabla \hat{g}(x_t)$$  \hspace{1cm} (21.6)

Therefore, $x_{t+1}$ is in the ellipsoid centered on $x_t$. 

$$x_{t+1} \in (\nabla^2_x R)\text{-ellipsoid}$$

**Back to Bandit Optimization**  
Previously, our update rule was 

$$x_t = \arg\min_{x \in K} \sum_{s=1}^{t-1} \tilde{f}_s \cdot x + \lambda R(x)$$  \hspace{1cm} (21.7)

We approximate $f_t$ as eigenpoles of $\nabla^2_x R$-ellipsoid. 

$$\tilde{f}_t \approx \lambda_j^{1/2} e_j$$  \hspace{1cm} (21.8)

where $\lambda_j$ and $e_j$ are eigenvalues and unit eigenvalues of $\nabla^2_x R$

![Figure 1: eigenpoles approximation](image)

Then, the new regret bound becomes 

$$\text{Regret}_T \leq \sum_{t=1}^{T} \lambda D_R(x_t, x_t + 1) + \lambda R(x^*)$$

$$\leq \sum_{t=1}^{T} \frac{\tilde{f}_t^T (\nabla^2_x R) \tilde{f}_t}{\lambda} + \lambda R(x^*)$$  \hspace{1cm} (21.9)

$$\leq \frac{n \sqrt{\sigma_i} o_i^{-1} \sqrt{\sigma_i}}{\lambda} + \lambda D \theta \log T$$

$$\leq 2 \sqrt{n \cdot G \cdot D \cdot \theta \cdot T \log T}$$
However, this results are "in expectation" only, and only work against "oblivious adversaries". The general problem is still hard.

21.2 Blackwell Approachability

In standard 2-player o-sum game, the game matrix $M$ satisfies

- $M \in [0, 1]^{n \times m}$,
- $M_{ij} \in \mathbb{R}$ is the payoff for $P_1$, when $P_1$ and $P_2$ play $i$ and $j$ respectively.

The minimax theorem is

$$\min_p \max_q p^T M q = \max_q \min_p p^T M q$$

or equivalently, (strong duality)

$$\forall p \exists q : p^T M q \geq c$$
$$\exists q \forall p : p^T M q \geq c$$

Generation Now, we want to generalize this to the case when $M_{ij} \in \mathbb{R}^d$.

Let $r(i, j)$ be the payoff vector for $P_1$,

$$r : \Delta_n \times \Delta_m \to \mathbb{R}^d$$

This is biaffine!

1. $r(\alpha p_1 + (1-\alpha)p_2, q) = \alpha r(p_1, q) + (1-\alpha)r(p_2, q)$
2. $r(p, \alpha q_1 + (1-\alpha)q_2) = \alpha r(p, q_1) + (1-\alpha)r(p, q_2)$

In this generalized version, we can define the minimax theorem by

$$\forall p \exists q : r(p, q) \in S$$
$$\exists q \forall p : r(p, q) \in S$$

$S$ is a certain convex set applying some constraints. In general, this condition is not satisfied.

Example: A bad case

$$r(p, q) = (p, q)$$
$$S = \{(x, y) : x = y\}$$

$\forall p$, there exists $q = p$ such that $r(p, q) \in S$.
However, there is no $q$ satisfying $\forall p : r(p, q) \in S$
21.2.1 Blackwell Approachability Theorem

If \( r, S, p, q \) satisfies
\[
\forall p, \exists q : r(p, q) \in S
\]  
(21.13)

Then, \( \exists \) an adaptive strategy
\[
q_t \leftarrow f(p_1, p_2, ..., p_t)
\]  
(21.14)
such that
\[
\frac{1}{T} \sum r(p_t, q_t) \rightarrow S
\]  
(21.15)
or equivalently,
\[
dist\left(\frac{1}{T} \sum r(p_t, q_t), S\right) \rightarrow 0
\]  
(21.16)

Example In above bad case example, one possible strategy for \( q \) is to choose previous \( p \).

\[
q_t \leftarrow p_{t-1}
\]  
(21.17)

This strategy satisfies Blackwell Approachability theorem,
\[
dist\left(\frac{1}{T} \sum_{t=1}^{T} p_t, \frac{1}{T} \sum_{t=0}^{T-1} p_t\right) \rightarrow S
\]  
(21.18)

where \( p_0 = q_1 \). (initial choice of \( q \))

21.2.2 Halfspace condition

\( \forall \) halfspaces \( H \supset S, \exists q \forall p \)
\[
r(p, q) \in H
\]  
(21.19)

Lemma: The followings are equivalent

1. The halfspace condition
2. \( \forall \) halfspaces \( H \supset S, \forall p, \exists q: r(p, q) \in H \)
3. \( \forall p, \exists q: r(p, q) \in S \)
Proof:

1. 1 and 2 are equivalent
   Project \( r(p, q) \) into the normal of \( H \), and apply minimax theorem.

2. \( \exists H, \exists p_{bad}, \forall q \)
   \[ r(p_{bad}, q) \notin H \]
   which implies \( r(p_{bad}, q) \notin S \) (contradiction)

3. \( \exists p_{bad}, \forall q: r(p_{bad}, q) \notin S \)
   \( \Rightarrow \exists \) hyperplane separating \( S \) and \( \{ r(p_{bad}, q) : q \in \Delta_m \} \), but this hyperplane violates 1.
   (contradiction)

![Figure 3: Separating Hyperplane](image-url)