Online Geometric Optimization in the Bandit Setting Against an Adaptive Adversary

A presentation by Evan Ettinger

March 2, 2006
Outline

Minimax and Vector Payoffs

Follow the Perturbed Leader (FPL)

Bandit Setting

Conclusions
Two-person zero-sum games

Two players - \( \{I, II\} \) and a payoff matrix \( M \).

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>-4</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>-8</td>
</tr>
</tbody>
</table>
Two-person zero-sum games

Two players - \{I, II\} and a payoff matrix $M$.

- Player I chooses a row $i \sim p = (p_1, p_2, p_3)$
Two-person zero-sum games

Two players - \{I, II\} and a payoff matrix $M$.

- Player I chooses a row $i \sim p = (p_1, p_2, p_3)$
- Player II chooses a column $j \sim q = (q_A, q_B, q_C)$. 

\[ M = \begin{pmatrix}
1 & 5 & -4 & 10 \\
2 & -2 & 0 & 2 \\
3 & 4 & 5 & -8
\end{pmatrix} \]
Two-person zero-sum games

Two players - \( \{I, II\} \) and a payoff matrix \( M \).

- Player I chooses a row \( i \sim p = (p_1, p_2, p_3) \)
- Player II chooses a column \( j \sim q = (q_A, q_B, q_C) \).
- I receives gain \( M(i, j) \) and II incurs loss \( -M(i, j) \).

\[
M : \begin{array}{ccc}
1 & 5 & -4 & 10 \\
2 & -2 & 0 & 2 \\
3 & 4 & 5 & -8 \\
\end{array}
\]
von Neumann Minimax Theorem

**Theorem:** For any $M$ with real elements, $\exists v \in \mathbb{R}$ and distributions $p, q$ such that for all $i,j$:

$$\sum_i p_i M(i, j) \geq v \geq \sum_j q_j M(i, j)$$
von Neumann Minimax Theorem

**Theorem:** For any $M$ with real elements, $\exists v \in \mathbb{R}$ and distributions $p, q$ such that for all $i,j$:

$$\sum_i p_i M(i,j) \geq v \geq \sum_j q_j M(i,j)$$

**Alternative Interpretation:**

- After a long series of plays, with probability approaching 1:
  - Player I’s average per play gain exceeds $v - \varepsilon$.
  - Player II’s average per play loss is no more than $v + \varepsilon$. 

Online Geometric Optimization in the Bandit Setting Against an Adaptive Adversary
Blackwell’s Generalization of the Game

- Let $X \subseteq \mathbb{R}^n$ be a closed bounded set.
- Each element $M(i, j)$ is a probability distribution over $X$. 
Blackwell’s Generalization of the Game

- Let $X \subseteq \mathbb{R}^n$ be a closed bounded set.
- Each element $M(i, j)$ is a probability distribution over $X$.
- At iteration $i$ of the game:
  - Player I chooses a row $i \sim f_i(x_1, \ldots, x_{i-1})$
  - Player II chooses column $j \sim q_i(x_1, \ldots, x_{i-1})$
  - $x_i \in X$ is generated according to $M(i, j)$. 

How can we generalize the minimax theorem to this more general setting?
Blackwell’s Generalization of the Game

- Let $X \subseteq \mathbb{R}^n$ be a closed bounded set.
- Each element $M(i, j)$ is a probability distribution over $X$.
- At iteration $i$ of the game:
  - Player I chooses a row $i \sim f_i(x_1, \ldots, x_{i-1})$
  - Player II chooses column $j \sim q_i(x_1, \ldots, x_{i-1})$
  - $x_i \in X$ is generated according to $M(i, j)$.
- strategy - $f = \{f_i\}$ (Player I) and $q = \{q_i\}$ (Player II).
Blackwell’s Generalization of the Game

- Let $X \subseteq \mathbb{R}^n$ be a closed bounded set.
- Each element $M(i, j)$ is a probability distribution over $X$.
- At iteration $i$ of the game:
  - Player I chooses a row $i \sim f_i(x_1, \ldots, x_{i-1})$
  - Player II chooses column $j \sim q_i(x_1, \ldots, x_{i-1})$
  - $x_i \in X$ is generated according to $M(i, j)$.
- strategy - $f = \{f_i\}$ (Player I) and $q = \{q_i\}$ (Player II).

\[
M: \begin{array}{ccc}
A & B & C \\
1 & p_{11} & p_{12} & p_{13} \\
2 & p_{21} & p_{22} & p_{23} \\
3 & p_{31} & p_{32} & p_{33} \\
\end{array} \Rightarrow x_k \in X, x_k \sim p_{ij}
\]
Blackwell’s Generalization of the Game

- Let $X \subseteq \mathbb{R}^n$ be a closed bounded set.
- Each element $M(i, j)$ is a probability distribution over $X$.
- At iteration $i$ of the game:
  - Player I chooses a row $i \sim f_i(x_1, ..., x_{i-1})$
  - Player II chooses column $j \sim q_i(x_1, ..., x_{i-1})$
  - $x_i \in X$ is generated according to $M(i, j)$.
- Strategy: $f = \{ f_i \}$ (Player I) and $q = \{ q_i \}$ (Player II).

$$M: \begin{array}{ccc} A & B & C \\ 1 & p_{11} & p_{12} & p_{13} \\ 2 & p_{21} & p_{22} & p_{23} \\ 3 & p_{31} & p_{32} & p_{33} \\ \end{array} \Rightarrow x_k \in X, x_k \sim p_{ij}$$

- How can we generalize the minimax theorem to this more general setting?
Approachability

- Consider $\bar{x}_n = \frac{1}{n} \sum_k x_k$, set $S \subseteq \mathbb{R}^n$, and $\delta_n$ the distance from $\bar{x}_n$ to $S$. 
Consider \( \bar{x}_n = \frac{1}{n} \sum_k x_k \), set \( S \subseteq \mathbb{R}^n \), and \( \delta_n \) the distance from \( \bar{x}_n \) to \( S \).

A set \( S \) is approachable with strategy \( f^* \) in \( M \) if \( \forall \varepsilon > 0 \), \( \exists N_0 \) s.t. for every \( g \)

\[
P(\delta_n \geq \varepsilon \text{ for some } n \geq N_0) < \varepsilon
\]
Approachability

Consider $\bar{x}_n = \frac{1}{n} \sum_k x_k$, set $S \subseteq \mathbb{R}^n$, and $\delta_n$ the distance from $\bar{x}_n$ to $S$.

- A set $S$ is approachable with strategy $f^*$ in $M$ if $\forall \varepsilon > 0$, $\exists N_0$ s.t. for every $g$

  $$P(\delta_n \geq \varepsilon \text{ for some } n \geq N_0) < \varepsilon$$

- A set $S$ is excludable with strategy $g^*$ in $M$ if $\exists d > 0$, $\forall \varepsilon > 0$, $\exists N_0$ s.t. for every $f$

  $$P(\delta_n \geq d \forall n \geq N_0) > 1 - \varepsilon$$
Sufficient Condition for Approachability

- Let $\bar{M}$ be the matrix with entries $\bar{m}(i, j)$ - the mean values of $p_{ij}$.
- $H(p)$ is convex hull of means according to strategy $p$ – $\sum_i p_i \bar{m}(i, j)$
Sufficient Condition for Approachability

Let $\tilde{M}$ be the matrix with entries $\tilde{m}(i,j)$ - the mean values of $p_{ij}$.

$H(p)$ is convex hull of means according to strategy $p$ – $\sum_i p_i \tilde{m}(i,j)$

**Theorem**: If for every $x \notin S$ there is a $p$ such that the hyperplane through $y$, the closest point in $S$ to $x$, perpendicular to line segment $xy$ separates $x$ from $H(p)$, then $S$ is approachable with $f_n$

$$f_n = \begin{cases} p & \bar{x}_n \notin S \\ \text{arbitrary} & \bar{x}_n \in S \end{cases}$$
Sufficient Condition for Approachability

Let $\bar{M}$ be the matrix with entries $\bar{m}(i,j)$ - the mean values of $p_{ij}$.

$H(p)$ is convex hull of means according to strategy $p$ – $\sum_j p_j \bar{m}(i,j)$

**Theorem:** If for every $x \notin S$ there is a $p$ such that the hyperplane through $y$, the closest point in $S$ to $x$, perpendicular to line segment $xy$ separates $x$ from $H(p)$, then $S$ is approachable with $f_n$

$$f_n = \begin{cases} p & \bar{x}_n \notin S \\ \text{arbitrary} & \bar{x}_n \in S \end{cases}$$
Properties of Certain S

- Every convex S is either approachable or excludable.
Properties of Certain $S$

- Every convex $S$ is either approachable or excludable.
  - If $S$ intersects every convex hull then $S$ is approachable.
Properties of Certain S

- Every convex S is either approachable or excludable.
  - If S intersects every convex hull then S is approachable.
  - If S fails to intersect some convex hull then S is excludable.
Properties of Certain $S$

- Every convex $S$ is either approachable or excludable.
  - If $S$ intersects every convex hull then $S$ is approachable.
  - If $S$ fails to intersect some convex hull then $S$ is excludable.

- For 1-dimensional case, every $S$ is either approachable or excludable.
- False for 2+-dimensional case.
Properties of Certain $S$

- Every convex $S$ is either approachable or excludable.
  - If $S$ intersects every convex hull then $S$ is approachable.
  - If $S$ fails to intersect some convex hull then $S$ is excludable.
- For 1-dimensional case, every $S$ is either approachable or excludable.
- False for 2+-dimensional case.
- Keep cost vectors close to some “good set” (or away from “bad”).
FPL Setting Reminder

- Make decision $d_t \in \mathcal{D} \subseteq \mathbb{R}^n$
FPL Setting Reminder

- Make decision $d_t \in D \subseteq \mathbb{R}^n$
- Cost vector $c_t \in \mathbb{R}^n$ is observed.
FPL Setting Reminder

- Make decision $d_t \in D \subseteq \mathbb{R}^n$
- Cost vector $c_t \in \mathbb{R}^n$ is observed.
- Cost $c_t \cdot d_t$ is incurred.
FPL Setting Reminder

- Make decision \( d_t \in \mathcal{D} \subseteq \mathbb{R}^n \)
- Cost vector \( c_t \in \mathbb{R}^n \) is observed.
- Cost \( c_t \cdot d_t \) is incurred.
- \( M \) is a function that computes the best single decision in hindsight:

\[
M(c) = \arg \min_{d \in S} \sum_t c \cdot d_t
\]
FPL Setting Reminder

- Make decision \( d_t \in D \subseteq \mathbb{R}^n \)
- Cost vector \( c_t \in \mathbb{R}^n \) is observed.
- Cost \( c_t \cdot d_t \) is incurred.
- \( M \) is a function that computes the best single decision in hindsight:
  \[
  M(c) = \arg \min_{d \in S} \sum_t c \cdot d_t
  \]
- Minimize regret: \( \sum_t d_t \cdot c_t - M(c) \cdot c \)
FPL Setting Reminder

- Make decision $d_t \in \mathcal{D} \subseteq \mathbb{R}^n$
- Cost vector $c_t \in \mathbb{R}^n$ is observed.
- Cost $c_t \cdot d_t$ is incurred.
- $M$ is a function that computes the best single decision in hindsight:

$$M(c) = \arg \min_{d \in S} \sum_t c \cdot d_t$$

- Minimize regret: $\sum_t d_t \cdot c_t - M(c) \cdot c$
- Generalizes expert advice problem - $S$ not explicit, only need existence of $M$. 
FPL Setting Reminder

- Make decision $d_t \in D \subseteq \mathbb{R}^n$
- Cost vector $c_t \in \mathbb{R}^n$ is observed.
- Cost $c_t \cdot d_t$ is incurred.
- $M$ is a function that computes the best single decision in hindsight:
  $$M(c) = \arg \min_{d \in S} \sum_t c \cdot d_t$$

- Minimize regret: $\sum_t d_t \cdot c_t − M(c) \cdot c$
- Generalizes expert advice problem - $S$ not explicit, only need existence of $M$.
- Example: Online shortest path
  - choose edge set $d_t \in D = \{0, 1\}^{|E|}$.
  - Can have $K = \mathcal{O}(2^{|V|})$ paths. Using “flat” bandit algorithm has regret $\mathcal{O}(\sqrt{TK \log K})$
FPL Algorithm

FPL(\(\varepsilon\)):
1. Choose \(p_t \sim [0, 1/\varepsilon]^n\) uniformly.
2. Use \(M(c_1 + \ldots + c_{t-1} + p_t)\)
FPL Algorithm

FPL(ε):
1. Choose $p_t \sim [0, 1/\varepsilon]^n$ uniformly.
2. Use $M(c_1 + \ldots + c_{t-1} + p_t)$

Define the following:

$$D \geq \|d - d'\|_1 \quad \forall d, d' \in \mathcal{D}$$
$$R \geq |d \cdot c| \quad \forall d \in \mathcal{D}, c \in S$$
$$A \geq \|c\|_1 \quad \forall c \in S$$

Regret: $E[L_{FPL}(T)] \leq L_{OPT}(T) + 2\sqrt{DRA}T$
A new setting...

- Make decision $d_t \in \mathcal{D} \subseteq \mathbb{R}^n$
A new setting...

- Make decision $d_t \in \mathcal{D} \subseteq \mathbb{R}^n$
- Cost vector $c_t \in \mathbb{R}^n$ is NOT observed.
A new setting...

- Make decision $d_t \in \mathcal{D} \subseteq \mathbb{R}^n$
- Cost vector $c_t \in \mathbb{R}^n$ is NOT observed.
- Cost $c_t \cdot d_t$ is incurred and observed.
A new setting...

- Make decision $d_t \in D \subseteq \mathbb{R}^n$
- Cost vector $c_t \in \mathbb{R}^n$ is NOT observed.
- Cost $c_t \cdot d_t$ is incurred and observed.
- Example: Online shortest path
  - At each iteration of the game we only observe how long the path took.
  - We do NOT see individual edge costs adversary picked.
Algorithm Idea

- Remember FPL uses the sum of previous cost vectors.
Algorithm Idea

- Remember FPL uses the sum of previous cost vectors.
- Goal: estimate $c^{1:T} = \sum_{t} c_t$ closely with $\hat{c}^{1:T}$
Algorithm Idea

- Remember FPL uses the sum of previous cost vectors.
- Goal: estimate $c^{1:T} = \sum_t c_t$ closely with $\hat{c}^{1:T}$
- Method: *Explore* by making “basis” decisions of the space $\mathcal{D}$ and *Exploit* using the FPL algorithm.
Algorithm Idea

- Remember FPL uses the sum of previous cost vectors.
- Goal: estimate \( c^{1:T} = \sum_t c_t \) closely with \( \hat{c}^{1:T} \)
- Method: Explore by making "basis" decisions of the space \( \mathcal{D} \) and Exploit using the FPL algorithm.
- Details: Choosing basis \( \{b_i\} \) "nicely" and choosing exploration rate, \( \gamma \), matters...
Bandit FPL (BFPL) Algorithm

1. Fix basis $B = \{b_1, ..., b_n\}$ of $\mathcal{D}$. 
Bandit FPL (BFPL) Algorithm

1. Fix basis $B = \{b_1, \ldots, b_n\}$ of $\mathcal{D}$.
2. For each $t$, with probability $\gamma$ do (Exploit):
   2.1 Select $d_t = \text{FPL}(\hat{c}_1: t-1)$.
   2.2 Observe cost $c_t \cdot x_t$.
   2.3 Set $\hat{c}_t = 0$.
3. Else do (Explore):
   3.1 Choose $x_t = b_j$ uniformly at random.
   3.2 Observe cost $c_t \cdot x_t$.
   3.3 Define $\hat{L}_t[j] = (n/\gamma) z_t$, and $\hat{L}_t[i] = 0$ for all $i \neq j$.
   3.4 $\hat{c}_t = (B^T)_{t-1} \hat{L}_t$. 
4. $\hat{c}_1:t = \hat{c}_1:t-1 + \hat{c}_t$. 

Online Geometric Optimization in the Bandit Setting Against an Adaptive Adversary
Bandit FPL (BFPL) Algorithm

1. Fix basis $B = \{b_1, \ldots, b_n\}$ of $\mathcal{D}$.

2. For each $t$, with probability $\gamma$ do (Exploit):
   2.1 select $d^t = FPL(\hat{c}^{1:t-1})$.
   2.2 Observe cost $c_t \cdot x_t$
   2.3 Set $\hat{c}^t = 0$.
Bandit FPL (BFPL) Algorithm

1. Fix basis $B = \{b_1, ..., b_n\}$ of $\mathcal{D}$.

2. For each $t$, with probability $\gamma$ do (Exploit):
   2.1 select $d^t = \text{FPL}(\hat{c}^{1:t-1})$.
   2.2 Observe cost $c_t \cdot x_t$
   2.3 Set $\hat{c}^t = 0$.

3. Else do (Explore):
   3.1 Choose $x^t = b_j$ uniformly at random.
   3.2 Observe cost $c^t \cdot x^t$
   3.3 Define $\hat{L}_j^t = (n/\gamma)z^t$, and $\hat{L}_i^t = 0 \forall i \neq j$
   3.4 $\hat{c}^t = (B^T)^{-1} \hat{L}^t$
Bandit FPL (BFPL) Algorithm

1. Fix basis $B = \{b_1, \ldots, b_n\}$ of $D$.

2. For each $t$, with probability $\gamma$ do (Exploit):
   2.1 select $d^t = FPL(\hat{c}_1^{1:t-1})$.
   2.2 Observe cost $c_t \cdot x_t$
   2.3 Set $\hat{c}^t = 0$.

3. Else do (Explore):
   3.1 Choose $x^t = b_j$ uniformly at random.
   3.2 Observe cost $c^t \cdot x^t$
   3.3 Define $\hat{L}^t_j = (n/\gamma)z_t$, and $\hat{L}^t_i = 0 \forall i \neq j$
   3.4 $\hat{c}^t = (B^T)^{-1}\hat{L}^t$

4. $\hat{c}^{1:t} = \hat{c}^{1:t-1} + \hat{c}^t$
Idea of Regret Analysis

- We plan on showing the following:
  1. \( E[L_{FPL}(T)] \leq E[\hat{c}^{1:T} \cdot M(\hat{c}^{1:T})] + \text{(terms)} \)
Idea of Regret Analysis

- We plan on showing the following:
  1. \( E[L_{FPL}(T)] \leq E[\hat{c}^{1:T} \cdot M(\hat{c}^{1:T})] + \) (terms)
  2. \( E[L_{BFPL}(T)] \leq E[L_{FPL}(T)] + \) (terms)
Idea of Regret Analysis

- We plan on showing the following:
  1. $E[L_{FPL}(T)] \leq E[\hat{c}^{1:T} \cdot M(\hat{c}^{1:T})] + \text{(terms)}$
  2. $E[L_{BFPL}(T)] \leq E[L_{FPL}(T)] + \text{(terms)}$
  3. $E[\hat{c}^{1:t} \cdot M(\hat{c}^{1:T})] \leq E[L_{OPT}(T)] + \text{(terms)}$
Idea of Regret Analysis

We plan on showing the following:

1. $E[L_{FPL}(T)] \leq E[\hat{c}^{1:T} \cdot M(\hat{c}^{1:T})] + (\text{terms})$
2. $E[L_{BFPL}(T)] \leq E[L_{FPL}(T)] + (\text{terms})$
3. $E[\hat{c}^{1:t} \cdot M(\hat{c}^{1:T})] \leq E[L_{OPT}(T)] + (\text{terms})$

The first bound follows from the analysis of FPL (previous talk).
Showing one of these parts...

**Theorem:** BFPL(\(\gamma\)) for \(T\) timesteps yields:

\[
E[L_{BFPL}] \leq (1 - \gamma)E[L_{FPL}] + \gamma RT
\]
Theorem: BFPL(γ) for T timesteps yields:

\[ E[L_{BFPL}] \leq (1 - \gamma)E[L_{FPL}] + \gamma RT \]

- Let \( G_t^{-1} = [b_1, d_1, ..., b_{t-1}, d_{t-1}] \) the full history of the algorithms decisions.
Theorem: BFPL(γ) for T timesteps yields:

\[ \mathbb{E}[L_{BFPL}] \leq (1 - \gamma)\mathbb{E}[L_{FPL}] + \gamma RT \]

- Let \( G^{t-1} = [b_1, d_1, \ldots, b_{t-1}, d_{t-1}] \) the full history of the algorithms decisions.

\[ \mathbb{E}[L_{BFPL}^t | G^{t-1}] = (1 - \gamma)(c^t \cdot \bar{x}^t) + \gamma \sum_{i=1}^{n} \frac{1}{n} (c^t \cdot b_i) \]
Theorem: BFPL($\gamma$) for $T$ timesteps yields:

$$E[L_{BFPL}] \leq (1 - \gamma)E[L_{FPL}] + \gamma RT$$

Let $G^{t-1} = [b_1, d_1, \ldots, b_{t-1}, d_{t-1}]$ the full history of the algorithms decisions.

$$E[L_{BFPL}^t \mid G^{t-1}] = (1 - \gamma)(c^t \cdot \bar{x}^t) + \gamma \sum_{i=1}^{n} \frac{1}{n}(c^t \cdot b_i)$$

$$\leq (1 - \gamma)E[L_{FPL}^t \mid G^{t-1}] + \gamma R$$
Bound proof (cont.)...

From last slide...

\[ E[L_{BFPL}^t | G^{t-1}] \leq (1 - \gamma) E[L_{FPL}^t | G^{t-1}] + \gamma R \]
Bound proof (cont.)...

From last slide...

$$E[L^t_{BFPL} | G^{t-1}] \leq (1 - \gamma)E[L^t_{FPL} | G^{t-1}] + \gamma R$$

Now...
Bound proof (cont.)...

From last slide...

\[ E[L_{BFPL}^t | G^{t-1}] \leq (1 - \gamma) E[L_{FPL}^t | G^{t-1}] + \gamma R \]

Now...

\[ E[L_{BFPL}^t] = E[E[L_{BFPL}^t | G^{t-1}] \]

Online Geometric Optimization in the Bandit Setting Against an Adaptive Adversary
Bound proof (cont.)...

From last slide...

\[
E[L_{BFPL}^t | G^{t-1}] \leq (1 - \gamma) E[L_{FPL}^t | G^{t-1}] + \gamma R
\]

Now...

\[
E[L_{BFPL}^t] = E[E[L_{BFPL}^t | G^{t-1}]] \\
\leq E[(1 - \gamma) E[L_{FPL}^t | G^{t-1}] + \gamma R]
\]
Bound proof (cont.)...

From last slide...

\[ E[L^t_{BFPL} \mid G^{t-1}] \leq (1 - \gamma)E[L^t_{FPL} \mid G^{t-1}] + \gamma R \]

Now...

\[ E[L^t_{BFPL}] = E[E[L^t_{BFPL} \mid G^{t-1}]] \]
\[ \leq E[(1 - \gamma)E[L^t_{FPL} \mid G^{t-1}] + \gamma R] \]
\[ = (1 - \gamma)E[E[L^t_{FPL} \mid G^{t-1}]] + \gamma R \]
Bound proof (cont.)...

From last slide...

\[ E[L_{BFPL}^t | G^{t-1}] \leq (1 - \gamma) E[L_{FPL}^t | G^{t-1}] + \gamma R \]

Now...

\[ E[L_{BFPL}^t] = E[E[L_{BFPL}^t | G^{t-1}]] \]
\[ \leq E[(1 - \gamma) E[L_{FPL}^t | G^{t-1}] + \gamma R] \]
\[ = (1 - \gamma) E[E[L_{FPL}^t | G^{t-1}]] + \gamma R \]
\[ = (1 - \gamma) E[L_{FPL}^t] + \gamma R \]
Bound proof (cont.)...

From last slide...

$$E[L_{BFPL}^t | G^{t-1}] \leq (1 - \gamma) E[L_{FPL}^t | G^{t-1}] + \gamma R$$

Now...

$$E[L_{BFPL}^t] = E[E[L_{BFPL}^t | G^{t-1}]]$$

$$\leq E[(1 - \gamma) E[L_{FPL}^t | G^{t-1}] + \gamma R]$$

$$= (1 - \gamma) E[E[L_{FPL}^t | G^{t-1}]] + \gamma R$$

$$= (1 - \gamma) E[L_{FPL}^t] + \gamma R$$

Summing over all $t$ from 1 to $T$ gives us:

$$E[L_{BFPL}] \leq (1 - \gamma) E[L_{FPL}] + \gamma RT$$
Regret of BFPL

\[ D \geq \|d - d'\|_1 \quad \forall d, d' \in \mathcal{D} \]

\[ R \geq |d \cdot c| \quad \forall d \in \mathcal{D}, \ c \in \mathcal{S} \]
Regret of BFPL

\[ D \geq \| d - d' \|_1 \quad \forall d, d' \in \mathcal{D} \]
\[ R \geq |d \cdot c| \quad \forall d \in \mathcal{D}, c \in \mathcal{S} \]

For all \( \delta \in (0, 1) \):
\[
E[L_{BFPL}] \leq E[L_{OPT}] + \mathcal{O}(D^{1/2}nR\sqrt{2\ln(2n/\delta)}\sqrt{T} + \delta RT + \frac{\epsilon}{\gamma^2}n^3R^2 T + \frac{n}{\epsilon} + \gamma RT)
\]
Regret of BFPL

\[ D \geq \|d - d'\|_1 \quad \forall d, d' \in D \]
\[ R \geq |d \cdot c| \quad \forall d \in D, \ c \in S \]

For all \( \delta \in (0, 1) \):

\[
E[L_{BFPL}] \leq E[L_{OPT}] + \mathcal{O}(\frac{nR}{\gamma} \sqrt{2 \ln(2n/\delta)} \sqrt{T} + \delta RT + \frac{\epsilon}{\gamma^2} n^3 R^2 T + \frac{n}{\epsilon} + \gamma RT)
\]

\[ \text{Ignoring dependence on } n, R \text{ and } D, \text{ and if we set} \]
\[ \gamma = \delta = T^{-1/4} \text{ and } \epsilon = T^{-3/4}: \]
\[ E[L_{BFPL}] \leq E[L_{OPT}] + \mathcal{O}(T^{3/4} \sqrt{\ln T}) \]
In conclusion...

- For “flat” bandits we know regret bounds of \( O(\sqrt{T}) \) achievable.
In conclusion...

- For “flat” bandits we know regret bounds of $O(\sqrt{T})$ achievable.
- Awerbuch and Kleinberg give a $O(T^{2/3})$ against an *oblivious* adversary.
In conclusion...

- For “flat” bandits we know regret bounds of $O(\sqrt{T})$ achievable.
- Awerbuch and Kleinberg give a $O(T^{2/3})$ against an *oblivious adversary*
  - *oblivious adversary* - must commit to cost vector sequence in advance
  - *adaptive adversary* - can determine next cost vector based on algorithm’s previous play

---

Online Geometric Optimization in the Bandit Setting Against an Adaptive Adversary
In conclusion...

- For “flat” bandits we know regret bounds of $O(\sqrt{T})$ achievable.
- Awerbuch and Kleinberg give a $O(T^{2/3})$ against an **oblivious adversary**
  - **oblivious adversary** - must commit to cost vector sequence in advance
  - **adaptive adversary** - can determine next cost vector based on algorithm’s previous play
- This work achieves $O(T^{3/4} \sqrt{\ln T})$. 

Note: this algorithm throws away $(1-\gamma)$ of the information received. Can we use this?
Can we tighten these bounds down to $O(\sqrt{T})$?
In conclusion...

- For “flat” bandits we know regret bounds of $\mathcal{O}(\sqrt{T})$ achievable.
- Awerbuch and Kleinberg give a $\mathcal{O}(T^{2/3})$ against an *oblivious adversary*
  - *oblivious adversary* - must commit to cost vector sequence in advance
  - *adaptive adversary* - can determine next cost vector based on algorithm’s previous play
- This work achieves $\mathcal{O}(T^{3/4} \sqrt{\ln T})$.
- Note: this algorithm throws away $(1 - \gamma)$ of the information received. Can we use this?
In conclusion...

- For “flat” bandits we know regret bounds of $O(\sqrt{T})$ achievable.
- Awerbuch and Kleinberg give a $O(T^{2/3})$ against an *oblivious adversary*:
  - *oblivious adversary* - must commit to cost vector sequence in advance
  - *adaptive adversary* - can determine next cost vector based on algorithm’s previous play
- This work achieves $O(T^{3/4}\sqrt{\ln T})$.
- Note: this algorithm throws away $(1 - \gamma)$ of the information received. Can we use this?
- Can we tighten these bounds down to $O(\sqrt{T})$?
Citations

