Minimax Regret of Finite Partial-Monitoring Games in Stochastic Environments

Gábor Bartók, Dávid Pál, Csaba Szepesvári

COLT2011, Budapest
Finite Stochastic Partial-Monitoring Games

learner

environment
Finite Stochastic Partial-Monitoring Games

learner

referee

environment

action $I_t$

outcome $J_t$
Finite Stochastic Partial-Monitoring Games

learner environment

referee

action \( I_t \) outcome \( J_t \)

loss \( \ell_t = L(I_t, J_t) \)

feedback \( h_t = H(I_t, J_t) \)

\( L, H \in \mathbb{R}^{N \times M} \) publicly known

L, H ∈ \( \mathbb{R}^{N \times M} \) publicly known

environment
Finite Stochastic Partial-Monitoring Games

learner

feedback $h_t = H(I_t, J_t)$

referee

loss $\ell_t = L(I_t, J_t)$

environment

action $I_t$

outcome $J_t$

Finitely many actions, outcomes; Stochastic environment

Bartók, Pál, Szepesvári (UofA)

Partial Monitoring

COLT2011, Budapest
Finite Stochastic Partial-Monitoring Games

learner

environment

referee

action \( I_t \)

outcome \( J_t \)

loss \( \ell_t = L(I_t, J_t) \)

feedback \( h_t = H(I_t, J_t) \)

Finitely many actions, outcomes; **Stochastic environment**
Examples
Examples

Bandits

\[ L = H \]
Examples

Bandits

\[ L = H \]

Full info

\[ H = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \]
Examples

Bandits

\[ L = H \]

Full info

\[ H = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \]

Dynamic pricing

\[ L(i, j) = c \mathbb{1}_{i > j} + (j - i) \mathbb{1}_{i \leq j} \]
\[ H(i, j) = \mathbb{1}_{i \leq j} \]
### Examples

#### Bandits

\[
L = H
\]

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

#### Full info

\[
H = \begin{pmatrix}
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
\end{pmatrix}
\]

#### Dynamic pricing

\[
L = \begin{pmatrix}
0 & 1 & \cdots & N - 1 \\
c & 0 & \cdots & N - 2 \\
\vdots & \vdots & \ddots & \vdots \\
c & \cdots & c & 0 \\
\end{pmatrix}
\]

\[
H = \begin{pmatrix}
1 & \cdots & \cdots & 1 \\
0 & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1 \\
\end{pmatrix}
\]
Expected regret $R_T(A) = \mathbb{E}[\sum_{t=1}^{T} L(I_t, J_t)] - \min_i \mathbb{E}[\sum_{t=1}^{T} L(i, J_t)]$
Performance measure

Expected regret $R_T(A) = \mathbb{E}[\sum_{t=1}^{T} L(I_t, J_t)] - \min_i \mathbb{E}[\sum_{t=1}^{T} L(i, J_t)]$

The problem: $(L, H)$ given, determine the minimax expected regret $\hat{R}_T$
### Performance measure

Expected regret \( R_T(A) = \mathbb{E}[\sum_{t=1}^{T} L(I_t, J_t)] - \min_i \mathbb{E}[\sum_{t=1}^{T} L(i, J_t)] \)

The problem: \((L, H)\) given, determine the minimax expected regret \( \hat{R}_T \)

A typical result: \( \hat{R}_T = O(T^\alpha) \) for some \( 0 \leq \alpha \leq 1 \)
Performance measure

Expected regret $R_T(A) = \mathbb{E}[\sum_{t=1}^{T} L(I_t, J_t)] - \min_{i} \mathbb{E}[\sum_{t=1}^{T} L(i, J_t)]$

The problem: $(L, H)$ given, determine the minimax expected regret $\hat{R}_T$

A typical result: $\hat{R}_T = O(T^\alpha)$ for some $0 \leq \alpha \leq 1$
Goal

Performance measure

Expected regret $R_T(A) = \mathbb{E}[\sum_{t=1}^{T} L(I_t, J_t)] - \min_i \mathbb{E}[\sum_{t=1}^{T} L(i, J_t)]$

The problem: $(L, H)$ given, determine the minimax expected regret $\hat{R}_T$

A typical result: $\hat{R}_T = O(T^\alpha)$ for some $0 \leq \alpha \leq 1$
Previous work

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>$T^{1/2}$</th>
<th>$?$</th>
<th>$T^{2/3}$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>trivial</td>
<td>trivial</td>
<td>easy</td>
<td>full-info</td>
<td>hard</td>
<td>hopeless</td>
</tr>
<tr>
<td>bandits</td>
<td>dynamic pricing</td>
<td>l.e.p.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Previous work

<table>
<thead>
<tr>
<th>0</th>
<th>$T^{1/2}$</th>
<th>?</th>
<th>$T^{2/3}$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>trivial</td>
<td>easy</td>
<td>full-info</td>
<td>hard</td>
<td>hopeless</td>
</tr>
<tr>
<td></td>
<td>bandits</td>
<td>dynamic pricing</td>
<td>l.e.p.</td>
<td></td>
</tr>
</tbody>
</table>

**Full-info, bandits:** [Littlestone and Warmuth, 1994, Auer et al., 2002]
Previous work

<table>
<thead>
<tr>
<th>0</th>
<th>$T^{1/2}$</th>
<th>$\omega$</th>
<th>$T^{2/3}$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>trivial</td>
<td>easy</td>
<td>full-info</td>
<td>hard</td>
<td>hopeless</td>
</tr>
<tr>
<td>bandits</td>
<td>dynamic pricing</td>
<td>l.e.p.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Full-info, bandits: [Littlestone and Warmuth, 1994, Auer et al., 2002]

Algorithm (non-hopeless): $\tilde{O}(T^{3/4})$ [Piccolboni and Schindelhauer, 2001]
### Full-info, bandits

[Littlestone and Warmuth, 1994, Auer et al., 2002]

Algorithm (non-hopeless): $\tilde{O}(T^{3/4})$ [Piccolboni and Schindelhauer, 2001]

Reducing regret to $O(T^{2/3})$ [Cesa-Bianchi et al., 2006]
Full-info, bandits: [Littlestone and Warmuth, 1994, Auer et al., 2002]
Algorithm (non-hopeless): $\tilde{O}(T^{3/4})$ [Piccolboni and Schindelhauer, 2001]
Reducing regret to $O(T^{2/3})$ [Cesa-Bianchi et al., 2006] + l.e.p. lower
Previous work

Full-info, bandits: [Littlestone and Warmuth, 1994, Auer et al., 2002]
Algorithm (non-hopeless): $\tilde{O}(T^{3/4})$ [Piccolboni and Schindelhauer, 2001]
Reducing regret to $O(T^{2/3})$ [Cesa-Bianchi et al., 2006] + l.e.p. lower
non-trivial $\rightarrow \Omega(\sqrt{T})$ [Antos et al., 2011]
Previous work

<table>
<thead>
<tr>
<th>0</th>
<th>$T^{1/2}$</th>
<th>?</th>
<th>$T^{2/3}$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>trivial</td>
<td>easy</td>
<td>full-info</td>
<td>hard</td>
<td>hopeless</td>
</tr>
<tr>
<td></td>
<td>full-info</td>
<td>bandits</td>
<td>l.e.p.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dynamic pricing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Full-info, bandits:** [Littlestone and Warmuth, 1994, Auer et al., 2002]

**Algorithm (non-hopeless):** $\tilde{O}(T^{3/4})$ [Piccolboni and Schindelhauer, 2001]

**Reducing regret to** $O(T^{2/3})$ [Cesa-Bianchi et al., 2006] + l.e.p. lower non-trivial $\rightarrow \Omega(\sqrt{T})$ [Antos et al., 2011]

**Non-stochastic results, apply to stochastic**
Our contribution

<table>
<thead>
<tr>
<th>0</th>
<th>$T^{1/2}$</th>
<th>?</th>
<th>$T^{2/3}$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>trivial</td>
<td>easy</td>
<td>full-info</td>
<td>hard</td>
<td>hopeless</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bandits</td>
<td>l.e.p.</td>
<td></td>
</tr>
</tbody>
</table>

What about the grey area? $\Omega(\sqrt{T})$ and $O\left(\frac{T^2}{3}\right)$

What is in between? Is there a game with $\Theta\left(\frac{T^3}{5}\right)$?

No! We eliminate the grey area.

Dynamic pricing is hard!

Main Theorem

The minimax regret of any finite partial-monitoring game against stochastic opponent can be 0 (trivial), $\tilde{\Theta}(\sqrt{T})$ (easy), $\Theta\left(\frac{T^2}{3}\right)$ (hard) or $\Theta(T)$ (hopeless).
Our contribution

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$T^{1/2}$</td>
<td>?</td>
<td>$T^{2/3}$</td>
<td>$T$</td>
<td></td>
</tr>
<tr>
<td>trivial</td>
<td>easy</td>
<td>full-info</td>
<td>hard</td>
<td>hopeless</td>
<td></td>
</tr>
<tr>
<td></td>
<td>bandits</td>
<td>dynamic pricing</td>
<td>l.e.p.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- What about the grey area? $\Omega(\sqrt{T})$ and $O(T^{2/3})$
Our contribution

What about the grey area? $\Omega(\sqrt{T})$ and $O(T^{2/3})$

What is in between? Is there a game with, say, $\Theta(T^{3/5})$?
What about the grey area? $\Omega(\sqrt{T})$ and $O(T^{2/3})$

What is in between? Is there a game with, say, $\Theta(T^{3/5})$?

No! We eliminate the grey area
### Our contribution

<table>
<thead>
<tr>
<th>0</th>
<th>$T^{1/2}$</th>
<th>$T^{2/3}$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>trivial</td>
<td>easy</td>
<td>hard</td>
<td>hopeless</td>
</tr>
<tr>
<td></td>
<td>full-info</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>bandits</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dynamic pricing</th>
<th>l.e.p.</th>
</tr>
</thead>
</table>

What about the grey area? $\Omega(\sqrt{T})$ and $O(T^{2/3})$

What is in between? Is there a game with, say, $\Theta(T^{3/5})$?

No! We eliminate the grey area

Dynamic pricing is hard!
Our contribution

What about the grey area? $\Omega(\sqrt{T})$ and $O(T^{2/3})$
What is in between? Is there a game with, say, $\Theta(T^{3/5})$?
No! We eliminate the grey area
Dynamic pricing is hard!

Main Theorem

The minimax regret of any finite partial-monitoring game against stochastic opponent can be 0 (trivial), $\tilde{\Theta}(\sqrt{T})$ (easy), $\Theta(T^{2/3})$ (hard) or $\Theta(T)$ (hopeless).
Main tools 1: using $L$

Cell decomposition of the probability simplex (the space of outcome distributions)

$$L = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \end{pmatrix}$$
Main tools 1: using $L$

Cell decomposition of the probability simplex (the space of outcome distributions)

$$L = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \end{pmatrix} \rightarrow (0, 0, 1) \quad (0, 1, 0) \quad (0, 0, 1) \quad (1, 0, 0)$$
Main tools 1: using $L$

Cell decomposition of the probability simplex (the space of outcome distributions)

$$L = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \end{pmatrix} \rightarrow (1, 0, 0)$$

Boundary $\subseteq (\ell_i - \ell_j)_{\perp}$
Row of action $i$ in $H$: $(a \ b \ a \ c)$

Given opponent strategy $p$, probability of observing $a, b, c$?
Main tools 2: using $H$

Row of action $i$ in $H$: $(a \ b \ a \ c)$

Given opponent strategy $p$, probability of observing $a, b, c$?

$$\begin{pmatrix} q_a \\ q_b \\ q_c \end{pmatrix} = \begin{pmatrix} p_1 + p_3 \\ p_2 \\ p_4 \end{pmatrix}$$
Row of action $i$ in $H$: $(a \ b \ a \ c)$

Given opponent strategy $p$, probability of observing $a, b, c$?

$$
\begin{pmatrix}
q_a \\
nb \\
q_c
\end{pmatrix}
= 
\begin{pmatrix}
p_1 + p_3 \\
p_2 \\
p_4
\end{pmatrix}
= 
\begin{pmatrix}
? \\
p_1 \\
p_2 \\
p_3 \\
p_4
\end{pmatrix}
$$
Main tools 2: using $H$

Row of action $i$ in $H$: $(a \ b \ a \ c)$

Given opponent strategy $p$, probability of observing $a, b, c$?

\[
\begin{pmatrix}
q_a \\
q_b \\
q_c
\end{pmatrix} = \begin{pmatrix}
p_1 + p_3 \\
p_2 \\
p_4
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 1 & 0
\end{pmatrix} \begin{pmatrix}
p_1 \\
p_2 \\
p_3 \\
p_4
\end{pmatrix}
\]
Main tools 2: using $H$

Row of action $i$ in $H$: $(a\ b\ a\ c)$

Given opponent strategy $p$, probability of observing $a$, $b$, $c$?

\[
\begin{pmatrix}
q_a \\
q_b \\
q_c
\end{pmatrix} =
\begin{pmatrix}
p_1 + p_3 \\
p_2 \\
p_4
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
p_1 \\
p_2 \\
p_3 \\
p_4
\end{pmatrix}
\]
Row of action $i$ in $H$: $(a \ b \ a \ c)$

Given opponent strategy $p$, probability of observing $a, b, c$?

\[
\begin{pmatrix}
q_a \\
q_b \\
q_c
\end{pmatrix} = \begin{pmatrix} p_1 + p_3 \\ p_2 \\ p_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}
\]
Main tools 2: using $H$

Row of action $i$ in $H$: $(a \ b \ a \ c)$
Given opponent strategy $p$, probability of observing $a, b, c$?

$$
\begin{pmatrix}
q_a \\
q_b \\
q_c
\end{pmatrix} =
\begin{pmatrix}
p_1 + p_3 \\
p_2 \\
p_4
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
p_1 \\
p_2 \\
p_3 \\
p_4
\end{pmatrix}
$$

Indicator rows, signal matrix $S_i$
Row of action $i$ in $H$: $(a \ b \ a \ c)$

Given opponent strategy $p$, probability of observing $a, b, c$?

$$
\begin{pmatrix}
q_a \\
q_b \\
q_c
\end{pmatrix} =
\begin{pmatrix}
p_1 + p_3 \\
p_2 \\
p_4
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}\begin{pmatrix}
p_1 \\
p_2 \\
p_3 \\
p_4
\end{pmatrix}
$$

Indicator rows, signal matrix $S_i$

For more actions: $S_{i,i'} = \begin{pmatrix} S_i \\ S_{i'} \end{pmatrix}$
Main tools 2: using $H$

Row of action $i$ in $H$: $(a \ b \ a \ c)$

Given opponent strategy $p$, probability of observing $a, b, c$?

$$
\begin{pmatrix}
  q_a \\
  q_b \\
  q_c 
\end{pmatrix} = 
\begin{pmatrix}
  p_1 + p_3 \\
  p_2 \\
  p_4 
\end{pmatrix} = 
\begin{pmatrix}
  1 & 0 & 1 & 0 \\
  0 & 1 & 0 & 0 \\
  0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  p_1 \\
  p_2 \\
  p_3 \\
  p_4
\end{pmatrix}
$$

Indicator rows, signal matrix $S_i$

For more actions: $S_{i,i'} = \begin{pmatrix} S_i \\ S_{i'} \end{pmatrix}$

$S_{i,i'}p = S_{i,i'}p' \rightarrow$ no way we can distinguish them
Main tools 2: using $H$

Row of action $i$ in $H$: $(a \ b \ a \ c)$

Given opponent strategy $p$, probability of observing $a$, $b$, $c$?

$$\begin{pmatrix} q_a \\ q_b \\ q_c \end{pmatrix} = \begin{pmatrix} p_1 + p_3 \\ p_2 \\ p_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}$$

Indicator rows, signal matrix $S_i$

For more actions: $S_{i,i'} = \begin{pmatrix} S_i \\ S_{i'} \end{pmatrix}$

$S_{i,i'} p = S_{i,i'} p' \rightarrow$ no way we can distinguish them

nullspace of $S_{i,i'}$ “dangerous”
What makes a game easy?

- “Local observability”
What makes a game easy?

- “Local observability”
- Two neighbor actions, which is better?
What makes a game easy?

- “Local observability"
- Two neighbor actions, which is better?
- Decide without using other actions
What makes a game easy?

- “Local observability”
- Two neighbor actions, which is better?
- Decide without using other actions

The condition: local observability

For every neighboring action pair \( i, i' \), \( \ell_i - \ell_{i'} \) is in the row space of \( S_{i,i'} \).
What makes a game easy?

- “Local observability”
- Two neighbor actions, which is better?
- Decide without using other actions

The condition: local observability

For every neighboring action pair $i, i'$, $\ell_i - \ell_{i'}$ is in the row space of $S_{i,i'}$.

- Why?
What makes a game easy?

- “Local observability”
- Two neighbor actions, which is better?
- Decide without using other actions

The condition: local observability

For every neighboring action pair $i, i'$, $\ell_i - \ell_{i'}$ is in the row space of $S_{i,i'}$.

- Why?
“Local observability”
Two neighbor actions, which is better?
Decide without using other actions

The condition: local observability
For every neighboring action pair $i, i'$, $\ell_i - \ell_{i'}$ is in the row space of $S_{i,i'}$.

Why? Unbiased estimate of $\langle \ell_i - \ell_{i'}, p^* \rangle$: “Which action is better?”
Algorithm outline

- Maintain set of “alive” actions

\[ (0, 0, 1) \]

\[ (0, 1, 0) \]

\[ (1, 0, 0) \]
Algorithm outline

- Maintain set of “alive” actions
- In every “round”, choose each alive action
Algorithm outline

- Maintain set of “alive” actions
- In every “round”, choose each alive action
- Update estimates of loss differences

\[ \hat{O}(\sqrt{T}) \text{ if local observability} \]

\[ (1, 0, 0) \]

\[ (0, 1, 0) \]

\[ (0, 0, 1) \]

\[ p^* \]
Algorithm outline

- Maintain set of “alive” actions
- In every “round”, choose each alive action
- Update estimates of loss differences
- If a loss difference is significant (Bernstein stopping),

\[ p^* \]
Algorithm outline

- Maintain set of “alive” actions
- In every “round”, choose each alive action
- Update estimates of loss differences
- If a loss difference is significant (Bernstein stopping), eliminate suboptimal halfspace
Algorithm outline

- Maintain set of “alive” actions
- In every “round”, choose each alive action
- Update estimates of loss differences
- If a loss difference is significant (Bernstein stopping), eliminate suboptimal halfspace
- Do until only one action, or time step $T$
Algorithm outline

- Maintain set of “alive” actions
- In every “round”, choose each alive action
- Update estimates of loss differences
- If a loss difference is significant (Bernstein stopping), eliminate suboptimal halfspace
- Do until only one action, or time step $T$
- Achieves $\tilde{O}(\sqrt{T})$ if local observability
Lower bound for hard games

- Actions $i$ and $j$ not enough feedback

![Diagram of a triangle with vertices labeled (0,0,1), (0,1,0), (1,0,0), and partially labeled with $i$ and $j$.]

$\Omega\left(\frac{T^2}{3}\right)$ bound

$(1,0,0)$

$(0,1,0)$

$(0,0,1)$

Bartók, Pál, Szepesvári (UofA)
Lower bound for hard games

- Actions $i$ and $j$ not enough feedback
- “dangerous line” crosses $(\text{Ker } S_{i,j})$

![Diagram showing a triangle with vertices (0,0,1), (0,1,0), and (1,0,0) and dashed lines indicating the "dangerous line" crossing through (0,0,1)]
Actions $i$ and $j$ not enough feedback

“dangerous line” crosses ($\text{Ker } S_{i,j}$)

Third action needed, but costly
Lower bound for hard games

- Actions $i$ and $j$ not enough feedback
- “dangerous line” crosses $(\text{Ker } S_{i,j})$
- Third action needed, but costly
- When does this line exist?

$$\text{Coincidence! When no local observability } (\ell_i - \ell_j) \not\in \text{Im } S_{i,j} \Rightarrow \text{unobservable } \iff \text{Ker } S_{i,j} \not\subseteq (\ell_i - \ell_j)$$

Gives $\Omega(\frac{T}{\sqrt{3}})$ bound

$(1, 0, 0)$  $(0, 1, 0)$  $(0, 0, 1)$
Actions $i$ and $j$ not enough feedback

“dangerous line” crosses ($\text{Ker } S_{i,j}$)

Third action needed, but costly

When does this line exist?

Coincidence! When no local observability

$$(\ell_i - \ell_j) \not\in \text{Im } S_{i,j}^\top \iff \text{Ker } S_{i,j} \not\subseteq (\ell_i - \ell_j)^\perp$$

unobservable

line crosses

Gives $\Omega(T^2/3)$ bound

$$(0, 0, 1)$$

$$(0, 1, 0)$$

$$(0, 0, 1)$$

$$(1, 0, 0)$$
Lower bound for hard games

- Actions \(i\) and \(j\) not enough feedback
- “dangerous line” crosses (\(\text{Ker } S_{i,j}\))
- Third action needed, but costly
- When does this line exist?
- Coincidence! When no local observability
  \((\ell_i - \ell_j) \not\in \text{Im } S_{i,j}^\top \iff \text{Ker } S_{i,j} \not\subseteq (\ell_i - \ell_j)^\perp\)
  unobservable \hspace{1cm} line crosses
- Gives \(\Omega(T^{2/3})\) bound
Discussion

- Finite stochastic partial monitoring fully classified
Finite stochastic partial monitoring fully classified

- trivial, easy, hard, hopeless

Key condition separating easy and hard: local observability

New algorithm achieves minimax regret rate for easy games

Computational efficiency: verifying the condition

Scaling with the number of actions? Lower bound: does not scale. Upper bound: $O\left(\frac{N^3}{2}\right)$

Scaling with the number of outcomes? Nope!

Non-stochastic opponent? Conjecture: the classification holds

Algorithm for easy games wanted

Bartók, Pál, Szepesvári (UofA)
Discussion

Finite stochastic partial monitoring fully classified
- trivial, easy, hard, hopeless
- Key condition separating easy and hard: local observability
Finite stochastic partial monitoring fully classified
- trivial, easy, hard, hopeless
- Key condition separating easy and hard: local observability
- New algorithm achieves minimax regret rate for easy games
Finite stochastic partial monitoring fully classified
  - trivial, easy, hard, hopeless
  - Key condition separating easy and hard: local observability
  - New algorithm achieves minimax regret rate for easy games

Computational efficiency
Discussion

- Finite stochastic partial monitoring fully classified
  - trivial, easy, hard, hopeless
  - Key condition separating easy and hard: local observability
  - New algorithm achieves minimax regret rate for easy games

- Computational efficiency
  - verifying the condition
Discussion

- Finite stochastic partial monitoring fully classified
  - trivial, easy, hard, hopeless
  - Key condition separating easy and hard: local observability
  - New algorithm achieves minimax regret rate for easy games

- Computational efficiency
  - verifying the condition
  - the algorithm
Discussion

- Finite stochastic partial monitoring fully classified
  - trivial, easy, hard, hopeless
  - Key condition separating easy and hard: local observability
  - New algorithm achieves minimax regret rate for easy games
- Computational efficiency
  - verifying the condition
  - the algorithm
- Scaling with the number of actions?
Finite stochastic partial monitoring fully classified
- trivial, easy, hard, hopeless
- Key condition separating easy and hard: local observability
- New algorithm achieves minimax regret rate for easy games

Computational efficiency
- verifying the condition
- the algorithm

Scaling with the number of actions?
- lower bound: does not scale
Discussion

- Finite stochastic partial monitoring fully classified
  - trivial, easy, hard, hopeless
  - Key condition separating easy and hard: local observability
  - New algorithm achieves minimax regret rate for easy games

- Computational efficiency
  - verifying the condition
  - the algorithm

- Scaling with the number of actions?
  - lower bound: does not scale
  - upper bound: $O(N^{3/2})$
Finite stochastic partial monitoring fully classified
- trivial, easy, hard, hopeless
- Key condition separating easy and hard: local observability
- New algorithm achieves minimax regret rate for easy games

Computational efficiency
- verifying the condition
- the algorithm

Scaling with the number of actions?
- lower bound: does not scale
- upper bound: $O(N^{3/2})$

Scaling with the number of outcomes? Nope!
Discussion

- Finite stochastic partial monitoring fully classified
  - trivial, easy, hard, hopeless
  - Key condition separating easy and hard: local observability
  - New algorithm achieves minimax regret rate for easy games

- Computational efficiency
  - verifying the condition
  - the algorithm

- Scaling with the number of actions?
  - lower bound: does not scale
  - upper bound: $O(N^{3/2})$

- Scaling with the number of outcomes? Nope!

- Non-stochastic opponent?
Discussion

- Finite stochastic partial monitoring fully classified
  - trivial, easy, hard, hopeless
  - Key condition separating easy and hard: local observability
  - New algorithm achieves minimax regret rate for easy games

- Computational efficiency
  - verifying the condition
  - the algorithm

- Scaling with the number of actions?
  - lower bound: does not scale
  - upper bound: $O(N^{3/2})$

- Scaling with the number of outcomes? Nope!

- Non-stochastic opponent?
  - Conjecture: the classification holds
Discussion

- Finite stochastic partial monitoring fully classified
  - trivial, easy, hard, hopeless
  - Key condition separating easy and hard: local observability
  - New algorithm achieves minimax regret rate for easy games

- Computational efficiency
  - verifying the condition
  - the algorithm

- Scaling with the number of actions?
  - lower bound: does not scale
  - upper bound: $O(N^{3/2})$

- Scaling with the number of outcomes? Nope!

- Non-stochastic opponent?
  - Conjecture: the classification holds
  - Algorithm for easy games wanted
Thank you!

Questions?
Toward a classification of finite partial-monitoring games.

The nonstochastic multiarmed bandit problem.


Regret minimization under partial monitoring.

The weighted majority algorithm.

Discrete prediction games with arbitrary feedback and loss.