Online Combinatorial Optimization with Stochastic Decision Sets and Adversarial Losses

Gergely Neu
gergely.neu@gmail.com
Michal Valko
michal.valko@inria.fr
Sequel Team
INRIA Lille – Nord Europe France

The learning problem

For each time step \( t = 1, 2, \ldots, T \):
- Environment chooses decision set \( S_t \subseteq S \)
- Learner chooses action \( V_t \in S_t \subseteq \{0, 1\}^d \)
- Environment chooses loss vector \( \ell_t \in \{0, 1\}^d \)
- Learner suffers loss \( V_t^T \ell_t \)
- Learner observes feedback based on \( V_t \) and \( \ell_t \)

Loss estimation by Counting Asleep-Times (CAT)

Loss x:
\[
\eta \log \text{component} \cdot 1 \leq \eta \log N_t / \ell \cdot t \leq \eta \log N_t / \ell.
\]

Expected downtime:
\[
E[V_{t+1}^T \ell_{t+1}^T] = 1 / a_t
\]

Analysis idea:
- Define independent copies \( \tilde{S} \sim P \) and \( \tilde{Z} \sim Z \)
- Define hypothetical forecaster \( \tilde{V} = \text{argmin} \ v^T (L - \tilde{Z}) \)
- so \( \tilde{V} \sim V_{t+1} \) given history up to round \( t \)
- Then,
\[
E[V_{t+1}^T \tilde{V}_{t+1}] = E[V_t^T \ell_t]
\]

Lemma 1: for any policy \( \pi \in \Pi \),
\[
E \sum_{t=1}^{T} (\tilde{V}_{t} - \pi(S_t))^T \ell_t \leq \frac{m (\log d + 1)}{\eta}
\]

Lemma 2: for any \( \xi \geq 0 \),
\[
E \left[ (V_{t+1} - \pi(S_{t+1}))^T \ell_{t+1} \right] \leq E \left[ (V_{t} - \pi(S_t))^T \ell_t \right]
\]

Regret definition

Usual goal: minimize regret against \( S \)
\[
R_T = \max E \sum_{t=1}^{T} (V_t - \pi(S_t))^T \ell_t
\]

A sensible comparator class:
\[
\text{The set of policies } \Pi = \{ \pi : \mathbb{R}^2 \rightarrow S \} \text{ s.t. } \pi(S) \in S \subseteq S
\]

Our goal: minimize regret against \( \Pi \)
\[
R_T = \max \text{min} E \sum_{t=1}^{T} (V_t - \pi(S_t))^T \ell_t
\]

Algorithm: SleepingCAT (FPL)

Parameter: learning rate \( \eta > 0 \), \( L_0 = 0 \)
For each time step \( t = 1, 2, \ldots, T \):
- Draw perturbation vector \( Z_t \) with \( \mathbb{P}(Z_t) \sim \text{Exp}(\eta) \) i.i.d.
- Play
\[
V_t = \arg\text{min} \ v^T (L_{t-1} - Z_t)
\]
- Compute \( \tilde{V}_t \) and let \( L_t = L_{t-1} + \tilde{V}_t \)

Computing \( K_{t+1} \):
- use Geometric Resampling (Neu & Bartók, 2013)
- requires \( d \) calls to oracle on average

Results

<table>
<thead>
<tr>
<th></th>
<th>Full info</th>
<th>Restricted info</th>
<th>Semi-bandit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kanade et al. 2009 (( m = 1 ) only)</td>
<td>( \sqrt{T} \log d )</td>
<td>( \eta \log d )</td>
<td>( (dT)^{2/3} (\log d)^{1/3} )</td>
</tr>
<tr>
<td>SleepingCAT general bound</td>
<td>( m^{3/2} \sqrt{L_T} \log d )</td>
<td>( m \sqrt{dT \log d} )</td>
<td>( (mdT)^{2/3} (\log d)^{1/3} )</td>
</tr>
<tr>
<td>SleepingCAT when ( a_t \geq \beta )</td>
<td>( m \sqrt{T \log d / \beta} )</td>
<td>( m \sqrt{T \log d / \beta} )</td>
<td>( (mdT)^{2/3} (\log d)^{1/3} )</td>
</tr>
</tbody>
</table>

Experiments

Setup follows Kanade et al. (2009):
- 5 arms, available independently of each other w.p. \( p \)
- Losses are symmetric random walks truncated to [0,1]

Experiments

Scaling with grid size in a shortest-path problem:
\[
3 \times 3 \quad 10 \times 10
\]