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

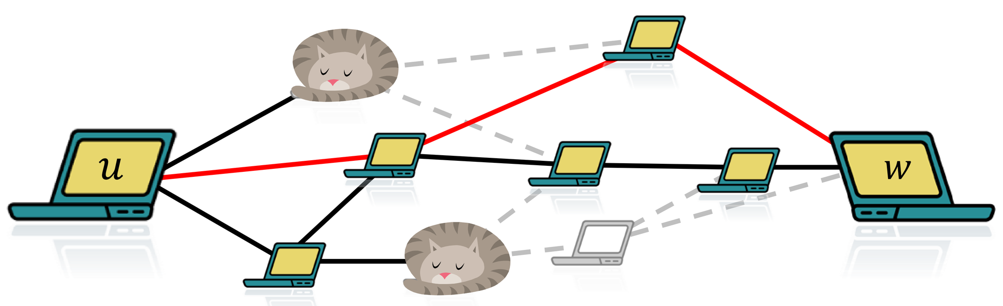
The learning problem

For each time step $t = 1,2,\ldots,T$
- Environment chooses decision set $S_t \subseteq S$
- Learner chooses action $\boldsymbol{V}_t \in S \subseteq \{0,1\}^d$
- Environment chooses loss vector $\ell_t \in [0,1]^d$
- Learner suffers loss $\boldsymbol{V}_t^\top \ell_t$
- Learner observes feedback based on $\boldsymbol{V}_t$ and ℓ_t

Decision set: $S = \{\boldsymbol{v}_i\}_{i=1}^N \subseteq \{0,1\}^d$, $\|\boldsymbol{v}_i\|_1 \le m$

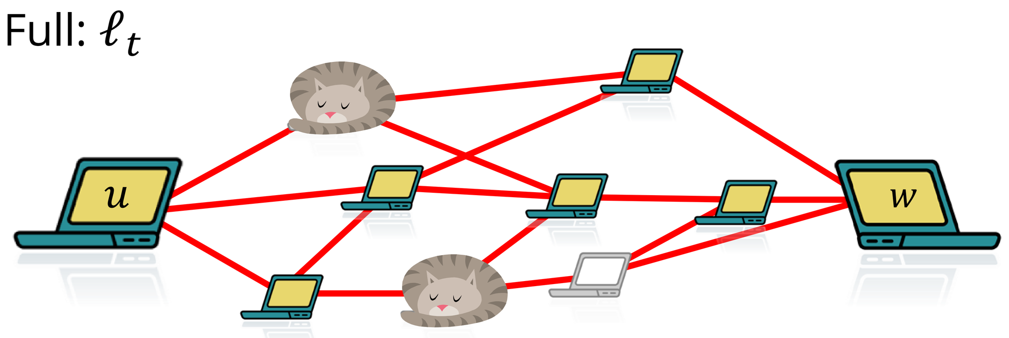
IID availability: $S_t \sim P$

Example: sequential routing

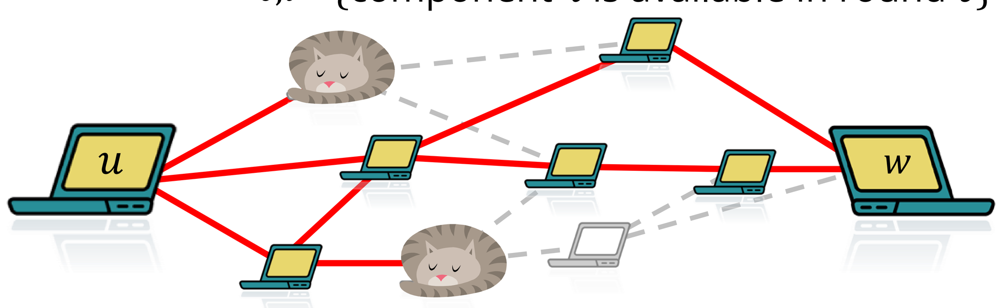


Feedback assumptions

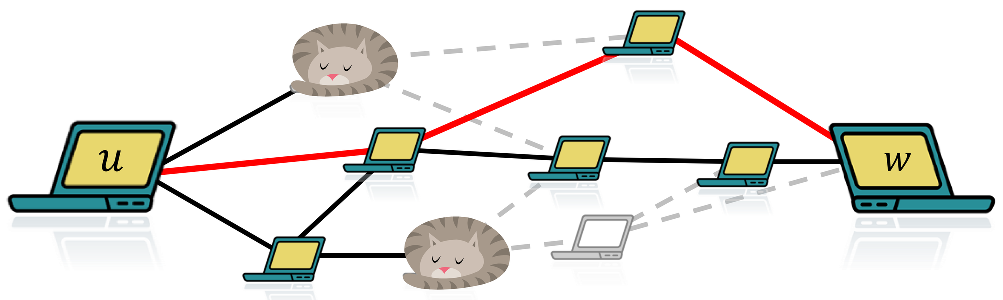
- Full: ℓ_t



- Restricted: $\ell_{t,i}\mathbf{1}_{\{\text{component } i \text{ is available in round } t\}}$



- Semi-bandit: $\ell_{t,i}V_{t,i}$



Regret definition

Usual goal: minimize regret against S

$$R_T = \max_{\boldsymbol{v}\in S} \mathbf{E}\left[\sum_{t=1}^T (\boldsymbol{V}_t - \boldsymbol{v})^\top \ell_t\right]$$

Comparator might not be available!

A sensible comparator class:

The set of policies $\Pi = \{\pi : 2^S \to S\}$ s.t. $\pi(\bar{S}) \in \bar{S} \subseteq S$

Our goal: minimize regret against Π

$$R_T = \max_{\pi\in\Pi} \mathbf{E}\left[\sum_{t=1}^T \left(\boldsymbol{V}_t - \pi(S_t)\right)^\top \ell_t\right]$$

Loss estimation by Counting Asleep-Times (CAT)

Restricted feedback

Define
- $A_{t,i} = \mathbf{1}_{\{\text{component } i \text{ is available in round } t\}}$
- $a_i = \mathbf{E}_t[A_{t,i}]$, and

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{a_i} A_{t,i}$$

… but how do we compute a_i?

Semi-bandit feedback

Define $q_{t,i}^* = \mathbf{E}_t[V_{t,i}]$ and

$$\hat{\ell}_{t,i}^* = \frac{\ell_{t,i}}{q_{t,i}^*} V_{t,i}$$

… but this requires perfect knowledge of P!

Another idea: define $q_{t,i} = \mathbf{E}_t\left[V_{t,i} \middle| S_t\right]$ and

$$\hat{\ell}_{t,i} = \frac{\ell_{t,i}}{a_i q_{t,i}} A_{t,i}$$

… but that needs a_i again.

Estimating $1/a_i$

Kanade, McMahan & Bryan (2009):
- devote first K steps for exploration!
- use samples to construct $\hat{a}_i$ s.t. $\mathbf{E}[\hat{a}_i] = a_i$!

The magic trick: Counting Asleep-Times

Expected downtime: $\mathbf{E}[N_{t,i}] = 1/a_i$

Restricted: $\hat{\ell}_{t,i} = \ell_{t,i}A_{t,i}N_{t,i}$

Semi-bandit: $\hat{\ell}_{t,i} = \ell_{t,i}V_{t,i}N_{t,i}K_{t,i}$

$\mathbf{E}_t[K_{t,i}] = 1/q_{t,i}$

Algorithm: SleepingCAT (FPL)

Parameter: learning rate $\eta > 0$, $\hat{\boldsymbol{L}}_0 = 0$
For each time step $t = 1,2,\ldots,T$
- Draw perturbation vector $\boldsymbol{Z}_t$ with $Z_{t,i} \sim \operatorname{Exp}(\eta)$ i.i.d. for all $i \in \{1,2,\ldots,d\}$
- Play $$\boldsymbol{V}_t = \operatorname*{argmin}_{\boldsymbol{v}\in S_t} \boldsymbol{v}^\top\left(\hat{\boldsymbol{L}}_{t-1} - \boldsymbol{Z}_t\right)$$
- Compute $\hat{\ell}_t$ and let $\hat{\boldsymbol{L}}_t = \hat{\boldsymbol{L}}_{t-1} + \hat{\ell}_t$

Efficient whenever the optimization $\min_{\boldsymbol{v}\in D} \boldsymbol{v}^\top \ell$ can be solved efficiently for all $D \in S$

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

Results

	Full info	Restricted info	Semi-bandit
Kanade et al. 2009 ($m = 1$ only)	$\sqrt{T\log d}$	??? ($T^{3/4}$ conjectured)	$(dT)^{4/5}\log T$
SleepingCAT general bound	$m^{3/2}\sqrt{L_T^*\log d}$	$m\sqrt{dT\log d}$	$(mdT)^{2/3}(\log d)^{1/3}$
SleepingCAT when $a_i \ge \beta$		$m\sqrt{\frac{T\log d}{\beta}}$	$(mT)^{2/3}\left(\frac{d\log d}{\beta}\right)^{1/3}$

Analysis idea

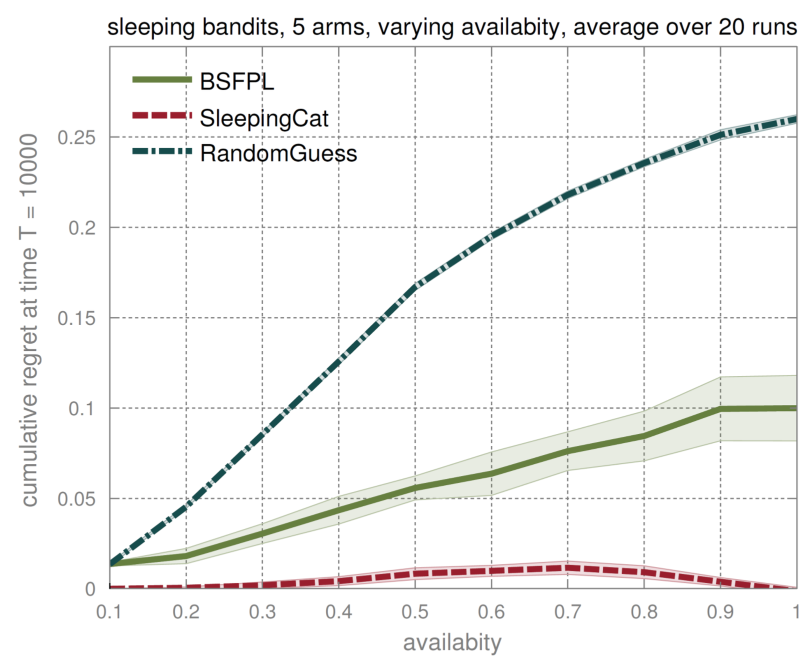
- Define independent copies $\tilde{S} \sim P$ and $\tilde{\boldsymbol{Z}} \sim \boldsymbol{Z}_1$
- Define hypothetical forecaster $$\tilde{\boldsymbol{V}}_t = \operatorname*{argmin}_{\boldsymbol{v}\in\tilde{S}} \boldsymbol{v}^\top\left(\hat{\boldsymbol{L}}_t - \tilde{\boldsymbol{Z}}\right)$$ so $\tilde{\boldsymbol{V}}_t \sim \boldsymbol{V}_{t+1}$ given history up to round t
- Then, $$\mathbf{E}_t\left[\tilde{\boldsymbol{V}}_{t-1}^\top \hat{\ell}_t\right] = \mathbf{E}_t\left[\boldsymbol{V}_t^\top \ell_t\right]$$ $$\mathbf{E}_t\left[\pi(\tilde{S})^\top \hat{\ell}_t\right] = \mathbf{E}_t\left[\pi(S_t)^\top \ell_t\right]$$

Lemma 1: for any policy $\pi \in \Pi$,
$$\mathbf{E}\left[\sum_{t=1}^T \left(\tilde{\boldsymbol{V}}_t - \pi(\tilde{S})\right)^\top \hat{\ell}_t\right] \le \frac{m(\log d + 1)}{\eta}$$

Lemma 2: for any $\hat{\ell}_t \geqslant 0$,
$$\mathbf{E}_t\left[\left(\tilde{\boldsymbol{V}}_{t-1} - \tilde{\boldsymbol{V}}_t\right)^\top \hat{\ell}_t\right] \le \mathbf{E}_t\left[\left(\tilde{\boldsymbol{V}}_{t-1}{}^\top \hat{\ell}_t\right)^2\right]$$

Experiments

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



Scaling with grid size in a shortest-path problem:

3×3 $\qquad$ 10×10