Online Learning For Pandora's Box and Min Sum Set Cover

Evangelia Gergatsouli, Christos Tzamos
UW Madison

A Search Problem

Driving to work, which route to choose? (Traffic/road closures etc)

Spend time finding route with less traffic

Find the best alternative!

- Information is not free
- Explore alternatives (open boxes) } **Strategy**
- Stop anytime and take best so far

Goal: Find minimum cost strategy

Opening cost: 3
Final option: Box 4
Total cost: 35+3

Our setting: repeated search problem

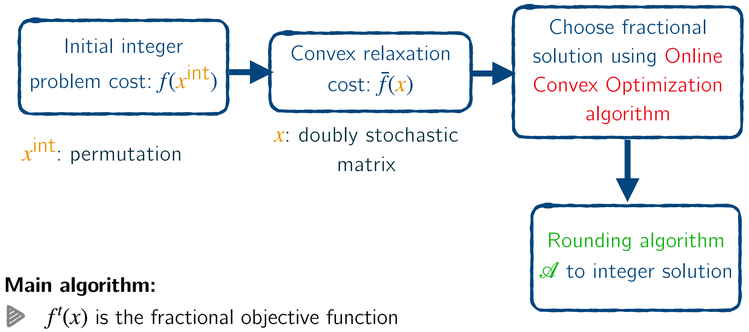
For each day t:

1. New realization of values in boxes
2. Pick a strategy $\mathscr{A}^t$
3. Play x_t according to $\mathscr{A}^t$
4. Receive loss $f^t(x_t)$
5. See loss function f^t on all x's

Full Information Setting

Goal: Obtain α-approximate no regret algorithm vs hindsight optimal

Our framework



Main algorithm:

- $f^t(x)$ is the fractional objective function
- For each round $t \in [T]$ do
  - Set $x_t = \text{OCO}(f^1, \ldots, f^{t-1})$
  - Round x_t to x_t^{int} according to algorithm $\mathscr{A}$
  - Receive loss $f^t(x_t^{\text{int}})$

Regret guarantee $R(T)$ + α-approx rounding algorithm = **Theorem 3.1** $\alpha R(T)$ approx. regret

Optimize the two components independently!

Regret guarantee $R(T)$

- Use **F**ollow **T**he **R**egularized **L**eader family
- Set $OCO = \min_x \sum_{\tau=1}^{t-1} f^\tau(x) + \text{Regularizer}(x)$
- Choose regularizer $= \sum_{i,t} x_{it} \log x_{it}$

α-approximation algorithm

- Use randomized rounding[1]
- Guarantee $\mathbb{E}[\bar{f}(x)] \leq \alpha f(x^{\text{int}})$
- Does not depend on f^t

Corollary 3.3.1
Algorithm is 9.22-approx no regret

Bandit Setting

For each day t:

1. New realization of values in boxes
2. Pick a strategy $\mathscr{A}^t$
3. Play x_t according to $\mathscr{A}^t$
4. Receive loss $f^t(x_t)$
5. See loss function **only** on x_t

Idea:
Balance explore/FTRL steps

Main algorithm:

- Split $[T]$ into intervals $\mathscr{I}_i$, choose uniformly random $t_p \in [\mathscr{I}_i]$, $\mathscr{R} = \emptyset$
- For each interval $\mathscr{I}_i$ and each time $t \in \mathscr{I}_i$
  - If $t = t_p$
    - Open all boxes, include t_p in $\mathscr{R}$
  - Else
    - Set $x_t = \min_x \sum_{\tau \in \mathscr{R}} f^\tau(x) + \text{Regularizer}(x)$
    - Round x_t to x_t^{int} according to algorithm $\mathscr{A}$

Theorem 4.1
In the bandit setting, OCO Algorithm is no regret

Summary of Results

	1 box	k boxes	Matroid basis, size k
α-approx. Regret	$\alpha = 9.22$	$\alpha = O(1)$	$\alpha = O(\log k)$
α-approx. Regret	$\alpha = 3.16$	$\alpha = 12.64$	$\alpha = O(\log k)$

Full information & bandit
Against PA
Against NA

Different ellipsoid-based algorithm

[1] Chawla, Gergatsouli, Teng, Tzamos, Zhang, *FOCS 2020*