Contextual Bandits with Packing & Covering Constraints: A Modular Lagrangian Approach via Regression

Alex Slivkins (MSR-NYC)
Dylan J. Foster (MSR-NYC)
Karthik A. Sankararaman (Meta)

Courtesy Speaker: Jon Schneider (Google)
Contextual Bandits with Linear Constraints

- $T$ rounds, $K$ arms, $d$ resources
- $\forall$ round $t$: algorithm observes context $x_t$, chooses arm $a_t \in [K]$ and observes outcome $\mathbf{o}_t = (r_t; \mathbf{c}_t)$
  - outcome is a vector: $(r_t; c_{t1}, \ldots, c_{td}) \in [0,1] \times [-1,1]^d$
  - Constraint $\sigma_i(C_i - B_i) \leq 0$ for each resource $i$
  - packing constraints: $\leq B_i$; covering constraints: $\geq B_i$
- Stochastic CBwLC: (context & outcome vector $\forall$ arm) drawn IID
- Objectives: regret & constraint violations

Special case: Contextual Bandits with Knapsacks (CBwK) packing constraints, non-negative consumption, null arm (“skip”)
Contextual Bandits (CB) model

CB with **regression oracle**: recent line of work on CB

- (arbitrary) algorithm for **online regression**
  - inputs datapoints (context, arm, real-valued “score”)
  - outputs “regression functions”: contexts $\times$ arms $\rightarrow$ reals
  - achieves low “squared regression error” under realizability

- “true” regression function contained in the “target” class of functions

- Good approach for CB: practicality, theory, practice

The oracle (& realizability) applied separately to rewards & each resource

- CBwLC (or CBwK) with regression oracles not studied in prior work
Some background

BwK are challenging because …

- per-round expected reward $\rightarrow$ total expected reward
- Best arm $\rightarrow$ best distribution over arms

CBwLC (& CBwK) are well-understood …

- without contexts
- for linear rewards & consumptions
- for CB with classification oracles

Another (earlier) standard paradigm for CB
Our contributions

First algorithm for CBwLC with regression oracles

- $T^{3/4}$ regret & constraint violations (possibly suboptimal)
- optimal $\sqrt{T}$ regret for CBwK
- extends to bandit convex optimization

Non-stationary environment: bounded #switches

- even 1 switch $\Rightarrow$ approx. ratio w.r.t. standard benchmarks!
- our algorithm $\Rightarrow$ vanishing regret w.r.t. non-standard benchmark

For each round, best distribution over arms for this round

- new even for BwK (i.e., without contexts, only packing constraints)

concurrent work: Han et al. (2023)

Prior work: “spend or save dilemma”
Techniques

SquareCB to handle CB with regression oracles
- randomization technique: estimated rewards $\rightarrow$ distribution over arms

LagrangeBwK to handle resource constraints
- solves BwK via repeated zero-sum game:
  “primal algorithm” chooses arms, “dual algorithm” chooses resources,
  game payoffs given by a natural Lagrangian relaxation
- This paper: (non-trivial) extension from BwK to BwLC

Modularity of both techniques
  crucial for design, analysis & extensions

- Regression oracle $\rightarrow$ SquareCB $\rightarrow$ “primal algorithm”