Approximately Stationary Bandits with Knapsacks

Giannis Fikioris and Éva Tardos

Cornell University

Introduction and Model

Bandits with Knapsacks (Bwk)

Generalizes Multi-armed Bandits: online decision-making with global constraints

T rounds

d resources

Budget $B = \rho T$ for each resource, $\rho \in [0,1]$

Action $a_t \in [K]$ in round t:

- reward $R_t(a_t) \in [0,1]$
- consumption $C_{t,i}(a_t) \in [0,1]$ of resource i

T_A: round any resource runs out or T

$\mathsf{REW} = \sum_{t=1}^{T_A} r_t(a_t)$

$\exists$ null action with $R_t(\emptyset) = C_{t,i}(\emptyset) = 0$

Stochastic/Adversarial BwK

Stochastic BwK: (R_t, C_t) drawn from distribution. Guarantee (tight):

$$\mathsf{REW} \geq \mathsf{OPT}_{\mathsf{FD}} - \tilde{O}\left(\frac{1}{\rho}\sqrt{KT\log d}\right) \qquad (1)$$

Adversarial BwK: (R_t, C_t) picked by adversary. Guarantee (tight):

$$\mathsf{REW} \geq \rho\,\mathsf{OPT}_{\mathsf{FD}} - \tilde{O}\left(\frac{1}{\rho}\sqrt{KT\log d}\right) \qquad (2)$$

Best-of-both-worlds: Algorithm with both (1) and (2), unaware of environment [Castiglioni et al., ICML'22]

"In-between" BwK

(R_t, C_t) not fully adversarial

E.g., sequential auctions:

- Seasonal changes: item value and price fluctuate
- Bidding against other players: item price unpredictable but not adversarial

(1) inapplicable, (2) too pessimistic

Approximately Stationary BwK

Expectations conditioned on past actions and rewards/consumptions:

- $r_t(a) = \mathbb{E}\left[R_t(a) | \mathcal{H}_{t-1}\right]$
- $c_{t,i}(a) = \mathbb{E}\left[C_{t,i}(a) | \mathcal{H}_{t-1}\right]$

(σ_r, σ_c)**-stationary adversary:**

- σ_r limits rewards: $\forall a$,

$$\min_t r_t(a) \geq \sigma_r \max_t r_t(a)$$

- σ_c limits consumptions: $\forall a, i$,

$$\min_t c_{t,i}(a) \geq \sigma_c \max_t c_{t,i}(a)$$

$(0,0)$-stationary $\iff$ Adversarial BwK

$(1,1)$-stationary $\iff$ Stochastic BwK

Benchmark

Best-fixed distribution of arms:

$$\mathsf{OPT}_{\mathsf{FD}} = \max_{\substack{A \in \Delta([K]) \\ T^* \in [T]}} \sum_{t=1}^{T^*} r_t(A) \quad \text{s.t.} \quad \sum_{t=1}^{T^*} c_{t,i}(A) \leq \rho T \quad \forall i \in [d]$$

Lagrangian Algorithm

Slight modification of algorithms of [Immorlica et al., FOCS'19], [Castiglioni et al., ICML'22]

Idea: find saddle point of "Lagrangian":

$$\mathcal{L}_t(a,i) = R_t(a) + \frac{1}{\rho}\mathbb{1}\left[i \neq 0\right]\left(\rho - C_{t,i}(a)\right)$$

$\mathsf{Alg}_{\mathsf{max}}$ chooses $a_t \in [K]$ to maximize $\sum_t \mathcal{L}(a_t, i_t)$

$\mathsf{Alg}_{\mathsf{min}}$ chooses $i_t \in [d] \cup \{0\}$ to minimize $\sum_t \mathcal{L}(a_t, i_t)$

$\mathsf{Alg}_{\mathsf{max}}$: no-regret algorithm (bandit feedback) with regret $\mathsf{Reg}_{\max}$

$\mathsf{Alg}_{\mathsf{min}}$: no-regret algorithm (full-information) with regret $\mathsf{Reg}_{\max}$

$\mathsf{Reg} = \mathsf{Reg}_{\max} + \mathsf{Reg}_{\min}$

Lower and Upper Bounds

Algorithmic Bound 1

Against (σ_r, σ_c)-stationary adversary:

$$\mathsf{REW} \geq \left(\rho + \sigma_r(\sigma_c - \rho)^+\right)\mathsf{OPT}_{\mathsf{FD}} - \mathsf{Reg}$$

- Continuous and increasing in σ_r, σ_c
- Interpolates (1) and (2)
- $\sigma_r\sigma_c$ fraction of $\mathsf{OPT}_{\mathsf{FD}}$ when $\rho \ll \sigma_r\sigma_c$
- Adversary can be adaptive

Novel key Lemma: For any $A \in \Delta([K])$

$$\mathsf{REW} \geq \min\left\{1, \frac{\rho}{\max_{t,i} c_{t,i}(A)}\right\}\sum_{t=1}^{T} r_t(A) - \mathsf{Reg}$$

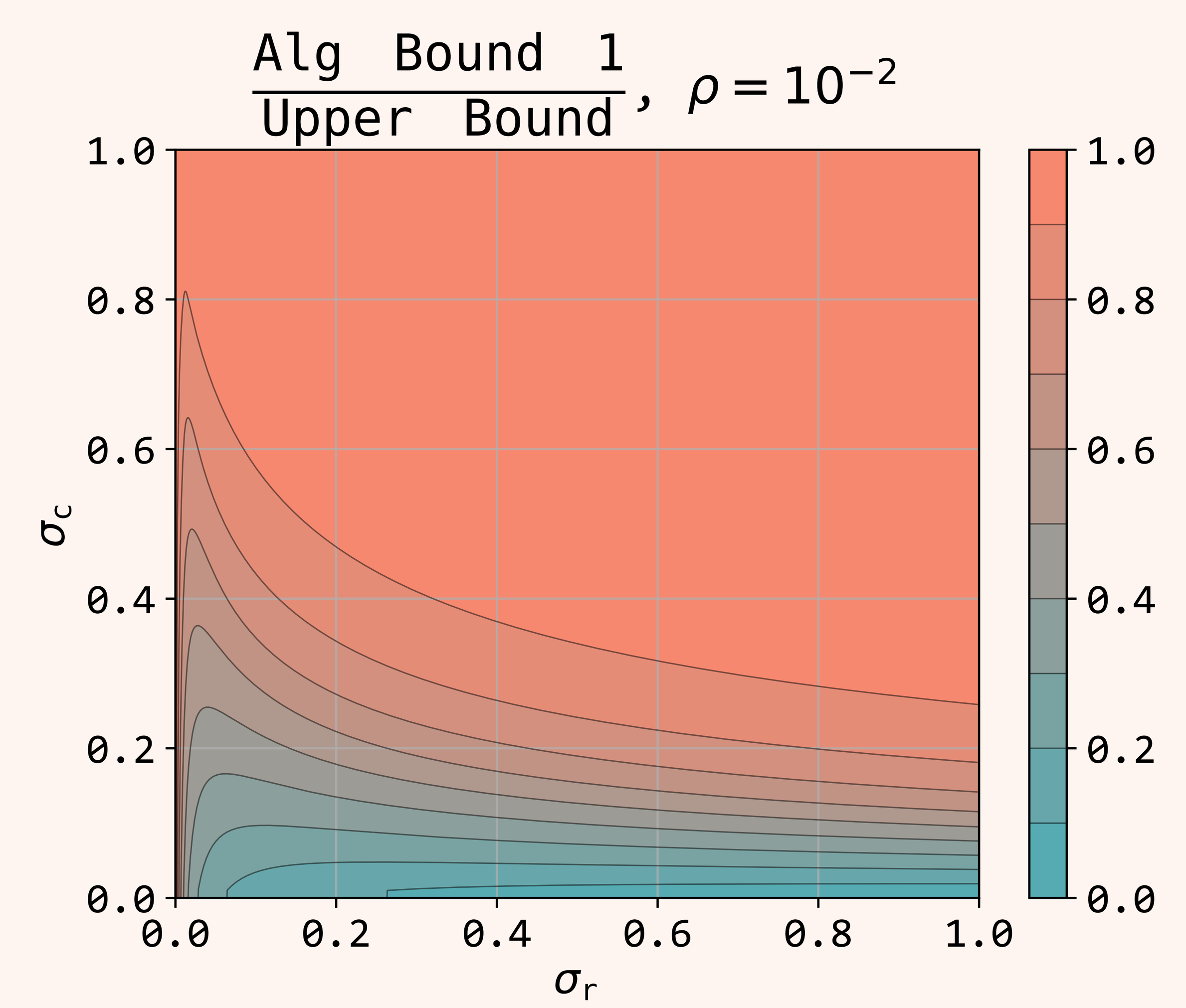
Implies (1) and (2)

Upper Bound

If σ_r, σ_c **unknown** and algorithm guarantees $\alpha_\rho(\sigma_r,\sigma_c)\mathsf{OPT}_{\mathsf{FD}} - o(T)$ with $\alpha_\rho(0,0) \geq \rho$, then

$$\alpha_\rho(\sigma_r,\sigma_c) \leq \begin{cases} \sigma_r + \rho(1-\sigma_r), & \sigma_r \in [0,\rho] \\ 2\sqrt{\sigma_r\rho} - \sigma_r\rho, & \sigma_r \in \left[\rho, \frac{\rho}{\sigma_c^2}\right] \\ \sigma_r\sigma_c + \rho(1/\sigma_c - \sigma_r), & \sigma_r \in \left[\frac{\rho}{\sigma_c^2}, 1\right] \end{cases}$$

- $\alpha_\rho(\sigma_r,\sigma_c) \approx \sigma_r\sigma_c$ when $\rho \ll \sigma_r\sigma_c^2$
- $\alpha_\rho(\sigma_r,\sigma_c) = O(\rho)$ when $\sigma_r = O(\rho)$
- $\alpha_\rho(\sigma_r,\sigma_c) = \rho$ when $\sigma_r = \sigma_c = 0$



Algorithmic Bound 2

Restart Lagrangian algorithm periodically

If $\sum_{t=1}^{T-1}\left|c_{t,i}(a) - c_{t+1,i}(a)\right| \leq o(T)$ guarantee

$$\min_{x\in[\rho,1]}\left(\max\left\{\rho, x\sigma_c, \sigma_r\frac{x}{d+x}\right\} + \max\left\{\rho\sigma_r\frac{1-x}{x}, \sigma_r\sigma_c(1-x)\right\}\right)$$

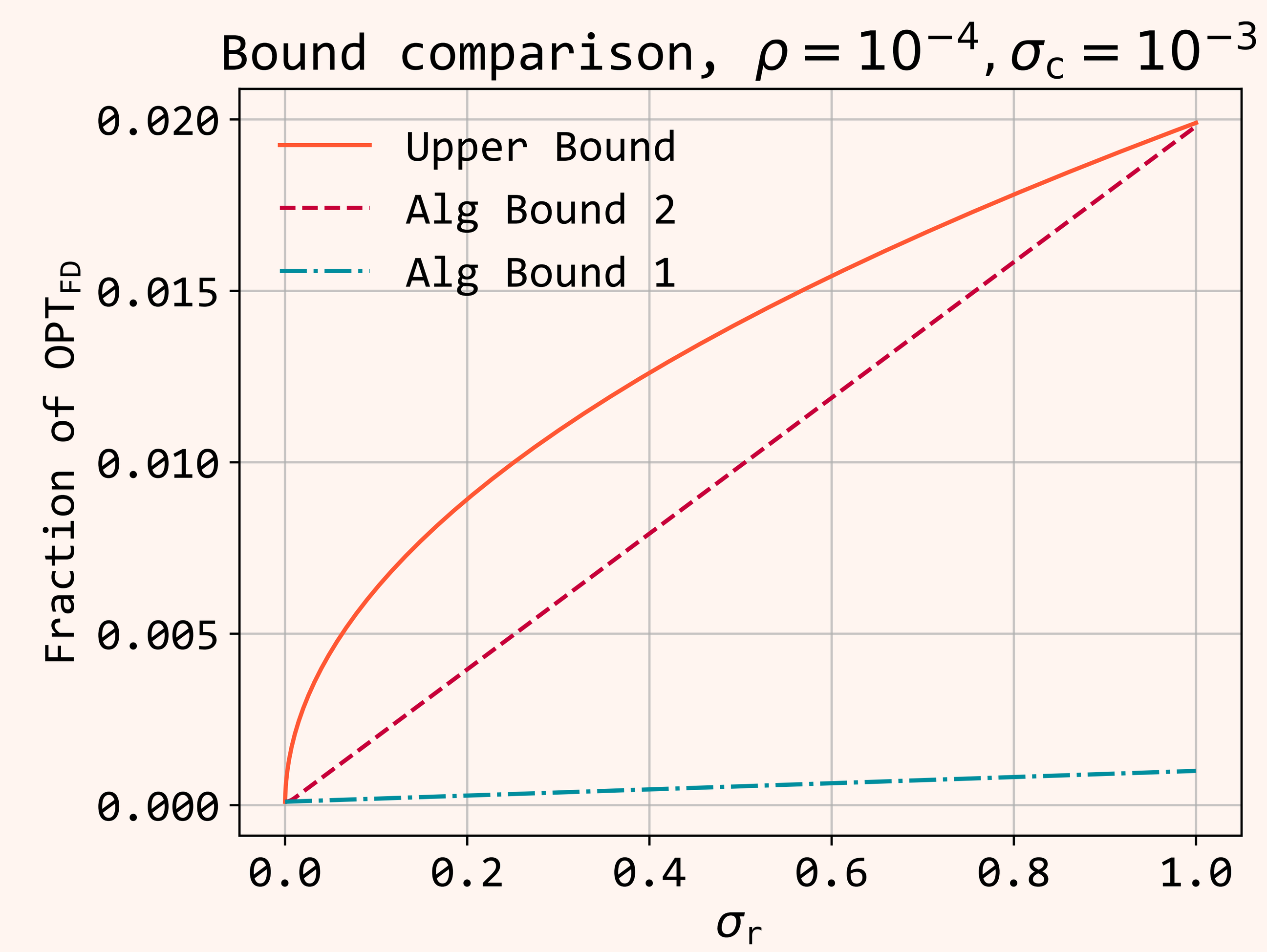
fraction of $\mathsf{OPT}_{\mathsf{FD}}$

- Large improvement over Algorithmic Bound 1 when σ_c, ρ small
- If $\sigma_c \leq \rho \ll \sigma_r$ then

$$\begin{cases} 2\sigma_r(\sqrt{\rho} - \rho), & \sigma_r^2 \geq \rho \\ \sigma_r^2 + \rho - 2\rho\sigma_r, & \sigma_r^2 \leq \rho \end{cases}$$

Improved key Lemma: For any $A \in \Delta([K])$

$$\mathsf{REW} \geq \sum_{t=1}^{T} r_t(A)\min\left\{1, \frac{\rho}{\max_i c_{t,i}(A)}\right\} - o(T)$$



arXiv link