Combinatorial Bandit Optimization for Channel Allocation in Wireless Networks

M. Sadegh Talebi

(Joint work with Alexandre Proutiere and Marc Lelarge)
Channel allocation under uncertainty

- Frequency selective fading
- Unknown average channel qualities
- Packet transmission feedback:
  \[ r_{ij}(t) \sim \text{Ber}(\theta_{ij}) \]
Channel allocation under uncertainty

- Frequency selective fading
- Unknown average channel qualities
- Packet transmission feedback:

\[ r_{ij}(t) \sim \text{Ber}(\theta_{ij}) \]

Sequential learning: At each step, select a mapping from users to channels.
Outline

- Model
- Stochastic combinatorial bandits
- Fundamental performance limit
- Algorithm
- Summary and future directions
Model

- $n$ transmitters (links) and $c$ channels
- Interference: conflict graph $G = ([n], E)$
- Set of feasible allocations (configurations): $\mathcal{M} \subset \{0, 1\}^{n \times c}$
- Configuration $M \in \mathcal{M}$ satisfies:

$$\forall i \in [n], \sum_{j \in [c]} M_{ij} \in \{0, 1\},$$

$$\forall (i, i') \in E \implies \forall j \in [c] : M_{ij} M_{i'j} = 0,$$

- Full interference case: When $G = ([n], E)$ is a complete graph.
A combinatorial bandit problem

- Learn the best allocation sequentially:
  - At each time choose a feasible configuration $M \in \mathcal{M}$
  - Receive feedback of chosen (user, channel) pairs

$$M(t) : \quad r(t) = [r_{ij}(t), (i, j) \in M(t)]$$
A combinatorial bandit problem

- Learn the best allocation sequentially:
  - At each time choose a feasible configuration $M \in \mathcal{M}$
  - Receive feedback of chosen (user, channel) pairs

$$M(t) : \quad r(t) = [r_{ij}(t), (i, j) \in M(t)]$$

- $\mathcal{M}$ has a combinatorial structure
  - E.g., full interference case: $n = c \implies |\mathcal{M}| = n!$
  - Goal: minimize the cumulative loss of not choosing the best allocation (regret)
  - Need to optimally exploit structure, how?
Unstructured Stochastic Bandits

Robbins 1952

- \(K\) arms
- Rewards: \(X_{i,t} \sim \text{Ber} (\mu_i), \quad \mu^* = \max_i \mu_i = \mu_{i^*}\)
- Regret:

\[
R^\pi(T) = \max_{i=1,\ldots,K} \mathbb{E} \sum_{t=1}^T X_{i,t} - \mathbb{E} \sum_{t=1}^T X_{I_i^\pi,t}
\]

**Theorem** (Lai-Robbins 1985) For any uniformly good policy \(\pi\)

\[
\lim_{T \to \infty} \inf \frac{R^\pi(T)}{\log(T)} \geq \sum_{i \neq i^*} \frac{\mu^* - \mu_i}{\text{KL}(\mu_i, \mu^*)}
\]

Regret scales linearly in the number of arms.
Stochastic combinatorial bandits

- Define: \( r_M(t) = \sum_{i,j} M_{ij} r_{ij}(t) = M \cdot r(t) \)

- Regret \( R^\pi(T) = \max_{M \in \mathcal{M}} \mathbb{E}\left[ \sum_{t=1}^{T} r_M(t) \right] - \mathbb{E}\left[ \sum_{t=1}^{T} r_{M^\pi(t)}(t) \right] \)

- Naïve approach: treat each configuration as an arm;
  \( R(T) \sim |\mathcal{M}| \log(T) \)

- State-of-the-art: UCB (Kalathil et al. 2012) for full interference case and \( n=c \):
  \( R^{UCB}(T) \sim n^5 \log(T) \)
Fundamental performance limit

Let $\theta = [\theta_{ij}, i \in [n], j \in [c]]$.

**Theorem**: For any uniformly good policy $\pi$:

$$\lim_{T \to \infty} \inf \frac{R^\pi(T)}{\log(T)} \geq C(\theta),$$

$$C(\theta) = \inf_{x_M \geq 0, M \in \mathcal{M}} \sum_{M \in \mathcal{M}} x_M (M^* \bullet \theta - M \bullet \theta)$$

s.t.

$$\inf_{\lambda \in B_M(\theta)} \sum_{Q \neq M^*} x_Q \sum_{i,j} M_{ij} \text{KL} (\theta_{ij}, \lambda_{ij}) \geq 1, \quad \forall M \neq M^*,$$

where

$$B_M(\theta) = \{\lambda \in [0,1]^{n \times c} : (\forall i, j : M^*_{ij} = 1, \lambda_{ij} = \theta_{ij}), \text{ and } M^* \bullet \theta < M \bullet \lambda\}.$$  

*Graves-Lai 1997*
Fundamental performance limit

Scaling of the regret lower bound in the full interference case:

**Theorem:** For almost every $\theta \in [0, 1]^{n \times c}$:

$$C(\theta) = \Theta(n \times c), \quad \text{as } n, c \to \infty$$

Regret scales with number of unknown parameters rather than size of the decision space.

$$n = c, \quad |\mathcal{M}| = n! \quad \implies R(T) \sim n^2 \log(T)$$

**Remark:** Can be strengthened for general interference graphs.
Algorithm

- Identify a set $\mathcal{A} \subset \mathcal{M}$ that covers all $(i, j)$ pairs.
- $\hat{r}(t)$ : empirical average reward

$\varepsilon$-Greedy algorithm

Let $\varepsilon_t = \min(1, d/t)$.

Select $M(t) \in \arg\max_{M \in \mathcal{M}} M \cdot \hat{r}(t)$ w.p. $1 - \varepsilon_t$ and a configuration uniformly at random from $\mathcal{A}$ w.p. $\varepsilon_t$. 
Theorem: For $d > 10A \frac{n^2}{\Delta_{\min}^2}$:

$$R^{\epsilon-greedy}(T) \leq \Delta_{\max} d \log(T) + O(1) \quad \text{as} \quad T \to \infty$$

where

$$\Delta_{\min} = \min_{M \neq M^*} (M^* \cdot \theta - M \cdot \theta),$$

$$\Delta_{\max} = \max_M (M^* \cdot \theta - M \cdot \theta).$$

For the full interference case and $n=c$:

$$R^{\epsilon-greedy}(T) \sim n^3 \log(T)$$

Remark: Improves over existing algorithms

$$R^{UCB}(T) \sim n^5 \log(T)$$
Summary

- Spectrum sharing under uncertainty is modeled as a combinatorial bandit problem in stochastic setting.
- We established regret lower bound as an optimization problem and analyzed its scaling for the full interference case.
- We provided an algorithm that improves upon existing algorithms.
Future work

- Optimal algorithm (that enjoys simplicity)
- Scalability of regret lower bound for general interference graphs
- Non-stationary environment
Thank you
Optimal Static Allocation

- Identify $M^*$ by solving the following ILP:

$$\max \sum_{i \in [n], j \in [c]} \gamma_{i,j} M_{i,j}$$

over $M \in \mathcal{M}$.

- Adversarial setting: $\gamma_{i,j} = \sum_{t=1}^{T} r_{i,j}(t)$
- Stochastic setting: $\gamma_{i,j} = \theta_{i,j}$

- NP-complete for general interference graphs
- Polynomial for full interference case
Controlled Markov Chains

Graves – Lai 1997

- Finite state and action spaces
- Unknown parameter $\theta \in \Theta$
  $\Theta$ : compact metric space

- Control: finite set of irreducible control laws $g : \mathcal{X} \rightarrow \mathcal{U}$
  $$\mu_g(\theta) = \sum_{x \in \mathcal{X}} \pi^g_\theta(x) r(x, g(x))$$

- Optimal control law: $g^* \in \arg \max_g \mu_g(\theta)$

- Regret:
  $$R^\pi(T) = T \mu_{g^*}(\theta) - \mathbb{E} \sum_{t=1}^T r(X_t, g^\pi(X_t))$$
Stochastic combinatorial bandits: Fundamental performance limit

- KL number: \[ I^g(\theta, \lambda) = \sum_{x,y} \pi^g_\theta(x)p(x, y; g(x), \theta) \log \frac{p(x, y; g(x), \theta)}{p(x, y; g(x), \lambda)} \]

- Bad parameter set: \[ B(\theta) = \{ \lambda \in \Theta : g^* \text{ not opt.}, I^{g^*}(\theta, \lambda) = 0 \} \]

- Lower bound:

\[ \lim_{T \to \infty} \inf \frac{R(T)}{\log(T)} \geq C(\theta) \]

\[ C(\theta) = \inf \sum_{g \neq g^*} c_g (\mu_{g^*}(\theta) - \mu_g(\theta)) \]

\[ \text{s.t. } \inf_{\lambda \in B(\theta)} \sum_{g \neq g^*} c_g I^g(\theta, \lambda) \geq 1 \]