Fast Pure Exploration via Frank-Wolfe

Po-An Wang[1], Ruo-Chun Tzeng[2], and Alexandre Proutiere[1]

Conference on Neural Information Processing Systems, 2021

[1]EECS, Division of Decision and Control System

[2]EECS, Division of Theoretical Computer Science
KTH Royal Institute of Technology

Pure exploration on structured bandits

Stochastic Multi-Armed Bandit (MAB)

K arms (K prob. distribution $\nu_1, \ldots, \nu_K$), the mean of ν_k is μ_k

ν_1

ν_2

ν_3

ν_4

ν_5

Stochastic Multi-Armed Bandit (MAB)

K arms (K prob. distribution $\nu_1, \ldots, \nu_K$), the mean of ν_k is μ_k





ν_1 ν_2 ν_3 ν_4 ν_5

In round t, an agent

1. pulls arm $A_t \in [K]$
2. receives the reward $X_{A_t}(t) \sim \nu_{A_t}$

Sequential sampling strategy: $A_t \in \mathcal{F}_t = \sigma[A_1, X_1, \ldots, A_{t-1}, X_{t-1}]$

Pure exploration with fixed confidence

Goal: Identify a certain answer $i^\star(\boldsymbol{\mu}) \in \mathcal{I}$
Example: Identify the best arm $i^\star(\boldsymbol{\mu}) = \text{argmax}_{k \in [K]} \mu_k$

A strategy consists of

- a sampling rule A_t (arm to explore)
- a stopping rule τ (time to stop)
- a $\mathcal{F}_\tau$-measurable decision rule $\hat{\imath} \in \mathcal{I}$ (answer to return)

Pure exploration with fixed confidence

Goal: Identify a certain answer $i^\star(\boldsymbol{\mu}) \in \mathcal{I}$
Example: Identify the best arm $i^\star(\boldsymbol{\mu}) = \mathrm{argmax}_{k\in[K]}\, \mu_k$

A strategy consists of

- a sampling rule A_t (arm to explore)
- a stopping rule τ (time to stop)
- a $\mathcal{F}_\tau$-measurable decision rule $\hat{\imath} \in \mathcal{I}$ (answer to return)

We wish to minimize $\mathbb{E}_{\boldsymbol{\mu}}[\tau]$ subject to $\mathbb{P}_{\boldsymbol{\mu}}[\hat{\imath} \neq i^\star(\boldsymbol{\mu})] < \delta$

Structured bandits

"Side information" is encoded by the **structure**

Popular structures: Unstructured, Linear, Lipschitz, Dueling, Combinatorial, Unimodal, Monotone, Spectral and Cascading

Structured bandits

"Side information" is encoded by the **structure**

Popular structures: Unstructured, Linear, Lipschitz, Dueling, Combinatorial, Unimodal, Monotone, Spectral and Cascading

Question 1. What is the sample complex gain achievable when exploiting the structure?

Structured bandits

"Side information" is encoded by the **structure**

Popular structures: Unstructured, Linear, Lipschitz, Dueling, Combinatorial, Unimodal, Monotone, Spectral and Cascading

Question 1. What is the sample complex gain achievable when exploiting the structure?

Question 2. Can we devise a computational efficient algorithm achieving the promised gains for all structures?

Lower bound [GK16]

For any **good** strategy,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau]}{\log(\frac{1}{\delta})} \geq T^\star(\boldsymbol{\mu}),$$

where $T^\star(\boldsymbol{\mu})^{-1} = \sup_{\boldsymbol{\omega} \in \boldsymbol{\Sigma}} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k)$

- Σ :$K-1$ simplex
- $\mathrm{Alt}(\boldsymbol{\mu}) = \{\boldsymbol{\lambda} \in \boldsymbol{\Lambda} : i^\star(\boldsymbol{\lambda}) \neq i^\star(\boldsymbol{\mu})\}$
- $d(\mu_k, \lambda_k)$: KL-divergent of arm-k reward distribution under $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$

Lower bound [GK16]

For any **good** strategy,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau]}{\log(\frac{1}{\delta})} \geq T^\star(\boldsymbol{\mu}),$$

where $T^\star(\boldsymbol{\mu})^{-1} = \sup_{\boldsymbol{\omega} \in \boldsymbol{\Sigma}} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k)$

- Σ :$K - 1$ simplex
- $\mathrm{Alt}(\boldsymbol{\mu}) = \{\boldsymbol{\lambda} \in \boldsymbol{\Lambda} : i^\star(\boldsymbol{\lambda}) \neq i^\star(\boldsymbol{\mu})\}$
- $d(\mu_k, \lambda_k)$: KL-divergent of arm-k reward distribution under $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$

$\Rightarrow$ An optimal algorithm has a sampling strategy described by

$$\boldsymbol{\omega}^\star(\boldsymbol{\mu}) \in \operatorname*{argmax}_{\boldsymbol{\mu} \in \boldsymbol{\Sigma}} F_{\boldsymbol{\mu}}(\boldsymbol{\omega}),$$

$$\text{where } F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k).$$

Generalized Likelihood Ratio Test (GLRT)

For each $k \in [K]$, $t \geq 1$, denote

- $N_k(t) = \sum_{s=1}^{t} \mathbb{1}\{A_s = k\}$,
- $\omega_k(t) = N_k(t)/t$,
- $\hat{\mu}_k(t) = \sum_{s=1}^{t} X_k(s)\mathbb{1}\{A_s = k\}/N_k(t)$ (when $N_k(t) > 0$),

where $X_k(s)$ is the reward by pulling arm k at time s.

Generalized Likelihood Ratio Test (GLRT)

For each $k \in [K]$, $t \geq 1$, denote

- $N_k(t) = \sum_{s=1}^{t} \mathbb{1}\{A_s = k\}$,
- $\omega_k(t) = N_k(t)/t$,
- $\hat{\mu}_k(t) = \sum_{s=1}^{t} X_k(s) \mathbb{1}\{A_s = k\}/N_k(t)$ (when $N_k(t) > 0$),

where $X_k(s)$ is the reward by pulling arm k at time s.

GLRT is the stopping rule s.t.

$\tau = \inf\{t \geq 1 : t F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \geq \beta(t, \delta)\}$, where $\beta(t, \delta)$ satisfies:

$$\forall t \geq 1, \quad \left(t F_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \geq \beta(t, \delta) \right) \Longrightarrow \left(\mathbb{P}_{\boldsymbol{\mu}} \left[i^\star(\hat{\boldsymbol{\mu}}(t)) \neq i^\star(\boldsymbol{\mu}) \right] \leq \delta \right),$$

$$\exists c_1(\Lambda), c_2(\Lambda) > 0 \ : \ \forall t \geq c_1(\Lambda), \quad \beta(t, \delta) \leq \log \left(\frac{c_2(\Lambda) t}{\delta} \right).$$

Generalized Likelihood Ratio Test (GLRT)

For each $k \in [K]$, $t \geq 1$, denote

- $N_k(t) = \sum_{s=1}^{t} \mathbb{1}\{A_s = k\}$,
- $\omega_k(t) = N_k(t)/t$,
- $\hat{\mu}_k(t) = \sum_{s=1}^{t} X_k(s) \mathbb{1}\{A_s = k\}/N_k(t)$ (when $N_k(t) > 0$),

where $X_k(s)$ is the reward by pulling arm k at time s.

GLRT is the stopping rule s.t.

$\tau = \inf\{t \geq 1 : tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \geq \beta(t,\delta)\}$, where $\beta(t,\delta)$ satisfies:

$$\forall t \geq 1, \quad \left(tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(t)) \geq \beta(t,\delta)\right) \Longrightarrow \left(\mathbb{P}_{\boldsymbol{\mu}}\left[i^\star(\hat{\boldsymbol{\mu}}(t)) \neq i^\star(\boldsymbol{\mu})\right] \leq \delta\right),$$

$$\exists c_1(\Lambda), c_2(\Lambda) > 0 \ : \ \forall t \geq c_1(\Lambda), \quad \beta(t,\delta) \leq \log\left(\frac{c_2(\Lambda)t}{\delta}\right).$$

Many literatures[GK16, KK18, JP20, Mén19] provide such $\beta(t,\delta)$

To reach the optimality

Challenges for sampling rules:

(i). $\boldsymbol{\mu}$ is unknown initially

(ii). **No** oracle for $\max_{\boldsymbol{\omega}\in\boldsymbol{\Sigma}} \inf_{\boldsymbol{\lambda}\in\mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k)$ in general

To reach the optimality

Challenges for sampling rules:

(i). $\boldsymbol{\mu}$ is unknown initially

(ii). **No** oracle for $\max_{\boldsymbol{\omega}\in\boldsymbol{\Sigma}} \inf_{\boldsymbol{\lambda}\in\mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k)$ in general

Previous works solve (i). by using forced exploration [GK16, Mén19] or building a confidence interval [DKM19]

To reach the optimality

Challenges for sampling rules:

(i). $\boldsymbol{\mu}$ is unknown initially

(ii). **No** oracle for $\max_{\boldsymbol{\omega}\in\boldsymbol{\Sigma}} \inf_{\boldsymbol{\lambda}\in\mathrm{Alt}(\boldsymbol{\mu})} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k)$ in general

Previous works solve (i). by using forced exploration [GK16, Mén19] or building a confidence interval [DKM19]
For (ii)., Previous works either solve maxmini program for a single problem or converge to the saddle with overly conservative approach[Mén19, DKM19]

Frank-Wolfe based sampling (FWS)

Best challenger (BC) in unstructured bandits [GK16, Mén19]

Let $i^\star = i^\star(\boldsymbol{\mu})$ and $\mathrm{Alt}(\boldsymbol{\mu}) = \cup_{j \neq i^\star}\{\boldsymbol{\lambda} \in \Lambda : \lambda_j \geq \lambda_{i^\star}\}$, then $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \min_{j \neq i^\star} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$, where

$$
\begin{aligned}
f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) &= \inf_{\lambda_j \geq \lambda_{i^\star}} \sum_{k=1}^{K} \omega_k d(\mu_k, \lambda_k) \\
&= \omega_j d(\mu_j, \frac{\omega_{i^\star}\mu_{i^\star} + \omega_j \mu_j}{\omega_{i^\star} + \omega_j}) + \omega_{i^\star} d(\mu_{i^\star}, \frac{\omega_{i^\star}\mu_{i^\star} + \omega_j \mu_j}{\omega_{i^\star} + \omega_j})
\end{aligned}
$$

Also, $\nabla_\omega f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = d(\mu_j, \frac{\omega_{i^\star}\mu_{i^\star} + \omega_j \mu_j}{\omega_{i^\star} + \omega_j})\boldsymbol{e}_j + d(\mu_{i^\star}, \frac{\omega_{i^\star}\mu_{i^\star} + \omega_j \mu_j}{\omega_{i^\star} + \omega_j})\boldsymbol{e}_{i^\star}$

After pulling each arm once, BC repeatedly does:

1. Assign $C_t \leftarrow \mathrm{argmin}_{j \neq i^\star(\hat{\boldsymbol{\mu}}(t))} f_j(\boldsymbol{\omega}(t), \hat{\boldsymbol{\mu}}(t))$
2. Play

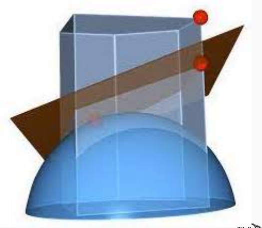
$$
A_t \leftarrow \begin{cases} \hat{\imath} = i^\star(\hat{\boldsymbol{\mu}}(t)), \text{ if } d(\mu_j, \frac{\omega_{\hat{\imath}}\mu_{\hat{\imath}} + \omega_j \mu_j}{\omega_{\hat{\imath}} + \omega_j}) > d(\mu_{\hat{\imath}}, \frac{\omega_{\hat{\imath}}\mu_{\hat{\imath}} + \omega_j \mu_j}{\omega_{\hat{\imath}} + \omega_j}) \\ C_t, \text{otherwise.} \end{cases}
$$

Frank-Wolfe algorithm (FW)

In the view of updating $\boldsymbol{\omega}(t)$, BC corresponds to FW iteration as if the objective function is smooth **(unfortunately, it is not)**

FW for $\max_{\boldsymbol{x}\in\Sigma} F(\boldsymbol{x})$ when F is smooth
Take $\boldsymbol{x}(1) \in \Sigma$ arbitrarily
For $t = 1, \ldots, T$ do:

1. $\boldsymbol{z}(t+1) \leftarrow \text{argmax}_{\boldsymbol{z}\in\Sigma}\langle \boldsymbol{z}, \nabla F(\boldsymbol{x}(t))\rangle$
2. $\boldsymbol{x}(t+1) \leftarrow \frac{t}{t+1}\boldsymbol{x}(t) + \frac{1}{t+1}\boldsymbol{z}(t+1)$



Curvature

For a compact set $\mathcal{K}$ and a concave function $\psi : \mathcal{K} \mapsto \mathbb{R}$, we define

$$C_\psi(\mathcal{K}) = \sup_{\substack{\boldsymbol{x},\boldsymbol{z}\in\mathcal{K} \\ \alpha\in(0,1] \\ \boldsymbol{y}=\boldsymbol{x}+\alpha(\boldsymbol{z}-\boldsymbol{x})}} \min_{h\in\partial\psi(\boldsymbol{x})} \frac{1}{\alpha^2}\left[\psi(\boldsymbol{x}) - \psi(\boldsymbol{y}) + \langle \boldsymbol{y}-\boldsymbol{x}, h\rangle\right] \quad (1)$$

Curvature

For a compact set $\mathcal{K}$ and a concave function $\psi : \mathcal{K} \mapsto \mathbb{R}$, we define

$$C_\psi(\mathcal{K}) = \sup_{\substack{\boldsymbol{x},\boldsymbol{z}\in\mathcal{K} \\ \alpha\in(0,1] \\ \boldsymbol{y}=\boldsymbol{x}+\alpha(\boldsymbol{z}-\boldsymbol{x})}} \min_{h\in\partial\psi(\boldsymbol{x})} \frac{1}{\alpha^2}\left[\psi(\boldsymbol{x}) - \psi(\boldsymbol{y}) + \langle \boldsymbol{y} - \boldsymbol{x}, h\rangle\right] \quad (1)$$

When $\mathcal{K}$ is a convex domain, a finite curvature permits the convergence of FW (see e.g. [Jag13]). The intuition is that $C_\psi(\mathcal{K})$ provides a controlled bound for each iteration as

$$\psi(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, h\rangle - \frac{C_\psi(\mathcal{K})}{\alpha^2} \leq \psi(\boldsymbol{y}) \leq \psi(\boldsymbol{x}) + \langle \boldsymbol{y} - \boldsymbol{x}, h\rangle,$$

where $h \in \partial\psi(\boldsymbol{x})$ is the one attaining minimum in (1)

Why does BC fail to reach the optimal allocation?

BC faces three issues:

(i). $F_{\boldsymbol{\mu}}$ is not smooth

(ii). Each f_j has an unbounded curvature close to the boundary of Σ

(iii). $\boldsymbol{\mu}$ is unknown initially

Why does BC fail to reach the optimal allocation?

BC faces three issues:

(i). $F_{\boldsymbol{\mu}}$ is not smooth

(ii). Each f_j has an unbounded curvature close to the boundary of Σ

(iii). $\boldsymbol{\mu}$ is unknown initially

We devise a simple algorithm (FW-based) to track $\boldsymbol{x(t)} \xrightarrow{t\to\infty} \boldsymbol{\omega}^\star(\boldsymbol{\mu})$ by circumventing these issues.

Non-smoothness issue

For **(i).** F_μ is not smooth

Non-smoothness issue

For **(i).** $F_{\boldsymbol{\mu}}$ is not smooth

- Under mild assumption, we show $F_{\boldsymbol{\mu}}$ is the minimum of a finite number of smooth concave functions f_j by *envelop theorem*

Non-smoothness issue

For **(i).** $F_{\boldsymbol{\mu}}$ is not smooth

- Under mild assumption, we show $F_{\boldsymbol{\mu}}$ is the minimum of a finite number of smooth concave functions f_j by *envelop theorem*
- Leveraging this fact, we have a novel and computational efficient construction which continuously approximates the non-smooth points

Unbounded curvature and unknown μ

The remaining issues,

(ii). Each f_j has an unbounded curvature close to the boundary of Σ

(iiii). μ is unknown initially,

are solved by a single trick

Unbounded curvature and unknown μ

The remaining issues,

(ii). Each f_j has an unbounded curvature close to the boundary of Σ

(iiii). $\boldsymbol{\mu}$ is unknown initially,

are solved by a single trick

Let updated direction $z(t)$ cover $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K$ sufficiently often so that the tracked allocation, $\boldsymbol{x}(t)$, is kept away from the boundary and each action is forced to be played frequently enough

Assumption 1 and an example

Assumption 1

$\forall i \in \mathcal{I}$, $\mathcal{S}_i = \{\boldsymbol{\mu} \in \Lambda : i^\star(\boldsymbol{\mu}) = i\}$ is open and its complementary $\Lambda \setminus \mathcal{S}_i$ is a finite union of convex set. Namely, a finite collection $\mathcal{J}_i$ of convex set $\mathcal{C}_j^i$ s.t. $\Lambda \setminus \mathcal{S}_i = \cup_{j \in \mathcal{J}_i} \mathcal{C}_j^i$

Assumption 1 and an example

Assumption 1

$\forall i \in \mathcal{I}$, $\mathcal{S}_i = \{\boldsymbol{\mu} \in \Lambda : i^\star(\boldsymbol{\mu}) = i\}$ is open and its complementary $\Lambda \setminus \mathcal{S}_i$ is a finite union of convex set. Namely, a finite collection $\mathcal{J}_i$ of convex set $\mathcal{C}_j^i$ s.t. $\Lambda \setminus \mathcal{S}_i = \cup_{j \in \mathcal{J}_i} \mathcal{C}_j^i$

Example: BAI for unstructured bandit

Here $\Lambda = \{\boldsymbol{\mu} \in (0,1)^K : \exists i \in [K] \text{ s.t. } \mu_i > \mu_k, \forall k \neq i\}$, and for each $i \in [K]$, $\mathcal{S}_i = \{\boldsymbol{\mu} \in \Lambda : \mu_i > \mu_k, \forall k \neq i\}$

We can see that $\Lambda \setminus \mathcal{S}_i = \cup_{j \neq i} \mathcal{C}_j^i$, where $\mathcal{C}_j^i = \{\boldsymbol{\lambda} \in \Lambda : \lambda_j > \lambda_i\}$ is a convex set $\forall j \neq i$

Assumption 1 and an example

Assumption 1

$\forall i \in \mathcal{I}$, $\mathcal{S}_i = \{\boldsymbol{\mu} \in \Lambda : i^\star(\boldsymbol{\mu}) = i\}$ is open and its complementary $\Lambda \setminus \mathcal{S}_i$ is a finite union of convex set. Namely, a finite collection $\mathcal{J}_i$ of convex set $\mathcal{C}_j^i$ s.t. $\Lambda \setminus \mathcal{S}_i = \cup_{j \in \mathcal{J}_i} \mathcal{C}_j^i$

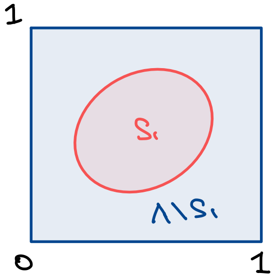
Example: BAI for unstructured bandit

Here $\Lambda = \{\boldsymbol{\mu} \in (0,1)^K : \exists i \in [K] \text{ s.t. } \mu_i > \mu_k, \forall k \neq i\}$, and for each $i \in [K]$, $\mathcal{S}_i = \{\boldsymbol{\mu} \in \Lambda : \mu_i > \mu_k, \forall k \neq i\}$
We can see that $\Lambda \setminus \mathcal{S}_i = \cup_{j \neq i} \mathcal{C}_j^i$, where $\mathcal{C}_j^i = \{\boldsymbol{\lambda} \in \Lambda : \lambda_j > \lambda_i\}$ is a convex set $\forall j \neq i$

With Assumption 1, we define $f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \inf_{\boldsymbol{\lambda} \in \mathcal{C}_j^i} \sum_{k=1}^K \omega_k d(\mu_k, \lambda_k)$ for any $(\omega, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$ and $j \in \mathcal{J}_i$, where $\mathring{\Sigma}$ is the interior of Σ

A counterexample for Assumption 1

Though most pure exploration and structures satisfy **Assumption 1**, it may not hold for an arbitrary parameter set. For example,



Envelop theorem

Proposition 1.

Let $i \in \mathcal{I}$, $j \in \mathcal{J}_i$. Define for all $(\boldsymbol{\omega}, \boldsymbol{\mu}) \in \Sigma \times \mathcal{S}_i$,

$$\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})} = \arg \min_{\boldsymbol{\lambda} \in \mathrm{cl}(\mathcal{C}_j^i)} \sum_k \omega_k d(\mu_k, \lambda_k), \tag{2}$$

where $\mathrm{cl}(\mathcal{C}_j^i)$ is the closure of $\mathcal{C}_j^i$. Then under Assumption 1, $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$ is unique for all $(\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$. In addition, f_j is continuously differentiable on $\mathring{\Sigma} \times \mathcal{S}_i$, and $\forall (\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$,

$$\nabla_\omega f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_k d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k) \boldsymbol{e}_k, \tag{3}$$

Envelop theorem

Proposition 1.

Let $i \in \mathcal{I}$, $j \in \mathcal{J}_i$. Define for all $(\boldsymbol{\omega}, \boldsymbol{\mu}) \in \Sigma \times \mathcal{S}_i$,

$$\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})} = \arg \min_{\boldsymbol{\lambda} \in \mathrm{cl}(\mathcal{C}_j^i)} \sum_k \omega_k d(\mu_k, \lambda_k), \tag{2}$$

where $\mathrm{cl}(\mathcal{C}_j^i)$ is the closure of $\mathcal{C}_j^i$. Then under Assumption 1, $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$ is unique for all $(\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$. In addition, f_j is continuously differentiable on $\mathring{\Sigma} \times \mathcal{S}_i$, and $\forall (\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$,

$$\nabla_\omega f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_k d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k) \boldsymbol{e}_k, \tag{3}$$

Example (Unstructured BAI)

$$\begin{aligned} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) &= \omega_j d(\mu_j, \tfrac{\omega_{i^\star}\mu_{i^\star} + \omega_j \mu_j}{\omega_{i^\star} + \omega_j}) + \omega_{i^\star} d(\mu_{i^\star}, \tfrac{\omega_{i^\star}\mu_{i^\star} + \omega_j \mu_j}{\omega_{i^\star} + \omega_j}) \\ \nabla_\omega f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) &= d(\mu_j, \tfrac{\omega_{i^\star}\mu_{i^\star} + \omega_j \mu_j}{\omega_{i^\star} + \omega_j}) \boldsymbol{e}_j + d(\mu_{i^\star}, \tfrac{\omega_{i^\star}\mu_{i^\star} + \omega_j \mu_j}{\omega_{i^\star} + \omega_j}) \boldsymbol{e}_{i^\star} \end{aligned}$$

Envelop theorem

Proposition 1.

Let $i \in \mathcal{I}$, $j \in \mathcal{J}_i$. Define for all $(\boldsymbol{\omega}, \boldsymbol{\mu}) \in \Sigma \times \mathcal{S}_i$,

$$\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})} = \arg \min_{\boldsymbol{\lambda} \in \mathrm{cl}(\mathcal{C}_j^i)} \sum_k \omega_k d(\mu_k, \lambda_k), \tag{2}$$

where $\mathrm{cl}(\mathcal{C}_j^i)$ is the closure of $\mathcal{C}_j^i$. Then under Assumption 1, $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$ is unique for all $(\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$. In addition, f_j is continuously differentiable on $\mathring{\Sigma} \times \mathcal{S}_i$, and $\forall (\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$,

$$\nabla_\omega f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_k d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k) \boldsymbol{e}_k, \tag{3}$$

Hence,

- Once we can solve (2), we have $\nabla_\omega f_j$ and f_j

Envelop theorem

Proposition 1.

Let $i \in \mathcal{I}$, $j \in \mathcal{J}_i$. Define for all $(\boldsymbol{\omega}, \boldsymbol{\mu}) \in \Sigma \times \mathcal{S}_i$,

$$\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})} = \arg \min_{\boldsymbol{\lambda} \in \mathrm{cl}(\mathcal{C}_j^i)} \sum_k \omega_k d(\mu_k, \lambda_k), \tag{2}$$

where $\mathrm{cl}(\mathcal{C}_j^i)$ is the closure of $\mathcal{C}_j^i$. Then under Assumption 1, $\overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}$ is unique for all $(\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$. In addition, f_j is continuously differentiable on $\mathring{\Sigma} \times \mathcal{S}_i$, and $\forall (\boldsymbol{\omega}, \boldsymbol{\mu}) \in \mathring{\Sigma} \times \mathcal{S}_i$,

$$\nabla_\omega f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) = \sum_k d(\mu_k, \overline{\boldsymbol{\lambda}_j(\boldsymbol{\omega}, \boldsymbol{\mu})}_k) \boldsymbol{e}_k, \tag{3}$$

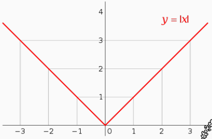
Hence,

- Once we can solve (2), we have $\nabla_\omega f_j$ and f_j
- $F_{\boldsymbol{\mu}} = \min_{j \in \mathcal{J}} f_j(\cdot, \boldsymbol{\mu})$ is the minimum of finite set of smooth

Non-smooth point may cause infinite curvature

Recall that $C_\psi(\mathcal{K}) = \sup_{\substack{\boldsymbol{x},\boldsymbol{z}\in\mathcal{K} \\ \alpha\in(0,1] \\ \boldsymbol{y}=\boldsymbol{x}+\alpha(\boldsymbol{z}-\boldsymbol{x})}} \min_{h\in\partial\psi(\boldsymbol{x})} \frac{1}{\alpha^2}\left[\psi(\boldsymbol{x}) - \psi(\boldsymbol{y}) + \langle \boldsymbol{y}-\boldsymbol{x}, h\rangle\right]$

Suppose $\psi(x) = |x|$ over $\mathcal{K} = [-1, 1]$. Let $x \in (0, \frac{1}{2}), z = 1 - x$ and $\alpha = 2x$. By definition, $\partial\psi(x) = \{1\}$

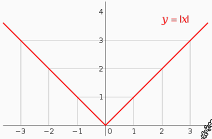


Non-smooth point may cause infinite curvature

Recall that $C_\psi(\mathcal{K}) = \sup_{\substack{\boldsymbol{x},\boldsymbol{z}\in\mathcal{K} \\ \alpha\in(0,1] \\ \boldsymbol{y}=\boldsymbol{x}+\alpha(\boldsymbol{z}-\boldsymbol{x})}} \min_{h\in\partial\psi(\boldsymbol{x})} \frac{1}{\alpha^2}\left[\psi(\boldsymbol{x}) - \psi(\boldsymbol{y}) + \langle \boldsymbol{y}-\boldsymbol{x}, h\rangle\right]$

Suppose $\psi(x) = |x|$ over $\mathcal{K} = [-1, 1]$. Let $x \in (0, \frac{1}{2}), z = 1 - x$ and $\alpha = 2x$. By definition, $\partial\psi(x) = \{1\}$

$$
\begin{aligned}
C_\psi(\mathcal{K}) &\geq \min_{h\in\partial\psi(\boldsymbol{x})} \frac{1}{\alpha^2}\left[\psi(\boldsymbol{x}) - \psi(\boldsymbol{y}) + \langle \boldsymbol{y}-\boldsymbol{x}, h\rangle\right] \\
&= \frac{1}{4x^2}(|-x| - |x| + 2x) = \frac{1}{2x}.
\end{aligned}
$$



Hence, $C_\psi(\mathcal{K}) = \infty$

Existing FW works for non-smooth functions

From the last example, $\partial\psi(x)$ is a discontinuous set-function, and hence fails to provide a proper bound across the non-smooth point.

Existing FW works for non-smooth functions

From the last example, $\partial\psi(x)$ is a discontinuous set-function, and hence fails to provide a proper bound across the non-smooth point. Existing FW-based works on non-smooth functions are:

- smoothing the objective function [FKM05, HK12]
- enlarging the set of differential [RCS19, CL18] (collect the set of the subdifferentials in the neighborhood)

Existing FW works for non-smooth functions

From the last example, $\partial\psi(x)$ is a discontinuous set-function, and hence fails to provide a proper bound across the non-smooth point. Existing FW-based works on non-smooth functions are:

- smoothing the objective function [FKM05, HK12]
- enlarging the set of differential [RCS19, CL18] (collect the set of the subdifferentials in the neighborhood)

Nevertheless, they are all too **computational expensive** to solve in our problem

r-subdifferential space

Recall that $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \min_{j \in \mathcal{J}_i} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$ for $\boldsymbol{\mu} \in \mathcal{S}_i$, we define

$$H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r) = \operatorname{cov}\left\{\nabla_{\omega} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) : j \in \mathcal{J}_i, f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) < F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) + r\right\}, \forall r \in \mathbb{R}_+,$$

where $\operatorname{cov}(S)$ is the convex hull of a set S.

r-subdifferential space

Recall that $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \min_{j \in \mathcal{J}_i} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$ for $\boldsymbol{\mu} \in \mathcal{S}_i$, we define

$$H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r) = \operatorname{cov}\left\{\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) : j \in \mathcal{J}_i, f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) < F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) + r\right\}, \forall r \in \mathbb{R}_+,$$

where $\operatorname{cov}(S)$ is the convex hull of a set S.
The advantages include:

- $H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r)$ is a continuous set-mapping
- $H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r)$ is very easy to be computed

r-subdifferential space

Recall that $F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \min_{j \in \mathcal{J}_i} f_j(\boldsymbol{\omega}, \boldsymbol{\mu})$ for $\boldsymbol{\mu} \in \mathcal{S}_i$, we define

$$H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r) = \operatorname{cov}\left\{\nabla_{\boldsymbol{\omega}} f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) : j \in \mathcal{J}_i, f_j(\boldsymbol{\omega}, \boldsymbol{\mu}) < F_{\boldsymbol{\mu}}(\boldsymbol{\omega}) + r\right\}, \forall r \in \mathbb{R}_+,$$

where $\operatorname{cov}(S)$ is the convex hull of a set S.
The advantages include:

- $H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r)$ is a continuous set-mapping
- $H_{F_{\boldsymbol{\mu}}}(\boldsymbol{\omega}, r)$ is very easy to be computed

Our modified FW update is

$$\begin{cases} \boldsymbol{z}(t+1) \leftarrow \operatorname{argmax}_{\boldsymbol{z} \in \boldsymbol{\Sigma}} \min_{h \in H_{F_{\boldsymbol{\mu}}(\boldsymbol{x}(t), r_t)}} \langle \boldsymbol{z} - \boldsymbol{x}(t), h \rangle, \\ \boldsymbol{x}(t+1) \leftarrow \frac{t}{t+1} \boldsymbol{x}(t) + \frac{1}{t+1} \boldsymbol{z}(t+1) \end{cases}$$

FWS

Input: Confidence level δ, sequence $\{r_t\}_{t\geq 1}$

Initialization: Sample each arm once and update $\boldsymbol{\omega}(K), \boldsymbol{x}(K) = (\frac{1}{K}, \ldots, \frac{1}{K})$, and $\hat{\boldsymbol{\mu}}(K)$
$t \leftarrow$K

While $tF_{\hat{\boldsymbol{\mu}}(t)}(\boldsymbol{\omega}(\boldsymbol{t}) < \beta(\delta, t)$ ←Stopping criteria or $\hat{\boldsymbol{\mu}}(t-1) \notin \Lambda$

IF $\sqrt{\lfloor t/K \rfloor} \in \mathbb{N}$ or $\hat{\boldsymbol{\mu}}(t-1) \notin \Lambda$, (Forced exploration) $\boldsymbol{z}(t) \leftarrow (\frac{1}{K}, \ldots, \frac{1}{K})$
Else, (FW update)

$$\boldsymbol{z}(t) \leftarrow \underset{z \in \boldsymbol{\Sigma}}{\operatorname{argmax}} \min_{h \in H_{F_{\hat{\boldsymbol{\mu}}(t-1)}}(\boldsymbol{x}(t-1), r_t)} \langle \boldsymbol{z} - \boldsymbol{x}(t-1), h \rangle$$

Update $\boldsymbol{x}(t) \leftarrow \frac{t-1}{t}\boldsymbol{x}(t-1) + \frac{1}{t}\boldsymbol{z}(t)$

Sample $A_t \leftarrow \operatorname{argmax}_k \boldsymbol{x}_k(t)/\boldsymbol{\omega_k}(t-1)$ (ties broken arbitrarily)
Update $\boldsymbol{\omega}(t)$ and $\hat{\boldsymbol{\mu}}(t)$

Output: $i^\star(\hat{\boldsymbol{\mu}}(t))$

Theoretical Results

Assumption 2

For $\gamma \in (0, \frac{1}{K})$, let $\Sigma_\gamma = \{\boldsymbol{\omega} \in \Sigma : \min_k \omega_k \geq \gamma\}$

Assumption 2.

For all $\boldsymbol{\mu} \in \Lambda$, there exist $L, D > 0$ s.t

(i). $\forall j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}, \boldsymbol{\omega} \in \Sigma$, $\|\nabla_\omega f_j(\boldsymbol{\omega}, \boldsymbol{\mu})\|_\infty \leq L$

(ii). $\forall \gamma \in (0, 1/K)$ and $\forall j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$, $C_{f_j(\cdot,\boldsymbol{\mu})}(\Sigma_\gamma) \leq \frac{D}{\gamma}$

Assumption 2

For $\gamma \in (0, \frac{1}{K})$, let $\Sigma_\gamma = \{\boldsymbol{\omega} \in \Sigma : \min_k \omega_k \geq \gamma\}$

Assumption 2.

For all $\boldsymbol{\mu} \in \Lambda$, there exist $L, D > 0$ s.t

(i). $\forall j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}, \boldsymbol{\omega} \in \Sigma$, $\left\|\nabla_\omega f_j(\boldsymbol{\omega}, \boldsymbol{\mu})\right\|_\infty \leq L$

(ii). $\forall \gamma \in (0, 1/K)$ and $\forall j \in \mathcal{J}_{i^\star(\boldsymbol{\mu})}$, $C_{f_j(\cdot, \boldsymbol{\mu})}(\Sigma_\gamma) \leq \frac{D}{\gamma}$

We also provide a generic method to verify Assumption 2.

Asymptotic optimality of FWS

Theorem 1

Consider the `FWS` algorithm with a sequence $\{r_t\}_{t\geq 1}$ strictly positive reals satisfying

- $\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} r_s = 0$,
- $\lim_{t\to\infty} t r_t = \infty$.

Under Assumptions 1, 2., the algorithm terminates in finite time a.s. and is δ-PAC. Its sample complexity τ satisfies: $\forall \boldsymbol{\mu} \in \Lambda$,

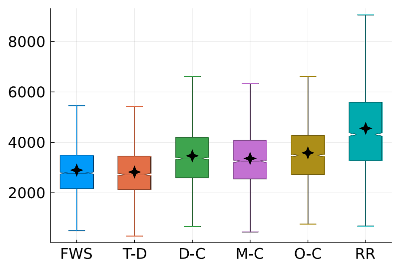
$$\mathbb{P}_{\boldsymbol{\mu}}\left[\overline{\lim_{\delta\to 0}} \frac{\tau}{\log(1/\delta)} \leq T^\star(\boldsymbol{\mu})\right] = 1, \text{ and } \overline{\lim_{\delta\to 0}} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau]}{\log(1/\delta)} \leq T^\star(\boldsymbol{\mu}).$$

With further assumptions, we can provide **non-asymptotic** upper bound for $\mathbb{E}_{\boldsymbol{\mu}}[\tau]$

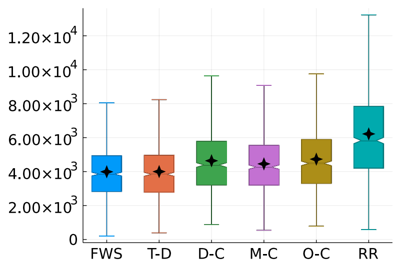
Numerical Results

Experiment (i) Unstructured bandits

Averaged sample complexity at $\delta = 0.01$



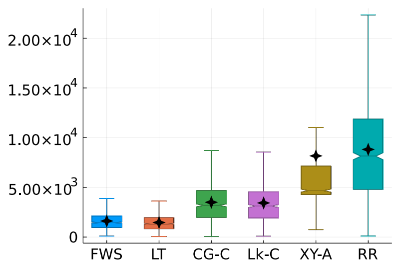
Bernoulli



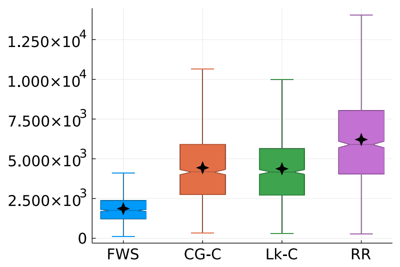
Gaussian

Experiment (ii) Linear bandits

Averaged sample complexity at $\delta = 0.01$



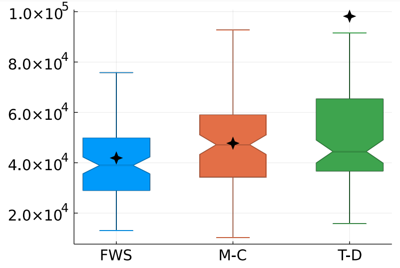
BAI



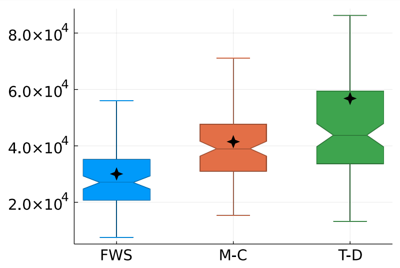
ThresholdingBandit

Experiment (iii) Lipschitz bandits

Averaged Sample complexity at $\delta = 0.01$



Experiment 1



Experiment 2

This is the first result for Lipschitz bandits in literatures

Related work and conclusion

Related works:

- TaS [GK16]: Compute and track the optimal allocation
- LMA [Mén19]: Apply mirror ascent to update $\boldsymbol{x}(t)$
- Gamification [DMSV20, Sha21, JMKK21]: Use 2 player game to reach $\omega^\star(\mu)$

Unclear to extend the above approaches to general structures

Related work and conclusion

Related works:

- TaS [GK16]: Compute and track the optimal allocation
- LMA [Mén19]: Apply mirror ascent to update $\boldsymbol{x}(t)$
- Gamification [DMSV20, Sha21, JMKK21]: Use 2 player game to reach $\boldsymbol{\omega}^\star(\boldsymbol{\mu})$

Unclear to extend the above approaches to general structures

Conclusion:

- FWS is computationally and statistically efficient for general pure exploration problems
- Theoretically, FWS matchs the instance-specific lower bounds
- Numerically, FWS is competitive to the state-of-art algorithms in structured bandits

Reference

- Edward Cheung and Yuying Li, *Solving separable nonsmooth problems using frank-wolfe with uniform affine approximations*, Proc. of IJCAI, 2018.
- Rémy Degenne, Wouter M Koolen, and Pierre Ménard, *Non-asymptotic pure exploration by solving games*, Proc. of NeurIPS, 2019.
- Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko, *Gamification of pure exploration for linear bandits*, Proc. of ICML, 2020.
- Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan, *Online convex optimization in the bandit setting: gradient descent without a gradient*, Proc. of SODA, 2005.
- Aurélien Garivier and Emilie Kaufmann, *Optimal best arm identification with fixed confidence*, Proc. of COLT, 2016.