Best-of-both-worlds FTRL for Linear Bandits:

The Interplay of Information Acquisition, Corruption, and Geometry

Canzhe Zhao

Department of Mathematics, National University of Singapore

Joint work with

Vincent Y. F. Tan

Department of Mathematics
Department of Electrical and Computer Engineering
National University of Singapore

STOC 2027 submission video

Linear Bandits

- **Known** compact action set $\mathcal{A} \subset \mathbb{R}^d$.
- Interaction over T rounds.

Linear Bandits

- **Known** compact action set $\mathcal{A} \subset \mathbb{R}^d$.
- Interaction over T rounds.
- In each round $t \in [T]$:
  - The environment chooses an **unknown** loss vector θ_t.

Linear Bandits

- **Known** compact action set $\mathcal{A} \subset \mathbb{R}^d$.
- Interaction over T rounds.
- In each round $t \in [T]$:
  - The environment chooses an **unknown** loss vector θ_t.
  - The learner chooses $A_t \in \mathcal{A}$ and observes only a scalar Y_t:

$$\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}, \theta_t, A_t] = A_t^\top \theta_t, \qquad |Y_t| \leq 1 \,.$$

Linear Bandits

- **Known** compact action set $\mathcal{A} \subset \mathbb{R}^d$.
- Interaction over T rounds.
- In each round $t \in [T]$:
  - The environment chooses an **unknown** loss vector θ_t.
  - The learner chooses $A_t \in \mathcal{A}$ and observes only a scalar Y_t:

$$\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}, \theta_t, A_t] = A_t^\top \theta_t, \qquad |Y_t| \leq 1\,.$$

- **Learning objective:** Minimize regret against the best fixed action.

$$\mathrm{Reg}_T = \sup_{u \in \mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T} (A_t - u)^\top \theta_t\right].$$

One observation contains information about many actions.

Why Linear Bandits Matter

Efficient learning sequential decision making with large, or even infinite, action sets [BCBK12]

- In practice, the number of actions may be exponentially large (*e.g.*, in recommender systems) or infinite (*i.e.*, in financial scenarios).
- However, each action has a d-dimensional feature.
- Feedback reveals an unknown linear preference or cost.

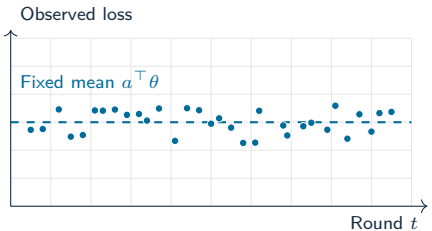
The statistical benefit

Exploit shared structure in dimension d, instead of learning every action independently.

Stochastic vs. Adversarial Regimes

For a fixed action a, observations have mean $a^\top \theta_t$.

Stochastic regime

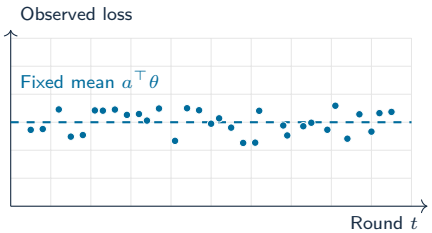


$\theta_t = \theta$ for all t: a fixed unknown loss vector.

Stochastic vs. Adversarial Regimes

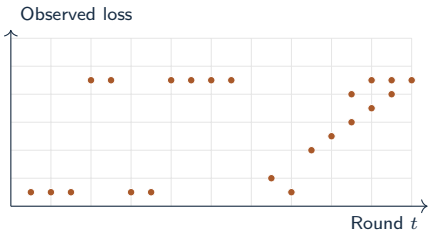
For a fixed action a, observations have mean $a^\top \theta_t$.

Stochastic regime



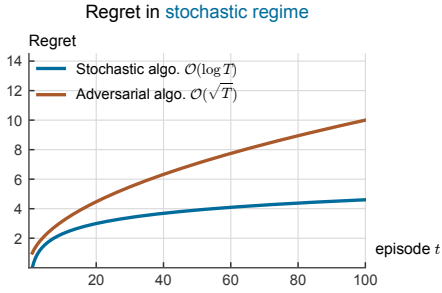
$\theta_t = \theta$ for all t: a fixed unknown loss vector.

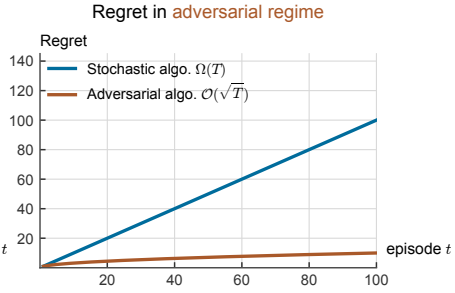
Adversarial regime



θ_t may change with t and depend on the past.

Disadvantages of Regime-specific Algorithms

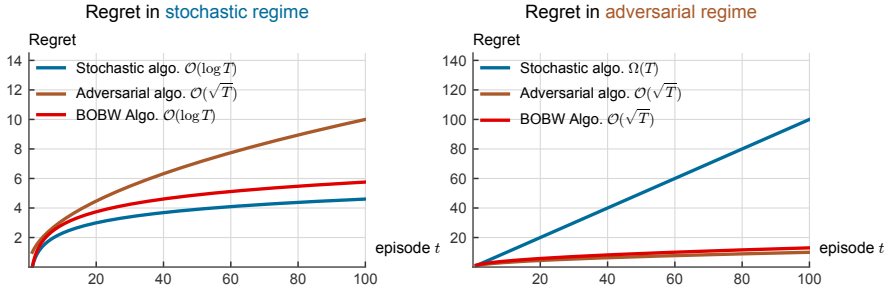




Stochastic algorithms learn quickly under stable losses, but can incur linear regret when losses change.

Adversarial algorithms remain robust, but their $\mathcal{O}(\sqrt{T})$ worst-case guarantee is too conservative compared with the $\mathcal{O}(\log T)$ rate under stable losses.

Advantages of Best-of-both-worlds Algorithms



Simultaneously guarantee $\mathcal{O}(\log T)$-type regret in the stochastic regime and $\mathcal{O}(\sqrt{T})$-type regret in the adversarial regime. [BS12]

Corrupted Stochastic Regime

- A fixed stochastic loss vector θ is perturbed by corruption c_t^{vec}.
- The total corruption budget measures how far the regime departs from that fixed instance.
- Varying C from 0 to order T interpolates between the stochastic and adversarial regimes.

Corrupted stochastic regime

$$\theta_t = \theta + c_t^{\mathrm{vec}}, \qquad \sum_{t=1}^{T} \kappa_t \le C, \qquad \kappa_t = \sup_{a \in \mathcal{A}} |a^\top c_t^{\mathrm{vec}}| .$$

The budget C is unknown to the learner.

Information Acquisition and Instance-dependent Optimality

For finite $\mathcal{A}$ and a unique optimum a^*, denote by $\Delta(a) = (a - a^*)^\top \theta$ the suboptimality gap of a.

Definition (Information-acquisition complexity [LS17])

$$c^* = \inf_{n \geq 0} \sum_{a \in \mathcal{A}} n_a \Delta(a) \quad \text{s.t.} \quad H_n = \sum_a n_a a a^\top \succ 0, \quad \|a\|^2_{H_n^{-1}} \leq \frac{\Delta(a)^2}{2} \ (a \neq a^*) \,.$$

An allocation n collects enough information to distinguish suboptimal actions, while minimizing its regret cost.

Information Acquisition and Instance-dependent Optimality

For finite $\mathcal{A}$ and a unique optimum a^*, denote by $\Delta(a) = (a - a^*)^\top \theta$ the suboptimality gap of a.

Definition (Information-acquisition complexity [LS17])

$$c^* = \inf_{n \geq 0} \sum_{a \in \mathcal{A}} n_a \Delta(a) \quad \text{s.t.} \quad H_n = \sum_a n_a a a^\top \succ 0, \quad \|a\|^2_{H_n^{-1}} \leq \frac{\Delta(a)^2}{2} \ (a \neq a^*) \,.$$

An allocation n collects enough information to distinguish suboptimal actions, while minimizing its regret cost.

Theorem (Stochastic lower bound [LS17, LLW+21])

For Gaussian or normalized Bernoulli feedback, every consistent learner satisfies

$$\liminf_{T \to \infty} \frac{\mathrm{Reg}_T}{\log T} \geq \Omega(c^*) \,.$$

[LS17] show c^* can be much smaller than $1/\Delta_{\min}$!

Progress on BOBW Linear Bandits

Approach	Stochastic leading term	Main limitation
Detect and switch [LLW+21]	$\mathcal{O}(c^* \log T \log(KT))$	Explicit regime tests
Purely FTRL-based [KZL23, IT23a, DWZ23] [IT23b, ITH24]	$\mathcal{O}\left(\frac{\mathrm{poly}(\log T)}{\Delta_{\min}}\right)$	Does not capture c^*

- Both lines retain $\sqrt{T}$-type adversarial guarantees.
- Follow-the-regularized-leader (FTRL) methods also adapt to corruption, typically with

$$\mathcal{O}\left(L_T + \sqrt{CL_T}\right), \qquad L_T = \frac{\mathrm{poly}(\log T)}{\Delta_{\min}}.$$

Questions 1 and 2: Information and Corruption

Q1: Information acquisition through FTRL

An open problem posed in [LLW+21, KZL23, IT23a, ZIL25]: can a **purely FTRL-based algorithm** (without a detect-and-switch scheme) achieve both guarantees below?

$$\mathcal{O}(c^* \log T)\text{-type stochastic regret} \quad \text{and} \quad \mathcal{O}(\sqrt{T})\text{-type adversarial regret}.$$

Questions 1 and 2: Information and Corruption

Q1: Information acquisition through FTRL

An open problem posed in [LLW+21, KZL23, IT23a, ZIL25]: can a **purely FTRL-based algorithm** (without a detect-and-switch scheme) achieve both guarantees below?

$$\mathcal{O}(c^* \log T)\text{-type stochastic regret} \quad \text{and} \quad \mathcal{O}(\sqrt{T})\text{-type adversarial regret}.$$

Q2: The cost of corruption

In the corrupted stochastic regime, can the same algorithm retain the $c^* \log T$ main term with the following additional cost?

$$J\sqrt{C} \qquad \text{with } J \text{ independent of } T.$$

More strongly, can we take $J = \mathcal{O}(\sqrt{c^*})$?

BOBW Linear Bandits with Infinite-action Sets

For finite $\mathcal{A}$ and a unique optimum,

$$\Delta_{\min} = \min_{a \neq a^*} \Delta(a) > 0\,.$$

For a compact infinite set, suboptimal actions can approach a^* arbitrarily close:

$$a_n \longrightarrow a^*, \qquad \Delta(a_n) \downarrow 0\,.$$

The limitation of previous BOBW guarantees

A bound proportional to $1/\Delta_{\min}$ becomes vacuous when $\inf_{a \neq a^*} \Delta(a) = 0$.

- Some methods require finitely many actions algorithmically. [KZL23, ITH24]
- Others allow infinite actions in the adversarial regime, but need a positive gap in the stochastic regime. [IT23a, IT23b]

Question 3: What Role Does Geometry Play?

Q3: Information acquisition and corruption on compact action sets

- On compact action sets, is there an analogue of c^* yielding instance-dependent $\mathcal{O}(\log T)$ regret within a BOBW algorithm?
- When can we also obtain a $\sqrt{C}$ corruption term, up to logarithmic factors? Can this hold on every compact instance?

Q1 and Q2 (Part 1): IA-FTRL Guarantees

Theorem 1.1 (Formal statement: Theorem 2.1)

For finite $\mathcal{A}$ ($K = |\mathcal{A}|$), our **Information Acquisition FTRL (IA-FTRL)** algorithm simultaneously guarantees:

- **Adversarial regime:**

$$\mathrm{Reg}_T = \mathcal{O}\Big(\sqrt{dT \log K} + d \log K\Big) .$$

Q1 and Q2 (Part 1): IA-FTRL Guarantees

Theorem 1.1 (Formal statement: Theorem 2.1)

For finite $\mathcal{A}$ ($K = |\mathcal{A}|$), our **Information Acquisition FTRL (IA-FTRL)** algorithm simultaneously guarantees:

- **Adversarial regime:**

$$\mathrm{Reg}_T = \mathcal{O}\Big(\sqrt{dT \log K} + d \log K \Big) .$$

- **Corrupted stochastic regime:**

$$\mathrm{Reg}_T = \mathcal{O}\Big(c^* \log K \log T + M_{\mathcal{A},\theta} + J_{d,K,\Delta_{\min}} \sqrt{C} \Big) ,$$

where $J_{d,K,\Delta_{\min}} = \mathcal{O}\Big(\sqrt{d \log K}\, \Delta_{\min}^{-128/255} \Big)$ and $M_{\mathcal{A},\theta} < \infty$ is independent of T, C.

- **Q1: yes, with an information-acquisition leading term.**
- **Q2 (Part 1): a T-independent coefficient for the $\sqrt{C}$ term.**

Q2 (Part 2): The Cost of Corruption

Information cost can stay bounded while corruption sensitivity grows.

Consider a fixed action set and two candidate instances:

$$\mathcal{A} = \{-e_1, e_1, e_2\}, \qquad \theta_\sigma = (\sigma\Delta_{\min}/2, 1/2), \quad \sigma \in \{-1, +1\}\,.$$

Here $c^* = 4/(1+\Delta_{\min}) = \Theta(1)$, even as $\Delta_{\min} \downarrow 0$.

Q2 (Part 2): The Cost of Corruption

Information cost can stay bounded while corruption sensitivity grows.

Consider a fixed action set and two candidate instances:

$$\mathcal{A} = \{-e_1, e_1, e_2\}, \qquad \theta_\sigma = (\sigma\Delta_{\min}/2, 1/2), \quad \sigma \in \{-1, +1\}.$$

Here $c^* = 4/(1+\Delta_{\min}) = \Theta(1)$, even as $\Delta_{\min} \downarrow 0$.

Theorem 1.2 (Finite-action lower bound; formal: Theorem 2.2)

For every policy, some corrupted environment in this family forces

$$\mathrm{Reg}_T = \Omega\Big(\min\Big\{\sqrt{T}, \sqrt{C/\Delta_{\min}}\Big\}\Big),$$

for $C \geq \Delta_{\min}^{-1}$ and $T \geq \lceil \Delta_{\min}^{-2} \rceil$.

- **Q2 (Part 2): in general, an $\mathcal{O}(\sqrt{c^* C})$ corruption term cannot be guaranteed.**
  - **a stochastic instance may be cheap to learn because its actions share information, but that does not immediately imply that it is cheap to protect against corruption!**

Q3: Geometry, Information Acquisition, and Corruption

Let $\mathcal{K} = \mathrm{conv}(\mathcal{A})$ and $\mathcal{S} = \mathrm{conv}(\mathcal{A} \cup (-\mathcal{A}))$.

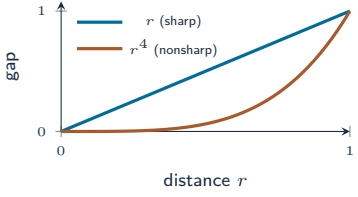
Definition 1.1 (Geometric Sharpness)

The loss gap is lower bounded linearly by distance:

$$\Delta(a) \geq \varsigma \|a - a^*\|_{\mathcal{S}} \quad \text{for all } a \in \mathcal{K}, \qquad \varsigma > 0\,,$$

where $\|\cdot\|_{\mathcal{S}}$ is the Minkowski norm induced by $\mathcal{S}$.

- Gaps may approach zero.
- However, they cannot vanish faster than distance.
- Small perturbations of θ preserve the optimum.



Q3 (Part 1): Centered Information Acquisition

Control **loss differences from** a^*, rather than individual action losses.

Definition 1.2 (Centered Information-acquisition Complexity (this work))

$$c^*_{\mathrm{ctr}} = \inf_{\substack{n \text{ finite} \\ n \geq 0}} \int_{\mathcal{A}} \Delta(a)\, n(da)\,,$$

$$\text{s.t.} \quad H_n = \int_{\mathcal{A}} aa^\top n(da) \succ 0, \qquad \| u - a^* \|^2_{H_n^{-1}} \leq \frac{\Delta(u)^2}{2} \quad (\forall u \in \mathcal{K})\,.$$

Q3 (Part 1): Centered Information Acquisition

Control **loss differences from** a^*, rather than individual action losses.

Definition 1.2 (Centered Information-acquisition Complexity (this work))

$$c^*_{\mathrm{ctr}} = \inf_{\substack{n \text{ finite} \\ n \geq 0}} \int_{\mathcal{A}} \Delta(a)\, n(da)\,,$$

$$\text{s.t.} \quad H_n = \int_{\mathcal{A}} aa^\top n(da) \succ 0, \qquad \| u - a^* \|^2_{H_n^{-1}} \leq \frac{\Delta(u)^2}{2} \quad (\forall u \in \mathcal{K})\,.$$

Compact stochastic instances (unique optimal action a^*)

Finite action sets
$c^* = c^*_{\mathrm{ctr}} < \infty$

Sharp instances
$c^*_{\mathrm{ctr}} < \infty$

Nonsharp instances: $c^*_{\mathrm{ctr}} = \infty$

Q3 (Part 1): BOBW under Sharpness

Theorem 1.3 (Formal statement: Theorem 3.1)

Our **Geometric Information Acquisition FTRL (GIA-FTRL)** algorithm simultaneously guarantees:

- **Adversarial regime:**

$$\mathrm{Reg}_T = \mathcal{O}(d\sqrt{T}\log T)\,.$$

Q3 (Part 1): BOBW under Sharpness

Theorem 1.3 (Formal statement: Theorem 3.1)

Our **Geometric Information Acquisition FTRL (GIA-FTRL)** algorithm simultaneously guarantees:

- **Adversarial regime:**

$$\mathrm{Reg}_T = \mathcal{O}(d\sqrt{T}\log T)\,.$$

- **Corrupted stochastic regime:** if the underlying instance satisfies the **sharpness condition**,

$$\mathrm{Reg}_T = \mathcal{O}\Big(d\,c^*_{\mathrm{ctr}}\log T + M_\theta + J_\theta\sqrt{C}\log(e+C)\Big)\,.$$

M_θ, J_θ are finite instance-dependent constants, independent of T, C.

- **Q3 (Part 1): yes under sharpness, even when** $\inf_{a\neq a^*}\Delta(a) = 0$.

Q3 (Part 1): Information and Corruption Costs

Lower bounds within the class of sharp compact instances.

Theorem 1.4 (Formal statement: Theorem 3.2)

- **Stochastic information cost.** There is a fixed-set class on which every consistent policy incurs
$$\Omega(\, c^*_{\mathrm{ctr}} \log T \,) .$$
- **Geometric sensitivity to corruption.** A different fixed-set class has $c^*_{\mathrm{ctr}} = \Theta(1)$ and $\varsigma = \Theta(\phi)$, yet forces
$$\Omega\Big(\min\Big\{\sqrt{T}, \sqrt{C/\phi}\Big\}\Big) ,$$
for every policy on a suitable corrupted instance, for sufficiently large C, T.

- **Q3 (Part 1): c^*_{ctr} captures information cost; corruption also depends on the sharpness modulus ς.**

Q3 (Part 2): Can the Guarantee Extend beyond Sharpness?

A fixed two-dimensional compact instance:

$$\mathcal{A} = \{(1/2 + s, s^4) : |s| \le 1/4\}, \qquad \theta_0 = (0, 1)\,.$$

Its unique optimum is $(1/2, 0)$; the gap is s^4, so sharpness fails.

Q3 (Part 2): Can the Guarantee Extend beyond Sharpness?

A fixed two-dimensional compact instance:

$$\mathcal{A} = \{(1/2 + s, s^4) : |s| \le 1/4\}, \qquad \theta_0 = (0, 1)\,.$$

Its unique optimum is $(1/2, 0)$; the gap is s^4, so sharpness fails.

Theorem 1.5 (Nonsharp lower bound; formal: Theorem 4.1)

For $T \ge 4$ and every policy, suitable deterministic corruption of budget at most C gives

$$\mathrm{Reg}_T = \Omega\left(\min\left\{\frac{C^{4/3}}{T^{1/3}},\ T^{1/3}\right\}\right).$$

At $C \asymp \sqrt{T}$, this is $\Omega(C^{2/3})$.

- **Q3 (Part 2): no. An $\mathcal{O}(\log T + \sqrt{C})$-type regret guarantee cannot hold on all compact instances.**

Our Idea: Optimal Exploration, T-independent FTRL Regret

Let q_t be the **FTRL update**, and p_t the **sampling distribution** obtained from q_t. Write $\ell_{t,a} = a^\top \theta_t$ and let H denote Shannon entropy.

$$\mathrm{Reg}_T \le \underbrace{\beta_1 \log K + \sum_{t=2}^{T} (\beta_t - \beta_{t-1}) \mathbb{E}[H(q_t)]}_{\text{penalty}} + \underbrace{\sum_{t=1}^{T} \mathbb{E}[D_t]}_{\text{stability}} + \underbrace{\sum_{t=1}^{T} \mathbb{E}\langle p_t - q_t, \ell_t \rangle}_{\text{exploration}},$$

$$\text{where } D_t = \langle q_t, \widehat{\ell}_t \rangle + \beta_t \log\left(\sum_a q_{t,a} e^{-\widehat{\ell}_{t,a}/\beta_t} \right).$$

Our Idea: Optimal Exploration, T-independent FTRL Regret

Let q_t be the **FTRL update**, and p_t the **sampling distribution** obtained from q_t. Write $\ell_{t,a} = a^\top \theta_t$ and let H denote Shannon entropy.

$$\mathrm{Reg}_T \le \underbrace{\beta_1 \log K + \sum_{t=2}^{T} (\beta_t - \beta_{t-1}) \mathbb{E}[H(q_t)]}_{\text{penalty}} + \underbrace{\sum_{t=1}^{T} \mathbb{E}[D_t]}_{\text{stability}} + \underbrace{\sum_{t=1}^{T} \mathbb{E}\langle p_t - q_t, \ell_t \rangle}_{\text{exploration}},$$

$$\text{where } D_t = \langle q_t, \widehat{\ell}_t \rangle + \beta_t \log\left(\sum_a q_{t,a} e^{-\widehat{\ell}_{t,a}/\beta_t} \right).$$

	Stochastic target	Adversarial target
Exploration	$\mathcal{O}(c^* \log K \log T)$	$\mathcal{O}(\sqrt{dT \log K})$
Penalty + stability	T-independent constants	$\mathcal{O}(\sqrt{dT \log K} + d \log K)$

IA-FTRL: Optimize Exploration around an FTRL Update

1. FTRL with Shannon entropy

$$q_t = \arg\min_q \left\{ \langle q, \widehat{L}_{t-1} \rangle + \beta_t \sum_a q_a \log q_a \right\}, \qquad q_{t,a} \propto e^{-\widehat{L}_{t-1,a}/\beta_t}.$$

IA-FTRL: Optimize Exploration around an FTRL Update

1. FTRL with Shannon entropy

$$q_t = \arg\min_{q} \left\{ \langle q, \widehat{L}_{t-1} \rangle + \beta_t \sum_a q_a \log q_a \right\}, \qquad q_{t,a} \propto e^{-\widehat{L}_{t-1,a}/\beta_t}.$$

2. Choose the cheapest admissible sampling distribution

$$\min_{p} \langle p, \widehat{\Delta}_t \rangle \quad \text{s.t.} \quad \begin{cases} p = (1-\gamma) q_t + w, \quad w \geq 0, \quad \sum_a w_a = \gamma \leq \Gamma_t, \\ \text{gap-calibrated magnitude and second-moment constraints,} \\ \text{an exponential stability constraint} \end{cases}.$$

IA-FTRL: Optimize Exploration around an FTRL Update

1. FTRL with Shannon entropy

$$q_t = \arg\min_q \left\{ \langle q, \widehat{L}_{t-1} \rangle + \beta_t \sum_a q_a \log q_a \right\}, \qquad q_{t,a} \propto e^{-\widehat{L}_{t-1,a}/\beta_t}.$$

2. Choose the cheapest admissible sampling distribution

$$\min_p \langle p, \widehat{\Delta}_t \rangle \quad \text{s.t.} \quad \begin{cases} p = (1-\gamma) q_t + w, \quad w \geq 0, \quad \sum_a w_a = \gamma \leq \Gamma_t, \\ \text{gap-calibrated magnitude and second-moment constraints,} \\ \text{an exponential stability constraint} \end{cases}.$$

3. Sample and update an unbiased estimate

$$A_t \sim p_t, \qquad \widehat{\theta}_t = M(p_t)^{-1} A_t Y_t, \qquad \widehat{\ell}_{t,a} = a^\top \widehat{\theta}_t,$$

$$\text{where } M(p) = \sum_a p_a a a^\top, \qquad \beta_t \asymp \sqrt{t}, \qquad \Gamma_t \asymp d/\beta_t.$$

Challenge 1: Unknown Gaps and Optimal Sampling Allocation

$$r_t \in \arg\min_a \widehat{L}_{t-1,a}, \qquad Z_{t,a} = \widehat{L}_{t-1,a} - \widehat{L}_{t-1,r_t}, \qquad \widehat{\Delta}_t(a) = \min\{2, Z_{t,a}/t\}\,.$$

Technique 1: empirical gap ratio control

Choose the sampling distribution p so that every $a \in \mathcal{A}$ satisfies:

- **Gap-calibrated magnitude constraint:**

$$\max_b \left|(a - r_t)^\top M(p)^{-1} b\right| \le \beta_t + \frac{Z_{t,a}}{B}\,.$$

- **Gap-calibrated second-moment constraint:**

$$\|a - r_t\|^2_{M(p)^{-1}} \le \beta_t + \frac{Z_{t,a}^2}{Bt}\,.$$

- Upper bound $\widehat{\Delta}_t(a)/\Delta(a)$ and $\Delta(a)/\widehat{\Delta}_t(a)$ with high probability ($a \neq a^*$).

$B = \Theta(1 + \log K)$. Good events: gap ratios $\widehat{\Delta}_t(a)/\Delta(a)$ and $\Delta(a)/\widehat{\Delta}_t(a)$ are upper bounded.

Challenge 1: Unknown Gaps and Optimal Sampling Allocation

Fix a feasible allocation n for c^*; write $c^n = \sum_a n_a \Delta(a)$.

Technique 2: compare with every feasible allocation

$$p_t^n = (1 - \gamma_t^n) q_t + \rho_t n_-, \qquad \gamma_t^n = \rho_t \sum_{a \neq a^*} n_a, \qquad \rho_t \asymp 1/t\,.$$

n_- removes the mass of n at a^*; the scale also adjusts for gap-estimation errors.

Challenge 1: Unknown Gaps and Optimal Sampling Allocation

Fix a feasible allocation n for c^*; write $c^n = \sum_a n_a \Delta(a)$.

Technique 2: compare with every feasible allocation

$$p_t^n = (1-\gamma_t^n) q_t + \rho_t n_-, \qquad \gamma_t^n = \rho_t \sum_{a \neq a^*} n_a, \qquad \rho_t \asymp 1/t\,.$$

n_- removes the mass of n at a^*; the scale also adjusts for gap-estimation errors.

On good events, p_t^n satisfies **all** exploration constraints.

$$\underbrace{\langle p_t, \widehat{\Delta}_t \rangle \leq \langle p_t^n, \widehat{\Delta}_t \rangle}_{\text{optimality of the sampling distribution } p_t} \quad \Longrightarrow \quad \text{exploration cost at scale } \frac{c^n}{t}\,.$$

- Sum $1/t$, then choose a fixed near-optimal allocation: $\mathcal{O}(c^* \log K \log T)$.

Challenge 2: FTRL Regret Must Not Dominate

- Both penalty and stability must stay below the $c^* \log T$ leading term.
- With $\beta_t \asymp \sqrt{t}$, the crude bound $H(q_t) \leq \log K$ gives $\sqrt{T}$ growth.
- Existing works [ITH22, KZL23] design data-dependent learning rate β_t, but need to use $\langle q_t, \Delta \rangle \geq \Delta_{\min}(1 - q_{t,a^*})$, introducing the coarse $1/\Delta_{\min}$ dependence.

Challenge 2: FTRL Regret Must Not Dominate

- Both penalty and stability must stay below the $c^* \log T$ leading term.
- With $\beta_t \asymp \sqrt{t}$, the crude bound $H(q_t) \le \log K$ gives $\sqrt{T}$ growth.
- Existing works [ITH22, KZL23] design data-dependent learning rate β_t, but need to use $\langle q_t, \Delta \rangle \ge \Delta_{\min}(1 - q_{t,a^*})$, introducing the coarse $1/\Delta_{\min}$ dependence.

Technique 3: constant regret control of the penalty term

On good events, suboptimal-action weights decay exponentially fast in $t^{1/4}$:

$$q_{t,a} \le e^{-X_{t,a}/\beta_t} \le e^{-ct^{1/4}}, \qquad X_{t,a} = \widehat{L}_{t-1,a} - \widehat{L}_{t-1,a^*} .$$

The penalty term in the stochastic regime is then bounded by a T-independent constant M_{pen}:

$$\beta_1 \log K + \sum_{t=2}^{T} (\beta_t - \beta_{t-1}) \mathbb{E}[H(q_t)] \le M_{\text{pen}} .$$

Challenge 2: FTRL Regret Must Not Dominate

Cheap exploration may produce $\widehat{\ell}_{t,a} \ll -\beta_t$, so the usual quadratic stability bound is unavailable.

Technique 4: constant stability regret with large loss estimates

We directly bound the one-round FTRL stability term D_t as

$$\mathbb{E}[D_t \mid \mathcal{F}_{t-1}] \le \beta_t \mathbb{E}_{b\sim p}\mathbb{E}_{a\sim q_t}\left[\chi\left(\frac{(a-\bar{a}_t)^\top M(p)^{-1} b}{\beta_t}\right)\right] =: \mathcal{C}_t(p)\,,$$

where $\bar{a}_t = \mathbb{E}_{q_t}[a]$ and $\chi(z) = e^z + e^{-z} - 2$.

The program also enforces $\mathcal{C}_t(p_t) \le 2d/\beta_t$ on every history.

Challenge 2: FTRL Regret Must Not Dominate

Cheap exploration may produce $\widehat{\ell}_{t,a} \ll -\beta_t$, so the usual quadratic stability bound is unavailable.

Technique 4: constant stability regret with large loss estimates

We directly bound the one-round FTRL stability term D_t as

$$\mathbb{E}[D_t \mid \mathcal{F}_{t-1}] \le \beta_t \mathbb{E}_{b\sim p}\mathbb{E}_{a\sim q_t}\left[\chi\left(\frac{(a-\bar{a}_t)^\top M(p)^{-1} b}{\beta_t}\right)\right] =: \mathcal{C}_t(p)\,,$$

where $\bar{a}_t = \mathbb{E}_{q_t}[a]$ and $\chi(z) = e^z + e^{-z} - 2$.

On good events, the magnitude constraint gives

$$\mathcal{C}_t(p_t) \lesssim \beta_t \sum_{a\neq a^*} \underbrace{q_{t,a}}_{\substack{\text{exponentially}\\ \text{small weight}}} \times \underbrace{e^{2X_{t,a}/(B\beta_t)}}_{\substack{\text{very negative}\\ \text{loss estimates}}} \le \beta_t \sum_{a\neq a^*} \underbrace{e^{-(1-2/B)X_{t,a}/\beta_t}}_{\substack{\text{large update absorbed by}\\ \text{exponentially small action weight}}} .$$

The program also enforces $\mathcal{C}_t(p_t) \le 2d/\beta_t$ on every history.

Challenge 2: FTRL Regret Must Not Dominate

Cheap exploration may produce $\widehat{\ell}_{t,a} \ll -\beta_t$, so the usual quadratic stability bound is unavailable.

Technique 4: constant stability regret with large loss estimates

We directly bound the one-round FTRL stability term D_t as

$$\mathbb{E}[D_t \mid \mathcal{F}_{t-1}] \le \beta_t \mathbb{E}_{b\sim p}\mathbb{E}_{a\sim q_t}\left[\chi\left(\frac{(a-\bar{a}_t)^\top M(p)^{-1} b}{\beta_t}\right)\right] =: \mathcal{C}_t(p)\,,$$

where $\bar{a}_t = \mathbb{E}_{q_t}[a]$ and $\chi(z) = e^z + e^{-z} - 2$.

On good events, the magnitude constraint gives

$$\mathcal{C}_t(p_t) \lesssim \beta_t \sum_{a\neq a^*} \underbrace{q_{t,a}}_{\substack{\text{exponentially}\\ \text{small weight}}} \times \underbrace{e^{2X_{t,a}/(B\beta_t)}}_{\substack{\text{very negative}\\ \text{loss estimates}}} \le \beta_t \sum_{a\neq a^*} \underbrace{e^{-(1-2/B)X_{t,a}/\beta_t}}_{\substack{\text{large update absorbed by}\\ \text{exponentially small action weight}}}\,.$$

The stability term in the stochastic regime is then bounded by a T-independent constant:

$$\sum_{t=1}^{T} \mathbb{E}[D_t] \le M_{\mathrm{stab}}\,.$$

The program also enforces $\mathcal{C}_t(p_t) \le 2d/\beta_t$ on every history.

The Same Constraints Preserve Adversarial Robustness

Let π be a G-optimal design and $\Gamma_t = 4d/\beta_t$.

A feasible distribution on every history

$$p_t^0 = (1 - \Gamma_t)q_t + \Gamma_t \pi\,,$$

satisfies all constraints, including $\mathcal{C}_t(p_t^0) \le 2d/\beta_t$.

- The sampling program is always feasible; each regret term has a worst-case bound.

Recall $\kappa_t = \sup_{a \in \mathcal{A}} |a^\top c_t^{\mathrm{vec}}|$ is the maximum loss perturbation at round t. The direct exploration cost obeys

$$2\sum_t \gamma_t \kappa_t \;\lesssim\; \sum_t \frac{\kappa_t}{\sqrt{t}} \;\lesssim\; \sqrt{C}, \qquad 0 \le \kappa_t \le 2, \quad \sum_t \kappa_t \le C\,.$$

Challenge 3: Infinite Action Sets

The finite-action argument used uniform decay over all $a \neq a^*$. When $\inf_{a \neq a^*} \Delta(a) = 0$, that argument breaks down.

FTRL with relative entropy

$$q_t = \arg\min_q \left\{ \int \widehat{L}_{t-1,a}\, q(da) + \beta_t \,\mathrm{KL}(q, \mu) \right\}, \qquad q_t(da) \propto e^{-\widehat{L}_{t-1,a}/\beta_t}\, \mu(da)\,.$$

The reference measure μ must provide:

1. Positive mass at the **optimal action of every sharp instance**.
2. Very little mass on suboptimal actions with tiny gaps.
3. Coverage near every action for the adversarial comparison.

Challenge 3: Infinite Action Sets

Technique 5: relative entropy with a nonuniform mixture reference measure

$$\mu = \sum\nolimits_j \pi_j \delta_{b_j} + \omega_\mu \mu_{16} \,.$$

Challenge 3: Infinite Action Sets

Technique 5: relative entropy with a nonuniform mixture reference measure

$$\mu = \sum_j \pi_j \delta_{b_j} + \omega_\mu \mu_{16} \,.$$

Discrete part: retain the optimal action a^*

- Rational directions give points b_j, including a^* under sharpness.
- **Atom:** positive mass π_j at a single point b_j; δ_{b_j} is the Dirac measure.
- **Downweight:** assign smaller π_j to new points near earlier ones.

Challenge 3: Infinite Action Sets

Technique 5: relative entropy with a nonuniform mixture reference measure

$$\mu = \sum_j \pi_j \delta_{b_j} + \omega_\mu \mu_{16} \, .$$

Discrete part: retain the optimal action a^*

- Rational directions give points b_j, including a^* under sharpness.
- **Atom:** positive mass π_j at a single point b_j; δ_{b_j} is the Dirac measure.
- **Downweight:** assign smaller π_j to new points near earlier ones.

Continuous part: retain coverage

- $\omega_\mu = 1 - \sum_j \pi_j > 0$.
- μ_{16} is the law of $\frac{1}{16} \sum_{i=1}^{16} U_i$, where $U_i \overset{\text{iid}}{\sim} \text{Unif}(\mathcal{K})$.
- A small average gap requires every sample to have a small gap.

Challenge 3: Infinite Action Sets

Technique 5: relative entropy with a nonuniform mixture reference measure

$$\mu = \sum\nolimits_j \pi_j \delta_{b_j} + \omega_\mu \mu_{16}\,.$$

Discrete part: retain the optimal action a^*

- Rational directions give points b_j, including a^* under sharpness.
- **Atom:** positive mass π_j at a single point b_j; δ_{b_j} is the Dirac measure.
- **Downweight:** assign smaller π_j to new points near earlier ones.

Continuous part: retain coverage

- $\omega_\mu = 1 - \sum_j \pi_j > 0$.
- μ_{16} is the law of $\frac{1}{16}\sum_{i=1}^{16} U_i$, where $U_i \overset{\text{iid}}{\sim} \mathrm{Unif}(\mathcal{K})$.
- A small average gap requires every sample to have a small gap.

$$\boxed{\mu\{a : 0 < \Delta(a) \le \varepsilon\} \lesssim \varepsilon^{16}}\,.$$

Near-optimal mass control gives T-independent penalty and stability constants.

Geometry-dependent Exploration Constraints

Let b_{r_t} be the prior-adjusted empirical reference and $r(a) = \|a - b_{r_t}\|_{\mathcal{S}}$.

Scale estimation budgets by distance

Define $Z_t(a) = [\widehat{L}_{t-1,a} - \widehat{L}_{t-1,b_{r_t}}]_+$.

- **Gap-calibrated magnitude budget:**

$$U_t(a) = \frac{\beta_t}{2} r(a) + \frac{Z_t(a)}{B}.$$

- **Gap-calibrated second-moment budget:**

$$W_t(a) = \frac{\beta_t}{4} r(a)^2 + \frac{Z_t(a)^2}{Bt}.$$

Geometry-dependent Exploration Constraints

Let b_{r_t} be the prior-adjusted empirical reference and $r(a) = \|a - b_{r_t}\|_{\mathcal{S}}$.

Scale estimation budgets by distance

Define $Z_t(a) = [\widehat{L}_{t-1,a} - \widehat{L}_{t-1,b_{r_t}}]_+$.

- **Gap-calibrated magnitude budget:**

$$U_t(a) = \frac{\beta_t}{2} r(a) + \frac{Z_t(a)}{B}.$$

- **Gap-calibrated second-moment budget:**

$$W_t(a) = \frac{\beta_t}{4} r(a)^2 + \frac{Z_t(a)^2}{Bt}.$$

Sharpness gives $\left\| \frac{a - a^*}{\Delta(a)} \right\|_{\mathcal{S}} \leq \frac{1}{\varsigma}$.

Allocation comparison still works even as $\Delta(a) \to 0$, with cost c^*_{ctr}.

Information Acquisition, Corruption, and Geometry

Question	Answer
Information acquisition	Pure FTRL attains a $c^* \log T$-type leading term while retaining adversarial robustness.
Corruption	A T-independent coefficient J of $\sqrt{C}$ is possible, but J can grow with $\Delta_{\min}^{-\frac{1}{2}}$ even when $c^* = \Theta(1)$.
Infinite action sets	Under sharpness, c^*_{ctr} gives $c^*_{\mathrm{ctr}} \log T$-type regret within a BOBW guarantee. Without sharpness, $\sqrt{C}$ robustness can fail.

Information Acquisition, Corruption, and Geometry

Question	Answer
Information acquisition	Pure FTRL attains a $c^* \log T$-type leading term while retaining adversarial robustness.
Corruption	A T-independent coefficient J of $\sqrt{C}$ is possible, but J can grow with $\Delta_{\min}^{-\frac{1}{2}}$ even when $c^* = \Theta(1)$.
Infinite action sets	Under sharpness, c^*_{ctr} gives $c^*_{\mathrm{ctr}} \log T$-type regret within a BOBW guarantee. Without sharpness, $\sqrt{C}$ robustness can fail.

Key idea: use optimal allocations for $c^* \log T$-type exploration, and ensure FTRL penalty and stability contribute only T-independent constants.

Thank you for your attention!

Selected References I

[BCBK12] Sébastien Bubeck, Nicolò Cesa-Bianchi, and Sham M. Kakade.
Towards minimax policies for online linear optimization with bandit feedback.
In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 41.1–41.14. PMLR, 2012.

[BS12] Sébastien Bubeck and Aleksandrs Slivkins.
The best of both worlds: Stochastic and adversarial bandits.
In *Conference on Learning Theory*, pages 42–1. JMLR Workshop and Conference Proceedings, 2012.

[DWZ23] Christoph Dann, Chen-Yu Wei, and Julian Zimmert.
A blackbox approach to best of both worlds in bandits and beyond.
In *The Thirty Sixth Annual Conference on Learning Theory, COLT 2023, 12-15 July 2023, Bangalore, India*, Proceedings of Machine Learning Research. PMLR, 2023.

[IT23a] Shinji Ito and Kei Takemura.
Best-of-three-worlds linear bandit algorithm with variance-adaptive regret bounds.
In *The Thirty Sixth Annual Conference on Learning Theory, COLT 2023, 12-15 July 2023, Bangalore, India*, Proceedings of Machine Learning Research. PMLR, 2023.

Selected References II

[IT23b] Shinji Ito and Kei Takemura.
An exploration-by-optimization approach to best of both worlds in linear bandits.
In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

[ITH22] Shinji Ito, Taira Tsuchiya, and Junya Honda.
Nearly optimal best-of-both-worlds algorithms for online learning with feedback graphs.
In *Advances in Neural Information Processing Systems*, 2022.

[ITH24] Shinji Ito, Taira Tsuchiya, and Junya Honda.
Adaptive learning rate for follow-the-regularized-leader: Competitive analysis and best-of-both-worlds.
In *The Thirty Seventh Annual Conference on Learning Theory*, pages 2522–2563. PMLR, 2024.

[KZL23] Fang Kong, Canzhe Zhao, and Shuai Li.
Best-of-three-worlds analysis for linear bandits with follow-the-regularized-leader algorithm.
In *The Thirty Sixth Annual Conference on Learning Theory, COLT 2023, 12-15 July 2023, Bangalore, India*, Proceedings of Machine Learning Research. PMLR, 2023.

[LLW+21] Chung-Wei Lee, Haipeng Luo, Chen-Yu Wei, Mengxiao Zhang, and Xiaojin Zhang.
Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously.
In *International Conference on Machine Learning*, pages 6142–6151. PMLR, 2021.

Selected References III

[LS17] Tor Lattimore and Csaba Szepesvári.
The end of optimism? An asymptotic analysis of finite-armed linear bandits.
In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 54 of *Proceedings of Machine Learning Research*, pages 728–737, 2017.
https://proceedings.mlr.press/v54/lattimore17a.html.

[ZIL25] Canzhe Zhao, Shinji Ito, and Shuai Li.
Heavy-tailed linear bandits: Adversarial robustness, best-of-both-worlds, and beyond.
arXiv preprint arXiv:2508.13679, 2025.