Lasso Bandit with Compatibility Condition on Optimal Arm

Min-hwan Oh

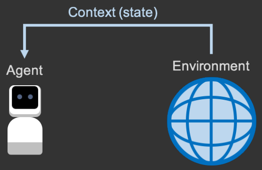
Seoul National University

Joint work with Harin Lee and Taehyun Hwang

Contextual Bandits Problem

For each round:

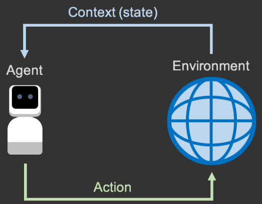
- Agent (decision maker) is presented with a context



Contextual Bandits Problem

For each round:

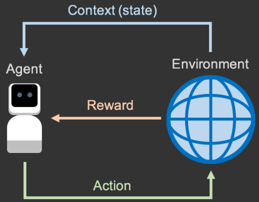
- Agent (decision maker) is presented with a context
- Agent chooses an action



Contextual Bandits Problem

For each round:

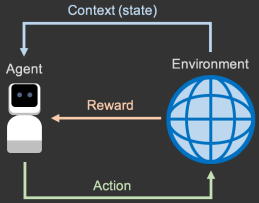
- Agent (decision maker) is presented with a context
- Agent chooses an action
- Agent observes reward (but only for chosen action)



Contextual Bandits Problem

For each round:

- Agent (decision maker) is presented with a context
- Agent chooses an action
- Agent observes reward
  (but only for chosen action)



Goal: Learn actions that maximize rewards

- Fundamental problem: How to efficiently use the experience?

Key Challenges of Contextual Bandits

Balancing exploration & exploitation

- Exploit: maximize reward given what is known
- Explore: collect more information for (potentially) higher reward

Key Challenges of Contextual Bandits

Balancing exploration & exploitation

- Exploit: maximize reward given what is known
- Explore: collect more information for (potentially) higher reward

Generalization

- May never see same context twice: use effectively
- Need to generalize across contexts

Key Challenges of Contextual Bandits

Balancing exploration & exploitation

- Exploit: maximize reward given what is known
- Explore: collect more information for (potentially) higher reward

Generalization

- May never see same context twice: use effectively
- Need to generalize across contexts

Statistical efficiency & computational efficiency

Key Challenges of High-Dimensional Contextual Bandits

Need to deal with high-dimensional context

- Context dimension is potentially larger than the time horizon
- Exploration duration cannot scale with ambient context dimension

However, the reward model is typically sparse

- Only small number of features are relevant w.r.t reward model.

But, this sparse structure is unknown!

Key challenge: How can we ensure statistical efficiency?

Sparse Linear Contextual Bandit

Stochastic linear contextual bandits

For each round $t = 1, \ldots, T$

1. Contexts $\{\mathbf{x}_{t,k} \in \mathbb{R}^d \mid k \in [K]\}$ drawn from (unknown) $\mathcal{P}_\mathcal{X}$
2. Agent selects an arm $a_t \in [K]$
3. Agent observes reward:

$$r_{t,a_t} = \underbrace{\mathbf{x}_{t,a_t}^\top \boldsymbol{\beta}^*}_{\text{expected reward}} + \eta_t$$

η_t sub-Gaussian noise with parameter σ

$\boldsymbol{\beta}^* \in \mathbb{R}^d$ unknown to agent

Sparse Linear Contextual Bandit

Stochastic linear contextual bandits

For each round $t = 1, ..., T$

1. Contexts $\{\mathbf{x}_{t,k} \in \mathbb{R}^d \mid k \in [K]\}$ drawn from (unknown) $\mathcal{P}_\mathcal{X}$
2. Agent selects an arm $a_t \in [K]$
3. Agent observes reward:

$$r_{t,a_t} = \underbrace{\mathbf{x}_{t,a_t}^\top \boldsymbol{\beta}^*}_{\text{expected reward}} + \eta_t$$

η_t sub-Gaussian noise with parameter σ

$\boldsymbol{\beta}^* \in \mathbb{R}^d$ unknown to agent

Sprase linear contextual bandits

- Context dimension is large ($d \gg 1$), even potentially $d > T$
- $\boldsymbol{\beta}^*$ is sparse, i.e., $\|\boldsymbol{\beta}^*\|_0 = s_0$ with $s_0 \ll d$

Sparse Linear Contextual Bandit (cont'd)

Optimal action at period t: $a_t^* = \arg\max_{k \in [K]} \mathbf{x}_{t,k}^\top \boldsymbol{\beta}^*$

Goal: Choose a policy $\pi = \{a_t : t = 1, 2, ...\}$ that minimizes the following cumulative regret

$$\text{Regret}_T(\pi) := \sum_{t=1}^{T} \underbrace{\mathbf{x}_{t,a_t^*}^\top \boldsymbol{\beta}^*}_{\text{optimal reward}} - \underbrace{\mathbf{x}_{t,a_t}^\top \boldsymbol{\beta}^*}_{\text{agent's reward}}$$

Maximizing cumulative reward $\equiv$ minimizing cumulative regret

Related Literature

Emerging body of work on sparse linear contextual bandit

- Multiple-parameter setting: each arm has its own underlying parameter (K parameters), and only one context vector is given.

  (Bastani and Bayati, 2020; Wang et al., 2018)

- Single-parameter setting: arms have one shared parameter, and K different contexts vectors are given.

  (Kim and Paik, 2019; Hao et al., 2020b; Oh et al., 2021; Li et al., 2021; Ariu et al., 2022; Chakraborty et al., 2023)

[1] Alternative form of regularity is required, e.g., minimum eigenvalue of $\mathbf{\Sigma}$ (Hao et al., 2020b), bounded sparse eigenvalue of $\mathbf{\Sigma}$ (Li et al., 2021; Chakraborty et al., 2023)

Related Literature

Emerging body of work on sparse linear contextual bandit

- Multiple-parameter setting: each arm has its own underlying parameter (K parameters), and only one context vector is given.

  (Bastani and Bayati, 2020; Wang et al., 2018)

- Single-parameter setting: arms have one shared parameter, and K different contexts vectors are given.

  (Kim and Paik, 2019; Hao et al., 2020b; Oh et al., 2021; Li et al., 2021; Ariu et al., 2022; Chakraborty et al., 2023)

To achieve regret bound that only depends logarithmically on d,

- Compatibility condition[1] on $\mathbf{\Sigma} := \frac{1}{K}\mathbb{E}[\sum_{k\in[K]} \mathbf{x}_k \mathbf{x}_k^\top]$

  (Kim and Paik, 2019; Oh et al., 2021; Ariu et al., 2022)

- Margin condition (Bastani and Bayati, 2020; Wang et al., 2018; Li et al., 2021; Ariu et al., 2022; Chakraborty et al., 2023)
- Relaxed symmetry & balanced covariance (Oh et al., 2021; Ariu et al., 2022)
- Anti-concentration (Li et al., 2021; Chakraborty et al., 2023)

---

[1] Alternative form of regularity is required, e.g., minimum eigenvalue of $\mathbf{\Sigma}$ (Hao et al., 2020b), bounded sparse eigenvalue of $\mathbf{\Sigma}$ (Li et al., 2021; Chakraborty et al., 2023)

Existing Assumptions in Related Literature

Compatibility condition on averaged arm

[Compatibility condition on $\boldsymbol{\Sigma}$]

Let $\boldsymbol{\Sigma} := \frac{1}{K}\mathbb{E}[\sum_{k\in[K]} \mathbf{x}_{t,k}\mathbf{x}_{t,k}^\top]$. For support set $S_0 := \{j \in [d] : \beta_j^* \neq 0\}$, $\exists \phi_0^2 > 0$ such that

$$\phi_0^2 \leq \frac{s_0 \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}}{\|\boldsymbol{\beta}_{S_0}\|_1^2} \text{ for all } \boldsymbol{\beta} \text{ with } \|\boldsymbol{\beta}_{S_0^c}\|_1 \leq 3\|\boldsymbol{\beta}_{S_0}\|_1$$

- Introduced to ensure ℓ_1-error bound of Lasso estimate with *i.i.d.* data (Bühlmann and Van De Geer, 2011)
- Extended to Lasso estimate with non-i.i.d. data (Kim and Paik, 2019; Oh et al., 2021; Li et al., 2021; Ariu et al., 2022)

Existing Assumptions in Related Literature

Margin condition

[α-margin condition]
For $\alpha > 0$, $\exists \Delta_* > 0$ such that for any $h > 0$ and $t \in [T]$,

$$\mathbb{P}\left(\mathbf{x}_{t,a_t^*}^\top \boldsymbol{\beta}^* - \max_{k \neq a_t^*} \mathbf{x}_{t,k}^\top \boldsymbol{\beta}^* \leq h \right) \leq \left(\frac{h}{\Delta_*} \right)^\alpha$$

- Probabilistic relaxation of usual “gap” assumption in multi-armed bandit literature (Goldenshluger and Zeevi, 2013)
- $\alpha = 0$ represents no additional condition imposed.
- $\alpha = \infty$ is equivalent to minimum gap condition.
- Utilized to achieve logarithmic depedence on both d and T

  (Bastani and Bayati, 2020; Wang et al., 2018; Li et al., 2021; Ariu et al., 2022; Chakraborty et al., 2023)

Existing Assumptions in Related Literature

Stochastic assumptions on context vector distribution

[Relaxed symmetry]

For $\mathcal{P}_{\mathcal{X}}$, $\exists 1 \leq \nu < \infty$ such that $\frac{\mathcal{P}_{\mathcal{X}}(-\mathbf{x})}{\mathcal{P}_{\mathcal{X}}(\mathbf{x})} \leq \nu \ \forall \mathbf{x}$ with $\mathcal{P}_{\mathcal{X}}(\mathbf{x}) \neq 0$

- Skewness of context distribution is bounded

[Balanced covariance]

Consider a permutation $(i_1, ..., i_K)$ of $(1, ..., K)$. For any $k \in \{2, ..., K-1\}$ and fixed $\boldsymbol{\beta}$, there exists $C_{\mathcal{X}} < \infty$ such that

$$\mathbb{E}\left[\mathbf{x}_{i_k}\mathbf{x}_{i_k}^\top \mathbb{1}\{\mathbf{x}_{i_1}^\top\boldsymbol{\beta} < ... < \mathbf{x}_{i_K}^\top\boldsymbol{\beta}\}\right] \preccurlyeq C_{\mathcal{X}}\mathbb{E}\left[(\mathbf{x}_{i_1}\mathbf{x}_{i_1}^\top + \mathbf{x}_{i_K}\mathbf{x}_{i_K}^\top)\mathbb{1}\{\mathbf{x}_{i_1}^\top\boldsymbol{\beta} < ... < \mathbf{x}_{i_K}^\top\boldsymbol{\beta}\}\right]$$

- Sufficient randomness in observed features compared to non-observed features

[Anti-concentration]

$\exists \xi > 0$ such that for each $k \in [K], t \in [T], \mathbf{v} \in \mathbb{R}^d$, and $h > 0$,

$$\mathbb{P}(|\mathbf{x}_{t,k}^\top \mathbf{v}|^2 \leq h\|\mathbf{v}\|_2^2) \leq \xi h$$

- Prohibits context features to fall along a sigular direction

Research Motivation

- Some combination of the aformentioned assumptions are needed to achieve $\mathcal{O}(\text{poly}\log dT)$ regret.
  - Margin condition is commonly assumed.
- However, their complexity often obscures relative strength of one assumption over another.

Research Motivation

- Some combination of the aformentioned assumptions are needed to achieve $\mathcal{O}(\text{poly}\log dT)$ regret.
  - Margin condition is commonly assumed.
- However, their complexity often obscures relative strength of one assumption over another.

Question: Can construct a weaker condition than existing assumptions to derive $\mathcal{O}(\text{poly}\log dT)$ regret?
Question: Can design a statistical efficient algorithm under such a new condition?

Compatibility Condition on Optimal Arm

HLS condition in stochastic linear bandits

Let $\Sigma^* := \mathbb{E}[\mathbf{x}_{t,a_t^*}\mathbf{x}_{t,a_t^*}^\top]$ where $\mathbf{x}_{t,a_t^*}$ is context feature for optimal arm.
Context feature $\mathcal{P}_{\mathcal{X}}$ is said to be HLS[2] if

$$\lambda_{\min}(\boldsymbol{\Sigma}^*) > 0$$

- Sufficient & necessary condition for achieving constant regret in stochastic linear bandit setting (Hao et al., 2020a; Papini et al., 2021)

[Compatibility condition on optimal arm]
There exists $\phi_*^2 > 0$ such that

$$\phi_*^2 \leq \frac{s_0 \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^* \boldsymbol{\beta}}{\|\boldsymbol{\beta}_{S_0}\|_1^2} \quad \text{for all } \boldsymbol{\beta} \text{ with } \|\boldsymbol{\beta}_{S_0^c}\|_1 \leq 3\|\boldsymbol{\beta}_{S_0}\|_1$$

- Generalization of HLS condition
- WANT: Strictly weaker than existing stochastic assumptions on context distributions

[2] The acronym refers to the last names of the authors of Hao et al. (2020a)

Usual pipeline of theoretical research is...

Assumptions

Towards the Weakest Conditions in Lasso Bandits

Usual pipeline of theoretical research is...

Assumptions

$\Downarrow$ derive

Theorem (Regret Bound)

How can we show that our assumptions are strictly weaker than the existing assumptions?

Existing Assumptions

Towards the Weakest Conditions in Lasso Bandits

How can we show that our assumptions are strictly weaker than the existing assumptions?

Existing Assumptions

$\Downarrow$ implies

Our Assumptions

Towards the Weakest Conditions in Lasso Bandits

How can we show that our assumptions are strictly weaker than the existing assumptions?

Existing Assumptions

$\Downarrow$ implies

Our Assumptions

$\Downarrow$ derive

Theorem (Regret bound)

Towards the Weakest Conditions in Lasso Bandits

How can we show that our assumptions are strictly weaker than the existing assumptions?

Existing Assumptions

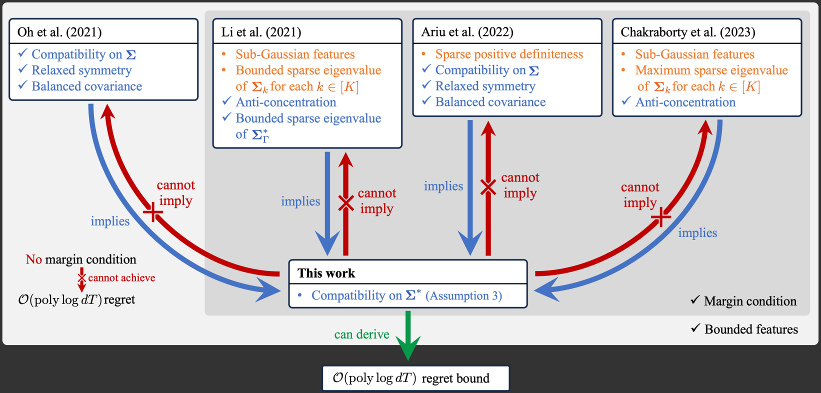
$\Downarrow$ implies

Our Assumptions

$\Downarrow$ derive

Theorem (Sharpest Regret Bound)

Relationships among Distributional Assumptions



- Blue arrows: implication relationships
- Red arrows: *infeasible* implication relationships
- Orange bullets: additional assumptions not needed by our analysis
- Our new assumption is the mildest condition that allows $\mathcal{O}(\text{poly}\log dT)$ regret (in single-parameter Lasso bandit problem)

Relationships among Distributional Assumptions (Cont'd)

Definition (Greedy diversity)
For $\boldsymbol{\beta} \in \mathbb{R}^d$, let $\pi_{\boldsymbol{\beta}}(\{\mathbf{x}_k\}_{k=1}^K) = \arg\max_k \mathbf{x}_k^\top \boldsymbol{\beta}$ and the chosen feature with respect to $\pi_{\boldsymbol{\beta}}$ be $\mathbf{x}_{\boldsymbol{\beta}}$. Context distribution $\mathcal{P}_{\mathcal{X}}$ satisfies the greedy diversity if $\exists \phi_G^2 > 0$ such that

$$\phi_G^2 \leq \frac{s_0 \boldsymbol{\beta}^\top \mathbb{E}_{\{\mathbf{x}_k\}_{k=1}^K \sim \mathcal{P}_{\mathcal{X}}}[\mathbf{x}_{\boldsymbol{\beta}} \mathbf{x}_{\boldsymbol{\beta}}^\top] \boldsymbol{\beta}}{\|\boldsymbol{\beta}_{S_0}\|_1^2} \quad \text{for any } \boldsymbol{\beta} \in \mathbb{R}^d$$

- Greedy diversity implies compatibility condition on optimal arm
  $\because$ optimal arm is a greedy policy with respect to $\boldsymbol{\beta}^*$

Lemma (Anti-concentration $\Rightarrow$ ours)

Anti-concentration condition implies the greedy diversity with $\phi_G^2 = \frac{1}{4\xi K}$.

Lemma (Relaxed symmetry & Balanced covariance $\Rightarrow$ ours)

Relaxed symmetry & Balanced covariance conditions imply the greedy diversity with $\phi_G^2 = \frac{\phi_0^2}{2\nu C_{\mathcal{X}}}$.

Challenges

Under compatibility condition on optimal arm,

- theoretical guarantee of Lasso estimator can be derived <u>only if sufficient selections of optimal arms is guaranteed</u>

To ensure sufficient selections of optimal arms,

- Choose an arm randomly while expecting optimal arm to be chosen

How many times? Is it enough?

Forced Sampling then Weighted Loss Lasso (FS-WLasso)

Input parameter: Number of exploration M_0, Weight w, Regularization param $\{\lambda_t\}_{t\geq 0}$

For each round $t = 1, ..., T$ do:

1. Observe $\mathbf{x}_{t,k}$ for all $k \in [K]$
2. **If** $t \leq M_0$ **then** ▷ *Forced sampling stage*

   Choose $a_t \sim \text{Unif}(\mathcal{A})$ and observe r_{t,a_t}
3. **Else** ▷ *Greedy selection stage*

   Compute $\hat{\boldsymbol{\beta}}_{t-1} = \arg\min_{\boldsymbol{\beta}} wL_0(\boldsymbol{\beta}) + L_{t-1}(\boldsymbol{\beta}) + \lambda_{t-1}\|\boldsymbol{\beta}\|_1$

   Select $a_t = \arg\max_{k\in[K]} \mathbf{x}_{t,k}^\top \hat{\boldsymbol{\beta}}_{t-1}$ and observe r_{t,a_t}

$L_0(\boldsymbol{\beta}) := \sum_{i=1}^{M_0} (\mathbf{x}_{i,a_i}^\top \boldsymbol{\beta} - r_{i,a_i})^2$: samples from forced sampling stage

$L_{t-1}(\boldsymbol{\beta}) := \sum_{i=M_0+1}^{t-1} (\mathbf{x}_{i,a_i}^\top \boldsymbol{\beta} - r_{i,a_i})^2$: samples from greedy selection stage

Regret Bound of FS-WLasso

Assumptions

- [Boundedness] $\mathbf{x} \in \mathcal{X}$, $\|\mathbf{x}\|_\infty \le x_{\max}$, and $\|\boldsymbol{\beta}^*\|_1 \le b$
- [α-margin condition] $\mathbb{P}(\mathbf{x}_{t,a_t^*}^\top \boldsymbol{\beta}^* - \max_{k \neq a_t^*} \mathbf{x}_{t,k}^\top \boldsymbol{\beta}^* \le h) \le (h/\Delta_*)^\alpha$
- [Compatibility condition on $\boldsymbol{\Sigma}^*$] $\exists \phi_*^2 > 0$ such that

$$\phi_*^2 \le \frac{s_0 \boldsymbol{\beta}^\top \boldsymbol{\Sigma}^* \boldsymbol{\beta}}{\|\boldsymbol{\beta}_{S_0}\|_1^2} \quad \text{for all } \boldsymbol{\beta} \text{ with } \|\boldsymbol{\beta}_{S_0^c}\|_1 \le 3\|\boldsymbol{\beta}_{S_0}\|_1$$

Definition

- [Compatibility constant ratio] $\rho := \phi_*^2/\phi_0^2$
  - Ratio of compatibility constant for $\boldsymbol{\Sigma}^*$ to compatibility constant for $\boldsymbol{\Sigma}$
  - $0 < \rho \le K$: compatibility condition on $\boldsymbol{\Sigma}^*$ $\Rightarrow$ compatibility condition on $\boldsymbol{\Sigma}$

Regret Bound of FS-WLasso (Cont'd)

Theorem (Regret bound of FS-WLasso)

For $\delta \in (0,1]$, set input parameters of FS-WLasso *by*

$$\tau \geq \textit{poly}\left(x_{\max}, s_0, \phi_*, \sigma, \alpha, \Delta_*, \log d, \log \delta\right), M_0 = \widetilde{\mathcal{O}}(\rho^2 \sigma^2 x_{\max}^{4+\frac{4}{\alpha}} s_0^{2+\frac{2}{\alpha}} \phi_*^{-4-\frac{4}{\alpha}}),$$
$$\lambda_t = \widetilde{\mathcal{O}}(\sigma x_{\max}(\sqrt{t - M_0} + w\sqrt{M_0})), w = \sqrt{\tau / M_0},$$

then with high probability, regret of FS-WLasso *policy π over round T is upper-bounded by*

$$\textit{Regret}_T(\pi) = \begin{cases} \mathcal{O}(s_0^{\alpha+1} T^{\frac{1-\alpha}{2}} (\log d + \log\log T)^{\frac{\alpha+1}{2}}) & \textit{for } \alpha \in (0,1), \\ \mathcal{O}(s_0^2 \log T (\log d + \log\log T)) & \textit{for } \alpha = 1, \\ \mathcal{O}(s_0^{2+\frac{2}{\alpha}} \log d) & \textit{for } 1 < \alpha \leq \infty. \end{cases}$$

- Matches lower bound of $\mathcal{O}(T^{\frac{1-\alpha}{2}} (\log d)^{\frac{\alpha+1}{2}} + \log T)$ for $\alpha \in (0,1]$ in Li et al. (2021) up to $\log T$ factor
- Mildest condition that allows $\mathcal{O}(\text{poly} \log dT)$ regret
- Expand range of α that logarithmic regret is attainable even for $(s_0 = d)$ low-dimensional setting (previously only known for $\alpha > 2$)

Regret Analysis of `FS-WLasso`

Cyclic structure induced by our assumptions

- optimal arms were chosen sufficiently until $t-1$

  $\Rightarrow$ small estimation error of $\hat{\boldsymbol{\beta}}_t$

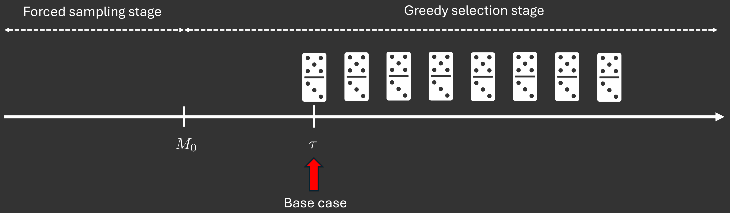
  $\Rightarrow$ high probability of choosing optimal arm at $t+1$

Domino-like phenomenon that propagates forward in time

- Mathematical induction argument: $P(n)$ holds $\Rightarrow P(n+1)$ holds
- Controlling probability of failing to propagate good event

Regret Analysis of FS-WLasso (Cont'd)

(1) Initial condition of induction must be satisfied (base case)



Lemma

Let $\hat{\mathbf{V}}_{M_0} := w \sum_{i=1}^{M_0} \mathbf{x}_{i,a_i} \mathbf{x}_{i,a_i}^\top$. *Suppose number of exploration* M_0 *is set to*

$$M_0 \gtrsim \max \left\{ \rho^2 \left(\frac{\sigma x_{\max}^2 s_0}{\Delta_* \phi_*^2} \right)^2 \left(\frac{x_{\max}^2 s_0}{\phi_*^2} \right)^{\frac{2}{\alpha}} \left(\log\log\tau + \log\frac{d}{\delta} \right), \frac{\rho^2 x_{\max}^4 s_0^2}{\phi_*^4} \log\frac{d^2}{\delta} \right\}.$$

Then with probability at least $1-\delta$,

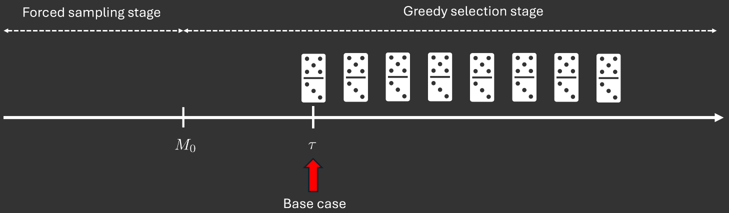
$$\phi^2\left(\hat{\mathbf{V}}_{M_0}\right) \geq \max\left\{ \frac{4x_{\max}s_0}{\Delta_*} \left(\frac{80x_{\max}^2 s_0}{\phi_*^2} \right)^{\frac{1}{\alpha}} \lambda_{M_0+\tau}, 64x_{\max}^2 s_0 \log\frac{1}{\delta} \right\}.$$

- optimal arms were chosen sufficiently

  $\Leftrightarrow$ empirical Gram matrix $\hat{\mathbf{V}}_{M_0}$ concentrates around $\mathbf{\Sigma}^*$

Regret Analysis of FS-WLasso (Cont'd)

(1) Initial condition of induction must be satisfied (base case)



Lemma

Let $\tau \geq$ *poly* $(x_{\max}, s_0, \phi_*, \sigma, \alpha, \Delta_*, \log d, \log \delta)$. *For* $M_0 \leq t \leq \tau$, *let*

$$\lambda_t \gtrsim \sigma x_{\max} \left(\sqrt{w^2 M_0 \log \frac{2d}{\delta}} + \sqrt{(t - M_0) \log \frac{d(\log 2(t-M_0))^2}{\delta}} \right).$$

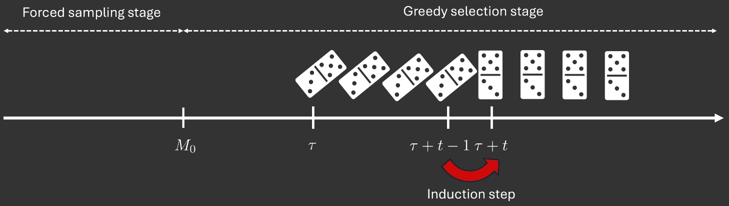
Then, with probability at least $1-\delta$, $\hat{\boldsymbol{\beta}}_t$ *satisfies*

$$\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}_t\|_1 \leq \frac{\Delta_*}{2x_{\max}} \left(\frac{\phi_*^2}{80 x_{\max}^2 s_0} \right)^{\frac{1}{\alpha}}.$$

- $\hat{\boldsymbol{\beta}}_t$ becomes sufficiently accurate rather than tighter with respect to t

Regret Analysis of FS-WLasso (Cont'd)

(2) Propagate good event to next round (induction step)



Lemma

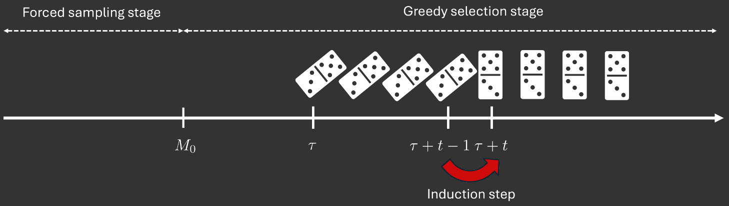
For any $t' \geq 0$, with high probability,

$$\overline{N}(t') := \sum_{t=M_0+1}^{M_0+t'} \left(\frac{2x_{\max}}{\Delta_*} \|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}_{t-1}\|_1 \right)^{\alpha} \leq \frac{\phi_*^2}{80 x_{\max}^2 s_0} t'$$

- $\overline{N}(t')$ is determined by errors of $\hat{\boldsymbol{\beta}}_{t'-1}$ up to $t = M_0 + t'$

Regret Analysis of FS-WLasso (Cont'd)

(2) Propagate good event to next round (induction step)



Lemma (Oracle Inequality for Weighted Loss Lasso Estimate)

For $t' \geq$ poly$(x_{\max}, s_0, \phi_*, \sigma, \alpha, \Delta_*, \log d, \log \delta)$, suppose $\overline{N}(t') \leq \frac{\phi_*^2}{80 x_{\max}^2 s_0} t'$. Then with probability at least $1 - \delta$,

$$\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}_{M_0+t'}\|_1 \leq \frac{C_0 \sigma x_{\max} s_0}{\phi_*^2} \sqrt{\frac{2 \log \log 2t' + \log \frac{7d}{\delta}}{t'}} .$$

- Confidence bound becomes tighter, as t' (number of samples obtained by greedy policy) increases

  ↳ Results in higher probability of choosing optimal arm at next round

Regret Analysis of FS-WLasso (Cont'd)

(3) Control probability of failing to propagate good event at every round

Stochasticity of problem induces small probability of failing to propagate good event

- $\mathcal{E}_e$: sub-Gaussian noise concentration for forced sampling stage
- $\mathcal{E}_g$: sub-Gaussian noise concentration for greedy selection stage
- $\mathcal{E}_N$: bounded number of sub-optimal arm selection for greedy selection stage
- $\mathcal{E}_\tau^*$: bounded compatibility constant of empirical Gram matrix of optimal arm for greedy selection stage

Lemma (High probability of jointly good events)

$$\mathbb{P}(\mathcal{E}_e \cap \mathcal{E}_g \cap \mathcal{E}_N \cap \mathcal{E}_\tau^*) \geq 1 - \delta\,.$$

- With high probability, good events occur independently of induction argument
- Under these good events, induction argument always holds!

Regret Analysis of FS-WLasso (Cont'd)

Divide the time horizon $[T]$ into three groups:

(1) $(t \leq M_0)$: Forced sampling stage

- incur max regret each round: $2x_{\max} b M_0 = \mathcal{O}(s_0^{2+\frac{2}{\alpha}} \log d)$

(2) $(M_0 < t \leq \tau)$: before cycle (base case) begins

- obtain samples with sufficiently accurate estimate
- incur $\mathcal{O}\left(\frac{\sigma^2}{\Delta_*}\left(\frac{x_{\max}^2 s_0}{\phi_*^2}\right)^{1+\frac{1}{\alpha}}\left(\log d + \log\frac{1}{\delta}\right)\right)$ regret

(3) $(t > \tau)$: induction argument holds

- Lasso estimates with tight confidence bound results in high probability of choosing optimal arm

$$\begin{cases} \mathcal{O}\left(\frac{1}{(1-\alpha)\Delta_*^{\alpha}}\left(\frac{\sigma x_{\max}^2 s_0}{\phi_*^2}\right)^{1+\alpha} T^{\frac{1-\alpha}{2}}\left(\log d + \log\frac{\log T}{\delta}\right)^{\frac{1+\alpha}{2}}\right) & \alpha \in (0,1)\,, \\ \mathcal{O}\left(\frac{1}{\Delta_*}\left(\frac{\sigma x_{\max}^2 s_0}{\phi_*^2}\right)^{2}(\log T)\left(\log d + \log\frac{\log T}{\delta}\right)\right) & \alpha = 1\,, \\ \mathcal{O}\left(\frac{\alpha}{(\alpha-1)^2}\cdot\frac{\sigma^2}{\Delta_*}\left(\frac{x_{\max}^2 s_0}{\phi_*^2}\right)^{1+\frac{1}{\alpha}}\left(\log d + \log\frac{1}{\delta}\right)\right) & \alpha > 1\,. \end{cases}$$

Efficiency of Forced Sampling

- What's happening during forced sampling stage
  - Compatibility condition of empirical Gram matrix is not guaranteed
    ↳ this period is also called "burn-in" phase
  - In previous Lasso bandits, compatibility condition after burn-in phase is ensured by diversity assumptions on context vectors, rather than exploration of algorithm
    ↳ Lasso estimator calculation (Oh et al., 2021; Ariu et al., 2022), UCB (Li et al., 2021), TS (Chakraborty et al., 2023)
  - `FS-WLasso` does not compute parameters but just samples arm
    ↳ Do not require additional diversity assumptions on context distribution

Theorem (Regret under Diversity Assumptions)

Suppose either anti-concentration or relaxed symmetry + balanced covariance assumptions hold. Then, `FS-WLasso` still achieves $\mathcal{O}(\textit{poly} \log dT)$ regret even if we set $M_0 = 0$.

- Forced sampling may not be required if diversity assumptions on context distribution are given

Details of Numerical Experiments

Benchmark algorithms

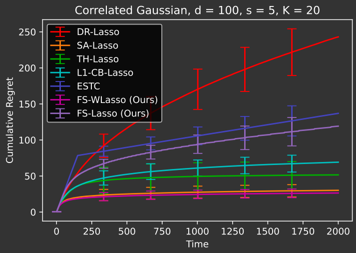
- `DR-Lasso` (Kim and Paik, 2019), `SA-Lasso` (Oh et al., 2021), `TH-Lasso` (Ariu et al., 2022), `L1-CB-Lasso` (Li et al., 2021), `ESTC` (Hao et al., 2020b)

Simulation set up

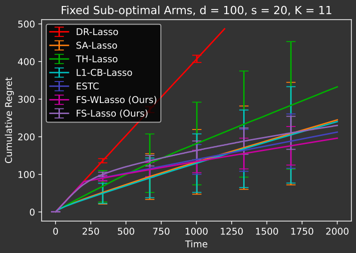
- Generate $\boldsymbol{\beta}^*$ with sparsity $s_0 = \|\boldsymbol{\beta}^*\|_0$ and $\boldsymbol{\beta}_{S_0} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$
- Multivariate correlated Gaussian context distribution (Experiment 1)
- Context feature vectors of sub-optimal arms are fixed, and only optimal arm has randomness (Experiment 2)

  ↳ Diversity assumptions on context distributions are not valid

Results of Numerical Experiments



Experiment 1



Experiment 2

- Report the average cumulative regret over 100 independent runs.
- The error bars represent the standard deviations.

Summary

- Suggest novel sufficient condition for deriving $\mathcal{O}(\text{poly}\log dT)$ regret for Lasso bandit algorithm
  - Compatibility on optimal arm is the weakest assumption on context distributions known in the (single-parameter) lasso bandit problem.
- Propose forced-sampling-based algorithm (`FS-WLasso`) for sparse linear bandit problem
  - Achieves $\mathcal{O}(\text{poly}\log dT)$ regret
  - Do not require additional diversity assumptions on context distribution
- Novel analysis technique based on high-probability analysis & mathematical induction
- `FS-WLasso` significantly outperforms benchmarks

References I

Ariu, K., Abe, K., and Proutière, A. (2022). Thresholded lasso bandit. In International Conference on Machine Learning, pages 878–928. PMLR.

Bastani, H. and Bayati, M. (2020). Online decision making with high-dimensional covariates. Operations Research, 68(1):276–294.

Bühlmann, P. and Van De Geer, S. (2011). Statistics for high-dimensional data: methods, theory and applications. Springer Science & Business Media.

Chakraborty, S., Roy, S., and Tewari, A. (2023). Thompson sampling for high-dimensional sparse linear contextual bandits. In International Conference on Machine Learning, pages 3979–4008. PMLR.

Goldenshluger, A. and Zeevi, A. (2013). A linear response bandit problem. Stochastic Systems, 3(1):230–261.

Hao, B., Lattimore, T., and Szepesvari, C. (2020a). Adaptive exploration in linear contextual bandit. In International Conference on Artificial Intelligence and Statistics, pages 3536–3545. PMLR.

Hao, B., Lattimore, T., and Wang, M. (2020b). High-dimensional sparse linear bandits. Advances in Neural Information Processing Systems, 33:10753–10763.

Kim, G.-S. and Paik, M. C. (2019). Doubly-robust lasso bandit. Advances in Neural Information Processing Systems, 32.

Li, K., Yang, Y., and Narisetty, N. N. (2021). Regret lower bound and optimal algorithm for high-dimensional contextual linear bandit. Electronic Journal of Statistics, 15(2):5652–5695.

Oh, M.-h., Iyengar, G., and Zeevi, A. (2021). Sparsity-agnostic lasso bandit. In International Conference on Machine Learning, pages 8271–8280. PMLR.

Papini, M., Tirinzoni, A., Restelli, M., Lazaric, A., and Pirotta, M. (2021). Leveraging good representations in linear contextual bandits. In International Conference on Machine Learning, pages 8371–8380. PMLR.

Wang, X., Wei, M., and Yao, T. (2018). Minimax concave penalized multi-armed bandit model with high-dimensional covariates. In International Conference on Machine Learning, pages 5200–5208. PMLR.