# Sparsity-Agnostic Lasso Bandit

Min-hwan Oh[1], Garud Iyengar[2], and Assaf Zeevi[2]

[1]Seoul National University, [2]Columbia University

# 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

# 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.

# 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 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?

# High-Dimensional Contextual Bandits

Generalized linear contextual (GLM) bandits

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

- Contexts $\{X_{t,i} \in \mathbb{R}^d \mid i \in [K]\}$ drawn from (unknown) $p_{\mathcal{X}}$
- Agent selects an arm $a_t \in [K]$
- Agent observes reward:

$$Y_t = \underbrace{\mu(X_{t,a_t}^\top \beta^*)}_{\text{expected reward}} + \epsilon_t$$

$\epsilon_t \sim$ sub-Gaussian with parameter $\sigma$

$\beta^* \in \mathbb{R}^d$ unknown to agent; $\mu$ link function of GLM

# High-Dimensional Contextual Bandits

Generalized linear contextual (GLM) bandits

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

- Contexts $\{X_{t,i} \in \mathbb{R}^d \mid i \in [K]\}$ drawn from (unknown) $p_{\mathcal{X}}$
- Agent selects an arm $a_t \in [K]$
- Agent observes reward:

$$Y_t = \underbrace{\mu(X_{t,a_t}^\top \beta^*)}_{\text{expected reward}} + \epsilon_t$$

$\epsilon_t \sim$ sub-Gaussian with parameter $\sigma$

$\beta^* \in \mathbb{R}^d$ unknown to agent; $\mu$ link function of GLM

High-dimensional GLM bandits

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

# Drawback of Existing Work

Emerging body of work on high-dimensional contextual bandit

- Abbasi-Yadkori et al. (2012); Gilton and Willett (2017); Wang et al. (2018); Kim and Paik (2019); Bastani and Bayati (2020)

Crucial drawback of existing work[1]

- They require prior knowledge of sparsity $s_0$!
- Information on $s_0$ is almost never available in practice.

Question: Can design a sparsity-agnostic algorithm and prove its regret?

# Drawback of Existing Work

Emerging body of work on high-dimensional contextual bandit

- Abbasi-Yadkori et al. (2012); Gilton and Willett (2017); Wang et al. (2018); Kim and Paik (2019); Bastani and Bayati (2020)

Crucial drawback of existing work[1]

- They require prior knowledge of sparsity $s_0$!
- Information on $s_0$ is almost never available in practice.

Question: Can design a sparsity-agnostic algorithm and prove its regret?

[1] Carpentier and Munos (2012) do not require to know sparsity, but both their algorithm and analysis are limited to the fixed $\ell_2$ unit ball arm set.

# Why do existing methods need sparsity $s_0$?

Bastani and Bayati (2020); Wang et al. (2018); Kim and Paik (2019), etc.

To ensure "suitable" concentration:

- Assume theoretical Gram matrix satisfies compatibility condition.
- Use forced-sampling[2] to draw sufficient i.i.d. samples to satisfy compatibility condition of empirical Gram matrix.
- Forced-sampling duration is calibrated with using $s_0$.

[2] Technique developed in Goldenshluger and Zeevi (2013)

# Why do existing methods need sparsity $s_0$?

Bastani and Bayati (2020); Wang et al. (2018); Kim and Paik (2019), etc.

To ensure "suitable" concentration:

- Assume theoretical Gram matrix satisfies compatibility condition.
- Use forced-sampling[2] to draw sufficient i.i.d. samples to satisfy compatibility condition of empirical Gram matrix.
- Forced-sampling duration is calibrated with using $s_0$.

Key findings in our analysis

- i.i.d. samples are in fact not essential.
- Empirical Gram matrix satisfies compatibility without i.i.d. samples!
- Even when $s_0$ is known, drawing i.i.d. samples can be wasteful.

[2] Technique developed in Goldenshluger and Zeevi (2013)

# Sparsity-Agnostic (SA) Lasso Bandit

Input parameter: $\lambda_0$
For each round $t = 1, ..., T$ do:

1. Observe $X_{t,i}$ for all $i \in [K]$
2. Compute $a_t = \arg\max_{i \in [K]} X_{t,i}^\top \hat{\beta}_t$
3. Pull arm $a_t$ and observe $Y_t$
4. Update penalty parameter $\lambda_t \leftarrow \lambda_0 \sqrt{\frac{4 \log t + 2 \log d}{t}}$
5. Lasso update $\hat{\beta}_{t+1} \leftarrow \arg\min_\beta \left\{ \ell_t(\beta) + \lambda_t \|\beta\|_1 \right\}$

The algorithm requires one parameter $\lambda_0$

- Establish regret bound with $\lambda_0 = 2\sigma \max_{t,i}\{\|X_{t,i}\|_2\}$.
- Need bound on noise variance $\sigma$ and the $\ell_2$ norm of $X_{t,i}$.
- It does not depend on the sparsity index $s_0$

# Sparsity-Agnostic (SA) Lasso Bandit

Input parameter: $\lambda_0$

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

1. Observe $X_{t,i}$ for all $i \in [K]$
2. Compute $a_t = \arg\max_{i \in [K]} X_{t,i}^\top \hat{\beta}_t$
3. Pull arm $a_t$ and observe $Y_t$
4. Update penalty parameter $\lambda_t \leftarrow \lambda_0 \sqrt{\frac{4 \log t + 2 \log d}{t}}$
5. Lasso update $\hat{\beta}_{t+1} \leftarrow \arg\min_\beta \left\{ \ell_t(\beta) + \lambda_t \|\beta\|_1 \right\}$

The algorithm requires one parameter $\lambda_0$

- Establish regret bound with $\lambda_0 = 2\sigma \max_{t,i}\{\|X_{t,i}\|_2\}$.
- Need bound on noise variance $\sigma$ and the $\ell_2$ norm of $X_{t,i}$.
- It does not depend on the sparsity index $s_0$

# Regret Analysis for Two-Armed Sparse Bandit

## Definitions

- Theoretical Gram matrix $\Sigma := \frac{1}{K}\mathbb{E}[\mathbf{X}^\top \mathbf{X}]$ where $\mathbf{X} \in \mathbb{R}^{K \times d}$
- Support set $S_0 := \{j : \beta_j^* \neq 0\}$

## Assumptions

- [Boundedness] $X \in \mathcal{X}$, $\|X\|_2 \leq 1$. $\kappa_{\min} \leq \dot{\mu}(X^\top \beta) \leq \kappa_{\max}$
- [Compatibility condition] For support set $S_0$, $\exists \phi_0^2 > 0$ such that

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

- [Relaxed symmetry][3] For $p_{\mathcal{X}}$, $\exists \rho_0 < \infty$ such that $\frac{p_{\mathcal{X}}(-\mathbf{x})}{p_{\mathcal{X}}(\mathbf{x})} \leq \rho_0 \; \forall \mathbf{x}$

We only need the parameters $(\kappa_{\min}, \kappa_{\max}, \phi_0, \rho_0)$ to exist. We do not need to know their values.

---

[3] In (non-sparse) low dimensional settings, this condition is equivalent to "covariate diversity" in Bastani et al. (2020)

# Regret Bound of SA Lasso Bandit for Two Arms

**Theorem (Regret bound for two arms)**

*The expected regret of the* SA Lasso Bandit *policy* $\pi$ *for two arms over horizon* $T$ *is upper-bounded by*

$$\textit{Regret}_T(\pi) = \mathcal{O}\left(s_0\sqrt{T}\log(dT)\right).$$

- SA Lasso Bandit achieves this bound without knowing $s_0$.
- Correct dependence on $d$ and $s_0$ based on offline Lasso results[4]
- Tighter than previously known $\mathcal{O}\left(s_0\sqrt{T}\log(dT)\right)$ regret (Kim and Paik, 2019) which assumes knowledge of $s_0$

[4] e.g., Theorem 6.1 in Bühlmann and Van De Geer (2011)

# Regret bound of SA Lasso Bandit for $K$ arms

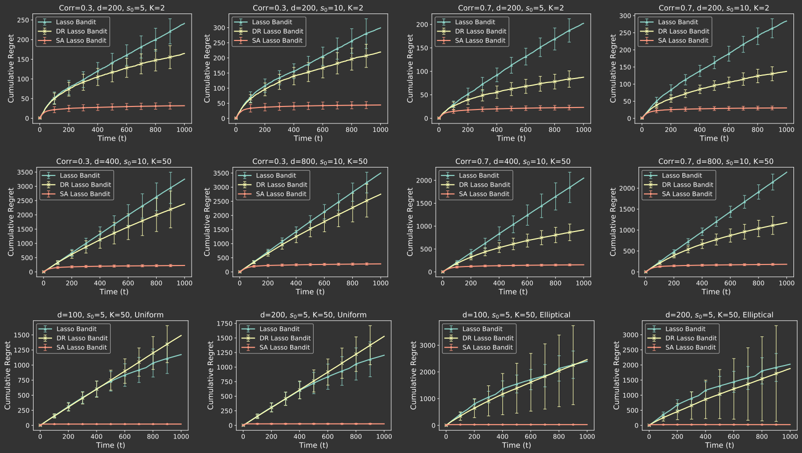
## Theorem (Regret bound for $K$ arms)

*The expected regret of the* SA Lasso Bandit *policy $\pi$ for $K$ arms over horizon $T$ is upper-bounded by*

$$\textit{Regret}_T(\pi) = \mathcal{O}\left(C_{\mathcal{X}} s_0 \sqrt{T \log(dT)}\right).$$

- Achieve the same rate as the regret bound for two-arm case – without prior knowledge on $s_0$
- Main difference is how $\Sigma_t$ is controlled with $\Sigma$.
- Both for two arms and $K$ arms, first regret analysis of a general sparse bandit algorithm that does not require the knowledge of $s_0$.

# Results of the numerical experiments




- Report the average cumulative regret over 20 independent runs.
- The error bars represent the standard deviations.

# References I


Abbasi-Yadkori, Y., Pal, D., and Szepesvari, C. (2012). Online-to-confidence-set conversions and application to sparse stochastic bandits. In Artificial Intelligence and Statistics, pages 1–9.

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

Bastani, H., Bayati, M., and Khosravi, K. (2020). Mostly exploration-free algorithms for contextual bandits. Management Science.

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

Carpentier, A. and Munos, R. (2012). Bandit theory meets compressed sensing for high dimensional stochastic linear bandit. In Artificial Intelligence and Statistics, pages 190–198.

Gilton, D. and Willett, R. (2017). Sparse linear contextual bandits via relevance vector machines. In 2017 International Conference on Sampling Theory and Applications (SampTA), pages 518–522. IEEE.

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

Kim, G.-S. and Paik, M. C. (2019). Doubly-robust lasso bandit. In Advances in Neural Information Processing Systems, pages 5869–5879.

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.