# A Simple Unified Framework for High Dimensional Bandit Problems

Wenjie Li[1], Adarsh Barik[1], and Jean Honorio[1]

[1]Department of Statistics, Purdue University

[2]Department of Computer Science, Purdue University

## Introduction

**Background**

- Stochastic multiarmed contextual bandits are useful models in various application domains, such as recommendation systems, online advertising, and personalized healthcare [1, 2, 3].
- In practice, such problems are often high-dimensional, but the unknown parameter is typically assumed to have low-dimensional structure, which in turns implies a succinct representation of the final reward. [5, 6, 4]
- However, prior works are scattered and different algorithms with different assumptions are proposed for these problems.

**Examples of High Dimensional Bandit Problems**

- LASSO Bandit.
- Low Rank Matrix Bandit.
- Group Sparse Matrix Bandit.

**Contributions**

- We present a simple and unified algorithm framework named Explore-the-Structure-Then-Commit (ESTC) for high dimensional stochastic bandit problems
- We provide a problem-independent regret analysis framework for our algorithm.
- We demonstrate the usefulness of our framework by applying it to different high dimensional bandit problems.

## Notations and Definitions

In modern multiarmed contextual bandit problems, a set of contexts $\{x_{t,a_i}\}_{i=1}^K$ for each arm is generated at every round $t$, and then the agent chooses an action $a_t$ from the $K$ arms. The contexts are assumed to be sampled i.i.d from a distribution $\mathcal{P}_X$ with respect to $t$, but the contexts for different arms can be correlated [3]. After the action is selected, a reward $y_t = f(x_{t,a_t}, \theta^*) + \epsilon_t$ for the chosen action is received.

Let $a_t^* = \text{argmax}_{i \in [K]} f(x_{t,a_i}, \theta^*)$ denote the optimal action at each round. We measure the performance of all algorithms by the expectation of the regret, denoted as

$$\mathbb{E}[R(T)] = \mathbb{E}\left[\sum_{t=1}^{T} f(x_{t,a_t^*}, \theta^*) - f(x_{t,a_t}, \theta^*)\right]$$

Many algorithms designed for multiarmed bandit problems involve solving an online optimization problem with a loss function $L_t(\theta; \mathrm{X}_t, \mathrm{Y}_t)$ and a regularization norm $R(\theta)$, i.e.,

$$\theta_t \in \text{argmin}_{\theta \in \Theta}\Big\{L_t(\theta; \mathrm{X}_t, \mathrm{Y}_t) + \lambda_t R(\theta)\Big\} \qquad (1)$$

where $\lambda_t$ is the regularization parameter chosen differently in different algorithms and $\Theta$ is the parameter domain.

## Assumptions

We use the following assumptions for the analysis of our algorithm, which are common in the analysis of high dimensional bandits [3, 6]

**Assumption 1. (Boundedness)** $x$ is normalized with respect to the norm $\|\cdot\|_\cdot$, i.e. $\|x\| \leq k_1$ for some constant $k_1$ .

**Assumption 2. (Lipschitzness)** $f(x, \theta)$ is $C_1$-Lipschitz over $x$ and $C_2$-Lipschitz over $\theta$ with respect to $\|\cdot\|$. i.e.,

$$f(x_1, \theta) - f(x_2, \theta) \leq C_1 \Big\|x_1 - x_2\Big\|,$$
$$f(x, \theta_1) - f(x, \theta_2) \leq C_2 \Big\|\theta_1 - \theta_2\Big\|$$

**Assumption 3. (Restricted Eigenvalue Condition)** Let $\mathrm{X}$ denote the matrix where each row is a context vector from an arm. The population Gram matrix $\Sigma = \frac{1}{K}\mathbb{E}[\mathrm{X}^T\mathrm{X}]$ satisfies that there exists some constant $\alpha_0 > 0$ such that $\beta^T\Sigma\beta \geq \alpha_0\|\beta\|^2$, for all $\beta \in \mathbb{C}$.

## Algorithm Framework

We propose the following Explore the Structure then Commit (ESTC) algorithm framework for high dimensional bandit problems.

**Algorithm 1** Explore-the-Structure-Then-Commit (ESTC)

```
1: Input: λ_{T_0}, K ∈ ℕ, L_t(θ), R(θ), f(x, θ), θ_0, T_0
2: Initialize X_0, Y_0 = (∅, ∅), θ_t = θ_0
3: for t = 1 to T_0 do
4:     Observe K contexts, x_{t,1}, x_{t,2}, ··· , x_{t,K}
5:     Choose action a_t uniformly randomly
6:     Receive reward y_t = f(x_{t,a_t}, θ*) + ε_t
7:     X_t = X_{t−1} ∪ {x_{t,a_t}}, Y_t = Y_{t−1} ∪ {y_{a_t}}
8: end for
9: Compute the estimator θ_{T_0}:
       θ_{T_0} ∈ argmin_{θ∈Θ} {L_{T_0}(θ; X_{T_0}, Y_{T_0}) + λ_{T_0} R(θ)}
10: for t = T_0 + 1 to T do
11:     Choose action a_t = argmax_a f(x_{t,a}, θ_{T_0})
12: end for
```

Our algorithm generalizes over the prior efforts on different high dimensional bandit problems.

**Advantages of the ESTC Algorithm**

- it is very simple
- it does not require strong assumptions
- it can be applied to different problems

## Regret Bounds

We provide a regret bound of Algorithm 1 that is independent of the high dimensional bandit problem.

**Theorem 1: Problem Independent Regret Bound**

The expected cumulative regret of Algorithm 1 satisfies the bound

$$\mathbb{E}[R(T)] = \mathcal{O}\left(\sum_{t=T_0}^{T}\sqrt{9\frac{\lambda_{T_0}^2}{\alpha^2}\phi^2 + \frac{1}{\alpha}[2Z_{T_0}(\theta^*) + 4\lambda_{T_0}R(\theta^*_{\mathcal{M}^\perp})]}\right)$$

Given the specific high dimensional bandit problems, we obtain the following regret bounds in different problems.

Table 1: Summary of Regret Bounds of Our ESTC Algorithm Framework in Different High Dimensional Bandit Problems

| High Dimensional Bandit Problem | Regret Bound |
|---|---|
| LASSO Bandit (sanity check) | $\mathcal{O}(s^{1/3}T^{2/3}\sqrt{\log(dT)})$ |
| Low-rank Matrix Bandit | $\mathcal{O}(r^{1/3}T^{2/3}\log((d_1+d_2)T)$ |
| Group-Sparse Matrix Bandit | $\mathcal{O}(s^{1/3}\sqrt{d_2}T^{2/3}\sqrt{\log d_1 T})$ |
| Multi-agent Matrix Bandit | $\mathcal{O}(d_2 s^{1/3}T^{2/3}\sqrt{\log(d_1 T)})$ |

## Experiments

The following figures validate the correctness of our theory.

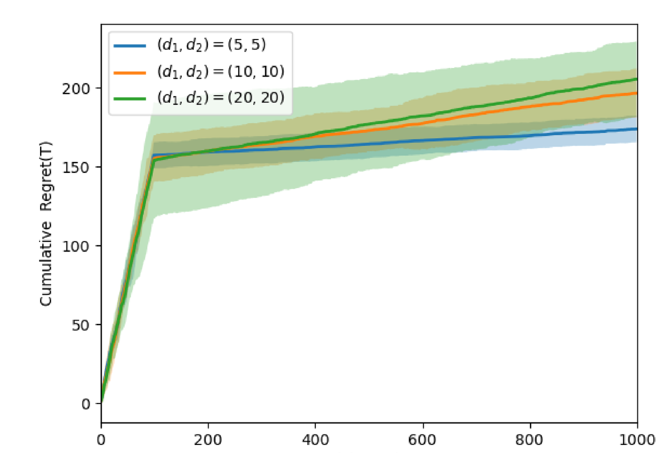


Figure 1: Low-rank Matrix Bandit $R(T)$

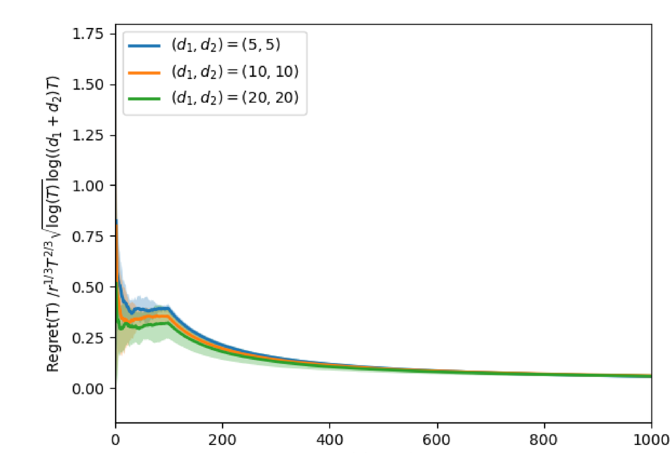


Figure 2: $R(T)/\text{Bound}(T)$

## References

[1] Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. "Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits". In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics.* 2011.

[2] Peter Auer. "Using confidence bounds for exploitationexploration trade-offs". In: *Journal of Machine Learning Research* (2002b), 3:397–422.

[3] Wei Chu et al. "Contextual Bandits with Linear Payoff Functions". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)* (2011).

[4] Nicholas Johnson, Vidyashankar Sivakumar, and Arindam Banerjee. *Structured Stochastic Linear Bandits.* 2016. arXiv: 1606.05693 [stat.ML].

[5] Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. "Low-Rank Generalized Linear Bandit Problems". In: *Proceedings of International Conference on Artificial Intelligence and Statistics.* 2021.

[6] Min-Hwan Oh, Garud Iyengar, and Assaf Zeevi. "Sparsity-Agnostic Lasso Bandit". In: *Proceedings of the 38th International Conference on Machine Learning.* 2021.