Efficient Low-Rank Matrix Estimation, Experimental Design, and Arm-Set-Dependent Low-Rank Bandits

Kyoungseok Jang[1] Chicheng Zhang[2] Kwang-Sung Jun[2]

[1]Università degli Studi di Milano

[2]University of Arizona

UNIVERSITÀ DEGLI STUDI DI MILANO

THE UNIVERSITY OF ARIZONA

Table of contents

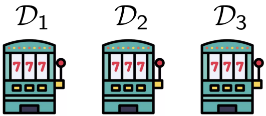
Table of Contents

Intro - Bandit problem

Scenario: At every time step $t \in \{1, \ldots, T\}$,

1. The learner chooses a machine A_t
2. and receives reward $y_t \sim \mathcal{D}_{A_t}$,
3. The learner does not know $\mathcal{D}_{A_t}$, and should learn by trials.

Objective: Maximize reward



Exploration vs Exploitation

- Exploration: spend enough chances to learn each $\mathcal{D}_i$
- Exploitation: believe your estimate and try to earn.

Adding low-rank structure

- Many applications on this bandit problem.
- Naturally, researchers tried to extend it by adding some structures over it.
- One good candidate is the low-rank structure.
    - In many cases, data exhibit low-rank structure.
- We call this problem as a low-rank bandit problem, and has the following applications.

NETFLIX

Problem Setup

At every time step $t \in \{1, \ldots, T\}$,

1. The learner chooses an arm A_t from the arm set $\mathcal{A} \subset \mathbb{R}^{d_1 \times d_2}$
2. and receives reward $y_t = \langle \Theta^*, A_t \rangle + \eta_t$,
   - where Θ^* is an unknown matrix with a known upper bound of the rank at most $r \ll \min(d_1, d_2)$.
   - η_t is an independent zero-mean σ-subgaussian noise
   - The inner product of two matrices are defined as $\langle A, B \rangle = \langle \text{vec}(A), \text{vec}(B) \rangle = \text{tr}(A^\top B)$.

Objective: Minimize its (pseudo-)regret:

$$\text{Reg}(T) := T \max_{A \in \mathcal{A}} \langle \Theta^*, A \rangle - \sum_{t=1}^{T} \langle \Theta^*, A_t \rangle.$$

Assumptions

Traditional boundedness on arms and Θ^*, but on $\|\cdot\|_{\text{op}}$ and $\|\cdot\|_{\text{nuc}}$

Assumption 1 (operator norm-bounded arm set)

The arm set $\mathcal{A} \subseteq \mathcal{B}_{\text{op}}(1) := \left\{ A \in \mathbb{R}^{d_1 \times d_2} : \|A\|_{\text{op}} \leq 1 \right\}$.

Assumption 2 (Bounded norm on reward predictor)

The reward predictor has a bounded nuclear norm: $\|\Theta^*\|_{nuc} \leq S_*$.

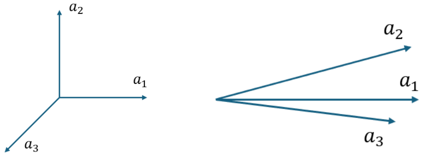
In many cases we will use the following weaker assumption than Assumption 2.

Assumption 3 (Bounded expected reward)

For all $A \in \mathcal{A}$, $\left|\langle \Theta^*, A \rangle\right| \leq R_{\max}$.

Previous Works - generality

- This field has been receiving a lot of attention recently.
- but many of them only deals with specific arm sets.
  - Frobenius norm ball $\mathcal{A} = \{M : \|M\|_F \leq 1\}$ [1]
  - Symmetric unit vector pairs $\{uu^\top : u \in \mathbb{S}^{d-1}\}$ [6, 8],
  - Entrywise canonical actions
    $\mathcal{A} = \{e_{ij}$: Only (i,j)-th entry is 1, and 0 otherwise$\}$ [5, 11, 2]
  - Perfectly symmetric and easy to think about exploration.
- In reality, it isn't!
  - (e.g.) Finite-armed low-rank bandit - usually arms are 'skewed'.



Well-distributed Skewed

Previous Works - estimation perspective

- Estimation: relied on the low-rank estimation literature.
  - Traditionally use nuclear norm regularized least squares ($\|\cdot\|_{\mathrm{nuc}}$-RLS)

$$\hat{\Theta}_t^{(nuc)} = \arg \min_{\Theta \in \mathbb{R}^{d_1 \times d_2}} \sum_{s=1}^{t} \left(\langle \Theta, A_s \rangle - y_s\right)^2 + \lambda \|\Theta\|_{nuc}$$

  - These low-rank estimation studies are mainly for offline setting - the data is **given**, and the learner just estimates.
  - In the setting **where the learner collects the data (such as bandits)**, experimental designs are also important.

Previous works - experimental design perspective

- Experimental design: find out appropriate distribution $\pi \in \mathcal{P}(\mathcal{A})$ for optimal estimation.
  - **It is hard to optimize experimental design for $\|\cdot\|_{\text{nuc}}$-RLS!**
  - Instead: just maximizing minimum eigenvalue [3].

Definition 1

For each distribution π over $\mathcal{A}$ (i.e., $\pi \in \mathcal{P}(\mathcal{A})$), define its covariance matrix $Q(\pi) := \mathbb{E}_{a\sim\pi}\left[\text{vec}(a)\text{vec}(a)^\top\right] = \sum_{a\in\mathcal{A}} \pi(a)\text{vec}(a)\text{vec}(a)^\top$.

$$C_{\min}(\mathcal{A}) = \max_{\pi\in\mathcal{P}(\mathcal{A})} \lambda_{\min}(Q(\pi)).$$

  - or simply assume some nice exploration distribution [9, 4].
  - **Not enough discussion on the experimental design!**

Question

For low-rank trace regression, can we design estimation algorithms with experimental designs that can outperform the classical nuclear norm penalized least squares?

Contribution

1. **A novel and computationally efficient low-rank estimation** method called LowPopArt .
   - We show that the estimation error of LowPopArt is *not worse and can orderwisely better than* the classical $\|\cdot\|_{\text{nuc}}$-RLS.

2. **A computationally tractable design of experiment objective** $B(Q(\pi))$ optimized for LowPopArt .

3. **Two computationally efficient and arm set geometry-adaptive algorithms using LowPopArt,** for low-rank bandits with general arm sets:
   - LPA-ETC (LowPopArt-Explore-Then-Commit), algorithm that works with fewer assumption.
   - LPA-ESTR (LowPopArt-Explore-Subspace-Then-Refine), strictly better performance compared to SOTA.

Table of Contents

LowPopArt - Assumption

- LowPopArt takes the **population covariance matrix** $Q(\pi)$ as its main input.
  - In some settings (such as bandits) where the learner should also collect data by himself, it is natural.
- LowPopArt also takes the pilot estimator Θ_0 and pilot estimation error bound R_0 as input, if possible.
  - For the case that we already have an appropriate candidate.
  - If one does not have such candidate, one can simply set $\Theta_0 = 0_{d\times d}$ and $R_0 = R_{\max}$.

LowPopArt - Algorithm

1. For each sample (A_i, Y_i), compute one-sample estimator $\tilde{\Theta}_i$.
   - Unbiased estimator of $\Theta^* - \Theta_0$
2. Using $\{\tilde{\Theta}_i\}_{i=1}^{n_0}$, compute the matrix Catoni estimator Θ_1 [10].
   - Lightening the tail distribution of singular values.
3. Run SVD on Θ_1, and zero out all the singular values which are under the threshold.

LowPopArt - Algorithm

Algorithm 5 LowPopArt

1: **Input:** Samples $\{A_i, Y_i\}_{i=1}^{n_0}$, sample size n_0, the population covariance matrix of the vectorized matrix $Q(\pi)$, pilot estimator Θ_0 and pilot estimation error bound R_0.

`Step 1: Compute one-sample estimators.`

2: **for** $t = 1, \ldots, n_0$ **do**

3: Compute $\tilde{\Theta}_i := Q(\pi)^{-1}(Y_i - \langle \Theta_0, A_i \rangle)\mathrm{vec}(A_i)$.

4: **end for**

`Step 2: Compute the matrix Catoni estimator using` $\{\tilde{\Theta}_i\}_{i=1}^{n_0}$

5: Compute:

$$\Theta_1 = \Theta_0 + \left(\frac{1}{n_0 \nu} \sum_{i=1}^{n_0} \psi \left(\nu \mathcal{H} \left(\mathrm{reshape} \left(\tilde{\Theta}_i \right) \right) \right) \right)_{\mathrm{ht}}$$

where $\nu = \frac{1}{\sigma + R_0} \sqrt{\frac{2}{B(Q) n_0} \ln \frac{2d}{\delta}}$.

`Step 3: Hard-thresholding eigenvalues.`

6: Let $U_1 \Sigma_1 V_1^\top$ be Θ_1's SVD. Let $\tilde{\Sigma}_1$ be a modification of Σ that zeros out its diagonal entries that are at most $\lambda_{\mathrm{th}} := 2(R_0 + \sigma)\sqrt{\frac{(B(Q) \ln \frac{2d}{\delta})}{n_0}}$ where $B(Q)$ is in Eq. (4).

7: **Return:** Estimator $\hat{\Theta} = U_1 \tilde{\Sigma}_1 V_1^\top$.

Details for step 2 - matrix Catoni estimator [10]

Definition 2 (Catoni's estimator)

Given a symmetric matrix M with its eigenvalue decomposition $M = U \Lambda U^\top$ where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_d)$, we first define $\phi_0 : \mathbb{R} \to \mathbb{R}$ as

$$\psi_0(x) = \begin{cases} \log(1 + x + \frac{x^2}{2}) & \text{if } x > 0 \\ -\log(1 - x + \frac{x^2}{2}) & \text{otherwise} \end{cases}$$

and $\psi : \mathbb{R}^{d \times d} \to \mathbb{R}^{d \times d}$ as

$$\psi(M) = U \left[\mathrm{diag}(\psi_0(\lambda_1), \psi_0(\lambda_2), \cdots, \psi_0(\lambda_d)) \right] U^\top$$

Details for step 2 - matrix Catoni estimator [10]

Definition 3 (Dilation operator)

For any matrix $A \in \mathbb{R}^{d_1 \times d_2}$, define the dilation operator $\mathcal{H} : \mathbb{R}^{d_1 \times d_2} \to \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$ as

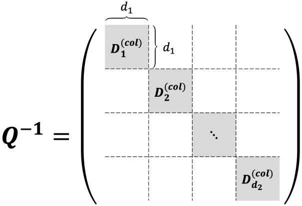
$$\mathcal{H}(A) = \begin{bmatrix} 0_{d_1 \times d_1} & A \\ A^\top & 0_{d_2 \times d_2} \end{bmatrix}.$$

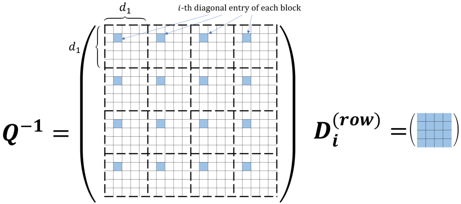
- This method gives an operator norm confidence bound given the operator norm variance.
  - Usually, it relies on the subgaussian proxy rather than variance.
  - For more details about how it works, please check [10].

New arm-set-dependent parameter $B(Q)$

- We found out that the following quantity, $B(Q(\pi))$, determines the **variance** (from the random arm selection) of the singular values of the one-sample estimator, $\tilde{\Theta}_i$.

$$B(Q(\pi)) := \max\left(\lambda_{\max}\left(\sum_{i=1}^{d_2} D_i^{(\mathrm{col})}\right), \lambda_{\max}\left(\sum_{i=1}^{d_1} D_i^{(\mathrm{row})}\right)\right)$$





New design of experiment and $B_{\min}(\mathcal{A})$

- One natural experimental design is to 'minimize variance.'

$$B_{\min}(\mathcal{A}) := \min_{\pi \in \mathcal{P}(\mathcal{A})} B(Q(\pi)) \tag{1}$$

- Turns out this is a convex optimization.
- Relationship between traditional $B(Q(\pi))$ and $\lambda_{\min}(Q(\pi))$:

Lemma 4

$B(Q(\pi)) \leq \frac{d}{\lambda_{\min}(Q)}$

Lemma 5

Suppose Assumption 1 holds. Then $d^2 \leq B_{\min}(\mathcal{A}) \leq \frac{d}{C_{\min}}$, *and there exists an arm set* $\mathcal{A}_{hard}$ *for which* $B_{\min}(\mathcal{A}_{hard}) \approx \frac{1}{C_{\min}}$.

Theoretical guarantee of LowPopArt

Theorem 6 (Theoretical guarantee of LowPopArt)

Suppose that Assumption 1 holds, and LowPopArt is run with arm set $\mathcal{A}$, sample size n_0, and failure rate δ. Then its output $\hat{\Theta}$ satisfies $\text{rank}(\hat{\Theta}) \le r$ and

$$\|\hat{\Theta} - \Theta^*\|_{\text{op}} \le \tilde{O}\left((\sigma + R_0)\sqrt{\frac{B(Q(\pi))}{n_0}}\right).$$

If we optimize π by Eq. (1), we can change $B(Q(\pi))$ to $B_{\min}(\mathcal{A})$.

Warm-LowPopArt

- To avoid the multiplicative R_0 term from the estimation.
- Run LowPopArt twice.
  - The first phase is to create a pilot estimator Θ_0 which satisfies $\max_{A\in\mathcal{A}}\left|\langle\Theta_0-\Theta, A\rangle\right| \le \sigma$
  - The second phase is to finish estimation.

Algorithm 2 Warm-LowPopArt: a bootstrapped version of LowPopArt

1: **Input:** Samples $\{X_i, Y_i\}_{i=1}^{n_0}$, sample size n_0, population covariance matrix of the vectorized matrix Q, failure rate δ.
2: $\Theta_0 \leftarrow$ LowPopArt$(\{X_i, Y_i\}_{i=1}^{\frac{n_0}{2}}, n_0/2, Q, 0_{d_1\times d_2}, S_*, \delta/2)$
3: $\hat{\Theta} \leftarrow$ LowPopArt$(\{X_i, Y_i\}_{i=\frac{n_0}{2}+1}^{n_0}, n_0/2, Q, \Theta_0, \sigma, \delta/2)$
4: **Return:** $\hat{\Theta}$

Warm-LowPopArt Analysis

Theorem 7

Suppose that Assumption 1 and 3 hold, and Warm-LowPopArt is run with arm set $\mathcal{A}$, sample size n_0, failure rate δ, and $n_0 \geq \tilde{O}\left(r^2 B(Q(\pi)) \cdot (\frac{\sigma + R_{\max}}{\sigma})^2\right)$, then its output $\hat{\Theta}$ is such that rank$(\hat{\Theta}) \leq r$, and:

$$\|\hat{\Theta} - \Theta^*\|_{op} \leq \tilde{O}\left(\sigma\sqrt{\frac{B(Q(\pi))}{n_0}}\right). \qquad (2)$$

How good LowPopArt is compare to $\|\cdot\|_{\text{nuc}}$-RLS?

- Theoretical estimation error: $\|\hat{\Theta} - \Theta^*\|_{\text{op}} \le \tilde{O}\left(\frac{\sigma}{\phi^2}\sqrt{\frac{1}{n_0}}\right)$.
  - ϕ: Compatibility constant, hard to optimize.
- Traditional estimation error ($\|\cdot\|_{\text{nuc}}$-RLS):

$$\|\hat{\Theta} - \Theta^*\|_{\text{op}} \le \tilde{O}\left(\sigma C_{\min}^{-1}(\mathcal{A})\sqrt{\frac{1}{n_0}}\right).$$

- In our setting, $C_{\min}{}^{-1}(\mathcal{A}) > d$, so $C_{\min}{}^{-1}(\mathcal{A}) > \sqrt{B_{\min}(\mathcal{A})}$.
- **Our LowPopArt is always better than traditional $\|\cdot\|_{\text{nuc}}$-RLS approach!**
- In many cases, LowPopArt is **orderwise** better
  - When $\mathcal{A} = \mathcal{B}_F(1)$, $C_{\min}^{-1} = d^2 \gg d^{1.5} = \sqrt{B_{\min}}$.
  - When $\mathcal{A} = \mathcal{A}_{\text{hard}}$, $C_{\min}^{-1} \approx d^3 \gg d^{1.5} \approx \sqrt{B_{\min}}$.

Table of Contents

New algorithms

If we replace the **exploration phases** of the existing experimental-design-based algorithms **with our LowPopArt -based method, it shows significantly better performance!**

LPA-ETC

Algorithm 6 LPA-ETC (LowPopArt based Explore then commit)

Input: time horizon T, arm set $\mathcal{A}$, exploration lengths n_0, regularization parameter ν, pilot estimator Θ_0
Solve the optimization problem in Eq. (1) and denote the solution as π^*
for $t = 1, \ldots, n_0$ **do**
    Independently pull the arm A_t according to π^* and receives the reward Y_t
end for
Run Warm-LowPopArt$(\{A_i, Y_i\}_{i=1}^{n_0}, n_0, Q(\pi^*), \delta)$ and get $\hat{\Theta}$
for $t = n_0 + 1, \ldots, T$ **do**
    Pull the arm $A_t = \arg\max_{A \in \mathcal{A}} \langle \hat{\Theta}, A \rangle$
end for

- This algorithm works with fewer and lenient assumptions
  - Needs only Assumption 1 and 3
  - No Assumption 2 or $\lambda_{\min}(\Theta^*)$ assumption needed.

LPA-ETC: Theoretical bound

Theorem 8 (Regret upper bound of LPA-ETC)

Suppose that Assumption 1 and 3 hold, and $T \geq r B_{\min}(\mathcal{A})(\frac{\sigma+R_{\max}}{\sigma})^4$. The regret upper bound of LPA-ETC with $n_0 = \min(T, \left(\sigma^2 r^2 B_{\min}(\mathcal{A}) T^2 / R_{\max}^2\right)^{1/3})$ is as follows:

$$\mathrm{Reg}(T) \leq \tilde{O}((\sigma^2 R_{\max} r^2 T^2 B_{\min}(\mathcal{A}))^{1/3}) \tag{3}$$

LPA-ESTR

Algorithm 7 LPA-ESTR (LowPopArt based Explore Subspace Then Refine)

1: **Input:** time horizon T, arm set $\mathcal{A}$, exploration lengths n_0, singular value lower bound S_r
2: Solve the optimization in Eq. (1) and denote the solution as π^*.
3: **for** $t = 1, \ldots, n_0$ **do**
4: Independently pull the arm A_t according to π^* and receives the reward Y_t
5: **end for**
6: Run Warm-LowPopArt$(\{A_i, Y_i\}_{i=1}^{n_0}, n_0, Q(\pi^*), \delta)$ and get $\hat{\Theta}$ with SVD result $\hat{\Theta} = \hat{U}\hat{\Sigma}\hat{V}^\top$.
7: Let $\hat{U}_\perp$ and $\hat{V}_\perp$ be the orthonormal bases of the orthogonal complement subspaces of $\hat{U}$ and $\hat{V}$, respectively.
8: Rotate whole arm feature set $\mathcal{A}' := \{[\hat{U}\ \hat{U}_\perp] A [\hat{V}\ \hat{V}_\perp]^\top : A \in \mathcal{A}\}$
9: Define a vectorized arm feature set so that the last $(d_1 - r)(d_2 - r)$components are from the complementary subspaces:

$$\mathcal{A}'_{vec} := \{(\text{vec}(A'_{1:r,1:r}); \text{vec}(A'_{r+1:d_1,1:r}); \\ \text{vec}(A'_{1:r,r+1:d_2}); \text{vec}(A'_{r+1:d_1,r+1:d_2})) : A' \in \mathcal{A}'\}$$

10: Invoke LowOFUL with time horizon $T - n_0$, arm set $\mathcal{A}'_{\text{vec}}$, the low dimension $k = r(d_1 + d_2 - r)$, $\lambda = \frac{\sigma^2}{S_*^2} dr$, $\lambda_\perp = \frac{T}{r\log(1+\frac{dT}{\lambda})}$, $B = S_*$, and $B_\perp = \frac{B_{\min}(\mathcal{A})\sigma^2 S_*}{n_0 S_r^2}$.

LPA-ESTR: Theoretical bound

Theorem 9

Suppose that Assumptions 1 and 2 hold, $\lambda_{\min}(\Theta^*) \geq S_r$ for some known $S_r > 0$, and $T \geq \frac{16 B_{\min}(\mathcal{A})\sigma^4}{d^{0.5} S_r (\Theta^*)^2}$. The regret upper bound of LPA-ESTR with $n_0 = \sqrt{\frac{d^{0.5} B_{\min}(\mathcal{A})}{S_r^2} T}$ is

$$\mathrm{Reg}(T) \leq \tilde{O}\left(\sigma \sqrt{\frac{S_*^2}{S_r^2} B_{\min}(\mathcal{A}) d^{0.5} T}\right)$$

with probability at least $1 - 2\delta$.

- Needs Assumption 2 and $\lambda_{\min}(\Theta^*) \geq S_r$, as other ESTR based algorithms do.
- Strictly better performance compare to SOTA.

Improvement of algorithms

	Regret bound	Regret when $\mathcal{A} = \mathcal{B}_{\text{op}}(1)$	Regret when $\mathcal{A} = \mathcal{A}_{\text{hard}}$	Limitation
OFUL (Abbasi-Yadkori et al., 2011)	$\tilde{O}(d^2\sqrt{T})$	$\tilde{O}(d^2\sqrt{T})$	$\tilde{O}(d^2\sqrt{T})$	
ESTR (Jun et al., 2019)	$\tilde{O}(\sqrt{\frac{rdT}{\lambda_{\min}(Q(\pi))}}\left(\frac{\lambda_1}{\lambda_r}\right)^3)$	-	-	Bilinear
ε-FALB (Jang et al., 2021)	$\tilde{O}(\sqrt{d^3T})$	-	-	Bilinear & Comp. intractable
rO-UCB (Jang et al., 2021)	$\tilde{O}(\sqrt{rd^3T})$	-	-	Bilinear & Requires oracle
LowLOC (Lu et al., 2021)	$\tilde{O}(\sqrt{rd^3T})$	$\tilde{O}(\sqrt{rd^3T})$	$\tilde{O}(\sqrt{rd^3T})$	Comp. intractable
LowESTR[1] (Lu et al., 2021)	$\tilde{O}(d^{1/4}\sqrt{r\frac{1}{\lambda_{\min}(Q(\pi))^2}T}\left(\frac{S_*}{\lambda_r}\right))$	$\tilde{O}(\sqrt{rd^{5/2}T})$	$\tilde{O}(\sqrt{rd^{13/2}T})$	
G-ESTT (Kang et al., 2022)	$\tilde{O}(d^{1/4}\sqrt{rdMT}\left(\frac{S_*}{\lambda_r}\right))$	$\tilde{O}(\sqrt{rd^{5/2}T})$	-[2]	
Lower bound (Lu et al., 2021)	$\Omega(rd\sqrt{T})$			
LPA-ETC (Algorithm 6)	$\tilde{O}((R_{\max}r^2B_{\min}(\mathcal{A})T^2)^{1/3})$	$\tilde{O}(r^{2/3}d^{2/3}T^{2/3})$	$\tilde{O}(r^{2/3}dT^{2/3})$	
LPA-ESTR (Algorithm 7)	$\tilde{O}(d^{1/4}\sqrt{B_{\min}(\mathcal{A})T}\left(\frac{S_*}{\lambda_r}\right)))$	$\tilde{O}(\sqrt{d^{5/2}T})$	$\tilde{O}(\sqrt{d^{7/2}T})$	

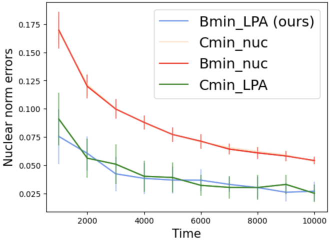
- We clarified the arm-set dependent constants of other algorithms.
- $\lambda_r := \lambda_{\min}(\Theta^*)$ in this table, for notational convenience.

Table of Contents

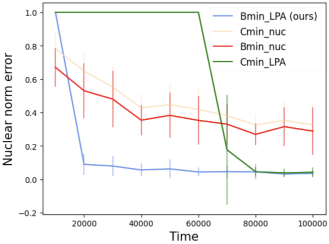
Estimation error

- The results on the $\|\cdot\|_{\text{nuc}}$ recovery error (y-axis) as a function of the sample size (x-axis).
  - **The prefix (Cmin, Bmin)**: the experimental design.
  - **The suffix (LPA, nuc)**: the estimation method (LowPopArt and $\|\cdot\|_{\text{nuc}}$-RLS, respectively.)

$\mathcal{A}$ from $\texttt{Unif}(\mathcal{B}_{Frob}(1))$



$\mathcal{A} = \mathcal{A}_{\text{hard}}$

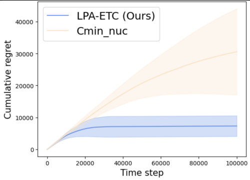


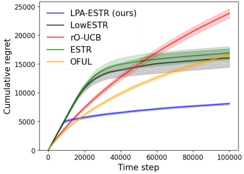
Bmin and **LPA** generally outperform **Cmin** and **nuc**, respectively.

Bandit experiments

- Trend of cumulative regret as time step increases
- In any situation, LowPopArt based algorithms work better than SOTA algorithms.

ETC: LPA vs $\|\cdot\|_{nuc}$	ESTR: Synthetic	ESTR: Real data





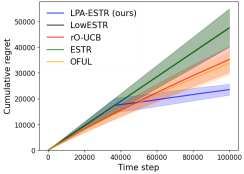


Table of Contents

Conclusion

- A novel low-rank estimation algorithm called LowPopArt
  - utilizes the geometry of the arm set to perform the estimation differently than conventional approaches.
- A novel experimental design for LowPopArt.
- Two new low-rank bandit algorithms based on LowPopArt , improving the dimensionality dependence in regret bounds.

Future work

- Designing general algorithms that can match guarantees in specialized settings [7, 1].
- Establishing tight regret lower bound that depends on the geometry of the arm set in the low-rank bandit problem.

Reference I

[1] B. Huang, K. Huang, S. Kakade, J. D. Lee, Q. Lei, R. Wang, and J. Yang. Optimal gradient-based algorithms for non-concave bandit optimization. *Advances in Neural Information Processing Systems*, 34:29101–29115, 2021.

[2] Y. Jedra, W. Réveillard, S. Stojanovic, and A. Proutiere. Low-rank bandits via tight two-to-infinity singular subspace recovery. *arXiv preprint arXiv:2402.15739*, 2024.

[3] K.-S. Jun, R. Willett, S. Wright, and R. Nowak. Bilinear Bandits with Low-rank Structure. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 97, pages 3163–3172, 2019.

Reference II

[4] Y. Kang, C.-J. Hsieh, and T. C. M. Lee. Efficient frameworks for generalized low-rank matrix bandit problems. *Advances in Neural Information Processing Systems*, 35:19971–19983, 2022.

[5] S. Katariya, B. Kveton, C. Szepesvári, C. Vernade, and Z. Wen. Bernoulli Rank-1 Bandits for Click Feedback. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2001–2007, 2017.

[6] W. Kotlowski and G. Neu. Bandit Principal Component Analysis. In A. Beygelzimer and D. Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1994–2024, Phoenix, USA, 2019. PMLR.

Reference III

[7] T. Lattimore and B. Hao. Bandit Phase Retrieval, 2021.

[8] T. Lattimore and C. Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[9] Y. Lu, A. Meisami, and A. Tewari. Low-rank generalized linear bandit problems. In *International Conference on Artificial Intelligence and Statistics*, pages 460–468. PMLR, 2021.

[10] S. Minsker. Sub-gaussian estimators of the mean of a random matrix with heavy-tailed entries. *The Annals of Statistics*, 46(6A):2871–2903, 2018.

[11] C. Trinh, E. Kaufmann, C. Vernade, and R. Combes. Solving bernoulli rank-one bandits with unimodal thompson sampling. In *Algorithmic Learning Theory*, pages 862–889. PMLR, 2020.

Thank you!

Email: ksajks@gmail.com

Web: http://jajajang.github.io