On High-dimensional and Low-rank Tensor Bandits

Chengshuai Shi, Cong Shen, and Nicholas D. Sidiropoulos
University of Virginia
Charlottesville, VA 22904, USA
{cs7ync, cong, nikos}@virginia.edu

***Abstract*—Most existing studies on linear bandits focus on a one-dimensional characterization of the overall system. While being representative, this formulation may fail to model applications with high-dimensional but favorable structures, such as the low-rank tensor representation for recommender systems. To address this limitation, this work studies a general tensor bandits model, where actions and system parameters are represented by tensors as opposed to vectors, and we particularly focus on the case that the unknown system tensor is low-rank. A novel bandit algorithm, coined `TOFU` (Tensor Optimism in the Face of Uncertainty), is developed. `TOFU` first leverages flexible tensor regression techniques to estimate low-dimensional subspaces associated with the system tensor. These estimates are then utilized to convert the original problem to a new one with norm constraints on its system parameters. Lastly, a norm-constrained bandit subroutine is adopted by `TOFU`, which utilizes these constraints to avoid exploring the entire high-dimensional parameter space. Theoretical analyses show that `TOFU` improves the best-known regret upper bound by a multiplicative factor that grows exponentially in the system order. A novel performance lower bound is also established, which further corroborates the efficiency of `TOFU`.**

I. Introduction

The multi-armed bandits (MAB) framework [2], [3] has attracted growing interest in recent years as it can characterize a broad range of applications requiring sequential decision-making. An active research area in MAB is linear bandits [4], [5], where the actions are characterized by feature vectors. While being representative, this one-dimensional (i.e., vectorized) formulation may fail to capture practical applications with high-dimensional but favorable structures. We use the recommender system model to illustrate this limitation. An online shopping platform needs an effective advertising mechanism for its products. However, instead of only deciding which item to promote (as typically considered in standard linear bandits studies), the marketer also needs to consider many other factors. For example, the marketer may plan where to place to promotion (e.g., on the sidebar or as a pop-up) and how to highlight the promotion (e.g., emphasizing the discounts or the product quality). The overall strategy with all these factors will determine the effectiveness of this promotion.

Traditional recommendation strategies often leverage tensor formulations to capture the joint decisions concerning many associated factors [6]–[8]. However, as mentioned, existing bandits strategies are largely restricted to vectorized systems. Although vectorizing multi-dimensional systems can preserve element-wise information, structural information is often lost. Especially, as recognized in [6]–[8], tensors formulated to characterize recommender systems often process the attractive property of *low-rankness* which, however, no longer exists in the vectorized systems and thus cannot be exploited.

A complete version can be found in [1]. The work of CSs was supported in part by the US National Science Foundation (NSF) under awards ECCS-2029978, ECCS-2143559, CNS-2002902, the Virginia Commonwealth Cyber Initiative (CCI) smart cities project, and the Bloomberg Data Science Ph.D. Fellowship. The work of NS was partially supported by NSF IIS-1908070.

TABLE I
RELATED WORKS AND REGRET COMPARISONS

Algorithm	Regret
Vectorized LinUCB [4]	$\tilde{O}(d^N\sqrt{T})$
Matricized ESTT/ESTS [10]	$\tilde{O}(d^{\lceil \frac{N}{2} \rceil} r^{\lfloor \frac{N}{2} \rfloor}\sqrt{T})$
Tensor Elim. [11]; modified to general actions	$\tilde{O}(d^{N-1}r\sqrt{T})$
`TOFU` (Corollary 1)	$\tilde{O}(d^2 r^{N-2}\sqrt{T})$
Lower bound (Theorem 2)	$\Omega(r^N\sqrt{T})$

The time horizon is T. The considered system tensor is order-N and of size $(d, d, \cdots, d)$. It also has a multi-linear rank $(r, r, \cdots, r)$, where $r \leq d$.

In this work, we study a general problem of tensor bandits for online decision-making, which extends the standard one-dimensional setting of linear bandits to a multi-dimensional and multi-linear one. In particular, each action is represented by a tensor (as opposed to a vector), and the mean reward of playing an action is the inner product between its feature tensor and an unknown system tensor. Then, motivated by various practical problems, a low-rank assumption is imposed on the system tensor, and this work aims at leveraging the low-rank knowledge to facilitate bandit learning. The main contributions are summarized in the following.

- The studied tensor bandits framework is general in the sense that it does not have restrictions on the system dimension and the action structure, which contributes to the generalization of linear bandits and extends the applicability of the MAB study; see Appendix A for related works.
- A novel learning algorithm, `TOFU` (Tensor Optimism in the Face of Uncertainty), is proposed for the challenging problem of low-rank tensor bandits. `TOFU` adopts flexible designs of tensor regressions to estimate low-dimensional subspaces associated with the unknown system tensor. Then, these estimates are utilized to convert the original problem into a new one, where the low-rank property is transformed into the knowledge of norm constraints on the system parameters. `TOFU` finally adopts the LowOFUL subroutine [9] to incorporate these norm constraints in bandit learning to avoid exploring the entire high-dimensional parameter space.
- Theoretical analyses demonstrate the effectiveness and efficiency of `TOFU` with performance guarantees. In particular,

the regret of TOFU improves the best-known regret upper bound by a multiplicative factor of order $O((d/r)^{\lceil N/2\rceil-2})$, where N is the order of the considered system tensor, d is the length of its modes, and $r \leq d$ denotes its multi-linear rank. Note that this improvement becomes more significant in high-dimensional problems, i.e., growing exponentially w.r.t. N. A novel regret lower bound is further established, and TOFU is shown to be sub-optimal only up to a factor of $O((d/r)^2)$, which does not scale with N. The baselines and the main results are summarized in Table I.

II. Problem Formulation

A. Preliminaries on Tensors

An order-N tensor $\mathcal{Y} \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_N}$ has $\prod_{n\in[N]} d_n$ elements and can be viewed as a hyper-rectangle with edges (referred to as modes) of lengths $(d_1, d_2, \cdots, d_N)$ (see [12], [13] for comprehensive reviews). The tensor elements are identified to by their indices along each mode, e.g., $\mathcal{Y}_{i_1,i_2,\cdots,i_N}$ denotes the $(i_1, i_2, \cdots, i_N)$-th element of $\mathcal{Y}$, while a block is denoted by the index set of its contained elements, e.g., the block $\mathcal{Y}_{I_1,I_2,\cdots,I_N}$ represents the elements with indices $(i_1, i_2, \cdots, i_N) \in I_1 \times I_2 \times \cdots \times I_N$. Moreover, fibers are one-dimensional sections of a tensor (as rows and columns in a matrix); thus an order-N tensor has N types of fibers.

Tensor operations. The inner product between tensor $\mathcal{Y}$ and a same-shape tensor $\mathcal{B} \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_N}$ is the sum of the products of their elements:

$$\langle \mathcal{B}, \mathcal{Y}\rangle = \sum_{i_1\in[d_1]}\sum_{i_2\in[d_2]}\cdots\sum_{i_N\in[d_N]} \mathcal{B}_{i_1,i_2,\cdots,i_N}\mathcal{Y}_{i_1,i_2,\cdots,i_N}.$$

The Frobenius norm is then defined as $\|\mathcal{Y}\|_F := \sqrt{\langle \mathcal{Y}, \mathcal{Y}\rangle}$.

The mode-n (matrix) product $\mathcal{Y} \times_n B$ between tensor $\mathcal{Y}$ and matrix $B \in \mathbb{R}^{d'_n\times d_n}$ outputs an order-N tensor of size $(d_1, \cdots, d_{n-1}, d'_n, d_{n+1}, \cdots, d_N)$ with elements:

$$(\mathcal{Y} \times_n B)_{i_1,\cdots,i_{n-1},i'_n,i_{n+1},\cdots,i_N} = \sum_{i_n\in[d_n]} B_{i'_n,i_n}\mathcal{Y}_{i_1,\cdots,i_n,\cdots,i_N}.$$

In addition, matricization is the process of reordering tensor elements into a matrix. The mode-n matricization of tensor $\mathcal{Y}$ is denoted as $\mathcal{M}_n(\mathcal{Y})$, whose columns are mode-n fibers of tensor $\mathcal{Y}$ and dimensions are $(d_n, \prod_{n'\in[N]/\{n\}} d_{n'})$. Similarly, vectorization converts a tensor to a vector with all its elements, which is denoted as vec$(\mathcal{Y})$ for tensor $\mathcal{Y}$.

Tucker decomposition. Similarly to matrices, tensor decomposition is a useful tool to characterize the structure of tensors. In this work, we mainly focus on the Tucker decomposition illustrated as follows: for tensor $\mathcal{Y}$, with r_n denoting the rank of its mode-n matricization, i.e., $r_n = \text{rank}(\mathcal{M}_n(\mathcal{Y}))$, and U_n the corresponding left singular vectors of $\mathcal{M}_n(\mathcal{Y})$, there exists a core tensor $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times\cdots\times r_N}$ such that

$$\mathcal{Y} = \mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 \cdots \times_N U_N =: \mathcal{G} \times_{n\in[N]} U_n,$$

which can be denoted as $\mathcal{Y} = [[\mathcal{G}; U_1, \cdots, U_N]]$, and the tuple $(r_1, \cdots, r_N)$ is called the multi-linear rank of tensor $\mathcal{Y}$.

Additional notations. Typically, lowercase characters (e.g., x) stand for scalars while vectors are denoted with bold lowercase characters (e.g., $\boldsymbol{x}$). Capital characters (e.g., X) are used for matrices, and calligraphic capital characters (e.g., $\mathcal{X}$) for tensors. In addition, $\|\cdot\|_2$ denotes the Euclidean norm for vectors and the spectral norm for matrices; for a vector $\boldsymbol{y}$ and a matrix Γ, we denote $\|\boldsymbol{y}\|_\Gamma := \sqrt{\boldsymbol{y}^\top\Gamma\boldsymbol{y}}$.

B. Tensor Bandits

This work considers the following multi-dimensional bandit problem. At each time step $t \in [T]$, the player has access to an action set $\mathbb{A}_t \subseteq \mathbb{R}^{d_1\times d_2\times\cdots\times d_N}$, i.e., the elements are tensors of size $(d_1, d_2, \cdots, d_N)$. She needs to select one action $\mathcal{A}_t$ from the set $\mathbb{A}_t$, and this action would bring her a reward of

$$r_t = \langle \mathcal{A}_t, \mathcal{X}\rangle + \varepsilon_t, \quad (1)$$

where $\mathcal{X} \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_N}$ is an unknown tensor of system parameters and ε_t is an independent 1-sub-Gaussian noise. We further denote $\mu_{\mathcal{A}} := \langle \mathcal{A}, \mathcal{X}\rangle$ as the expected reward of action $\mathcal{A}$ and, without loss of generality, assume that $\|\mathcal{X}\|_F \leq C$ for $C > 0$ and $\max\{\|\mathcal{A}\|_F : \mathcal{A} \in \cup_{t\in[T]}\mathbb{A}_t\} \leq 1$.

The agent's objective is to minimize her regret against the per-step optimal actions $\mathcal{A}^*_t := \arg\max_{\mathcal{A}\in\mathbb{A}_t}\langle \mathcal{A}, \mathcal{X}\rangle$ [2]:

$$R(T) := \sum\nolimits_{t\in[T]}(\langle \mathcal{A}^*_t, \mathcal{X}\rangle - \langle \mathcal{A}_t, \mathcal{X}\rangle).$$

C. The Low-rank Structure

It is possible to view the above problem as a $\prod_{n\in[N]} d_n$-dimensional linear bandits problem by vectorizing the action tensor $\mathcal{A}_t$ and the system tensor $\mathcal{X}$, which can then be solved by known algorithms [4], [5]. However, the high-dimensional structures of this system are not preserved by vectorization. Especially, one of the most commonly observed structures in real-world applications (e.g., recommender systems [6]–[8] and healthcare [14]–[17]) is the low-rankness. We give the general multi-linear rank assumption of $\mathcal{X}$ as follows.

Assumption 1. *The unknown system tensor $\mathcal{X}$ has a multi-linear rank of $(r_1, r_2, \cdots, r_N)$ and can be decomposed as $\mathcal{X} = [[\mathcal{G}; U_1, U_2, \cdots, U_N]]$.*

To simplify the notations, in the following, it is assumed that $d_1 = \cdots = d_N = d$ while $r_1 = \cdots = r_N = r$. In practice, the rank r is often much smaller than the mode length d, especially for very large d. Hence, the following problem is at the center of this work: *can bandit algorithms be designed to exploit the low-rank structure of the system tensor?* Especially, the key question is how much performance improvement we can achieve, compared with the naive regret of $\tilde{O}(d^N\sqrt{T})$ [4] that is obtained by directly vectorizing the actions and the system.

Note that the design and analysis can be extended to the general case of $d_1 \neq \cdots \neq d_N$ and $r_1 \neq \cdots \neq r_N$ with minor notation modifications. Also, without loss of generality, it is assumed that N is of order $O(1)$ (i.e., a constant) and $N \geq 3$.

III. The TOFU Algorithm

The TOFU algorithm (presented in Alg. 1) has two phases: A and B. Phase A aims at estimating the unknown system tensor $\mathcal{X}$ up to a certain precision, especially its low-dimensional subspaces. With this estimate, the original bandit problem can be

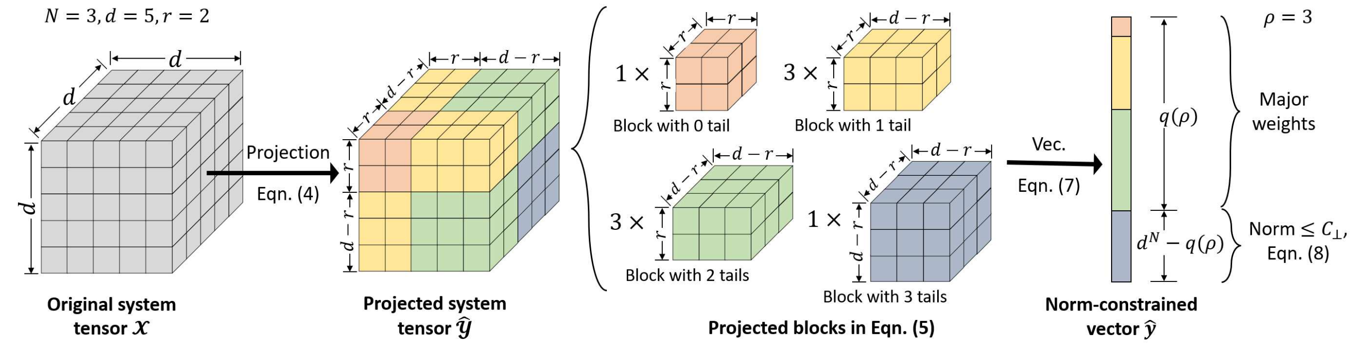

Fig. 1. An illustration of the projection in Eqn. (4), the blocks with varying amounts of tails in Eqn. (5), and the vectorization in Eqn. (7), where an order-3 tensor (i.e., $N = 3$) is adopted as an example with $d = 5$ and $r = 2$. The projection is performed with the low-dimensional subspaces estimated in Phase A (see Eqn. (4)) and the projected system tensor is shown to have blocks with zero to N tails. The value $q(\rho)$ is the number of elements in the projected blocks with less than ρ tails as specified in Eqn. (6), and here an input $\rho = 3$ is adopted which results in $q(3) = d^3 - (d - r)^3$. The norm constraint in Eqn. (8) is on the other $d^N - q(\rho)$ elements, i.e., the projected blocks with at least ρ tails. This constraint is leveraged in Phase B to avoid exploring the entire high-dimensional parameter space. Note that here with $N = 3$ and $\rho = 3$, the designed norm constraint is only on the block with three tails, while with a larger N, the constraint will cover more blocks if still using $\rho = 3$ (as in Corollary 1), e.g., blocks with three and four tails for $N = 4$.

reformulated, such that the new problem has (approximately) a small number of effective system parameters because the other parameters have small norms. Then, in Phase B, an OFU (optimism in the face of uncertainty)-style subroutine is adopted to solve this norm-constrained problem.

A. Phase A: Estimating Low-dimensional Subspaces

Phase A adopts techniques in low-rank tensor regression (also known as low-rank tensor factorization or completion from linear measurements) [17]–[21]. Especially, it considers the problem of estimating a low-rank tensor $\mathcal{X}$ by a collection of data $\{(\mathcal{A}_t, r_t) : t \in [T_1]\}$ that are associated with $\mathcal{X}$ through Eqn. (1), where T_1 is the amount of collected data samples.

Using the bandits terminology, Phase A is designed to last T_1 steps, during which a dataset of T_1 data samples is collected. With such a dataset, an estimate of $\mathcal{X}$, denoted as $\hat{\mathcal{X}}$, can be obtained via low-rank tensor regression techniques. From another perspective, Phase A can be interpreted as using forced explorations to estimate the system tensor $\mathcal{X}$.

Clearly, the estimation quality is related to the collected data, especially the selected arms and the noises. Also, different designs of low-rank tensor regression require different data collection procedures. To provide a general discussion and ease the presentation, we denote the adopted tensor regression algorithm as $\texttt{TRalg}(\cdot)$ and consider the following assumption:

Assumption 2. *The dataset* $\mathcal{D}_A = \{(A_t, r_t) : t \in [T_1]\}$ *and the tensor regression algorithm* $\texttt{TRalg}(\cdot)$ *are such that the output* $\hat{\mathcal{X}} \leftarrow \texttt{TRalg}(\mathcal{D}_A)$ *satisfies* $\|\hat{\mathcal{X}} - \mathcal{X}\|_F \leq \eta(T_1)$ *for a problem-dependent function* $\eta(T_1)$.

Under this assumption, regret bounds can be established to depend on the generic function $\eta(T_1)$. Specific dataset configurations and tensor regression algorithms can be incorporated to establish concrete forms of $\eta(T_1)$, which leads to the corresponding problem-dependent regret bounds. Examples of datasets and algorithms that satisfy Assumption 2 with a high probability can be found in Examples 1 and 2 in Sec. IV with $\eta(T_1) = \tilde{O}(\sqrt{d^N(dr + r^N)/T_1})$.

B. From Subspace Estimates to Norm Constraints

Intuitively, the estimated $\hat{\mathcal{X}}$ and its decomposition matrices $(\hat{U}_1, \hat{U}_2, \cdots, \hat{U}_N)$ from Phase A should help the task of bandit learning. To achieve this goal, the following projection is generalized from matrix bandits [9]. In particular, a new arm $\hat{\mathcal{B}}$ can be constructed from the original arm $\mathcal{A}$ as follows:

$$\hat{\mathcal{B}} = \mathcal{A} \times_{n\in[N]} [\hat{U}_n, \hat{U}_{n,\perp}]^\top \in \mathbb{R}^{d\times d\times\cdots\times d}, \tag{2}$$

where $\hat{U}_{n,\perp}$ is a set of orthogonal basis in the complementary subspace of $\hat{U}_n$ and $[\cdot,\cdot]$ denotes the concatenation of two matrices. In other words, Eqn. (2) projects the actions to the estimated low-dimensional subspaces and their complements.

After some algebraic manipulations, we can establish that

$$\mu_{\mathcal{A}} = \langle \mathcal{A}, \mathcal{X} \rangle = \langle \hat{\mathcal{B}}, \hat{\mathcal{Y}} \rangle \tag{3}$$

where $\hat{\mathcal{Y}} \in \mathbb{R}^{d\times d\times\cdots\times d}$ is a projected system tensor defined as

$$\begin{aligned} \hat{\mathcal{Y}} &:= \mathcal{X} \times_{n\in[N]} [\hat{U}_n, \hat{U}_{n,\perp}]^\top \\ &= \mathcal{G} \times_{n\in[N]} ([\hat{U}_n, \hat{U}_{n,\perp}]^\top U_n). \end{aligned} \tag{4}$$

Thus, the original tensor bandits problem can be reformulated with the action set $\hat{\mathbb{B}}_t := \{\hat{\mathcal{B}} = \mathcal{A} \times_{n\in[N]} [\hat{U}_n, \hat{U}_{n,\perp}]^\top : \mathcal{A} \in \mathbb{A}_t\}$ and the system tensor $\hat{\mathcal{Y}}$ defined above. While this problem still has d^N elements and the system tensor $\hat{\mathcal{Y}}$ is still unknown, it possesses norm constraints on elements in many blocks of $\hat{\mathcal{Y}}$. In other words, the above projection is capable of turning the low-rank property into the knowledge of parameter norms, which are specified in the following.

Especially, if $\hat{U}_n$ is estimated precisely enough, we can guarantee that $\|\hat{U}_{n,\perp}^\top U_n\|_2$ is relatively small. In particular, under Assumption 2, it holds that $\|\hat{U}_{n,\perp}^\top U_n\|_2 = \tilde{O}(\eta(T_1))$ (see Lemma 1). Then, with a closer look at the projected tensor $\hat{\mathcal{Y}}$, the following observation can be made: elements in many blocks are close to zero. In particular, the block

$$\begin{aligned} &\hat{\mathcal{Y}}_{\underbrace{:r,:r,\cdots,:r}_{N-k \text{ modes}},\underbrace{r+1:,r+1:,\cdots,r+1:}_{k \text{ modes}}} \\ &= \mathcal{G} \times_{n\in[N-k]} (\hat{U}_n^\top U_n) \underbrace{\times_{n'\in[N-k+1:N]} (\hat{U}_{n',\perp}^\top U_{n'})}_{\text{with } k \text{ tails}}, \end{aligned} \tag{5}$$

has a norm that scales with $\tilde{O}((\eta(T_1))^k)$ (see Lemma 2), where the notation $: r$ denotes the set $[r]$ while $r+1:$ represents the set $[r+1 : d]$ (thus the above block denotes the $r^{N-k}(d-r)^k$ tensor elements with indices $(i_1, \cdots, i_{N-k}, i_{N-k+1}, \cdots i_N) \in [r] \times \cdots \times [r] \times [r+1, d] \times \cdots \times [r+1, d]\}$). This property holds similarly for other symmetrical blocks. As $\eta(T_1)$ typically decays with T_1 (because the estimation quality should increase with more data samples), the norm of the above block will become smaller as the length of Phase A increases, which can be captured by a norm constraint that will be described later.

To ease the exposition, we refer to the above block and its symmetrical ones as blocks with k tails, meaning the indices of their elements have k modes in the interval $[r+1 : d]$ (i.e., the tail). An illustration of these blocks in an order-3 tensor is provided in Fig. 1. Furthermore, the number of tensor elements in blocks with less than k tails is denoted as

$$q(k) := \sum\nolimits_{i=0}^{k-1} \binom{N}{i} r^{N-i}(d-r)^i, \tag{6}$$

which is an important quantity in later designs and analyses.

Remark 1. Compared with previous works on matrix and tensor bandits [9]–[11], [24], the essence of this work is the observation that norm constraints commonly exist for blocks with different numbers of tails. In particular, [11] directly extends [9], [24] and only leverages the norm constraint on the block with N tails. Instead, Section IV will illustrate that the norm constraints on blocks with at least three tails can be leveraged together under a suitable $\eta(T_1)$, which then leads to the obtained performance improvement.

Algorithm 1 `TOFU`

Input: T; rank r; dimension N and d; tensor regression alg. `TRalg`; length of Phase A T_1; confidence parameter δ; tails ρ

1: Sample $\mathcal{A}_t \in \mathbb{A}_t$ following the arm selection rule required by `TRalg`$(\cdot)$ and observe reward r_t, for $t \in [T_1]$ ▷ *Phase A*
2: Estimate $\hat{\mathcal{X}} = [[\hat{\mathcal{G}}; \hat{U}_1, \cdots, \hat{U}_N]]$ with `TRalg` using $\mathcal{D}_A = \{(\mathcal{A}_t, r_t) : t \in [T_1]\}$, i.e., $\hat{\mathcal{X}} \leftarrow$ `TRalg`$(\mathcal{D}_A)$
3: Set $C_\perp, \lambda, \lambda_\perp$ as in Theorem 1 ▷ *Phase B*
4: Initialize $\Lambda(\rho) \leftarrow \mathrm{diag}(\lambda, \cdots, \lambda, \lambda_\perp, \cdots, \lambda_\perp)$, where the first $q(\rho)$ elements are λ; $\Psi_{T_1} \leftarrow \{\boldsymbol{y} \in \mathbb{R}^{d^N} : \|\boldsymbol{y}\|_2 \le C\}$
5: **for** $t = T_1 + 1, \cdots, T$ **do**
6: Set $\hat{\mathbb{B}}_t \leftarrow \{\mathcal{B}_t = \mathcal{A}_t \times_{n\in[N]} [\hat{U}_n, \hat{U}_{n,\perp}]^\top : \mathcal{A}_t \in \mathbb{A}_t\}$
7: Get $\hat{\boldsymbol{b}}_t \leftarrow \arg\max_{\hat{\boldsymbol{b}}_t \in \mathrm{vec}(\hat{\mathbb{B}}_t)} \max_{\boldsymbol{y} \in \Psi_{t-1}} \langle \hat{\boldsymbol{b}}_t, \boldsymbol{y} \rangle$
8: Pull arm $\mathcal{A}_t$ corresponding to $\hat{\boldsymbol{b}}_t$ and obtain reward r_t
9: Update $\hat{B}_t$ with rows $\{\boldsymbol{b}_\tau^\top : \tau \in (T_1, t]\}$
10: Update $\boldsymbol{r}_t$ with elements $\{r_\tau : \tau \in (T_1, t]\}$
11: Update $V_t \leftarrow \Lambda(\rho) + \hat{B}_t^\top \hat{B}_t$ and $\bar{\boldsymbol{y}} \leftarrow V_t^{-1} \hat{B}_t^\top \boldsymbol{r}_t$
12: Update $\sqrt{\beta_t} \leftarrow \sqrt{\log(\frac{\det(V_t)}{\det(\Lambda(\rho))\delta^2})} + \sqrt{\lambda} C + \sqrt{\lambda_\perp} C_\perp$
13: Update $\Psi_t \leftarrow \{\hat{\boldsymbol{y}} \in \mathbb{R}^{d^N} : \|\hat{\boldsymbol{y}} - \bar{\boldsymbol{y}}\|_{V_t} \le \sqrt{\beta_t}\}$
14: **end for**

C. Phase B: Solving the Norm-constrained Linear Bandits

As illustrated above, after the projection, norm constraints can be obtained on some blocks of tensor $\hat{\mathcal{Y}}$. For flexibility, we consider that Phase B aims to leverage such constraints on blocks with at least ρ tails, which contain $d^N - q(\rho)$ elements. The parameter ρ is an input with its value in $[N]$ that requires careful designs to balance losses from two phases and will be specified in Sec. IV (e.g., selected as $\rho = 3$ in Corollary 1). Equivalently, there exist norm constraints on parts of the elements in the unknown vector

$$\hat{\boldsymbol{y}} := \mathrm{vec}(\hat{\mathcal{Y}}) \in \mathbb{R}^{d^N}. \tag{7}$$

If the vectorization of $\hat{\mathcal{Y}}$ is performed first on the block with zero tail and then gradually on those with one and more tails (see Fig. 1 for an example), we can compactly express the norm constraint on blocks with at least ρ tails as

$$\|\hat{\boldsymbol{y}}_{q(\rho)+1:d^N}\|_2 \le C_\perp, \tag{8}$$

where the parameter $C_\perp$ will be specified later in Theorem 1. This condition can be interpreted as that there are approximately only $q(\rho)$ effective parameters in $\hat{\boldsymbol{y}}$ while the other parameters are nearly ignorable due to their constrained norm.

Then, a *norm-constrained linear bandits* problem with d^N parameters needs to be solved. In particular, the action set is $\Phi_t := \mathrm{vec}(\hat{\mathbb{B}}_t) \subseteq \mathbb{R}^{d^N}$ at step t, where $\mathrm{vec}(\hat{\mathbb{B}}_t) := \{\mathrm{vec}(\hat{\mathcal{B}}) : \hat{\mathcal{B}} \in \hat{\mathbb{B}}_t)\}$, and the expected reward for action $\hat{\boldsymbol{b}} \in \Phi_t$ is $\langle \hat{\boldsymbol{b}}, \hat{\boldsymbol{y}} \rangle$. Additionally, an important norm constraint on $\hat{\boldsymbol{y}}$, i.e., Eqn. (8), is available to the learner. Inspired by [25], the LowOFUL algorithm is designed in [9] to tackle such norm-constrained linear bandits. Especially, a weighted regularization is performed to estimate the system parameter: at time step t, the following estimate $\bar{\boldsymbol{y}}$ of $\hat{\boldsymbol{y}}$ is obtained as $\bar{\boldsymbol{y}} \leftarrow \arg\min_{\boldsymbol{y}} \|\hat{B}_t \boldsymbol{y} - \boldsymbol{r}_t\|_2^2 + \|\boldsymbol{y}\|_{\Lambda(\rho)}^2 = V_t^{-1} \hat{B}_t^\top \boldsymbol{r}_t$, where matrix $\hat{B}_t \in \mathbb{R}^{t \times d^N}$ is constructed with previous action vectors $\{\hat{\boldsymbol{b}}_\tau : \tau \in (T_1, t]\}$ as rows, vector $\boldsymbol{r}_t \in \mathbb{R}^t$ has elements $\{r_\tau : \tau \in (T_1, t]\}$, matrix $\Lambda(\rho) = \mathrm{diag}(\lambda, \cdots, \lambda, \lambda_\perp, \cdots, \lambda_\perp)$ (with λ as the first $q(\rho)$ elements and $\lambda_\perp$ as the others), and $V_t = \Lambda(\rho) + \hat{B}_t^\top \hat{B}_t$. Then, an OFU-style arm-selection subroutine is adopted (lines 5–14 of Alg. 1).

Remark 2. To better understand the projection performed in Eqns. (2) and (4), an ideal scenario is considered where the decomposition matrices $(U_1, \cdots, U_N)$ are *exactly* known. Then, the projected action $\hat{\mathcal{B}}$ and system parameter $\hat{\mathcal{Y}}$ both match their "exact" versions $\mathcal{B} = \mathcal{A} \times_{n=1}^N [U_n, U_{n,\perp}]^\top$ and $\mathcal{Y} = \mathcal{G} \times_{n\in[N]} ([U_n, U_{n,\perp}]^\top U_n) = \mathcal{G} \times_{n\in[N]} ([I_r, \mathbf{0}_{r\times(d-r)}]^\top)$. Although $\mathcal{Y}$ has d^N elements, there are only r^N non-zero ones in $\mathcal{G}$. However, for $\hat{\mathcal{Y}}$ projected via the imperfect estimates $(\hat{U}_1, \cdots, \hat{U}_N)$, we can only guarantee some blocks of elements have small norms instead of being exact nulls as in $\mathcal{Y}$.

IV. Theoretical Analysis

In this section, we formally establish the theoretical guarantee of the `TOFU` algorithm. First, the following assumption is adopted on the minimum singular value of the matricized system tensor, which is commonly used in the study of matrix bandits [9], [10], [24] and tensor bandits [22].

Assumption 3. *It holds that* $\min_{n\in[N]}\{\omega_{\min}(\mathcal{M}_n(\mathcal{X}))\} \ge \omega$ *for some parameter* $\omega > 0$*, where* $\omega_{\min}(\cdot)$ *returns the minimum positive singular value of a matrix.*

Then, the following regret upper bound can be established.

Theorem 1. *Under Assumptions 1, 2 and 3, with probability at least* $1-\delta$*, using* $\rho \in [N]$ *as input and* $\lambda = C^{-2}$*,* $\lambda_\perp = \frac{T}{q(\rho)\log(1+T/\lambda)}$*,* $C_\perp = 2^{N/2}C(\eta(T_1))^\rho\omega^{-\rho}$*, if* T_1 *is chosen such that* $\eta(T_1) \leq \omega$*, the regret of* `TOFU` *can be bounded as*

$$R(T) \leq \tilde{O}\Big(CT_1 + d^{\rho-1}r^{N-\rho+1}\sqrt{T} + C(\eta(T_1))^\rho\omega^{-\rho}T\Big).$$

It is worth noting that this theorem applies to any tensor regression technique satisfying Assumption 2 and any input ρ, which demonstrates the flexibility of `TOFU`. Furthermore, the above regret bound has three terms. The first term characterizes the dataset collection in Phase A. The second term represents the learning loss from the $q(\rho)$ major elements in Phase B. The third one is from the other $d^N - q(\rho)$ elements, which are nearly ignorable but still contribute to the regret.

According to function $\eta(T_1)$, parameters ρ and T_1 should be carefully selected such that the overall regret in Theorem 1 is minimized. Two specific tensor regression techniques from [22], [23] are considered to instantiate $\eta(T_1)$: the first one is established with the selected arms having sub-Gaussian elements, while the second selects random one-hot tensors as arms. To avoid complicated expressions, confidence parameters δ_1, δ_2, threshold parameters ι_1, ι_2 and scale parameters c_1, c_2 are adopted in the following, whose values are independent of T_1 and can be found in the corresponding references.

Example 1 (Section 4.2 of [22]). *If* $T_1 > \iota_1$*, all elements of* $\mathcal{A}_t$ *are i.i.d. drawn from* $1/d^N$*-sub-Gaussian distributions, and* ε_t *is an independent standard Gaussian noise, with probability at least* $1-\delta_1$*, an estimate* $\hat{\mathcal{X}} = [[\hat{\mathcal{G}}; U_1, \cdots, U_N]]$ *can be obtained from the tensor regression algorithm proposed in [22] such that* $\|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 \leq c_1 d^N(dr + r^N)/T_1$*.*

Example 2 (Corollary 2 of [23]). *If* $T_1 > \iota_2$*,* $\mathcal{A}_t$ *is a random one-hot tensor, and* ε_t *is an independent* 1*-sub-Gaussian noise, with probability at least* $1-\delta_2$*, an estimate* $\hat{\mathcal{X}} = [[\hat{\mathcal{G}}; U_1, \cdots, U_N]]$ *can be obtained from the tensor regression algorithm proposed in [23] such that* $\|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 \leq c_2 d^N(dr + r^N)/T_1$*.*

In these examples, it can be seen that Assumption 2 holds with a high probability for $\eta(T_1) = \tilde{O}(\sqrt{d^N(dr + r^N)/T_1})$. Then, Theorem 1 leads to the following corollary.

Corollary 1. *Under Assumptions 1 and 3, if the conditions in Example 1 (resp. Example 2) can be satisfied in Phase A, using the tensor regression algorithm from [22] (resp. [23]) as* `TRalg`$(\cdot)$*, the parameters from Theorem 1 with input* $\rho = 3$*, and the following length for Phase A (resp. with* ι_2, c_2*)*

$$T_1 = \max\left\{\iota_1, c_1 d^N(dr+r^N)\omega^{-2}, c_1^{\frac{3}{5}}d^{\frac{3N}{5}}(dr+r^N)^{\frac{3}{5}}\omega^{-\frac{6}{5}}T^{\frac{2}{5}}\right\},$$

with probability at least $1-\delta-\delta_1$ *(resp.* $1-\delta-\delta_2$*), the regret of* `TOFU` *can be bounded as*

$$R(T) \leq \tilde{O}\Big(CT_1 + d^2r^{N-2}\sqrt{T}\Big)$$

The above corollary adopts $\rho = 3$, i.e., the norm constraint in Eqn. (8) is on blocks with at least three tails. This choice is conscious with respect to the function $\eta(T_1)$ from Examples 1 and 2 as it lays aside as many parameters as possible without letting them negatively impact the bandit learning. In particular, with this choice, the length T_1 can be optimized as in Corollary 1 (which is of order $O(T^{2/5})$) and thus the dominating term (regarding the T-dependency) of the regret in Corollary 1 is the last one of order $\tilde{O}(d^2r^{N-2}\sqrt{T})$.

This obtained regret of order $\tilde{O}(d^2r^{N-2}\sqrt{T})$ is compared with several existing results in the following (see also Table I). First, if directly adopting linear bandits algorithms such as LinUCB [4] on the vectorized system, a regret of order $\tilde{O}(d^N\sqrt{T})$ would incur as the low-rank structure is not used. A second approach is to matricize the system and adopt algorithms for matrix bandits [9], [10], [24]. The state-of-the-art ESTT/ESTS [10] can then achieve a regret of order $\tilde{O}(d^{\lceil\frac{N}{2}\rceil}r^{\lfloor\frac{N}{2}\rfloor}\sqrt{T})$ (see Appendix E), which is still inefficient as matricization does not preserve all the structure information. At last, for [11] on tensor bandits, if we modify it to have general (instead of one-hot) tensors as actions, a regret of order $\tilde{O}(d^{N-1}r\sqrt{T})$ occurs as it does not fully consider the high-dimensional benefits (see Remark 1). Thus, compared with the best existing regret of order $\tilde{O}(d^{\lceil\frac{N}{2}\rceil}r^{\lfloor\frac{N}{2}\rfloor}\sqrt{T})$, `TOFU` has an improvement of a multiplicative factor of order $\tilde{O}((d/r)^{\lceil\frac{N}{2}\rceil-2})$, which grows exponentially in N. Hence, this benefit becomes more significant in higher-order problems.

While `TOFU` improves existing results, we further compare it against the following new regret lower bound.

Theorem 2. *Assume* $r^N \leq 2T$ *and for all* $t \in [T]$*, let* $\mathbb{A}_t = \mathbb{A} := \{\mathcal{A} \in \mathbb{R}^{d\times d\times\cdots\times d} : \|\mathcal{A}\|_F \leq 1\}$ *and* ε_t *be a sequence of independent standard Gaussian noise. Then, for any policy, there exists a system tensor* $\mathcal{X} \in \mathbb{R}^{d\times d\times d\times\cdots\times d}$ *with a multilinear rank* $(r, r, \cdots, r)$ *and* $\|\mathcal{X}\|_F^2 = O(r^{2N}/T)$ *such that* $\mathbb{E}_{\mathcal{X}}[R(T)] = \Omega(r^N\sqrt{T})$*, where the expectation is taken with respect the interaction of the policy and the system.*

Compared with this lower bound, `TOFU` is sub-optimal only up to an additional $O((d/r)^2)$ factor (which does not scale with N). We conjecture that a slightly tighter regret lower bound of order $\Omega(dr^{N-1}\sqrt{T})$ can be established, which reduces to that of $\Omega(dr\sqrt{T})$ in matrix bandits ($N = 2$) [24].

V. Conclusions

This work studied a general tensor bandits problem, where high-dimensional tensors characterize action and system parameters. Motivated by practical applications, the system tensor is modeled to be low-rank. To tackle this high-dimensional but low-rank problem, a novel algorithm named `TOFU` was proposed. `TOFU` adopts tensor regression techniques to estimate low-dimensional subspaces associated with the system tensor. The obtained estimates are then used to transform the challenging problem of low-rank tensor bandits into an equivalent but easier one of norm-constrained linear bandits. The theoretical analysis provided a regret guarantee of `TOFU`, which is shown to be exponentially more efficient than existing results. A novel performance lower bound was also established, further demonstrating the superiority of `TOFU`.

References

[1] C. Shi, C. Shen, and N. D. Sidiropoulos, "On high-dimensional and low-rank tensor bandits," *arXiv preprint arXiv:2305.02441*, 2023.

[2] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.

[3] S. Bubeck, N. Cesa-Bianchi *et al.*, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends® in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[4] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári, "Improved algorithms for linear stochastic bandits," *Advances in neural information processing systems*, vol. 24, 2011.

[5] W. Chu, L. Li, L. Reyzin, and R. Schapire, "Contextual bandits with linear payoff functions," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 208–214.

[6] E. Frolov and I. Oseledets, "Tensor methods and recommender systems," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, no. 3, p. e1201, 2017.

[7] P. Symeonidis and A. Zioupos, *Matrix and tensor factorization techniques for recommender systems*. Springer, 2016, vol. 1.

[8] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 2, pp. 1–44, 2016.

[9] K.-S. Jun, R. Willett, S. Wright, and R. Nowak, "Bilinear bandits with low-rank structure," in *International Conference on Machine Learning*. PMLR, 2019, pp. 3163–3172.

[10] Y. Kang, C.-J. Hsieh, and T. C. M. Lee, "Efficient frameworks for generalized low-rank matrix bandit problems," in *Advances in Neural Information Processing Systems*, 2022.

[11] J. Zhou, B. Hao, Z. Wen, J. Zhang, and W. W. Sun, "Stochastic low-rank tensor bandits for multi-dimensional online decision making," *arXiv e-prints*, pp. arXiv–2007, 2020.

[12] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[13] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.

[14] H. Zhou, L. Li, and H. Zhu, "Tensor regression with applications in neuroimaging data analysis," *Journal of the American Statistical Association*, vol. 108, no. 502, pp. 540–552, 2013.

[15] X. Li, D. Xu, H. Zhou, and L. Li, "Tucker tensor regression and neuroimaging analysis," *Statistics in Biosciences*, vol. 10, no. 3, pp. 520–545, 2018.

[16] B. Yaman, S. Weingärtner, N. Kargas, N. D. Sidiropoulos, and M. Akçakaya, "Low-rank tensor models for improved multidimensional mri: Application to dynamic cardiac t_1 mapping," *IEEE transactions on computational imaging*, vol. 6, pp. 194–207, 2019.

[17] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and M. Akçakaya, "Tensor completion from regular sub-nyquist samples," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1–16, 2019.

[18] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse problems*, vol. 27, no. 2, p. 025010, 2011.

[19] P. Jain and S. Oh, "Provable tensor factorization with missing data," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[20] A. R. Zhang, Y. Luo, G. Raskutti, and M. Yuan, "Islet: Fast and optimal low-rank tensor regression via importance sketching," *SIAM journal on mathematics of data science*, vol. 2, no. 2, pp. 444–479, 2020.

[21] T. Ahmed, H. Raja, and W. U. Bajwa, "Tensor regression using low-rank and sparse tucker decompositions," *SIAM Journal on Mathematics of Data Science*, vol. 2, no. 4, pp. 944–966, 2020.

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

[23] M. Valko, R. Munos, B. Kveton, and T. Kocák, "Spectral bandits for smooth graph functions," in *International Conference on Machine Learning*. PMLR, 2014, pp. 46–54.

[24] R. Han, R. Willett, and A. R. Zhang, "An optimal statistical and computational framework for generalized tensor estimation," *The Annals of Statistics*, vol. 50, no. 1, pp. 1–29, 2022.

[25] D. Xia, M. Yuan, and C.-H. Zhang, "Statistically optimal and computationally efficient low rank tensor completion from noisy entries," *The Annals of Statistics*, vol. 49, no. 1, 2021.

[26] K. Jang, K.-S. Jun, S.-Y. Yun, and W. Kang, "Improved regret bounds of bilinear bandits using action space analysis," in *International Conference on Machine Learning*. PMLR, 2021, pp. 4744–4754.

[27] T. Idé, K. Murugesan, D. Bouneffouf, and N. Abe, "Targeted advertising on social networks using online variational tensor regression," *arXiv preprint arXiv:2208.10627*, 2022.