Quasi Polynomial and Interpolative Models for Tensor Approximation

Jeongmin Chae†, Selin Bac⋆, Usama Saleem⋆, Shaama Mallikarjun Sharada⋆ and Urbashi Mitra†
† Department of Electrical and Computer Engineering, University of Southern California
⋆ Department of Chemical Engineering and Materials Science, University of Southern California

***Abstract*—A novel Tucker decomposition based tensor approximation is considered herein: observations based on only a few lateral slices and side structural information of a true tensor are exploited. This work is motivated by quantum chemistry problems wherein full Hessian computation is expensive, but partial computation is available. The proposed method successfully estimates the quasi-polynomial and interpolative structure of frontal and lateral slices given *a priori* knowledge of the true tensor. A theoretical error bound is provided, which characterizes the impact due to errors in the side information. To the best of our knowledge, this work proposes the first tensor approximation with side information and interpolation.**

***Index Terms*—Tensor approximation; Tucker decomposition; Interpolation; quasi-polynomial**

I. Introduction

Multidimensional data, or *tensors*, including medical images, sensor images and experimental data are becoming increasingly prevalent in domains such as neuroscience, quantum chemistry, machine learning and image processing [1]–[3]. If there is noisy or missing data within the tensor, one can exploit low-rank tensor completion/approximation methods where typically it is assumed that one has access to randomly collected slices, fibers or entries of the true tensor [4]–[6]. Recent work in magnetic resonance images (MRI), traffic estimation and quantum chemistry suggest that alternative sampling strategies may be more relevant [7]–[9]. Our motivating application is from quantum chemistry. Reaction rates are fundamental to catalyst design, optimizing industrial processes, and understanding biological mechanisms. To evaluate quantum tunneling effects [10], the computation of Hessian matrices along the pathway connecting reactants and products are essential [11], but computationally intensive. Models for this Hessian evolution do not currently exist. A challenge to overcome is that one cannot compute random samples of the Hessian tensor, but rather single Hessian matrices at one reaction point.

In this work, we develop a low rank Tucker based tensor approximation exploiting only a few lateral slices coupled with side information which is described by a quasi-polynomial slice model as well as an interpolative model for the evolution of the slices over the reaction. While our sampling framework is motivated by problems in quantum chemistry [12], [13], where full tensor computation is expensive, the approach can be also be applied to aerospace and network traffic [14], [15]. While [4], [5] share some features of our proposed framework (slice samples), they need **additional randomly sampled fibers** [4], [5] and thus cannot be applied to our quantum chemistry problem. In [7], knowledge of the relationship between sampled and unsampled slices is assumed. To the best of our knowledge, tensor approximation with lateral slices and structural information captured by a quasi- polynomial and interpolative structure has not been previously investigated.

Contributions. We propose a new framework for tensor approximation given the structure previously mentioned. The proposed algorithm, dubbed *quasi Polynomial and Interpolative Tensor Approximation (PITA)*, successfully integrates a known interpolative and quasi-polynomial structure between frontal and lateral slices of a true tensor. A theoretical error bound is provided. Simulation studies validate the capabilities of PITA outperforming other strategies.

II. Preliminaries

We employ the following notation. Tensors and matrices are denoted by $\mathcal{X}$ and $\mathbf{X}$ respectively. Slices are second-order tensors generated by fixing all modes except two. For a third order tensor $\mathcal{X}$, the slices $\mathcal{X}(:,:,k)$, $\mathcal{X}(:,j,:)$, and $\mathcal{X}(i,:,:)$ are called *frontal, lateral* and *horizontal* slices, respectively. We next introduce key definitions to be used throughout the paper.

Definition 1. (Mode-n Tensor unfolding and folding) *Given a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, we denote the mode-$n$ unfolding of $\mathcal{X}$ as $\mathbf{X}_{(n)}$, whose (i_n, j)-th element in the lexicographical order is mapped from the $(i_1, i_2, \ldots, i_N)$-th entry of tensor $\mathcal{X}$, where $j = 1 + \sum_{1 \le l \le N, l \neq n} (i_l - 1) J_l$ with $J_l = \Pi_{1 \le t \le l-1, t \neq n} I_t$. The inverse operator to express the tensor $\mathcal{X}$, is defined as $\mathcal{X} = fold_n\left(\mathbf{X}_{(n)}\right)$.*

Definition 2. (Tucker decomposition [16]) *Let $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ be a given data tensor, then the Tucker decomposition model is given by $\mathcal{X} = \mathcal{S} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \cdots \times_N \mathbf{A}_N$, where $\mathcal{S} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$ is called the core tensor, and the matrices $\mathbf{A}_n \in \mathbb{R}^{I_n \times R_n}$, $R_n \le I_n$, $n = 1, 2, \ldots, N$ are called* factor matrices*; they are multiplied by the core tensor $\mathcal{S}$ along each mode n. In short, we have $\mathcal{X} = (\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N) \cdot \mathcal{S}$.*

For a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and a matrix $\mathbf{B} \in \mathbb{R}^{J \times I_n}$, the tensor-matrix product along mode n is defined by $\mathcal{X} \times_n \mathbf{B} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$. In the unfolding form, $\mathcal{Y} = \mathcal{X} \times_n \mathbf{B} \iff \mathbf{Y}_{(n)} = \mathbf{B}\mathbf{X}_{(n)} \in \mathbb{R}^{J \times \Pi_{1 \le j \le N, j \neq n} I_j}$.

After straightforward manipulation of tensor relationships and unfolding of tensors, we have the following results: if $\mathcal{X} = \mathcal{S} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \cdots \times_N \mathbf{A}_N$,

$$\text{vec}\left(\mathcal{X}\right) = \left(\otimes_{n=1}^{N} \mathbf{A}_n\right) \text{vec}\left(\mathcal{S}\right) \tag{1}$$

$$\mathbf{X}_{(n)} = \mathbf{A}_n \mathbf{S}_{(n)} \left(\otimes_{i=1, i\neq n}^{N} \mathbf{A}_i\right)^{\mathsf{T}}, \tag{2}$$

where $\otimes_{n=1}^{N} \mathbf{A}_n \triangleq \mathbf{A}_N \otimes \cdots \otimes \mathbf{A}_1$ and $\mathbf{A} \otimes \mathbf{B}$ denotes Kronecker product of $\mathbf{A}$ and $\mathbf{B}$.

III. Problem formulation & optimization

A. System model

True tensor. Let $\mathcal{H} \in \mathbb{R}^{m \times n \times m}$ be the ground truth tensor of $n \leq m$ different Hessian matrices of dimension $m \times m$, where the lateral slices represent Hessian matrices generated from the chemical reaction systems. We assume the ground truth tensor $\mathcal{H}$ admits the following Tucker decomposition, for $1 \leq i_1, i_3 \leq m$ and $1 \leq i_2 \leq n$,

$$\mathcal{H}(i_1, i_2, i_3) \approx \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} \mathbf{X}_1(i_1, j_1)\, \mathbf{X}_2(i_2, j_2) \cdot \mathbf{X}_3(i_3, j_3)\, \mathcal{G}(j_1, j_2, j_3), \tag{3}$$

namely, $\mathcal{H} \approx (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) \cdot \mathcal{G}$, where $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ is the core tensor of multilinear rank $\mathbf{r} = (r_1, r_2, r_3)$, $r_k = \text{rank}(\mathbf{H}_{(k)})$ for $k = 1, 2, 3.$, and $\mathbf{X}_1 \in \mathbb{R}^{m \times r_1}$, $\mathbf{X}_2 \in \mathbb{R}^{n \times r_2}$, $\mathbf{X}_3 \in \mathbb{R}^{m \times r_3}$ are factor matrices of each mode. We shall fix the ground truth factors $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ as orthonormal matrices consisting of left singular vectors in each mode. The lateral slice sampling operation is defined as follows. Let $\mathbf{\Psi}$ be the sampling matrix for lateral slices, which samples d lateral slices from a total of n. Let $\mathcal{I} = \{i_1, \ldots, i_d\} \subset [n]$ denote the set of sampled lateral slice indices. Clearly $|\mathcal{I}| = d$. Then, $\mathbf{\Psi} \in \{0, 1\}^{d \times n}$ is $\mathbf{\Psi} := \mathbf{I}_{\mathcal{I}}$, where $\mathbf{I}$ is the identity matrix of dimension n and the notation $\mathbf{I}_{\mathcal{I}}$ means that we consider the sub-matrix of $\mathbf{I}$ formed by its rows indexed by entries in the set $\mathcal{I}$. Thus, a tensor of d sampled lateral slices $\mathcal{C} \in \mathbb{R}^{m \times d \times m}$ can be expressed as $\mathcal{C} = \text{fold}_2\left(\mathbf{\Psi}\mathbf{H}_{(2)}\right)$.

Quasi-polynomial structure. As per Section I, we have observed from numerical data that the ground truth tensor $\mathcal{H}$ exhibits a quasi-polynomial structure across all frontal slices. To describe this, given a list of reaction coordinate values, $\mathbf{s} = [s_1, s_2, \ldots, s_n]$ and polynomial order l, we assume that the true tensor $\mathcal{H}$ admits the following model

$$\mathcal{H} = \mathcal{Q} \times_1 \mathbf{I}_1 \times_2 \mathbf{S} \times_3 \mathbf{I}_2 + \mathcal{E}_{QS}, \tag{4}$$

where $\mathcal{E}_{QS} \in \mathbb{R}^{m \times n \times m}$ is an unknown perturbation matrix, and $\|\mathcal{E}_{QS}\|_F \leq \epsilon_1$. $\mathbf{S} \in \mathbb{R}^{n \times l}$ is a polynomial basis matrix and l indicates the degree of the polynomial. $\mathcal{Q} \in \mathbb{R}^{m \times l \times m}$ is the corresponding polynomial coefficient tensor. Both $\mathbf{I}_1 \in \mathbb{R}^{m \times m}$ and $\mathbf{I}_2 \in \mathbb{R}^{m \times m}$ are identity matrices. Note that our true tensor $\mathcal{H}$ in (4) admits a perturbed quasi-polynomial Tucker decomposition model, where the coefficient tensor $\mathcal{Q}$ serves as a core tensor, and $\mathbf{S}$, $\mathbf{I}_1$ and $\mathbf{I}_2$ correspond to a factor matrix along each mode. An individual element in the core tensor $\mathcal{Q}$ serves as the weight of the associated combination of polynomial basis in $\mathbf{S}$.

Interpolative structure. We next provide the interpolative structure that connects the lateral slices (a.k.a Hessian matrices) of $\mathcal{H}$. Given the set of d sampled lateral slice indices $\mathcal{I} = \{i_1, \ldots, i_d\} \subset [n]$ and the tensor of sampled lateral slices $\mathcal{C} \in \mathbb{R}^{m \times d \times m}$, we adopt the following model:

$$\mathcal{H} = \mathcal{C} \times_2 \mathbf{\Lambda}_{\mathcal{I}} + \mathcal{E}_{\mathcal{C}}, \tag{5}$$

where $\mathbf{\Lambda}_{\mathcal{I}} \in [0, 1]^{n \times d}$ is a structured interpolation coefficient matrix with respect to $\mathcal{I}$. Let λ_j for $j \in [n]/\mathcal{I}$ be interpolation coefficients that control interpolation between two sampled lateral slices.[1] $\mathbf{\Lambda}_{\mathcal{I}}$ models the change as the reaction is traversed. $\mathcal{E}_{\mathcal{C}}$ is an unknown perturbation tensor with respect to $\mathcal{C}$. The quantities $\mathbf{\Lambda}_{\mathcal{I}}$ and $\mathcal{E}_{\mathcal{C}}$ are defined depending on $\mathcal{I}$. Notice that the interpolative decomposition in (5) approximates the tensor $\mathcal{H}$ using only a few of its sampled lateral slices, $\mathcal{C}$. As described in Section I, $\mathcal{E}_{\mathcal{C}}$ and $\mathbf{\Lambda}_{\mathcal{I}}$ are unknown due to the nature of our chemistry application – the mathematical model for evolution of Hessian matrices is unknown. Solving for $\mathbf{\Lambda}_{\mathcal{I}}$ is an interesting future research direction. In this work, we treat the coefficients of $\mathbf{\Lambda}_{\mathcal{I}}$ as hyper-parameters.

B. Optimization

We next introduce the proposed optimization strategy. The proposed algorithm is comprised of three stages: (i) estimating the unknown polynomial coefficient tensor, $\mathcal{Q}$, leveraging the quasi-polynomial structure of frontal slices (ii) subsequently estimating the interpolative tensor between lateral slices and lastly, (iii) the final tensor estimation step via gradient descent constrained to the quasi-polynomial approximations and linear interpolation obtained previously.

Quasi-polynomial structure estimation. We first estimate the polynomial coefficient tensor $\mathcal{Q}$ using $\mathcal{C} = \text{fold}_2\left(\mathbf{\Psi}\mathbf{H}_{(2)}\right)$. We argue that as long as enough independent lateral slices are sampled, the following optimization gives us a good estimate for $\mathcal{Q}$, $\hat{\mathcal{Q}} = \text{argmin}_{\bar{\mathcal{Q}}} \left\|\mathbf{C}_{(2)} - \mathbf{\Psi}\left(\bar{\mathcal{Q}} \times_2 \mathbf{S}\right)_{(2)}\right\|_F^2$, where $\mathbf{C}_{(2)}$ is the mode-2 unfolding of $\mathcal{C}$. This is a standard tensor regression problem that admits a closed form solution, where $\mathcal{Q} := \text{fold}_{(2)}\left(\mathbf{\Psi}^{\dagger}\mathbf{C}_{(2)}\right) \times_2 \mathbf{S}^{\dagger}$, where $\dagger$ denotes Moore–Penrose inverse.

Interpolative tensor estimation. Next, in addition to the quasi-polynomial structure, we perform the interpolative tensor estimation to incorporate the structure between lateral slices into our low rank tensor completion step. Given the tensor of sampled lateral slices $\mathcal{C}$, our goal is to solve for $\mathbf{\Lambda}_{\mathcal{I}}$. However, solving for $\mathbf{\Lambda}_{\mathcal{I}}$ (5) explicitly is impossible due to the unknown mathematical relationship between lateral slices as explained

[1] For example, when $n = 4$ and $\mathcal{I} = \{1, 4\}$, namely $d = |\mathcal{I}| = 2$, $\mathbf{\Lambda}_{\mathcal{I}} \in \mathbb{R}^{4 \times 2}$ is defined as

$$\mathbf{\Lambda}_{\mathcal{I}} \in \mathbb{R}^{4 \times 2} = \begin{bmatrix} 1 & 0 \\ \lambda_2 & 1 - \lambda_2 \\ \lambda_3 & 1 - \lambda_3 \\ 0 & 1 \end{bmatrix}. \tag{6}$$

Algorithm 1 PITA Algorithm

Input: $\mathcal{C} \in \mathbb{R}^{m\times d\times m}$, $\mathbf{S} \in \mathbb{R}^{n\times l}$, $\mathbf{\Psi} \in \{0,1\}^{d\times n}$
Parameters: Degree of polynomial l, Step size η, Max iteration T, target rank r_1,r_2,r_3
Initialization: Generate each entry of $\hat{\mathcal{G}}$, $\hat{\mathbf{X}}_1$, $\hat{\mathbf{X}}_2$ and $\hat{\mathbf{X}}_3$ independently from $\mathcal{N}(0,1)$
Polynomial coefficient tensor estimation : Obtain the closed form solution as $\hat{\mathcal{Q}} := \text{fold}_{(2)}\left(\mathbf{\Psi}^{\dagger}\mathbf{C}_{(2)}\right) \times_2 \mathbf{S}^{\dagger}$.
Interpolative tensor estimation Design $\hat{\mathbf{\Lambda}}_{\mathcal{I}}$
Low rank tensor approximation: Using $\hat{\mathcal{Q}} \times_2 \mathbf{S}$ and $\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}$ obtained from the previous steps, for $t \in [T]$, update $\hat{\mathcal{G}}^t$ and $\hat{\mathbf{X}}_n^t$, $n = 1,2,3$ via gradient descent method where the gradients are defined in (8) and (9) respectively.
  Obtain $\hat{\mathcal{G}} = \hat{\mathcal{G}}^{T+1}$, $\hat{\mathbf{X}}_n = \hat{\mathbf{X}}_n^{T+1}$ for $n = 1,2,3$
Output: $\hat{\mathcal{H}} = \hat{\mathcal{G}} \times_1 \hat{\mathbf{X}}_1 \times_2 \hat{\mathbf{X}}_2 \times_3 \hat{\mathbf{X}}_3$.

in Section I. In fact, diverse samples of $\mathcal{H}(:,k,:)$, $k \notin \mathcal{I}$, can be synthesized from $\mathbf{\Lambda}_{\mathcal{I}}$ by adjusting λ_j for $j \in [n]/\mathcal{I}$. In this work, we set λ_j to be a hyper-parameter and control for the interpolation structure between lateral slices.

Low-rank tensor completion. Finally, we exploit the quasi-polynomial structure in (4) and the interpolative tensor estimates (5) to obtain a low-rank Tucker decomposition-based tensor approximation of $\mathcal{H}$. In what follows, we present the algorithm for representing low-rank tensors of $\mathcal{H}$ in terms of Tucker decomposition in (3), $\mathcal{H} = \mathcal{G} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \mathbf{X}_3$, where $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times r_3}$, $r_1, r_3 \le m$ and $r_2 \le n$, is a low rank core tensor with rank$(\mathcal{G}) \le [r_1, r_2, r_3]$, and $\mathbf{X}_1 \in \mathbb{R}^{m\times r_1}$, $\mathbf{X}_2 \in \mathbb{R}^{n\times r_2}$ and $\mathbf{X}_3 \in \mathbb{R}^{m\times r_3}$ are the factor matrices to be optimized. Leveraging the quasi-polynomial structure, $\hat{\mathcal{Q}} \times_2 \mathbf{S}$, and the interpolative structure, $\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}$, obtained from the previous step, we solve the following non-convex tensor regression problem

$$\underset{\hat{\mathcal{G}},\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3}{\text{argmin}} \ \alpha \left\|\hat{\mathcal{Q}} \times_2 \mathbf{S} - \hat{\mathcal{G}} \times_1 \hat{\mathbf{X}}_1 \times_2 \hat{\mathbf{X}}_2 \times_3 \hat{\mathbf{X}}_3\right\|_F^2 + (1-\alpha) \left\|\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}} - \hat{\mathcal{G}} \times_1 \hat{\mathbf{X}}_1 \times_2 \hat{\mathbf{X}}_2 \times_3 \hat{\mathbf{X}}_3\right\|_F^2, \tag{7}$$

where $0 < \alpha < 1$ is another hyper-parameter, which controls the trade-off between the quasi-polynomial and interpolative structures. The proposed algorithm iteratively updates the core tensor $\mathcal{G}$ and factor matrices in an alternating fashion according to the order of :$\mathcal{G}$, $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathcal{G}$, $\mathbf{X}_1$, $\ldots$, $\mathbf{X}_2$, $\mathbf{X}_3$ via gradient descent. Define $f(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ as $f(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) = \alpha \left\|\hat{\mathcal{Q}} \times_2 \mathbf{S} - \mathcal{G} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \mathbf{X}_3\right\|_F^2 + (1-\alpha)\left\|\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}} - \mathcal{G} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \mathbf{X}_3\right\|_F^2$. Then, according to (1), the gradient of $f(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ with respect to $\mathcal{G}$ is given by $\text{vec}\left(\nabla f_{\mathcal{G}}(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\right) = -2\alpha \left(\otimes_{n=1}^3 \mathbf{X}_n^{\mathsf{T}}\right) \cdot \text{vec}\left(\hat{\mathcal{Q}} \times_2 \mathbf{S}\right) - 2(1-\alpha)\left(\otimes_{n=1}^3 \mathbf{X}_n^{\mathsf{T}}\right) \text{vec}\left(\hat{\mathcal{C}} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}\right) + 2\left(\otimes_{n=1}^3 \mathbf{X}_n^{\mathsf{T}}\mathbf{X}_n\right) \text{vec}(\mathcal{G})$. The size of the reformulation for the Kronecker products in $\text{vec}\left(\nabla f_{\mathcal{G}}(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\right)$ is extremely large. Hence, from (1), we have,

System	PITA	Poly.	Interp.	QPMA [9]
CF3CH3	**0.0254**	0.048	0.042	0.3081
TSoxo	**0.0318**	0.0424	0.0322	0.3623
Sn2Ar1	**0.0053**	0.0068	0.0063	0.0272
Ir	**0.0271**	0.0601	0.0333	0.1275

TABLE I
ALGORITHM COMPARISON FOR DIFFERENT REACTIONS.

$$\begin{aligned}\nabla f_{\mathcal{G}}(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) &= -2\alpha\left(\hat{\mathcal{Q}} \times_2 \mathbf{S}\right) \times_1 \mathbf{X}_1^{\mathsf{T}} \times_2 \mathbf{X}_2^{\mathsf{T}} \times_3 \mathbf{X}_3^{\mathsf{T}} \\ &- 2(1-\alpha)\left(\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}\right) \times_1 \mathbf{X}_1^{\mathsf{T}} \times_2 \mathbf{X}_2^{\mathsf{T}} \times_3 \mathbf{X}_3^{\mathsf{T}} \\ &+ 2\mathcal{G} \times_1 \mathbf{X}_1^{\mathsf{T}}\mathbf{X}_1 \times_2 \mathbf{X}_2^{\mathsf{T}}\mathbf{X}_2 \times_3 \mathbf{X}_3^{\mathsf{T}}\mathbf{X}_3. \end{aligned} \tag{8}$$

We then repeat the following update rule at each iteration $t \in [T]$ until convergence $\hat{\mathcal{G}}^{t+1} = \hat{\mathcal{G}}^t - \eta \ \nabla f_{\hat{\mathcal{G}}}(\hat{\mathcal{G}}^t, \hat{\mathbf{X}}_1^t, \hat{\mathbf{X}}_2^t, \hat{\mathbf{X}}_3^t)$, where η is a learning rate. Next, we compute the gradient of $f(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ with respect to $\mathbf{X}_n$ for $n = 1,2,3$. According to (2), we have $f(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) = \alpha \left\|\left(\hat{\mathcal{Q}} \times_2 \mathbf{S}\right)_{(n)} - \mathbf{X}_n \mathbf{G}_{(n)} \left(\otimes_{i=1, i\neq n}^3 \mathbf{X}_i\right)^{\mathsf{T}}\right\|_F^2 + (1-\alpha) \left\|\left(\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}\right)_{(n)} - \mathbf{X}_n \mathbf{G}_{(n)} \left(\otimes_{i=1, i\neq n}^3 \mathbf{X}_i\right)^{\mathsf{T}}\right\|_F^2$. Hence, the gradient of $f(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ with respect to $\mathbf{X}_n$, $\nabla f_{\mathbf{X}_n}(\mathcal{G}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$, is obtained as,

$$\begin{aligned} &-2\alpha\left(\hat{\mathcal{Q}} \times_2 \mathbf{S}\right)_{(n)} \mathbf{B}_n^{\mathsf{T}} - 2(1-\alpha)\left(\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}\right)_{(n)} \mathbf{B}_n^{\mathsf{T}} \\ &-2\,\mathbf{X}_n\left(\mathbf{B}_n \mathbf{B}_n^{\mathsf{T}}\right), \text{ where } \mathbf{B}_n = \mathbf{G}_{(n)}\left(\otimes_{i=1, i\neq n}^3 \mathbf{X}_i\right)^{\mathsf{T}} \end{aligned} \tag{9}$$

Here, we do not explicitly reformulate the Kronecker product in $\mathbf{B}_n$, but let $\mathcal{T} = \mathcal{G} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \mathbf{X}_3$. Then, we have, $\mathbf{B}_n = \mathbf{T}_{(n)}$, where $\mathbf{T}_{(n)}$ is the mode-n unfolding of $\mathcal{T}$. Finally, we update $\mathbf{X}_n$, for $n = 1,2,3$, at each iteration $t \in [T]$ as follows, $\hat{\mathbf{X}}_1^{t+1} = \hat{\mathbf{X}}_1^t - \eta \ \nabla f_{\hat{\mathbf{X}}_1}(\hat{\mathcal{G}}^{t+1}, \hat{\mathbf{X}}_1^t, \hat{\mathbf{X}}_2^t, \hat{\mathbf{X}}_3^t)$, $\hat{\mathbf{X}}_2^{t+1} = \hat{\mathbf{X}}_2^t - \eta \ \nabla f_{\hat{\mathbf{X}}_2}(\hat{\mathcal{G}}^{t+1}, \hat{\mathbf{X}}_1^{t+1}, \hat{\mathbf{X}}_2^t, \hat{\mathbf{X}}_3^t)$ and $\hat{\mathbf{X}}_3^{t+1} = \hat{\mathbf{X}}_3^t - \eta \ \nabla f_{\hat{\mathbf{X}}_3}(\hat{\mathcal{G}}^{t+1}, \hat{\mathbf{X}}_1^{t+1}, \hat{\mathbf{X}}_2^{t+1}, \hat{\mathbf{X}}_3^t)$. Let $\hat{\mathcal{G}} := \hat{\mathcal{G}}^{T+1}$, $\hat{\mathbf{X}}_1 := \hat{\mathbf{X}}_1^{T+1}$, $\hat{\mathbf{X}}_2 := \hat{\mathbf{X}}_2^{T+1}$ and $\hat{\mathbf{X}}_3 := \hat{\mathbf{X}}_3^{T+1}$. This finally yields the reconstructed low rank Tensor, $\hat{\mathcal{H}} = \hat{\mathcal{G}} \times_1 \hat{\mathbf{X}}_1 \times_2 \hat{\mathbf{X}}_2 \times_3 \hat{\mathbf{X}}_3$. The proposed algorithm is summarized in Algorithm 1.

IV. Error Bound

In this section, we provide an interim bound on the error from solving (7). We assume that $\hat{\mathcal{Q}}$ and $\hat{\mathbf{\Lambda}}_{\mathcal{I}}$ are obtained from the previous quasi-polynomial estimation and interpolative tensor estimation steps, and are thus fixed.

Theorem 1. *Let $\hat{\mathcal{M}} = \mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}$. Assume $r_2 \le \{l, d\}$. With a proper initialization, for some small constant $\epsilon > 0$, and $\eta \le min(4/49, 1/(1+4(1+\epsilon)^{-6}(1+2\epsilon)^2))$, then for all $t \ge 0$, the iterates of Algorithm 1 satisfy*

$$\|(\hat{\mathbf{X}}_1^t, \hat{\mathbf{X}}_2^t, \hat{\mathbf{X}}_3^t) \cdot \hat{\mathcal{G}}^t - \mathcal{H}\|_F \le \epsilon(1-\eta)^t \left(4 + \frac{12\eta\delta}{(1+\epsilon)^3}\right)^t \cdot \sigma_{min}(\mathcal{H}) + \frac{12\eta^2\delta^2}{(1+\epsilon)^6}\left(\sum_{k=0}^{t-1}(1-\eta)^k \left(4 + \frac{12\eta\delta}{(1+\epsilon)^3}\right)^k\right)$$

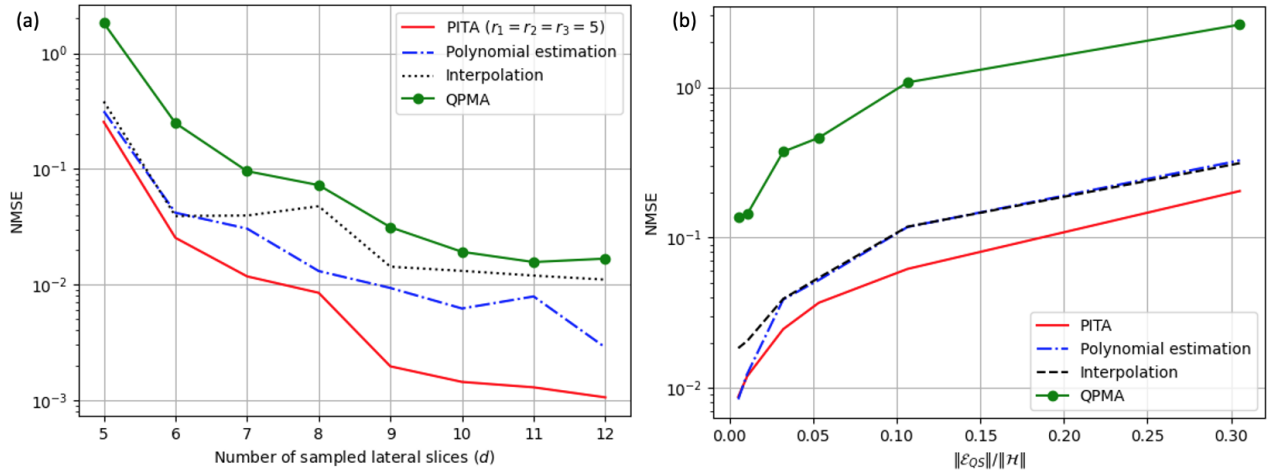

Fig. 1. (a). The performance of PITA varying the number of sampled lateral slices for $r_1 = r_2 = r_3 = 5$. and $l = 5$ (b). Sensitivity to noise ($\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}$) for all algorithms when $l = 5$, $d = 7$ and $r_1 = r_2 = r_3 = 5$.

where $\delta := \alpha\, \sigma_{max}\left(\mathcal{E}_{\mathcal{Q}\mathbf{S}}\right) + (1-\alpha)\, \sigma_{max}\left(\mathcal{E}_{\hat{\mathcal{M}}}\right)$ *and* $\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}} = \hat{\mathcal{Q}} \times_2 \mathbf{S} - \mathcal{H}$ *and* $\mathcal{E}_{\hat{\mathcal{M}}} = \mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}} - \mathcal{H}$ *are the estimation error between* $\hat{\mathcal{Q}} \times_2 \mathbf{S}$ *and* $\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}$ *and the ground truth tensor.*

The proof of Theorem 1 is extensive; we leverage results from [17]. There are key differences that need to be considered: the prior art [17] employs scaled gradient descent, whereas we are able to use classical gradient descent. We also must carefully incorporate the two forms of side information. For the detailed proof and initialization, see [18].

V. Numerical results

All numerical results are averaged over 200 independent runs. For all experiments, we set $m = 50$, $n = 25$, learning rate $\eta = 0.001$, $\alpha = 0.5$ and $l = 5$. We generate the data as follows. To ensure the interpolation structure between lateral Hessian matrices, we generate the first lateral Hessian matrices, $\bar{\mathcal{H}}(:,1,:)$. The remaining lateral Hessians $\bar{\mathcal{H}}(:,i,:)$, $2 \leq i \leq n$, are generated as $\bar{\mathcal{H}}(:,k+1,:) = \bar{\mathcal{H}}(:,k,:) + \mathbf{E}_k$, for $1 \leq k \leq n-1$, where $\mathbf{E}_k$ is generated from $\mathcal{N}(0,\sigma^2), \sigma = 0.01$. By taking the Tucker decomposition with multirank $[r_1, r_2, r_3]$ on this tensor, we obtain the ground truth $\mathcal{G}$, $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$. With this, we obtain the rank-$[r_1, r_2, r_3]$ ground truth matrices, $\mathcal{H} = \mathcal{G} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \mathbf{X}_3$. For the quasi-polynomial structure, $\mathbf{S} \in \mathbb{R}^{n \times l}$, we consider a polynomial basis transform, which is constructed using arbitrarily sampled polynomial coordinate values [9], [19]. The ground truth polynomial coefficient tensor $\mathcal{Q}$ is obtained by $\mathcal{Q} = \mathcal{H} \times_2 \mathbf{S}^{\dagger}$ and $\mathcal{E}_{QS} = \mathcal{H} - \mathcal{Q} \times_2 \mathbf{S}$. We use the normalized mean squared error (NMSE), defined as $\text{NMSE}(\hat{\mathcal{H}}, \mathcal{H}) = \frac{\|\mathcal{H} - \hat{\mathcal{H}}\|_F}{\|\mathcal{H}\|_F}$ to measure the performance.

Benchmark algorithms. We compare the performance of the proposed algorithm with our QPMA [9], Polynomial approximation ($\hat{\mathcal{Q}} \times_2 \mathbf{S}$) and Interpolation ($\mathcal{C} \times_2 \hat{\mathbf{\Lambda}}_{\mathcal{I}}$). As noted in Section I, QPMA approximates each frontal slice with a quasi-polynomial structure with a few number of sampled columns and is a slice-by-slice approach.

Performance analysis. In Fig. 1, we first investigate the effect of increasing the number of observed lateral slices; the true multirank is $r_1 = r_2 = r_3 = 5$. As expected, the NMSE for all algorithms reduces as we increase the number of observed slices. In both the low and high sample regime, PITA has the lowest reconstruction error; thus, the quasi-polynomial and interpolative structure are being well leveraged by PITA. Furthermore, the performance of PITA is investigated for 4 different chemical reactions on real data in Table 1. As shown, PITA achieves the lowest NMSE among all algorithms. It is notable that PITA shows almost 10 times better reconstruction error compared to QPMA (a non-tensor approach). Finally, we illustrate the sensitivity to noise, $\mathcal{E}_{QS}$ for all algorithms in Fig. 1. Notice that the performance of all algorithms gracefully degrade with increasing noise.

VI. Conclusions

We formulated a novel framework for low rank tensor approximation when only a few lateral slices are available. The proposed method, PITA, successfully integrates side structural information (quasi-polynomial and interpolative) of frontal and lateral slices of a true tensor. A bound on the estimation error is derived, which is characterized by the inaccuracy of the side information. Numerical simulations show that PITA effectively captures both quasi-polynomial and interpolative structures compared to a state-of-art method.

Acknowledgements

This work is funded in part by one or more of the following grants: NSF CCF-{1817200, 2008927,2200221, 2311653}, ARO {W911NF1910269, W911NF2410094}, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF RINGS-2148313, NSF A22-2666-S003, ONR 503400-78050, ONR N00014-15-1-2550, ONR N00014-22-1-2363, UCSD/NSF KR 705319, UCD/NSF CCF-2311653, ARMY W911NF1910269, UCD/NSF A25-0139-S009, USC Amazon Trust Center, and NSF PIPP phase II: Center for Pandemic Insights (CPI) 2412522.

REFERENCES

[1] Nicholas D Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E Papalexakis, and Christos Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on signal processing*, vol. 65, no. 13, pp. 3551–3582, 2017.

[2] Jie Lin, Ting-Zhu Huang, Xi-Le Zhao, Yong Chen, Qiang Zhang, and Qiangqiang Yuan, "Robust thick cloud removal for multitemporal remote sensing images using coupled tensor factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[3] Christian F Beckmann and Stephen M Smith, "Tensorial extensions of independent component analysis for multisubject fmri analysis," *Neuroimage*, vol. 25, no. 1, pp. 294–311, 2005.

[4] Michael W Mahoney, Mauro Maggioni, and Petros Drineas, "Tensor-cur decompositions for tensor-based data," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 327–336.

[5] HanQin Cai, Zehan Chao, Longxiu Huang, and Deanna Needell, "Fast robust tensor principal component analysis via fiber cur decomposition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 189–197.

[6] Pan Zhou, Canyi Lu, Zhouchen Lin, and Chao Zhang, "Tensor factorization for low-rank tensor completion," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1152–1163, 2017.

[7] Tatsuya Yokota, Burak Erem, Seyhmus Guler, Simon K Warfield, and Hidekata Hontani, "Missing slice recovery for tensors using a low-rank model in embedded space," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8251–8259.

[8] Yi Chang, Luxin Yan, Xi-Le Zhao, Houzhang Fang, Zhijun Zhang, and Sheng Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE transactions on cybernetics*, vol. 50, no. 11, pp. 4558–4572, 2020.

[9] Jeongmin Chae, Praneeth Narayanamurthy, Selin Bac, Shaama Mallikarjun Sharada, and Urbashi Mitra, "Matrix approximation with side information: When column sampling is enough," *IEEE Transactions on Signal Processing*, 2024.

[10] Antonio Fernandez-Ramos, Benjamin A Ellingson, Bruce C Garrett, and Donald G Truhlar, "Variational transition state theory with multidimensional tunneling," *Reviews in computational chemistry*, vol. 23, pp. 125, 2007.

[11] Donald G Truhlar and Bruce C Garrett, "Variational transition state theory," *Annual Review of Physical Chemistry*, vol. 35, no. 1, pp. 159–189, 1984.

[12] Jeongmin Chae, Praneeth Narayanamurthy, Selin Bac, Shaama Mallikarjun Sharada, and Urbashi Mitra, "Sketched column-based matrix approximation with side information," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 9816–9820.

[13] S. J. Quiton, J. Chae, S. Bac, K. Kron, U. Mitra, and S. M. Sharada, "Toward efficient direct dynamics studies of chemical reactions: A novel matrix completion algorithm," *Journal of Chemical Theory and Computation*, 2022.

[14] Hang Zheng, Zhiguo Shi, Chengwei Zhou, and André LF de Almeida, "Coarray tensor completion for doa estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 5, pp. 5472–5486, 2023.

[15] Kun Xie, Xiangge Wang, Xin Wang, Yuxiang Chen, Gaogang Xie, Yudian Ouyang, Jigang Wen, Jiannong Cao, and Dafang Zhang, "Accurate recovery of missing network measurement data with localized tensor completion," *IEEE/ACM Transactions on Networking*, vol. 27, no. 6, pp. 2222–2235, 2019.

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

[17] Tian Tong, Cong Ma, Ashley Prater-Bennette, Erin Tripp, and Yuejie Chi, "Scaling and scalability: Provable nonconvex low-rank tensor estimation from incomplete measurements," *Journal of Machine Learning Research*, vol. 23, no. 163, pp. 1–77, 2022.

[18] "Supplementary file for quasi polynomial and interpolative models for tensor approximation," https://github.com/JeongminChae/ICASSP2025.

[19] Stephen Jon Quiton, Jeongmin Chae, Selin Bac, Kareesa Kron, Urbashi Mitra, and Shaama Mallikarjun Sharada, "PVMC," Mar. 2022, (accessed 2022-03-07).