Quantized Higher-Order Tensor Recovery by Exploring Low-Dimensional Structures

1st Ren Wang
Rensselaer Polytechnic Institute
Troy, NY, USA
wangr8@rpi.edu, wangm7@rpi.edu

2nd Meng Wang
Rensselaer Polytechnic Institute
Troy, NY, USA
wangm7@rpi.edu

3rd Jinjun Xiong
IBM Thomas J. Watson Research Center
Yorktown Heights, NY, USA
jinjun@us.ibm.com

***Abstract*—This paper considers the recovery of higher-order tensor with intrinsic low-dimensional structure from quantized measurements. By introducing the low CANDECOMP/PARAFAC (CP) rank constraint, we propose nonconvex models for both the general tensor recovery and the recovery of the tensors with tensor singular value decomposition (TSVD). We prove that the recovery errors for both optimization models go to zero when the dimension lengths of tensors go to infinity, and tensors with TSVD can theoretically reach a lower error. This paper also establishes a lower bound for any tensor recovery algorithm. Subsequently, a tensor-based alternating proximal gradient descent algorithm (TBAPGD) and a TSVD-based projected gradient descent algorithm (TSVD-PGD) are proposed to solve the nonconvex optimization problems. We provide a convergence guarantee for the former algorithm, and demonstrate the effectiveness of the latter through simulations. We empirically extend both algorithms to scenarios of missing data and without quantization rule information. Finally, we present experimental results on both synthetic data and real datasets to demonstrate the effectiveness and efficiency of the proposed methods.**

***Index Terms*—higher-order tensor, tensor recovery, low-dimensional structures, quantization, CANDECOMP/PARAFAC (CP) rank**

I. Introduction

Though matrix techniques like low-rank matrix completion [1], [2] and low-rank matrix recovery [3], [4] have been widely used to process data in the real-world problems, some data contain intrinsic correlations that cannot be simply captured by matrices. For example, users might give different ratings to the same object under different contexts [5]. In contrast, higher-order tensors have the capacity to capture the additional correlations, and are leveraged to improve the performances on recovery/completion tasks [6]–[8].

Despite the impressive performance of matrix/tensor recovery/completion on the clean data, they are known to be brittle to low-quality data. In recent years, data quality issues extend to quantization scenarios, such as images with low-color depth [9], highly quantized rating systems [5], and data privacy protection in the power systems [10]. By considering the intrinsic data correlations, quantized matrix recovery techniques are leveraged to address the quantization error issue. 1-bit matrix recovery works consider recover the data from binary measurements, i.e. $\{1, -1\}$ [11], [12]. Then recovery from multi-bit measurements [13] and additional sparse corruptions [14] are further considered to generalize the framework of the quantizated recovery.

Motivated by the existing higher-order correlations within the real-world data and tensors' capacity of capturing additional correlations, it is natural to consider the quantized recovery in the higher-order framework. Early works mainly focus on the 1-bit and convex formulations [15], [16]. However, this convex formulation induces a larger recovery error theoretically and empirically. Such an issue is often mitigated by introducing the exact nonconvex formulation. In our paper, we first study the recovery under the exact rank constraint. In addition, we consider the more general multi-bit scenario.

a) Contributions: Compared to the existing quantized recovery methods, our work makes the following contributions.

- We for the first time solve the quantized tensor recovery problem with exact low-rank constraint in the multi-bit quantization framework.
- We consider both general tensors and a special tensor category named SVD-tensors. The theoretical recovery error bounds for both cases and the fundamental limitation are proved. The results improve the existing theoretical works.
- We propose a generalized quantized tensor recovery scheme in the scenario of unknown boundaries.
- With the aid of recent progress on proximal methods, we propose efficient algorithms that are proved to converge to critical points.
- Our method outperforms the existing works on synthetic data, real image data, and recommender data.

b) Related Work: Based on the needs of both application and theory, matrix/tensor recovery and completion have been widely studied [1]–[4], [6]–[8]. However, datasets in many real-world applications are highly quantized due to the measuring accuracy or storage limitation [10]. The problem of learning to recover the real-valued data from quantized measurements has generated much attention recently [11], [12]. 1-bit matrix completion probably is the first category of works that provides the theoretical recovery guarantee based on the convex relaxation of the original problem [11], [12]. Although the convex relaxation is widely used in solving nonconvex problems that are generally hard to reach the global optimum, the problem has been changed and the recovery error bound might not be tight. To recover the data in the tightest way, some works formulate the problem with the exact

nonconvex constraint [13], [14]. These works also extend the 1-bit framework to multi-bit matrix recovery. The recovery error is guaranteed to be in the order of $O(\sqrt{\frac{\bar{r}}{n}})$, where $\bar{r}$ is the matrix rank and n is the matrix dimension[1].

Motivated by the existences of correlations within different real-world data dimensions, restricting the quantized data recovery problem to the tensor region is further considered to improve the recovery performances [15], [16]. Again, this line of works starts from the most simple formulation, 1-bit and convex [16]. Exact tensor rank constraint is replaced by Tensor M-norm constraint, and the problem can be solved in a convex version. The recovery error is guaranteed to be bounded by $O((\frac{r^{3W-3}W}{n^{W-1}})^{1/4})$, where W is the order of a tensor. In this work, we consider the multi-bit and exact low-rank constraint setting. We also provide an empirically study that generalize the learning unknown boundary case.

c) Preliminaries And Notations: Higher-order tensors ($W > 2$) are denoted by capital calligraphic letters. For example, $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_W}$ represents a W-order tensor with the i-th dimensional size equaling to $n_i, i \in [W]$. $[W]$ is equivalent to the set $\{1, 2, \ldots, W\}$. The $(i_1, i_2, \ldots, i_W)$-th entry of $\mathcal{X}$ is denoted by $\mathcal{X}_{i_1,i_2,\ldots,i_W}$.

The boldface capital letters are used to denote matrices ($W = 2$). $\mathbf{X}_{(w)} \in \mathbb{R}^{n_w \times (n_1 \ldots n_{w-1} n_{w+1} \ldots n_W)}$ is formed by unfolding $\mathcal{X}$ along its w-th dimension, and is called the mode-w matricization of tensor $\mathcal{X}$.

The CANDECOMP/PARAFAC (CP) rank of $\mathcal{X}$ [17] is defined as

$$\begin{aligned}\text{rank}(\mathcal{X}) = \min\{R : \mathcal{X} = \sum_{i=1}^{R} \mathbf{A_1}_i \circ \mathbf{A_2}_i \circ \cdots \circ \mathbf{A_W}_i, \\ \mathbf{A_w} \in \mathbb{R}^{n_w \times R}, h \in [W]\},\end{aligned} \tag{1}$$

where $\mathbf{A_w}$ is the w-th component of the factorization, and $\mathbf{A_w}_i$ is the i-th column of $\mathbf{A_w}$. For W vectors $a_i \in \mathbb{R}^{n_i}, \forall i \in [W]$, $\mathcal{A} = a_1 \circ \cdots \circ a_W$ is a rank-1 W-order tensor with $\mathcal{A}_{i_1,i_2,\ldots,i_W} = a_{1i_1} a_{2i_2} \ldots a_{Wi_W}$. $\circ$ is the outer product. Throughout this paper, we will use $\mathbf{A_1} \circ \mathbf{A_2} \circ \cdots \circ \mathbf{A_W}$ to represent $\sum_{i=1}^{R} \mathbf{A_1}_i \circ \mathbf{A_2}_i \circ \cdots \circ \mathbf{A_W}_i$. when discussing tensor ranks, we refer to CP rank if not otherwise specified.

In this paper, we also consider a set of tensors that have the following form of decomposition

$$\begin{aligned}\mathcal{S}_{\text{tsvd}} := \{\mathcal{X} | \mathcal{X} = \sum_{i=1}^{R} \zeta_i \mathbf{V_1}_i \circ \mathbf{V_2}_i \circ \cdots \circ \mathbf{V_W}_i, \\ \mathbf{V_w} \in \mathbb{R}^{n_w \times R}, \zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_R > 0, \\ \langle \mathbf{V_w}_i, \mathbf{V_w}_j \rangle = \mathbf{1}_{[i=j]}, 1 \leq i, j \leq R, \forall R\}\end{aligned} \tag{2}$$

where $\langle , \rangle$ is the inner product operation. $\mathbf{1}_{[B]}$ denotes an indicator function that only takes value '1' when the event B is true. Definition (2) is the tensor singular value decomposition (TSVD), and generalizes the matrix SVD [18]. One can see that TSVD is a special case of the decomposition form in (1), and Lemma 3.3 in [18] implies that the tensor in (2) has CP rank R. We will frequently use SVD-tensors to represent tensors in $\mathcal{S}_{\text{tsvd}}$. We remark that SVD-tensors are within a subset of the tensor set.

[1]We use the notations $g = O(n)$, $g = \Theta(n)$ if as n goes to infinity, $g \leq c \cdot n$, $c_1 \cdot n \leq g \leq c_2 \cdot n$ eventually holds for some positive constants c, c_1 and c_2 respectively.

II. Problem Formulation

We use $\mathcal{X}^*, \mathcal{N} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_W}$ to denote the W-order tensors containing actual data and noise data, respectively. We assume $\mathcal{X}^*$ is a low-rank tensor and the values are bounded, i.e., $\mathcal{X}^* \leq r$ and $\|\mathcal{X}^*\|_\infty \leq \alpha$, where $r, \alpha > 0$. The values in noise tensor $\mathcal{N}$ are independent and identically distribute (i.i.d) sample realizations from a known cumulative distribution function Φ.

Consider the quantization operator M that maps real values to K discrete values $[K]$ given the intervals formed by boundary pairs $(\omega_0^*, \omega_1^*]; (\omega_1^*, \omega_2^*]; \ldots (\omega_{K-1}^*, \omega_K^*]$. Formally,

$$\begin{aligned}&\mathcal{Q}_{i_1,i_2,\ldots,i_W} = M(\mathcal{X}^*_{i_1,i_2,\ldots,i_W} + \mathcal{N}_{i_1,i_2,\ldots,i_W}) = l \\ &\text{if } \omega^*_{l-1} < \mathcal{X}_{i_1,i_2,\ldots,i_W} + \mathcal{N}_{i_1,i_2,\ldots,i_W} \leq \omega^*_l, \ l \in [K],\end{aligned} \tag{3}$$

where $\omega_0^* = -\infty$ and $\omega_K^* = \infty$. Figure 1 provides a visualization of this quantization operation on a 3-order tensor and $K = 3$. $\mathcal{Q}$ is the quantized matrix. In this paper, we also consider the missing data scenario, i.e., $\mathcal{Q}_\Omega$, where Ω is the observation set satisfying $\frac{|\Omega|}{n_1 n_2 \cdots n_W} \leq 1$.

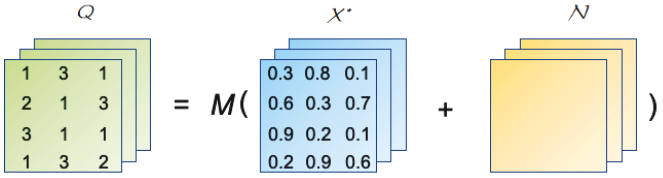


Fig. 1. Visualization of the quantization process on a 3-order tensor ($K = 3$).

Note that the noise tensor $\mathcal{N}$ provides uncertainty measurements. The probability for mapping $\mathcal{X}^*_{i_1,i_2,\ldots,i_W}$ to l is

$$\begin{aligned}&f_l(\mathcal{X}^*_{i_1,i_2,\ldots,i_W}) \\ &= P(\mathcal{Q}_{i_1,i_2,\ldots,i_W} = l | \mathcal{X}^*_{i_1,i_2,\ldots,i_W}) \\ &= \Phi(\omega^*_l - \mathcal{X}^*_{i_1,i_2,\ldots,i_W}) - \Phi(\omega^*_{l-1} - \mathcal{X}^*_{i_1,i_2,\ldots,i_W}),\end{aligned} \tag{4}$$

where $\sum_{l=1}^{K} f_l(\mathcal{X}^*_{i_1,i_2,\ldots,i_W}) = \Phi(\infty - \mathcal{X}^*_{i_1,i_2,\ldots,i_W}) - \Phi(-\infty - \mathcal{X}^*_{i_1,i_2,\ldots,i_W}) = 1$. We remark that the Probit model and Logistic model are two common choices of Φ [19].

In the rest of this paper, we will address the following questions. (P1): How can we estimate the ground truth $\mathcal{X}^*$ given $\mathcal{Q}_\Omega$ and Φ? (P2): Is it possible to estimate $\mathcal{X}^*$ when the quantization rule is unknown? The analysis will be applied on general tensors and SVD-tensors, respectively.

III. Theoretical Analysis

Note that $f_l(\mathcal{X}^*_{i_1,i_2,\ldots,i_W})$ is the probability of $\mathcal{Q}_{i_1,i_2,\ldots,i_W} = l$ given $\mathcal{X}^*_{i_1,i_2,\ldots,i_W}$. Therefore, we can estimate $\mathcal{X}^*$ by mini-

mizing a negative log-likelihood function in a feasible set. The negative log-likelihood function is given by

$$F_\Omega(\mathcal{X}) = -\frac{n_1 n_2 \cdots n_W}{|\Omega|} \sum_{(i_1,\cdots,i_W)\in\Omega} \sum_{l=1}^{K} \mathbf{1}_{[\mathcal{Q}_{i_1,i_2,\cdots,i_W}=l]} \log(f_l(\mathcal{X}_{i_1,i_2,\ldots,i_W})), \tag{5}$$

One can check that f_l is strictly log-concave if Φ is log-concave, which is true when noises following Gaussian and Logistic distributions. Therefore $F_\Omega(\mathcal{X})$ is convex. The optimization problem to estimate $\mathcal{X}^*$ is shown as follows.

$$\hat{\mathcal{X}} = \arg\min_{\mathcal{X}} F_\Omega(\mathcal{X}) \quad \text{s.t. } \mathcal{X} \in \mathcal{S}_f, \tag{6}$$

and the feasible set $\mathcal{S}_f$ contains the rank constraint and the infinity norm constraint.

$$\mathcal{S}_f := \{\mathcal{X} : \|\mathcal{X}\|_\infty \le \alpha, \text{rank}(\mathcal{X}) \le r\}. \tag{7}$$

Note that the feasible set is a nonconvex set, which leads to the nonconvexity of (6). We name the optimization (6) as (OPT1). If the tensor is the SVD-tensor, i.e. $\mathcal{X}^* \in \mathcal{S}_{\text{tsvd}}$, then we replace $\mathcal{S}_f$ with $\mathcal{S}_{fsvd} : \mathcal{X} \in \mathcal{S}_{\text{tsvd}} \cap \mathcal{S}_f$, and we name the new optimization problem as (OPT2).

We make the following assumptions before establishing our main results.

$$\gamma_\alpha = \min_{l\in[K]} \inf_{|x|\le 2\alpha} \{\frac{\dot{f}_l^2(x)}{f_l^2(x)} - \frac{\ddot{f}_l(x)}{f_l(x)}\},$$
$$L_\alpha = \max_{l\in[K]} \sup_{|x|\le 2\alpha} \{\frac{|\dot{f}_l(x)|}{f_l(x)}\}, \tag{8}$$

where $\dot{f}_l$ and $\ddot{f}_l$ denote the first and second order derivatives of f_l, respectively. L_α and γ_α are bounded by some fixed constants, and $\gamma_\alpha > 0$ [14].

A. Main Theoretical Results

a) General Tensor: We now introducing the recovery error upper bound for general tensors by solving (OPT1).

Theorem 1: Suppose $\mathcal{X}^* \in \mathcal{S}_f$ is fully observed, and $f_l(x)$ is strictly log-concave in x, $\forall l \in [K]$. Then with probability at least $1-\delta$, $\delta \in [0,1]$, any global minimizer $\hat{\mathcal{X}}$ of (OPT1) satisfies

$$\|\hat{\mathcal{X}} - \mathcal{X}^*\|_F / \sqrt{n_1 n_2 \ldots n_W} \le \min(2\alpha, U_\alpha) \tag{9}$$

where

$$U_\alpha = \sqrt{\frac{64 r^{W-1} L_\alpha^2 ((\sum_{w=1}^W n_w)\log(4W/3) + \log(2/\delta))}{n_1 n_2 \ldots n_W \cdot \gamma_\alpha^2}}, \tag{10}$$

Specifically, when $n_1, n_2, \ldots, n_W$ are all in the order of n, the recovery error of (9) can be represented as

$$\|\hat{\mathcal{X}} - \mathcal{X}^*\|_F / \sqrt{n_1 n_2 \ldots n_W} \le O(\sqrt{\frac{r^{W-1} W \log W}{n^{W-1}}}), \tag{11}$$

b) SVD-Tensor: When the tensor is SVD-tensor, then the recovery error of (OPT2) can be characterized by the following theorem.

Theorem 2: Under the same assumptions as Theorem 1, for $\mathcal{X}^* \in \mathcal{S}_{fsvd}$, any global minimizer $\hat{\mathcal{X}}$ of (OPT2) satisfies

$$\|\hat{\mathcal{X}} - \mathcal{X}^*\|_F / \sqrt{n_1 n_2 \ldots n_W} \le \min(2\alpha, U'_\alpha) \tag{12}$$

with probability at least $1-\delta$, where

$$U'_\alpha = \sqrt{\frac{64 r L_\alpha^2 ((\sum_{w=1}^W n_w)\log(4W/3) + \log(2/\delta))}{n_1 n_2 \ldots n_W \cdot \gamma_\alpha^2}}, \tag{13}$$

When $n_1, n_2, \ldots, n_W$ are all in the order of n, the recovery error of (12) can be represented as

$$\|\hat{\mathcal{X}} - \mathcal{X}^*\|_F / \sqrt{n_1 n_2 \ldots n_W} \le O(\sqrt{\frac{rW \log W}{n^{W-1}}}). \tag{14}$$

c) Analysis: Here we consider the low-rank case. When n increases to infinity, it is clear that the right-hand-side of (11) and (14) all diminish to zero, and the recovery error in the SVD-tensor scenario has a faster decreasing rate. Compared with the existing error bound $O((\frac{r^{3W-3}W}{n^{W-1}})^{1/4})$ on 1-bit tensor recovery [16], the recovery error bounds established by Theorems 1 and 2 have smaller order for any $W \ge 2$. As an example, the recovery error of 1-bit tensor recovery is $O((\frac{r^{3/2}}{n^{1/2}}))$ when $W = 3$. The recovery error bounds under the exact low-rank constraint are $O(\frac{r}{n})$ and $O(\frac{\sqrt{r}}{n})$ for general tensors and SVD-tensors. When $W = 2$, both (11) and (14) obtain $O(\sqrt{\frac{r}{n}})$, which is in the same order to the quantized matrix recovery.

One can also obtain the mode-w matricization $\mathbf{X}_{(w)}$ and then apply quantized matrix recovery methods. $\mathbf{X}_{(w)}$ is a $\Theta(n) \times \Theta(n^{W-1})$ matrix when the size of each dimension is $\Theta(n)$. To our best knowledge, the best theoretical recovery error bound obtained by matrix recovery is $O(\sqrt{\frac{\bar{r}}{n}})$ [14], where $\bar{r}$ is the rank of the matrix. Similar to the 1-bit tensor recovery [16], we assume r is a constant in this comparison. our methods yield lower recovery errors when $W \ge 3$.

d) Fundamental limitation of the recovery: Here we show the recovery error from unquantized measurements by any algorithm is at least $\Theta(\sqrt{\frac{r}{n^{W-1}}})$. We assume that the considered low-rank (CP rank) tensors are noisy, and the noise follows zero mean Gaussian distribution with variance σ^2. Let $n_{\max} = \max(n_1, n_2, \cdots, n_W)$, and we further assume $r n_{\max} > 64$.

Theorem 3: For any $\mathcal{X} \in \mathcal{S}_f$, consider any algorithm that takes $\mathcal{Q} = \mathcal{X} + \mathcal{N}$ as the input and returns an estimation $\hat{\mathcal{X}}$. Then there always exists $\mathcal{X} \in \mathcal{S}_f$ such that with probability at least $\frac{3}{4}$,

$$\frac{\|\hat{\mathcal{X}} - \mathcal{X}\|_F}{\sqrt{n_1 n_2 \cdots n_W}} \ge \min(\frac{\alpha}{4}, C_1 \sigma \sqrt{\frac{r n_{\max} - 64}{n_1 n_2 \cdots n_W}}) \tag{15}$$

holds for a fixed constant $C_1 < \sqrt{\frac{1}{512}}$.
The recovery error of (15) can be represented as

$$\|\hat{\mathcal{X}} - \mathcal{X}^*\|_F / \sqrt{n_1 n_2 \ldots n_W} \ge \Theta(\sqrt{\frac{r}{n^{W-1}}}) \tag{16}$$

when $n_1, n_2, \ldots, n_W$ are all in the order of n.

B. Recovery with unknown quantization rules

Here we consider the case that $\omega_l^*, \forall l \in [K-1]$ are unknown to the operator. We can treat the boundaries as decision variables, and estimate $\omega_l^*, \forall l \in [K-1]$ together with $\mathcal{X}^*$. The resulting optimization problem is

$$\begin{aligned}&(\hat{\mathcal{X}}, \hat{\omega}_1, \hat{\omega}_2, \cdots, \hat{\omega}_{K-1})\\ &= \arg\min_{\mathcal{X}, \omega_l, \forall l \in [K-1]} F_\Omega(\mathcal{X}, \omega_1, \omega_2, \cdots, \omega_{K-1}) \\ &\text{s.t. } \mathcal{X}, \omega_1, \omega_2, \cdots, \omega_{K-1} \in \mathcal{S}_{f\omega},\end{aligned} \tag{17}$$

where

$$\begin{aligned}\mathcal{S}_{f\omega} :=\{&\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \ldots n_W}, \omega_l, \forall l \in [K-1]: \\ &\|\mathcal{X}\|_\infty \le \alpha, \text{rank}(\mathcal{X}) \le r, \\ &\omega_0 < \omega_1 < \omega_2 < \cdots < \omega_{K-1} < \omega_K\}.\end{aligned} \tag{18}$$

The only difference between (6) and (17) is that we consider the estimation of quantization boundaries and their order relations. We name the optimization problem (17) as (OPT1$_\omega$). The revision is the same when we consider estimating boundaries in (OPT2), and we name the new problem as (OPT2$_\omega$).

IV. Algorithm

In this section, we develop fast algorithms to solve (OPT1$_\omega$) and (OPT2$_\omega$).

We first address how to handle the rank constraint $\text{rank}(\mathcal{X}) \le r$. A common practice to reformulate the low-rank property is leveraging the tensor factorization $\mathcal{X} = \mathbf{A_1} \circ \mathbf{A_2} \circ \cdots \circ \mathbf{A_W}$, where $\mathbf{A_w} \in \mathbb{R}^{n_w \times r}, \forall w \in [W]$. We then move the new term into $L(\mathcal{X}, \mathbf{A_w}) = \frac{\lambda}{2}\|\mathcal{X} - \mathbf{A_1} \circ \mathbf{A_2} \circ \cdots \circ \mathbf{A_W}\|_F^2$ in the objective, where $\lambda > 0$ is used to strike a balance between the low-rank property and the loss function reduction. The equality constraint holds when λ goes to infinity.

(OPT1$_\omega$) can be revised to the following problem

$$\begin{aligned}&(\hat{\mathcal{X}}, \hat{\omega}_1, \hat{\omega}_2, \cdots, \hat{\omega}_{K-1}) = \\ &\arg\min_{\mathcal{X}, \mathbf{A_w}, w \in [W], \omega_l, l \in [K-1]} F_\Omega(\mathcal{X}, \omega_1, \omega_2, \cdots, \omega_{K-1}) \\ &+ L(\mathcal{X}, \mathbf{A_w}) + \Psi_1(\mathcal{X}) + \sum_{l=1}^{K-1} \Psi_2(\omega_l)\end{aligned} \tag{19}$$

where

$$\begin{aligned}\Psi_1(\mathcal{X}) &= \begin{cases} \infty & \text{if } \|\mathcal{X}\|_\infty > \alpha \\ 0 & \text{otherwise} \end{cases} \\ \Psi_2(\omega_l) &= \begin{cases} \infty & \text{if } \omega_l > \min(\omega_{l+1} - \kappa_{l+1}, \alpha_{\text{upper}}) \\ & \quad \text{or } \ \omega_l < \max(\omega_{l-1} + \kappa_l, \alpha_{\text{low}}) \\ 0 & \text{otherwise} \end{cases}\end{aligned} \tag{20}$$

where $\Psi_1(\mathcal{X}), \Psi_2(\omega_l)$ are indicator forms of the constraints. $\kappa_l, \forall l \in \{2, 3, \cdots, K-1\}$ are some small positive numbers that avoid trivial solutions in practice. These hyperparmeters can be set as small positive constants. We set $\kappa_1 = \kappa_K = 0$. The lower and upper bound of the boundaries $\alpha_{\text{low}}, \alpha_{\text{upper}}$ could be chosen as $-\alpha$ and α, respectively.

Let $H = F_\Omega + \frac{\lambda}{2}\|\mathcal{X} - \mathbf{A_1} \circ \mathbf{A_2} \circ \cdots \circ \mathbf{A_W}\|_F^2$. Then we leverage the Alternating Proximal Gradient Descent to solve (19) [20]. Algorithm 1 shows our proposed algorithm - Tensor-based Alternating Proximal Gradient Descent (TAPGD). TAPGD updates $\mathcal{X}, \mathbf{A_w}, w \in [W], \omega_l, l \in [K-1]$ using gradient descent and projections. (21) - (23) provide the partial gradients of H with respect to $\mathbf{A_k}$, $\mathcal{X}$, and ω_l, where $\mathbf{B_w} = \mathbf{A_W} \odot \ldots \odot \mathbf{A_{w+1}} \odot \mathbf{A_{w-1}} \odot \ldots \odot \mathbf{A_1}$.

$$\nabla_{\mathbf{A_w}} H = (\mathbf{A_w}(\mathbf{B_w})^T - \mathbf{X}_{(w)})\mathbf{B_w}, \forall w \in [W], \tag{21}$$

$$\begin{aligned}\nabla_{\mathcal{X}} H =&\nabla_{\mathcal{X}} F_\Omega(\mathcal{X}, \omega_1, \omega_2, \cdots, \omega_{K-1}) \\ &+ \lambda(\mathcal{X} - \mathbf{A_1} \circ \mathbf{A_2} \circ \ldots \circ \mathbf{A_W}),\end{aligned} \tag{22}$$

$$\begin{aligned}\nabla_{\omega_l} H = &\Big(\sum_{(i_1, i_2, \cdots, i_W) \in \Omega} \\ &\frac{\mathbf{1}_{[\mathcal{Q}_{i_1, i_2, \ldots, i_W} = l+1]} \dot{\Phi}(\omega_l - \mathcal{X}_{i_1, i_2, \ldots, i_W})}{\Phi(\omega_{l+1} - \mathcal{X}_{i_1, i_2, \ldots, i_W}) - \Phi(\omega_l - \mathcal{X}_{i_1, i_2, \ldots, i_W})}\Big) \\ &- \Big(\sum_{(i_1, i_2, \cdots, i_W) \in \Omega} \\ &\frac{\mathbf{1}_{[\mathcal{Q}_{i_1, i_2, \ldots, i_W} = l]} \dot{\Phi}(\omega_l - \mathcal{X}_{i_1, i_2, \ldots, i_W})}{\Phi(\omega_l - \mathcal{X}_{i_1, i_2, \ldots, i_W}) - \Phi(\omega_{l-1} - \mathcal{X}_{i_1, i_2, \ldots, i_W})}\Big),\end{aligned} \tag{23}$$

where

$$\begin{aligned}&\nabla_{\mathcal{X}} F_\Omega(\mathcal{X}, \omega_1, \omega_2, \cdots, \omega_{K-1})_{i_1, i_2, \ldots, i_W} \\ &= \frac{\dot{\Phi}(\omega_l - \mathcal{X}_{i_1, i_2, \ldots, i_W}) - \dot{\Phi}(\omega_{l-1} - \mathcal{X}_{i_1, i_2, \ldots, i_W})}{\Phi(\omega_l - \mathcal{X}_{i_1, i_2, \ldots, i_W}) - \Phi(\omega_{l-1} - \mathcal{X}_{i_1, i_2, \ldots, i_W})}\end{aligned} \tag{24}$$

for any $(i_1, i_2, \cdots, i_W) \in \Omega$. Otherwise,

$$\nabla_{\mathcal{X}} F_\Omega(\mathcal{X}, \omega_1, \omega_2, \cdots, \omega_{K-1})_{i_1, i_2, \ldots, i_W} = 0. \tag{25}$$

The step sizes of the gradient descent in (26) can be calculated by Lipschitz constants of $\nabla_{\mathbf{A_w}} H$, $\nabla_{\mathcal{X}} H$, and $\nabla_{\omega_l} H$, i.e. $\|(\mathbf{B_w})^T \mathbf{B_w}\|$, $\frac{1}{\sigma\beta} + \lambda$, $\frac{\sqrt{G_l} + \sqrt{G_{l+1}}}{\sigma^2\beta^2}$. G_l, G_{l+1} are the number of entries in $\mathcal{Q}_\Omega$ that equal to l and $l+1$, and β is a small positive value.

$$\begin{aligned}&\tau_{\mathbf{A_w}} = \frac{1}{\|(\mathbf{B_w})^T \mathbf{B_w}\|}, \forall w \in [W], \quad \tau_{\mathcal{X}} = \frac{1}{\frac{1}{\sigma^2\beta^2} + \lambda}, \\ &\tau_{\omega_l} = \frac{\sigma^2\beta^2}{\sqrt{G_l} + \sqrt{G_{l+1}}}, \forall l \in [K-1],\end{aligned} \tag{26}$$

We now provide the convergence guarantee for TAPGD.

Theorem 4: Assume that the generated sequence $\{\mathbf{A_w}^t\}$ is bounded. Then TAPGD converges to a critical point of (19). The convergence rate is at least $O(t^{\frac{\theta-1}{2\theta-1}})$, $\theta \in (\frac{1}{2}, 1)$.

We further develop an algorithm named TSVD-based Alternating Projected Gradient Descent (TSVD-APGD) to solve the nonconvex problem (OPT2$_\omega$). We replace the term $L(\mathcal{X}, \mathbf{A_w})$ in (19) with $\frac{\lambda}{2}\|\mathcal{X} - \sum_{i=1}^r \zeta_i \mathbf{V_{1}}_i \circ \mathbf{V_{2}}_i \circ \cdots \circ \mathbf{V_{W}}_i\|_F^2$. The updates of $\mathcal{X}, \omega_l$ follow the same steps as TAPGD. When updating the decomposition components $\zeta_i, \mathbf{V_w}$, we use the similar QR decomposition of $\mathbf{X}_{(w)}\mathbf{B_w}$ as in Ref. [21]. We will not go into details due to the space limit.

Algorithm 1 Tensor Based Alternating Proximal Gradient Descent (TAPGD)

Input: Quantized measurements $\mathcal{Q}_\Omega \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_W}$; $\omega_l^0, l \in [K-1]$; $\mathcal{X}^0$; $\mathbf{A_w}^0, w \in [W]$; parameters $\kappa_l, l \in [K]$, α_{upper}, α_{low}, σ, β, r.

for $t = 1, 2, \ldots, T$ **do**
  for $w = 1, 2, \ldots, W$ **do**
    Calculate $\nabla_{\mathbf{A_w}} H$ by (21), and $\tau_{\mathbf{A_w}}^{t-1}$ by (26).
    $\mathbf{A_w}^t = \mathbf{A_w}^{t-1} - \tau_{\mathbf{A_w}}^{t-1} \nabla_{\mathbf{A_w}} H$.
  end for
  Calculate $\nabla_{\mathcal{X}} H$ by (22), and $\tau_{\mathcal{X}}^{t-1}$ by (26).
  $\mathcal{X}^t = \mathcal{X}^{t-1} - \tau_{\mathcal{X}}^{t-1} \nabla_{\mathcal{X}} H$.
  for $i_j = 0, 1, 2, \ldots, n_j, \forall j \in [W]$ **do**
    if $\mathcal{X}^t_{i_1,i_2,\ldots,i_W} > \alpha$, **then** set $\mathcal{X}^t_{i_1,i_2,\ldots,i_W} = \alpha$.
    else if $\mathcal{X}^t_{i_1,i_2,\ldots,i_W} < -\alpha$, **then** set $\mathcal{X}^t_{i_1,i_2,\ldots,i_W} = -\alpha$.
    end if
  end for
  for $l = 1, 2, \ldots, K-1$ **do**
    Calculate $\nabla_{\omega_l} H$, $\tau_{\omega_l}^{t-1}$ according to (23) and (26).
    $\omega_l^t = \omega_l^{t-1} - \tau_{\omega_l}^{t-1} \nabla_{\omega_l} H$.
    if $\omega_l^t > \min(\omega_{l+1}^{t-1} - \kappa_l, \alpha_{\text{upper}})$, **then** set $\omega_l^t = \min(\omega_{l+1}^{t-1} - \kappa_l, \alpha_{\text{upper}})$.
    else if $\omega_l^t < \max(\omega_{l-1}^t + \kappa_l, \alpha_{\text{low}})$, **then** set $\omega_l^t = \max(\omega_{l-1}^t + \kappa_l, \alpha_{\text{low}})$.
    end if
  end for
end for
Return: $\mathcal{X}, \omega_1, \omega_2, \cdots, \omega_{K-1}$.

V. Experimental Results

In this section, we conduct experiments on synthetic data, image data [22], and data from an in-car music recommender system [5]. We set $W = 3$, $T = 200$. The recovery error in the first two datasets is measured by $\|\mathcal{X}^* - \tilde{\mathcal{X}}\|_F^2 / \|\mathcal{X}^*\|_F^2$, where $\tilde{\mathcal{X}}$ is our recovered tensor.

A. Synthetic Data

We generate a rank-r, three-dimensional tensor by $\mathbf{A_1} \circ \mathbf{A_2} \circ \mathbf{A_3}$ where entries in $\mathbf{A_w} \in \mathbb{R}^{n_w \times r}, w = 1, 2, 3$ are uniformly sampled. A SVD-tensor is generated by (2), where $\zeta_i, i \in [r]$ are selected from right-half standard normal distribution, and $\mathbf{V_1}, \mathbf{V_2}, \mathbf{V_3}$ are obtained by transforming $\mathbf{A_1}, \mathbf{A_2}, \mathbf{A_3}$ to orthonormal matrices. We then scale all the values to $[-1, 1]$. The entries of $\mathcal{N}$ are i.i.d. generated from the zero mean Gaussian distribution with standard deviation σ of 0.25. We choose 1-bit and 2-bit, i.e. $K = 2, 4$, in our experiments. For $K = 2$, $\omega_1^* = 0$. For $K = 4$, $\omega_1^* = -0.4$, $\omega_2^* = 0$, $\omega_3^* = 0.4$.

In Figure 2, we show that TAPGD leads to substantial improvement compared with M-norm constrained 1-bit tensor recovery (MNC-1bit-TR) method [16] and the quantized matrix recovery method [14]. We vary one of the rank and dimension, and demonstrates that the relative recovery error increases when the dimension decreases or rank increases. We can draw the same conclusion of TSVD-APGD from Figure 2 when the tensor is a SVD-tensor.

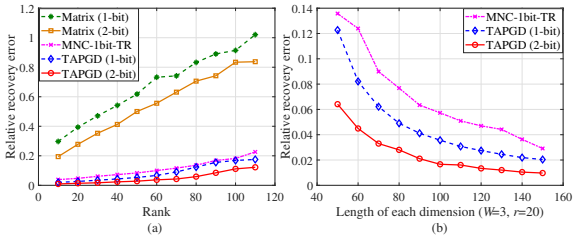


Fig. 2. (a) Relative recovery error when rank changes ($n_1 = n_2 = n_3 = 120$) (b) Relative recovery error when dimension changes.

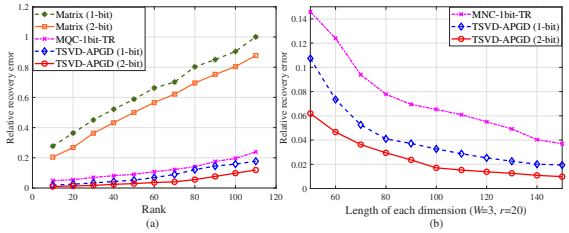


Fig. 3. (a) Relative recovery error when rank changes (SVD-tensor). (b) Relative recovery error when dimension changes (SVD-tensor).

B. Image Data

We next demonstrate the effectiveness of TAPGD on the Extend Yale Face Dataset B [22]. Each objective has 64 images, and we select two objectives to generate a three-order tensor with dimensions $192 \times 168 \times 128$. The noise tensor $\mathcal{N}$ contains values generated from Gaussian distribution with mean 0 and the standard deviation of 0.3. In the 1-bit case, $\omega_1^* = 0.4$. In the $\log_2$ 3-bit case, $\omega_1^* = 0.2$, $\omega_2^* = 0.4$. Figure 4 (a) demonstrates that TAPGD outperforms MNC-1bit-TR, the quantized matrix recovery method, and a nonconvex low-rank tensor recovery method named Nonconvex Regularized Tensor (NORT) [7]. Figure 4 (b) shows the performance of TAPGD when the quantization rule is unknown. One can see that TAPGD can reach a comparable result compared with scenario of known quantization rule.

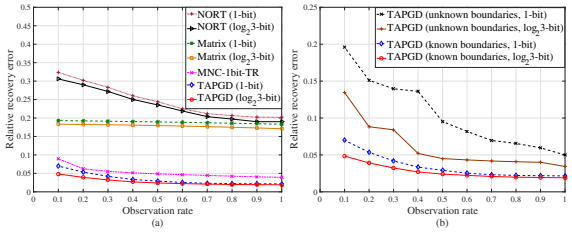


Fig. 4. (a) Relative recovery error when the observation rate changes (b) Relative recovery error of unknown boundaries.

C. Recommender Data

We apply TAPGD to an in-car music recommender dataset [5]. 42 users give ratings to 139 songs under 26 contexts

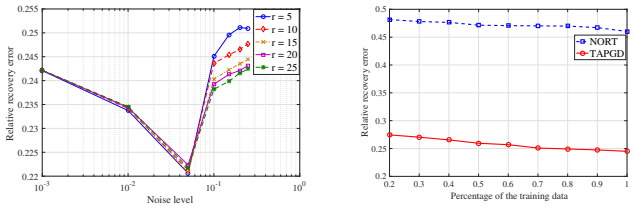
(driving, happy, sleepy, etc.). Then we can construct a tensor $\mathcal{Y}$ with dimensions $42 \times 139 \times 26$. We change the ratings from $0, 1, 2, 3, 4, 5$ to $1, 2, 3, 4, 5, 6$, and assign zeros to all missing locations. Among the total 4012 ratings, 2751 ratings have the corresponding context information, and these ratings will be used in the recovery process. The motivation here is that the actual ratings of users lie in the real-valued domain [12], [13], [16], while the measurements are quantized due to system limitation.

We separate the observed data (1.81% of the whole dataset) into two parts. The test data and training data occupy 20% and 80% of the observed data. The locations of test data are denoted by Ω_{predict}. We define the relative recovery error as

$$\frac{1}{|\Omega_{\text{predict}}|} \sum_{(i_1,i_2,i_3)\in\Omega_{\text{predict}}} \frac{|\mathcal{Y}_{i_1,i_2,i_3} - \bar{\mathcal{Y}}_{i_1,i_2,i_3}|}{5}, \qquad (27)$$

where $\bar{\mathcal{Y}}$ is the results that map the recovered values to their nearest quantized values.

Figure 5 (a) shows the recovery error when r is selected from the set $\{5, 10, 15, 20, 25\}$, and the estimated standard variation σ is selected from $\{0.001, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25\}$. When $r = 5$ and $\sigma = 0.05$, the recovery reaches its best performance. Figure 5 (b) shows the comparison between TAPGD and NORT [7]. One can see that TAPGD outperforms NORT under different percentage of the training data. The results support the hypothesis that the actual ratings are real-valued.



(a) (b)

Fig. 5. (a) Relative recovery error when the estimated noise level and rank change. (b) Comparison of TAPGD to the tensor recovery method NORT [7]: Relative recovery error when the percentage of the training data changes.

VI. Conclusion

In this paper, we propose a quantized recovery framework in the higher-order regime by exploring the low-dimensional structures. Specifically, we consider the low-CP-rank and formulate the recovery problem under the exact rank constraint. We provide the theoretical recovery errors for general tensors and SVD-tensors, as well as a fundamental limit for any recovery methods. We also build a recovery framework when the quantization boundaries are unknown and followed by two efficient algorithms. The effectiveness of our proposed methods are validated on synthetic and real datasets.

Acknowledgment

This research is supported in part by AFOSR FA9550-20-1-0122, ARO W911NF-17-1-0407, NSF 193196, and the Rensselaer-IBM AI Research Collaboration (http://airc.rpi.edu), part of the IBM AI Horizons Network (http://ibm.biz/AIHorizons).

References

[1] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[2] B. Recht, "A simpler approach to matrix completion." *Journal of Machine Learning Research*, vol. 12, no. Dec, 2011.

[3] E. J. Candes and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2342–2359, 2011.

[4] R. Kueng, H. Rauhut, and U. Terstiege, "Low rank matrix recovery from rank one measurements," *Applied and Computational Harmonic Analysis*, vol. 42, no. 1, pp. 88–116, 2017.

[5] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, A. Aydin, K.-H. Lüke, and R. Schwaiger, "Incarmusic: Context-aware music recommendations in a car," in *International Conference on Electronic Commerce and Web Technologies*. Springer, 2011, pp. 89–100.

[6] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 208–220, 2012.

[7] Q. Yao, J. T.-Y. Kwok, and B. Han, "Efficient nonconvex regularized tensor completion with structure-aware proximal iterations," in *International Conference on Machine Learning*, 2019, pp. 7035–7044.

[8] Y. Xu, R. Hao, W. Yin, and Z. Su, "Parallel matrix factorization for low-rank tensor completion," *arXiv preprint arXiv:1312.1254*, 2013.

[9] R. Wang, M. Wang, and J. Xiong, "Data recovery and subspace clustering from quantized and corrupted measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1547–1560, 2018.

[10] P. Gao, R. Wang, M. Wang, and J. H. Chow, "Low-rank matrix recovery from quantized and erroneous measurements: Accuracy-preserved data privatization in power grids," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 374–378.

[11] T. Cai and W.-X. Zhou, "A max-norm constrained minimization approach to 1-bit matrix completion," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 3619–3647, 2013.

[12] M. A. Davenport, Y. Plan, E. van den Berg, and M. Wootters, "1-bit matrix completion," *Information and Inference*, vol. 3, no. 3, pp. 189–223, 2014.

[13] S. A. Bhaskar, "Probabilistic low-rank matrix completion from quantized measurements," *The Journal of Machine Learning Research*, vol. 17, no. 60, pp. 1–34, 2016.

[14] P. Gao, R. Wang, M. Wang, and J. H. Chow, "Low-rank matrix recovery from noisy, quantized and erroneous measurements," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2918–2932, 2018.

[15] A. Aidini, G. Tsagkatakis, and P. Tsakalides, "1-bit tensor completion," *Electronic Imaging*, vol. 2018, no. 13, pp. 261–1, 2018.

[16] N. Ghadermarzy, Y. Plan, and O. Yilmaz, "Learning tensors from partial binary measurements," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 29–40, 2019.

[17] S. Friedland and L.-H. Lim, "Nuclear norm of higher-order tensors," *Mathematics of Computation*, vol. 87, no. 311, pp. 1255–1281, 2018.

[18] J. Chen and Y. Saad, "On the tensor svd and the optimal low rank orthogonal approximation of tensors," *SIAM journal on Matrix Analysis and Applications*, vol. 30, no. 4, pp. 1709–1734, 2009.

[19] R. Wang, M. Wang, and X. Jinjun, "Achieve data privacy and clustering accuracy simultaneously through quantized data recovery," *EURASIP Journal on Advances in Signal Processing*, vol. 2020, no. 1, 2020.

[20] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.

[21] J. Li and F. Huang, "Guaranteed simultaneous asymmetric tensor decomposition via orthogonalized alternating least squares," *arXiv preprint arXiv:1805.10348*, 2018.

[22] A. S. Georghiades, B. Peter N, and K. David J, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.