Inverse Problems and Imaging

doi:10.3934/ipi.2023059

ITERATIVE SINGULAR TUBE HARD THRESHOLDING ALGORITHMS FOR TENSOR RECOVERY

Rachel Grotheer[1], Shuang Li[2], Anna Ma[3]
Deanna Needell[4] and Jing Qin[*5]

[1]Department of Mathematics, Wofford College, Spartanburg, SC 29303, USA

[2]Department of Electrical and Computer Engineering,
Iowa State University, Ames, IA 50011, USA

[3]Department of Mathematics, University of California, Irvine, CA 92697, USA

[4]Department of Mathematics, University of California, Los Angeles, CA 90095, USA

[5]Department of Mathematics, University of Kentucky, Lexington, KY 40506, USA

(Communicated by Yifei Lou)

Abstract. Due to the explosive growth of large-scale data sets, tensors have been a vital tool to analyze and process high-dimensional data. Different from the matrix case, tensor decomposition has been defined in various formats, which can be further used to define the best low-rank approximation of a tensor to significantly reduce the dimensionality for signal compression and recovery. In this paper, we consider the low-rank tensor recovery problem when the tubal rank of the underlying tensor is given or estimated *a priori*. We propose a novel class of iterative singular tube hard thresholding algorithms for tensor recovery based on the low-tubal-rank tensor approximation, including basic, accelerated deterministic and stochastic versions. Convergence guarantees are provided along with the special case when the measurements are linear. Numerical experiments on tensor compressive sensing and color image inpainting are conducted to demonstrate convergence and computational efficiency in practice.

1. **Introduction.** Due to the fast development of sensing and transmission technologies, data has been growing explosively, which makes the data representation a bottleneck in signal analysis and processing. Unlike the traditional vector/matrix representation, tensor provides a versatile tool to represent large-scale multi-way data sets. It has also been playing an important role in a wide spectrum of application areas, including video processing, signal processing, machine learning and neuroscience. However, when transitioning from two-way matrices to multi-way tensors, it becomes challenging to extend certain operators and preserve properties, e.g., tensor multiplication. Moreover, it is desirable to design fast numerical algorithms since computation involving tensors is highly demanding. To address these

2020 *Mathematics Subject Classification.* Primary: 15A69; Secondary: 68W20, 15A83, 94A12, 65F55.

Key words and phrases. Tensor completion, tubal rank, hard thresholding, stochastic algorithm, image inpainting.

*Corresponding author: Jing Qin.

issues, decomposition of tensors becomes a feasible and popular approach to extract latent data structures and break the curse of dimensionality.

Many tensor decomposition methods have been developed by generalizing matrix decomposition. In particular, similar to the principal component decomposition for matrices, the CANDECOMP/PARAFAC (CP) decomposition [3, 12] decomposes a tensor as a sum of rank-one tensors. As a generalization of the matrix singular value decomposition (SVD), the Tucker form [29] decomposes a tensor as a k-mode product of a core tensor and multiple matrices. Two popular Tucker decomposition methods have been proposed, including Higher Order Singular Value Decomposition (HOSVD) [6] and Higher Order Orthogonal Iteration (HOOI) [7]. For the introduction of various tensor decomposition methods, we refer the readers to the comprehensive review paper [16] and references therein.

Finding the CP decomposition is an NP-hard problem and the low-CP-rank approximation is ill-posed [13]. Moreover, the Tucker decomposition is not unique as it depends on the representation order for each mode, which is also computationally expensive [16]. To make tensor decomposition more practically useful and possess uniqueness guarantees, a new type of tensor product, called t-product, and its corresponding related concepts including tubal rank, t-SVD, and tensor nuclear norm were developed for third-order tensors based on the fast Fourier transform [15, 14]. It has been applied to solve many high-dimensional data recovery problems, including tensor robust principal component analysis [20], tensor denoising and completion [36, 18, 25, 31, 30, 35], imaging applications [15, 14], image deblurring and video face recognition via order-p tensor decomposition [21], and hyperspectral image restoration [10]. In this paper, we focus on third-order tensors for simplicity and adopt the t-product. All our results can be extended to higher order tensors using a well-defined tensor product following the analysis in [21].

Due to the compressibility of high-dimensional data, a large amount of data recovery problems can be considered as a tubal-rank restricted minimization problem whose objective function depends on the generation of measurements. For example, in tensor compressive sensing, measurements are generated by taking a single or a sequence of tensor inner products of the sensing tensors and the tensor to be restored. Accordingly, the objective function is usually in the form of the ℓ_2-norm. The tensor restricted isometry property (RIP) and exact tensor recovery for the linear measurements is discussed in [33] as an extension of the matrix RIP. Rank-restricted RIP in the matrix case [11] can be extended to the tubal-rank-restricted RIP [34] in the tensor case, which will pave the theoretical foundation for our work. Motivated by the iterative singular tube thresholding (ISTT) algorithm [31], we propose iterative singular tube hard thresholding (ISTHT) algorithm which alternate gradient descent and singular tube hard thresholding (STHT). Note that ISTT uses the soft thresholding operator as the proximal operator of the tensor tubal nuclear norm, i.e., the solution to a convex relaxed tensor tubal nuclear norm minimization problem. Unlike ISTT, STHT serves as the proximal operator of the cardinality of the set consisting of nonzero tubes resulting from the tubal-rank constraint. In addition, STHT is based on the reduced t-SVD which provides a low tubal rank approximation of a given tensor [15]. Iterative hard thresholding algorithms for several tensor decompositions, based on the low Tucker rank approximation of a tensor, are discussed in the compressive sensing scenario with linear measurements [28]. Low-tubal-rank tensor recovery can also be solved by proximal gradient algorithm [19]. To further speed up the computation, we also develop an accelerated

version of ISTHT, which integrates the Nesterov scheme for updating the step size in an adaptive manner.

When the data size is extremely large, stochastic gradient descent can be embedded into the algorithm framework to reduce the computational cost. Recently, stochastic greedy algorithms (SGA) including stochastic iterative hard thresholding (StoIHT) and stochastic gradient matching pursuit (StoGradMP) have shown great potential in efficiently solving sparsity constrained optimization problems [24, 27] as well as in various applications [9]. By iteratively seeking the support and running stochastic gradient descent, SGA can preserve the desired sparsity of iterates while decreasing the objective function value. Based on this observation, we develop stochastic ISTHT algorithms in non-batched and batched versions. Theoretical discussions have shown the proposed stochastic algorithm achieves at a linear convergence rate. Note that we develop iterative hard thresholding algorithms based on the low tubal rank approximation in a more general setting where the objective function is separable and satisfies the tubal-rank restricted strong smoothness and convexity properties. Numerical experiments on synthetic and real third-order tensorial data sets, including synthetic linear tensor measurements and RGB color images, have demonstrated that the proposed algorithms are effective in high-dimensional data recovery. It is worth noting that proposed tensor algorithms perform excellent especially for color image inpainting in terms of computational efficiency and recovery accuracy when the underlying image has a low tubal rank structure.

The rest of the paper is organized as follows. In Section 2, fundamental concepts in tensor algebra are introduced, together with properties of tensor-variable functions, such as tensor restricted strong convexity and smoothness, and the tubal-rank restricted isometry property. In Section 3, we describe the proposed non-stochastic/stochastic ISTHT algorithms in detail. Section 4 provides convergence analysis of our proposed algorithms for tensor completion with linear measurements as a special case. Numerical experiments and corresponding performance results are reported in Section 5. Finally, we draw a conclusion and point out future works in Section 6.

2. **Preliminaries.**

In this section, we provide preliminary knowledge for the best low-tubal-rank approximation and then define new concepts based on the tubal rank, including tubal-rank restricted strong convexity and smoothness of functions on a tensor space, and the tubal-rank restricted isometry property.

2.1. **Tensor algebra.**

To comply with traditional notation, we use boldface lower case letters to denote vectors by default, e.g., a vector $\mathbf{x} \in \mathbb{R}^n$. Matrices are denoted by capital letters, e.g., $X \in \mathbb{R}^{n_1 \times n_2}$ represents a matrix with n_1 rows and n_2 columns. Calligraphy letters such as $\mathcal{X}$ are used to denote tensors or unless specified otherwise (e.g., linear map), and let $\mathbb{R}^{n_1 \times n_2 \times n_3}$ be the space consisting of all real third-order tensors of size $n_1 \times n_2 \times n_3$. The set of integers $\{1, 2, \ldots, n\}$ is denoted by $[n]$. Given a tensor, a fiber is a vector obtained by fixing two dimensions while a slice is a matrix obtained by fixing one dimension. If $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then the fiber $\mathcal{X}(i, j, :)$ along the third dimension is a n_3-dimensional vector, which is also called the (i, j)-th tube or tubal-scalar of length n_3. Likewise, $\mathcal{X}(i, :, :)$, $\mathcal{X}(:, i, :)$ and $\mathcal{X}(:, :, i)$ are used to denote the respective horizontal, lateral and frontal slices. The (i, j, k)-th component of $\mathcal{X}$ is denoted by $\mathcal{X}_{i,j,k}$ or $\mathcal{X}(i, j, k)$. A tensor $\mathcal{X}$ is called f-diagonal if all frontal slices are diagonal matrices. For further notational convenience, the frontal slice $\mathcal{X}(:, :, i)$ is denoted by $X^{(i)}$. Using frontal slices, block diagonalization and circular operators which convert an $n_1 \times n_2 \times n_3$ tensor to a

$(n_1n_3) \times (n_2n_3)$ matrix are defined as follows

$$\text{bdiag}(\mathcal{X}) = \begin{bmatrix} X^{(1)} & & & \\ & X^{(2)} & & \\ & & \ddots & \\ & & & X^{(n_3)} \end{bmatrix}, \text{bcirc}(\mathcal{X}) = \begin{bmatrix} X^{(1)} & X^{(n_3)} & \cdots & X^{(2)} \\ X^{(2)} & X^{(1)} & \cdots & X^{(3)} \\ & \cdots & \cdots & \\ X^{(n_3)} & X^{(n_3-1)} & \cdots & X^{(1)} \end{bmatrix}.$$

Without padding extra entries, another pair of operators are also defined to rewrite an $n_1 \times n_2 \times n_3$ tensor as an $(n_1n_3) \times n_2$ matrix and vice versa:

$$\text{unfold}(\mathcal{X}) = \begin{bmatrix} X^{(1)} \\ \vdots \\ X^{(n_3)} \end{bmatrix}, \quad \text{fold}(\text{unfold}(\mathcal{X})) = \mathcal{X}.$$

Definition 2.1. [14] Given two tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the tensor product (t-product) is defined as

$$\mathcal{A} * \mathcal{B} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})), \tag{1}$$

where $\cdot$ is the standard matrix multiplication. We can also rewrite (1) as

$$\text{unfold}(\mathcal{A} * \mathcal{B}) = \text{bcirc}(\mathcal{A})\text{unfold}(\mathcal{B}).$$

By letting $\mathcal{C} = \mathcal{A} * \mathcal{B}$, we get an equivalent form of the above equation [4]

$$\mathcal{C}(i, j, :) = \sum_{k=1}^{n_2} \mathcal{A}(i, k, :) \odot \mathcal{B}(k, j, :), \quad i = 1, \ldots, n_1, \; j = 1, \ldots, n_4,$$

where $\odot$ is the circular convolution of two vectors by treating $1 \times 1 \times n_3$ tensors as vectors. Using the convolution-based formulation, we can get an intuitive connection between the t-product and other imaging applications, such as image deblurring [4].

Remark. Using the definition of t-product and the block circulant operator, we can show that

$$\text{bcirc}(\mathcal{A} * \mathcal{B}) = \text{bcirc}(\mathcal{A})\, \text{bcirc}(\mathcal{B}).$$

Based on the definition of t-product, a lot of concepts in matrix algebra can be extended to the tensor case. For example, the transpose of $\mathcal{X}$ is denoted by $\mathcal{X}^T$, which is an $n_2 \times n_1 \times n_3$ tensor given by transposing each of the frontal slices and then reversing the order of transposed slices 2 through n_3, i.e., $(\mathcal{X})^T_{i,j,1} = \mathcal{X}_{j,i,1}$ and $(\mathcal{X})^T_{i,j,k} = \mathcal{X}_{j,i,n_3-k+1}$ for $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$ and $k = 2, \ldots, n_3$. The identity tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is a tensor whose first frontal slice is an $n_1 \times n_1$ identity matrix and all other frontal slices are zero matrices. A tensor $\mathcal{Q} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is called orthogonal if $\mathcal{Q} * \mathcal{Q}^T = \mathcal{Q}^T * \mathcal{Q} = \mathcal{I}$. By default, the space $\mathbb{R}^{n_1 \times n_2 \times n_3}$ is considered as a Hilbert space equipped with the inner product

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i,j,k} \mathcal{X}_{i,j,k} \mathcal{Y}_{i,j,k},$$

and the Frobenius norm given by

$$\|\mathcal{X}\| = \sqrt{\sum_{i,j,k} \mathcal{X}^2_{i,j,k}}.$$

Note that if $\boldsymbol{\theta}$ is a linear map from a tensor space to a vector space, then $\|\boldsymbol{\theta}\|$ stands for the operator norm unless otherwise stated. Further detailed definitions and examples can be found in [14]. Next, we introduce the singular value decomposition (t-SVD), based on the t-product.

Definition 2.2. [14] Given $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, there exist $\mathcal{U} \in \mathbb{R}^{n_1\times n_1\times n_3}$, $\mathcal{S} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2\times n_2\times n_3}$ such that

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T. \tag{2}$$

Here $\mathcal{S}$ is f-diagonal and the number of nonzero tubes in $\mathcal{S}$ is called the **tubal rank** of $\mathcal{X}$, denoted by $\mathrm{rank}_t(\mathcal{X})$.

Remark. Due to the relationship between tensor and matrix, the t-SVD form yields the matrix form

$$\mathrm{unfold}(\mathcal{X}) = \mathrm{bcirc}(\mathcal{U})\,\mathrm{bcirc}(\mathcal{S})\,\mathrm{unfold}(\mathcal{V}^T).$$

Note that $\mathrm{unfold}(\mathcal{V}^T) \neq (\mathrm{unfold}(\mathcal{V}))^T$.

Moreover, it can be shown that the reduced t-SVD of $\mathcal{X}$ denoted by

$$\mathcal{X}_r = \sum_{i=1}^{r} \mathcal{U}(:,i,:) * \mathcal{S}(i,i,:) * \mathcal{V}(:,i,:)^T \tag{3}$$

is a best tubal-rank-r approximation of $\mathcal{X}$ in the sense that [15, Theorem 4.3]

$$\mathcal{X}_r = \operatorname*{argmin}_{\mathcal{Z}\in\mathbb{T}_r} \|\mathcal{X} - \mathcal{Z}\| \tag{4}$$

where

$$\mathbb{T}_r = \{\mathcal{X}_1 * \mathcal{X}_2 \mid \mathcal{X}_1 \in \mathbb{R}^{n_1\times r\times n_3},\ \mathcal{X}_2 \in \mathbb{R}^{r\times n_2\times n_3}\} \tag{5}$$

is the set of all tubal-rank at most r tensors of the size $n_1 \times n_2 \times n_3$. In addition, due to the block diagonalization of block-circulant matrices with circulant blocks under the Fourier transform, t-SVD can be efficiently obtained by using the matrix SVD and the fast Fourier transform [15, 36]. More specifically, letting $F_{n_3} \in \mathbb{C}^{n_3\times n_3}$ be the unitary discrete Fourier transform matrix, we have

$$(F_{n_3} \otimes I_{n_1}) \cdot \mathrm{bcirc}(\mathcal{X}) \cdot (F_{n_3}^* \otimes I_{n_3}) = \mathrm{bdiag}(\mathcal{D})$$

where $\otimes$ is the Kronecker product of two matrices and

$$\widehat{\mathcal{D}}(:,:,k) = \widehat{\mathcal{U}}(:,:,k)\widehat{\mathcal{S}}(:,:,k)\widehat{\mathcal{V}}(:,:,k)^T.$$

where $\widehat{\mathcal{X}}$ is the Fourier transform of $\mathcal{X}$ along the third dimension, i.e., $\widehat{\mathcal{X}}(i,j,:) = \mathrm{fft}(\mathcal{X}(i,j,:))$. To make the paper self-contained, we include the t-SVD algorithm in Algorithm 1. The operators $\mathrm{fft}(\cdot,[\,],3)$ and $\mathrm{ifft}(\cdot,[\,],3)$ represent the Fourier transform along the third dimension. Note that the Fourier transform can be replaced by other unitary transforms, e.g., discrete cosine transform [32].

Algorithm 1 Tensor Singular Value Decomposition (t-SVD) [15]

Input: $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$.
Output: $\mathcal{U}, \mathcal{S}, \mathcal{V}$.
$\widehat{\mathcal{X}} = \mathrm{fft}(\mathcal{X},[\,],3)$.
for $i = 1, 2, \ldots, n_3$ **do**
  Find the SVD of $\widehat{\mathcal{X}}(:,:,i)$ such that $\widehat{\mathcal{U}}(:,:,i)\widehat{\mathcal{S}}(:,:,i)\widehat{\mathcal{V}}(:,:,i)^T = \widehat{\mathcal{X}}(:,:,i)$
end for
$\mathcal{U} = \mathrm{ifft}(\widehat{\mathcal{U}},[\,],3)$
$\mathcal{S} = \mathrm{ifft}(\widehat{\mathcal{S}},[\,],3)$
$\mathcal{V} = \mathrm{ifft}(\widehat{\mathcal{V}},[\,],3)$

Using the t-SVD, we define the following singular tube hard thresholding operator.

Definition 2.3. For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $r \leq \min\{n_1, n_2\}$, the singular tube hard thresholding (STHT) operator $\mathcal{H}_r$ is defined as

$$\mathcal{H}_r(\mathcal{X}) = \mathcal{U} * \mathcal{H}_r(\mathcal{S}) * \mathcal{V}^T, \quad \mathcal{H}_r(\mathcal{S})(i,i,:) = \mathbf{0}, \quad \text{for} \quad i = r+1, \ldots, \min\{n_1, n_2\}. \tag{6}$$

Since the zero singular tubes do not affect the t-product, we use the best tubal-rank-r approximation to express STHT as $\mathcal{H}_r(\mathcal{X}) = \mathcal{X}_r$.

2.2. **Functions on a tensor space.** In this section, we define novel concepts for a class of functions on a tensor space. By treating a tensor as a multi-dimensional array, we can define a differentiable function on a tensor space. Specifically, a function $f : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}$ is called differentiable if all partial derivatives $\frac{\partial f}{\partial \mathcal{X}_{i,j,k}}$ exist, and in addition the gradient is simply given by $\nabla f(\mathcal{X}) = (\frac{\partial f}{\partial \mathcal{X}_{i,j,k}})_{n_1 \times n_2 \times n_3}$. Alternatively, we could first convert a tensor function to a multivariate one, compute its gradient and then rewrite it in tensor form. For example, if $f(\mathcal{X}) = \langle \mathcal{A}, \mathcal{X} \rangle$ with $\mathcal{A}, \mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then $\nabla f(\mathcal{X}) = \mathcal{A}$.

Definition 2.4. The function $f : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}$ satisfies the tubal-rank restricted strong convexity (tRSC) if there exists $\rho_r^- > 0$ such that

$$f(\mathcal{X}_2) - f(\mathcal{X}_1) - \left\langle \nabla f(\mathcal{X}_1), \mathcal{X}_2 - \mathcal{X}_1 \right\rangle \geq \frac{\rho_r^-}{2} \left\| \mathcal{X}_2 - \mathcal{X}_1 \right\|^2 \tag{7}$$

for $\mathcal{X}_1, \mathcal{X}_2 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $\text{rank}_t(\mathcal{X}_1 - \mathcal{X}_2) \leq r$.

Remark. Based on the definition of tRSC, we switch the role of $\mathcal{X}_1$ and $\mathcal{X}_2$ in (7) and get

$$\langle \nabla f(\mathcal{X}_2) - \nabla f(\mathcal{X}_1), \mathcal{X}_2 - \mathcal{X}_1 \rangle \geq \rho_r^- \left\| \mathcal{X}_2 - \mathcal{X}_1 \right\|^2. \tag{8}$$

Definition 2.5. The function $f : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}$ satisfies the tubal-rank restricted strong smoothness (tRSS) if there exists $\rho_r^+ > 0$ such that

$$\left\| \nabla f(\mathcal{X}_1) - \nabla f(\mathcal{X}_2) \right\| \leq \rho_r^+ \left\| \mathcal{X}_1 - \mathcal{X}_2 \right\| \tag{9}$$

for $\mathcal{X}_1, \mathcal{X}_2 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $\text{rank}_t(\mathcal{X}_1 - \mathcal{X}_2) \leq r$. Note that ρ_r^+ becomes the Lipschitz constant of ∇f when there is no restriction on the tubal-rank of $\mathcal{X}_1$ and $\mathcal{X}_2$.

Remark. If f satisfies the tRSS property and $\Omega = \text{span}(\mathcal{X}_1, \mathcal{X}_2)$, i.e., the tensor space linearly spanned by $\mathcal{X}_1$ and $\mathcal{X}_2$, we can follow the proofs in [24, 27] to show the tubal-rank restricted co-coercivity

$$\left\| \mathcal{P}_\Omega(\nabla f(\mathcal{X}_2) - \nabla f(\mathcal{X}_1)) \right\|^2 \leq \rho_r^+ \langle \nabla f(\mathcal{X}_1) - \nabla f(\mathcal{X}_2), \mathcal{X}_1 - \mathcal{X}_2 \rangle \tag{10}$$

for $\mathcal{X}_1, \mathcal{X}_2 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with $\text{rank}_t(\mathcal{X}_1 - \mathcal{X}_2) \leq r$. Here $\mathcal{P}_\Omega(\cdot)$ projects a tensor onto the linear space Ω.

Definition 2.6. [34] Consider a linear map $\boldsymbol{\theta} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^m$. If there exists a constant $\delta_r \in (0, 1)$ such that

$$(1 - \delta_r) \left\| \mathcal{X} \right\|^2 \leq \left\| \boldsymbol{\theta}(\mathcal{X}) \right\|^2 \leq (1 + \delta_r) \left\| \mathcal{X} \right\|^2 \tag{11}$$

for any tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ whose tubal rank is at most r, then $\boldsymbol{\theta}$ is said to satisfy the tensor-tubal-rank r restricted isometry condition (tRIP) with the restricted isometry constant δ_r.

Lemma 2.7. *If* $f(\mathcal{X}) = \frac{1}{2}\|\boldsymbol{\theta}(\mathcal{X}) - \mathbf{y}\|^2$ *where* $\mathbf{y} \in \mathbb{R}^m$ *and* $\boldsymbol{\theta} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^m$ *is a linear map satisfying the tRIP with* δ_r*, then we have*

$$\rho_r^- = 1 - \delta_r, \quad \rho_r^+ = 1 + \delta_r.$$

The proof of this lemma is based on the matrix representation of the linear map $\boldsymbol{\theta}$, i.e., a matrix $A \in \mathbb{R}^{m \times (n_1 n_2 n_3)}$ exists with $A\,\mathrm{vec}(\mathcal{X}) = \mathrm{vec}(\boldsymbol{\theta}(\mathcal{X}))$ where $\mathrm{vec}(\cdot)$ is an operator that converts a tensor to a vector by column-wise stacking at each frontal slice followed by the slice stacking.

3. Proposed algorithms.

3.1. Iterative singular tube hard thresholding.

Let $f : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}$ be a differentiable function. Consider the minimization problem

$$\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} f(\mathcal{X}) \quad \text{subject to} \quad \mathrm{rank}_t(\mathcal{X}) \le r, \tag{12}$$

which can also be written as

$$\min_{\mathcal{X} \in \mathbb{T}_r} f(\mathcal{X})$$

where $\mathbb{T}_r$ is defined in (5). Following the idea of the alternating minimization algorithm, we can obtain an algorithm that alternates unconstrained minimization of f and projection to $\mathbb{T}_r$. Thus, based on the STHT operator defined in Definition (2.3), we propose the iterative singular tube hard thresholding algorithm (ISTHT) in Algorithm 2, which alternates gradient descent and singular tube hard thresholding using t-SVD. Empirically, we have observed that t-SVD returns an f-diagonal tensor whose first frontal slice has much larger diagonal entries than those in the remaining frontal slices, thus we recommend scaling the input tensor before performing STHT and then re-scaling back. Furthermore, the step size γ_t for gradient descent can be set as a constant or defined by an adaptive method during the iterations, such as line search, a trust region method or the Barzilai-Borwein method [1]. Moreover, the Nesterov acceleration technique [23] has been integrated into gradient descent, i.e., Nesterov Accelerated Gradient Descent, with further developments of popular momentum-based methods in deep learning, such as AdaGrad [8]. It has been shown that Nesterov Accelerated Gradient Descent can achieve a convergence rate with $O(1/N^2)$ while gradient descent typically converges with the rate $O(1/N)$ where N is the number of iterations [23]. More specifically, at the t-th iteration, the step size γ_t is updated by

$$\lambda_{t+1} = \frac{1 + \sqrt{1 + 4\lambda_t^2}}{2}, \quad \gamma_{t+1} = \frac{1 - \lambda_t}{\lambda_{t+1}}, \quad \text{for } t = 0, 1, \ldots, N-1,$$

with the initial $\lambda_0 = 1$. The gradient descent step thereby becomes the linear combination of the original gradient descent step and the previous step with weight specified by γ_{t+1}. The detailed algorithm is summarized in Algorithm 3. All the algorithms terminate when either the maximum number of iterations is reached or the stopping criterion $\left\|\mathcal{X}^{t+1} - \mathcal{X}^t\right\| / \left\|\mathcal{X}^t\right\| < tol$ with a preassigned tolerance tol is met. Empirical results in Section 5 will show that properly selected step size regime may significantly improve accuracy and convergence.

Algorithm 2 Iterative Singular Tube Hard Thresholding (ISTHT)

Input: tubal rank r, step size γ, tolerance tol.
Output: $\mathcal{X}^{t+1}$.
Initialize: $\mathcal{X}^0 = \mathbf{0} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.
for $t = 0, 1, \ldots, N-1$ **do**
  $\mathcal{Z}^{t+1} = \mathcal{X}^t - \gamma \nabla f(\mathcal{X}^t)$
  $\mathcal{X}^{t+1} = \mathcal{H}_r(\mathcal{Z}^{t+1})$
  Exit if the stopping criterion is met.
end for

Algorithm 3 Accelerated Iterative Singular Tube Hard Thresholding (aISTHT)

Input: tubal rank r, step size γ, tolerance tol.
Output: $\mathcal{X}^{t+1}$.
Initialize: $\mathcal{X}^0 = \widetilde{\mathcal{Z}}^0 = \mathbf{0} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\lambda_0 = 1$.
for $t = 0, 1, \ldots, N-1$ **do**
  $\lambda_{t+1} = \frac{1+\sqrt{1+4\lambda_t^2}}{2}$
  $\gamma_{t+1} = \frac{1-\lambda_t}{\lambda_{t+1}}$
  $\widetilde{\mathcal{Z}}^{t+1} = \mathcal{X}^t - \gamma \nabla f(\mathcal{X}^t)$
  $\mathcal{Z}^{t+1} = (1-\gamma_{t+1})\widetilde{\mathcal{Z}}^{t+1} + \gamma_{t+1}\widetilde{\mathcal{Z}}^t$
  $\mathcal{X}^{t+1} = \mathcal{H}_r(\mathcal{Z}^{t+1})$
  Exit if the stopping criterion is met.
end for

3.2. **Stochastic iterative singular tube hard thresholding.** Consider a collection of functions $f_j : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}$ with $j = 1, \ldots, M$ and their average

$$F(\mathcal{X}) = \frac{1}{M} \sum_{j=1}^{M} f_j(\mathcal{X}), \quad \mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}.$$

Now we consider the following tubal-rank constrained minimization problem

$$\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} F(\mathcal{X}) \quad \text{subject to} \quad \operatorname{rank}_t(\mathcal{X}) \leq r. \tag{13}$$

By combining the stochastic gradient descent and best tubal-rank approximation steps, we proposed the Stochastic Iterative Singular Tube Hard Thresholding (StoISTHT) in Algorithm 4.

Algorithm 4 Stochastic Iterative Singular Tube Hard Thresholding (StoISTHT)

Input: tubal rank r, step size γ, tolerance tol, probabilities $\{p(i)\}_{i=1}^{M}$.
Output: $\mathcal{X}^{t+1}$.
Initialize: $\mathcal{X}^0 = \mathbf{0} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.
for $t = 0, 1, \ldots, N-1$ **do**
  Randomly select an index $i_t \in \{1, 2, \ldots, M\}$ with probability $p(i_t)$
  $\mathcal{Z}^{t+1} = \mathcal{X}^t - \frac{\gamma}{Mp(i_t)} \nabla f_{i_t}(\mathcal{X}^t)$
  $\mathcal{X}^{t+1} = \mathcal{H}_r(\mathcal{Z}^{t+1})$
  Exit if the stopping criterion is met.
end for

Based on the mini-batch technique [22], we propose an accelerated version–Batched Stochastic Iterative Singular Tube Hard Thresholding (BStoISTHT), summarized in Algorithm 5.

Algorithm 5 Batched Stochastic Iterative Singular Tube Hard Thresholding (BStoISTHT)

Input: tubal rank r, step size γ, tolerance tol, batch probabilities $\{p(\tau)\}$.
Output: $\mathcal{X}^{t+1}$.
Initialize: $\mathcal{X}^0 = \mathbf{0} \in \mathbb{R}^{n_1\times n_2\times n_3}$.
for $t = 0, 1, \ldots, N-1$ **do**
Randomly select an index batch $\tau_t \subseteq \{1, 2, \ldots, d\}$ of size b with probability $p(\tau_t)$
$\mathcal{Z}^{t+1} = \mathcal{X}^t - \frac{\gamma}{Mp(\tau_t)}(\frac{1}{b}\sum_{j\in\tau_t}\nabla f_j(\mathcal{X}^t))$
$\mathcal{X}^{t+1} = \mathcal{H}_r(\mathcal{Z}^{t+1})$
Exit if the stopping criterion is met.
end for

4. **Convergence analysis.** In this section, we provide the convergence analysis for the proposed algorithms, which can are extended from the matrix case [24, 27] to the more general tensor setting [28].

Lemma 4.1. *Let $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{n_1\times n_2\times n_3}$ with tubal-rank r. Then we have*

$$\|\mathcal{H}_r(\mathcal{X}) - \mathcal{Y}\|^2 \le 2\langle \mathcal{H}_r(\mathcal{X}) - \mathcal{Y}, \mathcal{X} - \mathcal{Y}\rangle.$$

Proof. Since the operator $\mathcal{H}_r$ gives the best tubal-rank-r approximation, we have

$$\|\mathcal{H}_r(\mathcal{X}) - \mathcal{X}\| \le \|\mathcal{Y} - \mathcal{X}\|.$$

Then we get

$$\begin{aligned}\|\mathcal{H}_r(\mathcal{X}) - \mathcal{Y}\|^2 &= \|\mathcal{H}_r(\mathcal{X}) - \mathcal{X} + \mathcal{X} - \mathcal{Y}\|^2\\ &= \|\mathcal{H}_r(\mathcal{X}) - \mathcal{X}\|^2 + \|\mathcal{X} - \mathcal{Y}\|^2 + 2\langle \mathcal{H}_r(\mathcal{X}) - \mathcal{X}, \mathcal{X} - \mathcal{Y}\rangle\\ &\le 2\|\mathcal{Y} - \mathcal{X}\|^2 + 2\langle \mathcal{H}_r(\mathcal{X}) - \mathcal{X}, \mathcal{X} - \mathcal{Y}\rangle\\ &= 2\langle \mathcal{H}_r(\mathcal{X}) - \mathcal{Y}, \mathcal{X} - \mathcal{Y}\rangle.\end{aligned}$$

□

Theorem 4.2. *Let $f : \mathbb{R}^{n_1\times n_2\times n_3} \to \mathbb{R}^m$ satisfy the tRSC and tRSS, and $\mathcal{X}^*$ be a minimizer of f with the tubal-rank at most r. Then there exist $\kappa, \sigma > 0$ such that the recovery error at the t-th iteration of Algorithm 2 is bounded from above*

$$\left\|\mathcal{X}^t - \mathcal{X}^*\right\| \le \kappa^t \left\|\mathcal{X}^0 - \mathcal{X}^*\right\| + \frac{\sigma}{1-\kappa}.$$

Proof. Let $\mathcal{R}^t = \mathcal{X}^t - \mathcal{X}^*$, Ω be the linear space spanned by $\mathcal{X}^*$, $\mathcal{X}^t$ and $\mathcal{X}^{t+1}$, and $\mathcal{P}_\Omega$ be the orthogonal projection from $\mathbb{R}^{n_1\times n_2\times n_3}$ to Ω. Note that Ω and $\mathcal{P}_\Omega$ may change over the iterations. Then for any $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\mathcal{P}_\Omega(\mathcal{X})$ must have tubal rank $\le 3r$.

By Lemma 4.1, we calculate the norm square of $\mathcal{R}^{t+1}$ as follows

$$\begin{aligned}\left\|\mathcal{R}^{t+1}\right\|^2 &\leq 2\langle \mathcal{X}^{t+1}-\mathcal{X}^*, \mathcal{Z}^{t+1}-\mathcal{X}^*\rangle \\ &= 2\langle \mathcal{X}^{t+1}-\mathcal{X}^*, \mathcal{X}^t-\gamma\nabla f(\mathcal{X}^t)-\mathcal{X}^*\rangle \\ &= 2\langle \mathcal{X}^{t+1}-\mathcal{X}^*, \mathcal{X}^t-\mathcal{X}^*-\gamma(\nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*))\rangle \\ &\quad -2\langle \mathcal{X}^{t+1}-\mathcal{X}^*, \gamma\nabla f(\mathcal{X}^*)\rangle\end{aligned}$$

The definition of the orthogonal projection $\mathcal{P}_\Omega$ yields that

$$\langle \mathcal{R}^{t+1}, \mathcal{P}_{\Omega^c}(\nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*))\rangle = 0, \quad \langle \mathcal{R}^{t+1}, \mathcal{P}_{\Omega^c}\nabla f(\mathcal{X}^*)\rangle = 0.$$

Thus we have

$$\begin{aligned}\left\|\mathcal{R}^{t+1}\right\|^2 &\leq 2\langle \mathcal{R}^{t+1}, \mathcal{R}^t-\gamma\mathcal{P}_\Omega\big(\nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\big)\rangle - 2\langle \mathcal{R}^{t+1}, \gamma\mathcal{P}_\Omega\nabla f(\mathcal{X}^*)\rangle \\ &\leq 2\left\|\mathcal{R}^{t+1}\right\|\left(\left\|\mathcal{R}^t-\gamma\mathcal{P}_\Omega\big(\nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\big)\right\| + \|\gamma\mathcal{P}_\Omega\nabla f(\mathcal{X}^*)\|\right)\end{aligned}$$

which implies that

$$\left\|\mathcal{R}^{t+1}\right\| \leq 2\left\|\mathcal{R}^t-\gamma\mathcal{P}_\Omega\big(\nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\big)\right\| + 2\gamma\left\|\mathcal{P}_\Omega\nabla f(\mathcal{X}^*)\right\|.$$

By using the co-coercivity (10) and the reformulation (8) of tRSC, we have

$$\begin{aligned}&\left\|\mathcal{R}^t-\gamma\mathcal{P}_\Omega\big(\nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\big)\right\|^2 \\ &= \left\|\mathcal{R}^t\right\|^2 + \gamma^2\left\|\mathcal{P}_\Omega\big(\nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\big)\right\|^2 - 2\gamma\langle \mathcal{R}^t, \mathcal{P}_\Omega\big(\nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\big)\rangle \\ &\leq \left\|\mathcal{R}^t\right\|^2 + \gamma^2\rho_s^+\langle \mathcal{R}^t, \nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\rangle - 2\gamma\langle \mathcal{R}^t, \nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\rangle \\ &\leq \left\|\mathcal{R}^t\right\|^2 - (2\gamma-\gamma^2\rho_{3r}^+)\langle \mathcal{R}^t, \nabla f(\mathcal{X}^t)-\nabla f(\mathcal{X}^*)\rangle \\ &\leq \left(1-(2-\gamma\rho_{3r}^+)\gamma\rho_{3r}^-\right)\left\|\mathcal{R}^t\right\|^2.\end{aligned}$$

Therefore, we get

$$\left\|\mathcal{R}^{t+1}\right\| \leq 2\sqrt{1-(2-\gamma\rho_{3r}^+)\gamma\rho_{3r}^-}\left\|\mathcal{R}^t\right\| + 2\gamma\left\|\mathcal{P}_\Omega\nabla f(\mathcal{X}^*)\right\| := \kappa\left\|\mathcal{R}^t\right\| + \sigma.$$

By recursively applying the above inequality, we get

$$\left\|\mathcal{R}^t\right\| \leq \kappa^t\left\|\mathcal{R}^0\right\| + \sigma\sum_{i=0}^{t-2}\kappa^i \leq \kappa^t\left\|\mathcal{R}^0\right\| + \frac{\sigma}{1-\kappa}$$

which completes the proof. □

Theorem 4.3. *Let $f(\mathcal{X}) = \frac{1}{2}\|\mathcal{A}(\mathcal{X})-\mathbf{y}\|^2$ with $\mathbf{y}\in\mathbb{R}^m$. If the linear map $\mathcal{A}: \mathbb{R}^{n_1\times n_2\times n_3}\to\mathbb{R}^m$ satisfies the tRIP with the restricted isometry constant δ_s and the measurements $\mathbf{y} = \mathcal{A}(\mathcal{X}^*)+\mathbf{e}$ with $\mathrm{rank}_t(\mathcal{X}^*)\leq r$ and $\|\mathbf{e}\|\leq\varepsilon$, then the t-th iteration of Algorithm 2 has a bounded recovery error*

$$\left\|\mathcal{X}^t-\mathcal{X}^*\right\| \leq \kappa^t\left\|\mathcal{X}^0-\mathcal{X}^*\right\| + \frac{\sigma}{1-\kappa},$$

where

$$\kappa = 2(\mid 1-\gamma\mid +\gamma\delta_{3r}), \quad \sigma = 2\gamma\varepsilon\sqrt{1+\delta_{2r}}. \tag{14}$$

Proof. First the gradient of f is given by

$$\nabla f(\mathcal{X}^t) = \mathcal{A}^*(\mathcal{A}(\mathcal{X}^t)-\mathbf{y})$$

where the adjoint operator $\mathcal{A}^*: \mathbb{R}^m\to\mathbb{R}^{n_1\times n_2\times n_3}$ is defined by $\langle\mathcal{A}(\mathcal{X}),\mathbf{z}\rangle = \langle\mathcal{X},\mathcal{A}^*(\mathbf{z})\rangle$ for any $\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}$ and $\mathbf{z}\in\mathbb{R}^m$.

Let $\mathcal{R}^t = \mathcal{X}^t - \mathcal{X}^*$, $\Omega = \text{span}(\mathcal{X}^*, \mathcal{X}^t, \mathcal{X}^{t+1})$, and $\mathcal{P}_\Omega$ be the orthogonal projection from $\mathbb{R}^{n_1\times n_2\times n_3}$ onto Ω. To simplify the notation, we denote

$$\mathcal{A}_\Omega(\mathcal{X}) = \mathcal{A}(\mathcal{P}_\Omega(\mathcal{X})) = \mathcal{A}\circ\mathcal{P}_\Omega(\mathcal{X}), \quad \mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}.$$

Then we have $\mathcal{A}_\Omega(\mathcal{R}^{t+1}) = \mathcal{A}(\mathcal{R}^{t+1})$ and $\mathcal{A}_\Omega(\mathcal{R}^t) = \mathcal{A}(\mathcal{R}^t)$. Note that the operator $\mathcal{A}_\Omega$ is different from the composite operator $\mathcal{P}_\Omega\circ\nabla f$ even when ∇f is a linear map. By Lemma 4.1, we estimate $\left\|\mathcal{R}^{t+1}\right\|$ as follows

$$\begin{aligned}
\left\|\mathcal{R}^{t+1}\right\|^2 &\le 2\langle\mathcal{R}^{t+1}, \mathcal{Z}^{t+1} - \mathcal{X}^*\rangle \\
&= 2\langle\mathcal{R}^{t+1}, \mathcal{X}^t - \gamma\mathcal{A}^*(\mathcal{A}(\mathcal{X}^t) - \mathbf{y}) - \mathcal{X}^*\rangle \\
&= 2\langle\mathcal{R}^{t+1}, \mathcal{X}^t - \mathcal{X}^* - \gamma\mathcal{A}^*(\mathcal{A}(\mathcal{X}^t) - \mathcal{A}(\mathcal{X}^*))\rangle + 2\langle\mathcal{R}^{t+1}, \gamma\mathcal{A}^*(\mathbf{e})\rangle \\
&= 2\langle\mathcal{R}^{t+1}, (\mathcal{I} - \gamma\mathcal{A}^*\mathcal{A})(\mathcal{R}^t)\rangle + 2\gamma\langle\mathcal{A}(\mathcal{R}^{t+1}), \mathbf{e}\rangle \\
&= 2\langle\mathcal{R}^{t+1}, (\mathcal{I} - \gamma\mathcal{A}_\Omega^*\mathcal{A}_\Omega)(\mathcal{R}^t)\rangle + 2\gamma\langle\mathcal{A}(\mathcal{R}^{t+1}), \mathbf{e}\rangle \\
&\le 2\left\|\mathcal{R}^{t+1}\right\|\left(\left\|(\mathcal{I} - \gamma\mathcal{A}_\Omega^*\mathcal{A}_\Omega)(\mathcal{R}^t)\right\| + \gamma\varepsilon\sqrt{1+\delta_{2r}}\right),
\end{aligned}$$

where $\mathcal{I}:\mathbb{R}^{n_1\times n_2\times n_3}\to\mathbb{R}^{n_1\times n_2\times n_3}$ is the identity map. The last inequality follows from applying the Cauchy-Schwarz inequality to each term in the summation, the tRIP assumption on $\mathcal{A}$ for the second term (note that $\mathcal{R}^{t+1}$ has tubal-rank at most $2r$), and lastly, the assumption that $\|\mathbf{e}\|\le\epsilon$.

Therefore we have

$$\begin{aligned}
\left\|\mathcal{R}^{t+1}\right\| &\le 2\left\|(\mathcal{I} - \gamma\mathcal{A}_\Omega^*\mathcal{A}_\Omega)(\mathcal{R}^t)\right\| + 2\gamma\varepsilon\sqrt{1+\delta_{2r}} \\
&\le 2\left\|(1-\gamma)\mathcal{R}^t\right\| + 2\gamma\left\|(\mathcal{I} - \mathcal{A}_\Omega^*\mathcal{A}_\Omega)(\mathcal{R}^t)\right\| + 2\gamma\varepsilon\sqrt{1+\delta_{2r}} \\
&\le 2\left(\mid 1-\gamma \mid +\gamma\delta_{3r}\right)\left\|\mathcal{R}^t\right\| + 2\gamma\varepsilon\sqrt{1+\delta_{2r}}.
\end{aligned}$$

The last inequality holds due to the fact that the operator norm of the map $\mathcal{I} - \mathcal{A}_\Omega^*\mathcal{A}_\Omega$ can be bounded from above by

$$\|\mathcal{I} - \mathcal{A}_\Omega^*\mathcal{A}_\Omega\| \le \sup_{\substack{\mathcal{X}:\text{rank}_t(\mathcal{X})\le 3r \\ \|\mathcal{X}\|=1}} \mid \|\mathcal{X}\|^2 - \|\mathcal{A}(\mathcal{X})\|^2 \mid = \delta_{3r}.$$

Thus we get

$$\left\|\mathcal{R}^{t+1}\right\| \le \kappa\left\|\mathcal{R}^t\right\| + \sigma,$$

where $\kappa = 2(\mid 1-\gamma \mid +\gamma\delta_{3r})$ and $\sigma = 2\gamma\varepsilon\sqrt{1+\delta_{2r}}$. By the recursive relationship, we get

$$\left\|\mathcal{R}^t\right\| \le \kappa^t\left\|\mathcal{R}^0\right\| + \frac{\sigma}{1-\kappa},$$

provided that $\kappa < 1$, which completes the proof. □

Based on the tRIP of sub-Gaussian ensembles in [33] and Theorem 4.3, we get the following corollary about the convergence of Algorithm 2 for the sub-Gaussian measurements.

Corollary 4.4. *Let $f(\mathcal{X}) = \mathcal{A}(\mathcal{X})$ where $\mathcal{A}:\mathbb{R}^{n_1\times n_2\times n_3}\to\mathbb{R}^m$ is a linear sub-Gaussian measurement ensemble, and the solution $\mathcal{X}^*$ with tubal-rank r satisfy $\mathbf{y} = \mathcal{A}(\mathcal{X}^*) + \mathbf{e}$. If $\gamma = 1$ and the number of measurements satisfies*

$$m \ge C\max\{r(n_1+n_2+1)n_3, \log(\epsilon^{-1})\},$$

with $C > 0$ which depends on the sub-Gaussian parameter, then Algorithm 2 converges to the solution $\mathcal{X}^*$ with probability at least $1-\epsilon$.

Theorem 4.5. *Let $\mathcal{X}^*$ be a feasible solution of (13) and $\mathcal{X}^0$ be the initial guess. Assume that F satisfies tRSC with the constant ρ_r^- and each f_i satisfies tRSS with the constant $\rho_r^+(i)$. Then there exist a contraction coefficient $\kappa > 0$ and a tolerance coefficient $\sigma > 0$ such that the expectation of the recovery error at the t-th iteration of Algorithm 4 is bounded from above via*

$$E\left\|\mathcal{X}^t - \mathcal{X}^*\right\| \leq \kappa^t \left\|\mathcal{X}^0 - \mathcal{X}^*\right\| + \sigma. \tag{15}$$

If $\kappa < 1$, then Algorithm 4 generates a sequence $\{\mathcal{X}^t\}$ which converges to the desired solution $\mathcal{X}^*$.

Proof. Let $\mathcal{R}^t = \mathcal{X}^t - \mathcal{X}^*$, Ω be the linear space spanned by $\mathcal{X}^*$, $\mathcal{X}^t$ and $\mathcal{X}^{t+1}$, and $\mathcal{P}_\Omega$ be the orthogonal projection from $\mathbb{R}^{n_1\times n_2\times n_3}$ to Ω. Then for any $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\mathcal{P}_\Omega(\mathcal{X})$ must have tubal rank $\leq 3r$.

By Lemma 4.1, we calculate the norm square of R^{t+1} as follows

$$\begin{aligned}
\left\|\mathcal{R}^{t+1}\right\|^2 &\leq 2\langle \mathcal{X}^{t+1} - \mathcal{X}^*, \mathcal{Z}^{t+1} - \mathcal{X}^*\rangle \\
&= 2\langle \mathcal{X}^{t+1} - \mathcal{X}^*, \mathcal{X}^t - \frac{\gamma}{Mp(i_t)}\nabla f_{i_t}(\mathcal{X}^t) - \mathcal{X}^*\rangle \\
&= 2\langle \mathcal{X}^{t+1} - \mathcal{X}^*, \mathcal{X}^t - \mathcal{X}^* - \frac{\gamma}{Mp(i_t)}(\nabla f_{i_t}(\mathcal{X}^t) - \nabla f_{i_t}(\mathcal{X}^*))\rangle \\
&\quad - 2\langle \mathcal{X}^{t+1} - \mathcal{X}^*, \frac{\gamma}{Mp(i_t)}\nabla f_{i_t}(\mathcal{X}^*)\rangle
\end{aligned}$$

The definition of the orthogonal projection $\mathcal{P}_\Omega$ yields that

$$\langle \mathcal{R}^{t+1}, \mathcal{P}_{\Omega^c}(\nabla f_{i_t}(\mathcal{X}^t) - \nabla f_{i_t}(\mathcal{X}^*))\rangle = 0, \quad \langle \mathcal{R}^{t+1}, \mathcal{P}_{\Omega^c}\left(\nabla f_{i_t}(\mathcal{X}^*)\right)\rangle = 0.$$

Thus we have

$$\begin{aligned}
&\left\|\mathcal{R}^{t+1}\right\|^2 \\
&\leq 2\langle \mathcal{R}^{t+1}, \mathcal{R}^t - \frac{\gamma}{Mp(i_t)}\mathcal{P}_\Omega\big(\nabla f_{i_t}(\mathcal{X}^t) - \nabla f_{i_t}(\mathcal{X}^*)\big)\rangle - 2\langle \mathcal{R}^{t+1}, \frac{\gamma}{Mp(i_t)}\mathcal{P}_\Omega\nabla f_{i_t}(\mathcal{X}^*)\rangle \\
&\leq 2\left\|\mathcal{R}^{t+1}\right\|\left(\left\|\mathcal{R}^t - \frac{\gamma}{Mp(i_t)}\mathcal{P}_\Omega\big(\nabla f_{i_t}(\mathcal{X}^t) - \nabla f_{i_t}(\mathcal{X}^*)\big)\right\| + \left\|\frac{\gamma}{Mp(i_t)}\mathcal{P}_\Omega\nabla f_{i_t}(\mathcal{X}^*)\right\|\right).
\end{aligned}$$

Let $I_t = \{i_1, \ldots, i_t\}$ be the set of indices that are randomly drawn from the discrete distribution $\{p(i)\}_{i=1}^M$ after t iterations. By taking the expectation on both sides of the above inequality with respect to i_t conditioned on I_{t-1}, we get

$$\begin{aligned}
E_{i_t|I_{t-1}}\left\|\mathcal{R}^{t+1}\right\| \leq\ & 2E_{i_t|I_{t-1}}\left\|\mathcal{R}^t - \frac{\gamma}{Mp(i_t)}\mathcal{P}_\Omega\big(\nabla f_{i_t}(\mathcal{X}^t) - \nabla f_{i_t}(\mathcal{X}^*)\big)\right\| \\
&+ 2E_{i_t|I_{t-1}}\left\|\frac{\gamma}{Mp(i_t)}\mathcal{P}_\Omega\nabla f_{i_t}(\mathcal{X}^*)\right\|.
\end{aligned}$$

By adapting the proof in [24], we are able to show that

$$E_{i_t|I_{t-1}}\left\|\mathcal{R}^t - \frac{\gamma}{Mp(i_t)}\mathcal{P}_\Omega\big(\nabla f_{i_t}(\mathcal{X}^t) - \nabla f_{i_t}(\mathcal{X}^*)\big)\right\| \leq \sqrt{1 - (2 - \gamma\alpha_{3r})\gamma\rho_{3r}^-}\left\|\mathcal{R}^t\right\|,$$

where $\alpha_{3r} = \max_i \frac{\rho_{3r}^+(i)}{Mp(i)}$. In addition, we have

$$E_{i_t|I_{t-1}}\left\|\frac{\gamma}{Mp(i_t)}\mathcal{P}_\Omega\nabla f_{i_t}(\mathcal{X}^*)\right\| \leq \frac{\gamma}{M\max_i p(i)}E_{i_t}\left\|\mathcal{P}_\Omega f_{i_t}(\mathcal{X}^*)\right\|.$$

Then we get

$$E_{i_t|I_{t-1}}\left\|\mathcal{R}^{t+1}\right\| \leq \kappa\left\|\mathcal{R}^t\right\| + \sigma,$$

where

$$\kappa = 2\sqrt{1-(2-\gamma\alpha_{3r})\gamma\rho_{3r}^{-}}, \quad \text{and} \quad \sigma = \frac{2\gamma}{M\max_i p(i)} E_{i_t}\left\|\mathcal{P}_\Omega f_{i_t}(\mathcal{X}^*)\right\|.$$

By taking the expectation on both sides with respect to I_{t-1}, we get the desired inequality

$$E_{I_t}\left\|\mathcal{R}^{t+1}\right\| \le \kappa E_{I_{t-1}}\left\|\mathcal{R}^{t}\right\| + \sigma,$$

which further yields

$$E_{I_t}\left\|\mathcal{X}^{t}-\mathcal{X}^*\right\| \le \kappa^t\left\|\mathcal{X}^{0}-\mathcal{X}^*\right\| + \frac{\sigma}{1-\kappa}.$$

Applying this result recursively completes the proof. □

5. **Numerical experiments.** In this section, we show how to apply the proposed algorithms to solve multiple application problems, including tensor compressive sensing recovery and low-tubal-rank tensor completion, and conduct a variety of numerical experiments to show the performance of the proposed algorithms. For the performance comparison metric, we use the recovery error (RE) for tensor recovery defined by

$$RE = \|\mathcal{X}-\widetilde{\mathcal{X}}\|/\|\mathcal{X}\|$$

where $\widetilde{\mathcal{X}}$ is an estimation of the ground truth $\mathcal{X}$. All numerical experiments were run in MATLAB R2021b on a desktop computer with 64GB RAM and a 3.10GHz Intel Core i9-9960X CPU.

5.1. **Tensor compressive sensing.** By extending the matrix case, we consider the linear tensor compressive sensing problem where each measurement is generated by the inner product of two tensors, i.e., a sensing tensor and a low-tubal-rank tensor to be reconstructed. Let $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ be the tensor with low tubal rank. Given a collection of sensing tensors $\{\mathcal{A}_j\}_{j=1}^M$, the obtained data vector $\mathbf{y} \in \mathbb{R}^M$ is generated by a linear map $\boldsymbol{\theta}$ from $\mathbb{R}^{n_1\times n_2\times n_3}$ to $\mathbb{R}^M$ given by

$$\boldsymbol{\theta}(\mathcal{X}) = \begin{bmatrix} \langle \mathcal{A}_1, \mathcal{X}\rangle \\ \vdots \\ \langle \mathcal{A}_M, \mathcal{X}\rangle \end{bmatrix}, \quad \mathcal{A}_j \in \mathbb{R}^{n_1\times n_2\times n_3}, \quad j=1,\dots,M. \tag{16}$$

Let $\mathrm{vec}(\cdot)$ be the operator that converts a tensor into a vector by columnwise stacking of each frontal slice followed by slicewise stacking. The linear map $\boldsymbol{\theta}$ can be therefore represented as

$$\boldsymbol{\theta}(\mathcal{X}) = \begin{bmatrix} \mathrm{vec}(\mathcal{A}_1)^T \\ \vdots \\ \mathrm{vec}(\mathcal{A}_M)^T \end{bmatrix} \mathrm{vec}(\mathcal{X}) := A\,\mathrm{vec}(\mathcal{X})$$

where $A \in \mathbb{R}^{M\times(n_1n_2n_3)}$ and $\mathrm{vec}(\mathcal{X}) \in \mathbb{R}^{n_1n_2n_3}$. Then the compressive sensing low-tubal-rank tensor recovery problem boils down to solving (13) with the following objective function

$$F(\mathcal{X}) = \frac{1}{2M}\|\boldsymbol{\theta}(\mathcal{X})-\mathbf{y}\|^2 = \frac{1}{M}\sum_{\ell=1}^{M}\frac{1}{2}\Big|\sum_{i,j,k}\mathcal{A}_{ijk}^{\ell}\mathcal{X}_{ijk} - y_\ell\Big|^2 := \frac{1}{M}\sum_{\ell=1}^{M} f_\ell(\mathcal{X}). \tag{17}$$

As shown in [34], sub-Gaussian measurements satisfy tRIP and thereby satisfy tRSC and tRSS. Therefore, our proposed algorithms can be applied with guaranteed convergence.

To justify the performance of the proposed algorithms, we test synthetic data, where the sensing matrix A and the ground truth $\mathcal{X}$ with low tubal rank consist of independently identically distributed (i.i.d.) samples from the Gaussian distribution. In our first set of experiments, we compare the Algorithms 2 and 3 with various tensor tubal ranks, measurement sampling rates, and noise ratios. Specifically, all ground truth tensors are of the size $20 \times 20 \times 10$, the maximum number of iterations is set as 500 and the stepsize $\tau = 100$. In Figure 1, we plot the reconstruction error for the two algorithms versus the iteration number with the tubal rank of $\mathcal{X}$ ranging in $\{1, 2, 3, 4\}$. By varying the sampling rate $M/N \in \{0.3, 0.4, 0.5, 0.6\}$ with $N = n_1 n_2 n_3$ and fixing the tubal rank as 2, we get the results shown in Figure 2. From these two figures, one can see that Algorithm 3 converges faster than Algorithm 2 in terms of reconstruction error. To further test the robustness to noise, we add to the measurements $\mathbf{y}$ various types of noise with standard deviation $\max_{1 \le i \le M} |y_i| \sigma$ where $\sigma \in \{0.01, 0.02, 0.03, 0.04\}$. The corresponding results are displayed in Figure 3, which shows that Algorithm 3 yields faster decay in the error than Algorithm 2 but eventually converges to the same limit. Overall, if the tensor rank is relatively low and the sampling rate is high, then both algorithms have fast error decay and Algorithm 3 with Nesterov's adaptive stepsize selection strategy effectively accelerates the convergence.

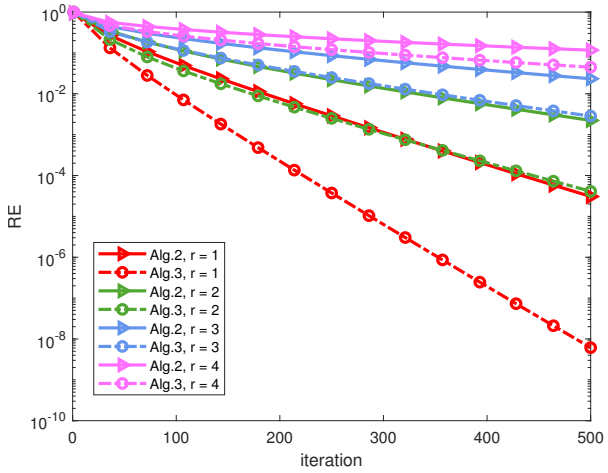


FIGURE 1. Reconstruction error comparison with various tensor tubal ranks. The ground truth is noise-free and the sampling rate is fixed as $M/N = 0.60$.

Furthermore, we test the noise-free tensor data of size $20 \times 20 \times 10$ with tubal rank one and a fixed sampling rate of 60%, and apply the Algorithm 5 with various batch sizes to recover the tensor. The results for using the batch size b ranging in $\{200, 400, 600, 800, 1000\}$, i.e., 0.05% to 0.25% of the entire tensor size, are shown in Figure 4. One can see that the stochastic version of the algorithm takes less iterations but more running time by increasing the batch size. If the batch size increases by 200, then about 4 more seconds in running time will be desired.

5.2. **Color image inpainting.** The second application for tensor recovery we show here is color image inpainting. Notice that color images naturally have three-dimensional tensor structures with the third dimension specifying the number of

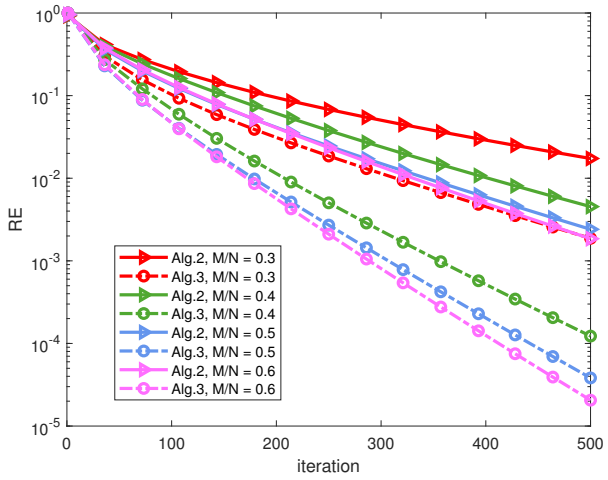


FIGURE 2. Reconstruction error comparison with various sampling rates. The ground truth is noise-free with fixed tubal rank as 2.

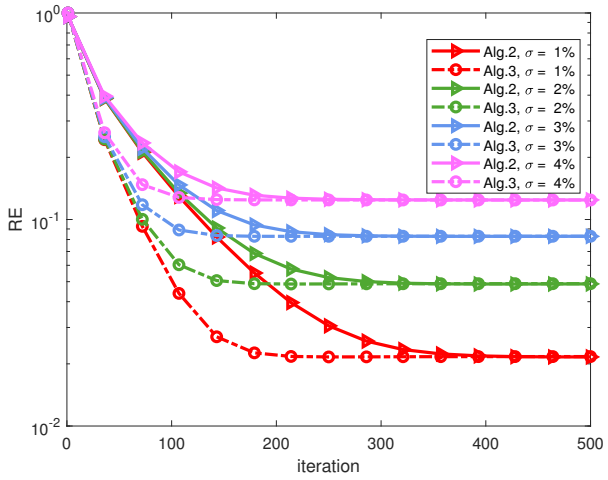


FIGURE 3. Reconstruction error comparison with various noisy measurements.

color channels. Assume that a RGB color image $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times 3}$ with $n_1 \times n_2$ pixels is contaminated by Gaussian noise with missing pixel intensities. Since the tubal rank of $\mathcal{X}$ is no greater than each dimension, we consider the following color image inpainting model

$$\min_{\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \frac{1}{2} \left\| \mathcal{P}_\Lambda(\mathcal{X} - \mathcal{Y}) \right\|^2, \quad \text{rank}_t(\mathcal{Y}) \leq k. \tag{18}$$

Although it can be reformulated as a linear map from $\mathbb{R}^{n_1 \times n_2 \times n_3}$ to $\mathbb{R}^M$, we express $\mathcal{P}$ implicitly, which is implemented by extracting entries in Λ. Here we compare the proposed Algorithm 2 with two popular color image inpainting methods: (1) Coherence Transport based Inpainting (CTI) [2] with Matlab command `inpaintCoherent`; (2) EXemplar-based inpainting in Tensor filling order (EXT) [5, 17] with the Matlab command `inpaintExemplar`.

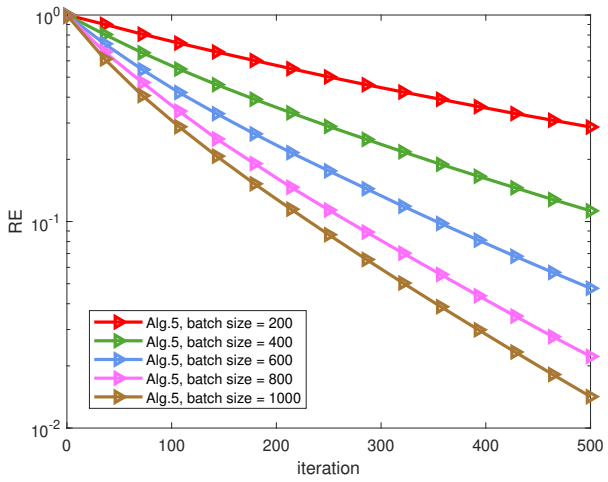
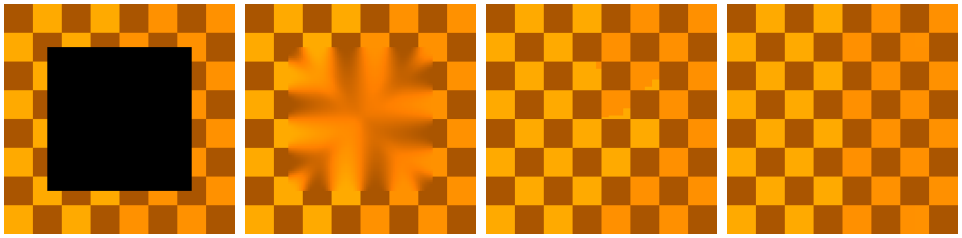


FIGURE 4. Reconstruction error comparison for Algorithm 5 with various batch sizes. The running times for the batch sizes $b = 200, 400, 600, 800, 1000$ are (in seconds): 8.38, 12.27, 16.51, 20.52, 24.66, respectively.

To start with, we create a synthetic color image, i.e., checkerboard image, with size $128 \times 128 \times 3$ and tubal rank 2. Note that the three color channels have different intensities. The observed image is then obtained by occluding all the pixels from a square of size 80×80 in the center of the image. In Figure 5, we show the observed image with occlusions and our result. Figure 7 contains the plots for the recovery error and the objective function values versus the iteration number in the Algorithm 2. One can see that our algorithm can achieve 10^{-4} for the recovery accuracy within 150 iterations while the other two comparing methods yield large errors $\geq 10^{-2}$. Notice that both CTI and EXT are based on exploiting the local image patch similarity while potentially skipping or putting less weight on preserving global repetitive patterns. They may struggle or fail when applied to fill large holes. In contrast, our algorithm prioritizes global similarity using the low tubal-rank data representation and performs better for this task. When the ground truth data has a relatively large tubal rank, Algorithm 2 will still converge to a decent result but very slowly. In the worst scenario when $r = \min\{n_1, n_2\}$, it behaves similar to gradient descent and $\mathcal{H}_r$ takes no effect.

For the natural image test, we download a facade image, "101_rectified_cropped", from https://people.ee.ethz.ch/~daid/FacadeSyn/ and then crop it to a color image of size $200 \times 200 \times 3$. Note that this image has tubal rank 3. In Figure 6, we show the observed image with an occluded box of size 80×80 in the center and our recovered result. Their recovery error and objective function value plots are shown in Figure 7, which show a swamp effect [26]. To achieve an RE of 10^{-4}, more than 300 iterations are desired. This implies that the increase of tubal rank would request more iterations to achieve the same accuracy.

6. **Conclusion.** The tensor has generalized the concept of the matrix but with more sophisticated features and computational challenges. Many tensor decompositions including CP, Tucker and t-SVD decompositions, have been developed to help analyze and manipulate large-scale data sets. Due to the computational efficiency and simple interpretation, t-SVD has recently attracted a lot of research



(a) Observed image (b) CTI RE=0.1945 (c) EXT RE=0.0385 (d) Our result RE=1.29×10^{-8}

FIGURE 5. Recovery of a checkerboard image with tubal rank 2 via Algorithm 2.

(a) Observed image (b) CTI RE= 0.1222 (c) EXT RE= 0.0658 (d) Our result RE= 3.89×10^{-12}

FIGURE 6. Recovery of a facade image with tubal rank 3 via Algorithm 2.

attention, especially in the imaging field. In this work, we develop the iterative singular tube hard thresholding algorithm, which uses the t-SVD, together with its stochastic and batched stochastic versions. Tubal rank-restricted strong convexity and strong smoothness yield the convergence of the proposed algorithms.

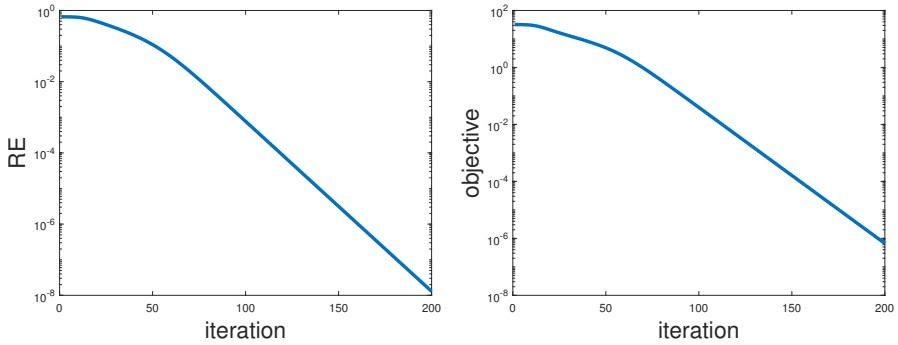
CRediT author contributions. Grotheer: validation, review & editing, funding acquisition; Li: validation, review & editing, revision, funding acquisition; Ma: investigation, formal analysis, review & editing, visualization, revision, funding acquisition; Needell: conceptualization, review & editing, revision, project administration, supervision, funding acquisition; Qin: conceptualization, methodology, software, formal analysis, data curation, visualization, investigation, original draft writing, review & editing, revision, funding acquisition.

Acknowledgments. This material is based upon work supported by the National Security Agency under Grant No. H98230-19-1-0119, The Lyda Hill Foundation, The McGovern Foundation, and Microsoft Research, while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the summer of 2019. In addition, Grotheer was supported by the Goucher College Summer Research grant, Needell was funded by NSF CAREER DMS #2011140 and NSF DMS #2108479, and Qin is supported by NSF DMS #1941197.

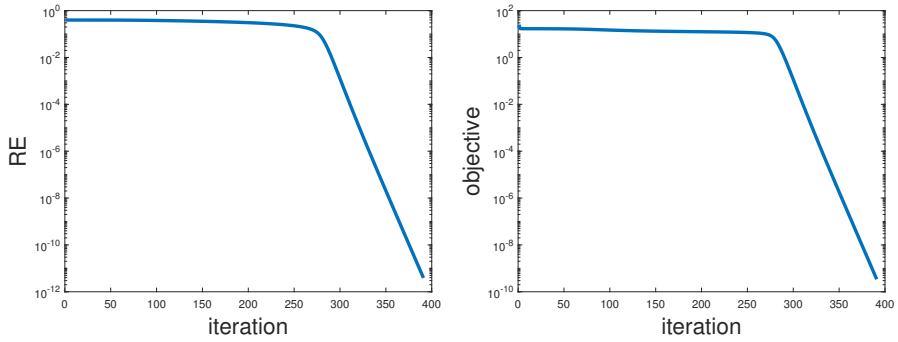
REFERENCES

[1] J. Barzilai and J. M. Borwein, Two-point step size gradient methods, *IMA Journal of Numerical Analysis*, **8** (1988), 141-148.

[2] F. Bornemann and T. März, Fast image inpainting based on coherence transport, *Journal of Mathematical Imaging and Vision*, **28** (2007), 259-278.



Convergence for the Checkerboard Image Test



Convergence for the Facade Image Test

FIGURE 7. Convergence of our method in color image inpainting.

[3] J. D. Carroll and J. J. Chang, Analysis of individual difference in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition, *Psychometrika*, **35** (2013), 283-319.

[4] X. Chen and J. Qin, Regularized Kaczmarz algorithms for tensor recovery, *SIAM Journal on Imaging Sciences*, **14** (2021), 1439-1471.

[5] A. Criminisi, P. Pérez and K. Toyama, Region filling and object removal by exemplar-based image inpainting, *IEEE Transactions on Image Processing*, **13** (2004), 1200-1212.

[6] L. De Lathauwer, B. De Moor and J. Vandewalle, A multilinear singular value decomposition, *SIAM Journal on Matrix Analysis and Applications*, **21** (2000), 1253-1278.

[7] L. De Lathauwer, B. De Moor and J. Vandewalle, On the best rank-1 and rank-$(r_1, r_2, ..., r_n)$ approximation of higher-order tensors, *SIAM Journal on Matrix Analysis and Applications*, **21** (2000), 1324-1342.

[8] J. Duchi, E. Hazan and Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *Journal of Machine Learning Research*, **12** (2011), 2121-2159.

[9] N. Durgin, R. Grotheer, C. Huang, S. Li, A. Ma, D. Needell and J. Qin, Fast hyperspectral diffuse optical imaging method with joint sparsity, In *41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, Berlin, Germany, 2019, 4758-4761.

[10] H. Fan, Y. Chen, Y. Guo, H. Zhang and G. Kuang, Hyperspectral image restoration using low-rank tensor recovery, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10** (2017), 4589-4604.

[11] S. Foucart and S. Subramanian, Iterative hard thresholding for low-rank recovery from rank-one projections, *Linear Algebra and its Applications*, **572** (2019), 117-134.

[12] R. A. Harshman, Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis, *UCLA Working Papers in Phonetics*, **16** (1970), 1-84.

[13] C. J. Hillar and L.-H. Lim, Most tensor problems are np-hard, *Journal of the ACM (JACM)*, **60** (2013), 1-39.

[14] M. E. Kilmer, K. Braman, N. Hao and R. C. Hoover, Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging, *SIAM Journal on Matrix Analysis and Applications*, **34** (2013), 148-172.
[15] M. E. Kilmer and C. D. Martin, Factorization strategies for third-order tensors, *Linear Algebra and its Applications*, **435** (2011), 641-658.
[16] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, *SIAM Review*, **51** (2009), 455-500.
[17] O. Le Meur, M. Ebdelli and C. Guillemot, Hierarchical super-resolution-based inpainting, *IEEE Transactions on Image Processing*, **22** (2013), 3779-3790.
[18] X.-Y. Liu, S. Aeron, V. Aggarwal and X. Wang, Low-tubal-rank tensor completion using alternating minimization, In *Modeling and Simulation for Defense Systems and Applications XI*, volume 9848, International Society for Optics and Photonics, 2016, 984809.
[19] Y. Liu, X. Zeng and W. Wang, Proximal gradient algorithm for nonconvex low tubal rank tensor recovery, *BIT Numerical Mathematics*, **63** (2023), 25.
[20] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin and S. Yan, Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 5249-5257.
[21] C. D. Martin, R. Shafer and B. LaRue, An order-p tensor factorization with applications in imaging, *SIAM Journal on Scientific Computing*, **35** (2013), A474-A490.
[22] D. Needell and R. Ward, Batched stochastic gradient descent with weighted sampling, In *International Conference Approximation Theory*, Springer, 2016, 279-306.
[23] Y. E. Nesterov, A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$, In *Doklady AN USSR*, **269** (1983), 543-547.
[24] N. Nguyen, D. Needell and T. Woolf, Linear convergence of stochastic iterative greedy algorithms with sparse constraints, *IEEE Transactions on Information Theory*, **63** (2017), 6869-6895.
[25] M. Nimishakavi, P. K. Jawanpuria and B. Mishra, A dual framework for low-rank tensor completion, In *Advances in Neural Information Processing Systems*, (2018), 5484-5495.
[26] L. Qi, Y. Chen, M. Bakshi and X. Zhang, Triple decomposition and tensor recovery of third order tensors, *SIAM Journal on Matrix Analysis and Applications*, **42** (2021), 299-329.
[27] J. Qin, S. Li, D. Needell, A. Ma, R. Grotheer, C. Huang and N. Durgin, Stochastic greedy algorithms for multiple measurement vectors, *Inverse Problems & Imaging*, **15** (2021), 79-107.
[28] H. Rauhut, R. Schneider and Ž. Stojanac, Low rank tensor recovery via iterative hard thresholding, *Linear Algebra and its Applications*, **523** (2017), 220-262.
[29] L. R. Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika*, **31** (1966), 279-311.
[30] A. Wang, Z. Lai and Z. Jin, Noisy low-tubal-rank tensor completion, *Neurocomputing*, **330** (2019), 267-279.
[31] A. Wang, D. Wei, B. Wang and Z. Jin, Noisy low-tubal-rank tensor completion through iterative singular tube thresholding, *IEEE Access*, **6** (2018), 35112-35128.
[32] W.-H. Xu, X.-L. Zhao and M. Ng, A fast algorithm for cosine transform based tensor singular value decomposition, arXiv preprint, `arXiv:1902.03070`, 2019.
[33] F. Zhang, W. Wang, J. Hou, J. Wang and J. Huang, Tensor restricted isometry property analysis for a large class of random measurement ensembles, *Sci. China Inf. Sci.*, **64** (2021), Paper No. 119101, 3 pp, arXiv preprint, `arXiv:1906.01198`, 2019.
[34] F. Zhang, W. Wang, J. Huang, Y. Wang and J. Wang, Rip-based performance guarantee for low-tubal-rank tensor recovery, *J. Comput. Appl. Math.*, **374** (2020), 112767, 14 pp, arXiv preprint, `arXiv:1906.01774`, 2019.
[35] X. Zhang and M. K. Ng, A corrected tensor nuclear norm minimization method for noisy low-rank tensor completion, *SIAM Journal on Imaging Sciences*, **12** (2019), 1231-1273.
[36] Z. Zhang, G. Ely, S. Aeron, N. Hao and M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-SVD, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 3842-3849.

Received March 2023; 1st revision August 2023; 2nd revision October 2023; early access January 2024.