

Generalized Nonconvex Approach for Low-Tubal-Rank Tensor Recovery

Hailin Wang^{ID}, Graduate Student Member, IEEE, Feng Zhang^{ID}, Jianjun Wang^{ID}, Member, IEEE,
Tingwen Huang^{ID}, Fellow, IEEE, Jianwen Huang^{ID}, and Xinling Liu

Abstract—The tensor–tensor product-induced tensor nuclear norm (t-TNN) (Lu *et al.*, 2020) minimization for low-tubal-rank tensor recovery attracts broad attention recently. However, minimizing the t-TNN faces some drawbacks. For example, the obtained solution could be suboptimal to the original problem due to its loose approximation. In this article, we extract a unified nonconvex surrogate of the tensor tubal rank as a tighter regularizer, which involves many popular nonconvex penalty functions. An iterative reweighted t-TNN algorithm is proposed to solve the resulting generalized nonconvex tubal rank minimization for tensor recovery. It converges to a critical point globally with rigorous proofs based on the Kurdyka–Łojasiewicz property. Furthermore, we provide the theoretical guarantees for exact and robust recovery by developing the tensor null space property. Extensive experiments demonstrate that our approach markedly enhances recovery performance compared with several state-of-the-art convex and nonconvex methods.

Index Terms—Convergence analysis, iterative reweighted algorithm, low-rank tensor recovery (LRTR), nonconvex optimization, recovery theory.

I. INTRODUCTION

DATA structures are becoming more and more complex in many fields, e.g., signal processing, machine learning, computer vision, and pattern recognition. Tensors, also called multidimensional arrays, as the natural and essential representation form of a wide range of real-world data including color images, videos, hyperspectral data, 3-D range data, reflectance data, and beyond, fully preserve the intrinsic structural information of these multirelational or multimodal data, making

Manuscript received 9 January 2020; revised 10 August 2020 and 11 December 2020; accepted 8 January 2021. Date of publication 29 January 2021; date of current version 4 August 2022. This work was supported by the National Natural Science Foundation of China under Grant 12071380, Grant 61673015, and Grant 11901476. (Corresponding author: Jianjun Wang.)

Hailin Wang, Feng Zhang, and Xinling Liu are with the School of Mathematics and Statistics, Southwest University, Chongqing 400715, China (e-mail: wanghailin97@163.com; zhangf@email.swu.edu.cn; fsluxl@163.com).

Jianjun Wang is with the School of Mathematics and Statistics, Southwest University, Chongqing 400715, China, and also with the Research Institute of Intelligent Finance and Digital Economics, Southwest University, Chongqing 400715, China (e-mail: wjj@swu.edu.cn).

Tingwen Huang is with the Department of Mathematics, Texas A&M University at Qatar, Doha 23874, Qatar (e-mail: tingwen.huang@qatar.tamu.edu).

Jianwen Huang is with the School of Mathematics and Statistics, Tianshui Normal University, Tianshui 741001, China (e-mail: hjwt303987297@126.com).

This article has supplementary material provided by the authors and color versions of one or more figures available at <https://doi.org/10.1109/TNNLS.2021.3051650>.

Digital Object Identifier 10.1109/TNNLS.2021.3051650

them have inherent advantages over vectors and matrices. Many practical problems can be converted to the low-rank tensor recovery (LRTR), such as images/videos inpainting [2], recommendation systems [3], background modeling [4], and 3-D light field displays [5]. It aims to recover an underlying tensor from its incomplete or disturbed observations, formulated as follows:

$$\min_{\mathcal{X}} \text{rank}(\mathcal{X}), \quad \text{s.t. } \Phi(\mathcal{X}) = \mathcal{T} \quad (1)$$

where \mathcal{T} is the observed measurement by a linear operator $\Phi(\cdot)$, and $\text{rank}(\cdot)$ represents the rank function of a tensor.

It is known that the most popular method for low-rank matrix recovery is nuclear norm minimization. The seemingly innocent task of generalizing this idea from matrices to tensors, whereas, turns out to be rather subtle, as basic notions of the tensor nuclear norm, and even tensor rank, are still ambiguous or nonuniform. Based on different tensor decompositions, there exist different types of tensor rank and corresponding nuclear norm. The CANDECOMP/PARAFAC (CP) rank equals the smallest number of rank-1 tensors achieving the CP decomposition [6] but generally NP-hard to estimate accurately [7]. Another popular rank is the Tucker rank induced by the Tucker decomposition [8]. It is a multilinear rank formed by matrix rank and, thus, computable, defined as a vector whose i th element equals the rank of the tensor's mode- i unfolding matrix. Accordingly, several notions of the tensor nuclear norm emerged. Liu *et al.* [9], [10] first proposed the sum of nuclear norm (SNN) as the convex surrogate of the Tucker rank, which significantly promotes the development of the LRTR problem. For example, by using the SNN, Zhang *et al.* [11], [12] proposed a general framework incorporating the features of rectification and alignment simultaneously for robust tensor recovery. Unfortunately, SNN is not a tight convex relaxation of the Tucker rank, and such a matricization technique fails to exploit the structure information completely [13]. A more well-established definition of tensor nuclear norm satisfying dual relationship with the defined spectral norm is proposed in [14] and [15] under the CP decomposition. However, it is usually intractable in computation [16]. Except for the above, some other definitions also exist (refer to [17]–[19]). Overall, exploring tensor rank and its efficient computational form within a certain decomposition scheme is always an open topic.

The algebraic framework of tensor Singular Value Decomposition (t-SVD) induced by tensor–tensor product (t-product)



Fig. 1. Illustrations of the low-tubal-rank property of natural color image and video. (a) and (d) Color image and video that can be modeled as tensor $\mathcal{A} \in \mathbb{R}^{400 \times 400 \times 3}$ and $\mathcal{B} \in \mathbb{R}^{144 \times 176 \times 100}$, respectively. (b) and (e) Plots of the singular values of \mathcal{A} and \mathcal{B} , respectively. (c) and (f) Approximations by tensor $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ with tubal rank 40 and 20, respectively.

[20] may give a more satisfying answer, as a more appropriate extension of singular value decomposition (SVD), which provides an alternative without drawbacks of CP rank and Tucker rank since t-SVD is always computable and does not need matricization. It decomposes a tensor into the t-product of two orthogonal tensors and one f-diagonal tensor (also called singular value tensor), and this motivates a new tensor rank, i.e., tensor tubal rank [21], defined as the number of nonzero singular tubes of the singular value tensor. Many multidimensional data in real world can be well approximated by a low-tubal-rank tensor with the fact that most of the singular values of the corresponding tensor are relatively small, and the few large ones contain the main information. See Fig. 1 for illustrations.¹ The tensor tubal rank can well characterize the intrinsic structure of a tensor, and thus, solving low-tubal-rank-related problems starts to attract more and more reasonable research interest these years [22]–[26].

This work focuses on the **Low-Tubal-Rank Tensor Recovery (LTRTR)** problem under the t-SVD framework. For the underlying tensor with low-tubal-rank assumption, it can be recovered by solving the following optimization:

$$\min_{\mathcal{X}} \lambda \text{rank}_t(\mathcal{X}) + \frac{1}{2} \|\Phi(\mathcal{X}) - \mathcal{T}\|_F \quad (2)$$

where $\lambda > 0$ is a regularization parameter, $\text{rank}_t(\cdot)$ denotes the tensor tubal rank, and $\|\cdot\|_F$ denotes the Frobenius norm. In general, solving problem (2) needs relaxation techniques since minimizing the rank function directly is NP-hard. In particular, in the recent work [1], Lu *et al.* first defined the tensor average rank and deduced that a tensor always has low average rank if it has low tubal rank and then proposed a novel tensor nuclear norm that is proven to be the convex envelope of the tensor average rank within the unit ball of the tensor spectral norm, which performed much better than other types of tensor nuclear norm in many tasks. For a clear distinction, we refer it as **t**-product-induced Tensor Nuclear Norm (**t**-TNN). Then, the LTRTR problem can be relaxed as

$$\min_{\mathcal{X}} \lambda \|\mathcal{X}\|_* + \frac{1}{2} \|\Phi(\mathcal{X}) - \mathcal{T}\|_F \quad (3)$$

where $\|\cdot\|_*$ denotes the t-TNN. The t-TNN minimization model (3) can be solved efficiently by convex optimization and achieves excellent performance in experiments [1], [22].

However, even though t-TNN relaxation is becoming a popular scheme for the LTRTR problem, it is also associated

with some shortcomings. First, there still exists a considerable gap separated from the tensor tubal rank minimization since t-TNN is a loose approximation of the tensor tubal rank, which usually results in the over-penalization of the optimization problem and the suboptimality of the obtained solution to the original problem. Besides, t-TNN minimization would also cause some unavoidable biases in real applications. For example, the theoretical assumption (typically like the incoherence condition [22]) could be violated for completing an extremely imbalanced tensor. To break the limits resulted from the convex relaxation method, we aim to solve the LTRTR problem by adopting the nonconvex relaxation strategy. We need to point out that the nonconvex method has already been well studied and applied in many sparse vectors and low-rank matrix recovery problems, adequately revealing its superiority in accuracy compared to the classical convex method. Nonetheless, studies on nonconvex LTRTR are rather limited at present. Moreover, existing works mainly focus on the optimization strategy, and the designed algorithms and corresponding convergence analysis, sadly, are only applicable to one certain nonconvex substitution [27]–[29]. Motivated by the above concerns, we propose a generalized framework for nonconvex tubal rank minimization in this article, including an iterative reweighted algorithm with global convergence and the theoretical guarantees for an exact and robust recovery, thereby dramatically enhancing the potential efficiency and capacity of the LTRTR in real-world applications.

A. Related Works

In order to introduce and highlight our contributions, we go over some related works from models and algorithms to theories.

1) Models: Some nonconvex models for the LTRTR problem need to be introduced first. In [30], by extending the Schatten- p norm, a t-Schatten- p norm is proposed to approximate the tensor tubal rank. Another nonconvex method approximates the tensor tubal rank by transforming each element in the sum form of t-TNN with the Laplace function [27]. Other similar models include [28], [29], which all use the nonconvex penalties studied in low-dimensional structured data, but only use one single function, and the solvers are varied with their corresponding penalty functions. Some other nonconvex surrogates induced from the t-TNN were proposed, e.g., weighted t-TNN [31] and the partial sum of t-TNN [32]. All these models can be viewed as special cases of this work,

¹The test image and video come from the USC-SIPI Image Database and the YUV Video Sequences Database, respectively.

and our approach can solve them all. To the best of our knowledge, there does not exist a generalized method for the LTRTR problem.

2) *Algorithms*: Solving a generalized nonconvex problem can be very hard. Still, there exist several algorithms that can be referred to. An early method is the multistage convex relaxation (or DC programming) [33]. This strategy solves the nonconvex problem by a sequence of convex optimizations but with a high computational cost. More efficiently, Gong *et al.* [34] proposed a general iterative shrinkage and thresholding algorithm. Continuing along this vein, Lu *et al.* [35] developed this general problem from sparse vector to low-rank matrix. Correlatively, they also developed the iteratively reweighted nuclear norm (IRNN) algorithm based on a key fact that the gradient of a concave function is nonincreasing [36], [37]. IRNN is truly a simple and efficient algorithm for nonconvex rank minimization and attracts a lot of attention. It is also a variant extension of the iterative reweighted least square (IRLS) in a 2-D matrix problem [38]. However, extending this idea of the iterative reweighted algorithm to multidimensional tensor recovery is still a big challenge on account of the incompleteness of the tensor algebraic framework and the nonuniqueness of the tensor rank function.

3) *Theories*: Finally, we add a brief investigation into the tensor recovery theories under the t-SVD framework. Typically, based on tensor incoherent condition, Zhang and Aeron [22] and Lu *et al.* [16] provided the exact recovery guarantees for two important problems, respectively, i.e., tensor completion and tensor robust principal component analysis. Lu *et al.* [39] give the sample conditions from the Gaussian measurement for exact and robust LTRTR. In our recent work, a tensor restricted isometry property (RIP) is proposed to guarantee the recovery performance of the t-TNN minimization. One thing in common is that all these works consider t-TNN-related models, and the theoretical results are motivated by sparse vector and low-rank matrix recovery. Note that some other theoretical tools have not been extended to LTRTR, such as the null space property (NSP) [40] and spherical section property (SSP) [41]. In addition, for the nonconvex problems, existing works mainly target the optimization algorithm for a certain specific model, e.g., [27]–[29], while the corresponding theory guarantees are rarely studied.

B. Our Contributions

The main contributions in this work are fourfold.

First, we extract a unified nonconvex surrogate of the tensor tubal rank, which not only equips universality for a large group of nonconvex penalty functions but also extends parallelly the nonconvex surrogates in sparse vector and low-rank matrix recovery. It achieves a better approximation than t-TNN and, thus, leads to a generalized nonconvex tubal rank minimization model. More broadly, the observation loss function is only to be Lipschitz gradient continuous. Many nonconvex relaxation models in related works are special cases of the proposed model.

Second, inspired by the IRNN, an Iterative Reweighted t-TNN (**IR-t-TNN**) algorithm is proposed to solve the generalized nonconvex problem, for which we exploit a weighted tensor singular value thresholding (WTSVT) operator. At the same time, the Barzilai–Borwein initialization is adopted in each outer iteration for accelerating convergence. The IR-t-TNN is proven to decrease the objective function value monotonically, and any accumulation point can be a critical point. Based on the Kurdyka–Łojasiewicz property, we further prove a stronger convergence that the solution sequence has a finite length and converges to a critical point globally. In particular, the time cost has been reduced efficiently compared to that of applying the IRNN for tensors slice by slice.

Third, for the generalized nonconvex tubal rank minimization, we establish the theoretical guarantees for exact and robust recovery with the developed tensor NSP conditions. Surprisingly, the results can be easily degraded to that of t-TNN minimization. Moreover, we rigorously prove that the theoretical conditions for the generalized nonconvex model are weaker than that of the convex t-TNN model.

Fourth, the proposed IR-t-TNN has a stronger recovery capacity in recovery than t-TNN and a faster speed in convergence than IRNN through experimental verification. At the same time, we apply it for various real-world applications, including natural image restoration, face inpainting, video recovery, and network traffic estimation, and the results confirm its superiority over the classical convex methods and state-of-the-art nonconvex methods.

C. Organization

This article is organized as follows. Section II gives the notations and some preliminaries. In Section III, we present the unified nonconvex surrogate and its resulting generalized nonconvex problem. The main algorithm and its convergence analysis are given in Section IV. Section V contains the theoretical guarantees, and Section VI reports the experimental results. Finally, we conclude our work in Section VII. Note that the whole proofs are given in appendixes.

II. NOTATIONS AND PRELIMINARIES

A. Notations

First, we introduce some notations throughout this article. For brevity, the main notations are listed in Table I. $t_+ = \max\{0, t\}$. $\lceil t \rceil$ denotes the nearest integer greater than or equal to t , and $\lfloor t \rfloor$ is the one less than or equal to t . The inner product of tensors \mathcal{T}_1 and \mathcal{T}_2 is defined as $\langle \mathcal{T}_1, \mathcal{T}_2 \rangle = \sum_k \langle \mathcal{T}_1^{(k)}, \mathcal{T}_2^{(k)} \rangle$. Operator unfold transforms $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ into a $n_1 n_3 \times n_2$ matrix along the third dimension, i.e., $\text{unfold}(\mathcal{T}) = [\mathcal{T}^{(1)}; \mathcal{T}^{(2)}; \dots; \mathcal{T}^{(n_3)}]$, and fold is its inverse operator. $\tilde{\mathcal{T}} = \text{fft}(\mathcal{T}, [], 3)$ denotes the discrete Fourier transformation (DFT) on \mathcal{T} along the third dimension using the MATLAB command fft. The inverse DFT is computed by command ifft satisfying $\mathcal{T} = \text{ifft}(\tilde{\mathcal{T}}, [], 3)$. We say \mathcal{T} is in original domain and $\tilde{\mathcal{T}}$ is in Fourier domain. Finally, the block diagonal matrix and block circulant matrix

TABLE I
SOME NOTATIONS

Notations	Descriptions	Notations	Descriptions
\mathbb{R}	real field	\mathbb{C}	complex field
t	scalar	\mathbf{t}	vector
\mathbf{T}	matrix	\mathcal{T}	tensor
\mathcal{T}_0	true tensor	\mathcal{T}_*	recovered tensor
\mathcal{T}_{ijk}	(i, j, k) -th entry	$\ \mathcal{T}\ _1$	$\sum_{ijk} \mathcal{T}_{ijk} $
$\mathcal{T}(i, j, :)$	(i, j) -th tube	$\ \mathcal{T}\ _F$	$\sqrt{\sum_{ijk} \mathcal{T}_{ijk}^2}$
$\mathcal{T}^{(k)}$	k -th frontal slice	$\ \mathcal{T}\ _\infty$	$\max_{ijk} \mathcal{T}_{ijk} $

are needed to be introduced, that is

$$\text{bdiag}(\bar{\mathcal{T}}) = \begin{bmatrix} \bar{\mathcal{T}}^{(1)} \\ & \bar{\mathcal{T}}^{(2)} \\ & & \ddots \\ & & & \bar{\mathcal{T}}^{(n_3)} \end{bmatrix}$$

and

$$\text{bcirc}(\mathcal{T}) = \begin{bmatrix} \mathcal{T}^{(1)} & \mathcal{T}^{(n_3)} & \dots & \mathcal{T}^{(2)} \\ \mathcal{T}^{(2)} & \mathcal{T}^{(1)} & \dots & \mathcal{T}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{T}^{(n_3)} & \mathcal{T}^{(n_3-1)} & \dots & \mathcal{T}^{(1)} \end{bmatrix}.$$

B. T-SVD Algebraic Framework

Now, we give the basic definitions on t-SVD and outline the associated algebraic framework. We first introduce the t-product, which is defined as a generalization of matrix product.

Definition 1 (T-Product [20]): Let $\mathcal{T}_1 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{T}_2 \in \mathbb{R}^{n_2 \times l \times n_3}$, and then, the t-product $\mathcal{T}_1 * \mathcal{T}_2 \in \mathbb{R}^{n_1 \times l \times n_3}$ is defined as

$$\mathcal{T}_1 * \mathcal{T}_2 := \text{fold}(\text{bcirc}(\mathcal{T}_1) \cdot \text{unfold}(\mathcal{T}_2)).$$

Before we introduce t-SVD, there are some notions that should be introduced, which are extended from the matrix cases.

Definition 2 (Conjugate Transpose [20]): The conjugate transpose of $\mathcal{T} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is denoted by $\mathcal{T}^\top \in \mathbb{C}^{n_2 \times n_1 \times n_3}$, obtained by conjugate transposing each frontal slice and then reversing the order of transposed frontal slices 2 through n_3 .

Definition 3 (Identity Tensor [20]): The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is the tensor whose first frontal slice is a $n \times n$ identity matrix and other frontal slices are all zeros.

Definition 4 (Orthogonal Tensor [20]): A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies $\mathcal{Q}^\top * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^\top = \mathcal{I}$.

Definition 5 (F-Diagonal Tensor [20]): A tensor is called f-diagonal if each of its frontal slices is a diagonal matrix.

As stated above, the t-SVD is proposed that any third-order tensor can be factorized into three components.

Theorem 1 (T-SVD [1], [20]): For any tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, it can be decomposed as

$$\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$$

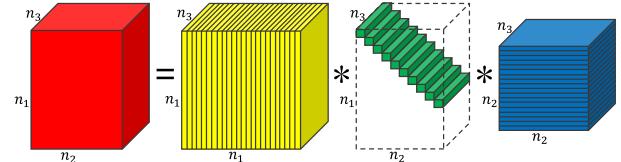


Fig. 2. Illustration of the t-SVD for a third-order tensor.

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f-diagonal tensor.

Fig. 2 gives an illustration of the t-SVD. One efficient algorithm for computing the t-SVD is presented in [1], which reduces the number of computing SVD from n_3 to $\lceil((n_3 + 1)/2)\rceil$ (see Algorithm 1). Finally, based on t-SVD, the definitions of tensor tubal rank and tensor nuclear norm can be derived.

Algorithm 1 t-SVD [1]

Input: $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.

Output: $\mathcal{U}, \mathcal{S}, \mathcal{V}$.

1. Compute $\bar{\mathcal{T}} = \text{fft}(\mathcal{T}, [], 3)$.
2. Compute each frontal slice of $\bar{\mathcal{U}}, \bar{\mathcal{S}}$ and $\bar{\mathcal{V}}$ by
 - for $i = 1, \dots, \lceil \frac{n_3+1}{2} \rceil$ do
 $[\bar{\mathcal{U}}^{(i)}, \bar{\mathcal{S}}^{(i)}, \bar{\mathcal{V}}^{(i)}] = \text{SVD}(\mathcal{T}^{(i)})$;
 - end for
 - for $i = \lceil \frac{n_3+1}{2} \rceil + 1, \dots, n_3$ do
 $\bar{\mathcal{U}}^{(i)} = \text{conj}(\bar{\mathcal{U}}^{(n_3+2-i)})$;
 - $\bar{\mathcal{S}}^{(i)} = \bar{\mathcal{S}}^{(n_3+2-i)}$;
 - $\bar{\mathcal{V}}^{(i)} = \text{conj}(\bar{\mathcal{V}}^{(n_3+2-i)})$;
- end for

3. Compute $\mathcal{U}, \mathcal{S}, \mathcal{V}$ by

$$\mathcal{U} = \text{ifft}(\bar{\mathcal{U}}, [], 3), \mathcal{S} = \text{ifft}(\bar{\mathcal{S}}, [], 3), \mathcal{V} = \text{ifft}(\bar{\mathcal{V}}, [], 3).$$

Definition 6 (Tensor Tubal Rank [21]): For a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with t-SVD that $\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, its tensor tubal rank, denoted as $\text{rank}_t(\mathcal{T})$, is defined as the number of nonzero singular tubes of \mathcal{S} , that is

$$\text{rank}_t(\mathcal{T}) := \#\{i, \mathcal{S}(i, i, :) \neq 0\}.$$

Equivalently, similar to the matrix case, $\text{rank}_t(\mathcal{T})$ is equal to the number of its nonzero singular values [1], i.e., $\text{rank}_t(\mathcal{T}) := \#\{i, \mathcal{S}(i, i, 1) \neq 0\}$, which uses the following property induced by inverse DFT:

$$\mathcal{S}(i, i, 1) = \frac{1}{n_3} \sum_{j=1}^{n_3} \bar{\mathcal{S}}(i, i, j). \quad (4)$$

Denote $\sigma(\mathcal{T}) \in \mathbb{R}^{m \times n_3}$ with $\sigma_{ij} := \bar{\mathcal{S}}(i, i, j)$ and $m = \min\{n_1, n_2\}$, and the t-TNN is defined as follows.

Definition 7 (t-TNN): Let $\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ be the t-SVD of $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and its nuclear norm induced by t-product (or t-SVD) is defined as

$$\|\mathcal{T}\|_* := \frac{1}{n_3} \sum_{i=1}^m \sum_{j=1}^{n_3} \sigma_{ij}.$$

Remark 1: The above t-TNN is defined as the sum of all singular values of all frontal slices up to a constant factor

TABLE II
FORMULAS OF POPULAR NONCONVEX PENALTY FUNCTIONS

Penalty	Formula ($x \geq 0, \lambda > 0, \gamma > 1$)
ETP [42]	$\frac{\lambda}{1-\exp(-\gamma)}(1-\exp(-\gamma x))$
Geman [43]	$\frac{\lambda x}{x+\gamma}$
Laplace [44]	$\lambda(1-\exp(-\frac{x}{\gamma}))$
Logarithm [45]	$\frac{\lambda}{\log(\gamma+1)} \log(\gamma x + 1)$
L_p [46]	$\lambda x^p, 0 < p < 1$
MCP [47]	$\begin{cases} \lambda x - \frac{x^2}{2\gamma}, & x \leq \gamma\lambda, \\ \frac{1}{2}\gamma\lambda^2, & x > \gamma\lambda. \end{cases}$
Capped L_1 [48]	$\begin{cases} \lambda x, & x < \gamma, \\ \lambda\gamma, & x \geq \gamma. \end{cases}$
SCAD [49]	$\begin{cases} \lambda x, & x < \lambda, \\ \frac{-x^2+2\gamma\lambda x-\lambda^2}{2(\gamma-1)}, & \lambda \leq x < \gamma\lambda, \\ \frac{\lambda^2(\gamma+1)}{2}, & x \geq \gamma\lambda. \end{cases}$

in the Fourier domain. Note that it is somewhat different in form from that in [1] but actually equivalent by using the property (4). This type of multisummation representation contains all basic singular values in the Fourier domain, which reveals a more complete form of SVD result in the multidimensional case. Besides, it degenerates to the matrix nuclear norm when $n_3 = 1$.

Definition 8 (Tensor Average Rank [1]): For a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor average rank, denoted as $\text{rank}_a(\mathcal{T})$, is defined as $\text{rank}_a(\mathcal{T}) := (1/n_3) \text{rank}(\text{brirc}(\mathcal{T}))$.

Definition 9 (Tensor Spectral Norm [1]): The tensor spectral norm of $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as $\|\mathcal{T}\| := \|\text{brirc}(\mathcal{T})\|$.

Proposition 1 (Relevant Properties): There exist some properties induced from t-SVD framework with t-product.

- 1) $\mathcal{T} = \mathcal{T}_1 * \mathcal{T}_2 \Leftrightarrow \text{bdiag}(\tilde{\mathcal{T}}) = \text{bdiag}(\tilde{\mathcal{T}}_1) \text{bdiag}(\tilde{\mathcal{T}}_2)$.
- 2) $\|\mathcal{T}\|_F^2 = \|\mathcal{S}\|_F^2 = (1/n_3) \|\text{bdiag}(\tilde{\mathcal{T}})\|_F^2$.
- 3) $(1/n_3) \|\text{bdiag}(\tilde{\mathcal{T}})\|_* = \|\mathcal{T}\|_* \leq \sqrt{m} \|\mathcal{T}\|_F$.

III. GENERALIZED NONCONVEX TUBAL RANK MINIMIZATION

In this section, we first extract a unified nonconvex surrogate of tensor tubal rank and then propose the generalized nonconvex tubal-rank minimization.

A. Unified Nonconvex Surrogate

Nonconvex relaxation methods have long been investigated in low-dimensional space. In compressed sensing (CS), there exist many nonconvex surrogates of the L_0 -norm of $\mathbf{x} \in \mathbb{R}^n$, that is $\|\mathbf{x}\|_0 := \sum_{i=1}^n \mathbf{I}_{\{x_i \neq 0\}}$, by applying a nonconvex penalty function ψ to $|x_i|$, that is

$$\sum_{i=1}^n \psi(|x_i|).$$

Continuing along this vein, many works exploit the nonconvex surrogate of the matrix rank function by extending the nonconvex functions onto the singular values of the matrix, formulated as

$$\sum_{i=1}^m \psi(\sigma_i(\mathbf{X}))$$

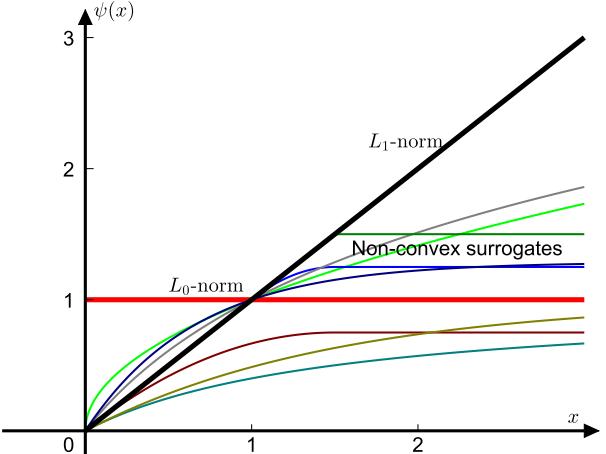


Fig. 3. Comparison of the L_0 -norm (red), L_1 -norm (black), and nonconvex surrogates (others) with $\lambda = 1$ and $\gamma = 1.5$.

where $\sigma_i(\mathbf{X})$ denotes the i th singular value of $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$. We summarize several popular nonconvex penalty functions in Table II. To some extent these penalty functions follow certain commonalities. More explicitly, they all coincidentally balance the gap between the essential sparse function and effective sparse function, that is, $\psi(x)$ always close to $|x|$ for small values and $\mathbf{I}\{x \neq 0\}$ for large values (see Fig. 3), which results in achieving better approximation of the L_0 -norm (rank function) compared with the famous L_1 -norm (nuclear norm). Heuristically, we propose a unified nonconvex surrogate of the tensor tubal rank.

Definition 10 (Unified Nonconvex Surrogate): Suppose that $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ has t-SVD $\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$. We define

$$\Psi(\mathcal{T}) := \frac{1}{n_3} \sum_{i=1}^m \sum_{j=1}^{n_3} \psi(\sigma_{ij}) \quad (5)$$

as the nonconvex surrogate of the tensor tubal rank, where $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a continuous, monotonically nondecreasing, and concave function that satisfies

$$\psi(0) = 0, \lim_{x \rightarrow +\infty} \frac{\psi(x)}{x} = 0.$$

Note that $\psi_{\lambda, \gamma}(\cdot)$ in (5) is usually a parametric function where λ works as scale parameter (also called regularization parameter in optimization models) and γ works as shape parameter. It is necessary to point that the assumptions in Definition 10 are actually quite simple, natural, and unified, which can be verified for all nonconvex penalty functions in Table II, which makes $\Psi(\mathcal{T})$ approximate the tensor tubal rank better than t-TNN. It can be viewed as a natural and generalized extension from sparse vector and low-rank matrix to low-tubal-rank tensor but not easy due to the complicity of the tensor decomposition. Table III summarizes the parallel concepts, which shows the coherence in form. In addition, $\Psi(\mathcal{T})$ satisfies the following properties.

Proposition 2: For $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\Psi(\mathcal{T})$ satisfies the following.

- 1) $\Psi(\mathcal{T})$ is continuous, monotonically nondecreasing, and concave with respect to σ_{ij} .

TABLE III
PARALLELISM OF SPARSE VECTOR, LOW-RANK MATRIX AND LOW-TUBAL-RANK TENSOR

	Essential Measure	Convex Surrogate	Non-convex Surrogate
Sparse Vector $\mathbf{x} \in \mathbb{R}^n$	$\ \mathbf{x}\ _0$	$\ \mathbf{x}\ _1 = \sum x_i $	$\sum \psi(x_i)$
Low-Rank Matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$	$\text{rank}(\mathbf{X})$	$\ \mathbf{X}\ _* = \sum \sigma_i(\mathbf{X})$	$\sum \psi(\sigma_i(\mathbf{X}))$
Low-Tubal-Rank Tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$	$\text{rank}_t(\mathcal{X})$	$\ \mathcal{X}\ _* = \frac{1}{n_3} \sum \sum \sigma_{ij}(\mathcal{X})$	$\frac{1}{n_3} \sum \sum \psi(\sigma_{ij}(\mathcal{X}))$

- 2) $\Psi(\mathcal{T})$ is separable and absolutely symmetric viewing it as a function of $\sigma(\mathcal{T})$.
- 3) $\Psi(\mathcal{T})$ is orthogonal invariant, i.e., for any orthogonal tensors $\mathcal{P} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{Q} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$, we have

$$\Psi(\mathcal{P} * \mathcal{T} * \mathcal{Q}^\top) = \Psi(\mathcal{T}).$$

Recall that the t-TNN is proven to be the convex envelope of the tensor average rank within the unit ball of the tensor spectral norm. We then claim that $\Psi(\mathcal{T})$ is a tighter but nonconvex envelope with appropriate parameters.

Proposition 3: Assume that $x \leq \psi_{\lambda,\gamma}(x) \leq 1$, ($x \leq 1$) by selecting proper parameters, and we have

$$\|\mathcal{T}\|_* \leq \Psi(\mathcal{T}) \leq \text{rank}_a(\mathcal{T})$$

on the set $\{\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \mid \|\mathcal{T}\| \leq 1\}$. In this sense, $\Psi(\mathcal{T})$ is a nonconvex envelope of the tensor average rank tighter than the t-TNN.

The tensor average rank is closely related to tensor tubal rank with $\text{rank}_a(\mathcal{T}) \leq \text{rank}_t(\mathcal{T})$; thus, it is expected that $\Psi(\mathcal{T})$ approximates tensor tubal rank better than the t-TNN. We point that the assumption in Proposition 3 can be easily satisfied. One can fix one of the parameters, and the other one can be solved from the equation $\psi_{\lambda,\gamma}(1) = 1$. Then, the assumption is verified based on the concavity of ψ .

B. Generalized Nonconvex Tubal Rank Minimization

Based on the advantage of the nonconvex surrogate in approximation, instead of the t-TNN minimization (3), in this article, we propose the following generalized nonconvex tubal rank minimization model

$$\min_{\mathcal{X}} \Psi(\mathcal{X}) + \frac{1}{2} \|\Phi(\mathcal{X}) - \mathcal{T}\|_F^2. \quad (6)$$

More broadly, we aim to solve a concave-convex problem contains a concave penalty term and a convex loss term, that is

$$\min_{\mathcal{X}} \mathcal{F}(\mathcal{X}) = \Psi(\mathcal{X}) + \ell(\mathcal{X}) \quad (7)$$

where $\ell(\mathcal{X})$ denotes a general loss function that satisfies the Lipschitz gradient continuous.

Definition 11 (Lipschitz Gradient Continuous): We call that a continuously differentiable function $\ell : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}$ is Lipschitz gradient continuous if there exists constant $L(\ell) > 0$ such that

$$\|\nabla \ell(\mathcal{X}) - \nabla \ell(\mathcal{Y})\|_F \leq L(\ell) \|\mathcal{X} - \mathcal{Y}\|_F$$

for any $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.

Note that model (7) is very general. For the penalty function, we can choose any known nonconvex penalty function

in Table II. For the loss function, most of the widely used ones, such as the least square and logistic loss functions, satisfy the assumption of Lipschitz gradient continuous. Thus, model (7) comprises of many models in related works [27]–[30].

IV. PROPOSED ALGORITHM SCHEME

This section presents the optimization scheme for solving the generalized problem (7). We first develop a proximal operator of the weighted t-TNN minimization and then propose the IR-t-TNN algorithm with detailed convergence analysis. Finally, we discuss the differences with the IRNN algorithm.

A. WTSVT Operator

We first give a WTSVT operator to minimize a weighted t-TNN optimization, which is the main subproblem for solving the original problem (7) as we will see later.

Definition 12 (Weighted t-TNN): For $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its weighted t-TNN is defined as the weighted sum of all singular values of all frontal slices in the Fourier domain, that is

$$\|\mathcal{T}\|_{W,*} := \frac{1}{n_3} \sum_{i=1}^m \sum_{j=1}^{n_3} w_{ij} \sigma_{ij}$$

where $\mathbf{W} = (w_{ij})_{m \times n_3}$ is the weight matrix.

The WTSVT operator is defined as follows to compute the proximal minimizer of the weighted t-TNN minimization.

Definition 13 (WTSVT Operator): For any $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with t-SVD that $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, we define the WTSVT operator as

$$\mathcal{D}_{W,\tau}(\mathcal{Y}) := \mathcal{U} * \mathcal{S}_{W,\tau} * \mathcal{V}^\top \quad (8)$$

where $\mathcal{S}_{W,\tau} = \text{ifft}((\bar{\mathcal{S}} - \tau \mathcal{W})_+, [\], 3)$, in which $\mathcal{W} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is f-diagonal whose diagonal entries of the j th frontal slice are equal to the j th column of weight matrix \mathbf{W} .

Theorem 2: For any $\tau > 0$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the WTSVT operator (8) obeys

$$\mathcal{D}_{W,\tau}(\mathcal{Y}) = \arg \min_{\mathcal{X}} \tau \|\mathcal{X}\|_{W,*} + \frac{1}{2} \|\mathcal{X} - \mathcal{Y}\|_F^2$$

if the weight matrix satisfies

$$0 \leq w_{1j} \leq w_{2j} \leq \dots \leq w_{mj}, \quad j = 1, 2, \dots, n_3.$$

Theorem 2 shows that the WTSVT operator $\mathcal{D}_{W,\tau}(\mathcal{Y})$ gives a close-from of the proximal minimizer of the weighted t-TNN minimization with monotone nonnegative weights, which is a natural extension of the weighted matrix SVD [50]. The feasibility of this extension is guaranteed by the t-SVD framework. Using Algorithm 1, we show how to compute $\mathcal{D}_{W,\tau}(\mathcal{Y})$ efficiently in Algorithm 2.

Algorithm 2 WTSVT

Input: $\tau > 0$, $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{n_3}] \in \mathbb{R}^{m \times n_3}$, and $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.

Output: $\mathcal{D}_{\mathbf{W}, \tau}(\mathbf{Y})$.

1. Compute $\bar{\mathbf{Y}} = \text{fft}(\mathbf{Y}, [], 3)$.
2. Compute each frontal slice of $\bar{\mathbf{D}}$ by

```

for  $i = 1, \dots, \lceil \frac{n_3+1}{2} \rceil$  do
     $[\bar{\mathbf{U}}^{(i)}, \bar{\mathbf{S}}^{(i)}, \bar{\mathbf{V}}^{(i)}] = \text{SVD}(\bar{\mathbf{Y}}^{(i)})$ ;
     $\bar{\mathbf{D}}^{(i)} = \bar{\mathbf{U}}^{(i)} \cdot (\bar{\mathbf{S}}^{(i)} - \tau \text{Diag}(\mathbf{w}_i))_+ \cdot \bar{\mathbf{V}}^{(i)}$ ;
end for
for  $i = \lceil \frac{n_3+1}{2} \rceil + 1, \dots, n_3$  do
     $\bar{\mathbf{D}}^{(i)} = \text{conj}(\bar{\mathbf{D}}^{(n_3+2-i)})$ ;
end for

```

3. Compute $\mathcal{D}_{\mathbf{W}, \tau}(\mathbf{Y}) = \text{ifft}(\bar{\mathbf{D}}, [], 3)$.

B. IR-t-TNN Algorithm

Now, we present the IR-t-TNN algorithm by updating alternately a weighted t-TNN optimization and corresponding weights. The k th iteration results are denoted by superscript k throughout this article.

For the general penalty function ψ in definition 10, based on its concavity, we have

$$\psi'(y) \leq \frac{\psi(x) - \psi(y)}{x - y} \leq \psi'(x) \quad (9)$$

for any $0 < x < y$. Note that when ψ is nondifferentiable at some nonsmooth points, the property (9) still holds, which becomes

$$q \leq \frac{\psi(x) - \psi(y)}{x - y} \leq p \quad (10)$$

where $p \in \partial\psi(x)$, $q \in \partial\psi(y)$. The above property is also called the antimonotone property in [37], which plays a key role for solving the subproblem of the proposed IR-t-TNN.

Our proposed IR-t-TNN algorithm solves the problem (7) by generating a iterative sequence $\{\mathbf{X}^{(k)}\}$, where $\mathbf{X}^{(k+1)}$ is the minimizer of a weighted t-TNN regularized problem that comes from the linearized approximation of the objective function $\mathcal{F}(\mathbf{X})$ at $\mathbf{X}^{(k)}$. To realize this, on the one hand, the property (10) yields that

$$\psi(\sigma_{ij}) \leq \psi(\sigma_{ij}^{(k)}) + w_{ij}^{(k)}(\sigma_{ij} - \sigma_{ij}^{(k)})$$

where $w_{ij}^{(k)} \in \partial\psi(\sigma_{ij}^{(k)})$. This makes

$$\Psi(\mathbf{X}) \leq \frac{1}{n_3} \sum_{i=1}^m \sum_{j=1}^{n_3} \left(\psi(\sigma_{ij}^{(k)}) + w_{ij}^{(k)}(\sigma_{ij} - \sigma_{ij}^{(k)}) \right). \quad (11)$$

On the other hand, for the loss function $\ell(\mathbf{X})$, we approximate it as a quadratic function

$$\begin{aligned} \ell(\mathbf{X}) &\doteq \ell(\mathbf{X}^{(k)}) + \langle \nabla \ell(\mathbf{X}^{(k)}), \mathbf{X} - \mathbf{X}^{(k)} \rangle \\ &\quad + \frac{\mu^{(k)}}{2} \|\mathbf{X} - \mathbf{X}^{(k)}\|_F^2. \end{aligned} \quad (12)$$

Then, by replacing the nonconvex penalty function Ψ and loss function ℓ by the right side of (11) and (12), respectively, we solve the following relaxed problem to update $\mathbf{X}^{(k+1)}$:

$$\begin{aligned} \mathbf{X}^{(k+1)} = \arg \min_{\mathbf{X}} & \frac{1}{n_3} \sum_{i=1}^m \sum_{j=1}^{n_3} \left(\psi(\sigma_{ij}^{(k)}) + w_{ij}^{(k)}(\sigma_{ij} - \sigma_{ij}^{(k)}) \right) \\ & + \ell(\mathbf{X}^{(k)}) + \langle \nabla \ell(\mathbf{X}^{(k)}), \mathbf{X} - \mathbf{X}^{(k)} \rangle \\ & + \frac{\mu^{(k)}}{2} \|\mathbf{X} - \mathbf{X}^{(k)}\|_F^2. \end{aligned}$$

After ignoring constant terms, it is equivalent to the following weighted t-TNN proximal operator problem:

$$\mathbf{X}^{(k+1)} = \arg \min_{\mathbf{X}} \frac{1}{\mu^{(k)}} \|\mathbf{X}\|_{W^{(k)},*} + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}^{(k)}\|_F^2 \quad (13)$$

where $\mathbf{Y}^{(k)} := \mathbf{X}^{(k)} - \nabla \ell(\mathbf{X}^{(k)})/\mu^{(k)}$. Since $\sigma_{1j}^{(k)} \geq \sigma_{2j}^{(k)} \geq \dots \geq \sigma_{mj}^{(k)} \geq 0$, using the antimonotone property, we have

$$0 \leq w_{1j}^{(k)} \leq w_{2j}^{(k)} \leq \dots \leq w_{mj}^{(k)}, \quad j = 1, 2, \dots, n_3 \quad (14)$$

which exactly satisfies Theorem 2. Thus, in IR-t-TNN, we first perform a weighted t-TNN minimization via the WTSVT operator and then update the weighting values.

Note that the parameter $\mu^{(k)}$ is should be selected properly and deliberately in each iteration since $\mu^{(k)}$ affects the accuracy of the approximated transformation (12) directly and $1/\mu^{(k)}$ acts as the step size in the proximal optimization (13). We use the popular backtracking strategy [51] that finds the smallest positive integer i_k such that, with $\mu^{(k)} = \rho^{i_k} \mu^{(k)}$, ($\rho > 1$), the following search criterion is satisfied:

$$\mathcal{F}(\mathbf{X}^{(k+1)}) + \frac{\alpha}{2} \mu^{(k)} \|\mathbf{X}^{(k+1)} - \mathbf{X}^k\|_F^2 \leq \mathcal{F}(\mathbf{X}^k) \quad (15)$$

where constant $0 < \alpha < 1$. In this article, instead of the initialization $\mu^{(k)} := \rho^{i_{k-1}} \mu^{(k-1)}$ in [51], we adopt the Barzilai–Borwein rule [52] to initialize $\mu^{(k)}$ as

$$\mu^{(k)} := \arg \min_{\mu} \|\Delta_1^{(k)} - \frac{1}{\mu} \Delta_2^{(k)}\|_F^2 = \frac{\langle \Delta_1^{(k)}, \Delta_2^{(k)} \rangle}{\langle \Delta_1^{(k)}, \Delta_1^{(k)} \rangle} \quad (16)$$

where $\Delta_1^{(k)} := \ell(\mathbf{X}^{(k)}) - \ell(\mathbf{X}^{(k-1)})$ and $\Delta_2^{(k)} := \nabla \ell(\mathbf{X}^{(k)}) - \nabla \ell(\mathbf{X}^{(k-1)})$. It has been demonstrated that the Barzilai–Borwein initialization makes finding an proper step size quickly and, thus, accelerates the algorithm convergence [34]. Finally, we give the detailed procedure of the IR-t-TNN in Algorithm 3.

C. Convergence Analysis

In this section, we establish a detailed convergence analysis for the proposed IR-t-TNN algorithm. Suppose that the sequence $\{\mathbf{X}^{(k)}, k = 0, 1, 2, \dots\}$ is generated by the IR-t-TNN, and we have the following convergence results.

Theorem 3 (Local Convergence): Assume that the objective function $\mathcal{F}(\mathbf{X})$ is coercive, i.e., $\mathcal{F}(\mathbf{X})$ is bounded from below, and $\mathcal{F}(\mathbf{X}) \rightarrow +\infty$ when $\mathbf{X} \rightarrow +\infty$ and the parameter $\mu^{(k)} \geq (L(\ell)/(1-\alpha))$, ($0 < \alpha < 1$) for any $k \geq 0$, and we have the following.

Algorithm 3 IR-t-TNN

Input: Formulas of Ψ and ℓ , $\rho > 1$, $0 < \alpha < 1$, and ϵ .
Initialize: $\mathcal{X}^{(0)} = \mathcal{O}$, $\mathbf{W}^{(0)} = \mathbf{0}$, $\mu^{(0)} > 0$, and $k = 0$.
Output: Solution $\mathcal{X}_* = \mathcal{X}^{(k)}$.

while not converge **do**

1. Update $\mathcal{X}^{(k+1)}$ by
- repeat**
- Compute $\mathcal{X}^{(k+1)}$ by solving (13) via Algorithm 2;
- Compute $\mu^{(k)} = \rho\mu^{(k)}$;
- until** the criterion (15) is satisfied.
2. Update $\mathbf{W}^{(k+1)}$ by $w_{ij}^{k+1} \in \partial\psi(\sigma_{ij}(\mathcal{X}^{(k+1)}))$.
3. Check convergence condition $\|\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}\|_\infty \leq \epsilon$.
4. Update $k \leftarrow k + 1$.
5. Barzilai-Borwein Initialization: computing $\mu^{(k)}$ by (16).

end while

- 1) The monotone search criterion (15) is always satisfied, which implies that $\mathcal{F}(\mathcal{X}^{(k)})$ is monotonically decreasing.
- 2) The sequence $\{\mathcal{X}^{(k)}\}$ is always bounded. Moreover, $\sum_{k=0}^{+\infty} \|\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}\|_F^2 < +\infty$, which implies the sequence's asymptotic regularity, that is, $\lim_{k \rightarrow +\infty} (\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}) = 0$.
- 3) Any accumulation point \mathcal{X}^* of the sequence $\{\mathcal{X}^{(k)}\}$ is a critical point of $\mathcal{F}(\mathcal{X})$.

Theorem 3 presents that the IR-t-TNN algorithm obtains a bounded sequence making the objective function value decreasing, and there exists a subsequence that converges to a critical point. To further seek a stronger convergence guarantee, we introduce the Kurdyka–Łojasiewicz (KL) property [53], [54], which is a very popular tool for proving algorithm convergence in many nonconvex problems [55]–[59].

Definition 14 (KL property [56]): A function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to have the KL property at point $x^* \in \text{dom}(f)$ if there exists $\eta \in \mathbb{R}^+$, a neighborhood U of x^* , and a function φ such that, for all $x \in U \cap \{x | f(x^*) < f(x) < f(x^*) + \eta\}$, the following KL inequality holds:

$$\varphi'(f(x) - f(x^*)) \text{dist}(0, \partial f(x)) \geq 1$$

where $\varphi : [0, \eta] \rightarrow \mathbb{R}^+$ satisfies: $\varphi(0) = 0$; φ is concave and C^1 -type on $[0, \eta]$; and $\varphi'(s) > 0$ for all $s \in (0, \eta)$.

A function that satisfies the KL property at every point in its effective domain is called a KL function. Specially, all semialgebraic functions satisfy the KL property, and its involved concave function can be chosen to be the form of $\varphi(s) = \gamma s^\theta$ for some $\gamma > 0$ and $\theta \in (0, 1]$.

Theorem 4 (Global Convergence): Assume that the objective function $\mathcal{F}(\mathcal{X})$ is a KL function, and then, the whole sequence $\{\mathcal{X}^{(k)}\}$ satisfies $\sum_{k=0}^{+\infty} \|\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}\|_F < +\infty$ and converges to a critical point of $\mathcal{F}(\mathcal{X})$.

Theorem 5 (Rate of Convergence): Assume that the objective function $\mathcal{F}(\mathcal{X})$ is semialgebraic, it satisfies the KL property with desingularising function $\varphi(s) = \gamma s^\theta$, and the convergence rate of the sequence $\{\mathcal{X}^{(k)}\}$ can be estimated as follows.

- 1) If $\theta \in (0, (1/2))$, then there exists $M_1 > 0$ such that $\|\mathcal{X}^{(k)} - \mathcal{X}^*\|_F \leq M_1(1/k)^{(\theta/(1-2\theta))}$.

- 2) If $\theta \in [(1/2), 1)$, then there exists $M_2 > 0$ and $\zeta \in [0, 1)$ such that $\|\mathcal{X}^{(k)} - \mathcal{X}^*\|_F \leq M_2\zeta^k$.
- 3) If $\theta = 1$, then $\{\mathcal{X}^{(k)}\}$ converges in finite iterations.

where \mathcal{X}^* is a critical point of $\mathcal{F}(\mathcal{X})$.

Theorem 4 shows that the sequence is globally convergent under the KL assumption, and moreover, it converges in some certain rates (sublinear, linear, or finite steps) with a further semialgebraic assumption by Theorem 5. Our analysis is mainly inspired by the pioneering work [55], [56]. The detailed connections and differences are discussed in Appendix I. By the way, utilizing the idea in [56] for proving algorithm convergence also exists in many other works, e.g., [57]–[59].

In the end, we discuss the reasonability of these assumptions in above theory analysis. First, it is easy to verify the assumption in Theorem 3 since $\Psi(\mathcal{X})$ is nonnegative and the loss function enjoys the property $\ell(\mathcal{X}) \rightarrow +\infty(\mathcal{X} \rightarrow +\infty)$ normally. As for Theorems 4 and 5, we show that the assumptions are satisfied in common cases.

Proposition 4: The objective function $\mathcal{F}(\mathcal{X})$ is semialgebraic with these common used penalty functions in Table II, as well as a KL function.

D. Differences With the IRNN

The proposed IR-t-TNN achieves the extension of the iterative reweighted algorithm framework from sparse vector and low-rank matrix recovery to LTRTR. It adopts the IRNN [36], [37] to some extent. Here, we discuss the main differences between each other.

- 1) The IR-t-TNN considers a more complicated LTRTR problem, for which we use a new formula of the t-TNN and propose its nonconvex surrogate with extensive analysis under the t-SVD framework.
- 2) To exploit the IR-t-TNN, the WTSVT operator is derived strictly, while the IRNN uses the existing result. Besides, we use adaptive parameter $\mu^{(k)}$ in each iteration via backtracking strategy to approximate the loss function better, which uses the Barzilai–Borwein rule for initialization, while the IRNN uses fixed parameter.
- 3) In [37], the authors only present some weak and local convergence results, and the function values are decreasing with a bounded sequence, while this article gives more abundant theoretical results, including the global convergence and estimation of convergence rate.
- 4) The main computation cost lies on the SVD operations both in the IR-t-TNN and IRNN, but our WTSVT operator needs $\lceil((n_3 + 1)/2)\rceil$ times per-iteration, which is nearly half of that of performing the IRNN slice by slice for tensors. Thus, our IR-t-TNN has a much lower time cost despite the process of backtracking, which is shown in the experiments later.

V. THEORY GUARANTEES

In this section, based on the NSP analysis, we establish the theoretical guarantees of the generalized nonconvex tubal rank minimization models in cases of exact and robust observation. With the usage of the proposed unified nonconvex surrogate of tensor tubal rank, the generalized nonconvex tubal rank

minimization model with linear measurement operator Φ can be formulated as

$$\min_{\mathcal{X}} \Psi(\mathcal{X}), \quad \text{s.t. } \Phi(\mathcal{X}) = \mathcal{T}. \quad (17)$$

As the baseline, the classical convex t-TNN minimization is

$$\min_{\mathcal{X}} \|\mathcal{X}\|_*, \quad \text{s.t. } \Phi(\mathcal{X}) = \mathcal{T}. \quad (18)$$

The null space analysis is known as one of the most important technical means to establish the theoretical guarantees of sparse vector or low-rank matrix recovery. Some representative works include [40], [60], [61]. In the following, we propose a tensor NSP condition as the exact recovery guarantee for the generalized nonconvex tubal rank minimization model (17). At the same time, we also give the NSP condition of the t-TNN minimization model (18) as a corollary of the main result. Moreover, the superiority of the nonconvex tubal rank minimization method is verified.

For tensor $\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, define

$$\mathcal{H}_\Gamma := \sum_{i \in \Gamma} \mathcal{U}(:, i, :) * \mathcal{S}(i, i, :) * \mathcal{V}(:, i, :)^\top$$

where $\Gamma := [r_t] \subset [m]$. Denote the null space of Φ as $\mathcal{N}(\Phi)$.

Theorem 6 (Exact Recovery): For any tensor \mathcal{X} of tubal rank at most r_t from observation $\Phi(\mathcal{X}) = \mathcal{T}$, it can be uniquely recovered by solving the generalized nonconvex tubal rank minimization model (17) if and only if

$$\Psi(\mathcal{H}_\Gamma) \leq \Psi(\mathcal{H}_{\Gamma^c}) \quad (19)$$

holds for any tensor $\mathcal{H} \in \mathcal{N}(\Phi) \setminus \mathcal{O}$.

A meaningful thing is that the process of Theorem 6 also holds when the penalty function $\psi(x) = x$. Thus, we obtain the corresponding tensor NSP condition for the t-TNN model (18).

Corollary 1: For any tensor \mathcal{X} of tubal rank at most r_t from observation $\Phi(\mathcal{X}) = \mathcal{T}$, it can be uniquely recovered by solving the t-TNN minimization model (18) if and only if

$$\|\mathcal{H}_\Gamma\|_* \leq \|\mathcal{H}_{\Gamma^c}\|_* \quad (20)$$

holds for any tensor $\mathcal{H} \in \mathcal{N}(\Phi) \setminus \mathcal{O}$.

Theorem 7: Compared to the standard tensor NSP condition (20), the generalized nonconvex tensor NSP condition (19) is weaker or less demanding by the following relationship:

$$\{\mathcal{H} \mid \|\mathcal{H}_\Gamma\|_* \leq \|\mathcal{H}_{\Gamma^c}\|_*\} \subset \{\mathcal{H} \mid \Psi(\mathcal{H}_\Gamma) \leq \Psi(\mathcal{H}_{\Gamma^c})\}.$$

Remark 2: Theorem 7 indicates that, in terms of the NSP condition, the potential recovery capability of the generalized nonconvex tubal rank minimization model (17) is stronger than the standard convex t-TNN minimization model (18).

Next, by extending the generalized nonconvex NSP condition, we establish the robust recovery guarantee when the observed tensor \mathcal{T} is contaminated by noise. At this time, model (17) becomes

$$\min_{\mathcal{X}} \Psi(\mathcal{X}), \quad \text{s.t. } \Phi(\mathcal{X}) + \mathcal{E} = \mathcal{T} \quad (21)$$

where \mathcal{E} is a noise tensor satisfying $\|\mathcal{E}\|_F \leq \varepsilon$ for some $\varepsilon > 0$.

Definition 15: For any tensor $\mathcal{H} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, if there exists constants $0 < a < 1$ and $b > 0$ such that

$$\Psi(\mathcal{H}_\Gamma) \leq a\Psi(\mathcal{H}_{\Gamma^c}) + b\|\Phi(\mathcal{H})\|_F \quad (22)$$

then the measurement operator Φ is said to satisfy the generalized nonconvex robust tensor NSP condition of order r_t .

Theorem 8 (Robust Recovery): Assume that operator Φ satisfies Definition 15. Then, for any tensor \mathcal{X} from observation $\Phi(\mathcal{X}) + \mathcal{E} = \mathcal{T}$, \mathcal{X}^\dagger is the solution of the model (21), and we have

$$\Psi(\mathcal{X} - \mathcal{X}^\dagger) \leq c_1\Psi(\mathcal{H}_{\Gamma^c}) + c_2\varepsilon \quad (23)$$

where constant $c_1 = ((2+2a)/1-a)$ and $c_2 = (4b/(1-a))$.

Corollary 2: With an additional low rank requirement $\text{rank}_t(\mathcal{X}) \leq r_t$ in Theorem 8, we have

$$\Psi(\mathcal{X} - \mathcal{X}^\dagger) \leq c_2\varepsilon.$$

In other words, the model (21) can robustly recover every tensor \mathcal{X} of tubal rank at most r_t , satisfying $\Phi(\mathcal{X}) + \mathcal{E} = \mathcal{T}$ if the generalized nonconvex robust tensor NSP condition of order r_t is satisfied, since $\Psi(\mathcal{X} - \mathcal{X}^\dagger) \rightarrow 0$ only if $\mathcal{X} \rightarrow \mathcal{X}^\dagger$.

Remark 3: The robust recovery guarantee of the t-TNN minimization model with noise observation can be derived from the proposed generalized nonconvex robust tensor NSP condition, which is also less demanding.

The above theoretical results provide the guarantees for the recoverability of the generalized nonconvex tubal rank minimization. To the best of our knowledge, such NSP-based theoretical analysis for tensors under the t-SVD framework has not been reported previously. The only existing works [62], [63] are based on the Tucker and CP decomposition. We emphasize that our results are not trivial because the algebra of tensors is fraught with hardness, and more challengingly, the problem is generalized and nonconvex. Finally, one thing that needs to be pointed out is that the NSP conditions are normally weaker than the RIP [64]; thus, our recent works [65], [66] developing the tensor RIP theories could provide a certain guarantee for the reasonability and feasibility of the established tensor NSP conditions.

VI. EXPERIMENTS

In this section, we conduct a series of experiments on both synthetic and real-world data to demonstrate the superiority of the proposed IR-t-TNN algorithm. The algorithm is tested for the tensor completion problem, that is

$$\min_{\mathcal{X}} \Psi(\mathcal{X}) + \frac{1}{2}\|\mathcal{P}_\Omega(\mathcal{X}) - \mathcal{P}_\Omega(\mathcal{T})\|_F^2 \quad (24)$$

where Ω denotes the support set of observations and $\mathcal{P}_\Omega : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a linear operator that keeps the entries in Ω unchanged and sets those outside Ω zeros. It is easy to know that the gradient of the squared loss function is a Lipschitz continuous function with a Lipschitz constant $L(\ell) = 1$. In experiments, we select the nonconvex penalty function ψ in $\Psi(\cdot)$ from these widely used ones in Table II. For its choice of the parameters, we search them from a candidate set and use the one which obtains good performance in most cases ($p = 0.5$ for L_p).

All the experiments are conducted on the platform of Windows 10 and MATLAB (R2016a) with an Intel Core i5-4200H 2.80-GHz CPU and 12-GB memory. More experimental results can be found in the Supplementary Materials II.

A. Synthetic Experiments

This section verifies the superiority of IR-t-TNN over the t-TNN [22] and IRNN [37] through numerical simulations. We synthesize a ground-truth $\mathcal{T}_0 \in \mathbb{R}^{n \times n \times n}$ with tubal rank r_t by performing t-product $\mathcal{T}_0 = \mathcal{T}_1 * \mathcal{T}_2$, where $\mathcal{T}_1 \in \mathbb{R}^{n \times r_t \times n}$ and $\mathcal{T}_2 \in \mathbb{R}^{r_t \times n \times n}$ are independently sampled from the standard Gaussian distribution. The recovered tensor is denoted as \mathcal{T}_* , and the performance is evaluated by the Relative Error (RelError), defined as

$$\text{RelError} := \frac{\|\mathcal{T}_* - \mathcal{T}_0\|_F}{\|\mathcal{T}_0\|_F}. \quad (25)$$

The following contains two tasks.

1) *Stronger Capacity in Recovery Than the t-TNN:* We first compare the model recovery ability with the classical convex t-TNN minimization method, that is

$$\min_{\mathcal{X}} \lambda \|\mathcal{X}\|_* + \frac{1}{2} \|\mathcal{P}_{\Omega}(\mathcal{X}) - \mathcal{P}_{\Omega}(\mathcal{T})\|_F^2. \quad (26)$$

We check the recoverability as a function of the tubal rank r_t and the sampling rate (SR) p . Set $n_1 = n_2 = 30$, $n_3 = 20$. For the task of recovering the generated tensor, we investigate the probability of recovery success varying the value of r_t and p . We set the parameter $\lambda = 0.01$ for model (26) in experiment. For each pair (r_t, p) , we simulate ten test instances and declare a trial to be successful when RelError is less than 10^{-3} . Fig. 4 reports the success percentage for each pair, in which the brown region reflects the full successful recovery, and the blue region reflects the full fail recovery. The bigger the brown area, the stronger ability to recovery. We can find that the size of the brown area obtained by IR-t-TNN with each different nonconvex surrogate is much bigger than that obtained by t-TNN obviously. It demonstrates that the nonconvex tubal rank minimization model (24) is more robust and efficient than the classical convex t-TNN minimization model (26).

2) *Faster Speed in Convergence Than the IRNN:* We then investigate the convergence behaviors. As mentioned before, our IR-t-TNN adopts the Barzilai–Borwein initialization in each inner loop, which is expected to accelerate the convergence. In addition, the IRNN cannot recover tensors directly and only process them slice by slice. This greatly increases the algorithmic complexity. Next, we show that our IR-t-TNN converges faster and also takes less time than the IRNN.

We generate a tensor of size $100 \times 100 \times 100$ with tubal rank $r_t = 20$, as the above mentioned, and sample 50% entries of the underlying tensor at random. With different nonconvex penalty functions, by using the IRNN and IR-t-TNN, respectively, we report the RelError values of each iteration. The algorithms stop when the RelError is less than 10^{-5} . At the same time, the running times up to every iteration are recorded. The results are presented in Fig. 5. It can be seen that IR-t-TNN reaches a smaller error at the beginning of iterations and stops earlier. This verifies the efficiency of

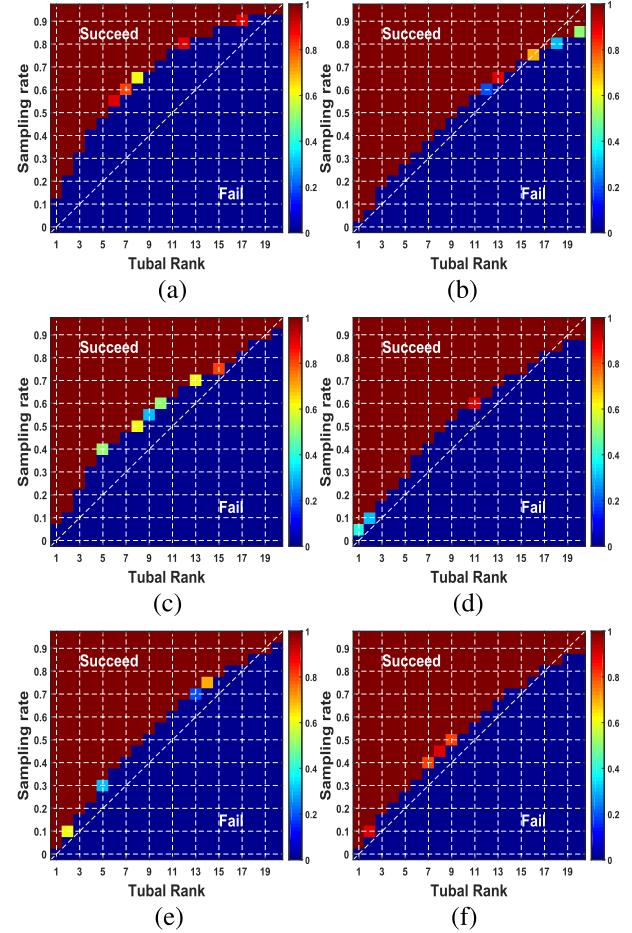


Fig. 4. Comparison of the recovery capacity varying tubal rank and sampling rate of the t-TNN and IR-t-TNN.

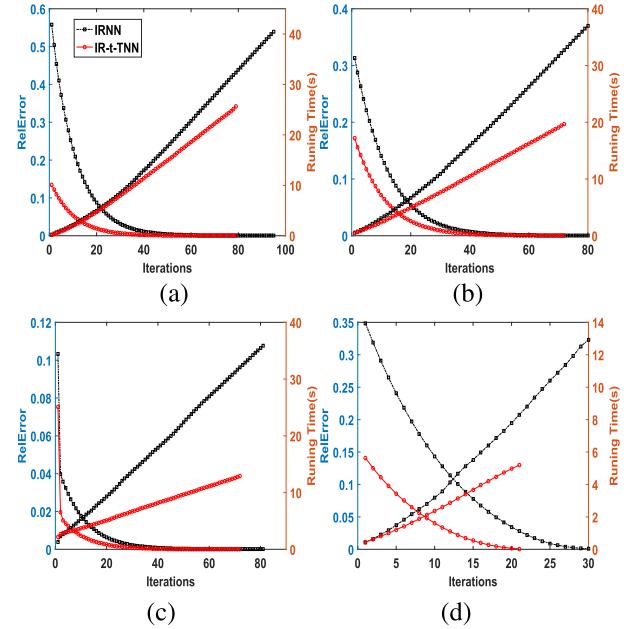


Fig. 5. Comparison of the relative error (fall) and running time (rise) versus iteration of the IRNN and IR-t-TNN.

the Barzilai–Borwein initialization technique. In terms of the running time, the IR-t-TNN's whole time nearly shortens by half compared to that of IRNN, and the time growth trend is

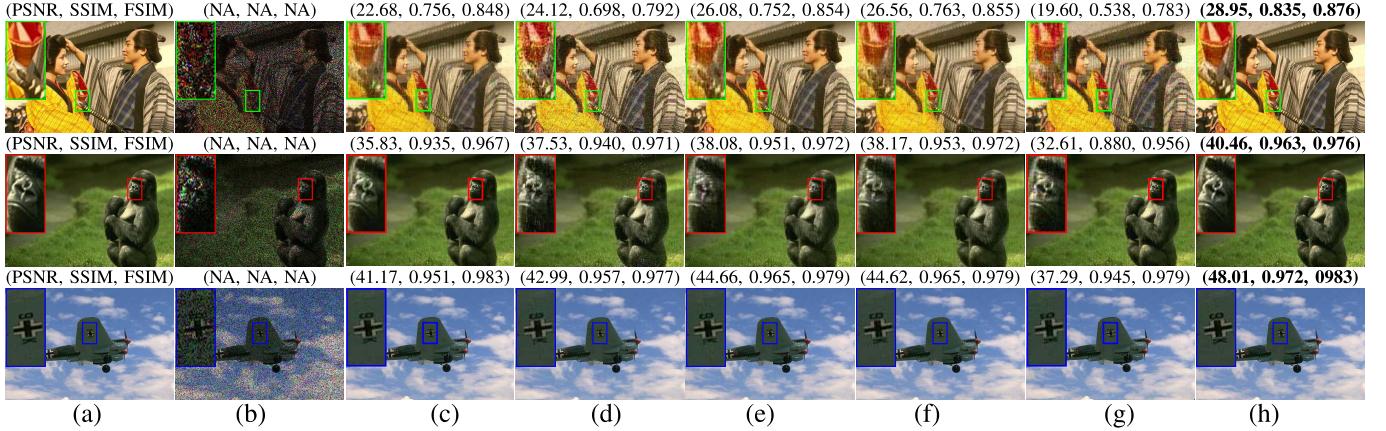


Fig. 6. Visual comparison of image restoration. From top to bottom, the SRs are 30%, 50%, and 70%, respectively. (a) Original, (b) Observed, (c) SNN, (d) t-TNN, (e) W-t-TNN, (f) PS-t-TNN, (g) IRNN, and (h) IR-t-TNN.

slower, which verifies that the time cost of IR-t-TNN is nearly half that of IRNN, as discussed in the algorithm analysis.

B. Real-World Applications

In this section, we apply the IR-t-TNN to recovery several real-world tensor data with the further comparison with the classical Tucker-based SNN [10] minimization and two state-of-the-art nonconvex methods for tubal rank minimization, including W-t-TNN [31] and PS-t-TNN [32]. Here, we only take the L_p penalty function for the test in the IR-t-TNN and IRNN since the results in other cases are very similar (see more experimental results in the Supplementary Materials II).

1) *Natural Image Restoration*: We apply the IR-t-TNN for natural image restoration from partial pixels since a color image with R, G, and B channels can be modeled as a third-order low-tubal-rank tensor, as illustrated in Fig. 1. We randomly select 100 images sized $481 \times 321 \times 3$ from the Berkeley Segmentation Database² (BSD) [67]. For each image, pixels are sampled randomly with SR 30%, 50%, and 70%. The image recovery performances are evaluated by the peak signal-to-noise (PSNR), structural similarity index (SSIM), and Feature SIMilarity (FSIM). Larger values of PSNR, SSIM, and FSIM indicate better image restoration performance.

The average performances over selected 100 images under different SRs are summarized in Table IV. From this table, we can find the PSNR, SSIM, and FSIM values of IRNN are the smallest mostly, which shows that the tensor-based methods are better than the matrix-based method. It is not difficult to explain this since IRNN itself cannot process tensors globally but only be applied for each channel independently. Another finding is that the performance of the Tucker-based method SNN is always inferior to other t-SVD-based methods except for the case that the SR is 0.3, which indicates that the tensor tubal rank minimization under the t-SVD framework would be a more effective modeling strategy. Among these t-SVD-based methods, nonconvex methods are more effective than the convex t-TNN method. More importantly, our IR-t-TNN almost outperforms all the other competing methods

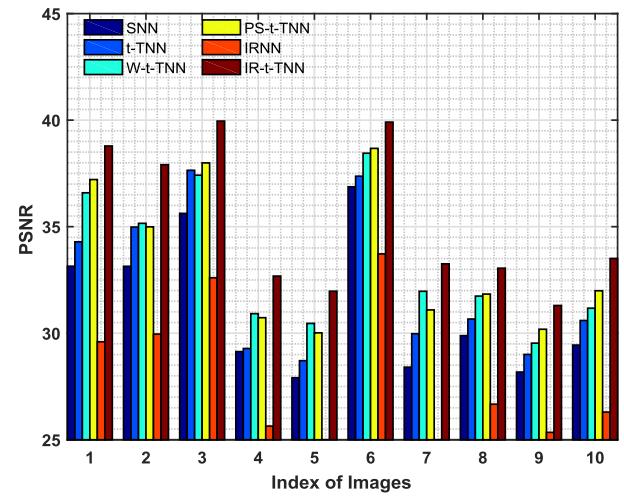


Fig. 7. PSNR values of various methods on BSD image data when the SR is 50%.

in terms of all performance evaluation indexes, whose PSNR, SSIM, and FSIM values are significantly larger than that of these convex and nonconvex ways currently in effect. Fig. 6 gives three visual examples with different SRs. It shows that our recovered images contain a clearer outline and fuller color and brightness compared with other competitors, exhibited in the enlarged views of the corresponding areas in color boxes. In addition, we show the PSNR values in Fig. 7 obtained by various methods on ten images when the SR is 50%, which also verifies that the IR-t-TNN outperforms other methods quantitatively for most of the images.

2) *Face Inpainting*: We then test the extended Yale Face Database.³ It contains 165 gray-scale images of 15 individuals, where there are 11 images of size 100×100 per subject with different facial expression or configuration. Each subject can be mapped to a third-order tensor of size $100 \times 100 \times 11$ with low-rank structure property [68]. Therefore, we apply IR-t-TNN to complete the face data in this experiment.

Fig. 8 gives the average PSNR values of the 15 subjects versus different SRs, and Fig. 9 shows the specific values

²<https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>

³<http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html>

TABLE IV
NATURAL IMAGES RESTORATION PERFORMANCES COMPARISON WHEN SR IS 30%, 50%, AND 70%

Method	SR = 30%			SR = 50%			SR = 70%		
	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
SNN	26.77	0.838	0.883	31.21	0.894	0.941	36.31	0.928	0.966
t-TNN	24.69	0.686	0.814	32.31	0.918	0.946	36.47	0.943	0.959
W-t-TNN	27.30	0.830	0.887	33.35	0.930	0.952	38.74	0.951	0.970
PS-t-TNN	27.45	0.843	0.891	33.47	0.922	0.949	40.58	0.955	0.972
IRNN	24.01	0.688	0.835	27.88	0.831	0.910	32.14	0.904	0.952
IR-t-TNN	28.09	0.841	0.899	35.21	0.935	0.958	43.08	0.957	0.974

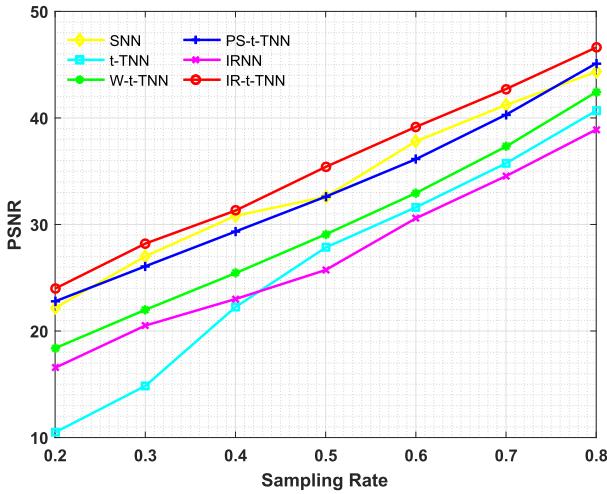


Fig. 8. Average PSNR values of various methods on the Yale Face Database versus different SRs.

per subjects when SR is 50%. Several recovered examples and corresponding PSNR values by different methods for the face data with 20%, 40%, 60%, and 80% observed measurements are presented in Fig. 10. From these results, we have the following observations. First, the nonconvex tensor methods, including W-t-TNN, PS-t-TNN, and IR-t-TNN, are better than the convex method t-TNN. Second, our method is better than other methods in most cases, especially when the sampling ratio is small. One interesting result is that SNN performs better than several t-SVD-based methods. This probably because these types of tensors are constructed by matrices superimposed on each other; thus, SNN would be more suitable to capture its low-rank structure.

3) *Video Recovery*: The gray video sequence is a third-order tensor in nature, and each frame is a gray image. It also can be approximated by a low-tuba-rank tensor. In this experiment, we consider the video recovery problem from partially observed entries via the IR-t-TNN algorithm.

We use ten videos from the YUR Video Database.⁴ Due to the computational limitation, we take the first 100 frames of each sequence. For the selected sequences, we use the file in the provided QCIF format, in which each frame has the size 144×176 . For the tensor of size $144 \times 176 \times 100$ constructed from a video, we randomly set 50% entries to be

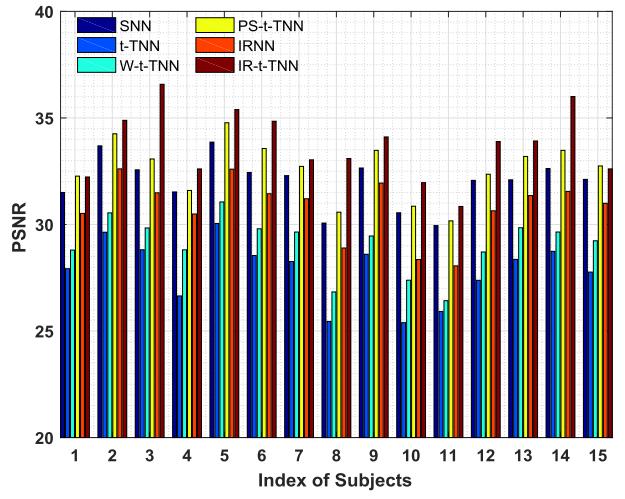


Fig. 9. PSNR values of various methods on the Yale Face Database when the SR is 50%.

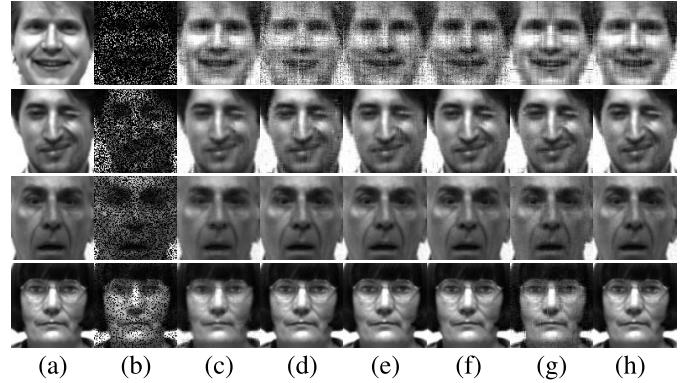


Fig. 10. Visual comparison of face inpainting. From top to bottom, the SRs are 20%, 40%, 60%, and 80%, respectively. (a) Original faces. (b) Observed faces. (c)–(h) Recovered faces by SNN, t-TNN, W-t-TNN, PS-t-TNN, IRNN, and IR-t-TNN, respectively. Best viewed in $\times 2$ sized pdf file.

observed. Then, we apply the IR-t-TNN algorithm to recover the original video. We compare our results with SNN, t-TNN, W-t-TNN, PS-t-TNN, and IRNN method, as mentioned earlier. The performance is evaluated by PSNR values. Table V shows the PSNR values of the compared methods on all ten video sequences. We can find that the IR-t-TNN has absolute superiority, whose PSNR values are much higher obviously. See some intuitive results in Fig. 11, which shows one frame from the recovered videos. Whether it is from a

⁴<http://trace.eas.asu.edu/yuv/>

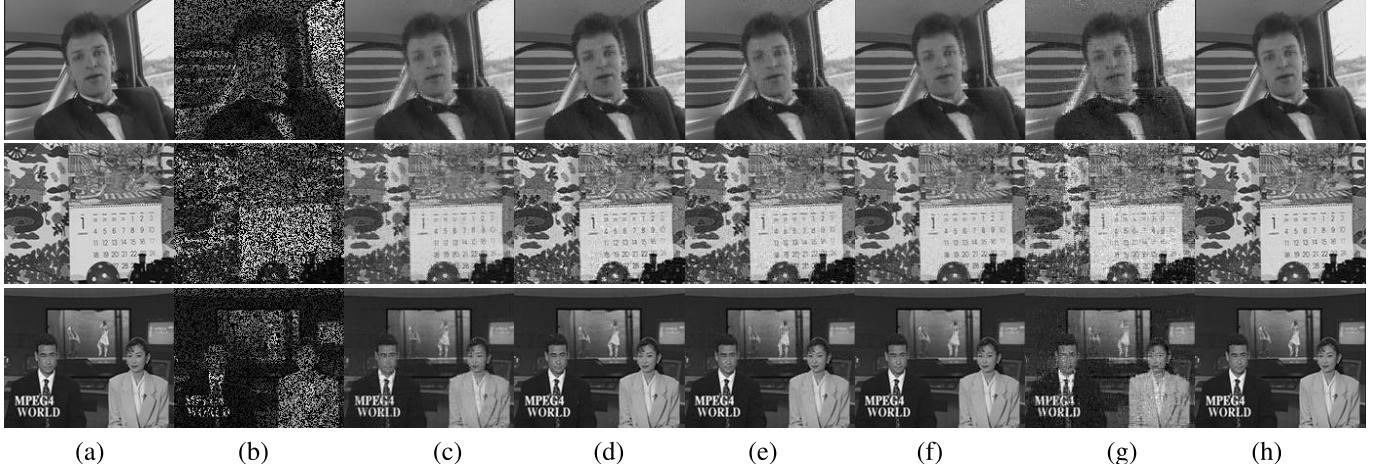


Fig. 11. Visual comparison of video recovery. Examples frames are from the sequences “Carphone,” “Mobile,” and “News”. (a) Original, (b) Observed, (c) SNN, (d) t-TNN, (e) W-t-TNN, (f) PS-t-TNN, (g) IRNN, and (h) IR-t-TNN.

TABLE V
PSNR VALUES OF THE VARIOUS METHODS ON YUR VIDEO DATABASE
FOR VIDEO RECOVERY

ID	Videos	SNN	t-TNN	W-t-TNN	PS-t-TNN	IRNN	IR-t-TNN
1	Akiyo	33.2	35.4	42.1	42.7	27.6	45.2
2	Bridge	30.9	30.3	34.5	35.3	26.7	36.1
3	Bus	28.7	25.1	26.9	27.3	27.2	26.8
4	Carphone	30.6	28.5	29.7	30.9	27.0	31.4
5	Claire	36.4	35.9	37.5	37.4	32.9	39.5
6	Container	30.4	31.0	35.8	36.6	25.1	38.8
7	Foreman	30.8	30.7	32.1	32.7	26.5	34.0
8	Hall	31.9	31.6	38.9	39.8	26.7	41.2
9	Mobile	23.8	21.5	21.9	23.5	20.0	24.1
10	News	30.7	32.8	35.1	35.6	25.7	38.0

visual or quantitative perspective, our IR-t-TNN outperforms other competing algorithms in video recovery.

4) *Network Traffic Estimation*: Network traffic data consist of traffic matrices, which records the amount of network data exchanged between the source and destination pairs, such as computers and routers, which is frequently used in computer network traffic analysis. However, it is usually corrupted by missing values due to hardware or software malfunctions and the expense of the data collection process. In this experiment, we apply our method to estimate the network traffic data from incomplete measurements.

The data used here come from the GÉANT network [69], which records the traffic exchanged between 23 routers per 15 min from January 1 to April 30, 2005, and can be downloaded from the Toolbox for Traffic Engineering Methods (TOTEM) project.⁵ Since the traffic matrices evolve over time, the network traffic data can be formed as a third-order tensor of size $23 \times 23 \times 11520$ with source routers, destination routers, and time modes, which has a low-rank structure because of the periodicity [70]. Therefore, the network traffic data estimation can be modeled as the LTRTR problem.

⁵<https://totem.info.ucl.ac.be/dataset.html>

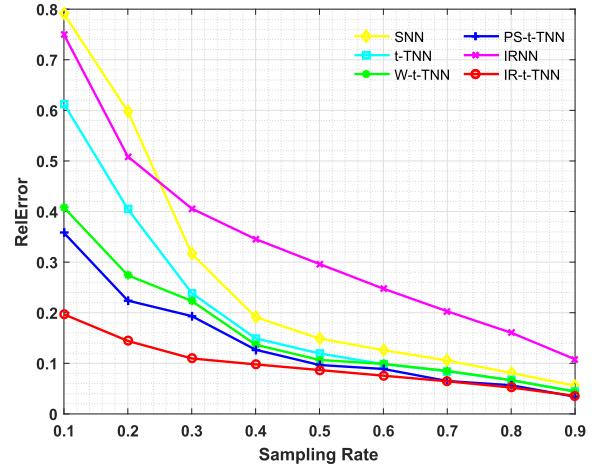


Fig. 12. Comparison of the estimation results obtained by various methods in terms of RelError on network traffic data with different SRs.

We recovery the formed tensor under different SRs. Fig. 12 gives the results in terms of the relative error. As excepted, IR-t-TNN always outshines other methods, especially in the case of low SR, say 0.1 and 0.2.

VII. CONCLUSION

This work solved the LTRTR problem via a generalized nonconvex approach, and extensive experiments demonstrate its superiority in many applications. It well established the nonconvex tensor recovery framework, including the model, algorithm, and theories, thereby greatly promoting the development of tensor tubal rank minimization under the t-SVD. For future works, one may develop the nonconvex strategy for certain tensor applications incorporating the prior information to achieve better accuracy and efficiency. Moreover, exploring more NSP-based theories for the LTRTR problem is also a feasible direction, for example, constructing the corresponding tensor NSP conditions for other heuristic models, and deducing the sampling complexity with probability tools making the NSP characteristics satisfied under the certain assumption of the measurement operator.

REFERENCES

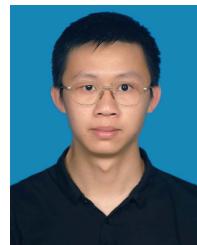
- [1] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Apr. 2020.
- [2] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3842–3849.
- [3] Y. Wu, H. Tan, Y. Li, J. Zhang, and X. Chen, "A fused CP factorization method for incomplete tensors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 751–764, Mar. 2019.
- [4] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S.-I. Amari, "Bayesian robust tensor factorization for incomplete multiway data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 736–748, Apr. 2016.
- [5] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–11, Aug. 2012.
- [6] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, vol. 14, no. 3, pp. 105–122, 2000.
- [7] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," *J. ACM*, vol. 60, no. 6, p. 45, Nov. 2009.
- [8] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1966.
- [9] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2114–2121.
- [10] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.
- [11] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, "Simultaneous rectification and alignment via robust recovery of low-rank tensors," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2013, pp. 1637–1645.
- [12] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, "Robust low-rank tensor recovery with rectification and alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 238–255, Jan. 2021.
- [13] B. Romeraparedes and M. Pontil, "A new convex relaxation for tensor completion," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2013, pp. 2967–2975.
- [14] L.-H. Lim and P. Comon, "Multiarray signal processing: Tensor decomposition meets compressed sensing," *Comp. Rendus Mécanique*, vol. 338, no. 6, pp. 311–320, Jun. 2010.
- [15] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1260–1280, Feb. 2014.
- [16] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5249–5257.
- [17] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, Jan. 2011.
- [18] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," 2016, *arXiv:1606.05535*. [Online]. Available: <http://arxiv.org/abs/1606.05535>
- [19] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, "The twist tensor nuclear norm for video completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2961–2973, Dec. 2017.
- [20] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, Aug. 2011.
- [21] 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 J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, Jan. 2013.
- [22] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2017.
- [23] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, Nov. 2018.
- [24] Z. Zhang, Y. Xie, W. Zhang, and Q. Tian, "Effective image retrieval via multilinear multi-index fusion," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2878–2890, Nov. 2019.
- [25] H. Wang, F. Zhang, J. Wang, and Y. Wang, "Estimating structural missing values via low-tubal-rank tensor completion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 3297–3301.
- [26] F. Zhang, J. Wang, W. Wang, and C. Xu, "Low-tubal-rank plus sparse tensor recovery with prior subspace information," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Apr. 13, 2020, doi: [10.1109/TPAMI.2020.2986773](https://doi.org/10.1109/TPAMI.2020.2986773).
- [27] W.-H. Xu *et al.*, "Laplace function based nonconvex surrogate for low-rank tensor completion," *Signal Process., Image Commun.*, vol. 73, pp. 62–69, Apr. 2019.
- [28] S. Cai, Q. Luo, M. Yang, W. Li, and M. Xiao, "Tensor robust principal component analysis via non-convex low rank approximation," *Appl. Sci.*, vol. 9, no. 7, p. 1411, Apr. 2019.
- [29] T. Li and J. Ma, "Non-convex penalty for tensor completion and robust PCA," 2019, *arXiv:1904.10165*. [Online]. Available: <http://arxiv.org/abs/1904.10165>
- [30] H. Kong, X. Xie, and Z. Lin, "T-schatten- p norm for low-rank tensor recovery," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1405–1419, Dec. 2018.
- [31] Y. Mu, P. Wang, L. Lu, X. Zhang, and L. Qi, "Weighted tensor nuclear norm minimization for tensor completion using tensor-SVD," *Pattern Recognit. Lett.*, vol. 130, pp. 4–11, Feb. 2020.
- [32] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, "Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm," *J. Comput. Appl. Math.*, vol. 372, Jul. 2020, Art. no. 112680.
- [33] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4686–4698, Dec. 2009.
- [34] P. Gong, C. Zhang, Z. Lu, J. Z. Huang, and J. Ye, "A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, vol. 28, no. 2, pp. 37–45.
- [35] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," in *Proc. 29th AAAI Conf. Artif. Intell. (AAAI)*, 2015, pp. 1805–1811.
- [36] C. Lu, J. Tang, S. Yan, and Z. Lin, "Generalized nonconvex nonsmooth low-rank minimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 4130–4137.
- [37] C. Lu, J. Tang, S. Yan, and Z. Lin, "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 829–839, Feb. 2016.
- [38] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, Jan. 2010.
- [39] C. Lu, J. Feng, Z. Lin, and S. Yan, "Exact low tubal rank tensor recovery from Gaussian measurements," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 2504–2510.
- [40] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k -term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, Jul. 2008.
- [41] Y. Zhang, "Theory of compressive sensing via ℓ_1 -minimization: A non-RIP analysis and extensions," *J. Oper. Res. Soc. China*, vol. 1, no. 1, pp. 79–105, Mar. 2013.
- [42] C. Gao, N. Wang, Q. Yu, and Z. Zhang, "A feasible nonconvex relaxation approach to feature selection," in *Proc. 25th AAAI Conf. Artif. Intell. (AAAI)*, 2011, pp. 356–361.
- [43] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
- [44] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0 -minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 106–121, Jan. 2009.
- [45] J. H. Friedman, "Fast sparse regression and classification," *Int. J. Forecasting*, vol. 28, no. 3, pp. 722–738, 2012.
- [46] L. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, May 1993.
- [47] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, Apr. 2010.
- [48] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *J. Mach. Learn. Res.*, vol. 11, pp. 1081–1107, Mar. 2010.
- [49] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.
- [50] K. Chen, H. Dong, and K.-S. Chan, "Reduced rank regression via adaptive nuclear norm penalization," *Biometrika*, vol. 100, no. 4, pp. 901–920, Dec. 2013.

- [51] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.
- [52] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.
- [53] K. Kurdyka, "On gradients of functions definable in o-minimal structures," *Annales de l'institut Fourier*, vol. 48, no. 3, pp. 769–781, 1988.
- [54] S. Łojasiewicz, "Une propriété topologique des sous-ensembles analytiques réels," *Les Équations aux Dérivées Partielles*, vol. 117, pp. 87–89, 1963.
- [55] H. Attouch and J. Bolte, "On the convergence of the proximal algorithm for nonsmooth functions involving analytic features," *Math. Program.*, vol. 116, nos. 1–2, pp. 5–16, Jan. 2009.
- [56] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, nos. 1–2, pp. 459–494, Aug. 2014.
- [57] T. Pock and S. Sabach, "Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems," *SIAM J. Imag. Sci.*, vol. 9, no. 4, pp. 1756–1787, Jan. 2016.
- [58] L. Stella, A. Themelis, and P. Patrinos, "Forward–backward quasi-newton methods for nonsmooth optimization problems," *Comput. Optim. Appl.*, vol. 67, no. 3, pp. 443–487, Jul. 2017.
- [59] Y. Xu and W. Yin, "A globally convergent algorithm for nonconvex optimization based on block coordinate update," *J. Sci. Comput.*, vol. 72, no. 2, pp. 700–734, Aug. 2017.
- [60] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, Mar. 2011.
- [61] B. Recht, W. Xu, and B. Hassibi, "Null space conditions and thresholds for rank minimization," *Math. Program.*, vol. 127, no. 1, pp. 175–202, Mar. 2011.
- [62] Z. Shi, J. Han, T. Zheng, and J. Li, "Guarantees of augmented trace norm models in tensor recovery," in *Proc. 23th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2013, pp. 1670–1676.
- [63] Q. Li, D. Schonfeld, and S. Friedland, "Generalized tensor compressive sensing," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2013, pp. 1–6.
- [64] J. Cahill, X. Chen, and R. Wang, "The gap between the null space property and the restricted isometry property," *Linear Algebra Appl.*, vol. 501, pp. 363–375, Jul. 2016.
- [65] 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.*, vol. 64, no. 1, Jan. 2021, Art. no. 119101.
- [66] F. Zhang, W. Wang, J. Huang, J. Wang, and Y. Wang, "RIP-based performance guarantee for low-tubal-rank tensor recovery," *J. Comput. Appl. Math.*, vol. 374, Aug. 2020, Art. no. 112767.
- [67] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, Jul. 2001, pp. 416–423.
- [68] X. Zhang, "A nonconvex relaxation approach to low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1659–1671, Jun. 2019.
- [69] S. Uhlig, B. Quoitin, J. Lepropre, and S. Balon, "Providing public intradomain traffic matrices to the research community," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 1, pp. 83–86, Jan. 2006.
- [70] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometric Intell. Lab. Syst.*, vol. 106, no. 1, pp. 41–56, Mar. 2011.



Hailin Wang (Graduate Student Member, IEEE) received the B.S. degree in statistics from Southwest University, Chongqing, China, in 2019, where he is currently pursuing the M.S. degree with the School of Mathematics and Statistics.

His research interests include neural networks, compressed sensing, and tensor sparsity.



Feng Zhang received the M.S. and Ph.D. degrees from the School of Mathematics and Statistics, Southwest University, Chongqing, China, in 2017 and 2020, respectively.

He is currently holding a postdoctoral position with the School of Mathematics and Statistics, Southwest University. His research focuses on compressed sensing and tensor sparsity.



Jianjun Wang (Member, IEEE) received the B.S. degree in mathematical education and the M.S. degree in fundamental mathematics from Ningxia University, Yinchuan, China, in 2000 and 2003, respectively, and the Ph.D. degree in applied mathematics from the Institute for Information and System Science, Xi'an Jiaotong University, Xi'an, China, in 2006.

He is currently a Professor with the School of Mathematics and Statistics, Southwest University, Chongqing, China. His research focuses on machine learning, data mining, neural networks, and sparse learning.



Tingwen Huang (Fellow, IEEE) received the B.S. degree in mathematics from Southwest Normal University (now, Southwest University), Chongqing, China, in 1990, the M.S. degree in mathematics from Sichuan University, Chengdu, China, in 1993, and the Ph.D. degree in mathematics from Texas A&M University, College Station, TX, USA, in 2002.

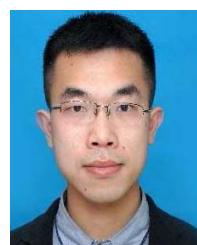
He worked as a Visiting Assistant Professor with Texas A&M University. In August 2003, he joined Texas A&M University at Qatar, Doha, Qatar, as an Assistant Professor, then he was promoted to a Professor in 2013. He is a Professor with the Texas A&M University at Qatar. He has published more than 300 peer-review reputable journal articles, including more than 100 articles in IEEE TRANSACTIONS. His research interests include neural networks-based computational intelligence, distributed control and optimization, nonlinear dynamics, and applications in smart grids.

Prof. Huang currently serves as an Action Editor for *Neural Networks*, and an Associate Editor for three journals, including the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, and the *Journal of Control and Decision*, and an Editorial Board Member of *Cognitive Computation*, *Journal of Nonlinear Systems and Applications* and *Numerical Algebra, Control and Optimization*.



Jianwen Huang received the B.S. degree in mathematics and applied mathematics from Tianshui Normal University, Tianshui, China, in 2010, and the M.S. degree in probability theory and mathematical statistics and the Ph.D. degree in statistics from Southwest University, Chongqing, China, in 2013 and 2019, respectively.

He is currently an Associate Professor with the School of Mathematics and Statistics, Tianshui Normal University. His research interests include compressed sensing, low-rank matrix recovery, and extreme value theory.



Xinling Liu received the B.S. degree in mathematics and applied mathematics and the M.S. degree in probability theory and mathematical statistics from Southwest University, Chongqing, China, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics.

His current research focuses on deep learning, approximation theory, and sparse learning.