



# Tensor Factorization with Total Variation and Tikhonov Regularization for Low-Rank Tensor Completion in Imaging Data

Xue-Lei Lin<sup>1</sup> · Michael K. Ng<sup>2</sup> · Xi-Le Zhao<sup>3</sup>

Received: 25 April 2019 / Accepted: 26 November 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

The main aim of this paper is to study tensor factorization for low-rank tensor completion in imaging data. Due to the underlying redundancy of real-world imaging data, the low-tubal-rank tensor factorization (the tensor–tensor product of two factor tensors) can be used to approximate such tensor very well. Motivated by the spatial/temporal smoothness of factor tensors in real-world imaging data, we propose to incorporate a hybrid regularization combining total variation and Tikhonov regularization into low-tubal-rank tensor factorization model for low-rank tensor completion problem. We also develop an efficient proximal alternating minimization (PAM) algorithm to tackle the corresponding minimization problem and establish a global convergence of the PAM algorithm. Numerical results on color images, color videos, and multispectral images are reported to illustrate the superiority of the proposed method over competing methods.

**Keywords** Tensor factorization · Hybrid regularization · Tensor completion · Proximal alternating minimization

**Mathematics Subject Classification** 90C26 · 90C30 · 90C90 · 65F22

## 1 Introduction

As a high-dimensional extension of the matrix, the tensor can express more complicated intrinsic structures of data, which plays an important role in many real-world applications; see, e.g., [22, 31, 37, 51]. Low-rank tensor completion (LRTC) is one of the most important problems in tensor

processing, which aims at filling in the missing entries of a partially observed low-rank tensor. Indeed, high-dimensional data, such as videos and images, usually have a low-rank structure or approximately so, based on which many works on low-rank tensor completion have been done; see, e.g., [32, 36, 42, 48].

The fundamental issue of LRTC is the definition of the rank of the tensor. However, unlike matrix rank, the definition of the rank of the tensor is not unique. Many research efforts have been devoted to designing the rank of tensor. The CANDECOMP/PARAFAC (CP) rank [10, 21, 50] is defined based on the CP decomposition, where an  $n$ th-order tensor is decomposed as the sum of rank-one tensors, i.e., the outer product of  $n$  vectors. The CP rank is defined as the minimal number of the rank-one tensors required to express the target tensor. However, the corresponding LRTC problems are NP-hard. The Tucker rank is based on the Tucker decomposition which decomposes a tensor into a core tensor multiplied by a matrix along each mode [17, 26, 27, 39]. The Tucker rank is defined as a vector, the  $k$ th element of which is the rank of the mode- $k$  unfolding matrix. The corresponding LRTC problems are also NP-hard. Liu et al. suggested the sum of the nuclear norm (SNN) of the mode- $k$  unfolding matrix as the convex surrogate [23, 30, 31]. However, unfolding the tensor

The research is supported by research grants HKRGC GRF (12306616, 12200317, 12300519, 12300218), HKU Grant (104005583), and NSFC (61876203, 61772003, 11801479).

✉ Xi-Le Zhao  
xlzhao122003@163.com

Xue-Lei Lin  
hxuellin@gmail.com

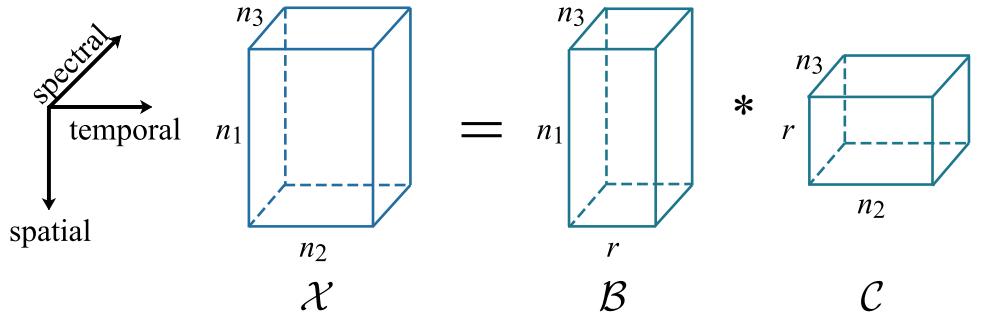
Michael K. Ng  
mng@maths.hku.hk

<sup>1</sup> Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

<sup>2</sup> Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong

<sup>3</sup> School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, People's Republic of China

**Fig. 1** Low-rank tensor factorization (tensor–tensor product of two factor tensors)



along each mode will inevitably destroy the intrinsic structures of the tensor [25,40].

Recently, the novel tensor–tensor product (*t*-product) of two tensors  $\mathcal{B} \in \mathbb{R}^{n_1 \times r \times n_3}$  and  $\mathcal{C} \in \mathbb{R}^{r \times n_2 \times n_3}$  is defined to be a tensor  $\mathcal{X} := \mathcal{B} * \mathcal{C} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  [25,46], where the mode-3 fibers of  $\mathcal{X}$  are defined as  $\mathcal{X}(i, j, :) = \sum_{k=1}^r \mathcal{B}(i, k, :) \star \mathcal{C}(k, j, :)$ . Here, “ $\star$ ” denotes the circular convolution of two vectors. Based on *t*-product, the tensor singular value decomposition (*t*-SVD), which generalizes the matrix singular value decomposition (SVD), is proposed in [25]. For any third-order tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , its *t*-SVD is defined as

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V},$$

where  $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$  and  $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$  are third-order orthogonal tensors and  $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is *f*-diagonal tensor whose frontal slices are diagonal matrices. Based on *t*-SVD, the corresponding tensor tubal rank has received an increasing attention [32,35,45,46,49,51].

The tubal rank of  $\mathcal{X}$  is defined as the number of nonzero tubes of  $\mathcal{S}$  [32,51], i.e.,

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

The nonzero tubes of  $\mathcal{S}$  in *t*-SVD are analogous to nonzero diagonal elements of the diagonal matrix in matrix SVD. As the natural generalization of the matrix SVD, the *t*-SVD also has the optimal tubal rank  $r$  approximation property [44]. Due to the combinational nature of the tensor tubal rank minimization problems, Semerci et al. [35] suggested the tensor nuclear norm (TNN) as the convex surrogate. Then, Zhang et al. [46] proposed the TNN-based LRTC model, and subsequently, Lu et al. [32] proposed the TNN-based tensor robust principal component analysis model. In order to avoid time-consuming *t*-SVD, Zhou et al. [51] proposed the low-tubal-rank tensor factorization model for the LRTC problem, which factorizes the target tensor into the *t*-product of two factor tensors. Interestingly, many real-world images exhibit spatial/temporal smoothness. We observed that the smoothness of the real-world imaging data is inherited by factor tensors. For example in Fig. 1, a color video can

be represented as a third-order tensor, whose factor tensors inherit the spatial and temporal smoothness along the first mode and the second mode, respectively. Motivated by the clear physical interpretations of factor tensors, we suggest the total variation- and Tikhonov-regularized tensor factorization model for the LRTC problem. Due to its ability to preserve edges, total variation (TV) regularization, firstly proposed in [34], is introduced to promote the smoothness of factor tensors. Moreover, Tikhonov regularization is introduced to guarantee the stability of the solution [19]. The TV regularization and Tikhonov regularization are complementary to each other, rather than isolated and uncorrelated. A hybrid regularization combining TV and Tikhonov regularization contributes to an excellent completion performance.

The rest of this paper is organized as follows: In Sect. 2, a new low-rank tensor completion model is introduced and the proximal alternating minimization (PAM) algorithm is developed for solving the model. In Sect. 3, a global convergence is established for the PAM algorithm. In Sect. 4, some efficient solvers and the fast implementations are introduced for the subproblems arising from the PAM algorithm. In Sect. 5, numerical results are reported to show the performance of the proposed method. Finally, some concluding remarks are given in Sect. 6.

## 2 The Proposed Model and the PAM Algorithm

Before presenting the model, we firstly introduce some notations which will be frequently used later. For a third-order tensor  $\mathcal{A}$ , we use the MATLAB notations, such as  $\mathcal{A}(:)$ ,  $\mathcal{A}(i, :, :)$ ,  $\mathcal{A}(:, i, :)$ ,  $\mathcal{A}(:, :, i)$ , and also define  $\|\mathcal{A}\|_{\ell_2} := \|\mathcal{A}(:)\|_2$ ,  $\|\mathcal{A}\|_{\ell_1} := \|\mathcal{A}(:)\|_1$ .

### 2.1 The Tensor Completion Model by Tensor Factorization with TV-Tikhonov Regularizations

The proposed tensor completion model by tensor factorization with TV-Tikhonov (TCTF-TVT) regularization is described as follows:

$$\begin{aligned} & \arg \min_{\substack{0 \leq \mathcal{X} \leq 1, \mathcal{B}, \mathcal{C}}} \frac{1}{2} \|\mathcal{X} - \mathcal{B} * \mathcal{C}\|_{\ell_2}^2 \\ & + \alpha \sum_{k=1}^{m_1} \|\mathcal{D}_{1k} * \mathcal{B}\|_{\ell_1} + \alpha \sum_{k=1}^{m_2} \|\mathcal{C} * \mathcal{D}_{2k}\|_{\ell_1} \\ & + \frac{\mu}{2} \|\mathcal{B}\|_{\ell_2}^2 + \frac{\mu}{2} \|\mathcal{C}\|_{\ell_2}^2, \\ \text{s.t. } & P_{\Omega}(\mathcal{X}) = P_{\Omega}(\mathcal{M}), \end{aligned} \quad (2.1)$$

where  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is the tensor to be filled;  $\mathcal{B} \in \mathbb{R}^{n_1 \times r \times n_3}$ ,  $\mathcal{C} \in \mathbb{R}^{r \times n_2 \times n_3}$ ;  $\Omega$  is an index set;  $\mathcal{M}$  is a given tensor;  $P_{\Omega}(\cdot)$  is a projection operator preserving entries with indices in  $\Omega$  while restricting entries with indices out of  $\Omega$  to be zeros;  $\mathcal{M}$  is a given tensor such that  $0 \leq \mathcal{M} \leq 1$ ;  $\mathcal{D}_{1k} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$  ( $k = 1, 2, \dots, m_1$ ) and  $\mathcal{D}_{2k} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$  ( $k = 1, 2, \dots, m_2$ ) are some tensor representations of backward partial difference operators with homogeneous Neumann boundary conditions, whose specific form will be given in Sect. 4;  $r$  is a given small positive integer such that  $r \leq \min\{n_1, n_2\}$ ;  $\alpha$  and  $\mu$  are given positive numbers.

Instead of imposing TV and Tikhonov regularizers on the original tensor  $\mathcal{X}$ , we consider TV and Tikhonov regularizers of the factor tensors  $\mathcal{B}$  and  $\mathcal{C}$  in (2.1) for two reasons. First, the smoothness of the original tensor can be inherited by factor tensors. Second, because the sizes of factor tensors are smaller than the size of the original tensor, imposing TV and Tikhonov regularizers on the factor tensors instead of the original tensor saves the computational cost and storage. Moreover, in this paper, we mainly consider completion of the color image, color video, and multispectral image, pixel values of which always belong to the interval  $[0, 1]$  up to a scaling constant. This is why we set a constraint  $0 \leq \mathcal{X} \leq 1$  in model (2.1).

For ease of statement, we rewrite (2.1) as the following equivalent unconstrained form

$$\arg \min_{\mathcal{X}, \mathcal{B}, \mathcal{C}} F(\mathcal{X}, \mathcal{B}, \mathcal{C}), \quad (2.2)$$

where

$$\begin{aligned} F(\mathcal{X}, \mathcal{B}, \mathcal{C}) &= \frac{1}{2} \|\mathcal{X} - \mathcal{B} * \mathcal{C}\|_{\ell_2}^2 + \delta_S(\mathcal{X}) + \alpha \sum_{k=1}^{m_1} \|\mathcal{D}_{1k} * \mathcal{B}\|_{\ell_1} \\ &+ \alpha \sum_{k=1}^{m_2} \|\mathcal{C} * \mathcal{D}_{2k}\|_{\ell_1} + \frac{\mu}{2} \|\mathcal{B}\|_{\ell_2}^2 + \frac{\mu}{2} \|\mathcal{C}\|_{\ell_2}^2 \end{aligned}$$

$$S = \{\mathcal{W} \in \mathbb{R}^{n_1 \times n_2 \times n_3} | \mathcal{W}(i, j, k) = \mathcal{M}(i, j, k) \text{ for } (i, j, k) \in \Omega, \mathcal{W}(i, j, k) \in [0, 1] \text{ for } (i, j, k) \notin \Omega\},$$

$$\delta_S(\mathcal{X}) := \begin{cases} 0, & \mathcal{X} \in S, \\ +\infty, & \mathcal{X} \notin S, \end{cases}.$$

It is clear that  $F$  is coercive with respect to  $\mathcal{B}$  and  $\mathcal{C}$ , i.e.,

$$\liminf_{\mathcal{B} \rightarrow \infty} F(\mathcal{X}, \mathcal{B}, \mathcal{C}) = +\infty, \quad \liminf_{\mathcal{C} \rightarrow \infty} F(\mathcal{X}, \mathcal{B}, \mathcal{C}) = +\infty.$$

By such coercivity and continuity of  $F$  on  $S \times \mathbb{R}^{n_1 \times r \times n_3} \times \mathbb{R}^{r \times n_2 \times n_3}$ , existence of (2.2) can be easily proven.

## 2.2 The PAM Algorithm for Solving (2.2)

Note that (2.2) is a multivariate optimization problem. Alternating minimization (AM) algorithm is commonly used to solve multivariate optimization problems due to its simplicity and efficiency; see, e.g., [3, 5, 9, 33, 38]. To enhance the theoretical convergence and numerical stability of AM algorithm, proximal terms are suggested to add in subproblems arising from AM algorithm, which is called PAM algorithm; see, e.g., [1, 2, 4]. In this subsection, a PAM algorithm is developed for solving (2.2).

Given an initial guess  $(\mathcal{X}^k, \mathcal{B}^k, \mathcal{C}^k)$  for problem (2.2), then the PAM iteration is defined as follows:

$$\mathcal{X}^{k+1} = \arg \min_{\mathcal{X}} F(\mathcal{X}, \mathcal{B}^k, \mathcal{C}^k) + \frac{\rho}{2} \|\mathcal{X} - \mathcal{X}^k\|_{\ell_2}^2, \quad (2.3)$$

$$\mathcal{B}^{k+1} = \arg \min_{\mathcal{B}} F(\mathcal{X}^{k+1}, \mathcal{B}, \mathcal{C}^k) + \frac{\rho}{2} \|\mathcal{B} - \mathcal{B}^k\|_{\ell_2}^2, \quad (2.4)$$

$$\mathcal{C}^{k+1} = \arg \min_{\mathcal{C}} F(\mathcal{X}^{k+1}, \mathcal{B}^{k+1}, \mathcal{C}) + \frac{\rho}{2} \|\mathcal{C} - \mathcal{C}^k\|_{\ell_2}^2, \quad (2.5)$$

where  $\rho > 0$  is a given primal parameter.

It is easy to see that (2.3)–(2.5) are all strongly convex optimization problems, whose existence and uniqueness are guaranteed. In Sect. 4, efficient algorithms will be introduced to solve (2.3)–(2.5).

## 3 Global Convergence Analysis of PAM Iteration

In this section, we will prove the global convergence of PAM algorithm.

To formalize the discussion, we express the tensor-form variables  $\mathcal{X}$ ,  $\mathcal{B}$  and  $\mathcal{C}$  as vectors in what follows.

For a positive integer  $n$ , denote  $\Xi(n) := \{1, 2, \dots, n\}$ . For positive integers,  $l_1, l_2, l_3$ , define bijections  $\mathbb{V}_{[l_1, l_2, l_3]} : \mathcal{A} \in \mathbb{R}^{l_1 \times l_2 \times l_3} \mapsto \mathbb{V}_{[l_1, l_2, l_3]}(\mathcal{A}) \in \mathbb{R}^{l_1 l_2 l_3 \times 1}$  and  $\mathbb{I}_{[l_1, l_2, l_3]} : (i, j, k) \in \Xi(l_1) \times \Xi(l_2) \times \Xi(l_3) \mapsto \mathbb{I}_{[l_1, l_2, l_3]}(i, j, k) \in \Xi(l_1 l_2 l_3)$  by

$$\mathbb{V}_{[l_1, l_2, l_3]}(\mathcal{A}) := \mathcal{A}(:),$$

$$\tilde{\mathcal{A}}(i, j, k) = [\tilde{\mathcal{A}}(:)](\mathbb{I}_{[l_1, l_2, l_3]}(i, j, k)) \text{ holding for each } \tilde{\mathcal{A}} \in \mathbb{R}^{l_1 \times l_2 \times l_3}.$$

From the above definition, it is easy to see that  $\mathbb{I}_{[l_1, l_2, l_3]}$  maps an index  $(i, j, k)$  of an entry in  $\mathcal{A} \in \mathbb{R}^{l_1 \times l_2 \times l_3}$  into the index of the same entry in  $\mathcal{A}(:)$ . Note that  $\mathbb{V}_{[l_1, l_2, l_3]}$  is a bi-continuous linear bijection.

In the following, we denote  $\mathbb{V}_{[n_1, n_2, n_3]}, \mathbb{V}_{[n_1, r, n_3]}, \mathbb{V}_{[r, n_2, n_3]}, \mathbb{I}_{[n_1, n_2, n_3]}, \mathbb{I}_{[n_1, r, n_3]}, \mathbb{I}_{[r, n_2, n_3]}$  by  $\mathbb{V}_1, \mathbb{V}_2, \mathbb{V}_3, \mathbb{I}_1, \mathbb{I}_2, \mathbb{I}_3$ , respectively. Denote  $M_1 = n_1 n_2 n_3, M_2 = n_1 r n_3, M_3 = r n_2 n_3, M = M_1 + M_2 + M_3$ . Then, (2.2) can be equivalently rewritten as the following vector form

$$\arg \min_{v \in \mathbb{R}^M} \tilde{F}(v), \quad (3.1)$$

where  $v = (v_1; v_2; v_3), v_i \in \mathbb{R}^{M_i}$  ( $i = 1, 2, 3$ ),

$$\begin{aligned} \tilde{F}(v) &:= G(v) + \delta_{\tilde{S}}(v_1) + g_2(v_2) + g_3(v_3), \\ G(v) &= \frac{1}{2} \|\mathbb{V}_1^{-1}(v_1) - \mathbb{V}_2^{-1}(v_2) * \mathbb{V}_3^{-1}(v_3)\|_{\ell_2}^2 \\ &\quad + \frac{\mu}{2} \left( \|v_2\|_2^2 + \|v_3\|_2^2 \right), \end{aligned}$$

$$\text{dom}(f) := \{x \in \mathbb{R}^n | f(x) < +\infty\},$$

$$\hat{\partial} f(x) := \begin{cases} \left\{ v \in \mathbb{R}^n \mid \liminf_{y \rightarrow x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|_2} \geq 0 \right\}, & x \in \text{dom}(f), \\ \emptyset, & x \notin \text{dom}(f), \end{cases},$$

$$\partial f(x) := \begin{cases} \left\{ v \in \mathbb{R}^n \mid \exists \{x^k\}_{k=1}^{+\infty} \text{ s.t. } x^k \rightarrow x, f(x^k) \rightarrow f(x), v^k \in \hat{\partial} f(x^k) \rightarrow v \right\}, & x \in \text{dom}(f), \\ \emptyset, & x \notin \text{dom}(f), \end{cases},$$

$$\text{dom}(\partial f) := \{x \in \mathbb{R}^n | \partial f(x) \neq \emptyset\}.$$

$$g_2(v_2) := \alpha \sum_{k=1}^{m_1} \|D_{1k}(v_2)\|_{\ell_1},$$

$$g_3(v_3) := \alpha \sum_{k=1}^{m_2} \|D_{2k}(v_3)\|_{\ell_1},$$

$$D_{1k}(v_2) := \mathcal{D}_{1k} * [\mathbb{V}_2^{-1}(v_2)],$$

$$D_{2k}(v_3) := \mathbb{V}_3^{-1}(v_3) * \mathcal{D}_{2k},$$

$$\begin{aligned} \tilde{S} &:= \{w \in \mathbb{R}^{M_1} | w(i) = [\mathbb{V}_1(\mathcal{M})](i) \text{ for } i \in \mathbb{I}_1(\Omega), \\ &\quad w(i) \in [0, 1] \text{ for } i \notin \mathbb{I}_1(\Omega)\}, \end{aligned}$$

$$\delta_{\tilde{S}}(v_1) := \begin{cases} 0, & v_1 \in \tilde{S}, \\ +\infty, & v_1 \notin \tilde{S}, \end{cases}.$$

It is clear that  $v_1, v_2$  and  $v_3$  in (3.1) are counterparts of  $\mathcal{X}, \mathcal{B}$  and  $\mathcal{C}$  in (2.2), respectively. Due to the bilinearity of “ $*$ ” operation, it is clear that  $D_{1k} : \mathbb{R}^{n_1 r n_3 \times 1} \rightarrow \mathbb{R}^{n_1 \times r \times n_3}$  ( $k = 1, 2, \dots, m_1$ ) and  $D_{2k} : \mathbb{R}^{r n_2 n_3 \times 1} \rightarrow \mathbb{R}^{r \times n_2 \times n_3}$  ( $k = 1, 2, \dots, m_2$ ) are both linear operators. Moreover, it is easy to see that  $\tilde{S}$  is a non-empty closed set, which means  $\delta_{\tilde{S}}(\cdot)$  is a proper lower semi-continuous (PLSC) function on  $\mathbb{R}^{M_1}$  and  $\tilde{F}(\cdot)$  is a PLSC function on  $\mathbb{R}^M$ .

(2.3)–(2.5) is then equivalent to

$$v_1^{k+1} \in \arg \min_{v_1} \tilde{F}(v_1; v_2^k; v_3^k) + \frac{\rho}{2} \|v_1 - v_1^k\|_2^2, \quad (3.2)$$

$$v_2^{k+1} \in \arg \min_{v_2} \tilde{F}(v_1^k; v_2; v_3^k) + \frac{\rho}{2} \|v_2 - v_2^k\|_2^2, \quad (3.3)$$

$$v_3^{k+1} \in \arg \min_{v_3} \tilde{F}(v_1^k; v_2^k; v_3) + \frac{\rho}{2} \|v_3 - v_3^k\|_2^2, \quad (3.4)$$

provided that  $(v_1^k; v_2^k; v_3^k)$  is an initial guess for (3.1).

The aim of this section is to establish a global convergence for the iteration process (3.2)–(3.4) or equivalently for the PAM iteration (2.3)–(2.5). We firstly introduce some notations and preliminaries.

For a real-extended-valued function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , define

$\partial f(x)$  is called sub-differential of  $f$  at  $x$ . Here,  $\langle \cdot, \cdot \rangle$  denotes the Euclidean inner product in  $\mathbb{R}^n$ .

**Proposition 1** [1] A necessary condition for  $x$  to be a minimizer of a PLSC function  $f$  is that

$$0 \in \partial f(x). \quad (3.5)$$

A point satisfying (3.5) is called limiting critical or simply critical.

For a PLSC function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , and  $-\infty < \eta_1 < \eta_2 \leq +\infty$ , we denote

$$[\eta_1 < f < \eta_2] := \{x \in \mathbb{R}^n | \eta_1 < f(x) < \eta_2\}.$$

**Definition 1** (Kurdyka–Łojasiewicz property, [1]) A PLSC function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is said to have the Kurdyka–Łojasiewicz property at  $\bar{x} \in \text{dom}(\partial f)$  if there exists  $\eta \in (0, +\infty]$ , a neighborhood  $U$  of  $\bar{x}$  and a continuous concave function  $\phi : [0, \eta) \mapsto [0, +\infty)$  such that:

- $\phi(0) = 0$ ,
- $\phi$  is  $C^1$  on  $(0, \eta)$ ,
- $\phi'$  is positive on  $(0, \eta)$ ,
- for each  $x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$ , the Kurdyka–Łojasiewicz inequality hold :  

$$\phi'(f(x) - f(\bar{x}))\text{dist}(0, \partial f(x)) \geq 1.$$

In the above definition, the norm involved in  $\text{dist}(\cdot, \cdot)$  is  $\|\cdot\|_2$  and the convention  $\text{dist}(0, \emptyset) := +\infty$  is used.

A PLSC function  $f$  satisfying Kurdyka–Łojasiewicz property at each point of  $\text{dom}(\partial f)$  is called a KL function.

We will make use of the convergence theory built in [2] stated in the following to analyze the convergence of PAM iteration.

**Lemma 1** ([2]) Let  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a PLSC function. Let  $\{x^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$  be a sequence such that

- H1.** (Sufficient decrease condition). For each  $k \in \mathbb{N}$ ,  $f(x^{k+1}) + a\|x^{k+1} - x^k\|_2^2 \leq f(x^k)$  holds for a constant  $a \in (0, +\infty)$ ;
- H2.** (Relative error condition). For each  $k \in \mathbb{N}$ ,  $\exists w^{k+1} \in \partial f(x^{k+1})$  such that  $\|w^{k+1}\|_2 \leq b\|x^{k+1} - x^k\|_2$  holds for a constant  $b \in (0, +\infty)$ ;
- H3.** (Continuity condition). There exists a subsequence  $\{x^{k_j}\}_{j \in \mathbb{N}}$  and  $\bar{x} \in \mathbb{R}^n$  such that

$$x^{k_j} \rightarrow \bar{x} \text{ and } f(x^{k_j}) \rightarrow f(\bar{x}), \quad \text{as } j \rightarrow \infty.$$

If  $f$  has the Kurdyka–Łojasiewicz property at  $\bar{x}$  appearing in H3, then

- (i)  $x^k \rightarrow \bar{x}$ ;
- (ii)  $\bar{x}$  is a critical point of  $f$ , i.e.,  $0 \in \partial f(\bar{x})$ ;
- (iii) the sequence  $\{x^k\}_{k \in \mathbb{N}}$  has a finite length, i.e.,

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|_2 < +\infty.$$

The rest of this section is devoted to proving that the objective function  $\tilde{F}$  in (3.1) and the iterative sequence  $(v_1^k; v_2^k; v_3^k)_{k \in \mathbb{N}}$  generated by PAM iteration (3.2)–(3.4) satisfy the assumptions imposed in Lemma 1, by which we establish the convergence of PAM iteration.

**Definition 2** [6]

- (a) A subset  $A \subset \mathbb{R}^n$  is called a semi-algebraic set if and only if there exists a finite number of real polynomial functions  $P_{ij}, Q_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}$  such that

$$A = \bigcup_{j=1}^p \bigcap_{i=1}^q \{x \in \mathbb{R}^n \mid P_{ij}(x) = 0, \quad Q_{ij}(x) < 0\}.$$

- (b) Let  $A_1 \subset \mathbb{R}^n$  and  $A_2 \subset \mathbb{R}^m$  be two subsets. A function  $f : A_1 \rightarrow A_2$  is called semi-algebraic if and only if its graph  $\{(x, f(x)) \in \mathbb{R}^{n+m} \mid x \in A_1\}$  is a semi-algebraic set in  $\mathbb{R}^{n+m}$ .

We call a function  $f$  is semi-algebraic on  $A$  if and only if  $\{(x, f(x)) \mid x \in A\}$  is a semi-algebraic set.

**Lemma 2** ([7,8,28]) A semi-algebraic real-valued function  $f$  is a KL function, i.e.,  $f$  satisfies Kurdyka–Łojasiewicz property at each  $x \in \text{dom}(f)$ .

**Proposition 2** ([2,6,14,41])

- (i) A subset of  $\mathbb{R}$  is semi-algebraic if and only if it is finite union of points and open intervals (bounded or unbounded).
- (ii) Complements, finite unions, finite intersections, finite Cartesian products of semi-algebraic sets are semi-algebraic.
- (iii) Finite sums and finite products of semi-algebraic functions are semi-algebraic.
- (iv) The composition  $g \circ f$  of semi-algebraic mappings  $f : A_1 \rightarrow A_2$  and  $g : A_2 \rightarrow A_3$  is semi-algebraic.
- (v) If  $f : A_1 \rightarrow A_2$  is semi-algebraic and  $A_3 \subset A_1$  is a semi-algebraic subset, then  $f(A_3)$  is a semi-algebraic set.
- (vi) Polynomials defined on  $\mathbb{R}^n$  are semi-algebraic functions.
- (vii) The absolute value function  $f(x) := |x|$  is semi-algebraic on  $\mathbb{R}$ .

**Proposition 3** If  $f : A_1 \rightarrow A_2$  is semi-algebraic and  $A_3 \subset A_1$  is a semi-algebraic subset, then  $f$  is semi-algebraic on  $A_3$ .

**Proof** By Proposition 2 (ii), (v),  $A_3 \times f(A_3)$  is semi-algebraic. Notice that

$$\{(x, f(x)) \mid x \in A_3\} = \{(x, f(x)) \mid x \in A_1\} \cap (A_3 \times f(A_3)),$$

which together with Proposition 2 (ii) proves the proposition.  $\square$

Denote

$$\hat{S} = \tilde{S} \times \mathbb{R}^{M_2} \times \mathbb{R}^{M_3}.$$

**Lemma 3**  $\tilde{F}$  is semi-algebraic on  $\hat{S}$ . Hence,  $\tilde{F}$  satisfies the Kurdyka–Łojasiewicz property at each  $v \in \hat{S}$ .

**Proof** On  $\hat{S}$ ,  $\tilde{F}$  can be expressed as

$$\tilde{F}(v) := G(v) + g_2(v_2) + g_3(v_3), \quad v \in \hat{S}.$$

By Proposition 2 (i), (ii), it is clear that  $\hat{S}$  is a semi-algebraic set.

Notice that  $\mathbb{V}_i^{-1}$  ( $i = 1, 2, 3$ ) are linear mappings between finite-dimensional spaces. Hence, each element of  $\mathbb{V}_i^{-1}(v_i)$  ( $i = 1, 2, 3$ ) is actually a linear polynomial of  $v := (v_1; v_2; v_3)$ . Additionally, by the definition of  $\|\cdot\|_{\ell_2}$ , we know that  $G(v)$  is a polynomial of  $v$ . Thus, by Proposition 2 (vi) and Proposition 3,  $G$  is semi-algebraic on  $\hat{S}$ .

Again,  $D_{1k}$  ( $k = 1, 2, \dots, m_1$ ) and  $D_{2k}$  ( $k = 1, 2, \dots, m_2$ ) are linear mappings between finite-dimensional spaces. Hence, each element of  $D_{1k}(v_2)$ 's and  $D_{2k}(v_3)$  can be regarded as a linear polynomial of  $v := (v_1; v_2; v_3)$  for each  $k$ . Therefore,  $g_2(v_2)$  and  $g_3(v_3)$  are simply finite sums of compositions of absolute value function and linear polynomial of  $v$ , which is a semi-algebraic function on  $\mathbb{R}^M$  and thus is semi-algebraic on  $\hat{S}$  by Propositions 2, 3.

Then, Proposition 2 (iii) implies that  $\tilde{F}$  is semi-algebraic on  $\hat{S}$ . The proof is complete.  $\square$

For two subsets  $A_1, A_2$  and an element  $x$  of the same vector space, denote

$$\begin{aligned} A_1 + A_2 &:= \{x_1 + x_2 | x_1 \in A_1, x_2 \in A_2\}, \\ x + A_1 &:= \{x + y | y \in A_1\}. \end{aligned}$$

**Theorem 4** Let  $v^k = (v_1^k; v_2^k; v_3^k)$  be the sequence generated by the PAM iteration (3.2)–(3.4) with an arbitrary initial guess  $v^0 \in \tilde{S}$ . Then, there exists  $\bar{v} \in \hat{S}$  such that

- (i)  $v^k \rightarrow \bar{v}$ ;
- (ii)  $0 \in \partial \tilde{F}(\bar{v})$ ;
- (iii)  $\{v^k\}_{k \in \mathbb{N}}$  has a finite length, i.e.,

$$\sum_{k=0}^{+\infty} \|v^{k+1} - v^k\|_2 < +\infty.$$

**Proof** It has been mentioned above that  $\tilde{F}$  is a PLSC function on  $\mathbb{R}^M$ . From (3.2)–(3.4), we see that

$$\begin{aligned} \tilde{F}(v_1^{k+1}; v_2^k; v_3^k) + \frac{\rho}{2} \|v_1^{k+1} - v_1^k\|_2^2 \\ \leq \tilde{F}(v_1^k; v_2^k; v_3^k), \quad k \in \mathbb{N}, \\ \tilde{F}(v_1^{k+1}; v_2^{k+1}; v_3^k) + \frac{\rho}{2} \|v_2^{k+1} - v_2^k\|_2^2 \\ \leq \tilde{F}(v_1^{k+1}; v_2^k; v_3^k), \quad k \in \mathbb{N}, \end{aligned}$$

$$\begin{aligned} \tilde{F}(v_1^{k+1}; v_2^{k+1}; v_3^{k+1}) + \frac{\rho}{2} \|v_3^{k+1} - v_3^k\|_2^2 \\ \leq \tilde{F}(v_1^k; v_2^k; v_3^k), \quad k \in \mathbb{N}. \end{aligned}$$

Summing over the three inequalities above, we obtain

$$\tilde{F}(v^{k+1}) + \frac{\rho}{2} \|v^{k+1} - v^k\|_2^2 \leq \tilde{F}(v^k), \quad k \in \mathbb{N}. \quad (3.6)$$

Hence, **H1** (sufficient decrease condition) is satisfied with  $a = \frac{\rho}{2}$ .

Recall that  $\tilde{F}$  contains the Tikhonov regularization terms  $\frac{\mu}{2} (\|v_2\|_2^2 + \|v_3\|_2^2)$ , which together with (3.6) implies that

$$\sup_{k \in \mathbb{N}} \frac{\mu}{2} (\|v_2^k\|_2^2 + \|v_3^k\|_2^2) \leq \tilde{F}(v^0) < +\infty.$$

Moreover, from the definition of the indicator function  $\delta_{\tilde{S}}(\cdot)$  and (3.2), we clearly see that  $\{v_1^k | k \in \mathbb{N}\} \subset \tilde{S}$ . Since  $\tilde{S}$  is a bounded set,  $\sup_{k \in \mathbb{N}} \|v_1^k\|_2 < +\infty$ . Thus, we conclude that the sequence  $\{v^k\}_{k \in \mathbb{N}}$  is bounded, i.e.,

$$\sup_{k \in \mathbb{N}} \|v^k\|_2 < +\infty.$$

Let  $\partial_{v_i}$  and  $\nabla_{v_i}$  denote sub-differential and gradient with respect to variable  $v_i$ , respectively, for  $i = 1, 2, 3$ . It has been mentioned in the proof of Lemma 3 that  $G$  is a polynomial function and thus is infinitely differentiable. Then, (3.2) and Proposition 1 imply that

$$\begin{aligned} 0 &\in \partial_{v_1} \left[ \tilde{F}(\cdot; v_2^k; v_3^k) + \frac{\rho}{2} \|\cdot - v_1^k\|_2^2 \right] \Big|_{v_1=v_1^{k+1}} \\ &= \partial_{v_1} \tilde{F}(v_1^{k+1}; v_2^k; v_3^k) + \rho(v_1^{k+1} - v_1^k) \\ &= \nabla_{v_1} G(v_1^{k+1}; v_2^k; v_3^k) + \partial \delta_{\tilde{S}}(v_1^{k+1}) + \rho(v_1^{k+1} - v_1^k), \quad k \in \mathbb{N} \end{aligned} \quad (3.7)$$

Similarly, (3.3)–(3.4) and Proposition 1 imply that

$$0 \in \nabla_{v_2} G(v_1^{k+1}; v_2^{k+1}; v_3^k) + \partial g_2(v_2^{k+1}) + \rho(v_2^{k+1} - v_2^k), \quad k \in \mathbb{N}, \quad (3.8)$$

$$0 \in \nabla_{v_3} G(v_1^{k+1}; v_2^{k+1}; v_3^{k+1}) + \partial g_3(v_3^{k+1}) + \rho(v_3^{k+1} - v_3^k), \quad k \in \mathbb{N}. \quad (3.9)$$

(3.7)–(3.9) imply that there exists  $w_1^{k+1} \in \partial \delta_{\tilde{S}}(v_1^{k+1})$ ,  $w_i^{k+1} \in \partial g_i(v_i^{k+1})$  ( $i = 2, 3$ ) such that

$$-\tilde{w}^{k+1} = u^k + \rho(v^{k+1} - v^k), \quad k \in \mathbb{N},$$

where

$$\tilde{w}^{k+1} = (w_1^{k+1}; w_2^{k+1}; w_3^{k+1}),$$

$$u^k = (\nabla_{v_1} G(v_1^{k+1}; v_2^k; v_3^k); \nabla_{v_2} G(v_1^{k+1}; v_2^{k+1}; v_3^k);$$

$$\nabla_{v_3} G(v_1^{k+1}; v_2^{k+1}; v_3^{k+1})).$$

Denote  $w^{k+1} = \tilde{w}^{k+1} + \nabla G(v^{k+1})$ . It is clear that  $w^{k+1} \in \partial \tilde{F}(v^{k+1})$  and

$$-w^{k+1} = u^k - \nabla G(v^{k+1}) + \rho(v^{k+1} - v^k), \quad k \in \mathbb{N}, \quad (3.10)$$

Denote  $E = \{v^k | k \in \mathbb{N}\}$ . For  $v = (v_1; v_2; v_3) \in \mathbb{R}^M$  with  $v_i \in \mathbb{R}^{M_i}$  ( $i = 1, 2, 3$ ), define the coordinate projections by

$$\Pi_i(v) := v_i, \quad i = 1, 2, 3.$$

Denote  $\hat{E} = \Pi_1(E) \times \Pi_2(E) \times \Pi_3(E)$ . Since  $E$  is bounded, it is trivial to see that  $\hat{E} \subset \mathbb{R}^M$  is also bounded. Since  $G$  is a polynomial, it is easy to prove that  $\nabla G$  is Lipschitz-continuous on any bounded subset of  $\mathbb{R}^M$ . Hence, there exists a constant  $c > 0$  such that

$$\|\nabla G(v) - \nabla G(w)\|_2 \leq c\|v - w\|_2, \quad v, w \in \hat{E}.$$

Therefore,

$$\begin{aligned} \|u^k - \nabla G(v^{k+1})\|_2 &= [\|\nabla_{v_1} G(v_1^{k+1}; v_2^k; v_3^k) - \nabla_{v_1} G(v^{k+1})\|_2^2 \\ &\quad + \|\nabla_{v_2} G(v_1^{k+1}; v_2^{k+1}; v_3^k) \\ &\quad - \nabla_{v_2} G(v^{k+1})\|_2^2]^{\frac{1}{2}} \\ &\leq [c^2(\|v_2^k - v_2^{k+1}\|_2^2 + \|v_3^k - v_3^{k+1}\|_2^2) \\ &\quad + c^2\|v_3^k - v_3^{k+1}\|_2^2]^{\frac{1}{2}} \\ &\leq \sqrt{2}c\|v^{k+1} - v^k\|_2, \quad k \in \mathbb{N}, \end{aligned}$$

which together with (3.10) implies that

$$\begin{aligned} \|w^{k+1}\|_2 &\leq \|u^k - \nabla G(v^{k+1})\|_2 + \rho\|v^{k+1} - v^k\|_2 \\ &\leq (\sqrt{2}c + \rho)\|v^{k+1} - v^k\|_2, \quad k \in \mathbb{N}. \end{aligned}$$

Hence, **H2** (Relative error condition) is satisfied with  $b = \sqrt{2}c + \rho$ .

Moreover, since  $\{v^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^M$  is bounded and thus relative compact, there exists a subsequence  $\{v^{k_j}\}_{j \in \mathbb{N}}$  and  $\bar{v} \in \mathbb{R}^M$  such that  $v^{k_j} \rightarrow \bar{v}$  as  $j \rightarrow +\infty$ . As mentioned above  $\{v_1^k | k \in \mathbb{N}\} \subset \hat{S}$ , it holds that  $\{v^k\}_{k \in \mathbb{N}} \subset \hat{S}$ . Since  $\hat{S}$  is closed,  $\bar{v} \in \hat{S}$ . Additionally, it is easy to see that  $\tilde{F}$  is continuous on  $\hat{S}$ . Therefore,  $\tilde{F}(v^{k_j}) \rightarrow \tilde{F}(\bar{v})$  as  $j \rightarrow +\infty$ . Hence, **H3** (Continuity condition) is fulfilled.

By Lemma 3,  $\tilde{F}$  is semi-algebraic on  $\hat{S}$ . Hence, by Lemma 2,  $\tilde{F}$  satisfies the Kurdyka–Łojasiewicz property at  $\bar{v} \in \hat{S}$ . Then, the proof is complete by applying Lemma 1.  $\square$

**Corollary 5** (Global convergence) *Let  $\{(\mathcal{X}^k, \mathcal{B}^k, \mathcal{C}^k)\}_{k \in \mathbb{N}}$  be the sequence generated by the PAM iteration (2.3)–(2.5) with an arbitrary initial guess  $(\mathcal{X}^0, \mathcal{B}^0, \mathcal{C}^0) \in S \times \mathbb{R}^{n_1 \times r \times n_3} \times \mathbb{R}^{r \times n_2 \times n_3}$ . Then, there exists  $(\tilde{\mathcal{X}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}) \in S \times \mathbb{R}^{n_1 \times r \times n_3} \times \mathbb{R}^{r \times n_2 \times n_3}$  such that*

- (i)  $(\mathcal{X}^k(:); \mathcal{B}^k(:); \mathcal{C}^k(:)) \rightarrow (\tilde{\mathcal{X}}(:); \tilde{\mathcal{B}}(:); \tilde{\mathcal{C}}(:))$ ;
- (ii)  $0 \in \partial F(\tilde{\mathcal{X}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}})$ ;
- (iii) the sequence  $\{(\mathcal{X}^k, \mathcal{B}^k, \mathcal{C}^k)\}_{k \in \mathbb{N}}$  has a finite length, i.e.,

$$\sum_{k=0}^{+\infty} \sqrt{\|\mathcal{X}^k - \tilde{\mathcal{X}}\|_{\ell_2}^2 + \|\mathcal{B}^k - \tilde{\mathcal{B}}\|_{\ell_2}^2 + \|\mathcal{C}^k - \tilde{\mathcal{C}}\|_{\ell_2}^2} < +\infty.$$

In the above corollary,  $\partial F(\mathcal{X}, \mathcal{B}, \mathcal{C})$  should be understood as the sub-differential of a multivariate function  $F$  whose variables are entries of  $\mathcal{X}$ ,  $\mathcal{B}$ , and  $\mathcal{C}$ .

## 4 Implementation

In this section, some efficient algorithms are introduced to solve subproblems (2.3)–(2.5) arising from PAM method.

Firstly, it is easy to check that (2.3) has an unique closed-form solution as follows:

$$\mathcal{X}^{k+1}(i, j, k) = \begin{cases} \min \left\{ 1, \max \left\{ \frac{(\mathcal{B}^k * \mathcal{C}^k + \rho \mathcal{X}^k)(i, j, k)}{1 + \rho}, 0 \right\} \right\}, & (i, j, k) \notin \Omega, \\ \mathcal{M}(i, j, k), & (i, j, k) \in \Omega, \end{cases} \quad (4.1)$$

It is clear that (2.4) and (2.5) are both strictly convex optimization problems with strongly convex objective function, which possess global and unique minimizers. There are many efficient solvers for finding global minimizers of (2.4) and (2.5); see, e.g., the Bregman methods [18], proximal splitting methods [12], primal-dual methods [11,15], Douglas–Rachford methods [13], and alternating direction of multiplier methods (ADMM) [16,43,47]. We choose ADMM as a solver for the minimization problems (2.4) and (2.5), since the convergence of ADMM for both (2.4) and (2.5) is theoretically guaranteed (see, e.g., [20]).

For introducing the ADMM iterations, we need to specify the expressions of  $\mathcal{D}_{1k}$ 's and  $\mathcal{D}_{2k}$ 's. In general, the backward partial difference linear operators  $\mathcal{D}_{1k}$ 's and  $\mathcal{D}_{2k}$ 's with homogeneous Neumann boundary conditions have the following representations:

$$\begin{aligned} \mathcal{D}_{1k}(:, :, i) &= \begin{cases} \bigotimes_{j=1}^{q_1} L_{n_{1j}}^{s_{kj}}, & i = 1, \\ O_{n_1}, & i > 1, \end{cases}, \\ \prod_{j=1}^{q_1} n_{1j} &= n_1, \quad \{n_{1j}\}_{j=1}^{q_1} \subset \mathbb{N}^+, \quad \sum_{j=1}^{q_1} s_{kj} = 1, \\ \{s_{kj}\}_{j=1}^{q_1} &\subset \mathbb{N}, \quad k = 1, 2, \dots, m_1. \quad (4.2) \end{aligned}$$

$$\begin{aligned} \mathcal{D}_{2k}(:, :, i) &= \begin{cases} \bigotimes_{j=1}^{q_2} (L_{n_{2j}}^{t_{kj}})^T, & i = 1, \\ O_{n_2}, & i > 1, \end{cases}, \\ \prod_{j=1}^{q_2} n_{2j} &= n_2, \quad \{n_{2j}\}_{j=1}^{q_2} \subset \mathbb{N}^+, \quad \sum_{j=1}^{q_2} t_{kj} = 1, \\ \{t_{kj}\}_{j=1}^{q_2} &\subset \mathbb{N}, \quad k = 1, 2, \dots, m_2. \end{aligned}$$

$$\prod_{j=1}^{q_2} n_{2j} = n_2, \quad \{n_{1j}\}_{j=1}^{q_2} \subset \mathbb{N}^+, \quad \sum_{j=1}^{q_2} t_{kj} = 1, \\ \{t_{kj}\}_{j=1}^{q_2} \subset \mathbb{N}, \quad k = 1, 2, \dots, m_2, \quad (4.3)$$

where

$$L_m = \begin{bmatrix} 0 & & & \\ -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{m \times m}, \quad L_m^0 := I_m,$$

$O_m$  and  $I_m$  denote  $m \times m$  zero matrix and  $m \times m$  identity matrix, respectively, “ $\otimes$ ” denotes Kronecker product.  $L_m(1, 1) = 0$  is due to the homogeneous Neumann boundary condition.

(2.4) is equivalent to the following equality constrained problem

$$\begin{aligned} \arg \min_{\mathcal{B}, \{\mathcal{Q}_{1j}\}_{j=1}^{m_1}} & \frac{1}{2} \|\mathcal{X}^{k+1} \\ & - \mathcal{B} * \mathcal{C}^k\|_{\ell_2}^2 + \frac{\rho}{2} \|\mathcal{B} - \mathcal{B}^k\|_{\ell_2}^2 + \frac{\mu}{2} \|\mathcal{B}\|_{\ell_2}^2 + \alpha \sum_{j=1}^{m_1} \|\mathcal{Q}_{1j}\|_{\ell_1}, \\ \text{s.t., } & \mathcal{Q}_{1j} = \mathcal{D}_{1j} * \mathcal{B}, \quad j = 1, 2, \dots, m_1. \end{aligned} \quad (4.4)$$

Note that the objective function in (4.4) is the summation of  $m_1 + 1$  single variable functions with  $\mathcal{B}, \{\mathcal{Q}_{1j}\}_{j=1}^{m_1}$  as their individual variables. Thus, ADMM is applicable. Let  $\{\mathcal{Y}_{1j}\}_{j=1}^{m_1}$  be Lagrangian multipliers for (4.4). Then, the ADMM iteration scheme for (4.4) can be described as follows:

$$\begin{aligned} \mathcal{Q}_{1j} & \leftarrow \arg \min_{\mathcal{Q}_{1j}} \alpha_1 \|\mathcal{Q}_{1j}\|_{\ell_1} + \frac{\beta}{2} \left\| \mathcal{Q}_{1j} - \mathcal{D}_{1j} * \mathcal{B} + \frac{\mathcal{Y}_{1j}}{\beta} \right\|_{\ell_2}^2, \\ j & = 1, 2, \dots, m_1, \end{aligned} \quad (4.5)$$

$$\begin{aligned} \mathcal{B} & \leftarrow \arg \min_{\mathcal{B}} \frac{1}{2} \|\mathcal{X}^{k+1} - \mathcal{B} * \mathcal{C}^k\|_{\ell_2}^2 + \frac{\rho}{2} \|\mathcal{B} - \mathcal{B}^k\|_{\ell_2}^2 + \frac{\mu}{2} \|\mathcal{B}\|_{\ell_2}^2 \\ & + \frac{\beta}{2} \sum_{j=1}^{m_1} \left\| \mathcal{Q}_{1j} - \mathcal{D}_{1j} * \mathcal{B} + \frac{\mathcal{Y}_{1j}}{\beta} \right\|_{\ell_2}^2, \end{aligned} \quad (4.6)$$

$$\mathcal{Y}_{1j} \leftarrow \mathcal{Y}_{1j} + \beta(\mathcal{Q}_{1j} - \mathcal{D}_{1j} * \mathcal{B}), \quad j = 1, 2, \dots, m_1. \quad (4.7)$$

Repeating the iteration process (4.5)–(4.7) sufficiently many times, the so-obtained  $\mathcal{B}$  is taken as an iterative solution to  $\mathcal{B}^{k+1}$ . Note that among these many rounds of iterative repeating of (4.5)–(4.7), the first round requires initial guesses of  $\mathcal{B}$  and  $\mathcal{Y}_{1j}$ 's, for which we set  $\mathcal{B}^k$  as the initial guess of  $\mathcal{B}$  and  $\mathcal{O}$  (zero tensor) as the initial guess of  $\mathcal{Y}_{1j}$ 's.

By the well-known soft thresholding, (4.5) has a unique solution as follows:

$$\begin{aligned} \mathcal{Q}_{1j}(l_1, l_2, l_3) & \leftarrow T_{\frac{\alpha_1}{\beta}} \left( \left( \mathcal{D}_{1j} * \mathcal{B} - \frac{\mathcal{Y}_{1j}}{\beta} \right) (l_1, l_2, l_3) \right), \\ (l_1, l_2, l_3) & \in \Xi(n_1) \times \Xi(r) \times \Xi(n_3), \\ j & = 1, 2, \dots, m_1, \end{aligned} \quad (4.8)$$

where

$$T_\mu(y) := \begin{cases} (|y| - \mu)\text{sign}(y), & |y| > \mu, \\ 0, & |y| \leq \mu, \end{cases}, \quad y \in \mathbb{R}.$$

In what follows, we introduce some notations and lemmas for fast implementation and solution of (4.6). For a third-order tensor  $\mathcal{Z} \in \mathbb{C}^{l \times m \times n}$ , define  $\mathcal{F}(\mathcal{Z}) \in \mathbb{C}^{l \times m \times n}$ ,  $\mathbb{B}(\mathcal{Z}) \in \mathbb{C}^{ln \times mn}$  as

$$\begin{aligned} \mathcal{F}(\mathcal{Z})(i, j, k) & := \sum_{q=1}^n \mathcal{Z}(i, j, q) \exp(-2\pi i(q-1)(k-1)/n), \\ (i, j, k) & \in \Xi(l) \times \Xi(m) \times \Xi(n), \\ \mathbb{B}(\mathcal{Z}) & := \text{blockdiag}(\mathcal{Z}(:, :, 1), \mathcal{Z}(:, :, 2), \dots, \mathcal{Z}(:, :, n)), \end{aligned}$$

where  $i = \sqrt{-1}$ . Then, by the properties of discrete Fourier transform, the inverse mapping  $\mathcal{F}^{-1}$  of  $\mathcal{F}$  is given by

$$\begin{aligned} \mathcal{F}^{-1}(\mathcal{Z})(i, j, k) & := \frac{1}{n} \sum_{q=1}^n \mathcal{Z}(i, j, q) \exp(2\pi i(q-1)(k-1)/n), \\ (i, j, k) & \in \Xi(l) \times \Xi(m) \times \Xi(n), \end{aligned}$$

for any  $\mathcal{Z} \in \mathbb{C}^{l \times m \times n}$ . For a block diagonal matrix  $Z = \text{blockdiag}(Z_1, Z_2, \dots, Z_n) \in \mathbb{C}^{ln \times mn}$  with  $Z_i \in \mathbb{C}^{l \times m}$  ( $i = 1, 2, \dots, n$ ), define  $\mathbb{T}(Z) \in \mathbb{C}^{l \times m \times n}$  such that

$$\mathbb{T}(Z)(i, j, k) := Z_k(i, j), \quad (i, j, k) \in \Xi(l) \times \Xi(m) \times \Xi(n).$$

By the above definitions, it is clear that  $\mathcal{F}(\mathcal{F}^{-1}(\mathcal{Z})) = \mathcal{F}^{-1}(\mathcal{F}(\mathcal{Z})) = \mathcal{Z}$ ,  $\mathcal{Z} = \mathbb{T}(\mathbb{B}(\mathcal{Z}))$ ,  $Z = \mathbb{B}(\mathbb{T}(Z))$  for any third-order tensor  $\mathcal{Z}$  and any block diagonal matrix  $Z$  whose blocks have the same size. Also, from the definition of  $\mathcal{F}$  and  $\mathcal{F}^{-1}$ , we know that the evaluations of  $\mathcal{F}(\mathcal{Z})$  and  $\mathcal{F}^{-1}(\mathcal{Z})$  can be fast implemented by fast Fourier transforms (FFTs) as follows:

$$\mathcal{F}(\mathcal{Z})(i, j, :) = \text{fft}(\mathcal{Z}(i, j, :)), \quad \mathcal{F}^{-1}(\mathcal{Z})(i, j, :) = \text{ifft}(\mathcal{Z}(i, j, :)), \quad (4.9)$$

for any third-order tensor  $\mathcal{Z}$ , where fft and ifft denote FFT and inverse FFT, respectively. Denote  $\tilde{\mathcal{F}} := \mathbb{B} \circ \mathcal{F}$ .

**Lemma 6** ([24])

- (i) For  $\mathcal{Z}_1 \in \mathbb{C}^{l \times m \times n}$  and  $\mathcal{Z}_2 \in \mathbb{C}^{m \times \tilde{l} \times n}$ ,  $\tilde{\mathcal{F}}(\mathcal{Z}_1 * \mathcal{Z}_2) = \tilde{\mathcal{F}}(\mathcal{Z}_1) \tilde{\mathcal{F}}(\mathcal{Z}_2)$ .

(ii) For  $\mathcal{Z} \in \mathbb{C}^{l \times m \times n}$ ,  $\|\mathcal{Z}\|_{\ell_2}^2 = \frac{1}{n} \|\tilde{F}(\mathcal{Z})\|_F^2$ .

It is clear that  $\mathcal{F}, \mathbb{B}$  are both linear bijections between underlying spaces. Thus,  $\tilde{\mathcal{F}}$  is also a linear bijection, and it is easy to see that  $\tilde{\mathcal{F}}^{-1} = \mathcal{F}^{-1} \circ \mathbb{T}$ . Hence, Lemma 6 (i) implies that (4.1) requires  $O(n_1 n_2 n_3(r + \log n_3))$  operation cost. Moreover, by Lemma 6, (4.6) is equivalent to

$$\begin{aligned} \bar{B} &\leftarrow \arg \min_{\bar{B}} \frac{1}{2} \|\bar{X}^{k+1} - \bar{B} \bar{C}^k\|_F^2 + \frac{\rho}{2} \|\bar{B} \\ &\quad - \bar{B}^k\|_F^2 + \frac{\mu}{2} \|\bar{B}\|_F^2 + \frac{\beta}{2} \sum_{j=1}^{m_1} \left\| \bar{Q}_{1j} - \bar{D}_{1j} \bar{B} + \frac{\bar{Y}_{1j}}{\beta} \right\|_F^2, \end{aligned}$$

with  $\bar{B} = \tilde{\mathcal{F}}(\mathcal{B})$ ,  $\bar{B}^k = \tilde{\mathcal{F}}(\mathcal{B}^k)$ ,  $\bar{X}^{k+1} = \tilde{\mathcal{F}}(\mathcal{X}^{k+1})$ ,  $\bar{C}^k = \tilde{\mathcal{F}}(\mathcal{C}^k)$ ,  $\bar{Q}_{1j} = \tilde{\mathcal{F}}(\mathcal{Q}_{1j})$ ,  $\bar{D}_{1j} = \tilde{\mathcal{F}}(\mathcal{D}_{1j})$ ,  $\bar{Y}_{1j} = \tilde{\mathcal{F}}(\mathcal{Y}_{1j})$ . Treating real part and imaginary part of  $\bar{B}$  as real variables of the objective function in the above problem, it is then easy to check that the unique solution of the above problem is actually the solution of the following matrix equation

$$\bar{B} [\bar{C}^k (\bar{C}^k)^* + (\rho + \mu) I_{rn_3}] + \beta \left( \sum_{j=1}^{m_1} \bar{D}_{1j}^* \bar{D}_{1j} \right) \bar{B} = R^k, \quad (4.10)$$

where  $R^k = \bar{X}^{k+1} (\bar{C}^k)^* + \rho \bar{B}^k + \sum_{j=1}^{m_1} \bar{D}_{1j}^* (\beta \bar{Q}_{1j} + \bar{Y}_{1j})$ . Notice that matrices appearing in (4.10) all have block diagonal structures. Additionally, from the definition of  $\mathcal{D}_{1j}$  in (4.2), we see that diagonal blocks of  $\bar{D}_{1j}$  are identical, i.e.,  $\bar{D}_{1j} = I_{n_3} \otimes \left( \bigotimes_{m=1}^{q_1} L_{n_{1m}}^{s_{jm}} \right)$  for  $j = 1, 2, \dots, m_1$ . Therefore, (4.10) is equivalent to  $n_3$  many matrix equations of smaller size as follows:

$$\bar{B}_l [\bar{C}_l^k (\bar{C}_l^k)^* + (\rho + \mu) I_r] + \beta \left( \sum_{j=1}^{m_1} \bigotimes_{m=1}^{q_1} H_{n_{1m}}^{s_{jm}} \right) \bar{B}_l = R_l^k, \quad l = 1, 2, \dots, n_3, \quad (4.11)$$

where  $\bar{B}_l$ ,  $\bar{C}_l^k$  and  $R_l^k$  denote  $l$ th diagonal block of  $\bar{B}$ ,  $\bar{C}^k$  and  $R^k$ , respectively,

$$H_m := L_m^T L_m = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{m \times m}, \quad H_m^0 := I_m.$$

Let  $\bar{C}_l^k (\bar{C}_l^k)^* + (\rho + \mu) I_r = U_l^k V_l^k (U_l^k)^*$  denote the singular value decomposition (SVD). Notice that  $r$  is a small number. Hence, the computation of the SVD,  $U_l^k V_l^k (U_l^k)^*$ , is economic, which requires  $O(r^3)$  operation cost. Also, it is

straightforward to verify that  $H_m$  has the following orthogonal diagonalization form

$$\begin{aligned} H_m &= K_m \Lambda_m K_m^T, \\ K_m &= \sqrt{\frac{2}{m}} \left[ \sqrt{(1 + \delta_{j,1})^{-1}} \cos \left( \frac{\pi(2i-1)(j-1)}{2m} \right) \right]_{i,j=1}^m, \\ \delta_{j,1} &= \begin{cases} 1, & j = 1, \\ 0, & \text{otherwise,} \end{cases}, \quad K_m K_m^T = I_m, \\ \Lambda_m &= 4 \times \text{diag} \left( \sin^2 \left( \frac{(i-1)\pi}{2m} \right) \right)_{i=1}^m. \end{aligned}$$

For a given vector  $v \in \mathbb{R}^m$ ,  $K_m v$  and  $K_m^T v$  can be computed by two types of discrete cosine transforms in MATLAB language such that  $K_m v = \text{dct}(v, \text{'Type'}, 3)$  and  $K_m^T v = \text{dct}(v, \text{'Type'}, 2)$ , which requires  $O(m \log m)$  operations. Denote  $\Lambda_m^0 := I_m$  and

$$\tilde{K}_1 = \bigotimes_{m=1}^{q_1} K_{n_{1m}}, \quad \tilde{\Lambda}_1 = \sum_{j=1}^{m_1} \bigotimes_{m=1}^{q_1} \Lambda_{n_{1m}}^{s_{jm}}.$$

Then, it is clear that  $\tilde{\Lambda}_1$  is a  $n_1 \times n_1$  nonnegative diagonal matrix and  $\sum_{j=1}^{m_1} \bigotimes_{m=1}^{q_1} H_{n_{1m}}^{s_{jm}} = \tilde{K}_1 \tilde{\Lambda}_1 \tilde{K}_1^T$ . Multiplying  $\tilde{K}_1^T$  from the left and multiplying  $U_l^k$  from the left on both sides of (4.11), we see that (4.11) is equivalent to

$$\hat{B}_l V_l^k + \beta \tilde{\Lambda}_1 \hat{B}_l = \hat{R}_l^k, \quad l = 1, 2, \dots, n_3, \quad (4.12)$$

where  $\hat{B}_l = \tilde{K}_1^T \bar{B}_l U_l^k$ ,  $\hat{R}_l^k = \tilde{K}_1^T \bar{R}_l^k U_l^k$ . Since  $V_l^k$  and  $\tilde{\Lambda}_1$  are both diagonal matrices,  $\hat{B}_l$  can be fast solved by

$$\hat{B}_l(m, n) = \frac{\hat{R}_l^k(m, n)}{\beta \tilde{\Lambda}_1(m) + V_l^k(n)}, \quad (m, n) \in \Xi(n_1) \times \Xi(r). \quad (4.13)$$

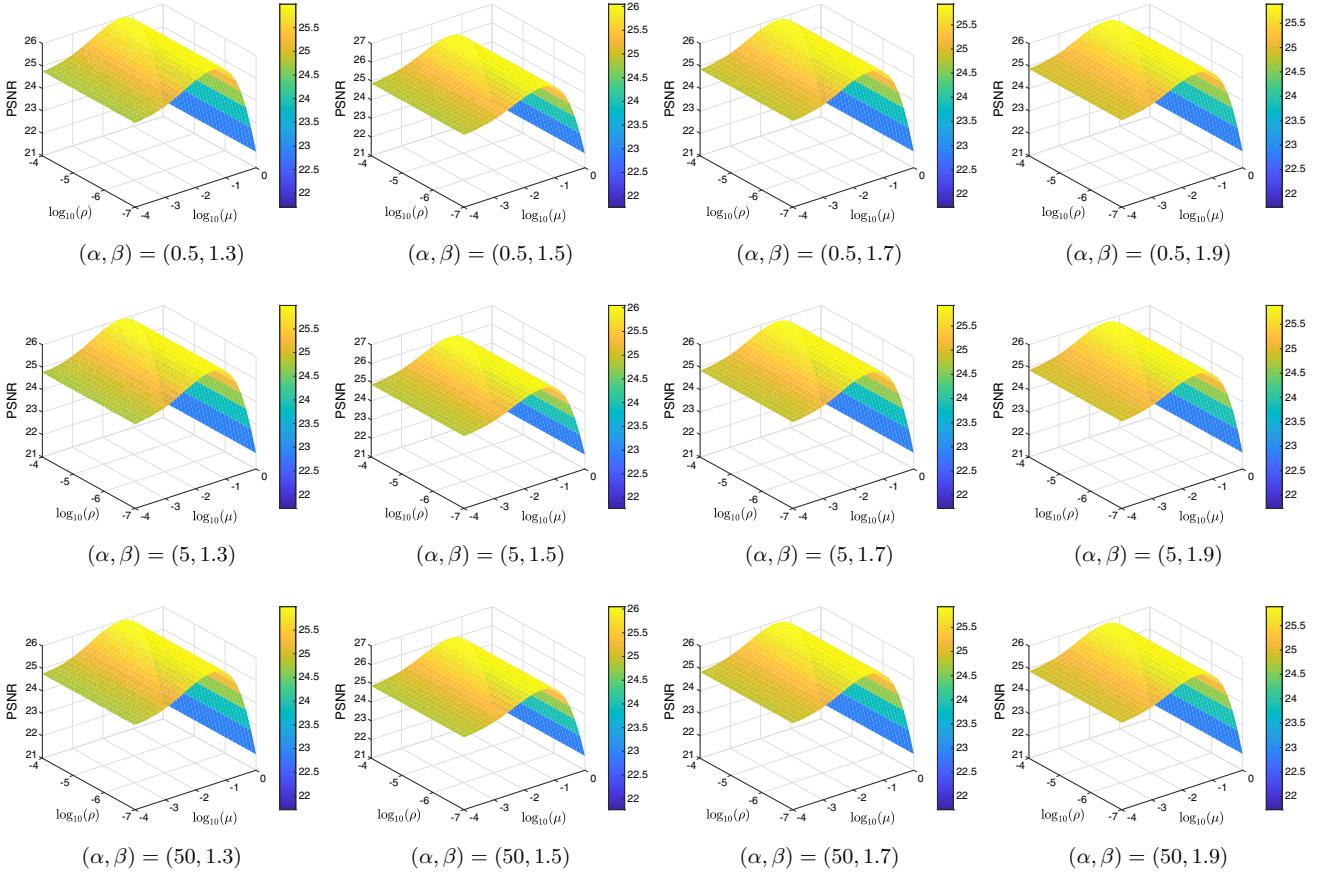
Hence,  $\mathcal{B}$  in (4.6) can be solved by

$$\mathcal{B} = \tilde{\mathcal{F}}^{-1}(\text{blockdiag}(\tilde{K}_1 \hat{B}_l (U_l^k)^*)_{l=1}^{n_3}). \quad (4.14)$$

To save the operation cost of each PAM iteration, the iterative process (4.5)–(4.7) is performed only one time. Then, it is easy to check that approximately solving  $\mathcal{B}$ -subproblem (2.4) requires  $O(n_1 n_3(r n_2 + n_2 \log n_3 + r \log n_1))$  operations.

Similarly to (4.5)–(4.7), one can perform one time of ADMM iteration with the same fashion of fast implementation to approximately solve (2.5), which requires  $O(n_2 n_3(r n_1 + n_1 \log n_3 + r \log n_2))$  operations.

To conclude, with the introduced fast implementation in this section, each PAM iteration (2.3)–(2.5) requires



**Fig. 2** The PSNR values of the recovered results by TCTF-TVT on the color image “Panda” with the sampling rate 0.1 and different  $(\alpha, \beta, \mu, \rho)$  (Color figure online)

$O(n_3(rn_1n_2 + n_1n_2 \log n_3 + rn_2 \log n_2 + rn_1 \log n_1))$  operations.

## 5 Numerical Experiment

In this section, we test the proposed TCTF-TVT model on several examples of image and video completion, comparing it with the TCTF model proposed in [51], TNN model proposed in [46], the LRTC-TV-I and LRTC-TV-II models proposed in [29]. All numerical experiments presented in this section are performed via MATLAB R2017a on a PC with the configuration: Intel(R) Core(TM) i5-6500 CPU 3.20 GHz and 32 GB RAM.

Let  $p \in (0, 1)$  be a given number. The set of indices of known entries,  $\Omega$ , is chosen randomly in MATLAB language as follows:

$$\Omega := \text{find}(\text{rand}(n_1n_2n_3, 1) < p). \quad (5.1)$$

$p$  is also called sampling rate.

Let  $\mathcal{X}_{ref}$  represent the ground truth of an  $n_b$ -band video with  $n_f$  frames (e.g., the case of  $n_f = 1$  and  $n_b = 3$  rep-

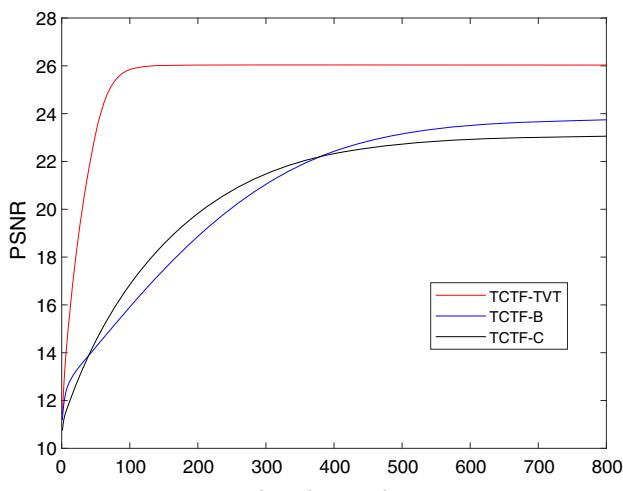
resents a color image). Let  $X_{ij}$  ( $1 \leq i \leq n_b$ ,  $1 \leq j \leq n_f$ ) denote the  $i$ th band of  $j$ th frame of  $\mathcal{X}_{ref}$ . Let  $\hat{\mathcal{X}}$  represent a recovered video from some incomplete observation of  $\mathcal{X}_{ref}$  whose  $i$ th band of  $j$ th frame is denoted by  $\hat{X}_{ij}$ . Then, two metrics are defined as follows to measure the quality of the recovered result  $\hat{\mathcal{X}}$ :

$$\begin{aligned} \text{PSNR}(\hat{\mathcal{X}}, \mathcal{X}_{ref}) &:= \frac{1}{n_b n_f} \sum_{i=1}^{n_b} \sum_{j=1}^{n_f} \text{PSNR}(\hat{X}_{ij}, X_{ij}, \|X_{ij}(\cdot)\|_\infty), \\ \text{SSIM}(\hat{\mathcal{X}}, \mathcal{X}_{ref}) &:= \frac{1}{n_b n_f} \sum_{i=1}^{n_b} \sum_{j=1}^{n_f} \text{SSIM}(\hat{X}_{ij}, X_{ij}), \end{aligned}$$

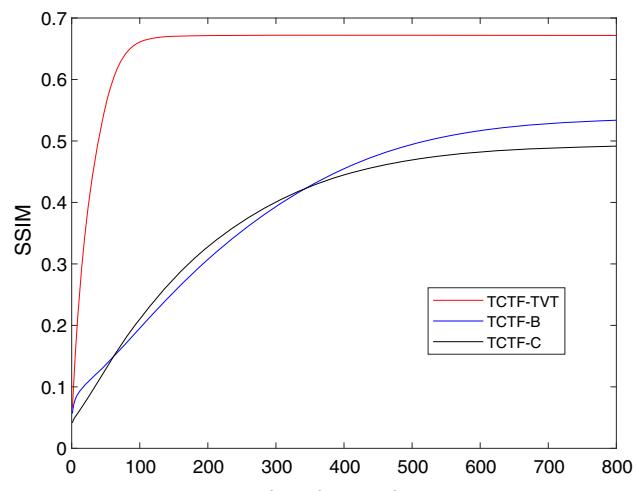
where PSNR and SSIM with images as inputs are two built-in functions in MATLAB.

Denote by “Time”, the computational time in unit of second. The maximum iteration number for TCTF-TVT is set to be 500 if not specified. The initial guess  $\mathcal{X}^0$  for TCTF-TVT is set as

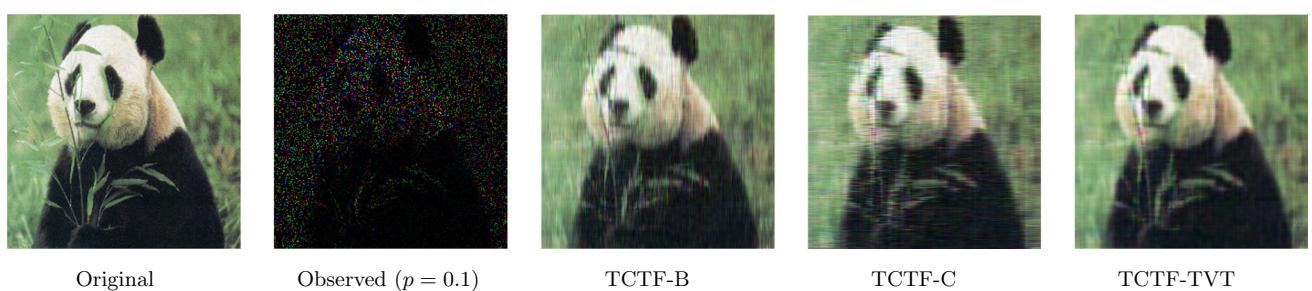
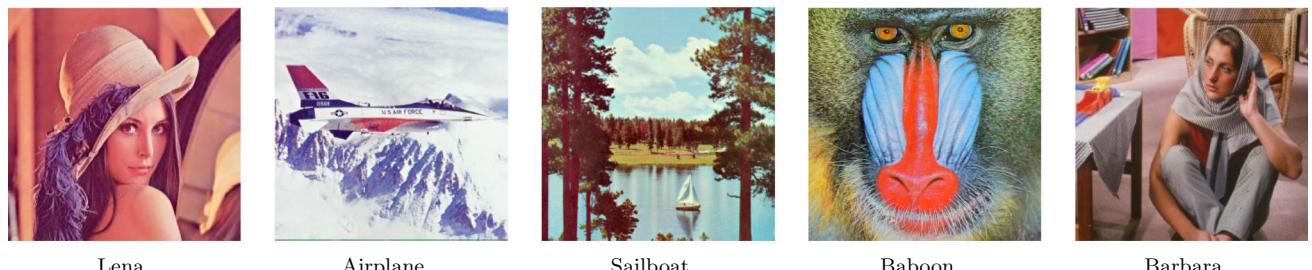
$$\mathcal{X}^0(i, j, k) = \begin{cases} \mathcal{M}(i, j, k), & (i, j, k) \in \Omega, \\ \text{random value}, & (i, j, k) \notin \Omega, \end{cases}, \quad (5.2)$$



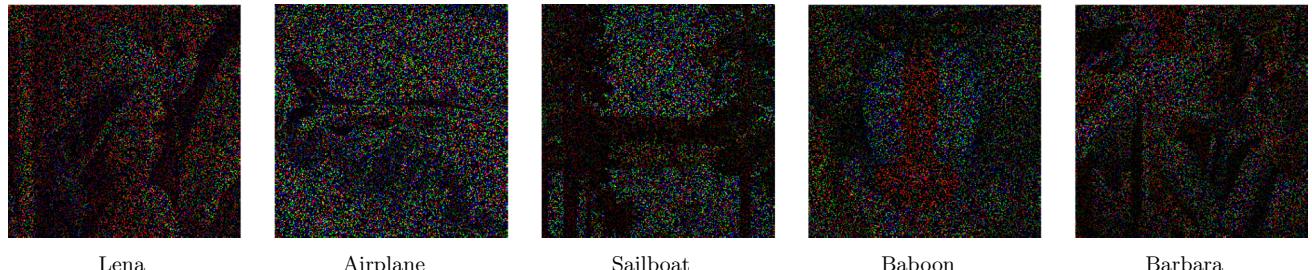
PSNR of first 800 iterations



SSIM of first 800 iterations

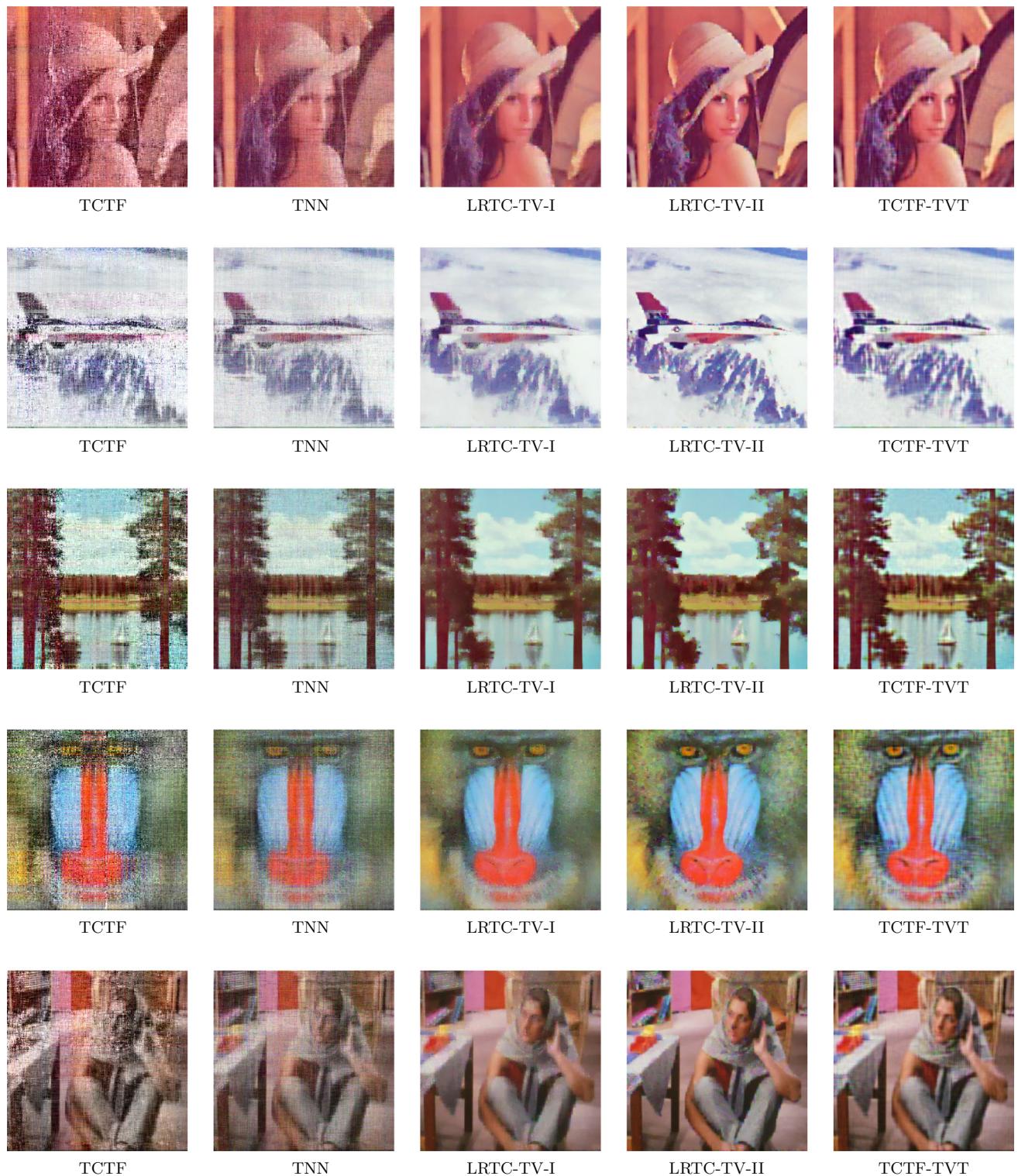
**Fig. 3** The PSNR and SSIM values of the recovered results by TCTF-TVT, TCTF-B, and TCTF-C on the color image “Panda” with the sampling rate 0.1 (Color figure online)Original                      Observed ( $p = 0.1$ )                      TCTF-B                      TCTF-C                      TCTF-TVT**Fig. 4** The original color image “Panda”, the observed image, and the recovered images by TCTF-B, TCTF-C, and TCTF-TVT (Color figure online)

Lena                      Airplane                      Sailboat                      Baboon                      Barbara

**Fig. 5** The original color images (Color figure online)

Lena                      Airplane                      Sailboat                      Baboon                      Barbara

**Fig. 6** The observed color images with the sampling rate 0.15 (Color figure online)



**Fig. 7** The recovered images by TCTF, TNN, LRTC-TV-I, LRTC-TV-II, and TCTF-TVT from observed data with sampling rate  $p = 0.15$  (Color figure online)

**Table 1** The PSNR and SSIM values of recovered images by different methods on color images “Lena,” “Airplane,” and “Sailboat” with different sampling rates

$p$	Model	Color image			Lena			Airplane			Sailboat		
		PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time
0.05	TCTF	11.32	0.051	2.75	10.30	0.049	2.00	10.77	0.060	2.22			
	TNN	16.36	0.225	5.51	17.01	0.301	5.27	15.32	0.230	5.33			
	LRTC-TV-I	18.41	0.540	24.07	18.55	0.567	22.73	16.91	0.448	24.26			
	LRTC-TV-II	22.52	<b>0.705</b>	83.87	21.50	<b>0.711</b>	83.98	20.12	<b>0.617</b>	85.87			
	TCTF-TVT	<b>22.71</b>	0.647	25.65	<b>21.91</b>	0.647	25.23	<b>20.42</b>	0.576	25.95			
0.1	TCTF	14.09	0.133	3.16	12.77	0.118	3.20	11.65	0.103	3.15			
	TNN	18.86	0.348	4.92	19.05	0.413	4.76	17.40	0.333	4.85			
	LRTC-TV-I	21.87	0.679	23.50	21.17	0.679	22.76	19.61	0.591	23.74			
	LRTC-TV-II	<b>25.02</b>	<b>0.796</b>	83.75	23.40	<b>0.785</b>	85.53	21.47	<b>0.691</b>	82.72			
	TCTF-TVT	24.67	0.743	26.00	<b>23.77</b>	0.751	26.69	<b>22.09</b>	0.674	27.93			
0.15	TCTF	17.69	0.306	3.19	17.06	0.351	3.08	15.97	0.315	3.19			
	TNN	20.81	0.461	4.75	20.78	0.510	4.63	18.99	0.426	5.20			
	LRTC-TV-I	23.73	0.758	23.41	22.92	0.757	22.60	21.30	0.682	24.35			
	LRTC-TV-II	<b>26.00</b>	<b>0.833</b>	55.42	23.80	<b>0.808</b>	85.49	21.91	0.717	84.25			
	TCTF-TVT	25.81	0.796	30.31	<b>24.97</b>	0.798	28.58	<b>23.21</b>	<b>0.731</b>	27.65			

The best PSNR and SSIM values are highlighted in bold

**Table 2** The PSNR and SSIM values of the recovered images by different methods on color images “Baboon” and “Barbara” with different sampling rates

$p$	Model	Color image			Baboon			Barbara		
		PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time
0.05	TCTF	11.54	0.066	3.07	11.53	0.062	2.90			
	TNN	15.46	0.171	5.48	15.83	0.228	5.45			
	LRTC-TV-I	17.29	0.277	24.03	17.20	0.414	24.12			
	LRTC-TV-II	<b>19.98</b>	<b>0.390</b>	84.59	22.10	<b>0.645</b>	86.37			
	TCTF-TVT	19.70	0.367	25.52	<b>22.23</b>	0.620	24.05			
0.1	TCTF	12.67	0.119	3.27	13.61	0.143	3.15			
	TNN	17.19	0.262	4.90	18.23	0.359	5.04			
	LRTC-TV-I	19.61	0.393	23.65	20.90	0.600	23.33			
	LRTC-TV-II	<b>20.79</b>	0.458	85.32	23.41	0.712	85.35			
	TCTF-TVT	20.68	<b>0.464</b>	27.57	<b>24.17</b>	<b>0.716</b>	26.69			
0.15	TCTF	15.20	0.242	3.15	17.68	0.378	3.22			
	TNN	18.52	0.347	4.74	20.28	0.478	4.96			
	LRTC-TV-I	20.79	0.481	23.43	23.26	0.707	23.39			
	LRTC-TV-II	21.34	0.516	89.51	24.51	0.762	93.14			
	TCTF-TVT	<b>21.36</b>	<b>0.531</b>	24.01	<b>25.35</b>	<b>0.769</b>	24.73			

The best PSNR and SSIM values are highlighted in bold

The initial guesses  $\mathcal{B}^0$  and  $\mathcal{C}^0$  for TCTF-TVT are set as

$$\mathcal{B}^0 = \mathcal{F}^{-1}(\hat{\mathcal{B}}^0), \quad \mathcal{C}^0 = \mathcal{F}^{-1}(\hat{\mathcal{C}}^0), \quad \hat{\mathcal{B}}^0(:, :, k) := \sum_{i=1}^r \sigma_{ki} u_{ki} e_i^T,$$

$$\hat{\mathcal{C}}^0(:, :, k) := \sum_{i=1}^r e_i v_{ki}^T, \quad k = 1, 2, \dots, n_3,$$

where  $\sigma_{ki}$  is the  $i$ th largest singular value of  $\mathcal{F}(\mathcal{X}^0)(:, :, k)$ ;  $u_{ki}$  and  $v_{ki}$  are the left and the right singular vectors corresponding to  $\sigma_{k,i}$ ;  $e_i$  denotes the  $i$ th column of  $r \times r$  identity

matrix. The stopping criterion for TCTF-TVT is set as follows:

$$\max \left\{ \frac{\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_{\ell_2}^2}{\|\mathcal{X}^{k+1}\|_{\ell_2}^2}, \frac{\|\mathcal{B}^{k+1} - \mathcal{B}^k\|_{\ell_2}^2}{\|\mathcal{B}^{k+1}\|_{\ell_2}^2}, \frac{\|\mathcal{C}^{k+1} - \mathcal{C}^k\|_{\ell_2}^2}{\|\mathcal{C}^{k+1}\|_{\ell_2}^2} \right\} < 10^{-6}.$$

Parameters of other models are specified by following the papers in which these approaches were proposed and developed.



**Fig. 8** The 50th frame of the original color video “Salesman” (Color figure online)

**Parameter Setting of TCTF-TVT:** Throughout this section, we set  $\alpha = 5$ ,  $\beta = 1.5$ ,  $\mu = 0.04$ ,  $\rho = 5e-6$ , if not specified. We mainly focus on the tensor completion from poorly observed data. Hence, in the numerical test, the  $\Omega$  given in (5.1) is equipped with small  $p$  such that  $p \in \{0.05, 0.1, 0.15\}$ .

For a color image with size  $n_v \times n_h \times 3$ , we set  $n_1 = n_v$ ,  $n_2 = n_h$ ,  $n_3 = 3$ ,  $m_1 = m_2 = q_1 = q_2 = 1$ . To demonstrate the optimality of the parameter setting  $(\alpha, \beta, \mu, \rho) = (5, 1.5, 0.04, 5e-6)$ , we compare the PSNR of the recovered

results by TCTF-TVT with different values of  $(\alpha, \beta, \mu, \rho)$  on completion of a color image. The tested color image “Panda” is of size  $256 \times 256 \times 3$ , the ground truth of which is listed in “Original” of Fig. 4. The results of the test are plotted in Fig. 2. Since changes of SSIM are quite similar to those of PSNR, we do not present values of SSIM in Fig. 2. Figure 2 shows that (i) the performance of TCTF-TVT is not sensitive to changes of  $\alpha$  and  $\rho$ ; (ii) performance of TCTF-TVT with  $\beta = 1.5$  is better than that with  $\beta \in \{1.3, 1.7, 1.9\}$ ; (iii) when  $\mu \in [0.03, 0.05]$ , TCTF-TVT achieves its best performance. For these reasons, we set the parameters as  $(\alpha, \beta, \mu, \rho) = (5, 1.5, 0.04, 5e-6)$ .

Let TCTF-B and TCTF-C denote model (2.1) but without  $\alpha \sum_{k=1}^{m_2} ||\mathcal{C} * \mathcal{D}_{2k}||_{\ell_1} + \frac{\mu}{2} ||\mathcal{C}||_{\ell_2}^2$  and  $\alpha \sum_{k=1}^{m_1} ||\mathcal{D}_{1k} * \mathcal{B}||_{\ell_1} + \frac{\mu}{2} ||\mathcal{B}||_{\ell_2}^2$ , respectively. The same PAM iteration is employed to solve TCTF-B and TCTF-C, the parameters of which are set to be the same as the one used for solving TCTF-TVT. To test the effectiveness of the regularization terms on  $\mathcal{B}$  and  $\mathcal{C}$ , we apply TCTF-TVT, TCTF-B, and TCTF-C on the color image “Panda”, the results of which are presented in Figs. 3 and 4. Figure 3 shows that the recovered result by TCTF-TVT has a better quality than those by TCTF-B and TCTF-C. From Fig. 4, we can observe that the recovered result by TCTF-B (TCTF-C) has a staircase effect along the horizontal (vertical) direction, while such staircase effects

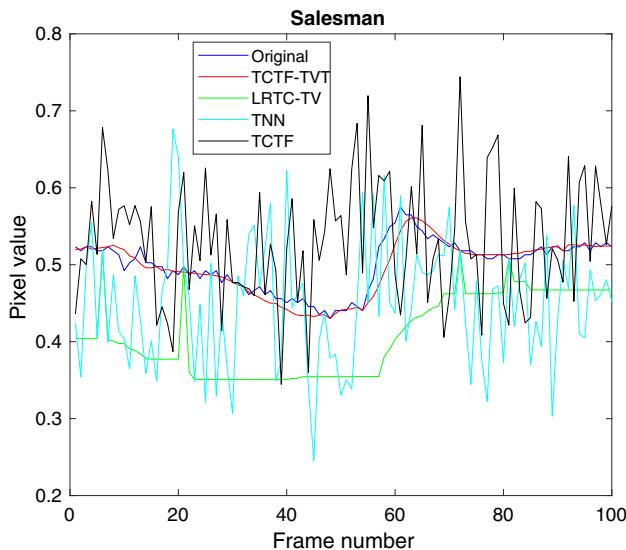


**Fig. 9** The 50th frame of the observed videos and the recovered videos by TCTF, TNN, LRTC-TV-I, and TCTF-TVT with different sampling rates (Color figure online)

**Table 3** The PSNR and SSIM values of the recovered videos by different methods on the color video “Salesman” with different sampling rates

Model	$p = 0.05$			$p = 0.1$			$p = 0.15$		
	PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time
TCTF	13.54	0.153	135.30	18.00	0.368	181.75	27.18	0.784	180.70
TNN	26.83	0.842	39.32	31.44	0.935	39.22	33.75	0.954	35.86
LRTC-TV-I	21.65	0.602	918.34	24.75	0.725	900.09	26.62	0.798	897.47
LRTC-TV-II	—	—	> 72 h	—	—	> 72 h	—	—	> 72 h
TCTF-TVT	<b>31.09</b>	<b>0.926</b>	655.31	<b>34.10</b>	<b>0.958</b>	658.99	<b>35.64</b>	<b>0.968</b>	676.23

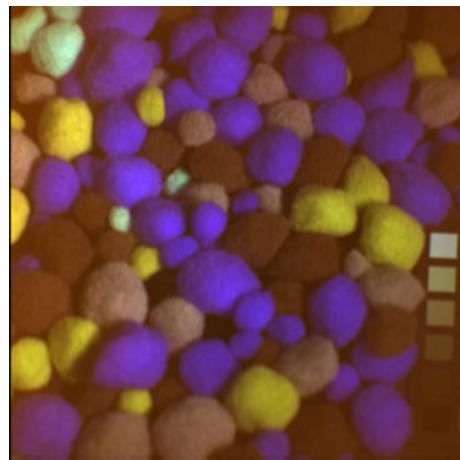
The best PSNR and SSIM values are highlighted in bold



**Fig. 10** The selected fibers along the temporal mode of the original video “Salesman” and the recovered videos by TCTF, TNN, LRTC-TV-I, and TCTF-TVT with the sampling rate 0.05 (Color figure online)

are significantly suppressed in the result by TCTF-TVT. These evidences illustrate that the regularization terms on both  $\mathcal{B}$  and  $\mathcal{C}$  contribute to the performance of TCTF-TVT, i.e., the regularizations on the two factor tensors are both effective.

**Example 1** In this example, we evaluate the performance of different methods on color images. For a color image with size  $n_v \times n_h \times 3$ , we set  $n_1 = n_v$ ,  $n_2 = n_h$ ,  $n_3 = 3$ ,  $m_1 = m_2 = q_1 = q_2 = 1$  and  $r = 40$ . We remark that not only in this paper and but also in the literature (see, e.g., [29, 51]), color image is always considered as a third-order tensor with the first two dimensions representing spatial dimensions and the third dimension representing band. All the color images tested in Example 1 are cited from [29]. The original color images and the observed color images of size  $256 \times 256 \times 3$  are shown in Figs. 5 and 6. The recovered images by TCTF, TNN, LRTC-TV-I, LRTC-TV-II and TCTF-TVT are displayed in Fig. 7. The PSNR and SSIM values of the recovered images are summarized in Tables 1 and 2. Figure 7 and Tables 1 and 2 show (i) TCTF-TVT and LRTC-TV-II have



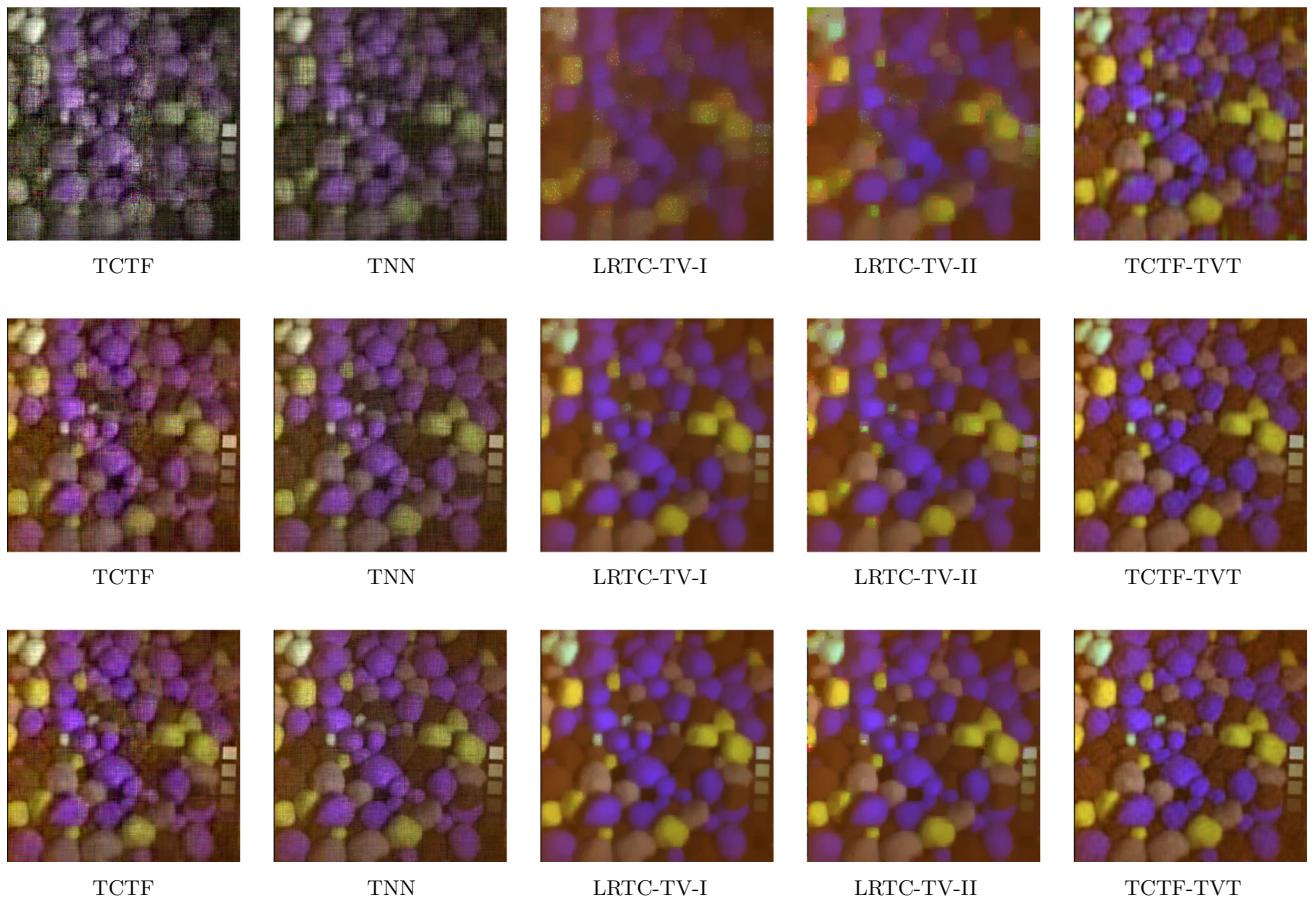
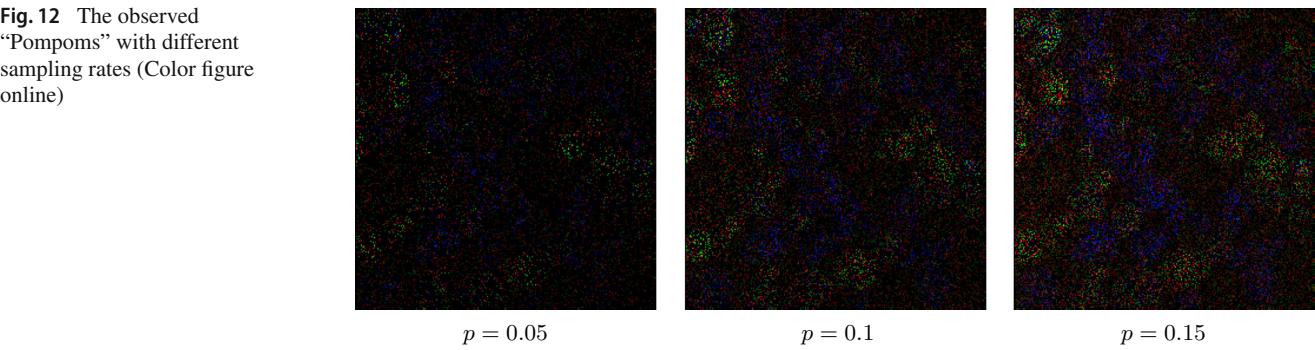
**Fig. 11** The original MSI “Pompoms” (Color figure online)

the best performance in terms of visual quality among the five models; (ii) the recovered results by TCTF-TVT and LRTC-TV-II are comparable; (iii) TCTF-TVT is more efficient than LRTC-TV-II in terms of computational time.

**Example 2** We evaluate the performance of different methods on the color video. We first form a third-order tensor by vectorizing each frame of the color video (the fourth-order tensor). Since the factor tensors of the formed third-order tensor have clear physical interpretations (the combined two spatial dimensions, the time dimension and the color dimension), Example 2 serves as a good example to justify our motivation in the introduction. For a color video of  $t$  frames with each frame of size  $n_v \times n_h \times 3$ , we set  $n_1 = n_v n_h$ ,  $n_2 = t$ ,  $n_3 = 3$ ,  $r = 30$ ,  $m_1 = q_1 = 2$ ,  $n_{11} = n_h$ ,  $n_{12} = n_v$ ,  $s_{11} = 0$ ,  $s_{12} = 1$ ,  $s_{21} = 1$ ,  $s_{22} = 0$ ,  $m_2 = q_2 = 1$ . The tested color video “Salesman”,<sup>1</sup> shown in Fig. 8, has 100 frames with each frame of size  $144 \times 176 \times 3$ . Since LRTC-TV-II is too time-consuming for color video completion, there is no result of LRTC-TV-II presented in Example 2. The 50th frame of the recovered videos is displayed in Fig. 9. The PSNR and SSIM values of the recovered videos are reported in Table 3. We can see that the recovered results

<sup>1</sup> <http://trace.eas.asu.edu/yuv/>.

**Fig. 12** The observed “Pompoms” with different sampling rates (Color figure online)



**Fig. 13** First row: the recovered results from observed data with  $p = 0.05$ . Second row: the recovered results from observed data with  $p = 0.1$ . Third row: the recovered results from observed data with  $p = 0.15$  (Color figure online)

**Table 4** The PSNR and SSIM values of the recovered MSIs by different methods on the MSI “Pompoms” with different sampling rates

Model	$p = 0.05$			$p = 0.1$			$p = 0.15$		
	PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time
TCTF	18.98	0.316	35.72	24.09	0.534	40.26	26.26	0.636	40.41
TNN	23.30	0.559	63.26	28.14	0.755	63.63	31.34	0.847	62.26
LRTC-TV-I	19.12	0.614	329.68	26.28	0.819	322.35	29.92	0.888	319.80
LRTC-TV-II	21.23	0.656	431.63	24.96	0.788	434.23	27.60	0.855	431.58
TCTF-TVT	<b>27.01</b>	<b>0.783</b>	153.84	<b>32.42</b>	<b>0.888</b>	152.48	<b>33.06</b>	<b>0.894</b>	153.23

The best PSNR and SSIM values are highlighted in bold

by TCTF-TVT are superior over those by TCTF, TNN, and LRTC-TV-I both qualitatively and quantitatively. Moreover, we display one selected fiber along the temporal mode of the original video and the recovered videos in Fig. 10. We can observe that the selected fibers of the recovered results by TCTF, TNN, and LRTC-TV-I exhibit strong oscillation phenomenon, while the selected fiber of the recovered result by TCTF-TVT is close to the true fiber.

**Example 3** We evaluate the performance of different methods on multispectral images (MSIs). For the MSI of  $n_b$  bands with each band of size  $n_v \times n_h$ , we set  $n_1 = n_v$ ,  $n_2 = n_h$ ,  $n_3 = n_b$ ,  $r = 30$ ,  $m_1 = m_2 = q_1 = q_2 = 1$ . Actually, not only in this paper, but also in the literature (see, e.g., [40]), MSI data are considered as a third-order tensor with the first two dimensions representing spatial dimensions and the third dimension representing band. The tested MSI “Pompoms”<sup>2</sup> is cited from [40], which has 31 bands with each band of size  $256 \times 256$ . A composition of 1st, 2nd and 31th bands of the ground truth, observed data and the recovered MSIs is displayed in Figs. 11, 12 and 13, respectively. The PSNR and SSIM values of the recovered MSIs are reported in Table 4. We can conclude from the figure and table that the recovered results by TCTF-TVT are superior over those by TCTF, TNN, LRTC-TV-I, and LRTC-TV-II both qualitatively and quantitatively.

## 6 Concluding Remark

A new tensor completion model based on the tensor factorization with TV–Tikhonov regularization has been proposed. The global convergence of the PAM algorithm for solving the proposed model has been established. We have also proposed a fast implementation for the PAM algorithm. Numerical results reported have shown the better performance of the proposed method compared with the state-of-the-art methods.

## References

- Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality. *Math. Oper. Res.* **35**(2), 438–457 (2010)
- Attouch, H., Bolte, J., Svaiter, B.-F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Math. Program.* **137**(1–2), 91–129 (2013)
- Auslender, A.: Méthodes numériques pour la décomposition et la minimisation de fonctions non différentiables. *Numer. Math.* **18**(3), 213–223 (1971)
- Auslender, A.: Asymptotic properties of the Fenchel dual functional and applications to decomposition problems. *J. Optim. Theory Appl.* **73**(3), 427–449 (1992)
- Bertsekas, D.P., Tsitsiklis, J.-N.: Parallel and Distributed Computation: Numerical Methods. Prentice Hall, Englewood Cliffs (1989)
- Bochnak, J., Coste, M., Roy, M.-F.: Real Algebraic Geometry. Springer, Berlin (2013)
- Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* **17**(4), 1205–1223 (2007a)
- Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. *SIAM J. Optim.* **18**(2), 556–572 (2007b)
- Cai, X.: Variational image segmentation model coupled with image restoration achievements. *Pattern Recognit.* **48**(6), 2029–2042 (2015)
- Carroll, J.-D., Pruzansky, S., Kruskal, J.-B.: CANDECOMP: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika* **45**(1), 3–24 (1980)
- Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011)
- Combettes, P.-L., Pesquet, J.-C.: Proximal splitting methods in signal processing. In: Bauschke, H., Burachik, R., Combettes, P., Elser, V., Luke, D., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer, New York (2011)
- Combettes, P.-L., Wajs, V.-R.: Signal recovery by proximal forward–backward splitting. *Multiscale Model. Simul.* **4**(4), 1168–1200 (2005)
- Coste, M.: An Introduction to O-Minimal Geometry. Istituti editoriali e poligrafici internazionali, Pisa (2000)
- Esser, E., Zhang, X., Chan, T.-F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.* **3**(4), 1015–1046 (2010)
- Gabay, D., Mercier, B.: A Dual Algorithm for the Solution of Non Linear Variational Problems Via Finite Element Approximation. Institut de recherche d'informatique et d'automatique, Rocquencourt (1975)
- Gandy, S., Recht, B., Yamada, I.: Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **27**(2), 025010 (2011)
- Goldstein, T., Osher, S.: The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.* **2**(2), 323–343 (2009)
- Hansen, P.-C., Nagy, J. G., O’leary, D.-P.: Deblurring Images: Matrices, Spectra, and Filtering. SIAM, Philadelphia (2006)
- Hong, M., Luo, Z.-Q.: On the linear convergence of the alternating direction method of multipliers. *Math. Program.* **162**(1–2), 165–199 (2017)
- Jain, P., Oh, S.: Provable tensor factorization with missing data. In: Proceedings of the International Conference on Neural Information Processing Systems (2014)
- Ji, T.-Y., Huang, T.-Z., Zhao, X.-L., Ma, T.-H., Liu, G.: Tensor completion using total variation and low-rank matrix factorization. *Inf. Sci.* **326**, 243–257 (2016)
- Ji, T.-Y., Huang, T.-Z., Zhao, X.-L., Ma, T.-H., Deng, L.-J.: A non-convex tensor rank approximation for tensor completion. *Appl. Math. Model.* **48**, 410–422 (2017)
- Kilmer, M., Martin, C.: Factorization strategies for third-order tensors. *Linear Algebra Appl.* **435**(3), 641–658 (2011)
- Kilmer, M.-E., Braman, K., Hao, N., Hoover, R.-C.: Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.* **34**(1), 148–172 (2013)
- Kolda, T.-G., Bader, B.-W.: Tensor decompositions and applications. *SIAM Rev.* **51**, 455–500 (2009)

<sup>2</sup> <http://www.cs.columbia.edu/CAVE/databases/multispectral/>.

27. Kressner, D., Steinlechner, M., Vandereycken, B.: Low-rank tensor completion by Riemannian optimization. *BIT Numer. Math.* **54**(2), 447–468 (2014)
28. Kurdyka, K.: On gradients of functions definable in o-minimal structures. *Annales de l'institut Fourier* **48**, 769–783 (1998)
29. Li, X., Ye, Y., Xu, X.: Low-rank tensor completion with total variation for visual data inpainting. In: Proceedings of Thirty-First AAAI Conference on Artificial Intelligence (2017)
30. Li, X.-T., Zhao, X.-L., Jiang, T.X., Zheng, Y.B., Ji, T.Y., Huang, T.Z.: Low-rank tensor completion via combined non-local self-similarity and low-rank regularization. *Neurocomputing* **367**, 1–12 (2019)
31. Liu, J., Musalski, P., Wonka, P., Ye, J.: Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(1), 208–220 (2013)
32. Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., Yan, S.: 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)
33. Ortega, J.-M., Rheinboldt, W.-C.: Iterative Solution of Nonlinear Equations in Several Variables. SIAM, Philadelphia (1970)
34. Rudin, L.-I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**(1–4), 259–268 (1992)
35. Semerci, O., Hao, N., Kilmer, M.-E., Miller, E.-L.: Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Trans. Image Process.* **23**(4), 1678–1693 (2014)
36. Tan, H., Cheng, B., Wang, W., Zhang, Y.-J., Ran, B.: Tensor completion via a multi-linear low-n-rank factorization model. *Neurocomputing* **133**, 161–169 (2014)
37. Tao, D., Li, X., Wu, X., Maybank, S.-J.: General tensor discriminant analysis and gabor features for gait recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(10), 1700–1715 (2007)
38. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109**(3), 475–494 (2001)
39. Tucker, L.: Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**(3), 279–311 (1966)
40. Xie, Q., Zhao, Q., Meng, D., Xu, Z.: Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(8), 1888–1902 (2018)
41. Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.* **6**(3), 1758–1789 (2013)
42. Xu, W.-H., Zhao, X.-L., Ji, T.-Y., Miao, J.-Q., Ma, T.-H., Wang, S., Huang, T.-Z.: Laplace function based nonconvex surrogate for low-rank tensor completion. *Signal Process. Image Commun.* **73**, 62–69 (2019)
43. Yang, J.-H., Zhao, X.-L., Ma, T.-H., Chen, Y., Huang, T.-Z., Ding, M.: Remote sensing images despeckling using unidirectional hybrid total variation and nonconvex low-rank regularization. *J. Comput. Appl. Math.* **363**, 124–144 (2020)
44. Zhang, Z.: A novel algebraic framework for processing multidimensional data: theory and application. Ph.D. dissertation, Tufts University (2017)
45. Zhang, Z., Aeron, S.: Exact tensor completion using t-SVD. *IEEE Trans. Signal Process.* **65**(6), 1511–1526 (2017)
46. Zhang, Z., Ely, G., Aeron, S., Hao, N., Kilmer, M.: 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)
47. Zhao, Xi-Le, Wang, Fan, Michael, K.Ng: A new convex optimization model for multiplicative noise and blur removal. *SIAM J. Imaging Sci.* **7**(1), 456–475 (2014)
48. Zheng, Y.-B., Huang, T.-Z., Zhao, X.-L., Jiang, T.-X., Ma, T.-H., Ji, T.-Y.: Mixed noise removal in hyperspectral image via low-fibered-rank regularization. *IEEE Trans. Geosci. Remote Sens.* (2019). <https://doi.org/10.1109/TGRS.2019.2940534>
49. Zhou, P., Feng, J.: Outlier-robust tensor PCA. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
50. Zhou, M., Liu, Y., Long, Z., Chen, L., Zhu, C.: Tensor rank learning in CP decomposition via convolutional neural network. *Signal Process. Image Commun.* **73**, 12–21 (2018a)
51. Zhou, P., Lu, C., Lin, Z., Zhang, C.: Tensor factorization for low-rank tensor completion. *IEEE Trans. Image Process.* **27**(3), 1152–1163 (2018b)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Xue-Lei Lin received the BS degree in Information and Computing Science from School of Mathematics and Statistics, Ningxia University, Ningxia, China. He received the MS degree in Mathematics from the Department of Mathematics, University of Macau, Macau, China. He is currently pursuing the PhD degree in the Department of Mathematics, Faculty of Science, Hong Kong Baptist University, Hong Kong, China. His research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies. His website is <http://scholar.google.com/citations?user=BqyuMZ8AAAAJ&hl=en>.



Michael K. Ng is the Director of Research Division for Mathematical and Statistical Science, and Chair Professor of Department of Mathematics, the University of Hong Kong. His research areas are data science, scientific computing, and numerical linear algebra. His website is <https://hkumath.hku.hk/~mng/>.



**Xi-Le Zhao** received the MS and PhD degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012. He is currently a professor with the School of Mathematical Sciences, UESTC. His research interests include image processing, computer vision, and machine learning. His website is <https://sites.google.com/site/xilezhao2018/home>.