International Journal of Computer Vision (2024) 132:1721–1747
https://doi.org/10.1007/s11263-023-01960-1

ORIGINAL PAPER

Convex–Concave Tensor Robust Principal Component Analysis

Youfa Liu[1,2] · Bo Du[1,2] · Yongyong Chen[3] · Lefei Zhang[1,2] · Mingming Gong[4] · Dacheng Tao[5]

Received: 2 September 2022 / Accepted: 13 November 2023 / Published online: 21 December 2023

Abstract

Tensor robust principal component analysis (TRPCA) aims at recovering the underlying low-rank clean tensor and residual sparse component from the observed tensor. The recovery quality heavily depends on the definition of tensor rank which has diverse construction schemes. Recently, tensor average rank has been proposed and the tensor nuclear norm has been proven to be its best convex surrogate. Many improved works based on the tensor nuclear norm have emerged rapidly. Nevertheless, there exist three common drawbacks: (1) the neglect of consideration on relativity between the distribution of large singular values and low-rank constraint; (2) the prior assumption of equal treatment for frontal slices hidden in tensor nuclear norm; (3) the missing convergence of whole iteration sequences in optimization. To address these problems together, in this paper, we propose a convex–concave TRPCA method in which the notion of convex–convex singular value separation (CCSVS) plays a dominant role in the objective. It can adjust the distribution of the first several largest singular values with low-rank controlling in a relative way and emphasize the importance of frontal slices collaboratively. Remarkably, we provide the rigorous convergence analysis of whole iteration sequences in optimization. Besides, a low-rank tensor recovery guarantee is established for the proposed CCSVS model. Extensive experiments demonstrate that the proposed CCSVS significantly outperforms state-of-the-art methods over toy data and real-world datasets, and running time per image is also the fastest.

Keywords Low-rank tensor recovery · Tensor nuclear norm · Convex–concave singular value separation · Fourier transformation · Convergence

1 Introduction

Principal component analysis (PCA) is one of the most popular unsupervised methods to deal with high-dimensional data, such as color images and videos. It can discover the low-rank components from high-dimensional data with small corruption. PCA has been used in many scenarios including data preprocessing (Belarbi et al., 2017) and image recovery (Bouwmans et al., 2018). PCA is also extended as KPCA (Schölkopf et al., 1997), MPCA (Lu et al., 2008) and 2D-PCA (Yu & Bennamoun, 2006) for specific applications.

However, PCA can not handle large noise corruption well and easily suffers from outliers in realistic applications. To alleviate this problem, robust PCA (RPCA) (Candès et al., 2011) is developed and is modeled by the convex program-

Communicated by Xavier Pennec.

✉ Bo Du
remoteking@whu.edu.cn

✉ Yongyong Chen
YongyongChen.cn@gmail.com

Youfa Liu
liuyoufa@mail.hzau.edu.cn

Lefei Zhang
zhanglefei@whu.edu.cn

Mingming Gong
gongmingnju@gmail.com

Dacheng Tao
dacheng.tao@sydney.edu.au

[1] School of Computer Science, Wuhan University, Bayi Road, Wuhan 430072, Hubei Province, China

[2] College of Informatics, Huazhong Agricultural University, Shizishan Street, Wuhan 430070, Hubei, China

[3] School of Computer Science, Harbin Institute of Technology (Shenzhen), Taoyuan Street, Shenzhen 518055, China

[4] School of Mathematics and Statistics, The University of Melbourne, Grattan Street, Parkville, VIC 3010, Australia

[5] School of Information Technologies and Faculty of Engineering and Information Technologies, University of Sydney, Sydney, NSW 2007, Australia

ming

$$\min_{\mathbf{X},\mathbf{E}} \|\mathbf{X}\|_* + \lambda\|\mathbf{E}\|_1$$
$$\text{s.t. } \mathbf{Y} = \mathbf{X} + \mathbf{E},$$

where $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$ is the observed matrix, $\mathbf{X}$ is the low-rank component, $\mathbf{E}$ is the sparse component, $\|\mathbf{X}\|_*$ denotes the nuclear norm of the matrix $\mathbf{X}$ and $\|\mathbf{E}\|_1$ denotes the l_1-norm of the matrix $\mathbf{E}$. The parameter $\lambda = \frac{1}{\sqrt{\max\{n_1,n_2\}}}$ is the theoretically suggested value in Candès et al. (2011). The nuclear norm and l_1-norm are the ideal convex surrogate of rank function and l_0-norm, respectively, which also brings the convenience of optimization because the original minimization problem is usually NP-hard. RPCA and its variants have been widely used in video denoising (Bouwmans et al., 2018), subspace clustering (Liu et al., 2013), and so on.

Despite the success of the RPCA, it can not analyze the tensor data perfectly. For example, when we apply RPCA to color images, the preprocessing of transforming them into gray-level images is required. The preprocessing can unfold the tensor data into matrix data but simultaneously destroys the intrinsic structure in the tensor. To remedy this shortage, tensor RPCA has been developed with the spirit of showing respect for the tensor-structural information. Tensor RPCA assumes that the observed tensor could be decomposed as the sum of a low-rank component and a sparse one. The aim is to recover the clean low-rank tensor. Existing works, such as Lu et al. (2020) and Imaizumi and Maehara (2017), typically explore the following framework

$$\min_{\mathcal{X},\mathcal{E}} r(\mathcal{X}) + \lambda_1\|\mathcal{E}\|_1$$
$$\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{E},$$

where $r(\mathcal{X})$ denotes a function with respect to a kind of tensor rank of $\mathcal{X}$ and $\lambda > 0$ is a trade-off parameter. There are many kinds of candidate tensor rank, such as CP rank (Kolda & Bader, 2009), Tucker rank (Malik & Becker, 2018), tensor tubal rank (Zhang et al., 2014), tensor average rank (Lu et al., 2020) and tensor train rank (Imaizumi & Maehara, 2017). For a better treatment of this framework, one concerns about its relaxed version

$$\min_{\mathcal{X},\mathcal{E}} g_r(\mathcal{X}) + \lambda_1\|\mathcal{E}\|_1$$
$$\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{E},$$

where $g_r(\mathcal{X})$ denotes the some ideal convex surrogate of the rank function $r(\mathcal{X})$.

For example, the tensor rank is chosen as the tensor average rank and the proposed tensor nuclear norm is set as the convex surrogate of the tensor average rank in TRPCA (Lu et al., 2020). Additionally, $\lambda = \frac{1}{\sqrt{\max\{n_1,n_2\}n_3}}$ is the theoretically suggested value in TRPCA (Lu et al., 2020). Subsequently, TPSSV (Zhang & Peng, 2019) and TRPCA-WTNN (Sun et al., 2020) make different improvements on the part of tensor nuclear norm respectively. Specifically, the former places the weights on different singular values to emphasize the different importance while the latter focuses on shrinking the k smallest singular values. Nevertheless, TRPCA (Lu et al., 2020) and TPSSV (Zhang & Peng, 2019) exhibit three common drawbacks. First, they just concern the low-rank property but neglect the ground-truth distribution of the first several largest eigenvalues, which are closely related to the important information of tensor, for example, the salient part of color images. Second, they equally treat frontal slices of tensor, which is not reasonable because the information level may be different among these slices (see Fig. 5 in experiments for intuition.). Third, they obtain the subsequence convergence for optimization, which is not satisfying because the quality of solution may be not ideal and lead to bad performance of models.

To address the aforementioned drawbacks, the convex–concave singular value separation (CCSVS) based tensor robust principal component analysis is proposed in this paper. CCSVS devotes to segregating all singular values into two parts, adjusting the distribution of the first several largest eigenvalues towards the ground truth, and treating the slices unequally in a collaborative manner. In this way, the tensor rank approximation, distribution of the first several largest singular values and the importance of different slices are simultaneously treated. The optimization is realized by the Augmented Lagrangian Multipliers (ALM) ((Lu et al., 2018)) methods in a non-convex setting, which just guarantees the convergence of subsequences. For the proposed CCSVS model, the convergence analysis of whole iteration sequences, instead of subsequences, is theoretically proved. We also demonstrate that the limit point of convergent sequences must be stationary points. It is worth mentioning that there exists a closed-form solution to the subproblem of solving the proximal operator with respect to the convex–concave singular value disparity of separation, which makes the optimization faster. Besides, we establish the recovery guarantee for the proposed CCSVS model.

In summary, the main contributions of this paper are as follows.

- A convex–convex TRPCA method is proposed, in which the notion of convex–concave singular value separation (CCSVS) is introduced. It treats the large and small singular values in an adversarial manner such that low-rank property and distribution of large singular values can be manipulated collaboratively.
- The proximal operator with respect to CCSVS is one of the crucial subproblems in the optimization procedure and provides an efficient and elegant analytic expression.

- We prove the convergence of whole iteration sequences of the convex–concave objective optimization. It is the first time to get the convergent result of whole sequences, instead of subsequences, by non-convex ALM method in the research of tensor RPCA problem when compared with previous works. Furthermore, we prove that the limit point pair satisfies the generalized (Karush–Kuhn–Tucker) KKT condition.
- We provide the theoretical recovery guarantee for the proposed CCSVS model, which shows that the underlying low-rank tensor can be recovered with high probability.

The remainder of this paper is organized as follows. In Sect. 2, several related works are briefly reviewed. In Sect. 3, necessary preliminaries are provided. In Sect. 4, the methodology is formulated. In Sect. 5, the solution of the proximal operator is discussed. The optimization procedure is followed in Sect. 6. In Sect. 7, the rigorous convergence analysis of the whole iteration sequences is done. We establish a recovery theorem for the proposed model in Sect. 8. In Sect. 9, extensive experiments on synthetic data, color images, and videos are conducted on simulation data and real data. Finally, we conclude this paper and provide some discussion about the proposed model in Sect. 10. Some lengthy proofs are deferred to the Appendix.

2 Related Works

(1) *TRPCA* (Lu et al., 2020): Lu et al. (2020) proposed a novel tenor rank called tensor average rank which is weaker than tensor tubal rank (Zhang et al., 2014). Further, they found its convex surrogate and derived the TRPCA model:

$$\begin{aligned} &\min_{\mathcal{X},\mathcal{E}} \|\mathcal{X}\|_{\odot} + \lambda_1 \|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{E}, \end{aligned} \tag{1}$$

where

$$\|\mathcal{X}\|_{\odot} = \langle \mathcal{S}, \mathcal{I} \rangle, \tag{2}$$

is the tensor nuclear norm, in which $\mathcal{S}$ comes from the t-SVD $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$ of $\mathcal{X}$. This model recovers the clean $\mathcal{X}$ from observation $\mathcal{Y}$ with a low tensor average rank.

(2) *TPSSV* (Zhang & Peng, 2019): The aforementioned TRPCA model (Lu et al., 2020) shrinks all singular values with the same parameter in solving the tensor nuclear norm minimization. Nevertheless, this way may not approximate the tensor average rank well in practice. To alleviate this drawback, TPSSV (Zhang & Peng, 2019) shrinks the smallest r singular values, which results in the 'cut-off' version of TRPCA as follows.

$$\begin{aligned} &\min_{\mathcal{X},\mathcal{E}} \|\mathcal{X}\|_{p=r} + \lambda_1 \|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{E}, \end{aligned} \tag{3}$$

where

$$\|\mathcal{X}\|_{p=r} = \sum_{i=1}^{n_3} \sum_{j=r+1}^{\min\{n_1,n_2\}} \sigma_j(\overline{\mathcal{X}}^{(i)}), \tag{4}$$

in which r is a controlling parameter which varies from 0 to $\min\{n_1, n_2\} - 1$.

(3) *TRPCA-WTNN:* TRPCA-WTNN (Sun et al., 2020) extends the TRPCA model (Lu et al., 2020) to weighted form as follows.

$$\begin{aligned} &\min_{\mathcal{X},\mathcal{E}} \|\mathcal{X}\|_{\mathbf{w},\odot} + \lambda_1 \|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{E}, \end{aligned} \tag{5}$$

where

$$\|\mathcal{X}\|_{\odot} = \sum_{i=1}^{n_3} \sum_{j=1}^{\min\{n_1,n_2\}} w_j \sigma_j(\overline{\mathcal{X}}^{(i)}), \tag{6}$$

in which $w = \{w_1, \cdots, w_{\min\{n_1,n_2\}}\}$, $\sum_{i=1}^{\min\{n_1,n_2\}} w_i = 1$ and $w_i \geq 0$.

(4) *Other works:* BTRTF (Zhou & Cheung, 2021) aims to establish the model

$$\mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{E}, \tag{7}$$

where $\mathcal{X}$ is the low-rank component, $\mathcal{S}$ is the sparse noisy component and $\mathcal{E}$ is error term. The entries of $\mathcal{E}$ are assumed as a group of i.i.d. Gaussian random variables. When $\mathcal{Y}$ is known, $\mathcal{X}$ and $\mathcal{S}$ in this decomposition model could be recovered in some way. BTRTF (Zhou & Cheung, 2021) is committed to recovering the low-tubal-rank (Zhang et al., 2014) structure of $\mathcal{X}$ by the t-product factorization. Then Bayesian framework is developed and a variational inference strategy is adopted.

Recently, tensor train rank (Imaizumi & Maehara, 2017) has gained much attention in the research community. It stems from a well-balanced matricization scheme. Let $l = 3$. Given an l-th order tensor $\mathcal{X}$, its tensor train rank is

$$rank_{tt}(\mathcal{X}) = (rank(\mathbf{X}_{[1]}), rank(\mathbf{X}_{[2]}), \cdots, rank(\mathbf{X}_{[l-1]})),$$

where

$$\mathbf{X}_{[i]} \in \mathbb{R}^{\prod_{k=1}^{i} n_k \times \prod_{k=i+1}^{l} n_k},$$

which is obtained by matricizing the first i modes and the remaining $k-i$ modes.

The tensor train rank has proven to be powerful in discovering and leveraging the latent correlations among different modes of a given tensor. Considering the weighted nuclear norm as the convex surrogate of tensor train rank, Imaizumi and Maehara (2017) proposed the TTNN-based tensor robust principal component analysis model:

$$\begin{aligned} &\min_{\mathcal{X},\mathcal{E}} \|\mathcal{X}\|_{\mathbf{w},TTNN} + \lambda_1 \|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{E}, \end{aligned} \tag{8}$$

where

$$\|\mathcal{X}\|_{\mathbf{w},TTNN} = \sum_{i=1}^{l-1} w_i \|\mathbf{X}_{[i]}\|_* \tag{9}$$

with $\mathbf{w} = (w_1, \cdots, w_{l-1})$, $\sum_{i=1}^{l-1} w_i = 1$ and $w_i \geq 0$ for $i \in [l-1]$.

3 Preliminaries

3.1 Notations

For a 3-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the i-th frontal slice $\mathcal{X}(:,:,i)$ is represented by $\mathcal{X}^{(i)}$, $i = 1, \cdots, n_3$. The tubes are the vectors like the form of $\mathcal{X}(i,j,:)$. The l_1-norm of $\mathcal{X}$ is $\|\mathcal{X}\|_1 = \sum_{i,j,k} |x_{ijk}|$. The Frobenius norm of $\mathcal{X}$ is $\|\mathcal{X}\|_F = \sqrt{\sum_{i,j,k} x_{ijk}^2}$. The infinity norm of $\mathcal{X}$ is $\|\mathcal{X}\|_\infty = \max_{i,j,k} |x_{ijk}|$. The inner product of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i=1}^{n_3} \langle \mathcal{X}^{(i)}, \mathcal{Y}^{(i)} \rangle$. The complex conjugate of $\mathcal{X}$ is denoted by $conj(\mathcal{X})$. The conjugate transpose of a complex tensor $\mathcal{X}_c \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is written as $\mathcal{X}_c^*$. For any $x \in \mathbb{R}$, x_+ denotes the positive part of x, i.e. $x_+ = \max\{x, 0\}$. For a matrix $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$, $\sigma_j(\mathbf{X})$ denotes its j-th largest singular value. $\#A$ denotes the cardinality of a set A. For any square matrix $\mathbf{T}_0$, $tr(\mathbf{T}_0)$ denotes its trace. We use $[m]$ to denote the set $\{1, 2, \cdots, m\}$.

The Discrete Fourier Transformation (DFT) of a tensor $\mathcal{Q} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ along the third-dimension is the result of performing DFT over all tubes $\{\mathcal{Q}(i,j,:)\}_{i \in [n_1], j \in [n_2]}$ and can be easily implemented by the MATLAB command

$$\overline{\mathcal{Q}} = fft(\mathcal{Q}, [\,], 3).$$

The underlying tensor $\mathcal{Q}$ can be recovered by inverse DFT as

$$\mathcal{Q} = ifft(\overline{\mathcal{Q}}, [\,], 3).$$

Make use of the convention from Kilmer and Martin (2011) that

$$\overline{\mathbf{Q}} = \text{bdiag}(\overline{\mathcal{Q}}) = \begin{bmatrix} \overline{\mathcal{Q}}^{(1)} & & & \\ & \overline{\mathcal{Q}}^{(2)} & & \\ & & \ddots & \\ & & & \overline{\mathcal{Q}}^{(n_3)} \end{bmatrix}. \tag{10}$$

The block circulant matrix (Lu et al., 2020) of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is

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

3.2 T-product and t-SVD

The '*unfold*' operator and its inverse operator '*fold*' (Kilmer & Martin, 2011) are prescribed by

$$unfold(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} \\ \mathcal{X}^{(2)} \\ \vdots \\ \mathcal{X}^{(n_3)} \end{bmatrix}$$

and

$$fold(unfold(\mathcal{X})) = \mathcal{X}$$

respectively.

We review the concepts of t-product from Kilmer and Martin (2011) and t-SVD from Lu et al. (2020), which extend the traditional matrix multiplication and singular value decomposition respectively.

Definition 1 (T-Product (Kilmer & Martin, 2011)) Let $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Z} \in \mathbb{R}^{n_2 \times l \times n_3}$. The t-product of $\mathcal{Y}$ and $\mathcal{Z}$ is a tensor of size $n_1 \times l \times n_3$

$$\mathcal{Y} * \mathcal{Z} = fold(bcirc(\mathcal{Y}) \cdot unfold(\mathcal{Z})). \tag{11}$$

Proposition 1 (Lu et al., 2020) *$\mathcal{C} = \mathcal{Y} * \mathcal{Z}$ is equivalent to $\overline{\mathbf{C}} = \overline{\mathbf{Y}} \cdot \overline{\mathbf{Z}}$.*

Definition 2 (F-Diagonal Tensor (Kilmer & Martin, 2011)) A tensor is called an f-diagonal tensor if each frontal slice is a diagonal matrix.

Definition 3 (T-SVD (Lu et al., 2020)) The t-SVD of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is like

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*, \tag{12}$$

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 orthogonal tensors under the t-product and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor.

When $n_3 = 1$, the t-SVD is actually the same as the singular value decomposition of matrices.

3.3 Tensor Tubal Rank and Tensor Average Rank

For a 3-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor tubal rank (Zhang et al., 2014) $rank_t(\mathcal{X})$ is expressed by

$$rank_t(\mathcal{X}) = \#\{i : \mathcal{S}(i, i, :) \neq \mathbf{0}\},$$

where $\mathcal{S}$ is an f-diagonal tensor coming from the t-SVD of $\mathcal{X}$, i.e. $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$.

The tensor average rank (Lu et al., 2020) of $\mathcal{X}$ is given by

$$rank_a(\mathcal{X}) = \frac{1}{n_3} rank(bcirc(\mathcal{X})).$$

As a matter of fact, the tensor average rank is less than or equal to the tensor tubal rank, i.e. $rank_a(\mathcal{X}) \leq rank_t(\mathcal{X})$. Therefore, the low tensor tubal rank must imply the low tensor average rank. Furthermore, the low average rank assumption is implied by the usual CP rank (Kolda & Bader, 2009) and Tucker rank (Malik & Becker, 2018).

4 Methodology

4.1 Motivation

The motivation for our work comes from the following three observations.

The first observation is that the clean (target) tensor which we want to recover typically bears with separation phenomenon of singular values. For example, the statistics of singular values for a color image are exhibited in Fig. 1 as follows. Large and small singular values exist a relatively clear dividing line in Fig. 1b. This phenomenon motivates us to explore the separation of singular values in the research of low-rank tensor recovery problems.

The second observation is that the distribution of large singular values has a significant impact on the fidelity of the recovered tensor. For example, the saturation and luminance of some areas (e.g. antlers, legs and weeds) of the recovered image with the classical TRPCA method (Lu et al., 2020) are actually distorted in Fig. 2c, which large singular values obey the uniformly distributed law. Although the noise is removed by TRPCA, the recovered image is not what we expect when compared with the original image. This motivates us to take the distribution of large singular values into our research.

The third observation is that the information in slices of a clean tensor may be unequally distributed. For example, the singular values with respect to the red channel are usually larger than the other two channels in color images in Fig. 1. This prior knowledge of color channels motivates us to attach importance to the slice weights in practice.

4.2 Convex–Concave Singular Value Disparity of Separation

Firstly, the convex–convex singular value separation is proposed in the matrix case. Then this notion is extended to the tensor case later.

Definition 4 (Convex–Concave Singular Value Disparity of Separation for Matrix) For a (latent) matrix $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$ with unknown rank, the convex–concave Ω-disparity of separation for $\mathbf{X}$ is

$$\text{Sep}(\mathbf{X}, \Omega) = \omega \sum_{i=r+1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}) - \sum_{j=1}^{r} \omega_j \sigma_j(\mathbf{X}), \tag{13}$$

where the separation parameter r satisfies the condition of $1 \leq r < \min\{n_1, n_2\}$ and the weight vector $\Omega = (\omega, \omega_1, \cdots, \omega_r)$ in which $\omega \in \mathbb{R}_+$ and $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_r \geq 0$.

The convex–concave meaning arises from the following structural theorem.

Theorem 1 *The weight vector $\Omega_r = (\omega_1, \cdots, \omega_r) \in \mathbb{R}^r_+$ satisfies the monotonic decreasing property $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_r$. Let $\omega \in \mathbb{R}_+$. Then, for any given $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$, the measurement* $\text{Sep}(\mathbf{X}, \Omega)$ *can be reformulated as the structural expression*

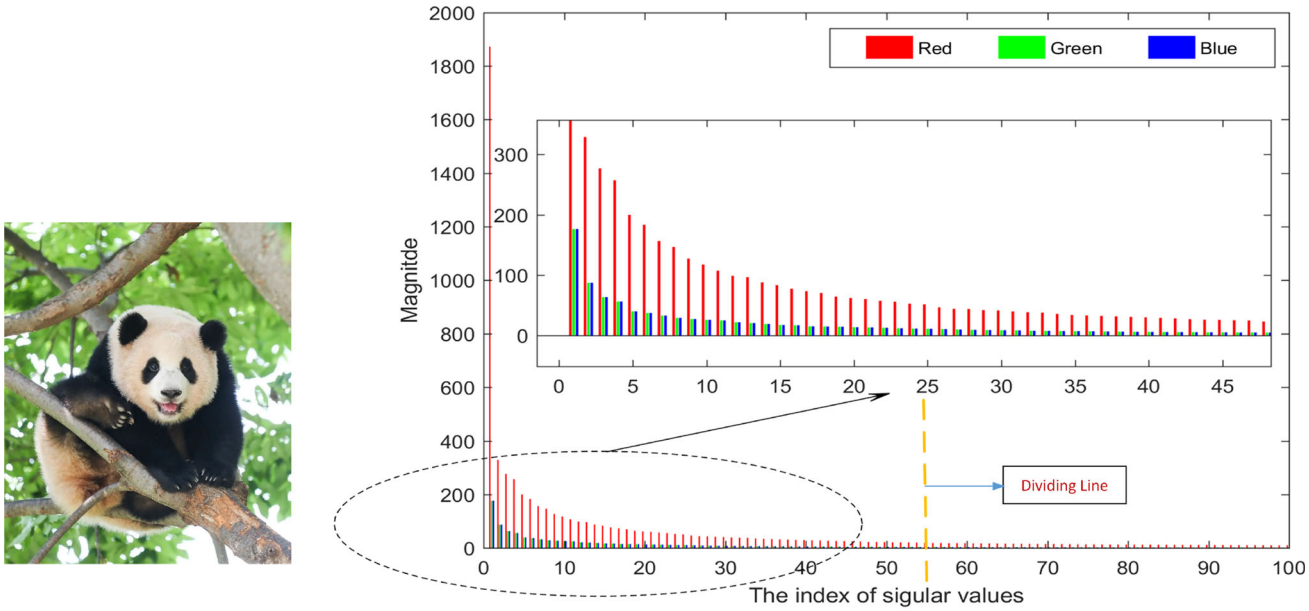
$$\text{Sep}(\mathbf{X}, \Omega) = \underbrace{\omega \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X})}_{f_{convex}(\mathbf{X})} + \underbrace{(-\sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j(\mathbf{X}))}_{g_{concave}(\mathbf{X})}.$$

where $\widetilde{\omega}_j = \omega_j + \omega$ for $j = 1, 2, \cdots, r$.

The proof is deferred to Sect. 11.2 in the Appendix.

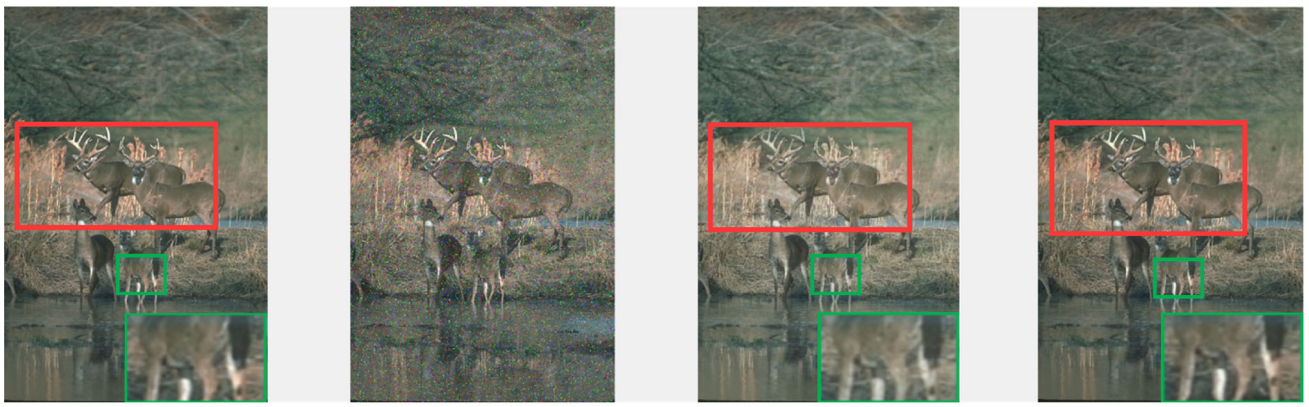
Remark 1 This theorem indicates that the quantity $\text{Sep}(\mathbf{X}, \Omega)$ bears with the difference-of-convex (DC) structure or convex–concave structure. This property leads to great difficulty in establishing the convergence analysis and the tensor recovery theory.

Remark 2 $f_{convex}(\mathbf{X})$ is indeed equal to $\omega\|\mathbf{X}\|_*$. It is well-known that the nuclear norm is the convex envelope of the function $rank(\mathbf{X})$ over the unit ball in the sense of the spectral norm (Lu et al., 2020). Hence the $f_{convex}(\cdot)$ is closely related to the underlying rank.



(a) Image (b) The distribution of singular values

Fig. 1 The illustration on the distribution of singular values for image 1 along three channels (i.e. red channel, green channel, and blue channel). For simplicity, we only exhibit the first 100 largest singular values. The dashed line represents a dividing line between large and small singular values (Color figure online)



(a) Original image (b) Noisy image (c) Recovered image by TRPCA [6] (d) Recovered image by our method

Fig. 2 The recovery comparison between TRPCA (Lu et al., 2020) and our method (preview). The main difference lies in the rectangular area within red and green bounding boxes. For TRPCA, the large singular values are treated uniformly, which leads to the worse fidelity of the recovered image (Color figure online)

Remark 3 This structural expression of $\mathrm{Sep}(\mathbf{X}, \Omega)$ plays an important role in the proof of recovery theorem in Sect. 8.

Motivated by the disparity of separation for the matrix version and the tensor average rank from (Lu et al., 2020), we propose the following disparity of separation for tensor, which is a crucial building block for the proposed model later.

Definition 5 (**Convex–Concave Singular Value Disparity of Separation for Tensor**) For a (latent) tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with unknown rank, the convex–concave (Ω, ϑ)-disparity of separation for $\mathcal{X}$ is

$$\mathrm{Sep}(\mathcal{X}, \Omega, \vartheta) = \sum_{k=1}^{n_3} \vartheta_k \mathrm{Sep}(\overline{\mathcal{X}}^{(k)}, \Omega), \tag{14}$$

where $\vartheta = (\vartheta_1, \vartheta_2, \cdots, \vartheta_{n_3}) \in \mathbb{R}_+^{n_3}$ and $\overline{\mathcal{X}}^{(k)} \in \mathbb{C}^{n_1 \times n_2}$ denotes the Discrete Fourier Transformation (DFT) of the k-th slice of $\mathcal{X}$.

The ϑ denotes the weight vector along the third dimension of tensor $\mathcal{X}$, which indicates the importance of different slices of $\mathcal{X}$. The Ω simultaneously controls the rank approximation and adjusts the distribution of the first several singular values of $\mathcal{X}$ towards the ground truth.

As a matter of fact, the consideration of introducing the weight vector ϑ is two-fold. The first consideration is motivated by a lot of empirical studies on the distribution of singular values along different slices in clean tensors. For example, the singular values are diversely distributed along different color channels of color images (see Fig. 1b for intuition.). The second consideration is motivated by Proposition 1 (in next page) which reveals that the proposed tensor nuclear norm in TRPCA (Lu et al., 2020) indeed conceals a prior assumption that the distribution of singular values along all frontal slices are equal, which is followed by other works, such as TPSSV (Zhang & Peng, 2019), TRPCA-WTNN (Sun et al., 2020) and ETRPCA (Gao et al., 2020).

4.3 The Proposed Model

Recently, a series of works on tensor average rank (Lu et al., 2020) have been done, such as TRPCA (Lu et al., 2020), TPSSV (Zhang & Peng, 2019), TRPCA-WTNN (Sun et al., 2020). Their corresponding models are formulated by Eqs. (1), (3) and (5), respectively. Although these methods make impressive progress in tensor robust PCA, there are three common drawbacks among them. Firstly, they are inflexible for approximating the tensor rank. Secondly, they only pay attention to the low-rank property and ignore the ground truth distribution of the first several largest eigenvalues, which represent the main information of tensor and hence may help to guide the tensor recovery in practice. For example, this distribution has an affinity with the salient part of the color images (see Fig. 2 for intuition.). Thirdly, they equally treat all the frontal slices of a tensor.

To address these drawbacks uniformly, the tensor robust principal component analysis model with convex–concave singular value separation (CCSVS) is proposed, which is formulated as

$$\begin{aligned} &\min_{\mathcal{X},\mathcal{E}} \text{Sep}(\mathcal{X}, \Omega, \vartheta) + \lambda\|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{E}, \end{aligned} \tag{15}$$

where $\mathcal{Y} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is the observed tensor, $\mathcal{X}$ is the low-rank component, $\mathcal{E}$ is the sparse noise, $\text{Sep}(\mathcal{X}, \Omega, \vartheta)$ is the convex–concave singular value (Ω, ϑ)-disparity of separation (i.e. Definition 5.) and $\lambda > 0$ is the trade-off parameter. In Eq. (15), ϑ controls the importance of each (frontal) slice. Ω simultaneously adjusts the quality of tensor rank approximation and the distribution of the first r largest singular values towards the ground truth, which makes a contribution to effective separation.

The proposed model Eq. (15) is significantly different from TRPCA (Lu et al., 2020). This can be revealed by the following proposition on the tensor nuclear norm.

Proposition 2 ***(Reformulation of Tensor Nuclear Norm)*** *For a 3-way tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, its tensor nuclear norm $\|\mathcal{X}\|_* = \langle \mathcal{S}, \mathcal{I}\rangle$ (i.e. Equation (2)) in Lu et al. (2020) is equivalent to*

$$\|\mathcal{X}\|_* = \sum_{i=1}^{n_3} \frac{1}{n_3} \left(\sum_{j=1}^{\min\{n_1,n_2\}} \sigma_j(\overline{\mathcal{X}}^{(i)}) \right). \tag{16}$$

Proof Recall from Lu et al. (2020) that

$$\|\mathcal{X}\|_* = \frac{1}{n_3}\|\overline{\mathbf{X}}\|_*.$$

By Eq. (10) and the meaning of matrix nuclear norm, we have

$$\begin{aligned} \|\mathcal{X}\|_* &= \frac{1}{n_3}\|bdiag(\overline{\mathcal{X}})\|_* \\ &= \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{\min\{n_1,n_2\}} \sigma_j(\overline{\mathcal{X}}^{(i)}) \\ &= \sum_{i=1}^{n_3} \frac{1}{n_3} \left(\sum_{j=1}^{\min\{n_1,n_2\}} \sigma_j(\overline{\mathcal{X}}^{(i)}) \right). \end{aligned}$$

□

From this proposition, it is readily seen that the distribution of singular values along all frontal slices bears the equal weight $\frac{1}{n_3}$. In other words, the tensor nuclear norm in TRPCA (Lu et al., 2020) signifies that all frontal slices are equal in prior with respect to the distribution of singular values. The proposed model Eq. (15) breaks this (universal) prior assumption and behaves better in practice.

5 Proximal Operator

If we choose the optimization method of Augmented Lagrange Multipliers (ALM) in nonconvex setting (Lu et al., 2018), the solving of Eq. (15) includes the subproblem of the proximal operator with respect to the function $\text{Sep}(\mathcal{X}, \Omega, \vartheta)$. In this section, it is proven to have a nice closed-form solution, which brings the hope of a fast solution to Eq. (15). The case of matrix is given first. Then it is reused in solving the tensor version.

5.1 The Matrix Version

The proximal operator of a given matrix $\mathbf{Q} \in \mathbb{C}^{n_1 \times n_2}$ with respect to $\mathrm{Sep}(\mathbf{X}, \Omega)$ is prescribed as

$$\begin{aligned} &\mathrm{prox}_{\gamma \mathrm{Sep}(\mathbf{X},\Omega)}(\mathbf{Q}) \\ &= \mathrm{argmin}_{\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}} \mathrm{Sep}(\mathbf{X}, \Omega) + \frac{1}{2\gamma}\|\mathbf{X} - \mathbf{Q}\|_F^2, \end{aligned} \tag{17}$$

where $\gamma > 0$ is a trade-off parameter.

Theorem 2 *If the list $\{\omega_i\}_{1 \le i \le r}$ satisfies the non-increasing property, namely,*

$$\omega_1 \ge \omega_2 \ge \cdots \ge \omega_r \ge 0,$$

the proximal operator Eq. (17) can be analytically represented by

$$\mathrm{prox}_{\gamma \mathrm{Sep}(\mathbf{X},\Omega)}(\mathbf{Q}) = \mathbf{U}_\mathbf{Q} S_{\gamma,\Omega}(\mathbf{Q}) \mathbf{V}_\mathbf{Q}^*, \tag{18}$$

where $\mathbf{Q} = \mathbf{U}_\mathbf{Q} \mathbf{\Sigma}_\mathbf{Q} \mathbf{V}_\mathbf{Q}^*$ is the singular value decomposition of $\mathbf{Q}$ and

$$\begin{aligned} S_{\gamma,\Omega}(\mathbf{Q}) = diag(&\{\sigma_i(\mathbf{Q}) + \gamma\omega_i\}_{1 \le i \le r}, \\ &\{(\sigma_j(\mathbf{Q}) - \gamma\omega)_+\}_{r+1 \le j \le \min\{n_1,n_2\}}), \end{aligned} \tag{19}$$

in which x_+ denotes the positive part of x, namely, $x_+ = \max\{x, 0\}$.

Proof Let $H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X})$ be the objective in Eq. (17). Note that

$$\mathrm{argmin}_{\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}} H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) = \mathrm{argmin}_{\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}} \gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}).$$

Hence we focus on $\gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X})$. With simple algebraic manipulation, we have

$$\begin{aligned} &\gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) \\ &= \gamma \mathrm{Sep}(\mathbf{X}, \Omega) + \frac{1}{2}\|\mathbf{X} - \mathbf{Q}\|_F^2 \\ &= \gamma\omega \sum_{i=r+1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}) - \sum_{i=1}^{r} (\gamma\omega_i)\sigma_i(\mathbf{X}) \\ &+ \frac{1}{2} \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i^2(\mathbf{X}) + \frac{1}{2} \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i^2(\mathbf{Q}) - \mathrm{Re}(\langle \mathbf{X}, \mathbf{Q} \rangle), \end{aligned} \tag{20}$$

where $\mathrm{Re}(\langle \mathbf{X}, \mathbf{Q} \rangle)$ is the real part of the complex number $\langle \mathbf{X}, \mathbf{Q} \rangle$.

In view of the classical von Neumann trace inequality for complex matrices (Von Neumann, 1937; Mirsky, 1975),

$$|tr(\mathbf{X}^T \mathbf{Q})| = |\langle \mathbf{X}, \mathbf{Q} \rangle| \le \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Q}), \tag{21}$$

where the equality is achieved if and only if there exist orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$ such that $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}_\mathbf{X}\mathbf{V}^*$ and $\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}_\mathbf{Q}\mathbf{V}^*$ are singular value decompositions of $\mathbf{X}$ and $\mathbf{Q}$, respectively. In other words, for the equality case, we have

$$|\langle \mathbf{X}, \mathbf{Q} \rangle| = \mathrm{Re}(\langle \mathbf{X}, \mathbf{Q} \rangle) = \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Q}). \tag{22}$$

Furthermore, since $\mathbf{Q} = \mathbf{U}_\mathbf{Q}\mathbf{\Sigma}_\mathbf{Q}\mathbf{V}_\mathbf{Q}^*$, there must be $\mathbf{U} = \mathbf{U}_\mathbf{Q}$ and $\mathbf{V} = \mathbf{V}_\mathbf{Q}$.

By the form of singular value decomposition of $\mathbf{X}$, we divide the whole space $\mathbb{C}^{n_1 \times n_2}$ into two subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$. Specifically,

$$\begin{aligned} \mathcal{S}_1 = \{\mathbf{X} \in \mathbb{C}^{n_1 \times n_2} : &\mathbf{X} = \mathbf{U}\mathbf{\Sigma}_\mathbf{X}\mathbf{V}^*, \text{where } \mathbf{U} \\ &= \mathbf{U}_\mathbf{Q} \text{ and } \mathbf{V} = \mathbf{V}_\mathbf{Q}.\} \end{aligned} \tag{23}$$

and

$$\begin{aligned} \mathcal{S}_2 = \{\mathbf{X} \in \mathbb{C}^{n_1 \times n_2} : &\mathbf{X} = \mathbf{U}\mathbf{\Sigma}_\mathbf{X}\mathbf{V}^*, \text{where} \\ &\mathbf{U} \ne \mathbf{U}_\mathbf{Q} \text{ or } \mathbf{V} \ne \mathbf{V}_\mathbf{Q}.\}. \end{aligned} \tag{24}$$

Then $\mathbb{C}^{n_1 \times n_2} = \mathcal{S}_1 \cup \mathcal{S}_2$.

For $\mathbf{X} \in \mathcal{S}_2 - \mathcal{S}_1$ (i.e. $\mathbf{X} \in \mathcal{S}_2$ but $\mathbf{X} \notin \mathcal{S}_1$), let $\mathbf{X} = \mathbf{U}_\mathbf{X}\mathbf{\Sigma}_\mathbf{X}\mathbf{V}_\mathbf{X}^*$ be its singular value decomposition. We construct a new matrix $\widetilde{\mathbf{X}} = \mathbf{U}_\mathbf{Q}\mathbf{\Sigma}_\mathbf{X}\mathbf{V}_\mathbf{Q}^*$. Obviously, $\widetilde{\mathbf{X}} \in \mathcal{S}_1$. We claim that, for $\mathbf{X} \in \mathcal{S}_2 - \mathcal{S}_1$,

$$H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) > H_{\gamma,\Omega,\mathbf{Q}}(\widetilde{\mathbf{X}}). \tag{25}$$

Since $\mathbf{X} \in \mathcal{S}_2 - \mathcal{S}_1$, by Eq. (20) and von Neumann trace inequality Eq. (21), we have

$$\begin{aligned} \gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) > \gamma\omega &\sum_{i=r+1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}) - \sum_{i=1}^{r} (\gamma\omega_i)\sigma_i(\mathbf{X}) \\ &+ \frac{1}{2} \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i^2(\mathbf{X}) + \frac{1}{2} \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i^2(\mathbf{Q}) \\ &- \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Q}), \end{aligned} \tag{26}$$

where the case of equality in the Neumann trace inequality is ruled out because $\mathbf{X} \in \mathcal{S}_2 - \mathcal{S}_1$ implies $\mathbf{X} \notin \mathcal{S}_1$. Note that $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ have the same singular values. Hence the right-hand side of Eq. (26) coincides with

$$\begin{aligned} \gamma\omega &\sum_{i=r+1}^{\min\{n_1,n_2\}} \sigma_i(\widetilde{\mathbf{X}}) - \sum_{i=1}^{r} (\gamma\omega_i)\sigma_i(\widetilde{\mathbf{X}}) \\ &+ \frac{1}{2} \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i^2(\widetilde{\mathbf{X}}) + \frac{1}{2} \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i^2(\mathbf{Q}) \end{aligned}$$

$$- \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\widetilde{\mathbf{X}})\sigma_i(\mathbf{Q}). \tag{27}$$

By the equality part of von Neumann trace inequality, this equation becomes $\gamma H_{\gamma,\Omega,\mathbf{Q}}(\widetilde{\mathbf{X}})$. Combing Eq. (26) and Eq. (27), we conclude that Eq. (25) holds. Then, for any $\mathbf{X} \in \mathcal{S}_2 - \mathcal{S}_1$, we can always find a point $\widetilde{\mathbf{X}}$ in $\mathcal{S}_1$ such that $H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) > H_{\gamma,\Omega,\mathbf{Q}}(\widetilde{\mathbf{X}})$. Therefore,

$$\begin{aligned}\operatorname{argmin}_{\mathbf{X}\in\mathbb{C}^{n_1\times n_2}} \gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) &= \operatorname{argmin}_{\mathbf{X}\in\mathcal{S}_1\cup\mathcal{S}_2} \gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) \\ &= \operatorname{argmin}_{\mathbf{X}\in\mathcal{S}_1} \gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}).\end{aligned} \tag{28}$$

In other words, the minimizers of $H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X})$ must fall into the space of $\mathcal{S}_1$.

Note that

$$\begin{aligned}&\frac{1}{2}\sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i^2(\mathbf{X}) + \frac{1}{2}\sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i^2(\mathbf{Q}) \\ &- \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Q}) = \frac{1}{2}\sum_{i=1}^{\min\{n_1,n_2\}} (\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{Q}))^2.\end{aligned} \tag{29}$$

Combining this fact with Eq. (28), we have

$$\begin{aligned}&\operatorname{argmin}_{\mathbf{X}\in\mathbb{C}^{n_1\times n_2}} \gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) \\ &= \arg\min_{\mathbf{X}\in\mathcal{S}_1} \gamma\omega \sum_{i=r+1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}) - \sum_{i=1}^{r}(\gamma\omega_i)\sigma_i(\mathbf{X}) \\ &\quad + \frac{1}{2}\sum_{i=1}^{\min\{n_1,n_2\}} (\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{Q}))^2.\end{aligned} \tag{30}$$

Equivalently,

$$\begin{aligned}&\arg\min_{\mathbf{X}\in\mathcal{S}_1} \gamma H_{\gamma,\Omega,\mathbf{Q}}(\mathbf{X}) \\ &= \arg\min_{\{\sigma_i(\mathbf{X})\}_{i\geq 1}} \frac{1}{2}\sum_{i=r+1}^{\min\{n_1,n_2\}} (\sigma_i(\mathbf{X}) - (\sigma_i(\mathbf{Q}) - \gamma\omega))^2 \\ &\quad + \frac{1}{2}\sum_{i=1}^{r}(\sigma_i(\mathbf{X}) - (\sigma_i(\mathbf{Q}) + \gamma\omega_i))^2 \\ &\text{s.t.} \quad \sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq \cdots \geq \sigma_{\min\{n_1,n_2\}}(\mathbf{X}) \geq 0.\end{aligned} \tag{31}$$

Hence, recalling the non-increasing condition $\omega_1 \geq \omega_2 \cdots \geq \omega_r \geq 0$, we have

$$\sigma_i(\mathbf{X}) = \begin{cases}\sigma_i(\mathbf{Q}) + \gamma\omega_i, & if \quad i = 1, 2, \cdots, r; \\ (\sigma_i(\mathbf{Q}) - \gamma\omega)_+, & if \quad i \geq r+1.\end{cases} \tag{32}$$

It easily checked that the constraint condition Eq. (31) is exactly satisfied.

Finally, Eq. (18) is achieved. □

5.2 The Tensor Version

The proximal operator of a given tensor $\mathcal{Q} \in \mathbb{R}^{n_1\times n_2\times n_3}$ with respect to $\mathrm{Sep}(\mathcal{X}, \Omega, \vartheta)$ is prescribed as

$$\begin{aligned}&\mathrm{prox}_{\gamma\mathrm{Sep}(\mathcal{X},\Omega,\vartheta)}(\mathcal{Q}) \\ &= \arg\min_{\mathcal{X}} \mathrm{Sep}(\mathcal{X}, \Omega, \vartheta) + \frac{1}{2\gamma}\|\mathcal{X} - \mathcal{Q}\|_F^2.\end{aligned} \tag{33}$$

Theorem 3 *If the list $\{\omega_i\}_{1\leq i\leq r}$ satisfies the non-increasing property, namely,*

$$\omega_1 \geq \omega_2 \geq \cdots \geq \omega_r \geq 0,$$

the proximal operator of Eq. (33) is solved by

$$\mathrm{prox}_{\gamma\mathrm{Sep}(\mathcal{X},\Omega,\vartheta)}(\mathcal{Q}) = \mathcal{U} * ifft(\mathcal{S}_{\gamma,\Omega,\vartheta}(\overline{\mathcal{Q}})) * \mathcal{V}^*, \tag{34}$$

where $\mathcal{Q} = \mathcal{U} * \mathcal{S}_{\mathcal{Q}} * \mathcal{V}^*$ is the t-SVD of $\mathcal{Q}$ and

$$\mathcal{S}_{\gamma,\Omega,\vartheta}(\overline{\mathcal{Q}}) = fold(\{S_{\gamma n_3\vartheta_i,\Omega}(\overline{\mathcal{Q}}^{(i)})\}_{i=1}^{n_3}). \tag{35}$$

Proof Note that the Frobenius norm of a tensor $\mathcal{R} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and its Discrete Fourier Transformation (DFT) $\overline{\mathcal{R}}$ satisfies $\|\mathcal{R}\|_F = \frac{1}{\sqrt{n_3}}\|\overline{\mathcal{R}}\|_F$ (Lu et al., 2020). With this in mind, we have

$$\begin{aligned}&\mathrm{prox}_{\gamma\mathrm{Sep}(\mathcal{X},\Omega,\vartheta)}(\mathcal{Q}) \\ &= \arg\min_{\mathcal{X}} \mathrm{Sep}(\mathcal{X}, \Omega, \vartheta) + \frac{1}{2\gamma n_3}\|\overline{\mathcal{X}} - \overline{\mathcal{Q}}\|_F^2 \\ &= \arg\min_{\mathcal{X}} \sum_{i=1}^{n_3}\left(\vartheta_i \mathrm{Sep}(\mathcal{X}^{(i)}, \Omega) + \frac{1}{2\gamma n_3}\|\overline{\mathcal{X}}^{(i)} - \overline{\mathcal{Q}}^{(i)}\|_F^2\right).\end{aligned} \tag{36}$$

It is obvious that the objective of the last term in Eq. (36) is separable along slices. By means of Theorem 2,

$$\overline{\mathrm{prox}_{\gamma\mathrm{Sep}(\mathcal{X},\Omega,\vartheta)}(\mathcal{Q})}^{(i)} = \overline{\mathcal{U}}^{(i)} S_{\gamma n_3\vartheta_i,\Omega}(\overline{\mathcal{Q}}^{(i)})(\overline{\mathcal{V}}^{(i)})^*,$$

for $i \in [n_3]$. Recall that $\mathcal{S}_{\gamma n_3,\Omega,\vartheta}(\overline{\mathcal{Q}})$ is formed by folding $\{S_{\gamma n_3\vartheta_i,\Omega}(\overline{\mathcal{Q}}^{(i)})\}_{i=1}^{n_3}$. By inverse DFT, we have

$$\mathrm{prox}_{\gamma\mathrm{Sep}(\mathcal{X},\Omega,\vartheta)}(\mathcal{Q}) = \mathcal{U} * ifft(\mathcal{S}_{\gamma n_3,\Omega,\vartheta}(\overline{\mathcal{Q}})) * \mathcal{V}^*.$$

□

6 Optimization Procedure

Although the objective in Eq. (15) bears with the convex–concave (hence non-convex) structure, the Augmented Lagrange Multipliers (ALM) method for non-convex setting (Wang et al., 2019) is considered here. For the proposed convex–concave model Eq. (15), the convergence analysis of whole iteration sequences, instead of subsequences, is provided in the next section. In particular, we prove that the limit point pair of convergence sequences is actually a stationary point for the proposed CCSVS model. The objective of the ALM method is

$$\Phi(\mathcal{X},\mathcal{E},\mathcal{L},\mu)=\mathrm{Sep}(\mathcal{X},\Omega,\vartheta)+\lambda\|\mathcal{E}\|_1+\langle\mathcal{L},\mathcal{Y}-\mathcal{X}-\mathcal{E}\rangle+\frac{\mu}{2}\|\mathcal{Y}-\mathcal{X}-\mathcal{E}\|_F^2, \tag{37}$$

where $\mathcal{L}$ is the Lagrangian multiplier and $\mu>0$ is the penalty factor. The optimization involves the following two main subproblems.

(1) $\mathcal{X}$-subproblem: Update $\mathcal{X}$ and fix the residual variables, minimizing the Φ is equivalent to

$$\min_{\mathcal{X}}\frac{1}{\mu^{(k)}}\mathrm{Sep}(\mathcal{X},\Omega,\vartheta)+\frac{1}{2}\|\mathcal{X}-\mathcal{N}^{(k)}\|_F^2, \tag{38}$$

where $\mathcal{N}^{(k)}=\mathcal{Y}+\frac{1}{\mu^{(k)}}\mathcal{L}^{(k)}-\mathcal{E}^{(k)}$. By Theorem 5, the optimal solution can be solved by

$$\mathcal{X}^{(k+1)}=\mathrm{prox}_{\frac{1}{\mu^{(k)}}\mathrm{Sep}(\mathcal{X},\Omega,\vartheta)}(\mathcal{N}^{(k)}). \tag{39}$$

(2) $\mathcal{E}$-subproblem: Update $\mathcal{E}$ and fix the residual variables, minimizing the Φ is equivalent to

$$\min_{\mathcal{E}}\frac{\lambda}{\mu^{(k)}}\|\mathcal{E}\|_1+\frac{1}{2}\|\mathcal{E}-\mathcal{J}^{(k)}\|_F^2, \tag{40}$$

where $\mathcal{J}^{(k)}=\mathcal{Y}+\frac{1}{\mu^{(k)}}\mathcal{L}^{(k)}-\mathcal{X}^{(k+1)}$. The optimal solution can be expressed by soft-thresholding operator (Wang et al., 2008) as

$$\mathcal{E}^{(k+1)}=\mathcal{T}_{\frac{\lambda}{\mu^{(k)}}}(\mathcal{J}^{(k)}), \tag{41}$$

where

The whole optimization procedure is given in Algorithm 1.

Algorithm 1 CCSVS

1: The observation data $\mathcal{Y}$, the separation parameter r, the weights Ω with non-increasing $\{\omega_i\}_{i=1}^r$ and ϑ.
2: $\mathcal{X}$ and $\mathcal{E}$.
3: Initialize $\mu_0=10^{-4}$, $\rho=1.1$, $k=0$, ϑ_0, Ω_0, $\mu_{max}=10^{10}$, $\epsilon=10^{-8}$, and $\mathcal{X}^{(k)}=\mathcal{E}^{(k)}=\mathcal{L}^{(k)}=0$.
4: **while** not convergent **do**
5: Update $\mathcal{X}$ by Eq. (39).
6: Update $\mathcal{E}$ by Eq. (41).
7: Update $\mathcal{L}$ by $\mathcal{L}^{(k+1)}=\mathcal{L}^{(k)}+\mu^{(k)}(\mathcal{Y}-\mathcal{X}^{(k+1)}-\mathcal{E}^{(k+1)})$.
8: $\mu^{(k+1)}=\rho\mu^{(k)}$ (In practice, $\mu^{(k+1)}$ is often set as $\min\{\rho\mu^{(k)},\mu_{max}\}$.).
9: Verify the stop conditions

$$\begin{cases}\|\mathcal{X}^{(k+1)}-\mathcal{X}^{(k)}\|_\infty\le\epsilon,\\ \|\mathcal{E}^{(k+1)}-\mathcal{E}^{(k)}\|_\infty\le\epsilon,\\ \|\mathcal{Y}-\mathcal{X}^{(k+1)}-\mathcal{E}^{(k+1)}\|_\infty\le\epsilon.\end{cases}$$

10: **end while**

7 The Convergence Analysis

In this section, the convergence of the whole iteration sequences, instead of subsequences, is generated by Algorithm 1. Above of all, the boundedness of the sequence $\{\mathcal{L}^{(k)}\}_{k\ge1}$ is achieved.

Theorem 4 *The sequence $\{\mathcal{L}^{(k)}\}_{k\ge1}$ generated by the Algorithm 1 is bounded.*

Proof By the updated formula of the Lagrangian multiplier in Algorithm 1, we have

$$\begin{aligned}\|\mathcal{L}^{(k+1)}\|_F&=\|\mathcal{L}^{(k)}+\mu^{(k)}(\mathcal{Y}-\mathcal{X}^{(k+1)}-\mathcal{E}^{(k+1)})\|_F\\&=\mu^{(k)}\|(\mathcal{Y}+\frac{1}{\mu^{(k)}}\mathcal{L}^{(k)}-\mathcal{E}^{(k+1)})-\mathcal{X}^{(k+1)}\|_F\\&=\mu^{(k)}\|\mathcal{N}^{(k)}-\mathcal{X}^{(k+1)}\|_F\\&=\frac{\mu^{(k)}}{\sqrt{n_3}}\|\overline{\mathcal{N}}^{(k)}-\overline{\mathcal{X}}^{(k+1)}\|_F\\&=\frac{\mu^{(k)}}{\sqrt{n_3}}\|\overline{\mathbf{N}}^{(k)}-\overline{\mathbf{X}}^{(k+1)}\|_F.\end{aligned} \tag{42}$$

In view of Eq. (39),

$$\mathcal{X}^{(k+1)}=\mathcal{U}^{(k)}*ifft(\mathcal{S}_{\frac{1}{\mu^{(k)}},\Omega,\vartheta}(\overline{\mathcal{N}}^{(k)}))*(\mathcal{V}^{(k)})^* \tag{43}$$

and $\mathcal{N}^{(k)}=\mathcal{U}^{(k)}*\mathcal{F}^{(k)}*(\mathcal{V}^{(k)})^*$ is the t-SVD of $\mathcal{N}^{(k)}$. Then

$$\begin{aligned}&\|\mathcal{L}^{(k+1)}\|_F\\&=\frac{\mu^{(k)}}{\sqrt{n_3}}\|\overline{\mathbf{F}}^{(k)}-\mathrm{bdiag}(\mathcal{S}_{\frac{1}{\mu^{(k)}},\Omega,\vartheta}(\overline{\mathcal{N}}^{(k)}))\|_F\\&\le\frac{\mu^{(k)}}{\sqrt{n_3}}\sqrt{\sum_{i=1}^{n_3}[(\min\{n_1,n_2\}-r)(\frac{n_3\vartheta_i\omega}{\mu^{(k)}})^2+\sum_{j=1}^{r}(\frac{n_3\vartheta_i\omega_j}{\mu^{(k)}})^2]}\\&=\sqrt{n_3\sum_{i=1}^{n_3}(\vartheta^{(i)})^2\cdot[(\min\{n_1,n_2\}-r)\omega^2+\sum_{j=1}^{r}\omega_j^2]}.\end{aligned} \tag{44}$$

As a consequence, the sequence $\{\mathcal{L}^{(k)}\}_{k\geq 1}$ is bounded. □

Theorem 5 *If the weights $\{\omega_i\}_{1\leq i\leq r}$ satisfy the non-increasing property, namely,*

$$\omega_1 \geq \omega_2 \geq \cdots \geq \omega_r \geq 0,$$

the sequences $\{\mathcal{E}^{(k)}\}_{k\geq 1}$ and $\{\mathcal{X}^{(k)}\}_{k\geq 1}$ generated by the Algorithm 1 satisfy

$$\lim_{k\to\infty} \|\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{E}^{(k+1)}\|_F = 0,$$
$$\lim_{k\to\infty} \|\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}\|_F = 0,$$
$$\lim_{k\to\infty} \|\mathcal{E}^{(k+1)} - \mathcal{E}^{(k)}\|_F = 0.$$

Proof By Theorem 4, $\|\mathcal{L}^{(k)}\|_F \leq C_0$, where C_0 is taken as the right hand side of Eq. (44). By Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&\|\mathcal{L}^{(k)} - \mathcal{L}^{(k-1)}\|_F^2 \\
&= \|\mathcal{L}^{(k)}\|_F^2 - 2\langle \mathcal{L}^{(k)}, \mathcal{L}^{(k-1)}\rangle + \|\mathcal{L}^{(k-1)}\|_F^2 \\
&\leq \|\mathcal{L}^{(k)}\|_F^2 + 2\|\mathcal{L}^{(k)}\|_F\|\mathcal{L}^{(k-1)}\|_F + \|\mathcal{L}^{(k-1)}\|_F^2 \\
&\leq 4C_0^2. \qquad (45)
\end{aligned}$$

It follows from Eq. (45) and the update formula of Lagrangian multiplier in Algorithm 1 that

$$\begin{aligned}
&\lim_{k\to\infty} \|\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{E}^{(k+1)}\|_F^2 \\
&= \lim_{k\to\infty} \frac{1}{\mu^{(k)}}\|\mathcal{L}^{(k+1)} - \mathcal{L}^{(k)}\|_F^2 \\
&\leq \lim_{k\to\infty} \frac{4C_0^2}{\mu^{(k)}} = 0.
\end{aligned}$$

In view of the $\mathcal{E}$-subproblem with solution Eq. (41), the update formula of Lagrangian multiplier and the boundedness of $\mathcal{L}^{(k)}$, we have

$$\begin{aligned}
&\lim_{k\to\infty} \|\mathcal{E}^{(k+1)} - \mathcal{E}^{(k)}\|_F \\
&\leq \lim_{k\to\infty} \|\mathcal{T}_{\frac{\lambda}{\mu^{(k)}}}(\mathcal{J}^{(k)}) - \mathcal{J}^{(k)} + \frac{1}{\mu^{(k)}}\mathcal{L}^{(k)} \\
&+ \frac{1}{\mu^{(k-1)}}(\mathcal{L}^{(k)} - \mathcal{L}^{(k-1)})\|_F \\
&\leq \lim_{k\to\infty} (\frac{\lambda n_1 n_2 n_3}{\mu^{(k)}} + \frac{\|\mathcal{L}^{(k)}\|_F}{\mu^{(k)}} + \frac{1}{\mu^{(k-1)}} \\
&\quad \|\mathcal{L}^{(k)} - \mathcal{L}^{(k-1)}\|_F) \\
&= 0. \qquad (46)
\end{aligned}$$

According to the update formula of $\mathcal{L}$, we know

$$\mathcal{X}^{(k+1)} = \mathcal{Y} + \frac{1}{\mu^{(k)}}(\mathcal{L}^{(k+1)} - \mathcal{L}^{(k)}) - \mathcal{E}^{(k+1)}. \qquad (47)$$

Then we deduce that

$$\begin{aligned}
&\lim_{k\to\infty} \|\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}\|_F^2 \\
&= \lim_{k\to\infty} \|\mathcal{Y} + \frac{1}{\mu^{(k)}}(\mathcal{L}^{(k+1)} - \mathcal{L}^{(k)}) - \mathcal{E}^{(k+1)} - \mathcal{X}^{(k)}\|_F^2 \\
&= \lim_{k\to\infty} \|\mathcal{Y} + \frac{1}{\mu^{(k)}}(\mathcal{L}^{(k+1)} - \mathcal{L}^{(k)}) - \mathcal{E}^{(k+1)} - \mathcal{X}^{(k)} \\
&+ (\mathcal{E}^{(k-1)} + \frac{1}{\mu^{(k-1)}}\mathcal{L}^{(k-1)}) - (\mathcal{E}^{(k-1)} + \frac{1}{\mu^{(k-1)}}\mathcal{L}^{(k-1)})\|_F^2 \\
&\leq \lim_{k\to\infty} (\|\mathcal{N}^{(k-1)} - \mathcal{X}^{(k)}\|_F + \|\mathcal{E}^{(k-1)} - \mathcal{E}^{(k)}\|_F \\
&\quad + \|\mathcal{E}^{(k)} - \mathcal{E}^{(k+1)}\|_F + \frac{1}{\mu^{(k)}}\|\mathcal{L}^{(k+1)} - \mathcal{L}^{(k)}\|_F \\
&+ \frac{1}{\mu^{(k-1)}}\|\mathcal{L}^{(k-1)}\|_F). \qquad (48)
\end{aligned}$$

It can be derived by combining Eq. (42) with Eq. (44) that

$$\begin{aligned}
&\|\mathcal{N}^{(k-1)} - \mathcal{X}^{(k)}\|_F \\
&\leq \frac{1}{\mu^{(k-1)}} \\
&\sqrt{n_3 \sum_{i=1}^{n_3} (\vartheta^{(i)})^2 \cdot [(\min\{n_1, n_2\} - r)\omega^2 + \sum_{j=1}^{r} \omega_j^2]}. \qquad (49)
\end{aligned}$$

According to the boundedness of the sequence $\{\mathcal{L}^{(k)}\}_{k\geq 1}$, Eq. (46), Eq. (48) and Eq. (49),

$$\lim_{k\to\infty} \|\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}\|_F^2 = 0.$$

□

Theorem 6 *The whole iteration sequences $\{\mathcal{X}^{(k)}\}_{k\geq 1}$ and $\{\mathcal{E}^{(k)}\}_{k\geq 1}$ generated by the Algorithm 1 are convergent. Let $\mathcal{X}^*$ and $\mathcal{E}^*$ be corresponding limit points and $\mathcal{L}^*$ be an accumulation point of $\{\mathcal{L}^{(k)}\}_{k\geq 1}$. Furthermore, the pair $(\mathcal{X}^*, \mathcal{E}^*)$ is a stationary point for the model Eq. (15), namely,*

$$\begin{cases} 0 \in \partial \mathrm{Sep}(\mathcal{X}^*, \Omega, \vartheta) - \mathcal{L}^*, \\ 0 \in \lambda \partial \|\mathcal{E}^*\|_1 - \mathcal{L}^*, \\ \mathcal{Y} = \mathcal{X}^* + \mathcal{E}^*, \end{cases} \qquad (50)$$

where the subdifferential refers to the limiting subdifferential in Rockafellar and Wets (1998) (See Definition 8 in the Appendix for details.).

Remark 4 Due to the convex–concave structure of $\mathrm{Sep}(\mathcal{X}^*, \Omega, \vartheta)$, the concept of subdifferential of convex functions can not be employed here. We need the more general definition of *limiting subdifferential* for non-convex and non-smooth functions in Rockafellar and Wets (1998) here (See Definition 6 in Sect. 11.3 for details). In this case, we call Eq. (50) the *generalized (Karush-Kuhn-Tucker) KKT condition* of the proposed CCSVS model Eq. (15).

Proof Due to the complexity of proving the Eq. (50), we just provide the convergence of whole sequence $\{\mathcal{X}^{(k)}\}_{k\geq 1}$ and $\{\mathcal{E}^{(k)}\}_{k\geq 1}$ here and the proof of Eq. (50) is deferred to the Sect. 11.3 in the Appendix. Note that the geometric series $\sum_{k=1}^{\infty}\frac{1}{\mu^{(k)}}$ is convergent because $\rho > 1$ in the Algorithm 1. As a consequence, for any $\varepsilon > 0$, there exists a positive integer K_0 such that, given any $k \geq K_0$,

$$\frac{1}{\mu^{(k+p-2)}} + \frac{1}{\mu^{(k+p-3)}} + \cdots + \frac{1}{\mu^{(k)}} + \frac{1}{\mu^{(k-1)}} \leq \varepsilon \tag{51}$$

holds for any positive integer p.

It can be intermediately derived from the proof of Theorem 5 that, for any $k \geq 1$,

$$\|\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}\|_F \leq \frac{C}{\mu^{(k-1)}}, \tag{52}$$

where C is a constant that is independent of k.

Then, for the aforementioned ε, we have

$$\begin{aligned}
&\|\mathcal{X}^{(k+p)} - \mathcal{X}^{(k)}\|_F \\
&\leq \|\mathcal{X}^{(k+p)} - \mathcal{X}^{(k+p-1)}\|_F + \|\mathcal{X}^{(k+p-1)} - \mathcal{X}^{(k+p-2)}\|_F \\
&+ \cdots + \|\mathcal{X}^{(k+1)} - \mathcal{X}^{(k)}\|_F \\
&\leq C\left(\frac{1}{\mu^{(k+p-2)}} + \frac{1}{\mu^{(k+p-3)}} + \cdots + \frac{1}{\mu^{(k)}} + \frac{1}{\mu^{(k-1)}}\right) \\
&\leq C\varepsilon,
\end{aligned} \tag{53}$$

where the first inequality holds by the triangular inequality, the second equality holds by Eq. (52) and the last inequality holds by Eq. (51). Hence $\{\mathcal{X}^{(k)}\}_{k\geq 1}$ is a Cauchy sequence. Since $(\mathbb{R}^{n_1\times n_2\times n_3}, \|\cdot\|_F)$ is a complete metric space, the Cauchy sequence must be a convergent sequence. In other words, $\{\mathcal{X}^{(k)}\}_{k\geq 1}$ is also convergent. With similar strategy, we can conclude that $\{\mathcal{E}^{(k)}\}_{k\geq 1}$ is convergent. Hence, $\lim_{k\to\infty}\mathcal{Y}^{(k)} = \mathcal{Y}^*$ and $\lim_{k\to\infty}\mathcal{E}^{(k)} = \mathcal{E}^*$. The proof of Eq. (50) is deferred to Sect. 11.3 in Appendix. □

It is worth mentioning that the nonconvex ALM scheme (Lu et al., 2018) only guarantees the convergence results of the subsequences in general. In other words, the boundedness of the whole iteration sequence is proved and hence it has at least one accumulation point. In the current setting, we go deeper into the rigorous convergence analysis in the sense that the convergence of whole iteration sequences is proved and the limit point of the convergent sequence is a stationary point. To this end, we first recall the concept of accumulation point from mathematics. A point is called an *accumulation point* of a given sequence if there exists a subsequence that converges to it. The convergence of the whole sequence means that, for sufficiently large iteration number, the current iterated points are close to some point. However, the convergence of some subsequences in the whole sequence cannot have this property. For sufficiently large iteration number, the current iterated points might be far away from some accumulation point of the whole sequence. In this sense, the convergence of the whole sequence will make the iteration algorithm become more reliable. The convergence of subsequences can not make sure that the iteration point of the current step is close to accumulation when the iteration number is sufficiently large. In other words, the output of the iteration algorithm may deviate from the stationary point of optimization heavily.

Theorem 6 reveals that the quality of output of the proposed Algorithm 1 is more reliable than previous works, such as Wang et al. (2021), Gu et al. (2017), Lu et al. (2020) and Gao et al. (2020), and makes a contribution to the performance stability of the learned model. In particular, we point out that the optimization method used in TRPCA may be not ideal. Specifically, Lu *et. al* mentioned that standard Alternating Direction Method of Multiplier (ADMM) (Lu et al., 2018) can be employed for solving the TRPCA model (Lu et al., 2020). Dating back to Lu et al. (2018), we find that there is no theorem indicating the convergence of iteration sequence (or subsequence). In Lu et al. (2018), the convergence is characterized by upper bounding the gap of object values and some additional terms.

8 Recovery Guarantee

We establish the recovery theorem in this section. Because the objective in Eq. (15) is convex–concave (hence nonconvex), we introduce the following (ε, δ)-asymptotic unique solution to characterize our recovery theorem.

Definition 6 (**(ε, δ)-Asymptotic Unique Solution**) $(\mathcal{X}_0, \mathcal{E}_0)$ is the (ε, δ)-unique local solution to the proposed CCSVS model Eq. (15) if it satisfies $\mathcal{Y} = \mathcal{X}_0 + \mathcal{E}_0$ and

$$\begin{aligned}
\mathrm{Sep}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) + \lambda\|\mathcal{E}_0 - \mathcal{H}\|_1 > \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) \\
+ \lambda\|\mathcal{E}_0\|_1 - \delta
\end{aligned} \tag{54}$$

for any $\mathcal{H} \in \mathrm{B}(0, \varepsilon) - \{0\}$ in the sense of tensor nuclear norm.

Remark 5 When $\delta = 0$, (ε, δ)-asymptotic unique solution becomes the unique local solution.

To get the recovery guarantee, the following tensor incoherence condition is necessary, which is an extension of the incoherence condition in the matrix recovery theory (De la Torre & Black, 2001).

Definition 7 (**Tensor Incoherence Condition** (Lu et al., 2020)) Given $\mathcal{X}_0 \in \mathbb{R}^{n_1\times n_2\times n_3}$ with $rank_t(\mathcal{L}_0) = r$, let $\mathcal{X}_0 = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$ be its skinny t-SVD expression, where

$\mathcal{U} \in \mathbb{R}^{n_1 \times r \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$ are orthogonal tensors in the sense of t-product, and $\mathcal{S}$ is f-diagonal. $\mathcal{X}_0$ is said to satisfy tensor incoherence conditions with the parameter μ if it holds that

$$\max_{i=1,2,\cdots,n_1} \|\mathcal{U}^* \otimes \mathbf{e}_i^{(1)}\|_F \leq \sqrt{\frac{\mu r}{n_1 n_3}}, \tag{55}$$

$$\max_{j=1,2,\cdots,n_2} \|\mathcal{V}^* \otimes \mathbf{e}_j^{(2)}\|_F \leq \sqrt{\frac{\mu r}{n_2 n_3}}, \tag{56}$$

$$\|\mathcal{U} * \mathcal{V}^*\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2 n_3^2}}, \tag{57}$$

where $\{\mathbf{e}_i^{(1)}\}_{i=1}^{n_1}$ and $\{\mathbf{e}_j^{(2)}\}_{j=1}^{n_2}$ are the standard orthonormal bases of $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$.

Now we establish the following recovery theorem for the proposed model Eq. (15).

Theorem 7 ***(Recovery Theorem)*** *Assume that $\mathcal{X}_0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ satisfies the tensor incoherence condition. Let $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a tensor of signs. Suppose that the support set $\mathbf{\Omega}$ of $\mathcal{E}_0$ is uniformly distributed over the sets of indices with cardinality m and that $sgn([\mathcal{E}_0]_{ijk}) = [\mathcal{M}]_{ijk}$ for all $(i, j, k) \in \mathbf{\Omega}$. Denote $\vartheta_{\max} = \max_{i \in [n_3]} \vartheta_i$ and $\omega_{\max} = \|\Omega\|_\infty$, where $\Omega = \{\omega, \omega_1, \cdots, \omega_r\}$ comes from Definition 4. Let $\delta = \varepsilon \left(3\vartheta_{\max}\omega_{\max} + \frac{1}{n_3\sqrt{\max\{n_1,n_2\}}}\right)$. Then there exit universal positive constants q_1 and q_2 such that with probability at least $1 - q_1(\max\{n_1 n_2\}n_3)^{-q_2}$ (over the choice of support of $\mathcal{E}_0$) $(\mathcal{X}_0, \mathcal{E}_0)$ is the (ε, δ)-asymptotic unique local solution to the proposed CCSVS model Eq. (15) with $\lambda = \frac{1}{\sqrt{\max\{n_1,n_2\}n_3}}$ provided that*

$$rank_t(\mathcal{X}_0) \leq \frac{C_1 \min\{n_1, n_2\}n_3}{\mu(\log(\max\{n_1, n_2\}n_3))^2}, \tag{58}$$

and

$$m \leq C_2 n_1 n_2 n_3, \tag{59}$$

where C_1 and C_2 are positive constants which only depend on $\vartheta_{\max}$ and $\omega_{\max}$.

The proof of Theorem 7 is deferred to Sect. 11.4 in Appendix.

9 Experiments

In this section, extensive experiments are done for verifying the efficiency of the proposed tensor robust principal component analysis method CCSVS. The selected seven comparison methods include RPCA (Candès et al., 2011), SNN (Liu et al., 2013), TRPCA (Lu et al., 2020), TPSSV (Zhang & Peng, 2019), TRPCA-WTNN (Sun et al., 2020), ETRPCA (Gao et al., 2020) and BTRTF (Zhou & Cheung, 2021). First, numerical experiments are conducted on synthetic data. Then the performance of the proposed CCSVS is tested on the tasks of color image recovery and background modeling of color videos. The code is available at https://github.com/Liuyouf/CCSVS.

9.1 Exact Recovery from Varying Fractions of Error

The recovery of low-rank and sparse components is testified on random tensor data under various error rate. We generate a tensor $\mathcal{G} = \mathcal{P} * \mathcal{Q}$ with low tubal rank rk by Theorem 4.1 in Lu et al. (2020), where $\mathcal{P} \in \mathbb{R}^{n \times rk \times n}$ and $\mathcal{Q} \in \mathbb{R}^{rk \times n \times n}$ are independently sampled from the normal distribution $\mathcal{N}(0, \frac{1}{n})$. It is readily seen that the tensor $\mathcal{G}$ belongs to $\mathbb{R}^{n \times n \times n}$, where n is selected from the set $\{100, 200\}$. Then, we generate a sparse tensor $\mathcal{E}(\|\mathcal{E}_0\|_0 = m)$ in which non-zero elements obey i.i.d. Bernoulli distribution ± 1. In experiments, two settings are taken into account. In the first setting, take $rk = rank_t(\mathcal{D}) = 0.1n$ and $m = \|\mathcal{E}\|_0 = 0.1n^3$. In the second setting, take $rk = rank_t(\mathcal{D}) = 0.1n$ and $m = \|\mathcal{E}\|_0 = 0.2n^3$.

The parameter setting of the proposed CCSVS is as follows. The separation parameter in the proposed CCSVS is set as $r = 2$ in experiments. Take the weights of frontal slices as $\vartheta_1 = \frac{0.55}{n}$ and $\vartheta_2 = \vartheta_3 = \cdots = \vartheta_n = \frac{1.55}{n}$. $\Omega = (\omega, \omega_1, \omega_2, \cdots, \omega_r)$ is specified by $\omega = 0.7$ and the decreasing arithmetic progression $\{\omega_i\}_{i=1}^r$ is generated by '`linspace(1e-05,1e-07,r)`' in MATLAB command, where r is the separation parameter in the proposed CCSVS. Note that r has been set as 2. λ in Eq. (15) is set as $\frac{1}{\sqrt{n^2}}$ by the suggested value in our recovery theorem (i.e. Theorem 7). Recovery results of the proposed CCSVS and seven selected comparison methods with different choices of n are reported in Table 1 and Table 2 respectively.

From Table 1 and Table 2, it can be seen that the proposed CCSVS achieves the best recovery precision. RPCA produces the false rank estimation. The reason is that it transforms a tensor into a matrix while the structural information is heavily destroyed. All the tensor based methods get the exact rank value. The recovery precision of TPSSV is worse than RPCA. The cause may be that the TPSSV neglects role of the first several largest eigenvalues which are closely related to some areas of the underlying tensor. The proposed CCSVS is advantageous over the TRPCA, TRPCA-WTNN, ETRPCA and BTRTF. The cause may be that the proposed CCSVS segregates the high and low singular values, adjusts the distribution of the first several largest singular values towards the ground truth and emphasizes the different importance of different frontal slices in a collaborative manner.

Table 1 Recovery results of different choices of n ($rk = rank_t(\mathcal{G}) = 0.1n$, $m = \|\mathcal{E}_0\|_0 = 0.1n^3$)

n	rk	m	Methods	$rank_t(\widehat{\mathcal{G}})$	$\|\widehat{\mathcal{E}}\|_0$	$\frac{\|\widehat{\mathcal{G}}-\mathcal{G}\|_F}{\|\mathcal{G}\|_F}$	$\frac{\|\widehat{\mathcal{E}}-\mathcal{E}\|_F}{\|\mathcal{E}\|_F}$
100	10	1e5	RPCA (Candès et al., 2011)	100	563273	4.75e−01	6.92e−03
			SNN (Liu et al., 2013)	10	237250	2.31e−02	9.14e−5
			TRPCA (Lu et al., 2020)	10	104328	2.31e−07	9.14e−10
			TPSSV (Zhang & Peng, 2019)	10	862738	3.35e+02	3.47e−01
			TRPCA-WTNN (Sun et al., 2020)	10	103325	2.01e−07	8.52e−10
			ETRPCA (Gao et al., 2020)	10	100000	1.98e−07	7.88e−10
			BTRTF (Zhou & Cheung, 2021)	10	100000	1.84e−07	7.14e−10
			CCSVS	10	100000	5.08e−08	5.81e−11
200	20	8e5	RPCA (Candès et al., 2011)	200	4723452	7.48e−01	3.84e−03
			SNN (Liu et al., 2013)	10	256720	2.28e−02	8.97e−5
			TRPCA (Lu et al., 2020)	20	820137	5.42e−07	1.12e−09
			TPSSV (Zhang & Peng, 2019)	10	104328	2.31e−07	9.14e−10
			TRPCA-WTNN (Sun et al., 2020)	20	800237	5.91e−07	8.12e−10
			ETRPCA (Gao et al., 2020)	20	800000	5.22e−07	7.45e−09
			BTRTF (Zhou & Cheung, 2021)	20	800000	4.31e−07	7.14e−10
			CCSVS	20	800000	2.73e−08	6.48e−11

Table 2 Recovery results of different choices of n ($rk = rank_t(\mathcal{G}) = 0.1n$,$m = \|\mathcal{E}_0\|_0 = 0.2n^3$)

n	rk	m	Methods	$rank_t(\widehat{\mathcal{G}})$	$\|\widehat{\mathcal{E}}\|_0$	$\frac{\|\widehat{\mathcal{G}}-\mathcal{G}\|_F}{\|\mathcal{G}\|_F}$	$\frac{\|\widehat{\mathcal{E}}-\mathcal{E}\|_F}{\|\mathcal{E}\|_F}$
100	10	10	RPCA (Candès et al., 2011)	100	559283	8.49e−01	5.81e−03
			SNN (Liu et al., 2013)	10	402839	3.92e−04	9.62e−07
			TRPCA (Lu et al., 2020)	10	201349	5.72e−07	3.41e−09
			TPSSV (Zhang & Peng, 2019)	10	104328	2.31e−07	9.14e−10
			TRPCA-WTNN (Sun et al., 2020)	10	203294	1.75e−07	8.92e−09
			ETRPCA (Gao et al., 2020)	10	200000	1.71e−07	8.62e−09
			BTRTF (Zhou & Cheung, 2021)	10	200000	1.66e−07	8.05e−09
			CCSVS	10	200000	1.02e−07	4.47e−10
200	20	10	RPCA (Candès et al., 2011)	200	4600283	8.72e−01	3.10e−03
			SNN (Liu et al., 2013)	10	3650231	2.69e−04	9.93e−07
			TRPCA (Lu et al., 2020)	20	1623203	6.88e−07	1.91e−09
			TPSSV (Zhang & Peng, 2019)	20	7174300	7.43e+01	2.81e−01
			TRPCA-WTNN (Sun et al., 2020)	20	1603874	5.22e−07	8.45e−09
			ETRPCA (Gao et al., 2020)	20	1600000	5.05e−07	8.19e−09
			BTRTF (Zhou & Cheung, 2021)	20	1600000	4.83e−07	7.80e−10
			CCSVS	20	1600000	1.48e−07	2.73e−10

9.2 Image Recovery

The performance of the proposed CCSVS is verified on the task of color image recovery. Experiments are conducted over Berkeley Segmentation Dataset (BSD500) (Martin et al., 2001) which contains 500 color images of size $481 \times 321 \times 3$ or $321 \times 481 \times 3$. Ten percent of pixels in each color image are set as random values in the interval [0, 255]. The positions of the corrupted pixels are unobservable. All three color channels (i.e. red, green and blue channel) of the color images are corrupted at the same positions. The corruptions happen on the whole tubes. This problem is much more difficult than the corruptions on three channels at inconsistent positions (Lu et al., 2020). For the recovered color image, the quality is measured by the following Peak Signal-to-Noise Ratio (PSNR) value.

$$\text{PSNR} = 10\log_{10}\left(\frac{\|\mathcal{X}\|_\infty^2}{\frac{1}{n_1n_2n_3}\|\widehat{\mathcal{X}}-\mathcal{X}\|_F^2}\right),$$

where $\|\cdot\|_\infty$ denotes the infinity norm of a tensor.

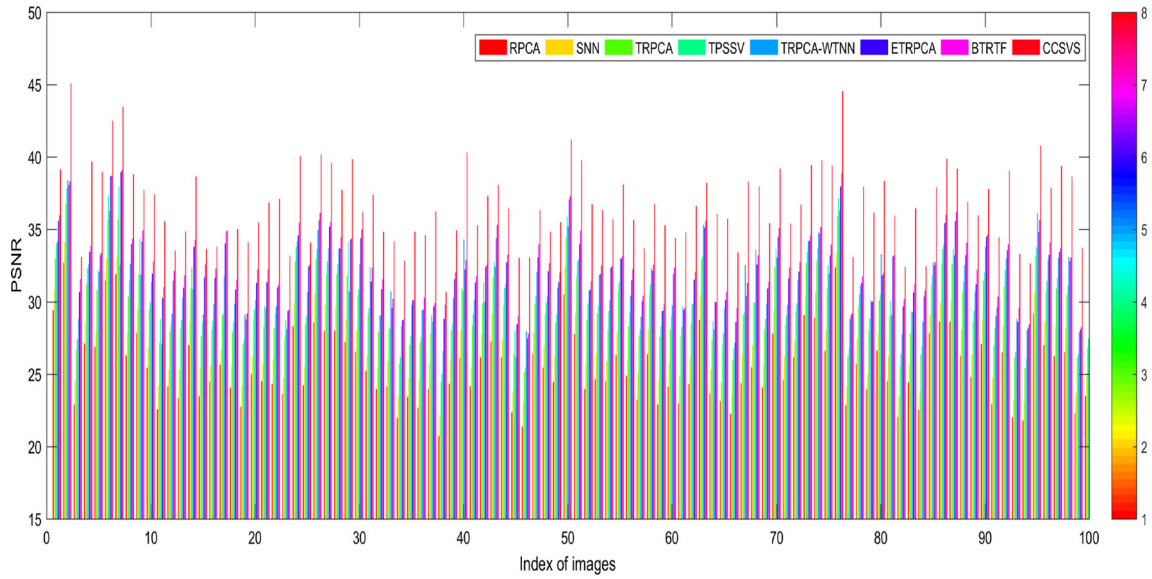


Fig. 3 The statistics of PSNR value of 100 images from the BSD500 datasets

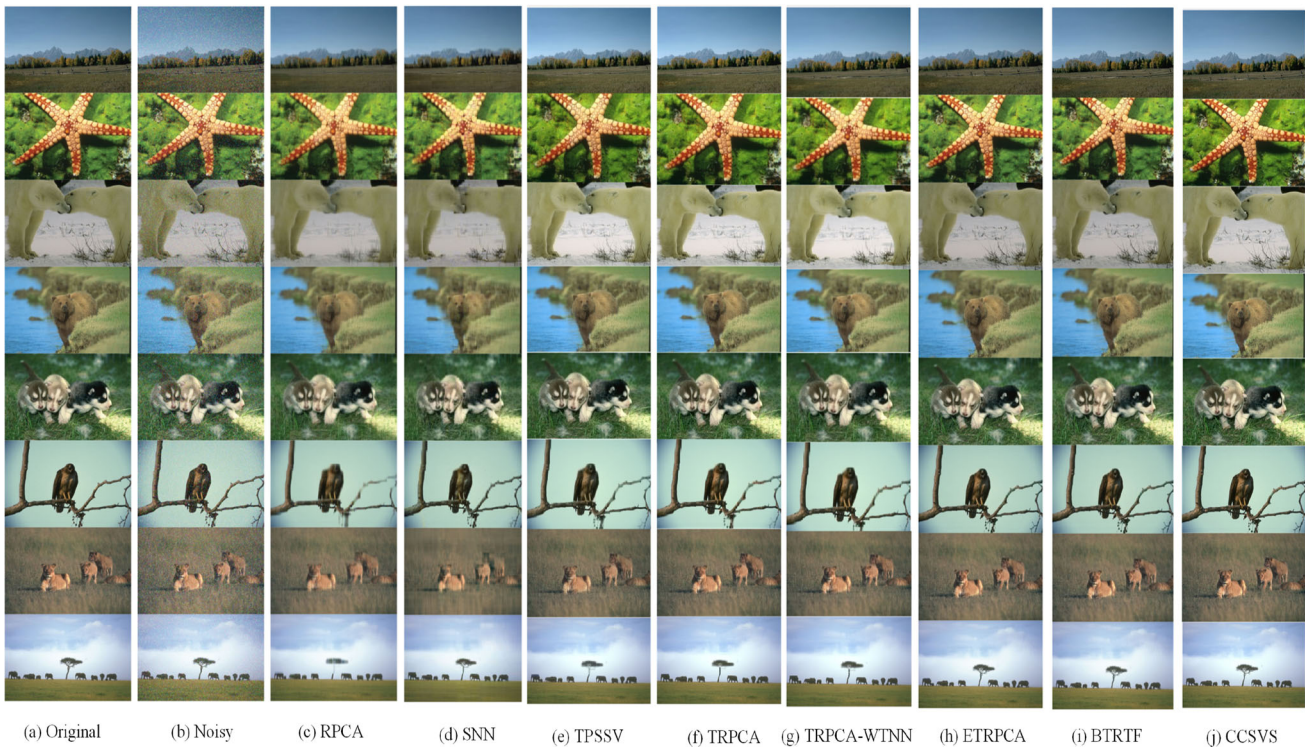


Fig. 4 Recovery result display over the selected eight images from the BSD500 datasets

Table 3 The setting of separation r in the proposed CCSVS for selected eight images from BSD500 dataset

Image Indices	1	2	3	4	5	6	7	8
The value of r	10	50	10	20	15	15	20	10

The recovery performance is positively related to the PSNR value. The statistics of PSNR value of selected methods on 100 images from the BSD500 datasets are displayed in Fig. 3, due to the limited space. The parameter setting of the proposed CCSVS model is as follows. For conve-

Table 4 The comparison of PSNR values over the above eight images

Image Indices	1	2	3	4	5	6	7	8
RPCA	28.67	25.53	25.51	32.28	27.32	23.49	32.45	27.85
SNN	30.82	28.25	26.70	34.21	30.23	25.17	34.82	29.66
TRPCA	32.56	28.81	29.09	35.10	32.29	27.17	35.66	31.45
TPSSV	33.68	29.47	29.28	36.48	32.74	27.61	37.84	32.46
TRPCA-WTNN	30.61	26.02	27.15	33.39	27.75	26.69	34.16	32.11
ETRPCA	35.57	31.23	32.87	37.71	34.88	32.40	38.12	37.25
BTRTF	36.52	31.36	33.10	37.85	35.01	32.84	38.55	37.64
CCSVS	**39.56**	**37.81**	**38.77**	**44.00**	**38.88**	**36.69**	**44.44**	**40.63**

Bold values represent the biggest PSNR values

Table 5 The comparison of running time in seconds about eight images

Image Indices	1	2	3	4	5	6	7	8
RPCA	25.74	26.12	26.14	26.0	25.97	26.14	26.60	26.31
SNN	27.73	28.12	28.01	27.93	29.12	28.05	28.01	28.07
TRPCA	13.10	13.21	13.58	14.01	13.88	12.94	13.78	13.91
TPSSV	13.00	12.84	12.58	13.25	13.04	12.01	12.77	13.21
TRPCA-WTNN	13.17	12.32	12.17	13.02	12.89	12.0	12.16	12.98
ETRPCA	39.45	39.07	41.25	39.53	41.23	37.80	41.25	40.18
BTRTF	32.91	32.27	33.83	32.60	33.07	31.72	33.78	32.84
CCSVS	**11.12**	**11.02**	**10.28**	**11.53**	**10.49**	**10.56**	**10.47**	**10.82**

Bold values represent the shortest running times

Table 6 The report on approximation quality of solutions to the proposed CCSVS model with respect to eight images

Image Indices	1	2	3	4	5	6	7	8
$rank_t(\widehat{\mathcal{X}})$	10	50	10	20	15	15	20	10
$\|\widehat{\mathcal{E}}\|_0$	46331	46029	46135	46411	46346	45999	46168	46386
$\frac{\|\widehat{\mathcal{X}}-\mathcal{X}_0\|_F}{\|\mathcal{X}_0\|_F}$	0.05	0.01	0.06	0.05	0.01	0.01	0.03	0.05
$\frac{\|\widehat{\mathcal{E}}-\mathcal{E}_0\|_F}{\|\mathcal{E}_0\|_F}$	0.13	0.07	0.15	0.12	0.08	0.04	0.04	0.18

nience, the separation parameter is set as $r = 10$ uniformly. $\Omega = (\omega, \omega_1, \omega_2, \cdots, \omega_r)$ is specified by $\omega = 0.7$ and the decreasing arithmetic progression $\{\omega_i\}_{i=1}^r$ is generated by '`linspace(1e-05,1e-07,r)`' in MATLAB command. The regularized coefficient λ in the proposed CCSVS is set as $\frac{1}{\sqrt{n_3 \max\{n_1,n_2\}}}$ by the suggested value in the recovery theorem (i.e. Theorem 7). The $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3)$ is specified by $\vartheta_1 = \frac{0.55}{n_3}$ and $\vartheta_2 = \vartheta_3 = \frac{1.55}{n_3}$. From Fig. 3, it can be readily seen that the performance of the proposed CCSVS significantly outperforms the state-of-the-art methods including RPCA (Candès et al., 2011), TRPCA (Lu et al., 2020), TPSSV (Zhang & Peng, 2019), TRPCA-WTNN (Sun et al., 2020), ETRPCA (Gao et al., 2020) and BTRTF (Zhou & Cheung, 2021).

Apart from the numerical viewpoint of the PSNR value, the realistic recovery result of the selected eight images is also exhibited in Fig. 4 for intuition. Figure 4 shows the original images, noisy images and the recovery images, which are produced by RPCA (Candès et al., 2011), SNN (Liu et al., 2013), TPSSV (Zhang & Peng, 2019), TRPCA (Lu et al., 2020), TRPCA-WTNN (Sun et al., 2020), ETRPCA (Gao et al., 2020) and BTRTF (Zhou & Cheung, 2021) respectively. Although the separation parameter $r = 10$ works well generally. Nevertheless, it may be not the best choice for concrete images. Therefore, it is adjusted according to the actual situation. The best separation parameter r for each image is set in Table 3 empirically.

The PSNR values of different methods are reported in Table 4. The running time is reported in Table 5. The RPCA behaves worse than the tensor methods: TRPCA (Lu et al., 2020), TPSSV (Zhang & Peng, 2019), TRPCA-WTNN (Sun et al., 2020), ETRPCA (Gao et al., 2020) and BTRTF (Zhou & Cheung, 2021). RPCA firstly runs for the color channel separately, which neglects structural information hidden in high-order tensor. The tensor methods improve the performance by leveraging this structural information. The proposed CCSVS achieves the best result in both performance and running time thanks to Table 4 and Table 5.

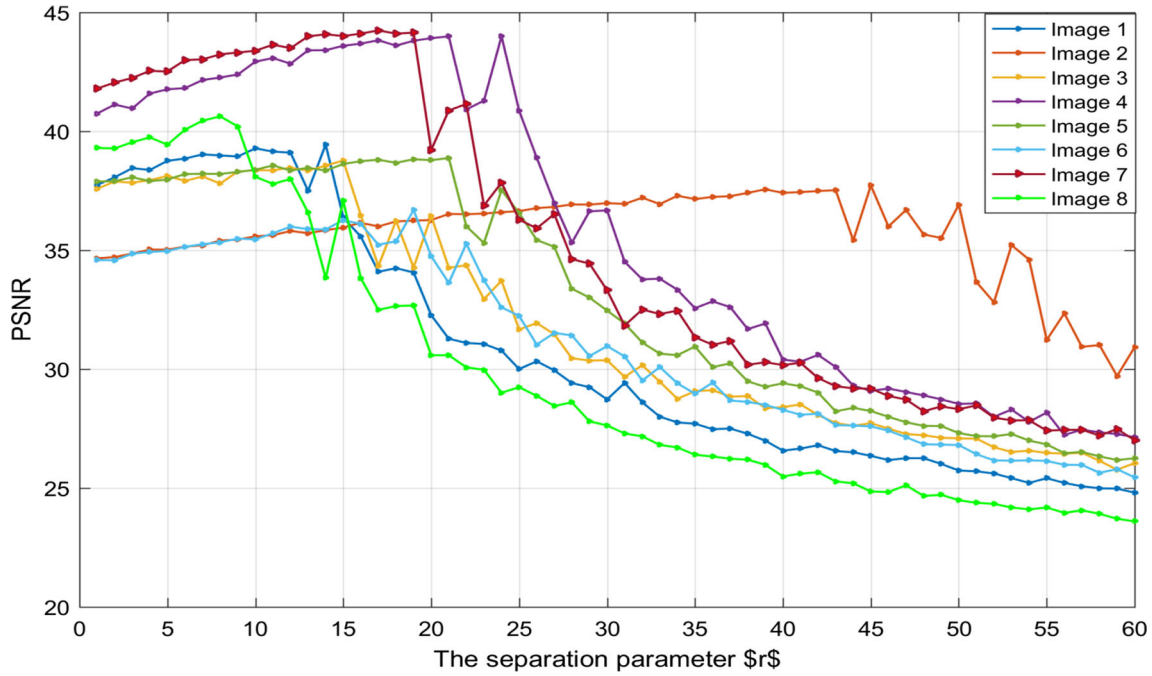
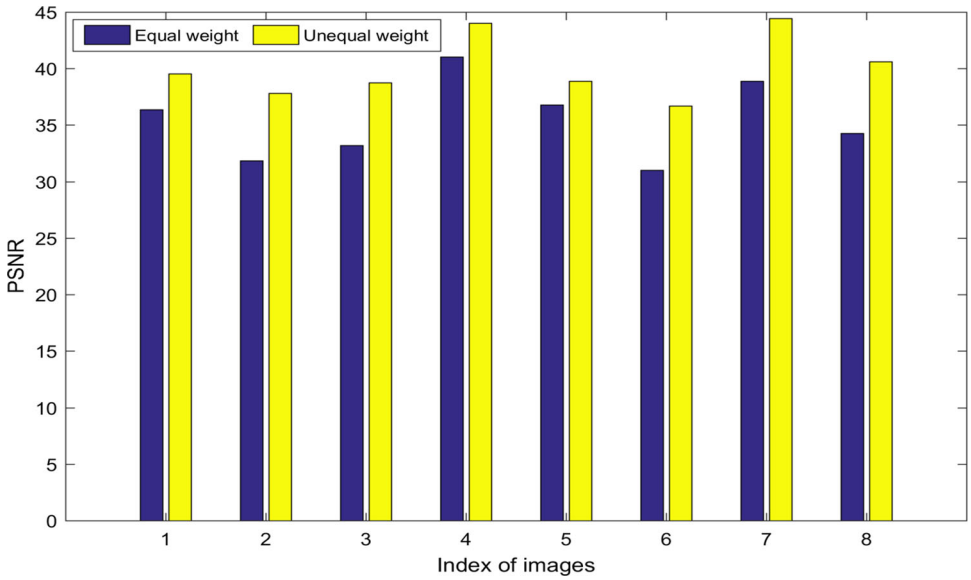


Fig. 5 The sensitivity curves of the proposed CCSVS with respect to the separation parameter r over eight selected images

Fig. 6 The performance comparison of the proposed CCSVS with respect to equal and unequal the frontal slice weight ϑ over eight selected images



The reason may be that the proposed CCSVS collaboratively segregates the singular values into high and low groups and adjusts the distribution of first several largest values towards the ground truth distribution with respect to the underlying low-rank component, and places different attention on frontal slices. The running time is satisfying and the key reason is that the proximal operator with respect to the proposed CCSVS can be efficiently solved by a closed-form expression. To see the approximation quality of solution to the proposed CCSVS model, we report the rank estimation, l_0-norm of sparse error and relative errors of low-rank and sparse terms in Table 6. It can be seen from Table 6 that the approximation of low-rank property and sparsity is good. This result confirms with convergence analysis in Theorem 6 and recovery guarantee in Theorem 7. Objectively speaking, the precision of relative errors is not high enough. This may be attributed to the fact that images just satisfy the approximate instead of absolute low-rank property.

Considering that the separation parameter r in the proposed CCSVS has heavy impact on the final performance, the ablation study on it is necessary. The sensitivity curves on the separation parameter r are exhibited in Fig. 5. It can

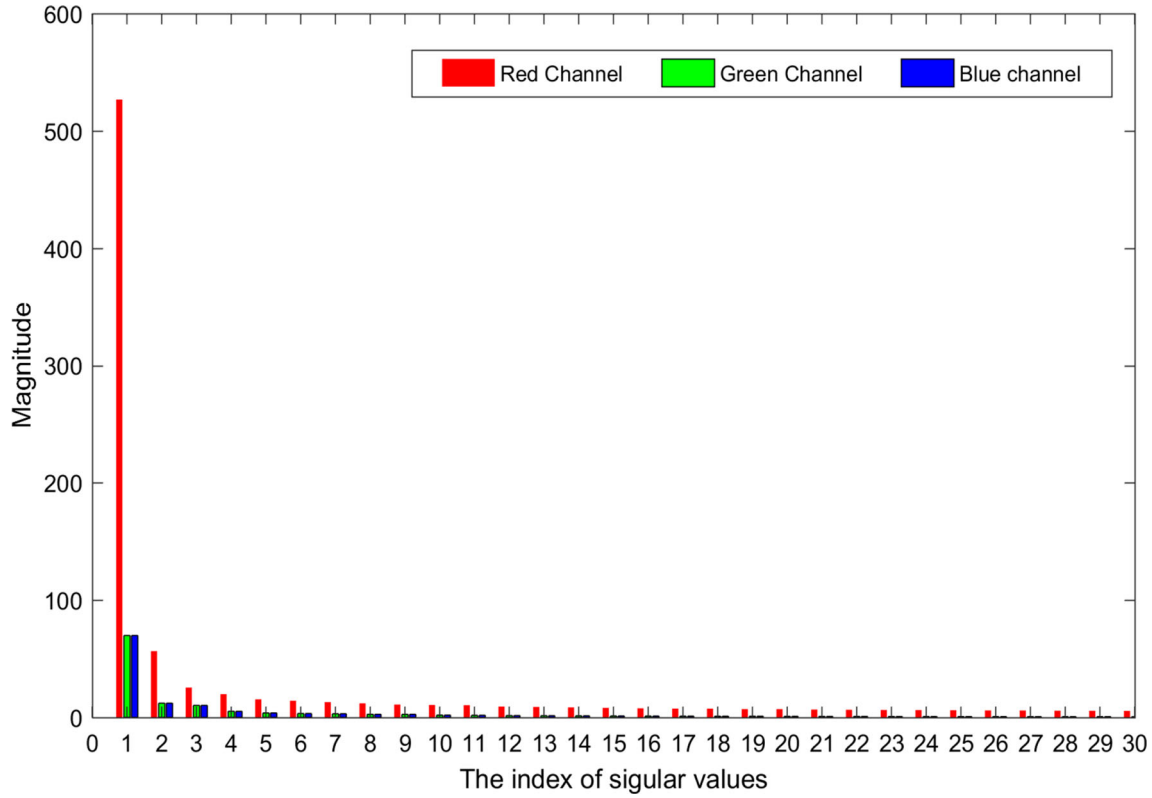


Fig. 7 The statistics of singular values for the image 1 along three channels. Due to limited space, only first 30 largest singular values are considered

be seen that the proposed CCSVS works well under the small value of the separation parameter r. When it takes larger values, the performance degrades rapidly. The cause is probably that the low value of r produces better low-rank approximation and makes the distribution of the first several largest singular values closer to the ground truth simultaneously.

The importance of different frontal slices also influences the performance of the proposed CCSVS. The comparison of equal and unequal frontal slice weight (i.e. $(\vartheta_1, \vartheta_2, \vartheta_3) = (1, 1, 1)$ and $(\vartheta_1, \vartheta_2, \vartheta_3) = (0.55, 1.55, 1.55)$) is displayed in Fig. 6. The performance under unequal frontal weight ϑ is better. The cause may be that much information of image is distributed along the first frontal slice. This can be supported by the observation of the distribution of singular values of the ground truth images as Fig. 7. Figure 7 takes image 1 as an example. It is revealed that the singular values along red channel are significantly larger than other channels. Therefore, considering the importance of different frontal slices is reasonable, which is neglected by all the previous works, such as TRPCA (Kilmer & Martin, 2011), TPSSV (Zhang & Peng, 2019), TRPCA-WTNN (Sun et al., 2020), ETRPCA (Gao et al., 2020) and BTRTF (Zhou & Cheung, 2021).

9.3 Background Modeling

We evaluate the performance of the proposed CCSVS model Eq. (15) on the task of video background modeling. For this task, the aim is to recover the background by removing foreground objects in color videos. As pointed out by Lu *et. al.* in Lu et al. (2020), this task can be roughly abstracted as the problem of low-rank tensor recovery.

Four color videos are selected from the public Scene Background Initialization (SBI) dataset (Maddalena & Petrosino, 2015), namely, Hall and Monitor, CAVIAR1, IBMtest2 and Candela. These videos have been preprocessed and a collection of images along frames is provided for each video. The image size in each frame is $w \times h \times 3$, where w and h represent the width and height respectively. Considering the dimension of frames, the experimental data is a fourth-order tensor. Hence a preprocessing procedure is necessary to make the proposed algorithm run successfully. Unlike previous works, e.g. TRPCA (Lu et al., 2020), which reshape the sampled video data as a third-order tensor $wh \times 3 \times f$, where f is the number of frames. Each image is reshaped as $wh \times f \times 3$. The reason why we do this is that each slice exactly corresponds to each color channel. With this format, the prior knowledge of color channels can be leveraged.

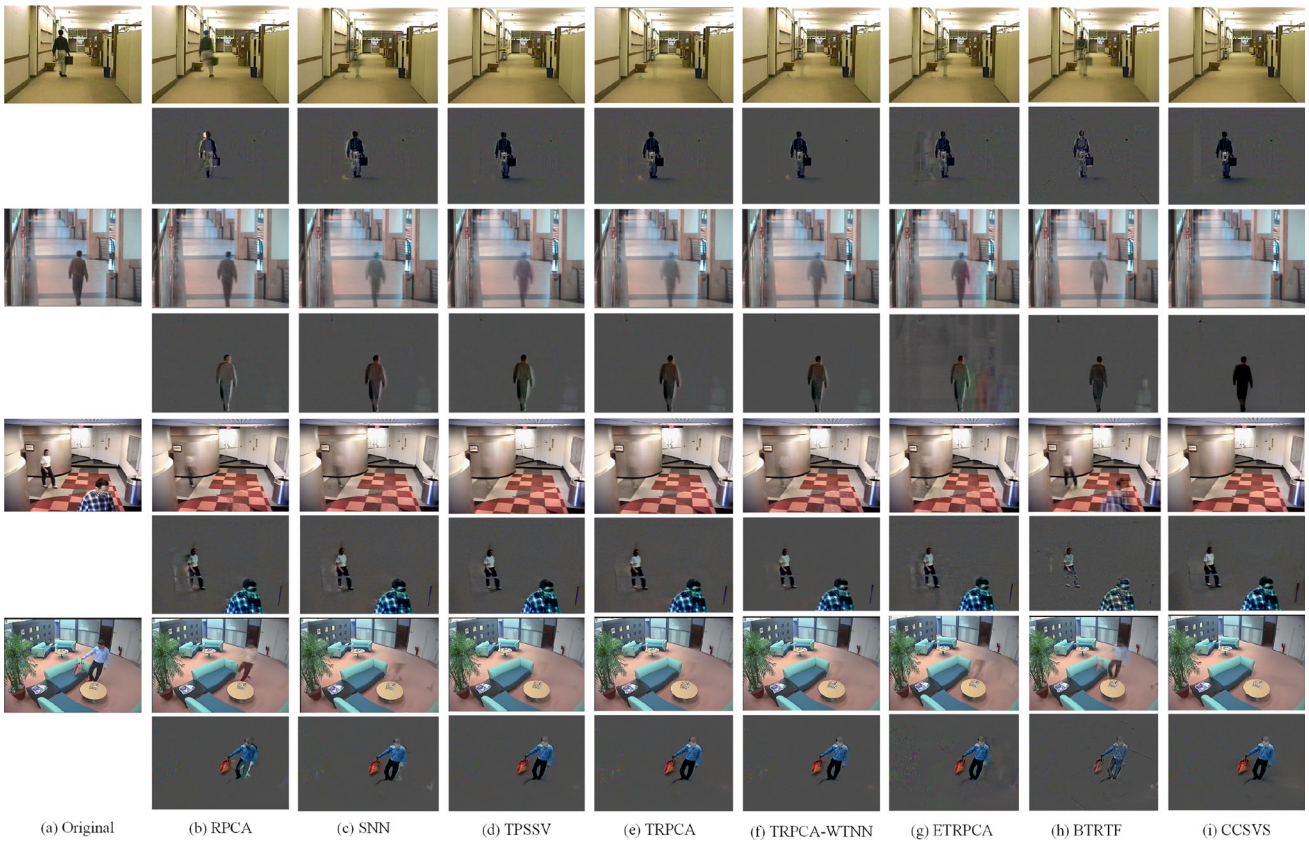


Fig. 8 Background modeling of four color video sequences. From top to down, the original image comes from Hall and Monitor, CAVIAR1, IBMtest2 and Candela datasets, respectively

The implementation details are as follows. In our experiments, we leverage the prior knowledge on color channels to set slice weights. Specifically, like Fig. 1b, we observe that the red channel usually carries more information amount than other channels and the residual channel bears with almost equal information amount. Hence, we set the weight ϑ_{red} of red channel as 0.55 and other weights $\vartheta_{\text{green}} (= \vartheta_{\text{blue}})$ as 1.55 in this experiment. The separation parameter r in Definition 4 is uniformly set as 1. The regularization parameter $\lambda = \frac{1}{\sqrt{3 \min\{wh, f\}}}$ by the suggested value in the recovery theorem (i.e. Theorem 7).

We compare the proposed method with seven baselines including RPCA (Candès et al., 2011), SNN (Liu et al., 2013), TRPCA (Lu et al., 2020), TPSSV (Zhang & Peng, 2019), TRPCA-WTNN (Sun et al., 2020), ETRPCA (Gao et al., 2020) and BTRTF (Zhou & Cheung, 2021). To make use of RPCA, we reshape it to a matrix of size $(3hw) \times f$. Due to the inherent restriction of BTRTF, we reshape the video sequences to third-order tensors of size $w \times 3h \times f$. It can be seen from Fig. 8 that the proposed CCSVS method achieves the best performance. The recovered background is almost the same as ground truth. It mainly attributes to the technique of separation of singular values, which guides the low-rank tensor approximation and the distribution of top-r singular values. Besides, the prior information on different color channels also plays an important role for the final successful background modeling.

10 Conclusion and Discussion

10.1 Conclusion

In this paper, a convex–concave singular value separation (CCSVS) based tensor robust principal component analysis method is proposed, which collaboratively segregates large and small singular values, controls the distribution of the first several largest singular values towards the ground truth, and emphasizes the different importance of each frontal slice. It breaks the prior assumption that frontal slices are treated equally which is hidden in tensor nuclear norm from previous works, e.g. TRPCA. The proximal operator with respect to the CCSVS is discussed and a closed-form solution is found. The convergence of whole sequences, instead of subsequences, generated by the optimization procedure is proved in the nonconvex setting and we show that the limit point

is the stationary point. Furthermore, the recovery guarantee is established for the proposed CCSVS model. Extensive experimental results on synthetic and real datasets indicate the superiority of the proposed CCSVS and the running time of algorithm is also the fastest in practice.

10.2 Discussion

10.2.1 The Limitations of Matricizing Methods

Roughly speaking, the proposed CCSVS method belongs to the class of matricizing methods (i.e. tensor slicing methods). that slice the tensor in different ways so that classical linear algebra tools such as SVD can be used to solve the problem efficiently. Seemingly, we decompose a tensor into matrices via slicing techniques and then treat these matrices separately. In fact, there exists a preprocessing step via Discrete Fourier Transformation (DFT) along all tubes. Because each tube exactly includes third-dimensional elements, the resulting tensor after DFT integrates the information from all slices of the original tensor. In other words, the information of slices in resulted tensor is not separate and they are correlated via the transformation of DFT. After the processing of DFT, we deal with each slice separately and employ tools from linear algebra naturally. One limitation is that the transformed tensor by DFT lies in the space of the complex number and the related proofs become more difficult. The DFT is a linear transformation and cannot possibly non-linear structural information hidden in tensor data.

10.2.2 The Setting of Slice Weights

In our experiments, we leverage the prior knowledge on color channels to determine slice weights. For color images, like Fig. 1 (b), the red channel usually carries more information amount than other channels and the residual two channels bear almost equal information amount. On the basis of this observation, we select the slice weights $\vartheta_{\text{red}} \in [0.35, 0.65]$ and $\vartheta_{\text{green}} = \vartheta_{\text{blue}} \in [1, 1.7]$ to balance information distribution. For the task of background modeling, we also discover a similar phenomenon for color channels, and hence we choose $\vartheta_{\text{red}} \in [0.3, 0.6]$ and $\vartheta_{\text{green}} = \vartheta_{\text{blue}} \in [1, 2]$.

Besides, we also discover a different way to set weights of slices as follows. The weights of slices can be sampled from exponential distributions in practice. The exponential distribution satisfies the memoryless property. Considering this good property and the inherent structure within tensors, we find that this weighting scheme also performs well in practice and produces almost the same experimental results (Hence we do not exhibit them repeatedly.). Auto-weighting schemes will be explored in our future works.

Acknowledgements We would like to thank the anonymous reviewers for their valuable comments and suggestions which helped to improve the quality of this paper. This work was supported in part by National Natural Science Foundation of China under Grants 62106081, 62225113, 62106063 and 62122060, by the Guangdong Natural Science Foundation under Grant 2022A1515010819 and by the Shenzhen Science and Technology Program under Grant RCBS20210609103708013.

Data Availability The Berkeley Segmentation dataset (BSD500) in experiments is publicly available and can be downloaded from the website https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/. The Scene Background Initialization (SBI) dataset can be found at https://sbmi2015.na.icar.cnr.it/SBIdataset.html.

11 Appendix

11.1 Preliminary Lemmas

Lemma 1 (Fan, 1951) *If* $\mathbf{H} \in \mathbb{C}^{n\times n}$ *is a complex matrix of order n, it holds that, for any* $1 \le k \le n$,

$$\sum_{i=1}^{k} \sigma_i(\mathbf{H}) = \max |\sum_{i=1}^{k} \langle \mathbf{U}\mathbf{H}\mathbf{z}_i, \mathbf{z}_i\rangle|, \tag{60}$$

where the maximization means that $\mathbf{U}$ *runs over all unitary matrices in* $\mathbb{C}^{n\times n}$ *and* $\{\mathbf{z}_i\}_{i=1}^{k}$ *runs over k orthonormal sets in* $\mathbb{C}^n$.

This lemma is also generalized to the case of multiple matrices as follows.

Lemma 2 (Fan, 1951) *If* $\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_t$ *are complex matrices from* $\mathbb{C}^{n\times n}$, *it holds that, for any* $1 \le k \le n$,

$$\begin{aligned}&\sum_{i=1}^{k} \sigma_i(\mathbf{H}_1)\sigma_i(\mathbf{H}_2)\cdots\sigma_i(\mathbf{H}_t)\\ &= \max |\sum_{i=1}^{k} \langle \mathbf{U}_1\mathbf{H}_1\mathbf{U}_2\mathbf{H}_2\cdots\mathbf{U}_t\mathbf{H}_t\mathbf{z}_i, \mathbf{z}_i\rangle|,\end{aligned} \tag{61}$$

where the maximization means that $\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_t$ *run over all unitary matrices in* $\mathbb{C}^{n\times n}$ *and* $\{\mathbf{z}_i\}_{i=1}^{k}$ *runs over k orthonormal sets in* $\mathbb{C}^n$.

11.2 The Proof of Theorem 1

Proof Denote $\widetilde{n} = \max\{n_1, n_2\}$ and $\widehat{n} = \min\{n_1, n_2\}$. We augment $\mathbf{X}$ as $\widetilde{\mathbf{X}} \in \mathbb{R}^{\widetilde{n}\times\widetilde{n}}$ by the following method. If column size n_2 is less than row size, i.e. $n_2 < n_1$, we form $\widetilde{\mathbf{X}}$ by zero padding scheme along columns. Otherwise, we perform zero padding scheme along rows. Either way, the first $\widehat{n}$ singular values of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ coincide.

It follows from the well-known Ky Fan inequality (Moslehian, 2012) that

$$\sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\widetilde{\mathbf{X}}_1+\widetilde{\mathbf{X}}_2) \le \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\widetilde{\mathbf{X}}_1) + \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\widetilde{\mathbf{X}}_2). \tag{62}$$

Equivalently,

$$\sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}_1+\mathbf{X}_2) \le \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}_1) + \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}_2).$$

Combining this fact with the positive homogeneity of singular value function $\sigma_i(\cdot)$, we have

$$\sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\theta\mathbf{X}_1+(1-\theta)\mathbf{X}_2) \le \theta \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}_1) + (1-\theta) \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}_2), \tag{63}$$

where $\theta \in [0,1]$. This implies the convexity of the function $f_{convex}(\cdot)$.

Let

$$F(\mathbf{X}) = \sum_{i=1}^{r} \omega_i \sigma_i(\mathbf{X}). \tag{64}$$

By Lemma 1, we have

$$\sum_{i=1}^{r} \sigma_i(\widetilde{\mathbf{X}}) = \max|\sum_{i=1}^{r} \langle \mathbf{U}\mathbf{H}\mathbf{z}_i, \mathbf{z}_i\rangle|. \tag{65}$$

Denote $\mathbf{D} = diag([\omega_1, \omega_2, \cdots, \omega_r, 0, \cdots, 0]) \in \mathbb{R}^{\tilde{n}\times\tilde{n}}$. Given $\mathbf{A}$ and $\mathbf{B}$ from $\mathbb{C}^{n_1\times n_2}$, we represent $\mathbf{C} = \mathbf{A}+\mathbf{B}$. By the aforementioned augmentation method, $\widetilde{\mathbf{C}} = \widetilde{\mathbf{A}}+\widetilde{\mathbf{B}}$. Let $\widetilde{\mathbf{X}} = \widetilde{\mathbf{P}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{Q}}^*$ be the singular value decomposition of $\widetilde{\mathbf{X}}$. Then

$$\begin{aligned}
&\sum_{i=1}^{r} \omega_i \sigma_i(\widetilde{\mathbf{C}}) \\
&= \max|\sum_{i=1}^{r} \langle \mathbf{U}\widetilde{\mathbf{P}}\widetilde{\mathbf{\Sigma}}\mathbf{D}\widetilde{\mathbf{Q}}^*\mathbf{z}_i, \mathbf{z}_i\rangle| \\
&= \max|\sum_{i=1}^{r} \langle \mathbf{U}(\widetilde{\mathbf{P}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{Q}}^*)\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*\mathbf{z}_i, \mathbf{z}_i\rangle| \\
&= \max|\sum_{i=1}^{r} \langle \mathbf{U}\widetilde{\mathbf{C}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*\mathbf{z}_i, \mathbf{z}_i\rangle| \\
&\le \max|\sum_{i=1}^{r} \langle \mathbf{U}\widetilde{\mathbf{A}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*\mathbf{z}_i, \mathbf{z}_i\rangle| \\
&\quad + \max|\sum_{i=1}^{r} \langle \mathbf{U}\widetilde{\mathbf{B}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*\mathbf{z}_i, \mathbf{z}_i\rangle|,
\end{aligned} \tag{66}$$

where the first equality holds by the decreasing property of the diagonal elements of $\mathbf{D}$ and the last inequality is attributed to triangle inequality.

Hence, in view of Lemma 1 and Eq. (68), we have

$$\sum_{i=1}^{r} \omega_i \sigma_i(\widetilde{\mathbf{C}}) \le \sum_{i=1}^{r} \sigma_i(\widetilde{\mathbf{A}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*) + \sum_{i=1}^{r} \sigma_i(\widetilde{\mathbf{B}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*). \tag{67}$$

Note that

$$\begin{aligned}
\sum_{i=1}^{r} \sigma_i(\widetilde{\mathbf{A}})\sigma_i(\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*) &= \max|\sum_{i=1}^{k} \langle \mathbf{U}_1\widetilde{\mathbf{A}}\mathbf{U}_2\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*\mathbf{z}_i, \mathbf{z}_i\rangle| \\
&\ge \max|\sum_{i=1}^{r} \langle \mathbf{U}_1\widetilde{\mathbf{A}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*\mathbf{z}_i, \mathbf{z}_i\rangle| \\
&= \sum_{i=1}^{k} \sigma_i(\widetilde{\mathbf{A}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*).
\end{aligned} \tag{68}$$

The left-hand side is further equivalent to

$$\begin{aligned}
\sum_{i=1}^{r} \sigma_i(\widetilde{\mathbf{A}})\sigma_i(\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*) &= \sum_{i=1}^{r} \sigma_i(\widetilde{\mathbf{A}})\sqrt{\lambda_i(\widetilde{\mathbf{Q}}\mathbf{D}^2\widetilde{\mathbf{Q}}^*)} \\
&= \sum_{i=1}^{r} \sigma_i(\widetilde{\mathbf{A}})\sqrt{\lambda_i(\mathbf{D}^2)} \\
&= \sum_{i=1}^{r} \omega_i \sigma_i(\widetilde{\mathbf{A}}).
\end{aligned} \tag{69}$$

Then

$$\sum_{i=1}^{k} \sigma_i(\widetilde{\mathbf{A}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*) \le \sum_{i=1}^{r} \omega_i \sigma_i(\widetilde{\mathbf{A}}). \tag{70}$$

With a similar argument, we have

$$\sum_{i=1}^{k} \sigma_i(\widetilde{\mathbf{B}}\widetilde{\mathbf{Q}}\mathbf{D}\widetilde{\mathbf{Q}}^*) \le \sum_{i=1}^{r} \omega_i \sigma_i(\widetilde{\mathbf{B}}). \tag{71}$$

Gathering Eqs. (67–71), it holds that

$$\sum_{i=1}^{r} \omega_i \sigma_i(\widetilde{\mathbf{C}}) \le \sum_{i=1}^{r} \omega_i \sigma_i(\widetilde{\mathbf{A}}) + \sum_{i=1}^{r} \omega_i \sigma_i(\widetilde{\mathbf{B}}). \tag{72}$$

We simplify Eq. (72) as

$$\sum_{i=1}^{r}\omega_i\sigma_i(\mathbf{C}) \leq \sum_{i=1}^{r}\omega_i\sigma_i(\mathbf{A}) + \sum_{i=1}^{r}\omega_i\sigma_i(\mathbf{B}). \tag{73}$$

In other words,

$$F(\mathbf{A}+\mathbf{B}) \leq F(\mathbf{A}) + F(\mathbf{B}). \tag{74}$$

Considering the positive homogeneity of singular value function $\sigma_i(\cdot)$, we have

$$F(\theta\mathbf{A} + (1-\theta)\mathbf{B}) \leq \theta F(\mathbf{A}) + (1-\theta)F(\mathbf{B}), \tag{75}$$

where $\theta \in [0,1]$. Therefore, the function $F(\cdot)$ is convex, which implies the concave property of $g_{concave}(\cdot)$. □

11.3 The Remaining Proof of Theorem 6

Since the objective in Eq. (15) is convex–concave, the traditional concept of subdifferential for convex functions is not suitable for convergence analysis here. As a matter of fact, the following notion of limiting subdifferential for non-convex and non-smooth functions is a necessary ingredient.

Definition 8 (The Limiting Subdifferential (Rockafellar & Wets, 1998) **in Tensor Form**) Let G be a proper and lower semi-continuous function $G : \mathbb{R}^{n_1\times n_2\times n_3} \to \mathbb{R}$. The Frechét subdifferential of G at $\mathcal{X}$ is

$$\widehat{\partial}G(\mathcal{X}) = \Big\{\mathcal{Z}\in\mathbb{R}^{n_1\times n_2\times n_3} : \lim_{\mathcal{Y}\neq\mathcal{X}}\inf_{\mathcal{Y}\to\mathcal{X}}\frac{G(\mathcal{Y}) - G(\mathcal{X}) - \langle\mathcal{Z}, \mathcal{Y}-\mathcal{X}\rangle}{\|\mathcal{Y}-\mathcal{X}\|_F} \geq 0.\Big\}.$$

The *limiting subdifferetial* of G at $\mathcal{X}$ is

$$\begin{aligned}\partial G(\mathcal{X}) = \{&\mathcal{Z}\in\mathbb{R}^{n_1\times n_2\times n_3} : \exists\mathcal{X}^{(k)}\to\mathcal{X}, G(\mathcal{X}^{(k)})\\ &\to G(\mathcal{X}), \mathcal{Z}^{(k)}\in\widehat{\partial}G(\mathcal{X})\to\mathcal{Z},\\ &\text{as } k\to\infty\}.\end{aligned} \tag{76}$$

The following three propositions is also necessary for convergence analysis.

Proposition 3 *(Fermat's Rule* (Rockafellar & Wets, 1998)*) In the non-smooth case, Fermat's rule still holds. This is, if $\mathcal{X}$ is a local minimizer of non-smooth function G, then $0\in\partial G(\mathcal{X})$, i.e. $\mathcal{X}$ is a stationary point of G.*

Proposition 4 (Rockafellar & Wets, 1998) *Let $\{\mathcal{X}_k\}_{k\geq1}$ and $\{\mathcal{Z}_k\}_{k\geq1}$ satisfy $\lim_{k\to\infty}\mathcal{X}_k = \mathcal{X}$, $\lim_{k\to\infty}\mathcal{Z}_k = \mathcal{Z}$, $\lim_{k\to\infty}G(\mathcal{X}_k) = G(\mathcal{X})$ and $\mathcal{Z}_k\in\partial G(\mathcal{X}_k)$. Then $\mathcal{Z}\in\partial G(\mathcal{X})$.*

Proposition 5 (Rockafellar & Wets, 1998) *Let G be a proper and lower semi-continuous function $G : \mathbb{R}^{n_1\times n_2\times n_3}\to\mathbb{R}$. Besides, suppose that $f : \mathbb{R}^{n_1\times n_2\times n_3}\to\mathbb{R}$ is continuously differentiable. Then $\partial(f+G)(\mathcal{X}) = \nabla f(\mathcal{X}) + \partial G(\mathcal{X})$.*

Proof (The remaining part of the proof of Theorem 6.) First of all, we prove the basic fact that the limiting point pair $(\mathcal{X}^*, \mathcal{E}^*)$ is actually a feasible point pair. By Theorem 5, we have

$$\lim_{k\to\infty}\|\mathcal{Y} - \mathcal{X}^{(k+1)} - \mathcal{E}^{(k+1)}\|_F = \|\mathcal{Y} - \mathcal{X}^* - \mathcal{E}^*\|_F = 0. \tag{77}$$

Then the constraint $\mathcal{Y} = \mathcal{X}^* + \mathcal{E}^*$ is satisfied, which implies that $(\mathcal{X}^*, \mathcal{E}^*)$ is a feasible point for the model Eq. (15).

By Proposition 3, it can be derived from Eq. (38) that

$$0\in\partial\left(\frac{1}{\mu^{(k)}}\mathrm{Sep}(\mathcal{X}^{(k+1)}, \Omega, \vartheta) + \frac{1}{2}\|\mathcal{X}^{(k+1)} - \mathcal{N}^{(k)}\|_F^2\right). \tag{78}$$

Further, by Proposition 5, Eq. (80) can be reformulated as

$$0\in\frac{1}{\mu^{(k)}}\partial\mathrm{Sep}(\mathcal{X}^{(k+1)}, \Omega, \vartheta) + \mathcal{X}^{(k+1)} - \mathcal{N}^{(k)}. \tag{79}$$

Recall that $\mathcal{N}^{(k)} = \mathcal{Y} + \frac{1}{\mu^{(k)}}\mathcal{L}^{(k)} - \mathcal{E}^{(k)}$. Then we get

$$\mathcal{L}^{(k)}\in\partial\mathrm{Sep}(\mathcal{X}^{(k+1)}, \Omega, \vartheta) + \mu^{(k)}(\mathcal{X}^{(k+1)} + \mathcal{E}^{(k+1)} - \mathcal{Y}). \tag{80}$$

Equivalently, considering the expression of $\mathcal{L}^{(k+1)}$ in the Algorithm 1, we have

$$\mathcal{L}^{(k+1)}\in\partial\mathrm{Sep}(\mathcal{X}^{(k+1)}, \Omega, \vartheta). \tag{81}$$

We claim that

$$\lim_{k\to\infty}\mathrm{Sep}(\mathcal{X}^{(k+1)}, \Omega, \vartheta) = \mathrm{Sep}(\mathcal{X}^*, \Omega, \vartheta). \tag{82}$$

According to the structural expression in Theorem 1 and the triangle inequality of absolute value, we have

$$\begin{aligned}&\Big|\mathrm{Sep}(\mathcal{X}^{(k+1)}, \Omega, \vartheta) - \mathrm{Sep}(\mathcal{X}^*, \Omega, \vartheta)\Big|\\ &\leq \underbrace{\omega\Big|\|\mathcal{X}^{(k+1)}\|_* - \|\mathcal{X}^*\|_*\Big|}_{J_0^{(k)}}\\ &+ \underbrace{\left|\sum_{i=1}^{n_3}\vartheta_i\sum_{j=1}^{r}\widetilde{\omega}_j\sigma_j(\overline{\mathcal{X}^{(k+1)}}^{(i)}) - \sum_{i=1}^{n_3}\vartheta_i\sum_{j=1}^{r}\widetilde{\omega}_j\sigma_j(\overline{\mathcal{X}^*}^{(i)})\right|}_{J_1^{(k)}}.\end{aligned} \tag{83}$$

On the one hand, it holds that

$$\begin{aligned} J_0^{(k)} &\le \omega \|\mathcal{X}^{(k+1)} - \mathcal{X}^*\|_* \\ &\le \omega \sqrt{\frac{\min\{n_1, n_2\}}{n_3}} \|\mathcal{X}^{(k+1)} - \mathcal{X}^*\|_F. \end{aligned} \tag{84}$$

where the first inequality is attributed to the triangle inequality of the tensor nuclear norm. Because $\lim_{k\to\infty} \mathcal{X}^{(k+1)} = \mathcal{X}^*$, we have $\lim_{k\to 0} J_0^{(k)} = 0$.

On the other hand, we have

$$\begin{aligned} J_1^{(k)} &\le \sum_{i=1}^{n_3} \vartheta_i \sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j (\overline{\mathcal{X}^{(k+1)}}^{(i)} - \overline{\mathcal{X}^*}^{(i)}) \\ &\le \vartheta_{\max}\omega_{\max} \sum_{i=1}^{n_3} \sum_{j=1}^{r} \sigma_j (\overline{\mathcal{X}^{(k+1)}}^{(i)} - \overline{\mathcal{X}^*}^{(i)}) \\ &\le \vartheta_{\max}\omega_{\max} \sum_{i=1}^{n_3} \sum_{j=1}^{\min\{n_1,n_2\}} \sigma_j (\overline{\mathcal{X}^{(k+1)}}^{(i)} - \overline{\mathcal{X}^*}^{(i)}) \\ &= n_3 \vartheta_{\max}\omega_{\max} \|\mathcal{X}^{(k+1)} - \mathcal{X}^*\|_* \\ &\le \sqrt{\min\{n_1, n_2\} n_3} \vartheta_{\max}\omega_{\max} \|\mathcal{X}^{(k+1)} - \mathcal{X}^*\|_F, \end{aligned} \tag{85}$$

where $\vartheta_{\max} = \max_{i\in[n_3]} \vartheta_i$ and $\omega_{\max} = \|\Omega\|_\infty$, in which $\Omega = \{\omega, \omega_1, \cdots, \omega_r\}$ comes from Definition 4. The first inequality holds by the convexity of top-r singular values like Eq. (74). According to the fact $\lim_{k\to\infty} \mathcal{X}^{(k+1)} = \mathcal{X}^*$, it holds that $\lim_{k\to 0} J_1^{(k)} = 0$.

Therefore, Eq. (84) holds. Recall that $\mathcal{L}^*$ is an accumulation point of the sequence $\{\mathcal{L}^{(k)}\}_{k\ge1}$. Then there must exist a subsequence $\{\mathcal{L}^{(k_j)}\}_{j\ge1}$ which converges to $\mathcal{L}^*$. Select the corresponding subsequences $\{\mathcal{X}^{(k_j)}\}_{j\ge1}$ and $\{\mathcal{X}^{(k_j)}\}_{j\ge1}$ in $\{\mathcal{X}^{(k)}\}_{k\ge1}$ and $\{\mathcal{E}^{(k)}\}_{k\ge1}$, respectively. Using these selected subsequences, By Proposition 4, we have

$$0 \in \partial \mathrm{Sep}(\mathcal{X}^*, \Omega, \vartheta) - \mathcal{L}^*. \tag{86}$$

By Proposition 3, it can be derived from Eq. (40) that

$$0 \in \partial \left(\frac{\lambda}{\mu^{(k)}} \|\mathcal{E}^{(k+1)}\|_1 + \frac{1}{2} \|\mathcal{E}^{(k+1)} - \mathcal{J}^{(k)}\|_F^2 \right). \tag{87}$$

Recall that $\mathcal{J}^{(k)} = \mathcal{Y} + \frac{1}{\mu^{(k)}} \mathcal{L}^{(k)} - \mathcal{X}^{(k+1)}$. Then it holds that

$$\mathcal{L}^{(k)} \in \partial \|\mathcal{E}^{(k+1)}\|_1 + \mu^{(k)} (\mathcal{X}^{(k+1)} + \mathcal{E}^{(k+1)} - \mathcal{Y}). \tag{88}$$

Equivalently, considering the expression of $\mathcal{L}^{(k+1)}$ in the Algorithm 1, we have

$$\mathcal{L}^{(k+1)} \in \partial \|\mathcal{E}^{(k+1)}\|_1. \tag{89}$$

We claim that

$$0 \in \partial \|\mathcal{E}^{(*)}\|_1 - \mathcal{L}^*. \tag{90}$$

Note that

$$\begin{aligned} &\left| \|\mathcal{E}^{(k+1)}\|_1 - \|\mathcal{E}^*\|_1 \right| \\ &\le \|\mathcal{E}^{(k+1)} - \mathcal{E}^*\|_1 \\ &\le \sqrt{n_1 n_2 n_3} \|\mathcal{E}^{(k+1)} - \mathcal{E}^*\|_F, \end{aligned} \tag{91}$$

where the first inequality holds by triangle inequality.

Considering the aforementioned subsequences $\{\mathcal{X}^{(k_j)}\}_{j\ge1}$, $\{\mathcal{E}^{(k_j)}\}_{j\ge1}$ and $\{\mathcal{X}^{(k_j)}\}_{j\ge1}$, it follows from Proposition 4 that

$$0 \in \partial \|\mathcal{E}^{(*)}\|_1 - \mathcal{L}^*. \tag{92}$$

□

11.4 The Proof of Theorem 7

Let $\mathcal{A}$ be a real tensor of size $n_1 \times n_2 \times n_3$ and $\Phi = diag([\theta_1, \theta_2, \cdots, \theta_{n_3}]) \in \mathbb{R}^{n_3\times n_3}$. We define the tensor product of Φ and $\mathcal{X}$ as

$$\Phi \otimes \mathcal{A} = fold \left(\begin{bmatrix} \theta_1 \mathcal{A}^{(1)} \\ \theta_2 \mathcal{A}^{(2)} \\ \vdots \\ \theta_{n_3} \mathcal{A}^{(n_3)} \end{bmatrix} \right). \tag{93}$$

Denote $\mathbf{T}$ by

$$\mathbf{T} = \{\mathcal{U} * \mathcal{Y}^* + \mathcal{W} * \mathcal{V}^* : \mathcal{Y}, \mathcal{W} \in \mathbb{R}^{n_1\times r\times n_3}\}. \tag{94}$$

Let $\mathbf{T}^\perp$ be the orthogonal component of $\mathbf{T}$ in $\mathbb{R}^{n_1\times n_2\times n_3}$. Given any $\mathcal{Z} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the projection of $\mathcal{Z}$ into $\mathbf{T}$ and $\mathbf{T}^\perp$ are prescribed as

$$\mathcal{P}_{\mathbf{T}} \mathcal{Z} = \mathcal{U} * \mathcal{U}^* * \mathcal{Z} + \mathcal{Z} * \mathcal{V} * \mathcal{V}^* - \mathcal{U} * \mathcal{U}^* * \mathcal{Z} * \mathcal{V} * \mathcal{V}^* \tag{95}$$

and

$$\mathcal{P}_{\mathbf{T}^\perp} \mathcal{Z} = (\mathcal{I}_{n_1} - \mathcal{U} * \mathcal{U}^*) * \mathcal{Z} * (\mathcal{I}_{n_2} - \mathcal{V} * \mathcal{V}^*), \tag{96}$$

respectively.

To prove the Theorem 7, we need the following auxiliary lemma.

Lemma 3 *For any tensor $\mathcal{H} \in \mathbb{R}^{n_1\times n_2\times n_3}$, we have*

$$\|\mathcal{H}\|_* \ge \|\mathcal{P}_{\mathbf{T}^\perp} \mathcal{H}\|_*. \tag{97}$$

Proof By the expression of the tensor nuclear norm (Please refer to Proposition 2 for details.),

$$\|\mathcal{P}_{\mathbf{T}^\perp}\mathcal{H}\|_* = \sum_{i=1}^{n_3}\frac{1}{n_3}\sum_{j=1}^{\min\{n_1,n_2\}} \sigma_j(\overline{\mathcal{P}_{\mathbf{T}^\perp}\mathcal{H}}^{(i)}). \tag{98}$$

By the definition of projection Eq. (96) and Marshall and Olkin (1979),

$$\sum_{j=1}^{\min\{n_1,n_2\}}\sigma_j(\overline{\mathcal{P}_{\mathbf{T}^\perp}\mathcal{H}}^{(i)}) \le \sum_{j=1}^{\min\{n_1,n_2\}}\sigma_j(\mathbf{I}_{n_1}\overline{\mathcal{U}}^{(j)}(\overline{\mathcal{U}}^{(j)})^*)\sigma_j(\mathcal{Z})\sigma_j(\mathbf{I}_{n_2}\overline{\mathcal{V}}^{(j)}(\overline{\mathcal{V}}^{(j)})^*), \tag{99}$$

where $\mathbf{I}_{n_1}$ and $\mathbf{I}_{n_2}$ are identity matrices of size $n_1 \times n_1$ and $n_2 \times n_2$, respectively.

Note that, for $1 \le j \le \min\{n_1, n_2\}$,

$$\sigma_j(\mathbf{I}_{n_1}\overline{\mathcal{U}}^{(j)}(\overline{\mathcal{U}}^{(j)})^*) \le 1, \quad \sigma_j(\mathbf{I}_{n_2}\overline{\mathcal{V}}^{(j)}(\overline{\mathcal{V}}^{(j)})^*) \le 1. \tag{100}$$

Then we have

$$\sum_{j=1}^{\min\{n_1,n_2\}}\sigma_j(\overline{\mathcal{P}_{\mathbf{T}^\perp}\mathcal{H}}^{(i)}) \le \sum_{j=1}^{\min\{n_1,n_2\}}\sigma_j(\mathcal{Z}). \tag{101}$$

Further, we have

$$\|\mathcal{H}\|_* \ge \|\mathcal{P}_{\mathbf{T}^\perp}\mathcal{H}\|_*. \tag{102}$$

□

Now we begin to prove Theorem 1.

Proof The deduction is divided into two steps as follows.

- *Step 1*: Under the assumption of $\|\mathcal{P}_{\mathbf{\Omega}}\mathcal{P}_{\mathbf{T}}\| \le \frac{1}{2}$ and $\lambda = \frac{1}{\sqrt{\max\{n_1,n_2\}n_3}}$, we prove that $(\mathcal{X}_0, \mathcal{E}_0)$ is the (ε, δ)-asymptotic unique local solution to the proposed CCSVS model Eq. (15) if there exists $(\mathcal{W}, \mathcal{F})$ such that

$$\omega\Phi \otimes (\mathcal{U} * \mathcal{V}^* + \mathcal{W}) = \lambda(\mathrm{sgn}(\mathcal{E}_0) + \mathcal{F} + \mathcal{P}_{\mathbf{\Omega}}\mathcal{D}), \tag{103}$$

where $\Phi = diag([\vartheta_1, \vartheta_2, \cdots, \vartheta_{n_3}]) \in \mathbb{R}^{n_3\times n_3}$ and

$$\mathcal{P}_{\mathbf{T}}\mathcal{W} = 0, \|\mathcal{W}\| \le \epsilon_0, \mathcal{P}_{\mathbf{\Omega}}(\mathcal{F}) = 0, \|\mathcal{F}\|_\infty \le \epsilon_0 \text{ and } \|\mathcal{P}_{\mathbf{\Omega}}\mathcal{D}\|_F \le \frac{1}{4}, \tag{104}$$

in which ϵ_0 is a positive constant such that $0 < \epsilon_0 < \frac{3}{4}$.

- *Step 2*: On the basis of *Step 1*, it suffices to produce a dual certification $\mathcal{W}$ which satisfies

$$\begin{cases} \mathcal{W} \in \mathbf{T}^\perp, \\ \|\mathcal{W}\| \le \epsilon_0, \\ \|\mathcal{P}_{\mathbf{\Omega}}(\omega\Theta \otimes (\mathcal{U} * \mathcal{V}^* + \mathcal{W}) - \lambda\mathrm{sgn}(\mathcal{E}_0))\|_F \le \frac{\lambda}{4}, \\ \|\mathcal{P}_{\mathbf{\Omega}^\perp}(\omega\Theta \otimes (\mathcal{U} * \mathcal{V}^* + \mathcal{W}))\|_\infty \le \lambda\epsilon_0. \end{cases} \tag{105}$$

First of all, we complete the *Step 1*. For any $\mathcal{H} \neq 0$, $(\mathcal{X}_0 + \mathcal{H}, \mathcal{E}_0 - \mathcal{H})$ is a feasible point because $(\mathcal{X}_0 + \mathcal{H}) + (\mathcal{E}_0 - \mathcal{H}) = \mathcal{X}_0 + \mathcal{E}_0$. To prove that is the (ε, δ)-asymptotic unique local solution (See Definition 6 for details.) to the proposed CCSVS model Eq. (15), we need to show

$$\mathrm{Sep}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) + \lambda\|\mathcal{E}_0 - \mathcal{H}\|_1 > \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) + \lambda\|\mathcal{E}_0\|_1 - \delta \tag{106}$$

for any $\mathcal{H} \in \mathrm{B}(0, \varepsilon) - \{0\}$ in the sense of tensor nuclear norm.

Note that

$$\mathrm{Sep}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) = \mathrm{Sep}_{ns}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) - \sum_{i=1}^{n_3}\vartheta_i\sum_{j=1}^{r}\widetilde{\omega}_j\sigma_j(\overline{\mathcal{X}}_0^{(i)} + \overline{\mathcal{H}}^{(i)}), \tag{107}$$

where $\widetilde{\omega}_j = \omega + \omega_j$, $j = 1, 2, \cdots, r$ and $r < \min\{n_1, n_2\}$.

For any $\omega\Phi \otimes (\mathcal{U} * \mathcal{V}^* + \mathcal{W}_0) \in \partial\mathrm{Sep}_{ns}(\mathcal{X}_0, \Omega, \vartheta)$ and $\mathrm{sgn}(\mathcal{E}_0) + \mathcal{F}_0 \in \partial\|\mathcal{E}_0\|_1$, we have

$$\begin{aligned}\mathrm{Sep}_{ns}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) + \quad &\lambda\|\mathcal{E}_0 - \mathcal{H}\|_1 \\ &\ge \mathrm{Sep}_{ns}(\mathcal{X}_0, \Omega, \vartheta) + \langle\omega\Phi \otimes (\mathcal{U} * \mathcal{V}^* + \mathcal{W}_0), \mathcal{H}\rangle \\ &\quad + \lambda\|\mathcal{E}_0\|_1 s - \lambda\langle\mathrm{sgn}(\mathcal{E}_0) + \mathcal{F}_0, \mathcal{H}\rangle.\end{aligned} \tag{108}$$

By the intermediate result in the proof of Theorem 1, we have

$$\sum_{i=1}^{n_3}\vartheta_i\sum_{j=1}^{r}\widetilde{\omega}_j\sigma_j(\overline{\mathcal{X}}_0^{(i)} + \overline{\mathcal{H}}^{(i)}) \le \sum_{i=1}^{n_3}\vartheta_i\sum_{j=1}^{r}\widetilde{\omega}_j\sigma_j(\overline{\mathcal{X}}_0^{(i)}) + \sum_{i=1}^{n_3}\vartheta_i\sum_{j=1}^{r}\widetilde{\omega}_j\sigma_j(\overline{\mathcal{H}}^{(i)}). \tag{109}$$

Therefore

$$\begin{aligned}&\mathrm{Sep}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) + \lambda\|\mathcal{E}_0 - \mathcal{H}\|_1 \\ &\ge \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) + \langle\omega\Phi \otimes (\mathcal{U} * \mathcal{V}^* + \mathcal{W}_0), \mathcal{H}\rangle \\ &\quad + \lambda\|\mathcal{E}_0\|_1 - \lambda\langle\mathrm{sgn}(\mathcal{E}_0) + \mathcal{F}_0, \mathcal{H}\rangle\end{aligned}$$

$$
\begin{aligned}
&- \sum_{i=1}^{n_3} \vartheta_i \sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j(\overline{\mathcal{H}}^{(i)}) \\
&= \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) + \langle \mathcal{U} * \mathcal{V}^* + \mathcal{W}_0, \omega\Phi \otimes \mathcal{H} \rangle \\
&\quad + \lambda \|\mathcal{E}_0\|_1 - \lambda \langle \mathrm{sgn}(\mathcal{E}_0) + \mathcal{F}_0, \mathcal{H} \rangle \\
&\quad - \sum_{i=1}^{n_3} \vartheta_i \sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j(\overline{\mathcal{H}}^{(i)}). \qquad (110)
\end{aligned}
$$

According to the property of dual norm on the tensor nuclear norm and l_1-norm,

$$
\sup_{\|\mathcal{W}_0\| \le 1} \langle \mathcal{W}_0, \omega\Phi \otimes \mathcal{H} \rangle = \|\omega\Phi \otimes \mathcal{H}\|_*. \qquad (111)
$$

and

$$
\sup_{\|\mathcal{F}_0\|_\infty \le 1} -\langle \mathcal{F}_0, \mathcal{H} \rangle = \|\mathcal{H}\|_1. \qquad (112)
$$

Note that

$$
\|\mathcal{H}\|_1 \ge \|\mathcal{P}_{\mathbf{\Omega}^\perp} \mathcal{H}\|_1. \qquad (113)
$$

Then we have

$$
\sup_{\|\mathcal{F}_0\|_\infty \le 1} -\langle \mathcal{F}_0, \mathcal{H} \rangle \ge \|\mathcal{P}_{\mathbf{\Omega}^\perp} \mathcal{H}\|_1. \qquad (114)
$$

Further, taking the supreme for the right-hand side of Eq. (110) over the set $\{\mathcal{W} : \|\mathcal{W}\| \le 1\}$ leads to

$$
\begin{aligned}
&\mathrm{Sep}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) + \lambda \|\mathcal{E}_0 - \mathcal{H}\|_1 \\
&\quad \ge \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) + \langle \mathcal{U} * \mathcal{V}^*, \omega\Phi \otimes \mathcal{H} \rangle \\
&\qquad + \lambda \|\mathcal{S}_0\|_1 - \lambda \langle \mathrm{sgn}(\mathcal{E}_0), \mathcal{H} \rangle \\
&\qquad - \sum_{i=1}^{n_3} \vartheta_i \sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j(\overline{\mathcal{H}}^{(i)}) + \sup_{\|\mathcal{W}\| \le 1} \langle \mathcal{W}, \omega\Phi \otimes \mathcal{H} \rangle \\
&\qquad + \lambda \sup_{\|\mathcal{F}\|_\infty \le 1} \langle -\mathcal{F}, \mathcal{H} \rangle \\
&\quad = \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) + \langle \omega\Phi \otimes (\mathcal{U} * \mathcal{V}^*), \mathcal{H} \rangle \\
&\qquad + \lambda \|\mathcal{E}_0\|_1 - \lambda \langle \mathrm{sgn}(\mathcal{E}_0), \mathcal{H} \rangle \\
&\qquad - \sum_{i=1}^{n_3} \vartheta_i \sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j(\overline{\mathcal{H}}^{(i)}) + \sup_{\|\mathcal{W}\| \le 1} |\langle \mathcal{W}, \omega\Phi \otimes \mathcal{H} \rangle| \\
&\qquad + \lambda \sup_{\|\mathcal{F}\|_\infty \le 1} |\langle \mathcal{F}, \mathcal{H} \rangle|. \qquad (115)
\end{aligned}
$$

By the assumption on Eq. (103),

$$
\begin{aligned}
&\langle \omega\Phi \otimes (\mathcal{U} * \mathcal{V}^*) - \lambda \mathrm{sgn}(\mathcal{E}_0), \mathcal{H} \rangle \\
&\quad = -\langle \mathcal{W}, \omega\Phi \otimes \mathcal{H} \rangle + \lambda \langle \mathcal{F}, \mathcal{H} \rangle + \lambda \langle \mathcal{P}_{\mathbf{\Omega}} \mathcal{D}, \mathcal{H} \rangle \\
&\quad \ge - \sup_{\|\mathcal{W}\| \le \epsilon_0} |\langle \mathcal{W}, \omega\Phi \otimes \mathcal{H} \rangle| - \lambda \sup_{\|\mathcal{F}\|_\infty \le \epsilon_0} |\langle \mathcal{F}, \mathcal{H} \rangle| \\
&\qquad + \lambda \langle \mathcal{P}_{\mathbf{\Omega}} \mathcal{D}, \mathcal{P}_{\mathbf{\Omega}} \mathcal{H} \rangle \\
&\quad \ge -\epsilon_0 \sup_{\|\widehat{\mathcal{W}}\| \le 1} |\langle \widehat{\mathcal{W}}, \omega\Phi \otimes \mathcal{H} \rangle| - \lambda\epsilon_0 \\
&\qquad \sup_{\|\widehat{\mathcal{F}}\|_\infty \le 1} |\langle \widehat{\mathcal{F}}, \mathcal{H} \rangle| - \frac{\lambda}{4} \|\mathcal{P}_{\mathbf{\Omega}} \mathcal{H}\|_F \qquad (116)
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\mathrm{Sep}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) + \lambda \|\mathcal{E}_0 - \mathcal{H}\|_1 \\
&\quad \ge \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) + \lambda \|\mathcal{E}_0\|_1 \\
&\qquad - \sum_{i=1}^{n_3} \vartheta_i \sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j(\overline{\mathcal{H}}^{(i)}) + \sup_{\|\mathcal{W}\| \le 1} \langle \mathcal{W}, \omega\Phi \otimes \mathcal{H} \rangle \\
&\qquad + \frac{\lambda}{2} \sup_{\|\mathcal{F}\|_\infty \le 1} \langle -\mathcal{F}, \mathcal{H} \rangle \\
&\quad \ge \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) + \lambda \|\mathcal{E}_0\|_1 \\
&\qquad - \sum_{i=1}^{n_3} \vartheta_i \sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j(\overline{\mathcal{H}}^{(i)}) + (1 - \epsilon_0) \\
&\qquad \left(\|\omega\Phi \otimes \mathcal{H}\|_* + \lambda \|\mathcal{P}_{\mathbf{\Omega}^\perp} \mathcal{H}\|_1 \right) - \frac{\lambda}{4} \|\mathcal{P}_{\mathbf{\Omega}} \mathcal{H}\|_F. \qquad (117)
\end{aligned}
$$

Note that

$$
\begin{aligned}
\|\mathcal{P}_{\mathbf{\Omega}} \mathcal{H}\|_F &\le \|\mathcal{P}_{\mathbf{\Omega}} \mathcal{P}_{\mathbf{T}} \mathcal{H}\|_F + \|\mathcal{P}_{\mathbf{\Omega}} \mathcal{P}_{\mathbf{T}^\perp} \mathcal{H}\|_F \\
&\le \frac{1}{2} \|\mathcal{H}\|_F + \|\mathcal{P}_{\mathbf{T}^\perp} \mathcal{H}\|_F \\
&\le \frac{1}{2} \|\mathcal{P}_{\mathbf{\Omega}} \mathcal{H}\|_F + \frac{1}{2} \|\mathcal{P}_{\mathbf{\Omega}^\perp} \mathcal{H}\|_F + \|\mathcal{P}_{\mathbf{T}^\perp} \mathcal{H}\|_F. \qquad (118)
\end{aligned}
$$

which implies that

$$
\begin{aligned}
\|\mathcal{P}_{\mathbf{\Omega}} \mathcal{H}\|_F &\le \|\mathcal{P}_{\mathbf{\Omega}} \mathcal{H}\|_F + 2 \|\mathcal{P}_{\mathbf{T}^\perp} \mathcal{H}\|_F \\
&\le \|\mathcal{P}_{\mathbf{\Omega}^\perp} \mathcal{H}\|_1 + 2\sqrt{n_3} \|\mathcal{P}_{\mathbf{T}^\perp} \mathcal{H}\|_* \\
&\le \|\mathcal{P}_{\mathbf{\Omega}^\perp} \mathcal{H}\|_1 + 2\sqrt{n_3} \|\mathcal{H}\|_* \qquad (119)
\end{aligned}
$$

The last inequality holds by Lemma 3.

Then

$$
\begin{aligned}
&\mathrm{Sep}(\mathcal{X}_0 + \mathcal{H}, \Omega, \vartheta) + \lambda \|\mathcal{E}_0 - \mathcal{H}\|_1 \\
&\ge \mathrm{Sep}(\mathcal{X}_0, \Omega, \vartheta) + \lambda \|\mathcal{E}_0\|_1 \\
&\underbrace{- \sum_{i=1}^{n_3} \vartheta_i \sum_{j=1}^{r} \widetilde{\omega}_j \sigma_j(\overline{\mathcal{H}}^{(i)}) + (1 - \epsilon_0) \|\omega\Phi \otimes \mathcal{H}\|_* - \frac{\lambda\sqrt{n_3}}{2} \|\mathcal{H}\|_*}_{I_0} \\
&\underbrace{+ \left(\frac{3}{4} - \epsilon_0 \right) \lambda \|\mathcal{P}_{\mathbf{\Omega}^\perp} \mathcal{H}\|_1}_{I_1}. \qquad (120)
\end{aligned}
$$

As a matter of fact,

$$I_0 = \sum_{i=1}^{n_3}\sum_{j=1}^{r}\kappa_{ij}\sigma_j(\overline{\mathcal{H}}^{(i)}), \tag{121}$$

where

$$\kappa_{ij} = \frac{1}{n_3}\left((1-\epsilon_0)\omega\vartheta_i - \frac{\lambda\sqrt{n_3}}{2}\right) - \vartheta_i\widetilde{\omega}_j. \tag{122}$$

Then

$$|\kappa_{ij}| \le 3\vartheta_{\max}\omega_{\max} + \frac{1}{n_3\sqrt{\max\{n_1,n_2\}}}. \tag{123}$$

By given condition $\|\mathcal{H}\|_* \le \varepsilon$, we have

$$|I_0| \le \varepsilon\left(3\vartheta_{\max}\omega_{\max} + \frac{1}{n_3\sqrt{\max\{n_1,n_2\}}}\right) = \delta. \tag{124}$$

Recall that $\epsilon_0 < \frac{3}{4}$. $\mathcal{H} \neq 0$ means that all singular values $\sigma_j(\overline{\mathcal{H}}^{(i)})$ are non-zero. Hence, I_1 must be positive, i.e. $I_1 > 0$. Then

$$\begin{aligned}&\mathrm{Sep}(\mathcal{X}_0+\mathcal{H},\Omega,\vartheta) + \lambda\|\mathcal{E}_0-\mathcal{H}\|_1\\ &> \mathrm{Sep}(\mathcal{X}_0,\Omega,\vartheta) + \lambda\|\mathcal{E}_0\|_1 - \delta.\end{aligned} \tag{125}$$

which holds for any $\mathcal{H} \in B(0,\varepsilon) - \{0\}$ in the sense of tensor nuclear norm.

The remaining *Step 2* is similar to the dual certification step in Lu et al. (2020) and omitted here. □

References

Barber, R. F., & Sidky, E. Y. (2020). Convergence for nonconvex ADMM, with applications to CT imaging. arXiv:2006.07278

Belarbi, M.A., Mahmoudi, S., & Belalem, G. (2017). PCA as dimensionality reduction for large-scale image retrieval systems. *Proceedings of International Joint Conference on Artificial Intelligence*.

Bhatia, R. (2013). *Matrix Analysis*. Springer.

Bouwmans, T., Javed, S., Zhang, H., Lin, Z., & Otazo, R. (2018). On the applications of robust PCA in image and video processing. *Proceedings of the IEEE, 106*(8), 1427–1457.

Candès, E. J., Li, X. D., Ma, Y., & Wright, J. (2011). Robust principal component analysis? *Journal of the ACM*, **58**(3).

De la Torre, F., & Black, M. J. (2001). Robust principal component analysis for computer vision. In *Proceedings of the IEEE International Conference on Computer Vision*.

Fan, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proceedings of the National Academy of Sciences, 37*(11), 760–766.

Gao, Q., Zhang, P., Xia, W., Xie, D., Gao, X., & Tao, D. (2020). Enhanced tensor RPCA and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(6), 2133–2140.

Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X., & Zhang, L. (2017). Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision, 121*(2), 183–208.

Imaizumi, M., & Maehara, T. (2017). On tensor train rank minimization: statistical efficiency and scalable algorithm. In *Proceedings of Advances in Neural Information Processing Systems*.

Kilmer, M. E., & Martin, C. D. (2011). Factorization strategies for third-order tensors. *Linear Algebra and Its Applications, 435*(3), 641–658.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review, 51*(3), 455–500.

Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., & Ma, Y. (2013). Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Recognition and Machine Intelligence, 35*(1), 171–185.

Liu, J., Musialski, P., Wonka, P., & Ye, J. (2013). Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Recognition and Machine Intelligence, 35*(1), 208–220.

Lu, C., Feng, J., Liu, W., Lin, Z., & Yan, S. (2020). Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Recognition and Machine Intelligence, 42*(4), 925–938.

Lu, C., Feng, J., Yan, S., & Lin, Z. (2018). A unified alternating direction method of multipliers by majorization minimization. *IEEE Transactions on Pattern Recognition and Machine Intelligence, 40*(3), 527–541.

Lu, H., Plataniotis, K. N., & Venetsanopoulos, A. N. (2008). MPCA: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks, 19*(1), 18–39.

Maddalena, L., & Petrosino, A. (2015). Towards benchmarking scene background initialization. In *New Trends in Image Analysis and Processing–ICIAP 2015 Workshops: ICIAP 2015 International Workshops*.

Malik, O.A., & Becker, S. (2018). Low-rank Tucker decomposition of large tensors using TensorSketch. In *Proceedings of Advances in Neural Information Processing Systems*.

Marshall, A., & Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications*. Academic.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in *Proceedings Eighth IEEE International Conference on Computer Vision*,

Mirsky, L. (1975). A trace inequality of John von Neumann. *Monatshefte für Mathematik, 79*(4), 303–306.

Moslehian, M. (2012). S, Ky fan inequalities. *Linear and multilinear algebra, 60*(11–12), 1313–1325.

Rockafellar, R. T., & Wets, R. (1998). *Variational analysis*. Springer.

Schölkopf, B., Smola, A., & Müller, K.-R. (1997). Kernel principal component analysis. In *Artificial Neural Networks-ICANN*.

Sun, M., Zhao, L., Zheng, J., & Xu. J. (2020). A Nonlocal Denoising Framework Based on Tensor Robust Principal Component Analysis with l_p norm. In *Proceedings of the IEEE International Conference on Big Data*.

Von Neumann, J. (1937). *Some matrix-inequalities and metrization of matric space*.

Wang, Y., Yin, W., & Zeng, J. (2008). Linear convergence of iterative soft-thresholding. *Journal of Fourier Analysis and Applications, 14*, 813–837.

Wang, Y., Yin, W., & Zeng, J. (2019). Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing, 78*, 29–63.

Wang, L., Zhang, S., & Huang, H. (2021). Adaptive dimension-discriminative low-rank tensor recovery for computational hyper-

spectral imaging. *International Journal of Computer Vision, 129*(10), 2907–2926.

Yu, H., & Bennamoun, M. (2006). 1D-PCA, 2D-PCA to nD-PCA. In *18th International Conference on Pattern Recognition (ICPR'06)*.

Zhang, Z., Ely, G., Aeron, S., Hao, N., & Kilmer, M. (2014). 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*.

Zhang, L., & Peng, Z. (2019). Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sensing, 11*(4), 382-392.

Zhou, Y., & Cheung, Y.-M. (2021). Bayesian low-tubal-rank robust tensor factorization with multi-rank determination. *IEEE Transacrions on Pattern Recognition and Machine Intelligence, 43*(1), 62–76.