Hyperspectral Image Recovery Using Non-Convex Low-Rank Tensor Approximation

Hongyi Liu 1, Hanyang Li 1, Zebin Wu 2* and Zhihui Wei 2

1 School of Science, Nanjing University of Science and Technology, Nanjing 210094, China; hyliu@njust.edu.cn (H.L.); lhy_holly@njust.edu.cn (H.L.)
2 School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; gswei@njust.edu.cn
* Correspondence: wuzb@njust.edu.cn

Received: 01 June 2020; Accepted: 13 July 2020; Published: 15 July 2020

Abstract: Low-rank tensors have received more attention in hyperspectral image (HSI) recovery. Minimizing the tensor nuclear norm, as a low-rank approximation method, often leads to modeling bias. To achieve an unbiased approximation and improve the robustness, this paper develops a non-convex relaxation approach for low-rank tensor approximation. Firstly, a non-convex approximation of tensor nuclear norm (NCTNN) is introduced to the low-rank tensor completion. Secondly, a non-convex tensor robust principal component analysis (NCTRPCA) method is proposed, which aims at exactly recovering a low-rank tensor corrupted by mixed-noise. The two proposed models are solved efficiently by the alternating direction method of multipliers (ADMM). Three HSI datasets are employed to exhibit the superiority of the proposed model over the low rank penalization method in terms of accuracy and robustness.

Keywords: hyperspectral image (HSI); non-convex relaxation; tensor completion; tensor robust principal analysis

1. Introduction

Hyperspectral image (HSI) has drawn a lot of attention on account of its rich spectral information. However, HSI can be easily disturbed by many external factors, such as missing entries, noise, and so on, which not only degrade the visual quality of the image but also limit the precision of the subsequent image interpretation and analysis. Therefore, HSI restoration has attracted an increasing interest in recent years and various HSI restoration models have been developed [1–3]. Among them, considering the spatial geometric similarity and spectral correlation of HSI, low-rank prior modeling has become one of the most popular technologies [4]. Besides, to enhance the structural details, a total variation regularized low-rank matrix factorization (LRTV) restoration model was proposed in Ref. [5], which obtains satisfying restoration results in case of Gaussian noise. As for removing the mixed-noise, enhancing edge and spectral signatures, a weighted total variation regularized low-rank (LRWTVC) model was proposed [6]. Candes et al. employed robust principle component analysis (RPCA) to separate clean HSI from corrupted sparse noise in polynomial time [7]. Chang et al. established a low-rank-based single-image decomposition model (LRSID) was established to remove the stripe noise in HSI [1].

With the development of low-rank approaches, non-convex low-rank approximation approaches, including weighted nuclear norm minimization (WNNM) [8], Schatten p-norm minimization [9], weighted Schatten p-norm minimization (WSNM) [10] and minimax concave penalty (MCP) function [11], etc., have also been studied due to the robustness and unbiasedness. Wang et al. [12] developed the MCP function for the rank minimization problem and to obtain a so-called γ-norm as a non-convex relaxation of the matrix rank. A non-convex relaxation HSI low-rank
recovery with TV regularized (NRLRWTV) model [13] was proposed in order to achieve high estimation for HSI as well as capture spatial and spectral information. NMC [14] was proposed as a non-convex matrix completion method to recover the observed image. By reweighting singular values of matrices reconstructed from HSI, a non-convex, non-smooth minimization denoising method, named as the iteratively reweighted nuclear norm (IRNN) was developed [15].

However, as HSI is 3D cube data, matrix-based methods may destroy its intrinsic spectral and spatial correlations. Therefore, tensor representation is proposed and low-rank tensor prior is also introduced to make full use of the inherent structure of HSI [16–24]. Different from the matrix rank, the tensor rank is diverse according to the non-uniqueness of tensor decomposition [16]. One of the most popular techniques to define the tensor rank is to take the unfolding matrices of the tensor into consideration. In this case, tensor n-rank based on Tucker decomposition (TD) was developed and the low-rank tensor completion (LRTC) model was proposed to recover the missing values in visual data [17]. As singular value decomposition (SVD) used in LRTC results in more time consumption or even non-applicability for large-scale problems, Xu proposed the TMac [22] method, which applies low-rank matrix factorization to each unfolding mode of the tensor. Although the n-rank can flexibly exploit the correlation along each mode by adjusting the weights [23], as pointed out in [24,25], unfolding the tensor as a matrix destroys the intrinsic structure of the tensor. Therefore, tensor tubal rank (multi-rank) based on tensor SVD (t-SVD) has received considerable attention [22,26–34]. In fact, the tubal-rank (multi-rank) is an NP-hard problem, and tensor nuclear norm (TNN) is developed as the convex surrogate in [26], and the exactly-recover-property for TNN-based tensor robust principle component analysis (TRPCA) model is also proved. Zhang et al. [30,31] proposed the t-SVD based tensor completion and TRPCA models in video restoration. Hu et al. [32] proposed a twist tensor nuclear norm (t-TNN) for video completion.

Although a low-rank of tensor model has been widely used, they are also based on matrix rank, which depends on the nuclear norm of the matrix. Therefore, for tensor low-rank approximation, the non-convex technique used in matrix can be extended to tensor [34,35]. Followed by this idea, several tensor low-rank relaxation methods have been developed. For instance, Cao [36] proposed a new tensor completion model via folded-concave penalty for estimating missing values in tensor data, in which tensor n-rank is substituted. A non-convex logDet function, as a smooth approximation for tensor rank instead of the convex tensor nuclear norm, was introduced [37]. The low-rank tensor completion using the MCP function as a relaxation of tensor tubal rank was proposed in our previous paper [38]. Zhang [39] proposed a novel non-convex relaxation tensor completion method, together with the Proximal Linear Minimization (PLM) algorithm. Cai [40] developed a t-Gamma norm-based non-convex tensor approximation, which can better keep the low-rank property of HSI. In Ref. [41], the Non-Convex Regularized Tensor (NORT) tensor completion method was proposed to improve the computation efficiency. Li [42] introduced a non-convex penalty of tensor rank based on t-SVD, which can effectively avoid the bias of $l_1$–norm. However, by integrating a non-convex function in [43] with an overlapped nuclear norm [44], the operator in NORT was computed by the unfolding matrix of the tensor, which can destroy the intrinsic structure of the tensor.

Considering the problems mentioned above in low-rank approximation and HSI restoration, we propose non-convex low-rank tensor approximation in two cases for HSI restoration. The main process of this paper is summarized in Figure 1, and the paper is organized as follows. Section 2 introduces some notations and preliminaries of tensors, where several algebraic structures of 3-order tensors are defined. In Section 3, the proposed non-convex tensor nuclear norm (tensor $\gamma$-norm) (NCTNN) penalized tensor completion is developed. Meanwhile, a non-convex low-rank approximation is merged into the tensor robust PCA (NCTRPCA), which separates low multi-rank components from sparse corruptions. Furthermore, the alternating direction method of multipliers (ADMM) algorithm [45] is developed to solve the two optimization models. Section 4 presents the simulation and real experiments for recovering HSI with limited sampled data and noise corruptions, meanwhile, parameters analysis and discussions are also given. Conclusions are drawn in Section 5.
2. Notations and Preliminaries

Throughout this paper, we use $\mathbf{X}, \mathbf{x}$, $\mathbf{x}$ and $\mathbf{x}$ to denote tensor, matrix, vector and scalar, respectively. For a 3-order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the $(i, j, k)$-th entry is denoted as $X_{ijk}$ and $X_{(i)}$ is the mode-$i$ matricization. We use the Matlab notation $\mathbf{X}(i,:,:),\mathbf{X}(:,i,:)\text{ and }\mathbf{X}(;,:;i)$ to denote, respectively, the $i$-th horizontal, lateral and frontal slice. More often, the frontal slice $\mathbf{X}(;,:;i)$ is denoted as $\mathbf{X}^{(i)}$. The tube is denoted as $\mathbf{X}(i,j,:)$. The inner product of matrices $\mathbf{X}$ and $\mathbf{Y}$ in $\mathbb{R}^{n_1 \times n_2}$ is defined as $<\mathbf{X},\mathbf{Y}> = \sum_{r=1}^{n_3} <\mathbf{X}^{(i)},\mathbf{Y}^{(i)}>$. 

Some norms of vector, matrix and tensor are used. We denote the $l_1$-norm, Frobenius norm and nuclear norm as $||\mathbf{X}||_1 = \sum_{ijk}|X_{ijk}|$, $||\mathbf{X}||_F = \sqrt{\sum_{ijk}|X_{ijk}|^2}$ and $||\mathbf{X}||_* = \sum_i ||\mathbf{X}^{(i)}||_*$, respectively. The above norms reduce to the vector or matrix norms if $\mathbf{X}$ is a vector or a matrix. Specifically, the nuclear norm of the matrix $\mathbf{X}$ is defined as $||\mathbf{X}||_* = \sum_i \sigma_i(\mathbf{X})$, in which $\sigma_i(\mathbf{X})$ is the singular values of matrix $\mathbf{X}$ with $\sigma_1 \gg \sigma_2 \gg \cdots \sigma_r$, and $r$ is the rank of matrix $\mathbf{X}$.

The Fourier transformation of tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is denoted by $\hat{\mathbf{X}}$. In particular, we denote $\hat{\mathbf{X}}$ as a block diagonal matrix with each block on diagonal as the frontal slice $\hat{\mathbf{X}}^{(i)}$ of $\hat{\mathbf{X}}$, i.e.,

$$\hat{\mathbf{X}} = \text{blkdiag}(\hat{\mathbf{X}}) = \begin{bmatrix} \hat{\mathbf{X}}^{(1)} \\ \hat{\mathbf{X}}^{(2)} \\ \vdots \\ \hat{\mathbf{X}}^{(n_3)} \end{bmatrix}$$

The block circulant matrix of tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ has size $n_1 n_3 \times n_2 n_3$, i.e.,

$$\text{brcirc}(\mathbf{X}) = \begin{bmatrix} \mathbf{X}^{(1)} & \mathbf{X}^{(n_3)} & \mathbf{X}^{(2)} \\ \mathbf{X}^{(2)} & \mathbf{X}^{(1)} & \mathbf{X}^{(3)} \\ \vdots & \vdots & \vdots \\ \mathbf{X}^{(n_3)} & \mathbf{X}^{(n_3-1)} & \mathbf{X}^{(1)} \end{bmatrix}$$

We also define the following operator

$$\text{unfold}(\mathbf{X}) = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \\ \vdots \\ \mathbf{X}^{(n_3)} \end{bmatrix}, \text{fold}\left(\text{unfold}(\mathbf{X})\right) = \mathbf{X}$$

Then, t-product between two 3-order tensors can be defined as follows.

**Definition 1. (t-product)** [19]: The t-product $\mathbf{X} \ast \mathbf{Y}$ of two tensors $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathbf{Y} \in \mathbb{R}^{n_2 \times n_3 \times n_4}$ is a tensor of size $n_1 \times n_3 \times n_4$, and $\mathbf{X} \ast \mathbf{Y} = \text{fold}\left(\text{brcirc}(\mathbf{X}) \cdot \text{unfold}(\mathbf{Y})\right)$.

**Definition 2. (f-diagonal tensor)** [19]: A tensor is called f-diagonal if each of its frontal slices is a diagonal matrix.

**Theorem 1. (t-SVD)** [19]: For a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD of $\mathbf{X}$ is given by $\mathbf{X} = \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathbf{V} \in \mathbb{R}^{n_2 \times n_3 \times n_3}$ are orthogonal, and $\mathbf{S}$ is a rectangular f-diagonal tensor of size $n_1 \times n_2 \times n_3$. 

Figure 1. The flowchart of the proposed method.
Definition 3. (Tensor multi rank and tubal rank) [33]: The tensor multi rank of $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a vector $r \in \mathbb{R}^{n_3}$ with its $i$-th entry as the rank of the $i$-th frontal slice of $\mathbf{X}$, i.e., $r_m = \text{rank}(\mathbf{X}^{(i)})$.

The tensor tubal rank, denoted as $\text{rank}_t(\mathbf{X})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$, where $\mathcal{S}$ is from the t-SVD of $\mathbf{X} = \mathbf{U} * \mathcal{S} * \mathbf{V}^T$.

That is

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

Definition 4. (TNN norm) [33]: Another definition of the tensor nuclear norm of $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\|\mathbf{X}\|_{\text{TNN}}$, is defined as the average of the nuclear norm of all the frontal slices of $\mathbf{X}$, i.e.,

$$\|\mathbf{X}\|_{\text{TNN}} = \frac{1}{n_3} \sum_{k=1}^{n_3} \|\mathbf{X}^{(k)}\|_*$$

which can be rewritten as

$$\|\mathbf{X}\|_{\text{TNN}} = \frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{r=1}^{r_k} \sigma_r(\mathbf{X}^{(k)}).$$

3. Proposed Methods

Minimizing the rank is equal to minimizing the $l_0$-norm; it is a NP-hard problem and can be solved by convex or non-convex relaxation methods. Nuclear norm and weighted nuclear norm are common convex approximations, while the convex surrogate will over-penalize larger singular values, resulting in the approximation deviation. Therein, non-convex surrogate function, which can keep large singular values unchanged, has a nearly unbiased approximation to the $l_0$-norm.

3.1. Non-Convex Low-Rank Relaxation

As it has been investigated in [10], the $\gamma$-norm $\|\mathbf{X}\|_\gamma$ ($\gamma$ is the positive factor) of matrix $\mathbf{X}$ with rank $r$ is defined as the sum of MCP functions of the matrix singular values

$$\|\mathbf{X}\|_\gamma = \sum_{i=1}^{r} \psi_{\text{MCP}}(\sigma_i(\mathbf{X}))$$

With

$$\psi_{\text{MCP}}(t) = \begin{cases} \frac{\gamma t^2}{2}, & |t| > \gamma \lambda \\ \lambda |t| - \frac{t^2}{2\gamma}, & \text{otherwise} \end{cases}$$

where $\psi_{\text{MCP}}(\cdot)$ is the so-called MCP function, $\lambda$ is a threshold parameter.

It can be derived that when $\gamma \to \infty$, the non-convex MCP function tends to $l_1$-norm and it satisfies $\psi_{\text{MCP}}(\mathbf{X}) \ll \lambda \|\mathbf{X}\|_{l_1}$, which leads to a nearly optimal non-convex approximation to $l_0$-norm. As such, the derived matrix $\gamma$-norm defined by the MCP function with regard to singular values generates a nearly optimal non-convex approximation to low-rank (NRLR). The solution of $\gamma$-norm results in singular value soft thresholding (SVT). Compared with the popular singular values thresholding (SVT) of matrix nuclear norm, SVST keeps large singular values of a matrix unchanged, which has the advantages of nearly unbiasedness and more accuracy, as illustrated in Figure 2.
Figure 2. Non-convex soft shrinkage and soft shrinkage ($\lambda = 1$).

Inspired by matrix $\gamma$-norm, for a third-order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we define the tensor $\gamma$-norm as follows:

$$
||\mathbf{X}||_\gamma = \frac{1}{n_3} \sum_{k=1}^{n_3} ||\tilde{\mathbf{X}}^{(k)}||_\gamma = \frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{r=1}^{r_k} \psi_{\text{MCP}}(\sigma_{kr}(\tilde{\mathbf{X}}^{(k)}))
$$

where $\sigma_{kr}(\tilde{\mathbf{X}}^{(k)})$ stands for the $r$-th singular values of the matrix. The defined tensor $\gamma$-norm is the average of $\gamma$-norm of all the frontal slices of $\mathbf{X}$, and can be computed by the sum of MCP function with singular values of $\tilde{\mathbf{X}}^{(k)}$ as variables, which is a non-convex relaxation of the matrix nuclear norm. Therefore, the defined tensor $\gamma$-norm can be regarded as a nearly unbiased non-convex relaxation of tensor rank, which leads to better approximation to tensor rank than the TNN norm or other convex relaxation methods. Moreover, the proposed tensor $\gamma$-norm has some good properties that can be summarized in Theorem 2.

**Theorem 2.** For a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, it has the decomposition of t-SVD with $\mathbf{X} = \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathbf{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, and $\mathbf{S}$ is a rectangular $f$-diagonal tensor of size $n_1 \times n_2 \times n_3$. The proposed tensor norm $||\mathbf{X}||_\gamma$ satisfies:

(i) $||\mathbf{X}||_\gamma \geq 0$, $||\mathbf{X}||_\gamma = 0$ iff $\mathbf{X} = 0$.
(ii) $||\mathbf{X}||_\gamma$ increases by $\gamma$.
(iii) $||\mathbf{X}||_\gamma$ is a concave function of $\sigma_{kr}(\tilde{\mathbf{X}}^{(k)})$.
(iv) $||\mathbf{X}||_\gamma$ is orthogonal unitary invariant, i.e., $||\mathbf{X}||_\gamma = ||\mathbf{P} \ast \mathbf{X} \ast \mathbf{Q}^T||_\gamma$, where $\mathbf{P} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathbf{Q} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal.

**Proof.**

(i) Since $\psi_{\text{MCP}}(t) \geq 0$, and $||\mathbf{X}||_\gamma$ is the sum of $\psi_{\text{MCP}}(t)$, we have $||\mathbf{X}||_\gamma \geq 0$. If $\mathbf{X} = 0$, then $\sigma_{kr}(\tilde{\mathbf{X}}^{(k)}) = 0$, that means $||\mathbf{X}||_\gamma = 0$. On the other hand, $||\mathbf{X}||_\gamma = 0$ means $\sigma_{kr}(\tilde{\mathbf{X}}^{(k)}) = 0$, that is $\tilde{\mathbf{X}} = 0$, so we have $\mathbf{X} = \text{iff}t(\tilde{\mathbf{X}}) = 0$.

(ii) Because $\psi_{\text{MCP}}(t)$ is increasing with respect to $\gamma$, and $||\mathbf{X}||_\gamma$ is the sum of $\psi_{\text{MCP}}(t)$, according to the properties of $\psi_{\text{MCP}}(t)$, we have $||\mathbf{X}||_\gamma$ is increasing with respect to $\gamma$.

(iii) Because $\psi_{\text{MCP}}(t)$ is concave over $[0, \infty)$, and $||\mathbf{X}||_\gamma$ is the sum of $\psi_{\text{MCP}}(t)$, by regarding $\sigma_{kr}(\tilde{\mathbf{X}}^{(k)})$ as $t$, we know $||\mathbf{X}||_\gamma$ is concave over $\sigma_{kr}(\tilde{\mathbf{X}}^{(k)})$.

(iv) Since $\mathbf{X}$ has t-SVD $\mathbf{X} = \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^T$, then $\mathbf{P} \ast \mathbf{X} \ast \mathbf{Q}^T = \mathbf{P} \ast \mathbf{U} \ast \mathbf{S} \ast \mathbf{V}^T \ast \mathbf{Q}^T$. Because $\mathbf{S}$ is $f$-diagonal, we only need to verify that $\mathbf{P} \ast \mathbf{U}$ and $(\mathbf{V}^T \ast \mathbf{Q}^T)^T = \mathbf{Q} \ast \mathbf{V}$ are orthogonal.

So we have

$$(\mathbf{P} \ast \mathbf{U}) \ast (\mathbf{P} \ast \mathbf{U})^T = \mathbf{P} \ast \mathbf{U} \ast \mathbf{U}^T \ast \mathbf{P}^T = \mathbf{I}$$

$$(\mathbf{Q} \ast \mathbf{V}) \ast (\mathbf{Q} \ast \mathbf{V})^T = \mathbf{Q} \ast \mathbf{V} \ast \mathbf{V}^T \ast \mathbf{Q}^T = \mathbf{I}$$
\[ \| \mathcal{P} \ast \mathbf{X} \ast \mathcal{Q} \|_y = \frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{r=1}^{r_k} \psi_{\text{MCP}}(\sigma_{kr}(\mathbf{X}^{(k)})) = \| \mathbf{X} \|_y \]  

(2)

then \( \| \mathbf{X} \|_y \) is the orthogonal invariant.

### 3.2. Non-Convex Soft Shrinkage Operator of Tensor

As it is stated in the matrix \( \gamma \)-norm, minimization of the \( \gamma \)-norm gets the solution in the form of SVST. For the proposed tensor \( \gamma \)-norm minimization, we can also infer the same result as the following theorem presented.

**Theorem 3.** The solution of non-convex tensor minimization:

\[ \min_{\mathbf{X}} \| \mathbf{X} \|_y \quad \text{s.t.} \quad \mathbf{X} = \mathbf{Y} \]

is a non-convex singular value soft shrinkage operator in the Fourier domain, as follows:

\[ \mathbf{X}^* = D_{\lambda,y}^2(\mathbf{X}) = \mathbf{U} \ast (S_{\lambda,y} \ast \mathcal{M}) \ast \mathbf{V}^T \]  

(3)

where \( \ast \) means t-product, \( \mathbf{U}, \mathbf{V} \) comes from the t-SVD of \( \mathbf{Y} \) with \( \mathbf{Y} = \mathbf{U} \ast S \ast \mathbf{V}^T \). \( S_{\lambda,y} \) is a \( f \)-diagonal tensor with diagonal elements \( D_{\lambda,y}(\sigma_1), \ldots, D_{\lambda,y}(\sigma_m) \) as

\[ D_{\lambda,y}(\sigma_i) = \begin{cases} \sigma_i & \text{if } \sigma_i \geq \gamma \\ \sigma_i - \gamma & \text{if } \lambda \leq \sigma_i < \gamma \\ 1 - \lambda/\gamma & \text{otherwise} \end{cases} \]

\( \mathcal{M} \) is a \( f \)-diagonal tensor, its \( i \)-th diagonal tube \( \mathcal{M}(i,:,:) \) has Fourier transformation \( \hat{\mathcal{M}}(i,:,:) = [\hat{m}_{1,i}, \ldots, \hat{m}_{n_3,i}] \). The elements \( \hat{m}_{ik} \) are defined as

\[ \hat{m}_{ik} = \begin{cases} 1 & \text{if } \hat{S}(i,i,k) \geq \gamma \\ \frac{1 - \lambda/\gamma}{1 - \lambda/\gamma} & \text{if } \lambda \leq \hat{S}(i,i,k) < \gamma, \quad k = 1, \ldots, n_3 \end{cases} \]

**Proof.** Firstly, we consider a general form of minimization problem Equation (2) as follows:

\[ \phi(\mathbf{X}) = \frac{1}{2} \| \mathbf{Y} - \mathbf{X} \|_2^2 + \lambda \| \mathbf{X} \|_y \]  

(4)

where \( \| . \|_y \) is the Frobenius norm.

We rewrite Equation (4) as a form of the solution in terms of the frontal slice of \( \mathbf{X} \) as follows:

\[ \mathbf{X}^{(k)} = \arg \min_{\mathbf{X}^{(k)}} \frac{1}{2} \| \mathbf{Y}^{(k)} - \mathbf{X}^{(k)} \|_F^2 + \lambda \| \mathbf{X}^{(k)} \|_y \]  

(5)

where matrix \( \mathbf{Y}^{(k)} \) can be represented as \( \mathbf{Y}^{(k)} = \mathbf{U} \mathbf{S} \mathbf{V}^T \), \( \mathbf{U}, \mathbf{V} \) are orthogonal matrices, \( \mathbf{S} \) is a diagonal matrix with diagonal elements \( \sigma_1, \ldots, \sigma_r \).

According to matrix-based non-convex soft shrinkage operator in [44], the solution of Equation (5) is

\[ \mathbf{X}^{(k)} = \mathbf{U} D_{\lambda,y}(\mathbf{S}) \mathbf{V}^T \]  

(6)

where \( D_{\lambda,y}(\mathbf{S}) = \begin{cases} \sigma_i & \text{if } \sigma_i \geq \gamma \\ \frac{\sigma_i - \gamma}{1 - \lambda/\gamma} & \text{if } \lambda \leq \sigma_i < \gamma \\ 0 & \text{otherwise} \end{cases} \)

Let \( \mathbf{D}^{(1)}_{\lambda,y} = \begin{cases} \frac{1 - \lambda/\gamma}{1 - \lambda/\gamma} & \text{if } \lambda \leq \sigma_i < \gamma \end{cases} \), then \( D_{\lambda,y}(\mathbf{S}) = \sigma_i \times \mathbf{D}^{(1)}_{\lambda,y} \), which means \( D_{\lambda,y}(\mathbf{S}) \) is the multiplication of the \( i \)-th singular value of \( \mathbf{S} \) and \( \mathbf{D}^{(1)}_{\lambda,y} \). In the following, we will show how to solve \( \mathbf{X} \) through the expression of \( \mathbf{X}^{(k)} \) as Equation (6).
Assuming \( \Psi \) has t-SVD \( \Psi = U \ast S \ast V^T \), \( S \) has Fourier transform \( \tilde{S} \), then the elements in the singular tube of \( \tilde{X} \) are produced by multiplying \( \tilde{S}(i, k) \) and \( D_{k,y}^2 \), where

\[
D_{k,y}^2 = \begin{cases} 
1 & \text{if } \tilde{S}(i, i, k) \geq \gamma \\
1 - \gamma / \sigma_i & \text{if } \lambda \leq \tilde{S}(i, i, k) < \gamma \\
0 & \text{otherwise}
\end{cases}
\]

The above mentioned is processed in the Fourier domain; in the spatial domain, an equivalent formulation can be represented by \( \tilde{S} \ast \tilde{M} \), where \( \tilde{M} \) is a \( 3 \)-diagonal tensor and its Fourier transform of the \( i \)-th diagonal tube is denoted as \( \tilde{M}(i, i, :) = \text{fft}(\tilde{M}(i, i, :), [\cdot, 3]) \). Hence, the elements of \( \tilde{M}(i, i, :) \) are \( \tilde{m}_{ik} \) and are computed as follows:

We have \( \tilde{x} = U \ast (S_{t,y} \ast M) \ast V^T = D_{t,y}^2(\Psi) \).

\[
\tilde{m}_{ik} = \begin{cases} 
1 & \text{if } \tilde{S}(i, i, k) \geq \gamma \\
1 - \gamma / \sigma_i & \text{if } \lambda \leq \tilde{S}(i, i, k) < \gamma, \ k = 1, ..., n_3
\end{cases}
\]

This theorem proves that the optimization of the proposed tensor \( \gamma \)-norm leads to SVST. As illustrated in Figure 2, SVST retains more of the large singular values of frontal slice \( \tilde{X}^{(k)} \) than SVT does, which in turn captures the most important information of HSI. Taking this advantage into consideration, we utilize the proposed tensor \( \gamma \)-norm as a non-convex low-rank regularization to the following HSI restoration models.

3.3. Tensor \( \gamma \)-Norm in HSI Recovery

3.3.1. Non-Convex Low-Rank Tensor Completion

Assuming \( P_\Omega(\mathcal{O}) \) of size \( n_1 \times n_2 \times n_3 \) is the observed HSI in a set \( \Omega \), which has missing values compared with original HSI \( \tilde{X} \) of size \( n_1 \times n_2 \times n_3, \Omega \subset [1, n_1] \times [1, n_2] \times [1, n_3] \) denotes the so-called sampling set. The objective of HSI completion is to estimate the original HSI \( \tilde{X} \) from the observed HSI \( P_\Omega(\mathcal{O}) \). The HSI completion problem can be formulated as follows:

\[
P_\Omega(\mathcal{O}) = \begin{cases} 
P_\Omega(\tilde{X}) & \text{if } (i_1, i_2, i_3) \in \Omega \\
0 & \text{otherwise}
\end{cases}
\]

where \( P_\Omega: R^{n_1 \times n_2 \times n_3} \rightarrow R^{n_1 \times n_2 \times n_3} \) is a projection operator, \((i_1, i_2, i_3)\) is the pixel of \( \tilde{X} \).

By incorporating tensor \( \gamma \)-norm into the TNN minimization model [33] as a non-convex relaxation of low-rank, a novel low-rank tensor completion method named as non-convex TNN (NCTNN) minimization is developed as follows:

\[
\min_{\tilde{X}} \|\|X\|\|_\gamma \quad \text{s.t. } P_\Omega(\tilde{X}) = P_\Omega(\mathcal{O})
\]

(7)

According to Theorem 3, tensor \( \gamma \)-norm in problem (7) has a solution with the SVST operator, as Equation (3) has shown. As it is well known, low-rank prior represents the inherence spectral–spatial correlation of HSI, and large singular values stand for the most useful information. In SVST, the large singular values are nearly untouched, and other tiny values are shrunk, which provides nearly unbiased and more accurate low-rank approximation. Moreover, due to the low-rank relaxation term, the model can capture the spatial and spectral information of HSI simultaneously.

The optimization problem (7) is solved in the Fourier transform domain. Let \( \mathcal{Y} \) be the available sampled data \( \mathcal{Y} = P_\Omega(\mathcal{O}) \). Define \( \mathcal{G} = \mathcal{F}_3 P_\mathcal{O} \mathcal{F}_\mathcal{G} \), where \( \mathcal{F}_3 \) and \( \mathcal{F}_\mathcal{G} \) are Fourier and inverse Fourier transformation along the third dimension of tensors, respectively. Then, we have \( \mathcal{Y} = \mathcal{G}(\mathcal{O}) \), where \( \mathcal{G} \) and \( \mathcal{O} \) are the Fourier transforms of \( \mathcal{Y} \) and \( \mathcal{O} \) along the third mode, respectively. So Equation (7) is equivalent to the following:

\[
\min_{\tilde{X}} \|\|\text{blkdiag}(\tilde{X})\|\|_\gamma \quad \text{s.t. } \mathcal{Y} = \mathcal{G}(\tilde{X})
\]

(8)
The optimization problem (8) can be solved by various methods. Considering that t-SVD converts the TNN minimization into matrix rank minimization in the Fourier domain, the argument Lagrange formulation of Equation (8) in the Fourier domain is written as follows:

\[ l(\mathcal{X}, \mathcal{Z}, Q) = ||\text{blkdiag}(\hat{Z})||_F + 1_{y=p}\mathcal{X}^+Q + Q - \hat{\mathcal{X}} > + \lambda/2||\hat{\mathcal{X}} - \hat{Z}||_F^2 \]  

(9)

where \( \mathcal{Z} \) is the auxiliary variable which satisfies \( \hat{\mathcal{X}} = \hat{\mathcal{Z}} \), \( 1 \) is the indicator function, \( Q \) is Lagrange multiplier, \( \lambda \) is the parameter.

Then we incorporate t-SVD with the alternating direction method of multipliers (ADMM) [30–31,36-37,45] to solve the proposed model as shown in Algorithm 1. The iteration is ended if the error of two successive iterations is smaller than a constant \( \varepsilon \), which is usually sufficiently small. The smaller \( \varepsilon \) is, the more time consumption takes. In Algorithm 1, we set \( \varepsilon = 10^{-6} \) empirically.

Algorithm 1 Solve NCTNN (7) by ADMM

Input: observed image \( \mathcal{O} \), parameters \( \lambda, r, y \).
Initialize: \( \mathcal{X} = 0, \mathcal{Z} = 0, Q = 0, \varepsilon = 10^{-6}, k = 0 \).

while not converged do
1. Update \( \mathcal{X}^{k+1} \) by \( \mathcal{X}^{k+1} = \mathcal{Z}^k - Q^k \)
2. Update \( \mathcal{Z}^{k+1} \) by \( \mathcal{Z}^{k+1} = D^3_{x,r}(\hat{\mathcal{X}}^{k+1} + Q^k) \)
3. Update \( Q^{k+1} \) by \( Q^{k+1} = Q^k + (\mathcal{X}^{k+1} - Z^{k+1}) \)
4. Check the convergence conditions \( ||\mathcal{Z}^{k+1} - Z^k||_F^2 \leq \varepsilon \)
end while

3.3.2. Non-Convex Tensor Robust Principle Component Analysis

Although the aforementioned NCTNN model can recover degraded HSI well, it may have a weakness in removing mixed-noise efficiently, such as the sparse noise. Robust principle component analysis (RPCA) [7] fully considers the sparse noise in HSI by separating the image into signal and sparse noise or outliers, and gives low-rank and sparsity priors to the two parts, respectively. Unfortunately, RPCA in [7] regards HSI as a matrix, which destroys the intrinsic structure of HSI, while tensor robust principle component analysis (TRPCA) associated with TNN in [33] may obtain an image with residual noise on account of the limited low-rank approximation of TNN. Hence, to remove the sparse noise effectively, we introduce the proposed tensor \( \gamma \)-norm to the RPCA model, resulting in the second proposed method: non-convex tensor robust principle component analysis (NCTRPCA) as follows:

\[ \min_{\mathcal{X}, \mathcal{S}} ||\mathcal{X}||_F + \lambda||\mathcal{S}||_{1,1,2} \text{ s.t. } \mathcal{O} = \mathcal{X} + \mathcal{S} \]  

(10)

where \( \mathcal{O} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) is the observed image, \( n_1 \) and \( n_2 \) are the spatial ways and \( n_3 \) is the spectral way, \( \mathcal{X} \) is the original image, \( \mathcal{S} \) is the sparse noise, \( \lambda \) is a parameter, \( ||\mathcal{S}||_{1,1,2} \) is defined as \( \Sigma_i \Sigma_j ||\mathcal{S}(i,j,:)||_F \).

The non-convex optimization problem Equation (10) is solved by the same strategy as Algorithm 1, in which ADMM and Fourier transform are also employed.

The argumented Lagrange function of Equation (10) can be written as follows:

\[ l(\mathcal{X}, \mathcal{S}, W) = ||\mathcal{X}||_F + \lambda||\mathcal{S}||_{1,1,2} + < W, \mathcal{O} - \mathcal{X} - \mathcal{S} > + \rho/2||\mathcal{O} - \mathcal{X} - \mathcal{S}||_F^2 \]  

(11)

where \( W \) is Lagrangian multiplier, \( \lambda, \rho \) are parameters.

Algorithm 2 summarizes the solving steps. As mentioned in Algorithm 1, iteration termination condition is decided by \( \tau_X \) and \( \tau_S \), and we set \( \tau_X = 10^{-3}, \tau_S = 10^{-3} \) empirically in Algorithm 2.
Algorithm 2: Solve NCTRPCA (10) by ADMM

**Input:** observed image $O$, parameter $\lambda$, $\gamma$, $r$.

**Initialize:** $X = 0$, $S = 0$, $W = 0$, $\rho = \frac{2}{\tau}$, $k = 0$, $\tau_X = 10^{-3}$, $\tau_S = 10^{-3}$.

while not converged do

1. Update $X^{k+1}$ by

$$X^{k+1} = D_{\frac{1}{\rho}}(O - S^k - W^k)$$

2. Update $S^{k+1}$ by

$$S^{k+1} = D_{\frac{1}{\rho}}(X^{k+1} - O + W^k)$$

3. Update $W^{k+1}$ by

$$W^{k+1} = W^k + X^{k+1} + S^{k+1} - O$$

4. Check the convergence conditions

$$||X^{k+1} - X^k||^2_F \ll \tau_X$$

$$||S^{k+1} - S^k||^2_F \ll \tau_S$$

end while

where $D_{\frac{1}{\rho}}(x) = \text{sgn}(x)[|x| - \frac{\lambda}{\rho}]_+$ is the soft shrinkage operator.

4. Experimental Results and Analysis

In this section, to verify the effectiveness of the proposed methods, experiments are performed on three HSI datasets. Among them, Washington DC Mall with a size of $151 \times 151 \times 191$, and Pavia University with a size of $200 \times 200 \times 103$ are used for simulated experiments, while HYDICE Urban dataset with a size of $200 \times 200 \times 210$ is used for real experiments. The gray values of each HSI band are normalized to $[0, 1]$.

The parameters in the compared methods are adjusted manually to optimal. While in our method, parameter rank $r$, $\lambda$ and $\gamma$ are corresponding to the parameters analysis given later on.

The quantitative assessment indices, including the relative square error (RSE), the mean peak signal-to-noise ratio (MPSNR), the mean structural similarity (MSSIM) and the mean spectral angle distance (MSAD) are adopted.

4.1. Experiment 1: Comparative Experiments for NCTNN

In order to verify the effectiveness of NCTNN in the restoration of HSI, we applied some matrix and tensor completion methods from the perspective of non-convex relaxation and convex relaxation to HSI recovery. The observed HSI is obtained by randomly choosing a certain percentage element from the original HSI tensor, a large percentage denotes a high sampling rate (SR) and vice versa. Washington DC and Pavia University datasets are used in the experiments.

4.1.1. Non-Convex Methods Comparative Experiments

Firstly, the proposed NCTNN is compared with two non-convex methods, NMC [14] and NORT [41], among them, NMC is a matrix completion method, and NORT belongs to a tensor completion method.

The rank of the matrix in NMC is 15, and the tensor rank of NORT is $r = [10, 10, 10]$, and in the proposed NCTNN method, $r = [25, 25, 25], \gamma = 2$, respectively.

Table 1 summarizes the MPSNR, MSSIM values obtained from the three methods for Pavia University in different sampling rates (SR). The best index is in boldface. It is obvious that NORT and...
NCTNN, two tensor completion methods, perform better than the matrix-based NMC method in dealing with HSI recovery. NCTNN has improvements in MPSNR, MSSIM with nearly 0.7535 dB, 0.0122 over NORT on average, respectively. In addition, MSAD values of NCTNN perform best on spectral maintenance. Furthermore, the performance of NCTNN has a rising trend with the decrease in the sampling rate.

Figure 3 shows PSNR and SSIM curves of the three methods with a 50% sampling rate. We can find that NORT has more abnormal bands than the other two methods, while the curves of NMC, NCTNN are much more stable, and NCTNN performs best.

The recovery images of the three methods on band 100 with a 50% sampling rate are shown in Figure 4. NMC has more missing pixels, while the other two tensor-based methods can recover most information of the image, and the result of NCTNN has more details in the red box.

**Table 1.** Quantitative evaluation of completion results for Pavia University.

| SR    | Index | NMC  | NORT | NCTNN |
|-------|-------|------|------|-------|
| 10%   | MPSNR | 29.9454 | 31.0101 | 32.0785 |
|       | MSSIM | 0.8010 | 0.8179 | 0.8480 |
|       | MSAD  | 10.7925 | 7.6400 | 7.4383 |
|       | RSE   | 0.1249 | 0.1061 | 0.1024 |
| 50%   | MPSNR | 32.9078 | 33.1604 | 33.7900 |
|       | MSSIM | 0.8749 | 0.8773 | 0.8832 |
|       | MSAD  | 7.7770 | 6.1126 | 5.5622 |
|       | RSE   | 0.0504 | 0.0490 | 0.0466 |
| 70%   | MPSNR | 39.6411 | 39.9678 | 40.5303 |
|       | MSSIM | 0.9653 | 0.9699 | 0.9704 |
|       | MSAD  | 3.2117 | 2.8011 | 2.6770 |

4.1.2. Convex Methods Comparative Experiments

The state-of-the-art low-rank tensor convex relaxation methods, including LRTC [17], TMac [23] and TNN [30], are employed in this section. TNN uses tensor tubal-rank to tensor completion, while LRTC and TMac utilize the same tensor rank, which is defined by matrix factorization of each mode unfolding of the tensor. Moreover, LRTC provides algorithms to get efficient or high accuracy estimations, while TMac solves the tensor completion with faster algorithms as well as rank-adjusting strategies.

The parameters are set empirically. In TMac, a smaller rank is given due to its adaptation in the algorithm, e.g., \( r = [10, 10, 10] \). While for other three methods, the same rank \( r = [20, 20, 20] \) is chosen to give a relatively fair comparison. Parameters \( \lambda \) and \( \gamma \) in NCTNN are set as \( \lambda = 1/\sqrt{\lambda_1 \lambda_2}, \gamma = 2 \). Table 2 summarizes the RSE, MPSNR, MSSIM and MSAD values obtained from four tensor completion methods for Washington DC Mall in different sampling ratios. Obviously, the indices of TNN and NCTNN are superior to those of the LRTC and TMac methods, which is mainly due to the different definitions of tensor rank. The proposed NCTNN method achieves the best performance, and TNN ranks second best. Compared with TNN, NCTNN has higher indices in MPSNR, MSSIM with nearly 0.01 dB and 0.002 values, respectively. Meanwhile, NCTNN also achieves the lowest value in MSAD, which means an improvement in spectral fidelity.
Figure 3. PSNR and SSIM values of each band for Pavia University (SR = 50%): (a) PSNR; (b) SSIM.

Figure 4. Restoration results for Pavia University of band 100 (SR = 50%): (a) sampling; (b) NMC; (c) NORT; (d) NCTNN.

In order to highlight the advantages of NCTNN over TNN minimization, the PSNR and SSIM curves for Washington DC Mall with 30% sampling rates are given in Figure 5. The indices of NCTNN are greater than that of TNN’s, and the curve of NCTNN is more stable, which further demonstrates the outstanding performance of our proposed method in data completion with unbiasedness, as well as robustness.

Figure 5. PSNR and SSIM values of each band for Washington DC Mall (SR = 30%): (a) PSNR; (b) SSIM.

In Figure 5, the abnormal band 75 has a lower value than those of other normal bands. Figure 6 gives the restoration results. It can be obviously observed that TMac and LRTC perform worse in retaining the edges and textures. In comparison to NCTNN, TNN deletes some important textural information. In general, due to the non-convex relaxation of the tensor nuclear norm, NCTNN perfectly preserves the spatial geometric structures, which further illustrates that NCTNN has the highest recovery ability and the best stability compared with the other three methods.
Figure 6. Restoration results for Washington DC Mall of band 75 (SR = 10%): (a) sampling; (b) TMac; (c) LRTC; (d) TNN; (e) NCTNN.

Table 2. Quantitative evaluation of completion results for Washington DC Mall.

| SR  | Index | TMac | LRTC | TNN | NCTNN |
|-----|-------|------|------|-----|-------|
|     | RSE   | 0.1778 | 0.2549 | 0.1098 | 0.1085 |
| 10% | MPSNR | 36.4864 | 34.5769 | 34.8567 | 34.8649 |
|     | MSSIM | 0.8092 | 0.7499 | 0.8354 | 0.8378 |
|     | MSAD  | 10.6298 | 10.4791 | 6.1192 | 6.1147 |
| 30% | MPSNR | 38.2925 | 38.6530 | 39.9691 | 40.6430 |
|     | MSSIM | 0.9284 | 0.9239 | 0.9428 | 0.9360 |
|     | MSAD  | 5.1722 | 5.6094 | 3.1803 | 3.0635 |
| 50% | MPSNR | 39.5542 | 38.9234 | 40.1398 | 40.8033 |
|     | MSSIM | 0.9284 | 0.9099 | 0.9380 | 0.9361 |
|     | MSAD  | 6.5591 | 3.7911 | 3.1894 | 2.7657 |
| 80% | MPSNR | 42.1434 | 41.4434 | 40.5245 | 41.1576 |
|     | MSSIM | 0.9268 | 0.9232 | 0.9316 | 0.9443 |
|     | MSAD  | 4.0437 | 3.6204 | 3.9243 | 2.6206 |

To further verify the advantages of the proposed model, Figure 7 shows the spectral signatures before and after restoration of the Washington DC dataset with a 50% sampling rate. We choose the spectral signature of pixel (68, 42); it is obvious that the curve of NCTNN is the most similar to the original one, and the amplitude of vibration is the smallest, which shows that NCTNN has an improvement in spectral preservation as well as robustness.

4.1.3. Parameter Analysis

In order to illustrate the effect of rank selection on image restoration, we selected the Washington DC Mall dataset (SR = 10%) to carry out image restoration simulation experiments with different $r$, where $r$ is taken from $[20, 20, 20]$ to $[40, 40, 40]$, and the results are shown in Figure 8a,b. It can be seen that the overall trend of MPSNR and MSSIM increases with the increase in $r$. When $r > [25, 25, 25]$, the curve begins to decline. Therefore, $r = [25, 25, 25]$ may be as a reference for other datasets.

In order to study the approximation ability of the proposed tensor $\gamma$-norm on the tensor rank, we choose the Washington DC Mall dataset (SR = 10%) to carry out image restoration simulation experiments with different $\gamma$, where the value of $\gamma$ is from 0.5 to 6, and the results are shown in Figure 8c,d. It can be found that when $\gamma > 1$, the effect of $\gamma$ on the curve changes slowly. When $\gamma = 2$, the curve gets the peak. Therefore, in experiment 1, we fixed it to 2 empirically.
Figure 7. Spectrum of pixel (68, 42) in the restoration results for Washington DC Mall (SR = 50%).

Figure 8. Sensitivity analysis of parameter \( r \) and \( \gamma \) for Washington DC Mall (SR = 10%): (a) MPSNR of \( r \); (b) MSSIM of \( r \); (c) MPSNR of \( \gamma \); (d) MSSIM of \( \gamma \).

4.2. Experiment 2: Comparative Experiments for NCTRPCA

In this experiment, the proposed NCTRPCA is compared with four methods: two convex methods: RPCA [7], TRPCA [31] and two non-convex methods: IRNN[15], NLRWTV[13]. RPCA separates clean HSI from corrupted sparse noise with low-rank and sparsity regularization of the matrix, TRPCA is a convex relaxation based on TNN norm of tensor, IRNN is a non-convex non-smooth minimization denoising method by iteratively reweighted nuclear norm, and NLRWTV is a non-convex low-rank approximation along with a weighted total variation.

Both the simulated and real HSI restoration experiments are performed on three different datasets, and the parameters are fixed empirically during experiments. In IRNN, NLRWTV, and RPCA, \( r = 10 \), \( \gamma = 2 \). For TRPCA, \( r = [10,10,10] \). In our method of NCTRPCA, \( r = [15,15,15] \), \( \gamma = 1.5 \), \( \lambda = 1/\sqrt{\pi_1\pi_2} \), respectively.

4.2.1. Comparative Experiments

For simulated experiments, Washington DC Mall and Pavia University datasets are corrupted by mixed Gaussian and salt-and-pepper noise. Each band is added with the same noise intensity. Several different noise intensities are tested, specifically, for zero mean Gaussian noise, variances \( G \) are 0.025, 0.05, 0.075, and 0.1, while for salt-and-pepper noise, the percentages \( P \) are 0.05, 0.1, 0.15, and 0.2, respectively.

Table 3 summarizes the values of RSE, MPSNR, MSSIM and MSAD after the recovery of Washington DC Mall and Pavia University datasets by five methods. It can be seen that TRPCA and NCTRPCA are superior to other matrix-based methods in denoising performance, which benefits from structural properties of the tensor. At the same time, compared with convex method TRPCA, the MPSNR and MSSIM indices of NCTRPCA have relatively large values, which indicate a better performance in maintaining the spatial structure and spectral information.
Table 3. Quantitative evaluation of denoising results for simulated experiments.

| Dataset          | Noise Level | Index    | IRNN       | NRLRWTW | RPCA    | TRPCA | NCTRPCA |
|------------------|-------------|----------|------------|----------|---------|-------|---------|
| Washington DC    | G = 0.025   | RSE      | 0.0829     | 0.0556   | 0.2210  | 0.0439| 0.0402  |
|                  | P = 0.05    | MPSNR    | 40.8370    | 42.7526  | 26.9754 | 44.6603| 45.9740 |
|                  | P = 0.1     | MSSIM    | 0.9000     | 0.9323   | 0.5710  | 0.9774| 0.9796  |
|                  |             | MSAD     | 3.7665     | 2.2682   | 10.8807 | 2.2678| 2.2237  |
|                  | G = 0.05    | RSE      | 0.0800     | 0.0601   | 0.3466  | 0.0764| 0.0599  |
|                  | P = 0.1     | MPSNR    | 35.5132    | 38.0533  | 22.6086 | 43.3904| 42.0658 |
|                  |             | MSSIM    | 0.7973     | 0.8511   | 0.3902  | 0.9542| 0.9599  |
|                  |             | MSAD     | 5.3607     | 4.3987   | 18.7847 | 3.9803| 3.9644  |
|                  | G = 0.075   | RSE      | 0.1293     | 0.0860   | 0.1670  | 0.1401| 0.1293  |
|                  | P = 0.15    | MPSNR    | 32.3048    | 34.8965  | 35.5580 | 40.1497| 39.7160 |
|                  |             | MSSIM    | 0.7283     | 0.8013   | 0.8147  | 0.8775| 0.9266  |
|                  |             | MSAD     | 7.7141     | 6.3412   | 9.5087  | 5.4784| 5.4685  |
|                  | G = 0.1     | RSE      | 0.2085     | 0.1150   | 0.2041  | 0.1504| 0.2085  |
|                  | P = 0.2     | MPSNR    | 30.0710    | 32.8250  | 34.0711 | 37.7062| 37.6901 |
|                  |             | MSSIM    | 0.6736     | 0.7707   | 0.7965  | 0.8716| 0.8979  |
|                  |             | MSAD     | 9.7615     | 6.8418   | 12.0437 | 7.6575| 6.8022  |
| Pavia University | G = 0.025   | RSE      | 0.0688     | 0.0503   | 0.2233  | 0.0488| 0.0469  |
|                  | P = 0.05    | MPSNR    | 37.5136    | 39.6694  | 26.4945 | 40.0122| 40.4803 |
|                  |             | MSSIM    | 0.9490     | 0.9701   | 0.6772  | 0.9655| 0.9701  |
|                  |             | MSAD     | 3.5643     | 2.7967   | 8.0692  | 3.2130| 2.7195  |
|                  | G = 0.05    | RSE      | 0.1398     | 0.0811   | 0.2762  | 0.0752| 0.0724  |
|                  | P = 0.1     | MPSNR    | 34.0721    | 36.3067  | 24.4730 | 35.9387| 36.5155 |
|                  |             | MSSIM    | 0.8890     | 0.9314   | 0.6277  | 0.9266| 0.9320  |
|                  |             | MSAD     | 8.0235     | 4.8143   | 10.5837 | 4.7153| 4.7020  |
|                  | G = 0.075   | RSE      | 0.2242     | 0.1303   | 0.3670  | 0.1022| 0.1068  |
|                  | P = 0.15    | MPSNR    | 31.9070    | 32.8567  | 22.0202 | 33.1871| 33.7116 |
|                  |             | MSSIM    | 0.8314     | 0.8752   | 0.5017  | 0.8774| 0.8823  |
|                  |             | MSAD     | 11.8662    | 8.1859   | 14.4344 | 6.1167| 5.6251  |
|                  | G = 0.1     | RSE      | 0.2871     | 0.1702   | 0.4830  | 0.1309| 0.1149  |
|                  | P = 0.2     | MPSNR    | 29.6851    | 29.9220  | 19.6690 | 31.0240| 32.0927 |
|                  |             | MSSIM    | 0.7588     | 0.8044   | 0.3573  | 0.8197| 0.8487  |
|                  |             | MSAD     | 15.6410    | 10.7094  | 18.5420 | 7.4184| 6.8022  |

For the Washington DC Mall dataset, Figure 9 shows the PSNR and SSIM curves recovered by three denoising methods with the noise intensity of G = 0.05 and P = 0.1. The curves of NRLRWTW and NCTRPCA are more stable. Meanwhile, the index values of NCTRPCA and TRPCA are higher than those of IRNN and NRLRWTW. This is because the IRNN, NRLRWTW and RPCA methods decompose HSI into matrix and solve the calculation in the process of model construction, while NCTRPCA and TRPCA deal with HSI in the third-order tensor manner. The tensor representation makes better use of the structural information contained in the spatial similarity and spectral correlation of HSI. Figure 10 shows the recovered image of abnormal band 100 with five denoising methods, among which the image of IRNN and RPCA are very blurry, and the edges of NRLRWTW, TRPCA and NCTRPCA are clearer. However, the recovered images of NRLRWTW contain more noise, and TRPCA is somewhat blurry, as shown in the enlarged parts. The proposed NCTRPCA exhibits a restored image with the highest quality.

For the Pavia University dataset, the advantages of NCTRPCA and TRPCA over other matrix-based methods are illustrated in Figure 11 by PSNR and SSIM curves with G = 0.05, P = 0.1. As it has been shown previously, the tensor-based methods are more robust than matrix-based methods. At the same time, the fluctuation of the NCTRPCA curve has the highest value with the lowest range, which means it has the best denoising ability and achieves the best unbiasedness. The visual comparison of band 87 is shown in Figure 12, along with the enlarged red box. Compared with the
other four methods, NCTRPCA also obtains a noise-free image with more geometrical details simultaneously.

Figure 9. PSNR and SSIM values of each band for Washington DC Mall (G = 0.05, P = 0.1): (a) PSNR; (b) SSIM.

Figure 10. Restoration results for Washington DC Mall of band 100: (a) G = 0.05, P = 0.1; (b) IRNN; (c) NRLRWTV; (d) RPCA; (e) TRPCA; (f) NCTRPCA.

Figure 11. PSNR and SSIM values of each band for Pavia University (G = 0.025, P = 0.05): (a) PSNR; (b) SSIM.
Figure 12. Restoration results for Pavia University of band 87: (a) G = 0.025, P = 0.05; (b) IRNN; (c) NRLRWTV; (d) RPCA; (e) TRPCA; (f) NCTRPCA.

For real experiments that we performed on the HYDICE Urban dataset to compare the recovery results, the parameters are set the same as in the simulated experiments.

In the real experiments, the indices of RSE, PSNR and SSIM are incapable on account of missing groundtruth. However, MSAD between the noisy and denoised images can be computed to give a measure of the spectral similarity. Therefore, Table 4 gives the MSAD after the recovery of the HYDICE Urban dataset by five methods. It can be seen that compared with convex methods, non-convex methods perform better in spectral preservation. TRPCA and NCTRPCA are superior to matrix-based methods. Furthermore, NCTRPCA has larger values than TRPCA, which shows a better performance in spectral preservation.

A visual comparison of spectral similarity is shown by plotting the spectral signatures of the noisy and the resulted images in Figure 13. The closer to the original curve, the higher similarity is. As an illustration, pixel (60, 50) of HYDICE Urban is selected randomly. We can find that IRNN has the biggest spectral distortion, while the curves of other non-convex methods are more close to those of the original. In addition, most of the curves are more stable except some abnormal bands. The results of RPCA and TRPCA, NRLRWTV and NCTRPCA are comparable, due to the same low-rank relaxation they employed. On the whole, the results of NCTRPCA are best, both on visual and quantitative assessment, which further demonstrates the effectiveness of the proposed method in spectral distortion.

Figure 13. Spectrum of pixel (60, 50) in the restoration results for HYDICE Urban.
Table 4. MASD values for real experiments.

| Dataset  | Index | IRNN  | NLRWTV | RPCA  | TRPCA | NCTRPCA |
|----------|-------|-------|--------|-------|-------|---------|
| HYDICE Urban | MSAD  | 11.1242 | 8.8584 | 14.7328 | 8.1838 | 7.5465 |

4.2.2. Parameter Analysis

In order to illustrate the influence of rank selection on image restoration, as an example, we chose the Washington DC Mall dataset (G = 0.025, P = 0.05) to conduct noise removal simulation experiments with different r, where r is taken from [5,5,5] to [25,25,25], and the results are shown in Figure 14a,b. On the whole, the values of MPSNR and MSSIM first increase then decrease with the increase in r. When r = [15,15,15], the curves of MPSNR and MSSIM can reach an approximate peak value, and the model has the best denoising effect.

In order to study the approximate effect of the tensor γ-norm on the tensor rank, we chose the Washington DC Mall dataset (G = 0.025, P = 0.05) to carry out noise removal simulation experiments under different γ. Among them, the value of γ is from 0.5 to 2.5 as shown in Figure 14c,d. It can be found that when γ > 1.5, the curve changes slowly. When γ = 1.5, both MPSNR and MSSIM reach the maximum value. Therefore, we fixed γ to 1.5 in the experimental parameter selection.

![Figure 14. Sensitivity analysis of parameters for Washington DC Mall (G = 0.025, P = 0.05): (a) MPSNR of r; (b) MSSIM of r; (c) MPSNR of γ; (d) MSSIM of γ; (e) MPSNR of λ; (f) MSSIM of λ.](image)

λ is to adjust the regularization parameters of the sparse term; we chose the Washington DC Mall dataset (G = 0.025, P = 0.05) to conduct noise removal simulation experiments under different λ, where λ = p/√n1n2, and the value of p is from 0.5 to 15, as shown in Figure 14e,f. It is easy to see that when p > 1, the value curves of MPSNR and MSSIM are relatively stable, that is, NCTRPCA is robust in the selection of parameters λ. Therefore, when p = 1, the curve reaches the peak value, and the best denoising performance can also be obtained in visual effects.

Finally, time complexity is also analyzed. The proposed methods, along with other tensor-based methods, take more time than matrix-based methods. In addition, LRTC and TMac are faster than other methods; in LRTC, the Nesterov strategy is used to implement the nonsmooth optimization, while in TMac, low-rank matrix factorization is applied to speed up the algorithm. The other tensor-based methods, namely TNN, TRPCA, NCTNN, and NCTRPCA, are mainly negatively affected by the high complexity of SVD. More efficient algorithms for commutating SVD, such as random or parallel based algorithms, are encouraged for further investigation.
5. Conclusions

In this paper, to obtain an unbiased HSI restoration, we introduce the MCP penalization as a new non-convex penalty in tensor TNN minimization and TRPCA. On the whole, convex low-rank approximation methods are inferior to non-convex methods, and matrix-based methods also perform worse than tensor-based methods. Compared with the state-of-the-art non-convex and convex completion and denoising methods, the proposed methods, NCTNN and NCTRPCA, perform best both in visual and quantitative assessments, and achieve stable and robust results, especially on retaining spectral structure.

Although the proposed methods achieve stratifying restoration results, there are some aspects that need to be improved further. The low-rank-based methods focus on the global correlation while ignoring the nonlocal spatial similarity. Nonlocal tensor patches, which have been proven to be effective in image modeling, are going to be introduced to the proposed models in our future work. Besides, as mentioned above, the tensor-based methods are quite time-consuming, efficient processing techniques, e.g., parallel and random, are our on-going work.

Author Contributions: H.L. (Hanyang Li) conceptualized the work, carried out the data experiments and image processing. H.L. (Hongyi Liu) supervised the research and helped to write the manuscript. Z.W. (Zebin Wu) and Z.W. (Zhihui Wei) contributed to the review of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61971223, Grant 61772274, Grant 61671243 and Grant 61701238, in part by the Fundamental Research Funds for the Central Universities under Grant 30917015104, and in part by the Jiangsu Provincial Natural Science Foundation of China under Grant BK20170858.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chang, Y.; Yan, L.X.; Wu, T.; Zhong, S. Remote sensing image stripe noise Removal: From image decomposition perspective. IEEE Trans. Geosci. Remote Sens. 2016, 54, 7018–7031, doi:10.1109/tgrs.2016.2594080.
2. Huang, Z.H.; Li, S.T.; Hu, F. Hyperspectral image denoising with multiscale low-rank matrix recovery. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 23–28 July 2017; pp. 5442–5445.
3. Garzelli, A. A review of image fusion algorithms based on the super-resolution paradigm. Remote Sens. 2016, 8, 797.
4. Zhang, X.; Li, C.; Zhang, J.; Chen, Q.; Feng, J.; Jiao, L.; Zhou, H. Hyperspectral unmixing via low-rank representation with space consistency constraint and spectral library pruning. Remote Sens. 2018, 10, 339.
5. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. IEEE Trans. Geosci. Remote Sens. 2016, 54, 178–188, doi:10.1109/tgrs.2015.2452812.
6. Liu, H.Y.; Sun, P.P.; Du, Q.; Wu, Z.B.; Wei, Z.H. Hyperspectral image restoration based on low-rank recovery with a local neighborhood weighted spectral-spatial total variation model. IEEE Trans. Geosci. Remote Sens. 2019, 57, 1409–1422, doi:10.1109/tgrs.2018.2866439.
7. Candès, E.J.; Li, X.D.; Ma, Y.; Wright, J. Robust principal component analysis? J. ACM 2011, 58, 1–37, doi:10.1145/1970392.1970395.
8. Gu, S.H.; Zhang, L.; Zuo, W.M.; Feng, X.C. Weighted nuclear norm minimization with application to image denoising. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 2862–2869.
9. Liu, L.; Huang, W.; Chen, D.R. Exact minimum rank approximation via Schatten p-norm minimization. J. Comput. Appl. Math 2014, 267, 218–227, doi:10.1016/j.cam.2014.02.015.
10. Xie, Y.; Gu, S.H.; Liu, Y.; Zuo, W.M.; Zhang, W.S.; Zhang, L. Weighted schatten p-norm minimization for image denoising and background subtraction. IEEE Trans. Image Process. 2016, 25, 4842–4857, doi:10.1109/tip.2016.2599290.
11. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* 2010, 38, 894–942, doi:10.1214/09-aos729.

12. Wang, S.; Liu, D.; Zhang, Z. Nonconvex relaxation approaches to robust matrix recovery. In Proceedings of the Twenty-Third *International Joint Conference on Artificial Intelligence*, Barcelona, Spain, 16–22 July 2011.

13. Li, H.Y.; Sun, P.P.; Liu, H.Y.; Wu, Z.B.; Wei, Z.H. Non-convex low-rank approximation for hyperrectangular image recovery with weighted total variation regularization. In *Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 22–27 July 2018; pp. 2733–2736.

14. Wen, F.; Ying, R.; Liu, P.; Truong, T.-K. Matrix completion via nonconvex regularization: Convergence of the proximal gradient algorithm. *arXiv 2019*, arXiv:1903.00702.

15. Lu, C.Y.; Tang, J.H.; Yan, S.C.; Lin, Z.C. Generalized nonconvex nonsmooth low-rank minimization. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 24–27 June 2014; pp. 4130–4137.

16. Kolda, T.G.; Bader, B.W. Tensor decompositions and applications. *Siam Rev.* 2009, 51, 455–500, doi:10.1137/07070111X.

17. Liu, J.; Musialski, P.; Wonka, P.; Ye, J.P. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* 2013, 35, 208–220, doi:10.1109/tpami.2012.39.

18. Deng, Y.J.; Li, H.C.; Fu, K.; Du, Q.; Emery, W.J. Tensor low-rank discriminant embedding for hyperspectral image dimensionality reduction. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 7183–7194, doi:10.1109/tgrs.2018.2849085.

19. Fan, H.Y.; Chen, Y.J.; Guo, Y.L.; Zhang, H.Y.; Kuang, G.Y. Hyperspectral image restoration using low-rank tensor recovery. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* 2017, 10, 4589–4604, doi:10.1109/jstars.2017.2714338.

20. Gandy, S.; Recht, B.; Yamada, I. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* 2011, 27, 2, 025010, doi:10.1088/0266-5611/27/2/025010.

21. Kilmer, M.E.; Martin, C.D. Factorization strategies for third-order tensors. *Linear Algebra Appl.* 2011, 435, 641–658.

22. Xu, Y.Y.; Hao, R.R.; Yin, W.T.; Su, Z.X. Parallel matrix factorization for low-rank tensor completion. *Inverse Probl. Imaging* 2015, 9, 601–624, doi:10.3934/ippi.2015.9.601.

23. Ng, M.K.P.; Yuan, Q.Q.; Yan, L.; Sun, J. An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data. *IEEE Trans. Geosci. Remote Sens.* 2017, 55, 3367–3381, doi:10.1109/tgrs.2017.2670021.

24. Kilmer, M.E.; Braman, K.; Hao, N.; Hoover, R.C. Third-order tensors as operatoron matrices: A theoretical and computational framework with applications in imaging. *Siam J. Matrix Anal. Appl.* 2013, 34, 148–172, doi:10.1137/110837711.

25. Xie, Q.; Zhao, Q.; Meng, D.Y.; Xu, Z.B. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE Trans. Pattern Anal. Mach. Intell.* 2018, 40, 1888–1902, doi:10.1109/tpami.2017.2734888.

26. Lu, C.Y.; Feng, J.S.; Chen, Y.D.; Liu, W.; Lin, Z.C.; Yan, S.C. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 5249–5257.

27. Zhao, L.; Xu, Y.; Wei, Z.H.; Yu, R.P.; Qian, L. Hyperspectral image denoising via coupled spectral-spatial tensor representation. In *Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 22–27 July 2018; pp. 4784–4787.

28. Liu, Y.Y.; Shang, F.H. An Efficient matrix factorization method for tensor completion. *IEEE Signal Process. Lett.* 2013, 20, 307–310, doi:10.1109/isp.2013.2245416.

29. Goldfarb, D.; Qin, Z. Robust low-rank tensor recovery: Models and algorithms. *Siam J. Matrix Anal. Appl.* 2014, 35, 225–253.

30. Zhang, Z.M.; Aeron, S. Exact tensor completion using t-SVD. *IEEE Trans. Signal Process.* 2017, 65, 1511–1526, doi:10.1109/bsp.2016.2639466.

31. Zhang, Z.M.; Ely, G.; Aeron, S.; Hao, N.; Kilmer, M. Novel methods for multidimensional data completion and de-noising based on tensor-SVD. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 24–27 June 2014; pp. 3842–3849.
32. Hu, W.; Tao, D.; Zhang, W.; Xie, Y.; Yang, Y. The twist tensor nuclear norm for video completion. IEEE Trans. Neural Netw. Learn. Syst. 2016, 28, 2961–2973.
33. Zhou, P.; Feng, J. Outlier-robust tensor PCA. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–21 July 2017; pp. 2263–2271.
34. Zhou, P.; Lu, C.; Lin, Z.; Zhang, C. Tensor factorization for low-rank tensor completion. IEEE Trans. Image Process. 2017, 27, 1152–1163.
35. Jiang, T.-X.; Huang, T.-Z.; Zhao, X.-L.; Deng, L.-J. A novel nonconvex approach to recover the low-tubal-rank tensor data: When t-SVD meets PSSV. arXiv 2017, arXiv:1712.05870.
36. Cao, W.F.; Wang, Y.; Yang, C.; Chang, X.Y.; Han, Z.; Xu, Z.B. Folded-concave penalization approaches to tensor completion. Neurocomputing 2015, 152, 261–273, doi:10.1016/j.neucom.2014.10.069.
37. Ji, T.Y.; Huang, T.Z.; Zhao, X.L.; Ma, T.H.; Deng, L.J. A non-convex tensor rank approximation for tensor completion. Appl. Math. Model. 2017, 48, 410–422, doi:10.1016/j.apm.2017.04.002.
38. Li, H.; Liu, H.; Zhang, J.; Wu, Z.; Wei, Z. Non-convex relaxation low-rank tensor completion for hyperspectral image recovery. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1935–1938.
39. Zhang, X.J. A nonconvex relaxation approach to low-rank tensor completion. IEEE Trans. Neural Netw. Learn. Syst. 2019, 30, 1659–1671, doi:10.1109/tnnls.2018.2872583.
40. Cai, S.T.; Luo, Q.L.; Yang, M.; Li, W.; Xiao, M.Q. Tensor robust principal component analysis via non-convex low rank approximation. Appl. Sci. 2019, 9, 1411, doi:10.3390/app9071411.
41. Yao, Q.; Kwok, J.T.-Y.; Han, B. Efficient nonconvex regularized tensor completion with structure-aware proximal iterations. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 7035–7044.
42. Li, T.; Ma, J. Non-convex penalty for tensor completion and robust PCA. arXiv 2019, arXiv:1904.10165.
43. Gui, H.; Gu, Q. Towards faster rates and oracle property for low-rank matrix estimation. arXiv 2015, arXiv:1505.04780.
44. Gu, S.; Xie, Q.; Meng, D.; Zuo, W.; Feng, X.; Zhang, L. Weighted nuclear norm minimization and its applications to low level vision. Int. J. Comput. Vis. 2017, 121, 183–208.
45. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 2011, 3, 1–122.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).