

# LOW-RANK MATRIX COMPLETION WITH POISSON OBSERVATIONS VIA NUCLEAR NORM AND TOTAL VARIATION CONSTRAINTS\*

Duo Qiu

*School of Mathematics and Physics, Wuhan Institute of Technology, Wuhan 430205, China  
Email: qiuduoduo13@wit.edu.cn*

Michael K. Ng

*Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China  
Email: michael-ng@hkbu.edu.hk*

Xiongjun Zhang<sup>1)</sup>

*School of Mathematics and Statistics and Hubei Key Laboratory of Mathematical Sciences,  
Central China Normal University, Wuhan 430079, China  
Email: xjzhang@ccnu.edu.cn*

## Abstract

In this paper, we study the low-rank matrix completion problem with Poisson observations, where only partial entries are available and the observations are in the presence of Poisson noise. We propose a novel model composed of the Kullback-Leibler (KL) divergence by using the maximum likelihood estimation of Poisson noise, and total variation (TV) and nuclear norm constraints. Here the nuclear norm and TV constraints are utilized to explore the approximate low-rankness and piecewise smoothness of the underlying matrix, respectively. The advantage of these two constraints in the proposed model is that the low-rankness and piecewise smoothness of the underlying matrix can be exploited simultaneously, and they can be regularized for many real-world image data. An upper error bound of the estimator of the proposed model is established with high probability, which is not larger than that of only TV or nuclear norm constraint. To the best of our knowledge, this is the first work to utilize both low-rank and TV constraints with theoretical error bounds for matrix completion under Poisson observations. Extensive numerical examples on both synthetic data and real-world images are reported to corroborate the superiority of the proposed approach.

*Mathematics subject classification:* 15A83, 65K10, 90C30.

*Key words:* Low-rank matrix completion, Nuclear norm, Total variation, Poisson observations.

## 1. Introduction

The problem of low-rank matrix completion with Poisson observations is to estimate a low-rank matrix from given measurements at some subset of its locations, where the observations follow a Poisson distribution. Poisson observations appear in slew of practical applications in the areas of astronomical images, positron emission tomography, and magnetic resonance imaging [18, 33, 36]. Poisson noise is related to the count of photons recorded in the imaging devices, and can be modeled by a Poisson process, where the observations consist of counts of

---

\* Received February 25, 2023 / Revised version received June 19, 2023 / Accepted September 8, 2023 /

<sup>1)</sup> Corresponding author

photons arrivals at a detector [50]. The total number of photons collected by imaging sensors follows a Poisson distribution. Furthermore, many real-world image data are low-rank or approximately low-rank. And during the data acquisition and processing procedures, only partial observations are available due to mechanical failure or human-induced factors. In this paper, we focus on the low-rank matrix completion problem with Poisson observations, where the entries one observes follow a Poisson distribution.

Low-rank matrix completion has received broad practical interest in past decades, and been applied in a variety of fields, such as image processing, computer vision, machine learning, and data mining [8, 12, 19, 22, 25, 27, 29, 38, 47]. For example, Candès *et al.* [4] presented a nuclear norm minimization method for matrix completion without noise and showed that one can recover a low-rank matrix exactly from a small number of its sampled entries with high probability. Moreover, Candès *et al.* [3] studied the matrix completion with additive noise and showed that the error of the estimator of the resulting model is proportional to the noise level. In particular, for the underlying matrix corrupted by sparse noise, it is possible to recover both the low-rank and the sparse components exactly with high probability under suitable assumptions by solving a very convenient convex program [2]. Besides, Taherkhani *et al.* [46] proposed a matrix completion approach based on nuclear norm minimization to predict the missing label of unsupervised nodes for graph-based semi-supervised learning, which trained a convolutional neural network based classifier using a large amount of unlabeled data and a small amount of labeled data. However, these methods did not consider the local prior information of the underlying matrix.

Being different from additive Gaussian noise or additive sparse noise, Poisson noise is nonadditive and signal-dependent. For the problem of compressed sensing with Poisson observations, Raginsky *et al.* [37] proposed a method consisting of a negative Poisson logarithmic likelihood term and a penalty term, and then established an upper bound of the estimator, where the penalty term was utilized to measure signal sparsity. Then Jiang *et al.* [24] provided minimax lower bounds on mean square errors for sparse Poisson inverse problems under nonnegative and flux-preserving constraints. Furthermore, Cao *et al.* [5] proposed a novel model composed of Kullback-Leibler (KL) divergence in the objective and the nuclear norm constraint for matrix completion with Poisson observations, and established an upper bound of the estimator of their proposed approach, which is minimax optimal up to a logarithmic factor. Soni *et al.* [45] proposed a novel model with unified framework for structural low-rank matrix completion with general noise observations, where the underlying matrix is factorized into the product of two matrices and one factor matrix is sparse. Then the error bounds of the estimator of the resulting model were established, where the minimax lower bounds of this kind of models were also derived in [42]. Recently, McRae *et al.* [32] proposed a low-rank matrix completion method by utilizing Frobenius norm for the data-fitting term, where both the nuclear norm constraint and nuclear norm regularized least squares were studied. However, the method in [32] just utilized the Frobenius norm for the data-fitting term, which is just suboptimal for Poisson observations [26]. Recently, Zhang *et al.* [54] proposed a transformed tensor nuclear norm constraint method for low-rank tensor completion with Poisson observations, while it only utilized the low-rankness of the underlying tensor and the local prior information was not considered.

In image restoration, some work discussed and studied grey image recovery with Poisson noise, where there are no missing entries for the observations, see, e.g. [6, 7, 15, 26, 30, 43, 49, 51, 53, 55, 56]. The main model in the literature composed of the KL divergence data-fitting term and the total variation (TV) regularization term. The TV regularization is proposed to

generate piecewise smooth objects and preserve sharp edges for image restoration, where the TV was first proposed by Rudin *et al.* [41] for image denoising and then generalized to image deconvolution [40]. For image restoration with Poisson observations, Le *et al.* [26] proposed a variational model combined the TV for the underlying image and the KL divergence for the data-fitting term to denoise an image, where the TV can preserve good details of images and the KL divergence is suitable for Poisson noise by the maximum likelihood estimation. Zhao *et al.* [56] proposed a nonlocal low-rank model for Poisson noise removal, while it only utilized the low-rankness of the underlying images. Besides, some fast first-order algorithms were proposed and studied to solve this kind of models efficiently, see [15, 43, 49, 55] and references therein. However, these work cannot deal with the noisy observations with missing entries and only single prior information of the grey images is utilized.

On the other hand, the above existing work only discuss the low-rankness or piecewise smoothness of a matrix. However, many real-world matrices, e.g. grey images, are not only low-rank, but also piecewise smooth. In image restoration, some work were proposed and studied based on the TV and low-rank properties of images [20, 21, 31, 44, 52]. For example, in hyperspectral image restoration with mixed Gaussian and sparse noise removal, He *et al.* [21] proposed a spatial and spectral TV and low-rank factorization method for the underlying images, where hyperspectral images are not only assumed to lie in a low-dimensional subspace from the spectral perspective but also assumed to be piecewise smooth in the spatial dimension. Moreover, they further proposed a spatial-spectral TV regularized local low-rank matrix recovery method based on nonlocal self-similarity and low-rank constraint for the underlying hyperspectral images [20]. For medical image super-resolution, Shi *et al.* [44] proposed a novel image super-resolution method that integrates both local and global information for effective image recovery, whose information is achieved by the TV and low-rank regularization. However, these previous work did not consider the observations with Poisson noise and missing entries. More importantly, the theory guarantee combining the low-rankness and TV of the underlying matrix is not studied and analyzed in the literature.

In this paper, we propose a novel method composed of TV and nuclear norm constraints for low-rank matrix completion with Poisson observations. The objective of the proposed method is the KL divergence between the underlying matrix and noisy observations, which is derived by the maximum likelihood estimation of Poisson observations. The TV constraint is utilized to explore the piecewise smoothness and the nuclear norm constraint is to make use of the low-rankness of the underlying matrix. The advantage of both constraints in our proposed model is that the low-rankness and piecewise smoothness of the underlying matrix can be exploited simultaneously, where many real-world image data have the two properties. Then an upper error bound of the estimator of the proposed model is established with high probability under some assumptions, which is not larger than that of the single TV or nuclear norm constraint for low-rank matrix completion with Poisson observations. Moreover, an alternating direction method of multipliers (ADMM) is developed to solve the resulting model. Comprehensive numerical experiments on both synthetic data and real-world images demonstrate the recovery error theory of the estimator and the effectiveness of the proposed model compared with the models with single TV constraint or nuclear norm constraint.

The remaining parts of this paper are organized as follows. In Section 2, we introduce some preliminaries and notation which will be used throughout this paper. In Section 3, we propose a novel model consisted of the KL divergence in the objective and the nuclear norm and TV constraints for matrix completion with Poisson observations. An upper error bound of

the estimator of the proposed approach is established with high probability in Section 4. Then an ADMM based algorithm is designed to solve the resulting model in Section 5. Extensive numerical experiments are conducted to illustrate the effectiveness of the proposed approach in Section 6. We conclude our paper in Section 7. The proof of the main theorem is left to the Appendix B.

## 2. Preliminaries

We use  $\mathbb{R}^n$  and  $\mathbb{R}^{n_1 \times n_2}$  to denote the  $n$ -dimensional real Euclidean space and the set of all  $n_1 \times n_2$  matrices with real entries, respectively. For an arbitrary matrix  $X \in \mathbb{R}^{n_1 \times n_2}$ ,  $\|X\|$  denotes the spectral norm of  $X$ , which is the largest singular value of  $X$ .  $\|X\|_*$  denotes the nuclear norm of  $X$ , which is the sum of all singular values of  $X$ .  $\|X\|_F$  represents the Frobenius norm of  $X$ , which is defined as

$$\|X\|_F = \sqrt{\langle X, X \rangle}.$$

Let  $\mathbb{R}_+^{n_1 \times n_2}$  ( $\mathbb{R}_{++}^{n_1 \times n_2}$ ) represent the set of all  $n_1 \times n_2$  matrices with nonnegative (positive) and real entries. For any  $X \in \mathbb{R}^{n_1 \times n_2}$ ,  $X_{ij}$  stands for the  $(i, j)$ -th entry of  $X$ .

Now we recall several information-theoretic preliminaries, which play a pivotal role in the proof of error bounds of the proposed approach. The KL divergence between two Poisson distributions with parameters  $p, q > 0$  is defined as

$$K(p||q) = p \log \frac{p}{q} - (p - q).$$

The average KL divergence between Poisson distributions of two matrices  $P, Q \in \mathbb{R}_{++}^{n_1 \times n_2}$ , whose parameters are their entries  $P_{ij}, Q_{ij}$ , is defined as

$$K(P||Q) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} K(P_{ij}||Q_{ij}).$$

The square Hellinger distance between two Poisson distributions with parameters  $p, q > 0$  is defined as

$$H^2(p||q) = 2 - 2 \exp\left(-\frac{1}{2}(\sqrt{p} - \sqrt{q})^2\right).$$

Similarly, for two matrices  $P, Q \in \mathbb{R}_{++}^{n_1 \times n_2}$ , the average square Hellinger distance of two Poisson distributions, whose parameters are the entries of  $P, Q$ , is defined as

$$H^2(P||Q) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} H^2(P_{ij}||Q_{ij}).$$

It follows from [10, Theorem 2.1] that

$$H^2(p||q) \leq K(p||q). \tag{2.1}$$

For any  $v \geq 1$  and  $x, y \in \mathbb{R}$ , we have [39, Section 7.3]

$$|x + y|^v \leq (|x| + |y|)^v \leq 2^{v-1}(|x|^v + |y|^v). \tag{2.2}$$

For a set  $\mathfrak{C} \subseteq \mathbb{R}^{n_1 \times n_2}$ , the indicator function of  $\mathfrak{C}$  is defined as

$$\delta_{\mathfrak{C}}(X) = \begin{cases} 0, & \text{if } X \in \mathfrak{C}, \\ +\infty, & \text{otherwise.} \end{cases}$$

We list some notions used throughout this paper, which are summarized as follows.

- $\tilde{n} = \min\{n_1, n_2\}$  and  $n_m = \max\{n_1, n_2\}$ .
- $\mathbf{1}$  denotes a matrix with all entries being 1, whose corresponding dimension should often be clear from the context.

### 3. Matrix Completion with Poisson Observations

For an unknown low-rank matrix  $\bar{X} \in \mathbb{R}_+^{n_1 \times n_2}$ , which we aim to estimate, it is corrupted by Poisson noise and only partial entries are available. The entries of the observation  $Y$  is generated in the following way:

$$Y_{ij} = \text{Poisson}(\bar{X}_{ij} + c), \quad (i, j) \in \Omega,$$

where  $Y_{ij} = \text{Poisson}(\bar{X}_{ij} + c)$  represents that  $Y_{ij}$  follows a Poisson distribution with parameter  $\bar{X}_{ij} + c > 0, c > 0$  is a fixed background, and  $\Omega$  is a subset of indices  $\{1, 2, \dots, n_1\} \times \{1, 2, \dots, n_2\}$ . Here  $c$  is used to reflect the noise level of Poisson distribution since  $\bar{X}_{ij} + c$  is the parameter of Poisson distribution and the probability of the observed entry  $Y_{ij} = k$  is given by

$$\mathbb{P}\{Y_{ij} = k\} = \frac{(\bar{X}_{ij} + c)^k}{k!} \exp(-(\bar{X}_{ij} + c)), \quad k = 0, 1, 2, \dots$$

And the index set  $\Omega$  chosen at random with  $\mathbb{E}[\Omega] = m$ , where  $m = |\Omega|$  is the number of entries of observations. In particular, we assume that  $\Omega$  follows a Bernoulli model, i.e. each entry  $(i, j) \in \{1, 2, \dots, n_1\} \times \{1, 2, \dots, n_2\}$  is included in  $\Omega$  with probability  $m/(n_1 n_2)$  independently, which is denoted by  $\Omega \sim \text{Bern}(s)$  with  $s = m/(n_1 n_2)$ .

By the maximum likelihood estimation of Poisson observations (see, e.g. [26]), the data-fitting term between the observations and the underlying matrix  $X$  is the KL divergence given by

$$f_{\Omega, Y}(X) := \sum_{(i, j) \in \Omega} (X_{ij} + c) - Y_{ij} \log(X_{ij} + c). \quad (3.1)$$

Suppose that each entry of the underlying matrix  $X$  is nonnegative and bounded, i.e.  $0 \leq X_{ij} \leq \beta, 1 \leq i \leq n_1, 1 \leq j \leq n_2$ , which is commonly used in image processing since the pixels of real-world images are nonnegative and bounded. In many areas of real-world applications, the singular values of the underlying matrix exhibit only a gradual decay toward zero rather than an exact low-rank matrix. Therefore, we allow a relaxation of the assumption that  $X$  has rank  $r$  exactly. For any matrix with rank at most  $r$ , we have

$$\|X\|_* \leq \sqrt{r} \|X\|_F \leq \beta \sqrt{r n_1 n_2},$$

which is a relaxation of the conditions that  $\text{rank}(X) \leq r$  and  $0 \leq X_{ij} \leq \beta, 1 \leq i \leq n_1, 1 \leq j \leq n_2$ .

Besides the approximate low-rankness of the underlying matrix, we also assume that the underlying matrix is piecewise smooth, which is satisfied for many real-world matrices, such as image data. Then the TV constraint is employed to measure its piecewise smoothness, which is capable of preserving sharp edges of images. The TV was first proposed by Rudin *et al.* [41] for image denoising, and then generalized to image deconvolution [40]. By combining the nonnegativity, low-rankness and piecewise smoothness, we consider that the underlying matrix lies in the following set:

$$\mathfrak{D} := \{X \in \mathbb{R}^{n_1 \times n_2} \mid 0 \leq X_{ij} \leq \beta, \|X\|_* \leq \beta \sqrt{n_1 n_2 r}, \|X\|_{\text{TV}} \leq \tau\}, \quad (3.2)$$

where  $\|X\|_{\text{TV}}$  is the TV seminorm of  $X$  and  $\tau > 0$  is a given constant. Specifically, the discrete directional derivatives of  $X \in \mathbb{R}^{n_1 \times n_2}$  are defined as

$$(\nabla_x X)_{ij} = \begin{cases} X_{(i+1)j} - X_{ij}, & \text{if } 0 \leq i < n_1 - 1, \\ 0, & \text{otherwise,} \end{cases} \quad (3.3)$$

$$(\nabla_y X)_{ij} = \begin{cases} X_{i(j+1)} - X_{ij}, & \text{if } 0 \leq j < n_2 - 1, \\ 0, & \text{otherwise.} \end{cases} \quad (3.4)$$

Then the TV seminorm of  $X$  is defined as

$$\|X\|_{\text{TV}} := \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sqrt{(\nabla_x X)_{ij}^2 + (\nabla_y X)_{ij}^2}. \quad (3.5)$$

Combining (3.1) and (3.2), we propose the following constrained model for low-rank matrix completion with Poisson observations:

$$\begin{aligned} \min_X \quad & f_{\Omega, Y}(X) \\ \text{s.t.} \quad & X \in \mathfrak{D}, \end{aligned} \quad (3.6)$$

where  $f_{\Omega, Y}(X)$  is defined as (3.1).

The choice of TV in (3.5) is the isotropic version, which can also be applied to the anisotropic version of TV. The anisotropic TV just uses the  $\ell_1$  norm instead of the  $\ell_2$  norm in (3.5). The isotropic and anisotropic induced TV seminorms are thus equivalent up to a factor of  $\sqrt{2}$ . We emphasize here that our approach can be applied to both anisotropic and isotropic TV. For simplicity, we only discuss the isotropic TV version in detail. Moreover, the zero boundary conditions for the TV are used in (3.3) and (3.4) in our model. Other boundary conditions can also be employed in the TV seminorm and we refer the reader to [35] for a detailed discussion on how to address other boundary conditions.

**Remark 3.1.** In contrast to the model in [5], which only has nuclear norm constraint, our proposed model in (3.6) not only exploits the low-rankness, but also utilizes the piecewise smoothness of the underlying matrix. This is important to recover several real-world images with Poisson observations, where the underlying images are approximately low-rank and piecewise smooth simultaneously. The main advantage of both TV and nuclear norm constraints in (3.6) is that the low-rankness and piecewise smoothness of the underlying matrix can be exploited simultaneously, which are possessed for many real-world image data.

**Remark 3.2.** Recently, Zhang *et al.* [54] proposed a low-rank tensor completion approach for Poisson observations, where the transformed tensor nuclear norm constraint is utilized. However, the model in (3.6) incorporates the TV and nuclear norm constraints, which can explore the piecewise smoothness and global low-rankness of the underlying matrix simultaneously.

**Remark 3.3.** The constrained formulation of TV is preferable since many real-world matrices are not exactly piecewise smooth, e.g. images. The discrete directional derivatives of  $X$  exhibit only a gradual decay toward zero rather than an exactly sparse difference at each point [13].

**Remark 3.4.** In the literature, there are many work based on both low-rank and TV priors for matrix completion or tensor completion, see [23, 28, 31, 52, 57] and references therein. However, the type of noises are different between the above work and model (3.6). Besides, the above work did not analyze the error bounds of the corresponding models, while we will establish the error bound of model (3.6) in the following section.

## 4. Error Bounds

In this section, we establish an upper error bound between the optimal solution of (3.6) and the underlying matrix  $\bar{X}$ , whose proof follows the similar line of [5, Theorem 2]. Let  $\hat{X}$  be an optimal solution of (3.6). Then we have the following main result about the upper error bound of the estimator of (3.6).

**Theorem 4.1.** *Assume that  $\Omega$  is chosen at random following the Bernoulli sampling model, i.e.  $\Omega \sim \text{Bern}(s)$  with  $s = m/(n_1 n_2)$ . Then the following inequality holds with probability at least  $1 - 1/(n_1 n_2)$ :*

$$\begin{aligned} \frac{\|\hat{X} - \bar{X}\|_F^2}{n_1 n_2} &\leq C_{\beta c} (\beta + c + \log(n_1 n_2)) \min \left\{ \beta \sqrt{r} + \tilde{c}, \frac{\tau}{\sqrt{n_m}} + \beta + \tilde{c} \right\} \\ &\quad \times \sqrt{\frac{n_1 + n_2}{m}} \sqrt{1 + \frac{(n_1 + n_2) \log(n_1 + n_2)}{m}}, \end{aligned}$$

where

$$\tilde{c} = \max\{c, |c - 1|\}, \quad C_{\beta c} = \frac{C_2(\beta + c)}{e^{-T}}, \quad C_2 > 0, \quad T = \frac{\beta^2}{8c}.$$

If  $m \geq (n_1 + n_2) \log(n_1 + n_2)$ , then

$$\frac{\|\hat{X} - \bar{X}\|_F^2}{n_1 n_2} \leq \sqrt{2} C_{\beta c} (\beta + c + \log(n_1 n_2)) \min \left\{ \beta \sqrt{r} + \tilde{c}, \frac{\tau}{\sqrt{n_m}} + \beta + \tilde{c} \right\} \sqrt{\frac{n_1 + n_2}{m}}.$$

The detailed proof of Theorem 4.1 is delegated to the Appendix B. Theorem 4.1 shows that the mean square error of the estimator of model (3.6) is proportional to the minimum of the nuclear norm and TV seminorm constraints. Moreover, the upper error bound of the estimator may decrease if the rank of the underlying matrix is smaller. At the same time, in light of Theorem 4.1, we can see that the upper error bound of the estimator decreases as the number of the observed samples decreases.

**Remark 4.1.** In contrast to the upper error bound in [5, Theorem 2] for the single nuclear norm constraint, we add the TV constraint in the model (3.6), where both piecewise smoothness and low-rankness of the underlying matrix are utilized. In this case, the upper error bound is related to the minimum of the nuclear norm and TV seminorm constraints, which is not worse than that of only TV or nuclear norm constraint. In particular, if  $\beta \sqrt{r} > \tau/\sqrt{n_m} + \beta$ , the upper error bound in Theorem 4.1 is lower than that of [5, Theorem 2].

**Remark 4.2.** For fixed  $\beta$  and  $c$ , the upper error bound in Theorem 4.1 is on the order of

$$\min \left\{ \sqrt{r}, \frac{\tau}{\sqrt{n_m}} \right\} \sqrt{\frac{n_1 + n_2}{m}} \log(n_1 n_2),$$

when  $m \geq (n_1 + n_2) \log(n_1 + n_2)$ . Consequently, the nuclear norm and TV seminorm constraints will influence the upper error bound of the estimator in (3.6) simultaneously. In addition, if the TV seminorm of the underlying matrix is very small, the upper error bound of the estimator in Theorem 4.1 is smaller than that of [5, Theorem 2], which will be demonstrated by numerical experiments in Section 6.1.

## 5. Optimization Algorithm

In this section, we develop an ADMM based algorithm [14, 17] to solve the model in (3.6). Let

$$\begin{aligned}\mathfrak{D}_1 &:= \{X \in \mathbb{R}^{n_1 \times n_2} \mid 0 \leq X_{ij} \leq \beta\}, \\ \mathfrak{D}_2 &:= \{X \in \mathbb{R}^{n_1 \times n_2} \mid \|X\|_* \leq \beta\sqrt{n_1 n_2 r}\}, \\ \mathfrak{D}_3 &:= \{X \in \mathbb{R}^{n_1 \times n_2} \mid \|X\|_{\text{TV}} \leq \tau\}.\end{aligned}$$

Therefore, one has  $\mathfrak{D} = \mathfrak{D}_1 \cap \mathfrak{D}_2 \cap \mathfrak{D}_3$ . We let  $X = S, X = M, X = N$ . Then model (3.6) can be written equivalently in the following form:

$$\begin{aligned}\min_{X, S, M, N} \quad & f_{\Omega, Y}(X) + \delta_{\mathfrak{D}_1}(S) + \delta_{\mathfrak{D}_2}(M) + \delta_{\mathfrak{D}_3}(N) \\ \text{s.t.} \quad & X = S, \quad X = M, \quad X = N.\end{aligned}\tag{5.1}$$

The augmented Lagrangian function associated with (5.1) is given by

$$\begin{aligned}L(X, S, M, N, Z_1, Z_2, Z_3) &= f_{\Omega, Y}(X) + \delta_{\mathfrak{D}_1}(S) + \delta_{\mathfrak{D}_2}(M) + \delta_{\mathfrak{D}_3}(N) \\ &\quad + \langle Z_1, X - S \rangle + \langle Z_2, X - M \rangle + \langle Z_3, X - N \rangle \\ &\quad + \frac{\rho}{2}(\|X - S\|_F^2 + \|X - M\|_F^2 + \|X - N\|_F^2),\end{aligned}\tag{5.2}$$

where  $\rho > 0$  is a given parameter. Then the iteration template of ADMM for solving (5.1) is given as follows:

$$X^{k+1} = \arg \min_X \{L(X, S^k, M^k, N^k, Z_1^k, Z_2^k, Z_3^k)\},\tag{5.3}$$

$$(S^{k+1}, M^{k+1}, N^{k+1}) = \arg \min_{S, M, N} \{L(X^{k+1}, S, M, N, Z_1^k, Z_2^k, Z_3^k)\},\tag{5.4}$$

$$\begin{pmatrix} Z_1^{k+1} \\ Z_2^{k+1} \\ Z_3^{k+1} \end{pmatrix} = \begin{pmatrix} Z_1^k \\ Z_2^k \\ Z_3^k \end{pmatrix} + \varpi \rho \begin{pmatrix} X^{k+1} - S^{k+1} \\ X^{k+1} - M^{k+1} \\ X^{k+1} - N^{k+1} \end{pmatrix},\tag{5.5}$$

where  $\varpi \in (0, (1 + \sqrt{5})/2)$  is the step-length.

Let

$$A^k := \rho(S^k + M^k + N^k) - (\mathbf{1} + Z_1^k + Z_2^k + Z_3^k).$$

Then the optimal solution in (5.3) is given explicitly by

$$X^{k+1} = \mathcal{P}_\Omega \left( \frac{A^k - 3\rho c \mathbf{1} + \sqrt{(A^k - 3\rho c \mathbf{1})^2 + 12\rho(cA^k + \mathcal{P}_\Omega(Y))}}{6\rho} \right) + \mathcal{P}_{\bar{\Omega}} \left( \frac{A^k + \mathbf{1}}{3\rho} \right),\tag{5.6}$$

where the square and root are performed in a point-wise manner, and  $\mathcal{P}_\Omega : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^{n_1 \times n_2}$  is the projection onto the sampling pattern  $\Omega$ , i.e.

$$(\mathcal{P}_\Omega(X))_{ij} = \begin{cases} X_{ij}, & \text{if } (i, j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

Since each variable is separate in (5.4), we can solve the subproblems with respect to  $S, M, N$  independently. The optimal solution with respect to  $S$  is given by

$$(S^{k+1})_{ij} = \begin{cases} 0, & \text{if } W_{ij} < 0, \\ W_{ij}, & \text{if } 0 \leq W_{ij} \leq \beta, \\ \beta, & \text{otherwise,} \end{cases} \quad (5.7)$$

where  $W = X^{k+1} + Z_1^k/\rho$ . For the optimal solution with respect to  $M$ , it is the projection of  $X^{k+1} + Z_2^k/\rho$  onto  $\mathfrak{D}_2$ , which is given by

$$M^{k+1} = \Pi_{\mathfrak{D}_2} \left( X^{k+1} + \frac{Z_2^k}{\rho} \right), \quad (5.8)$$

where  $\Pi_{\mathfrak{D}_2}(X^{k+1} + Z_2^k/\rho)$  denotes the projection of  $X^{k+1} + Z_2^k/\rho$  onto  $\mathfrak{D}_2$ . Let  $Q = X^{k+1} + Z_2^k/\rho$ . Consider the SVD of  $Q$  as  $\sum_{i=1}^{\min\{n_1, n_2\}} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ . By [16, Lemma 2.1], the projection onto the nuclear norm ball constraint  $\mathfrak{D}_2$  is given by

$$\Pi_{\mathfrak{D}_2}(Q) = \begin{cases} Q, & \text{if } Q \in \mathfrak{D}_2, \\ \sum_{i=1}^{\min\{n_1, n_2\}} \max\{0, \sigma_i - t\} \mathbf{u}_i \mathbf{v}_i^T, & \text{otherwise,} \end{cases}$$

where  $t \geq 0$  is the unique solution to the equation

$$\sum_{i=1}^{\min\{n_1, n_2\}} \max\{0, \sigma_i - t\} = \beta \sqrt{n_1 n_2 r}.$$

Here we use a fast algorithm in [9] to search such unique  $t$  exactly and in finite time.

For the optimal solution with respect to  $N$ , it is the projection of  $X^{k+1} + Z_3^k/\rho$  onto the TV seminorm ball constraint  $\mathfrak{D}_3$ , i.e.

$$N^{k+1} = \Pi_{\mathfrak{D}_3} \left( X^{k+1} + \frac{Z_3^k}{\rho} \right). \quad (5.9)$$

However, one cannot get its exact solution. Recently, Fadili *et al.* [13] proposed a first-order algorithm for solving such projection problem efficiently. The projection onto the TV seminorm ball constraint is computed by a dual formulation, which yields an iterative soft thresholding algorithm to the dual vector field. We also refer the reader to [13] for more detailed discussions about the projection onto the TV seminorm ball constraint.

Now we are ready to state the ADMM for solving (5.1) in Algorithm 5.1.

**Algorithm 5.1:** Alternating Direction Method of Multipliers for Solving (5.1).

**Input.** Let  $\rho > 0, \varpi \in (0, (1 + \sqrt{5})/2)$  be given constants.

Choose  $S^0, M^0, N^0, Z_1^0, Z_2^0, Z_3^0$ .

For  $k = 0, 1, 2, \dots$ , perform the following steps:

**Step 1.** Compute  $X^{k+1}$  via (5.6).

**Step 2.** Compute  $S^{k+1}, M^{k+1}, N^{k+1}$  by (5.7), (5.8), (5.9), respectively.

**Step 3.** Update  $Z_1^{k+1}, Z_2^{k+1}, Z_3^{k+1}$  via (5.5).

Since there are only two blocks  $X$  and  $(S, M, N)$  in the objective of (5.1), the ADMM in Algorithm 5.1 is convergent [14, Appendix B]. For brevity, we omit the details of the convergence of ADMM here.

## 6. Numerical Results

In this section, numerical examples are presented to demonstrate the effectiveness of the proposed low-rank and TV constraints (LRTVC) method for matrix completion with Poisson observations. We compare LRTVC with only the TV constraint (TVC) method and only the nuclear norm constraint (NNC) method [5], respectively, where the objective functions of NNC and TVC are the same as that of LRTVC.

The Karush-Kuhn-Tucker (KKT) conditions of model (5.1) are given as follows:

$$\begin{cases} \nabla_X f_{\Omega, Y}(X) + Z_1 + Z_2 + Z_3 = 0, \\ 0 \in \partial\delta_{\mathfrak{D}_1}(S) - Z_1, \quad 0 \in \partial\delta_{\mathfrak{D}_2}(M) - Z_2, \quad 0 \in \partial\delta_{\mathfrak{D}_3}(N) - Z_3, \\ X - S = 0, \quad X - M = 0, \quad X - N = 0, \end{cases}$$

where  $\nabla_X$  denotes the gradient of  $f_{\Omega, Y}$  with respect to  $X$  and  $\partial g$  denotes the subdifferential of a function  $g$ . The following KKT residual is adopted to measure the accuracy of an approximate optimal solution for model (5.1):

$$\eta_{\max} := \max\{\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7\},$$

where

$$\begin{aligned} \eta_1 &= \frac{\|\mathcal{P}_{\Omega}(\mathbf{1} - \mathcal{P}_{\Omega}(Y)/(X + c\mathbf{1})) + Z_1 + Z_2 + Z_3\|_F}{1 + \|X\|_F + \|Z_1\|_F + \|Z_2\|_F + \|Z_3\|_F}, \\ \eta_2 &= \frac{\|S - \Pi_{\mathfrak{D}_1}(Z_1 + S)\|_F}{1 + \|S\|_F + \|Z_1\|_F}, \quad \eta_3 = \frac{\|M - \Pi_{\mathfrak{D}_2}(Z_2 + M)\|_F}{1 + \|M\|_F + \|Z_2\|_F}, \\ \eta_4 &= \frac{\|N - \Pi_{\mathfrak{D}_3}(Z_3 + N)\|_F}{1 + \|N\|_F + \|Z_3\|_F}, \quad \eta_5 = \frac{\|X - S\|_F}{1 + \|X\|_F + \|S\|_F}, \\ \eta_6 &= \frac{\|X - M\|_F}{1 + \|X\|_F + \|M\|_F}, \quad \eta_7 = \frac{\|X - N\|_F}{1 + \|X\|_F + \|N\|_F}. \end{aligned}$$

Here / stands for the point-wise division. In our experiments, Algorithm 5.1 will be terminated when  $\eta_{\max} < 10^{-4}$  or the maximum number of iterations reaches 200.

The signal-to-noise ratio (SNR) is used to evaluate the quality of the recovered matrix, which is defined as

$$\text{SNR} := 10 \log_{10} \frac{\|\overline{X}\|_F^2}{\|X - \overline{X}\|_F^2},$$

where  $X$  and  $\overline{X}$  denote the recovered and original matrices, respectively. Moreover, the structural similarity index (SSIM) [48] is also used to measure the quality of the recovered data.

The sampling ratio (SR) is defined as  $\text{SR} = |\Omega|/(n_1 n_2)$ , where  $\Omega$  is generated uniformly at random and  $|\Omega|$  denotes the cardinality of  $\Omega$ . Since Poisson noise is data-dependent [49], the noise level of the observed data is related to the pixel intensity of  $\overline{X}$ . In order to test different noise levels for Poisson observations, different peak intensities of the underlying data matrix are considered. The noisy and missing data in our experiments is simulated as follows: The original matrix is first scaled with the peak intensities and the background of the scaled matrix is added. Then the Poisson noise is added using the function `poissrnd` in MATLAB. Finally, we generate the index set  $\Omega$  and get the observed matrix.

In Algorithm 5.1, we set the parameter  $\varpi$  as 1.618 for fast convergence in the experiments. The penalty parameter  $\rho$  is chosen from the set  $\{0.1, 0.01, 0.001, 0.0001, 0.00001\}$  to get the

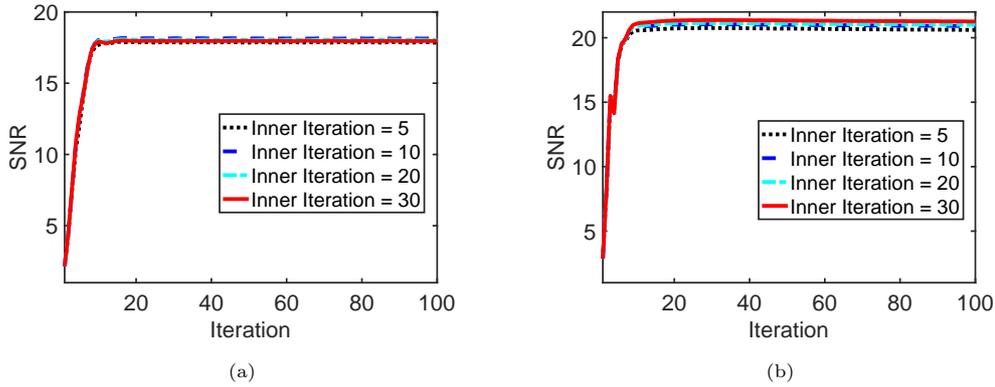


Fig. 6.1. SNR values versus number of outer iterations in Algorithm 5.1 for different number of inner iterations: (a) Cameraman, (b) Chart.

best recovery performance in all experiments. For computing the projection onto TV seminorm constraint, we employ the first-order projection algorithm in [13] to get an approximately optimal solution. We first investigate the performance of different number of inner iterations for computing the projection onto the TV seminorm constraint, which is chosen from the set  $\{5, 10, 20, 30\}$ . Fig. 6.1 shows the SNR values versus number of outer iteration in Algorithm 5.1 for different number of inner iterations, where  $\beta = 50$ ,  $c = 10$ , and  $SR = 0.5$  for the Cameraman and Chart images. Here the details of the testing images can be found in Section 6.2. It can be observed from this figure that the SNR values increase as the number of outer iterations increases. Moreover, the final performance of different number of inner iterations of computing the projection onto the TV seminorm constraint is very close for the two images. However, larger number of inner iterations implies to take more CPU time. Therefore, we choose the number of iterations as 20 for computing the projection onto the TV seminorm ball constraint in (5.9) in the following experiments for simplicity.

### 6.1. Synthetic data

In this subsection, we test three kinds of synthetic data to validate the effectiveness of the LRTVC. The details of the three data are given as follows:

- We constructed a low-rank matrix (called Data I) while its TV seminorm is large. Data I is generated as follows:  $X = UV^T$ , where  $U \in \mathbb{R}_+^{n_1 \times r}$ ,  $V \in \mathbb{R}_+^{n_2 \times r}$  and  $T$  stands for the transpose of a matrix. Here  $U$  and  $V$  are generated by the MATLAB commands `rand(n1, r)` and `rand(n2, r)`, respectively. We set  $n_1 = n_2 = 200$  and  $r = 2$ . In this case,  $X$  is low-rank while its TV seminorm is large.
- The second data (called Data II) is a symmetric Toeplitz matrix with size  $200 \times 200$ , where the first ten elements are 0 and others are 1 for the first row. The visual image of this matrix is shown in Fig. 6.2(a). The TV seminorm of Data II is small while it is a full rank matrix.
- The third data (called Data III) is a full one matrix with size  $200 \times 200$  but its central region with size  $51 \times 51$  is a zero matrix, whose visual image is shown in Fig. 6.2(b). Data III is not only low-rank but also its TV seminorm is small.

Table 6.1 shows the SNR and SSIM values of different methods for the three testing data with different  $\beta$  and  $c$ , where the sampling ratio is 50%. It can be seen from this table that the performance of LRTVC is best compared with that of NNC and TVC if the data is not only low-rank but also its TV seminorm is small, see the results of Data III, where the TVC outperforms NNC for the testing cases with different  $\beta$  and  $c$ . For Data I, which is low-rank but its TV seminorm is large, the NNC and LRTVC perform much better than TVC in terms of SNR and SSIM values. This phenomenon is due to the low-rankness of Data I while its TV seminorm is large. And the performance of NNC and LRTVC is nearly the same, where the SNR of LRTVC is just slightly better than that of NNC. The main reason is that the TV term has little influence on the performance of matrix recovery for this data and the parameters may also influence of the performance of LRTVC slightly. Similarly, for Data II, the performance obtained by TVC and LRTVC is much better than that obtained by NNC in terms of SNR and SSIM values, which is due to Data II is not low-rank while its TV seminorm is small. At the same time, the performance of TVC and LRTVC is almost the same since the nuclear norm constrain has little influence for this data, which also demonstrates that the performance of LRTVC is not worse than that of NNC and TVC in Theorem 4.1.

Fig. 6.3 shows the SNR values versus sampling ratios of different methods for the testing three data, where  $\beta = 100$  and  $c = 1$ . We can observe from the three figures that the SNR values obtained by the three methods increase as the sampling ratios increase for different data. When the data is just low-rank, the NNC and LRTVC perform much better than TVC in terms of SNR values, see Fig. 6.3(a). Moreover, the performance of LRTVC is almost the same as that of NNC for different sampling ratios since Data I is only low-rank and the TV term has little influence for this data. Similarly, when the TV seminorm of the data is small while it is not low-rank, the TVC and LRTVC perform much better than NNC in terms of SNR, see

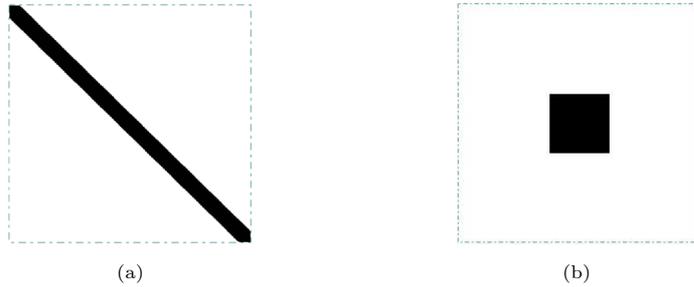


Fig. 6.2. The visual images of Data II and Data III: (a) Data II. (b) Data III.

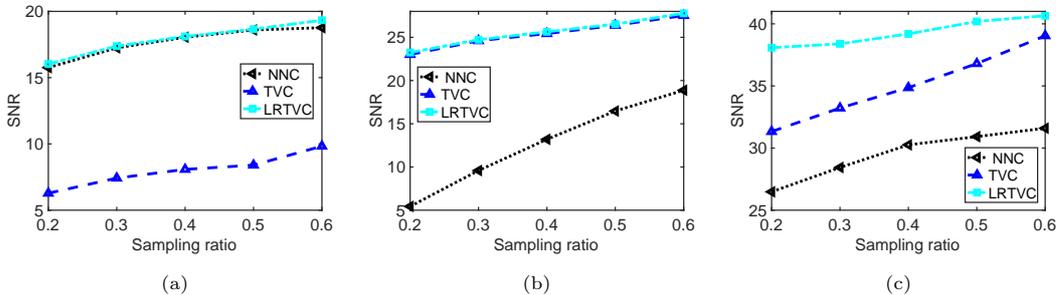


Fig. 6.3. SNR versus sampling ratio of different methods for synthetic data, where  $\beta = 100$  and  $c = 1$ : (a) Data I. (b) Data II. (c) Data III.

Fig. 6.3(b). It can be seen from this figure that the performance of TVC and LRTVC is almost the same for different sampling ratios since Data II is not low-rank and the nuclear norm term has little influence for this data. When the TV seminorm and nuclear norm of the data are both small (e.g. Data III), the performance of LRTVC is much better than that of NNC and TVC. At the same time, the performance of TVC is better than that of NNC in terms of SNR values, see Fig. 6.3(c).

In the following, we consider different relaxations for the nuclear norm and TV seminorm constraints. Let  $\beta^* = \|X_0\|_* / \sqrt{n_1 n_2 r}$ , where  $X_0$  is the true matrix and  $r$  is the rank of  $X_0$ . We set the nuclear norm constraint as  $\|X\|_* \leq \tilde{\beta} \sqrt{n_1 n_2 r}$  in our model and discuss different  $\tilde{\beta}$ . Table 6.2 shows the performance of different  $\tilde{\beta}$  of NNC and LRTVC for Data I. It can be seen that when  $\tilde{\beta} = \beta^*$ , the SNR values of recovered matrices are highest for both NNC and LRTVC. Moreover, the decreases of SNR of NNC and LRTVC for  $\tilde{\beta} > \beta^*$  are smaller than those for  $\tilde{\beta} < \beta^*$ , which implies that large  $\tilde{\beta}$  is more safe than small  $\tilde{\beta}$  in order to get good recovery performance. Besides, the TV seminorm of the recovered matrix of LRTVC is smaller than that of NNC for most cases, where the TV seminorm of the underlying matrix is  $1.02 \times 10^4$ .

Let  $\tau^* = \|X_0\|_{\text{TV}}$ . We set  $\|X\|_{\text{TV}} \leq \tilde{\tau}$  and discuss different  $\tilde{\tau}$  for the constraints of TVC and LRTVC, where Data II is tested with 50% sampling ratio and  $\beta = 50, c = 1$ . Here  $\tau^* = 3.80 \times 10^4$  for Data II. Table 6.3 shows the SNR, nuclear norm, and TV seminorm of the recovered matrix by TVC and LRTVC. It can be seen from this table that both TVC and LRTVC obtain highest SNR values when  $\tilde{\tau} = \tau^*$ . Moreover, the decreases of SNR values of TVC and LRTVC when

Table 6.1: SNR and SSIM values of different methods for the synthetic data with 50% sampling ratio.

	$\beta$	$c$	SNR			SSIM		
			NNC	TVC	LRTVC	NNC	TVC	LRTVC
Data I	50	1	17.47	8.29	17.67	0.9678	0.5177	0.9713
		10	16.90	7.90	16.97	0.9688	0.4350	0.9699
	100	1	18.20	8.95	18.41	0.9787	0.6327	0.9774
		10	17.13	8.68	18.19	0.9699	0.6128	0.9718
Data II	50	1	16.38	25.92	25.97	0.1721	0.9421	0.9427
		10	15.87	25.63	25.65	0.1572	0.9416	0.9428
	100	1	16.60	26.49	26.82	0.1911	0.9462	0.9493
		10	16.35	26.20	26.23	0.1799	0.9437	0.9455
Data III	50	1	27.84	36.58	38.36	0.7296	0.9938	0.9947
		10	26.92	35.92	38.29	0.6715	0.9853	0.9864
	100	1	30.94	38.38	39.99	0.8379	0.9958	0.9972
		10	30.18	37.16	39.89	0.8019	0.9926	0.9916

Table 6.2: SNR, nuclear norm, and TV seminorm of different methods for Data I with 50% sampling ratio, where  $\beta = 50$  and  $c = 1$ .

	$\tilde{\beta} - \beta^*$	NNC			LRTVC		
		SNR	$\ X\ _*$	$\ X\ _{\text{TV}}$	SNR	$\ X\ _*$	$\ X\ _{\text{TV}}$
Data I	-10	8.50	$3.85 \times 10^3$	$4.12 \times 10^5$	8.84	$6.86 \times 10^3$	$5.10 \times 10^5$
	5	13.89	$5.22 \times 10^3$	$5.60 \times 10^5$	14.00	$7.36 \times 10^3$	$5.65 \times 10^5$
	0	17.47	$6.45 \times 10^3$	$6.77 \times 10^5$	17.67	$7.32 \times 10^3$	$6.69 \times 10^5$
	5	17.14	$7.78 \times 10^3$	$7.10 \times 10^5$	17.60	$8.16 \times 10^3$	$6.99 \times 10^5$
	10	16.92	$9.39 \times 10^3$	$7.65 \times 10^5$	17.15	$9.18 \times 10^3$	$6.99 \times 10^5$

$\tilde{\tau} > \tau^*$  are smaller than those when  $\tilde{\tau} < \tau^*$ . In fact, when  $\tilde{\tau} = \tau^* + 20000$ , the SNR values of TVC and LRTVC are much higher than those when  $\tilde{\tau} = \tau^* - 20000$ . This implies that larger regions of the TV seminorm constraint are more safe than smaller regions in order to get good recovery performance.

In Table 6.4, we show the SNR, nuclear norm, and TV seminorm of the recovered matrices of different relaxations of the nuclear norm and TV seminorm constraints for Data III, where the sampling ratio is 50% and  $\beta = 50, c = 1$ . Here  $r = 2, \beta^* = 40.54, \tau^* = 1.02 \times 10^4$ . It can be seen

Table 6.3: SNR, nuclear norm, and TV seminorm of different methods for Data II with 50% sampling ratio, where  $\beta = 50$  and  $c = 1$ .

	$\tilde{\tau} - \tau^*$	TVC			LRTVC		
		SNR	$\ X\ _*$	$\ X\ _{\text{TV}}$	SNR	$\ X\ _*$	$\ X\ _{\text{TV}}$
Data II	-20000	18.12	$1.91 \times 10^4$	$2.87 \times 10^4$	18.58	$2.01 \times 10^4$	$3.05 \times 10^4$
	10000	24.16	$2.41 \times 10^4$	$3.06 \times 10^4$	24.44	$2.41 \times 10^4$	$3.25 \times 10^4$
	0	25.92	$2.59 \times 10^4$	$3.82 \times 10^4$	25.97	$2.62 \times 10^4$	$3.82 \times 10^4$
	10000	24.51	$2.60 \times 10^4$	$4.18 \times 10^4$	25.10	$2.61 \times 10^4$	$4.81 \times 10^4$
	20000	23.53	$2.60 \times 10^4$	$5.10 \times 10^4$	24.20	$2.62 \times 10^4$	$5.81 \times 10^4$

Table 6.4: SNR, nuclear norm, and TV seminorm of different methods for Data III with 50% sampling ratio, where  $\beta = 50$  and  $c = 1$ .

$\tilde{\beta} - \beta^*$	$\tilde{\tau} - \tau^*$	NNC			TVC			LRTVC		
		SNR	$\ X\ _*$	$\ X\ _{\text{TV}}$	SNR	$\ X\ _*$	$\ X\ _{\text{TV}}$	SNR	$\ X\ _*$	$\ X\ _{\text{TV}}$
-10	-4000	12.70	$8.63 \times 10^3$	$6.50 \times 10^4$	25.10	$1.22 \times 10^4$	$1.00 \times 10^4$	12.80	$8.62 \times 10^3$	$9.88 \times 10^3$
	-2000							12.73	$8.65 \times 10^3$	$9.88 \times 10^3$
	0							12.70	$8.64 \times 10^3$	$1.03 \times 10^4$
	2000							12.72	$8.64 \times 10^3$	$1.22 \times 10^4$
	4000							12.73	$8.64 \times 10^3$	$1.41 \times 10^4$
-5	-4000	18.51	$1.00 \times 10^4$	$6.10 \times 10^4$	25.10	$1.22 \times 10^4$	$1.00 \times 10^4$	18.23	$1.01 \times 10^4$	$1.04 \times 10^4$
	-2000							18.57	$1.00 \times 10^4$	$1.05 \times 10^4$
	0							18.67	$1.00 \times 10^4$	$1.05 \times 10^4$
	2000							18.67	$1.00 \times 10^4$	$1.22 \times 10^4$
	4000							18.67	$1.00 \times 10^4$	$1.41 \times 10^4$
0	-4000	28.13	$1.14 \times 10^4$	$6.21 \times 10^4$	25.10	$1.22 \times 10^4$	$1.00 \times 10^4$	25.44	$1.15 \times 10^4$	$1.03 \times 10^4$
	-2000							29.09	$1.18 \times 10^4$	$1.01 \times 10^4$
	0							36.13	$1.16 \times 10^4$	$1.09 \times 10^4$
	2000							35.25	$1.19 \times 10^4$	$1.15 \times 10^4$
	4000							34.25	$1.22 \times 10^4$	$1.25 \times 10^4$
5	-4000	25.641	$1.28 \times 10^4$	$1.07 \times 10^5$	25.10	$1.22 \times 10^4$	$1.00 \times 10^4$	25.34	$1.21 \times 10^4$	$1.04 \times 10^4$
	-2000							29.66	$1.17 \times 10^4$	$1.05 \times 10^4$
	0							36.03	$1.17 \times 10^4$	$1.08 \times 10^4$
	2000							34.59	$1.20 \times 10^4$	$1.25 \times 10^4$
	4000							33.49	$1.22 \times 10^4$	$1.44 \times 10^4$
10	-4000	22.96	$1.42 \times 10^4$	$1.56 \times 10^5$	25.10	$1.22 \times 10^4$	$1.00 \times 10^4$	24.92	$1.21 \times 10^4$	$1.04 \times 10^4$
	-2000							29.63	$1.18 \times 10^4$	$1.06 \times 10^4$
	0							34.33	$1.17 \times 10^4$	$1.08 \times 10^4$
	2000							34.30	$1.19 \times 10^4$	$1.25 \times 10^4$
	4000							32.23	$1.23 \times 10^4$	$1.44 \times 10^4$

from this table that the NNC, TVC, and LRTVC obtain highest SNR values when  $\tilde{\tau} = \tau^*$  and  $\tilde{\beta} = \beta^*$ . For larger  $\tilde{\tau}$ , e.g.  $\tilde{\tau} = \tau^* + 2000$ , the SNR values of TVC and LRTVC decrease slightly compared with those of smaller  $\tilde{\tau}$ . In fact, the SNR values obtained by TVC and LRTVC for  $\tilde{\tau} < \tau^*$  are much smaller than those for  $\tilde{\tau} = \tau^*$ , whose decreases are much larger than those for  $\tilde{\tau} > \tau^*$ . The same phenomenon occurs for different  $\tilde{\beta}$  for NNC and LRTVC. In particular, when  $\tilde{\tau} = \tau^*$  and  $\tilde{\beta} = \beta^*$ , the SNR values obtained by LRTVC are higher than those obtained by NNC and TVC, where TVC performs better than NNC in terms of SNR values.

### 6.2. Real-world image data

In this subsection, we test three real-world images to demonstrate the effectiveness of the LRTVC method. The testing images include Cameraman ( $256 \times 256$ ), Chart ( $256 \times 256$ ), and Airplane ( $256 \times 256$ )<sup>1</sup>, which are shown in Fig. 6.4. It can be seen from this figure that the low-rankness of Chart and Airplane is better than that of Cameraman.

Now we discuss the parameters  $\tau, \beta, r$  in the LRTVC model for the real-world images. In Fig. 6.5(a), we show the SNR values versus  $r$  of different  $\beta$  for the Chart image, where the peak value of the underlying image is 100,  $c = 10$ , and  $SR = 0.5$ . It can be seen from this figure that the SNR values obtained by LRTVC keep almost the same for different  $r$ . And  $\beta$  has much influence than  $r$  for the recovery performance. Therefore, we choose  $r = 20$  in all experiments in order to balance the CPU time and recovery quality. Moreover, when  $\beta$  is larger than the peak value 100, e.g.  $\beta = 110, 120$ , its performance is almost the same as that when  $\beta = 100$ .

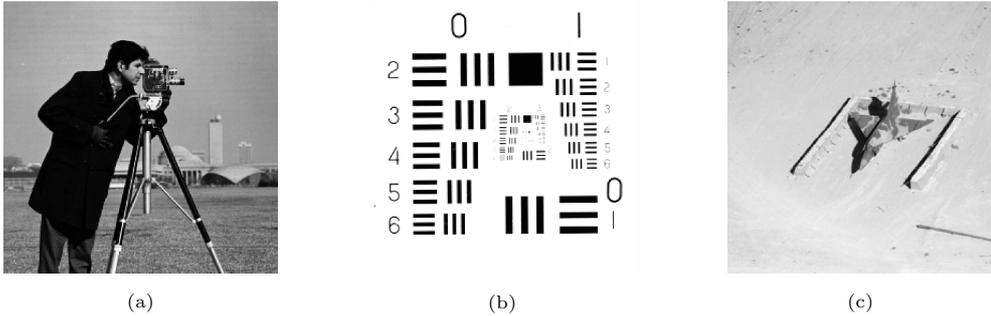


Fig. 6.4. Original images: (a) Cameraman. (b) Chart. (c) Airplane.

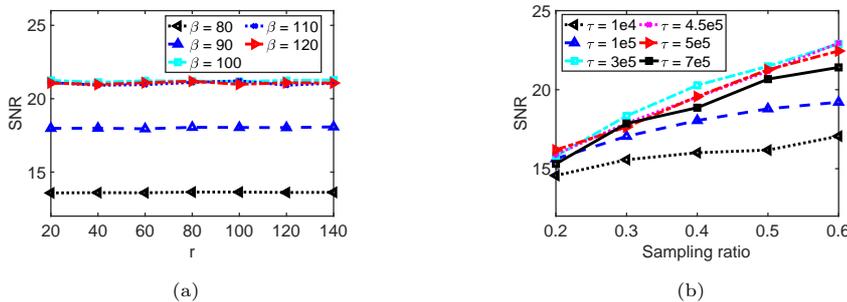


Fig. 6.5. Performance of different parameters of LRTVC for the Chart image: (a) SNR values versus  $r$  for different  $\beta$ . (b) SNR values versus sampling ratio for different  $\tau$ .

<sup>1</sup> <http://sipi.usc.edu/database/>

While the performance is worse when  $\beta$  is smaller than 100, e.g.  $\beta = 80, 90$ . As a consequence, slightly larger  $\beta$  is more safe to get good recovery performance. In Fig. 6.5(b), we show the SNR values versus sampling ratio for different  $\tau$ , where the true TV seminorm is almost  $4.5 \times 10^5$ . It can be observed from this figure that when the chosen value  $\tau$  is around the true value, e.g. from  $3 \times 10^5$  to  $5 \times 10^5$ , the SNR values obtained by LRTVC are very close. However, when  $\tau$  is too large or too small, the performance of LRTVC is not good in terms of SNR values. In our following experiments, we choose  $\tau$  from the set  $\{1 \times 10^5, 2 \times 10^5\}$  for the Cameraman image, the set  $\{2.5 \times 10^5, 5.5 \times 10^5\}$  for the Chart image, and the set  $\{4 \times 10^4, 9 \times 10^4\}$  for the Airplane image, to get the best recovery performance.

In Table 6.5, we show the SNR and SSIM values of different methods for the three testing images, where the sampling ratio is 40%. It can be observed that the TVC and LRTVC perform almost the same for the Cameraman for different  $\beta$  and  $c$ , which are better than NNC. This phenomenon is due to the fact that the low-rankness of the Cameraman image is not well. For the Chart and Airplane images, the SNR and SSIM values obtained by LRTVC are much higher than those obtained by TVC for different  $\beta$  and  $c$ , which are higher than those of NNC. These demonstrate that when the images are low-rank and piecewise smooth, the performance of LRTVC is much better than that of NNC and TVC.

In Fig. 6.6, we show the SNR versus sampling ratio for the testing three images, where  $\beta = 50$  and  $c = 1$ . It can be seen that the SNR values increase as the sampling ratios increase for the three images. Moreover, for the Cameraman image, the performance of TVC and LRTVC is almost the same, which performs much better than NNC in terms of SNR values. This is due to the fact that low-rankness of the Cameraman image is not well while the piecewise smoothness of this image is more obvious. For the Chart and Airplane images, their low-rankness is better than that of the Cameraman image. And the LRTVC outperforms TVC in terms of SNR values for different sampling ratios, where both LRTVC and TVC perform much better than NNC for different sampling ratios.

In Fig. 6.7, we show the visual quality of the recovered images by NNC, TVC, and LRTVC for the Cameraman, Chart, and Airplane images, where  $\beta = 50$ ,  $c = 10$ , and SR = 0.6. It can be

Table 6.5: SNR and SSIM values of different methods for the Cameraman, Chart, and Airplane images with 40% sampling ratio.

Image	$\beta$	$c$	SNR			SSIM		
			NNC	TVC	LRTVC	NNC	TVC	LRTVC
Cameraman	50	1	12.99	18.44	18.48	0.2679	0.7131	0.7134
		10	12.62	18.13	18.21	0.2422	0.6783	0.6977
	100	1	14.35	19.17	19.18	0.3314	0.7507	0.7540
		10	13.95	19.08	19.14	0.3071	0.7340	0.7349
Chart	50	1	17.39	19.29	20.66	0.4370	0.6636	0.6782
		10	16.88	19.11	20.17	0.4192	0.6539	0.6756
	100	1	19.06	19.71	21.29	0.5125	0.6351	0.6888
		10	18.79	19.14	21.10	0.5007	0.6251	0.6820
Airplane	50	1	19.28	25.62	26.51	0.2572	0.7318	0.8049
		10	18.86	25.33	25.87	0.2407	0.7210	0.7965
	100	1	20.69	26.73	27.23	0.3264	0.7999	0.8056
		10	20.41	26.71	27.10	0.3136	0.7959	0.8018

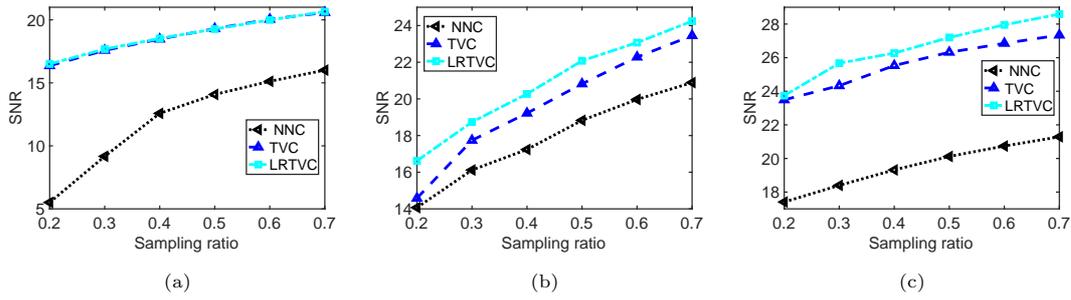


Fig. 6.6. SNR versus sampling ratio of different methods for the real-world images, where  $\beta = 50$  and  $c = 1$ : (a) Cameraman. (b) Chart. (c) Airplane.

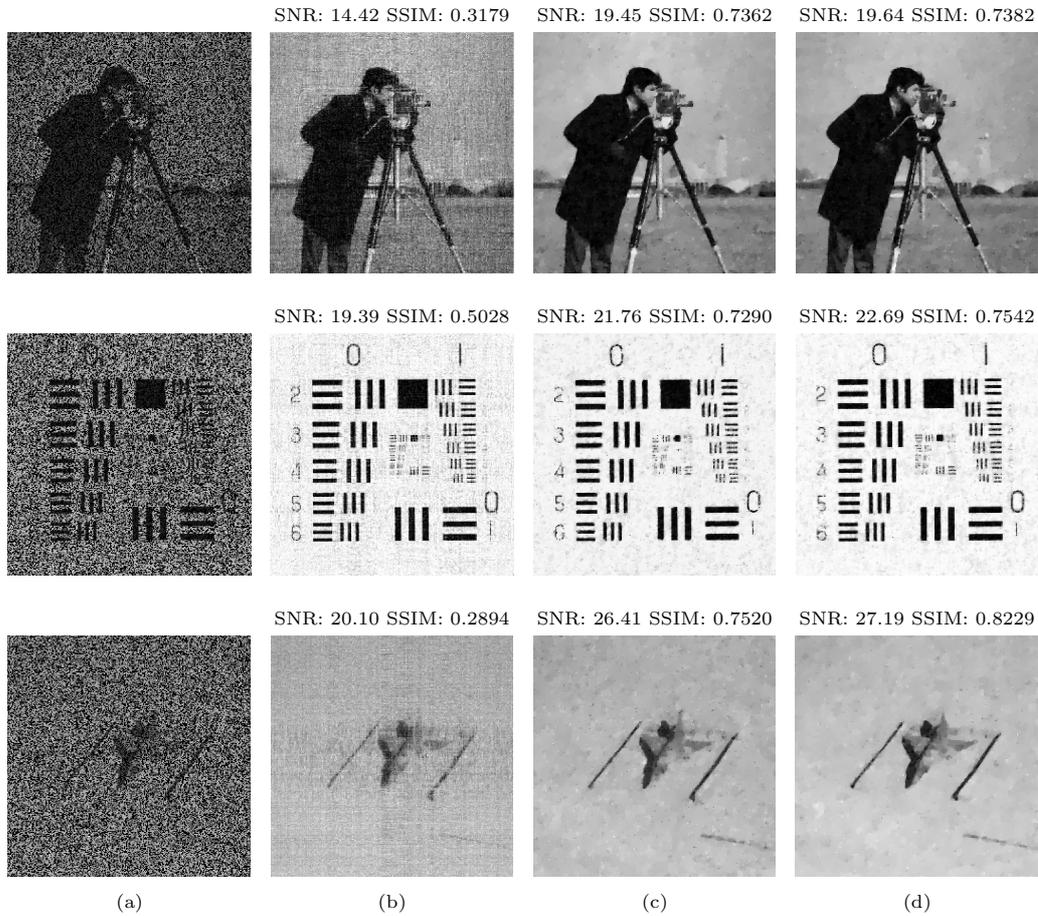


Fig. 6.7. Recovered images by NNC, TVC, and LRTVC on Cameraman, Chart, and Airplane, where  $\beta = 50$ ,  $c = 10$ , and  $SR = 0.6$ : (a) Observed images. (b) Recovered images by NNC. (c) Recovered images by TVC. (d) Recovered images by LRTVC.

seen from the figure that the visual quality recovered by TVC and LRTVC is almost the same for the Cameraman image, which is better than NNC. For the Chart and Airplane images, the images recovered by LRTVC are more clear than those recovered by TVC and NNC. Besides, the TVC performs better than NNC in terms of visual quality for the two images.

## 7. Concluding Remarks

In this paper, we have proposed a novel approach combined with TV and nuclear norm constraints for low-rank matrix completion with Poisson observations, where the KL divergence is derived by the maximum likelihood estimation of Poisson noise for the data-fitting term. The TV seminorm and nuclear norm constraints are utilized to explore the piecewise smoothness and low-rankness of the underlying matrix data, respectively. Furthermore, an upper error bound of the estimator of the LRTVC model is established with high probability, which is not larger than that of single nuclear norm constraint [5] or TV seminorm constraint. Then an ADMM based algorithm is designed to solve the resulting optimization model. The LRTVC is validated to outperform other compared methods both theoretically and empirically.

Since we only establish the upper bound of the estimator of model (3.6), we will discuss the minimax lower bound for our problem in the future. Besides, an interesting direction for future work is to extend the Poisson matrix completion problem with TV and nuclear norm constraints to the tensor case, which can address higher order data (cf. [54,57]). Another future research direction involves the fast algorithms by using matrix factorization methods instead of using the nuclear norm constraint for the low-rank matrix completion problem with Poisson observations, which can reduce the computational complexity further. Moreover, since many real-world images are corrupted by Gaussian noise and Poisson noise simultaneously, we will also go to extend our model to address this kind of mixed noise in the future.

## Appendix A

In the appendix, we establish the central inequality of  $f_{\Omega,Y}(X)$  under the constraint set  $\mathfrak{D}$  in (3.2).

**Lemma A.1.** *Let  $f_{\Omega,Y}(X)$  and  $\mathfrak{D}$  be defined as (3.1) and (3.2), respectively. Suppose that  $\Omega$  is chosen at random following the Bernoulli sampling model, i.e.  $\Omega \sim \text{Bern}(s)$  with  $s = m/(n_1 n_2)$ . Then the following inequality holds with probability at least  $1 - 1/(n_1 n_2)$ :*

$$\begin{aligned} & \sup_{X \in \mathfrak{D}} |f_{\Omega,Y}(X) - \mathbb{E}[f_{\Omega,Y}(X)]| \\ & \leq \frac{C_1}{c} ((\beta + c)(e^2 - 1) + 4 \log(n_1 n_2)) \min \left\{ \beta \sqrt{r} + \tilde{c}, \frac{\tau}{\sqrt{n_m}} + \beta + \tilde{c} \right\} \\ & \quad \times \sqrt{m(n_1 + n_2) + n_1 n_2 \log(n_1 + n_2)}, \end{aligned}$$

where  $C_1 > 0$  is a given constant and  $\tilde{c} = \max\{c, |c - 1|\}$ .

*Proof.* Let

$$\begin{aligned} \omega & := \frac{C_1}{c} ((\beta + c)(e^2 - 1) + 4 \log(n_1 n_2)) \min \left\{ \beta \sqrt{r} + \tilde{c}, \frac{\tau}{\sqrt{n_m}} + \beta + \tilde{c} \right\} \\ & \quad \times \sqrt{m(n_1 + n_2) + n_1 n_2 \log(n_1 + n_2)}. \end{aligned}$$

By Markov's inequality, for any  $p \geq 1$ , one has

$$\mathbb{P} \left( \sup_{X \in \mathfrak{D}} |f_{\Omega,Y}(X) - \mathbb{E}[f_{\Omega,Y}(X)]| \geq \omega \right) \leq \frac{\mathbb{E}}{\omega^p} \left[ \sup_{X \in \mathfrak{D}} |f_{\Omega,Y}(X) - \mathbb{E}[f_{\Omega,Y}(X)]|^p \right]. \quad (\text{A.1})$$

Next we estimate an upper bound of  $\mathbb{E}[\sup_{X \in \mathfrak{D}} |f_{\Omega, Y}(X) - \mathbb{E}[f_{\Omega, Y}(X)]|^p]$ . Let  $\epsilon_{ij}$  be a Rademacher sequence taking values 1 and  $-1$  with probability  $1/2$ , respectively,  $\sigma_{\Omega}(i, j)$  be 1 if  $(i, j) \in \Omega$ , otherwise 0. In this case,  $\sigma_{\Omega}(i, j)$  is a random variable taking value 1 with probability  $m/(n_1 n_2)$ , otherwise 0 with probability  $1 - m/(n_1 n_2)$ . Notice that

$$\begin{aligned} & \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} |f_{\Omega, Y}(X) - \mathbb{E}[f_{\Omega, Y}(X)]|^p \right] \tag{A.2} \\ & \leq 2^p \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \left| \sum_{i, j} \epsilon_{ij} \sigma_{\Omega}(i, j) ((X_{ij} + c) - Y_{ij} \log(X_{ij} + c)) \right|^p \right] \\ & \leq 2^{2p-1} \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \left| \sum_{i, j} \epsilon_{ij} \sigma_{\Omega}(i, j) Y_{ij} (-\log(X_{ij} + c)) \right|^p + \sup_{X \in \mathfrak{D}} \left| \sum_{i, j} \epsilon_{ij} \sigma_{\Omega}(i, j) (X_{ij} + c) \right|^p \right] \\ & = 2^{2p-1} \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \left| \sum_{i, j} \epsilon_{ij} \sigma_{\Omega}(i, j) Y_{ij} (-\log(X_{ij} + c)) \right|^p \right] + 2^{2p-1} \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \left| \sum_{i, j} \epsilon_{ij} \sigma_{\Omega}(i, j) (X_{ij} + c) \right|^p \right], \end{aligned}$$

where the first inequality follows from [1, Theorem 14.3] and the second inequality follows from (2.2). Let  $R, P_{\Omega} \in \mathbb{R}^{n_1 \times n_2}$  represent two random matrices whose entries are given by  $\epsilon_{ij}$  and  $\sigma_{ij}$ , respectively. Then we get that

$$\begin{aligned} & \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \left| \sum_{i, j} \epsilon_{ij} \sigma_{\Omega}(i, j) (X_{ij} + c) \right|^p \right] \\ & = \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} |\langle R \circ P_{\Omega}, X + c\mathbf{1} \rangle|^p \right] \\ & \leq \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \|R \circ P_{\Omega}\|^p \|X + c\mathbf{1}\|_*^p \right], \tag{A.3} \end{aligned}$$

where  $\circ$  denotes the Hadamard product of two matrices and the inequality holds by the Cauchy-Swartz inequality. Let  $a$  be the average value of all entries of  $X$ . It can be verified directly that

$$\begin{aligned} \|X + c\mathbf{1}\|_* & = \|X - a\mathbf{1} + a\mathbf{1} + c\mathbf{1}\|_* \\ & \leq \|X - a\mathbf{1}\|_* + \|(a + c)\mathbf{1}\|_* \\ & \leq \sqrt{\tilde{n}} \|X - a\mathbf{1}\|_F + (a + c) \sqrt{n_1 n_2} \\ & \leq \sqrt{\tilde{n}} \|X - a\mathbf{1}\|_{\text{TV}} + (\beta + c) \sqrt{n_1 n_2}, \tag{A.4} \end{aligned}$$

where the last inequality holds by [34, Proposition 7] and  $a \leq \beta$ . By the definition of TV, we know that  $\|X\|_{\text{TV}} = \|X - a\mathbf{1}\|_{\text{TV}}$ . Note that  $\|X + c\mathbf{1}\|_* \leq \|X\|_* + c\sqrt{n_1 n_2}$ . As a consequence, we have

$$\begin{aligned} \sup_{X \in \mathfrak{D}} \{\|X + c\mathbf{1}\|_*\} & \leq \min \{(\beta\sqrt{r} + c)\sqrt{n_1 n_2}, \sqrt{\tilde{n}}\tau + (\beta + c)\sqrt{n_1 n_2}\} \\ & = \min \left\{ \beta\sqrt{r} + c, \frac{\tau}{\sqrt{n_m}} + \beta + c \right\} \sqrt{n_1 n_2}. \tag{A.5} \end{aligned}$$

Plugging (A.5) into (A.3), we obtain that

$$\begin{aligned} & \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \left| \sum_{i, j} \epsilon_{ij} \sigma_{\Omega}(i, j) (X_{ij} + c) \right|^p \right] \\ & \leq \left( \min \left\{ \beta\sqrt{r} + c, \frac{\tau}{\sqrt{n_m}} + \beta + c \right\} \right)^p (n_1 n_2)^{\frac{p}{2}} \mathbb{E} [\|R \circ P_{\Omega}\|^p]. \tag{A.6} \end{aligned}$$

By the proof of [11, Lemma A.1], we know that

$$\mathbb{E}[\|R \circ P_\Omega\|^p] \leq C_0^p \left( \sqrt{\frac{m(n_1 + n_2) + n_1 n_2 \log(n_1 + n_2)}{n_1 n_2}} \right)^p, \quad (\text{A.7})$$

where  $C_0 > 0$  is a given constant.

Let  $Z = X + (c - 1)\mathbf{1}$  and  $a$  be the average value of all entries of  $X$ . Then we observe through simple calculations that

$$\begin{aligned} \|Z\|_* &= \|X + (c - 1)\mathbf{1}\|_* \\ &= \|X - a\mathbf{1} + a\mathbf{1} + (c - 1)\mathbf{1}\|_* \\ &\leq \|X - a\mathbf{1}\|_* + \|(a + c - 1)\mathbf{1}\|_* \\ &\leq \sqrt{\tilde{n}}\|X - a\mathbf{1}\|_{\text{TV}} + (a + |c - 1|)\sqrt{n_1 n_2} \\ &= \sqrt{\tilde{n}}\|X\|_{\text{TV}} + (a + |c - 1|)\sqrt{n_1 n_2} \\ &\leq \sqrt{\tilde{n}}\tau + (a + |c - 1|)\sqrt{n_1 n_2}, \end{aligned} \quad (\text{A.8})$$

where the second inequality follows from (A.4) and the third equality holds by  $\|X - a\mathbf{1}\|_{\text{TV}} = \|X\|_{\text{TV}}$ . For the first term of (A.2), by the proof of [5, Lemma 4], we have

$$\begin{aligned} &\mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \left| \sum_{i,j} \epsilon_{ij} \sigma_\Omega(i, j) Y_{ij} (-\log(X_{ij} + c)) \right|^p \right] \\ &\leq \left( \frac{2}{c} \right)^p \mathbb{E} \left[ \max_{i,j} Y_{ij}^p \right] \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} |\langle R \circ P_\Omega, Z \rangle|^p \right] \\ &\leq \left( \frac{2}{c} \right)^p \mathbb{E} \left[ \max_{i,j} Y_{ij}^p \right] \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} \|R \circ P_\Omega\|^p \|Z\|_*^p \right] \\ &\leq \left( \frac{2}{c} \right)^p \mathbb{E} \left[ \max_{i,j} Y_{ij}^p \right] \mathbb{E}[\|R \circ P_\Omega\|^p] \\ &\quad \times \left( \min \left\{ \beta\sqrt{r} + |c - 1|, \frac{\tau}{\sqrt{n_m}} + \beta + |c - 1| \right\} \sqrt{n_1 n_2} \right)^p, \end{aligned} \quad (\text{A.9})$$

where the last inequality holds by (A.8) and

$$\|Z\|_* \leq \|X\|_* + |c - 1|\sqrt{n_1 n_2} \leq (\beta\sqrt{r} + |c - 1|)\sqrt{n_1 n_2}.$$

For the Poisson observation  $Y_{ij}$ , by the proof of [5, Lemma 4], we get that

$$\begin{aligned} \mathbb{E} \left[ \max_{i,j} Y_{ij}^p \right] &\leq 2^{2p-1} \left( (\beta + c)^p + t_0^p + \mathbb{E} \left[ \max_{i,j} S_{ij}^p \right] \right) \\ &\leq 2^{2p-1} \left( (\beta + c)^p + t_0^p + 2^{p-1} (2p! + \log^p(n_1 n_2)) \right), \end{aligned} \quad (\text{A.10})$$

where  $t_0 = (\beta + c)(e^2 - 3)$  and  $S_{ij}$  is the independent and identically distributed exponential random variables with parameter 1, and the second inequality follows from [54, Lemma 2.1]. Combining (A.2), (A.6), (A.7), and (A.9), we get that

$$\left( \mathbb{E} \left[ \sup_{X \in \mathfrak{D}} |f_{\Omega, Y}(X) - \mathbb{E}[f_{\Omega, Y}(X)]|^p \right] \right)^{\frac{1}{p}}$$

$$\begin{aligned}
&\leq C_0 2^{2-\frac{1}{p}} \left\{ \min \left\{ \beta\sqrt{r} + c, \frac{\tau}{\sqrt{n_m}} + \beta + c \right\} + \frac{2}{c} \left( \mathbb{E} \left[ \max_{i,j} Y_{ij}^p \right] \right)^{\frac{1}{p}} \right. \\
&\quad \left. \times \min \left\{ \beta\sqrt{r} + |c-1|, \frac{\tau}{\sqrt{n_m}} + \beta + |c-1| \right\} \right\} \\
&\quad \times \sqrt{m(n_1 + n_2) + n_1 n_2 \log(n_1 + n_2)} \\
&\leq \frac{8C_0}{c} ((\beta + c)(e^2 - 1) + 4 \log(n_1 n_2)) \min \left\{ \beta\sqrt{r} + \tilde{c}, \frac{\tau}{\sqrt{n_m}} + \beta + \tilde{c} \right\} \\
&\quad \times \sqrt{m(n_1 + n_2) + n_1 n_2 \log(n_1 + n_2)},
\end{aligned}$$

where the second inequality follows from the fact that

$$(a^p + b^p + c^p + d^p)^{\frac{1}{p}} \leq a + b + c + d$$

for any  $a, b, c, d > 0$  and  $\tilde{c} = \max\{c, |c-1|\}$ ,  $p = \log(n_1 n_2)$ . Therefore, the probability in (A.1) is

$$\left( \frac{8C_0}{C_1} \right)^{\log(n_1 n_2)} \leq \frac{1}{n_1 n_2},$$

if  $C_1 \geq 8C_0 e$ . This completes the proof.  $\square$

## Appendix B. Proof of Theorem 4.1

For any  $X \in \mathfrak{D}$ , we have

$$\begin{aligned}
f_{\Omega, Y}(\bar{X}) - f_{\Omega, Y}(X) &= f_{\Omega, Y}(\bar{X}) - \mathbb{E}[f_{\Omega, Y}(\bar{X})] + \mathbb{E}[f_{\Omega, Y}(\bar{X})] \\
&\quad - \mathbb{E}[f_{\Omega, Y}(X)] + \mathbb{E}[f_{\Omega, Y}(X)] - f_{\Omega, Y}(X) \\
&\leq \mathbb{E}[f_{\Omega, Y}(\bar{X})] - \mathbb{E}[f_{\Omega, Y}(X)] + |f_{\Omega, Y}(X) - \mathbb{E}[f_{\Omega, Y}(X)]| \\
&\quad + |f_{\Omega, Y}(\bar{X}) - \mathbb{E}[f_{\Omega, Y}(\bar{X})]| \\
&\leq \mathbb{E}[f_{\Omega, Y}(\bar{X})] - \mathbb{E}[f_{\Omega, Y}(X)] \\
&\quad + 2 \sup_{X \in \mathfrak{D}} |f_{\Omega, Y}(X) - \mathbb{E}[f_{\Omega, Y}(X)]|.
\end{aligned} \tag{B.1}$$

Note that

$$\begin{aligned}
\mathbb{E}[f_{\Omega, Y}(\bar{X})] - \mathbb{E}[f_{\Omega, Y}(X)] &= \frac{m}{n_1 n_2} \sum_{i,j} -(\bar{X}_{ij} + c) \log \left( \frac{\bar{X}_{ij} + c}{X_{ij} + c} \right) + \bar{X}_{ij} - X_{ij} \\
&= -\frac{m}{n_1 n_2} \sum_{i,j} K(\bar{X}_{ij} + c || X_{ij} + c) \\
&= -mK(\bar{X} + c || X + c).
\end{aligned} \tag{B.2}$$

Since  $\hat{X}$  is an optimal solution of (3.6), we get that

$$0 \leq f_{\Omega, Y}(\bar{X}) - f_{\Omega, Y}(\hat{X}) \leq -mK(\bar{X} + c || \hat{X} + c) + 2 \sup_{X \in \mathfrak{D}} |f_{\Omega, Y}(X) - \mathbb{E}[f_{\Omega, Y}(X)]|,$$

which implies that

$$K(\bar{X} + c || \hat{X} + c) \leq \frac{2}{m} \sup_{X \in \mathfrak{D}} |f_{\Omega, Y}(X) - \mathbb{E}[f_{\Omega, Y}(X)]|. \tag{B.3}$$

It follows from (2.1) that the average KL divergence can be bounded below by the average square Hellinger distance for two matrices. Consequently, we obtain that

$$H^2(\bar{X} + c\|\hat{X} + c) \leq K(\bar{X} + c\|\hat{X} + c). \quad (\text{B.4})$$

For  $M, X \in \mathfrak{D}$ , by [5, Lemma 5], we can deduce that

$$H^2(M + c\|X + c) \geq \frac{1 - e^{-T}}{4(\beta + c)T} \frac{\|M - X\|_F^2}{n_1 n_2} \geq \frac{e^{-T}}{4(\beta + c)} \frac{\|M - X\|_F^2}{n_1 n_2}, \quad (\text{B.5})$$

where  $T = \beta^2/(8c)$  and the last inequality holds by the fact that  $1 - e^{-T} \geq T e^{-T}$ . Combining (B.3)-(B.5), Lemma A.1, and the fact that  $n_1 n_2 \leq (n_1 + n_2)^2$ , we get that the following inequality holds with probability at least  $1 - 1/(n_1 n_2)$ :

$$\begin{aligned} \frac{\|\hat{X} - \bar{X}\|_F^2}{n_1 n_2} &\leq \frac{8C_1(\beta + c)}{e^{-T}} ((\beta + c)(e^2 - 1) + 4 \log(n_1 n_2)) \min \left\{ \beta\sqrt{r} + \tilde{c}, \frac{\tau}{\sqrt{n_m}} + \beta + \tilde{c} \right\} \\ &\quad \times \sqrt{\frac{n_1 + n_2}{m}} \sqrt{1 + \frac{(n_1 + n_2) \log(n_1 + n_2)}{m}} \\ &\leq C_{\beta c}(\beta + c + \log(n_1 n_2)) \min \left\{ \beta\sqrt{r} + \tilde{c}, \frac{\tau}{\sqrt{n_m}} + \beta + \tilde{c} \right\} \\ &\quad \times \sqrt{\frac{n_1 + n_2}{m}} \sqrt{1 + \frac{(n_1 + n_2) \log(n_1 + n_2)}{m}}, \end{aligned}$$

where  $\tilde{c} = \max\{c, |c - 1|\}$  and  $C_{\beta c} = C_2(\beta + c)/e^{-T}$  with  $C_2 = 56C_1 > 0$ . In particular, when  $m \geq (n_1 + n_2) \log(n_1 + n_2)$ , we have

$$\sqrt{1 + \frac{(n_1 + n_2) \log(n_1 + n_2)}{m}} \leq \sqrt{2}.$$

As a consequence, the desired conclusion can be obtained easily.  $\square$

**Acknowledgments.** The research of D. Qiu was supported in part by the National Natural Science Foundation of China (Grant No. 12201473) and by the Science Foundation of Wuhan Institute of Technology (Grant No. K202256). The research of M.K. Ng was supported in part by the HKRGC GRF (Grant Nos. 12300218, 12300519, 17201020, 17300021). The research of X. Zhang was supported in part by the National Natural Science Foundation of China (Grant No. 12171189), by the Knowledge Innovation Project of Wuhan (Grant No. 2022010801020279), and by the Fundamental Research Funds for the Central Universities (Grant No. CCNU22JC023).

## References

- [1] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer, 2011.
- [2] E.J. Candès, X. Li, Y. Ma, and J. Wright, Robust principal component analysis? *J. ACM*, **58**:3 (2011), 1–37.
- [3] E.J. Candès and Y. Plan, Matrix completion with noise, *Proc IEEE Inst. Electr. Electron. Eng.*, **98**:6 (2010), 925–936.
- [4] E.J. Candès and B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.*, **9**:6 (2009), 717–772.

- [5] Y. Cao and Y. Xie, Poisson matrix recovery and completion, *IEEE Trans. Signal Process.*, **64**:6 (2016), 1609–1620.
- [6] R. Chan, H. Yang, and T. Zeng, A two-stage image segmentation method for blurry images with Poisson or multiplicative Gamma noise, *SIAM J. Imaging Sci.*, **7**:1 (2014), 98–127.
- [7] H. Chang, Y. Lou, Y. Duan, and S. Marchesini, Total variation-based phase retrieval for Poisson noise removal, *SIAM J. Imaging Sci.*, **11**:1 (2018), 24–55.
- [8] Y. Chi, Y.M. Lu, and Y. Chen, Nonconvex optimization meets low-rank matrix factorization: An overview, *IEEE Trans. Signal Process.*, **67**:20 (2019), 5239–5269.
- [9] L. Condat, Fast projection onto the simplex and the  $\ell_1$  ball, *Math. Program.*, **158**:1-2 (2016), 575–585.
- [10] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, Springer, 2008.
- [11] M.A. Davenport, Y. Plan, E. van den Berg, and M. Wootters, 1-Bit matrix completion, *Information and Inference: A Journal of the IMA*, **3**:3 (2014), 189–223.
- [12] M.A. Davenport and J. Romberg, An overview of low-rank matrix recovery from incomplete observations, *IEEE J. Sel. Top. Signal Process.*, **10**:4 (2016), 608–622.
- [13] J.M. Fadili and G. Peyré, Total variation projection with first order schemes, *IEEE Trans. Image Process.*, **20**:3 (2011), 657–669.
- [14] M. Fazel, T.K. Pong, D. Sun, and P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, *SIAM J. Matrix Anal. Appl.*, **34**:3 (2013), 946–977.
- [15] M.A. Figueiredo and J.M. Bioucas-Dias, Restoration of Poissonian images using alternating direction optimization, *IEEE Trans. Image Process.*, **19**:12 (2010), 3133–3145.
- [16] D. Garber, On the convergence of projected-gradient methods with low-rank projections for smooth convex minimization over trace-norm balls and related problems, *SIAM J. Optim.*, **31**:1 (2021), 727–753.
- [17] R. Glowinski and A. Marroco, Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires, *Rev. Fr. Autom. Inf. Rech. Oper.*, **9**:2 (1975), 41–76.
- [18] P.J. Green, Bayesian reconstructions from emission tomography data using a modified EM algorithm, *IEEE Trans. Med. Imaging*, **9**:1 (1990), 84–93.
- [19] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, Weighted nuclear norm minimization and its applications to low level vision, *Int. J. Comput. Vis.*, **121**:2 (2017), 183–208.
- [20] W. He, H. Zhang, H. Shen, and L. Zhang, Hyperspectral image denoising using local low-rank matrix recovery and global spatial-spectral total variation, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **11**:3 (2018), 713–729.
- [21] W. He, H. Zhang, L. Zhang, and H. Shen, Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration, *IEEE Trans. Geosci. Remote Sens.*, **54**:1 (2016), 178–188.
- [22] Z. Hu, F. Nie, R. Wang, and X. Li, Low rank regularization: A review, *Neural Netw.*, **136** (2021), 218–232.
- [23] T.Y. Ji, T.Z. Huang, X.L. Zhao, T.H. Ma, and G. Liu, Tensor completion using total variation and low-rank matrix factorization, *Inf. Sci.*, **326** (2016), 243–257.
- [24] X. Jiang, G. Raskutti, and R. Willett, Minimax optimal rates for Poisson inverse problems with physical constraints, *IEEE Trans. Inf. Theory*, **61**:8 (2015), 4458–4474.
- [25] R.H. Keshavan, A. Montanari, and S. Oh, Matrix completion from a few entries, *IEEE Trans. Inf. Theory*, **56**:6 (2010), 2980–2998.
- [26] T. Le, R. Chartrand, and T.J. Asaki, A variational approach to reconstructing images corrupted by Poisson noise, *J. Math. Imaging Vis.*, **27**:3 (2007), 257–263.
- [27] J. Li, J.F. Cai, and H. Zhao, Robust inexact alternating optimization for matrix completion with outliers, *J. Comput. Math.*, **38**:2 (2020), 337–354.
- [28] X. Li, Y. Ye, and X. Xu, Low-rank tensor completion with total variation for visual data inpainting, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, **31** (2017), 2210–2216.

- [29] Y. Li, C. Chen, N. Liu, H. Huang, Z. Zheng, and Q. Yan, A blockchain-based decentralized federated learning framework with committee consensus, *IEEE Netw.*, **35**:1 (2021), 234–241.
- [30] L. Ma, L. Moisan, J. Yu, and T. Zeng, A dictionary learning approach for Poisson image deblurring, *IEEE Trans. Med. Imaging*, **32**:7 (2013), 1277–1289.
- [31] L. Ma, L. Xu, and T. Zeng, Low rank prior and total variation regularization for image deblurring, *J. Sci. Comput.*, **70**:3 (2017), 1336–1357.
- [32] A.D. McRae and M.A. Davenport, Low-rank matrix completion and denoising under Poisson noise, *Information and Inference: A Journal of the IMA*, **10**:2 (2021), 697–720.
- [33] R. Molina, On the hierarchical Bayesian approach to image restoration: Applications to astronomical images, *IEEE Trans. Pattern Anal. Mach. Intell.*, **16**:11 (1994), 1122–1128.
- [34] D. Needell and R. Ward, Stable image reconstruction using total variation minimization, *SIAM J. Imaging Sci.*, **6**:2 (2013), 1035–1058.
- [35] M.K. Ng, R.H. Chan, and W.C. Tang, A fast algorithm for deblurring models with Neumann boundary conditions, *SIAM J. Sci. Comput.*, **21**:3 (1999), 851–866.
- [36] J.M. Ollinger and J.A. Fessler, Positron-emission tomography, *IEEE Signal Process. Mag.*, **14**:1 (1997), 43–55.
- [37] M. Raginsky, R.M. Willett, Z.T. Harmany, and R.F. Marcia, Compressed sensing performance bounds under Poisson noise, *IEEE Trans. Signal Process.*, **58**:8 (2010), 3990–4002.
- [38] B. Recht, M. Fazel, and P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Rev.*, **52**:3 (2010), 471–501.
- [39] H.L. Royden and P. Fitzpatrick, *Real Analysis*, Prentice-Hall, 2010.
- [40] L.I. Rudin and S. Osher, Total variation based image restoration with free local constraints, in: *Proc. 1st Int. Conf. Image Process.*, IEEE, **1** (1994), 31–35.
- [41] L.I. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, **60**:1-4 (1992), 259–268.
- [42] A.V. Sambasivan and J.D. Haupt, Minimax lower bounds for noisy matrix completion under sparse factor models, *IEEE Trans. Inf. Theory*, **64**:5 (2018), 3274–3285.
- [43] S. Setzer, G. Steidl, and T. Teuber, Deblurring Poissonian images by split Bregman techniques, *J. Vis. Commun. Image R.*, **21**:3 (2010), 193–199.
- [44] F. Shi, J. Cheng, L. Wang, P.T. Yap, and D. Shen, LRTV: MR image super-resolution with low-rank and total variation regularizations, *IEEE Trans. Med. Imaging*, **34**:12 (2015), 2459–2466.
- [45] A. Soni, S. Jain, J. Haupt, and S. Gonella, Noisy matrix completion under sparse factor models, *IEEE Trans. Inf. Theory*, **62**:6 (2016), 3636–3661.
- [46] F. Taherkhani, H. Kazemi, and N.M. Nasrabadi, Matrix completion for graph-based deep semi-supervised learning, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, **33** (2019), 5058–5065.
- [47] Y.X. Wang, C.M. Lee, L.F. Cheong, and K.C. Toh, Practical matrix completion and corruption recovery using proximal alternating robust subspace, *Int. J. Comput. Vis.*, **111**:3 (2015), 315–344.
- [48] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.*, **13**:4 (2004), 600–612.
- [49] Y. Wen, R.H. Chan, and T. Zeng, Primal-dual algorithms for total variation based image restoration under Poisson noise, *Sci. China Math.*, **59**:1 (2016), 141–160.
- [50] M.N. Wernick and J.N. Aarsvold, *Emission Tomography: The Fundamentals of PET and SPECT*, Academic Press, 2004.
- [51] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero, Efficient gradient projection methods for edge-preserving removal of Poisson noise, *Inverse Probl.*, **25**:4 (2009), 045010.
- [52] L. Zhang, L. Song, B. Du, and Y. Zhang, Nonlocal low-rank tensor completion for visual data, *IEEE Trans. Cybern.*, **51**:2 (2021), 673–685.
- [53] X. Zhang, B. Javidi, and M.K. Ng, Automatic regularization parameter selection by generalized cross-validation for total variational Poisson noise removal, *Appl. Opt.*, **56**:9 (2017), D47–D51.

- [54] X. Zhang and M.K. Ng, Low rank tensor completion with Poisson observations, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44**:8 (2022), 4239–4251.
- [55] X. Zhang, M.K. Ng, and M. Bai, A fast algorithm for deconvolution and Poisson noise removal, *J. Sci. Comput.*, **75**:3 (2018), 1535–1554.
- [56] M. Zhao, Y.W. Wen, M. Ng, and H. Li, A nonlocal low rank model for Poisson noise removal, *Inverse Probl. Imaging*, **15**:3 (2021), 519–537.
- [57] Y.B. Zheng, T.Z. Huang, T.Y. Ji, X.L. Zhao, T.X. Jiang, and T.H. Ma, Low-rank tensor completion via smooth matrix factorization, *Appl. Math. Model.*, **70** (2019), 677–695.