SELECTING REGULARIZATION PARAMETERS FOR NUCLEAR NORM TYPE MINIMIZATION PROBLEMS[*]

KEXIN LI[†], HONGWEI LI[‡], RAYMOND H. CHAN[§], AND YOU-WEI WEN[†]

Abstract. The reconstruction of low-rank matrix from its noisy observation finds its usage in many applications. It can be reformulated into a constrained nuclear norm minimization problem, where the bound η of the constraint is explicitly given or can be estimated by the probability distribution of the noise. When the Lagrangian method is applied to find the minimizer, the solution can be obtained by the singular value thresholding operator where the thresholding parameter λ is related to the Lagrangian multiplier. In this paper, we first show that the Frobenius norm of the discrepancy between the minimizer and the observed matrix is a strictly increasing function of λ. From that we derive a closed-form solution for λ in terms of η. The result can be used to solve the constrained nuclear-norm-type minimization problem when η is given. For the unconstrained nuclear-norm-type regularized problems, our result allows us to automatically choose a suitable regularization parameter by using the discrepancy principle. The regularization parameters obtained are comparable to (and sometimes better than) those obtained by Stein's unbiased risk estimator (SURE) approach while the cost of solving the minimization problem can be reduced by 11–18 times. Numerical experiments with both synthetic data and real MRI data are performed to validate the proposed approach.

Key words. low-rank matrix, singular value thresholding, discrepancy principle, nuclear norm, regularization parameter

AMS subject classifications. 68U10, 94A08, 90C99, 65K99

1. Introduction. In the past two decades, the low-rank matrix recovery problem has attracted much attention in computer vision, pattern recognition, image processing, machine learning, and optimization communities. The problem arises in many applications such as denoising [9, 19], inpainting [12, 34], matrix completion [7, 8], background subtraction [43, 50], and image alignment [45]. In these applications, we often need to solve the constrained minimization problem

$$\min_{\|X-Y\|_F^2 \leq \eta} f(X) \tag{1.1}$$

to recover an unknown low-rank matrix $X \in \mathbb{R}^{m \times n}$ from its noisy observation Y, which is corrupted by Gaussian white noise. Here $\|\cdot\|_F$ denotes the Frobenius norm defined as the square root of the sum of the squares of its elements. Without loss of generality, we assume that $m \leq n$. The function $f(X)$ is a regularization function of X which incorporates the prior information on the cleaned matrix X, and it can be the rank of X, the nuclear norm (NN) [6, 8], the truncated nuclear norm (TNN) [27, 31, 33] or the generalized weighted nuclear norm (GWNN) [22, 53] of the matrix X, etc. When $\|Y\|_F^2 \leq \eta$, we have a trivial solution $X = \mathbf{0}$ as $f(\mathbf{0}) = 0$. In order to

[†]Key Laboratory of Computing and Stochastic Mathematics (LCSM), School of Mathematics and Statistics, Hunan Normal University, Changsha, Hunan, China. (Corresponding author: Y. Wen Email: wenyouwei@gmail.com)

[‡]Beijing Advanced Innovation Center for Imaging Theory and Technology Capital Normal University, Beijing, China.

[§]Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong SAR, China; Hong Kong Centre for Cerebro-Cardiovascular Health Engineering.

exclude this case, we assume that

$$\|Y\|_F^2 > \eta > 0 \tag{1.2}$$

throughout this paper and we look for non-trivial solutions of (1.1).

Generally, a constrained minimization problem can be solved by projected gradient descent method [2, 29], Lagrangian multiplier method [2] and so on. When the projected gradient descent method is applied, the objective function $f(X)$ should be differentiable. However, the functions $f(X)$ considered in this paper are non-differentiable, so the projected gradient descent method cannot be directly applied to find the minimizers.

The Lagrangian multiplier method is also widely applied to solve the constrained minimization problem, which converts the constrained minimization problem (1.1) into a regularized one as follows

$$\min_X f(X) + \frac{1}{2\lambda}\|X - Y\|_F^2, \tag{1.3}$$

where $1/\lambda$ is the Lagrangian multiplier associated with the inequality constraint. In regularization theory, λ is also called the regularization parameter. If the singular value decomposition (SVD) of Y is given, the optimal solution of (1.3) can be obtained by soft thresholding of Y's singular values when the regularization function is the nuclear norm [6, 8]. The threshold value equals to the regularization parameter λ, see Theorem 2.2 below or [6]. By an abuse of notation, we use λ to refer to both the threshold value and the regularization parameter. It is a crucial issue to choose λ. If λ is too small, then the shrinkage is insufficient and the corresponding solution is still noisy; while if λ is too large, informative structures might be removed together with the noise and the solution is a poor approximation of the cleaned matrix.

Ideally, the optimal λ should minimize the mean-squared error (MSE) between the true matrix and the estimated matrix. It is difficult to implement this approach because the true matrix is unknown. Some literatures have discussed how to choose an optimal regularization parameter for the unconstrained minimization problem (1.3). In [18], Gavish and Donoho studied the asymptotic MSE (AMSE) and proposed the AMSE-optimal choice of hard threshold, which is $(4/\sqrt{3})\sqrt{n}\tau$ when $m = n$ and the standard deviation τ of the noise is known. In [19], Gavish *et al.* applied the asymptotic framework to find the optimal threshold value, either analytically or numerically, for a variety of loss functions, including Frobenius norm, the nuclear norm and the operator norm. In [52], Yadav *et al.* applied random matrix theory to infer the AMSE without the knowledge of the true matrix X. In [9], Candés *et al.* applied Stein's unbiased risk estimator (SURE), an unbiased estimate of MSE, to choose the soft threshold. They also gave an expression for the divergence of the estimated matrix with respect to the observation matrix. In [3], Bigot *et al.* derived the corresponding generalized SURE (GSURE) formula for different noise types, which further realized the adaptive thresholding of singular values. Deledalle *et al.* [10] proposed the Stein unbiased gradient estimator of the risk (SUGAR) and provided an asymptotically unbiased estimate of the gradient of the risk, which is an effective strategy to automatically optimize a collection of parameters.

In fact, the SURE approach requires the knowledge of the noise variance. When the knowledge of the noise variance is unknown, the generalized cross validation (GCV) method can be applied to choose the regularization parameter [20]. In [30], Josse *et al.* integrated the SURE method and the GCV method to handle the case

when the variance is unknown. Furthermore, Hansen *et al.* [25] proposed a method for choosing the regularization parameter based on the L-curve, which is a log-log plot of the norm of a regularized solution versus the norm of the corresponding residual norm. The optimal regularization parameter is the corner point of the L-curve, see [21, 26].

The discrepancy principle is also widely applied to choose the regularization parameter [1, 42, 46, 49]. It states that the regularization parameter λ should be chosen such that the minimizer $X(\lambda)$ of (1.3) has a discrepancy that equals to the bound η if η is explicitly given, that is

$$\|X(\lambda) - Y\|_F^2 = \eta. \tag{1.4}$$

Obviously, (1.4) is the complementary condition associated with the constrained minimization problem (1.1), which shows that there is a mapping between the parameter $\lambda > 0$ and the bound η. Given a regularization parameter λ, it is trivial to obtain the bound η through directly calculating $X(\lambda)$ and the squared Frobenius norm of $X(\lambda) - Y$. It is a forward problem. However, given the bound η it is not trivial to obtain the corresponding regularization parameter λ because it is the root of the nonlinear equation (1.4) and it is an inverse problem. Existing work on the discrepancy principle is mainly for Tikhonov [1, 5] and TV [49] regularization terms. To the best of our knowledge, there is no literature on considering how to obtain λ for a given η in the nuclear-norm-type regularization minimization problems.

In this paper, we aim precisely to obtain λ for any given η in (1.4) for nuclear-norm type minimization problems. We first show that the norm of the residual $\|X(\lambda) - Y\|_F^2$ is a strictly increasing function with respect to λ. Hence we derive a closed form formula for λ associated with the constraint (1.4). This result can be applied to select an optimal regularization parameter λ for the unconstrained minimization problem (1.3). Compared with the SURE methods or the GCV methods, the proposed method can directly compute λ without solving a complicated and intricate optimization problem. We will see that the cost of solving the unconstrained minimization problem (1.3) (and hence the constrained minimization problem (1.1) too) is reduced by 11 times when compared with SURE methods. We remark that our approach also provides a fast method for solving the constrained minimization problem (1.1) for any given η by solving the corresponding unconstrained problem (1.3) with the corresponding λ.

Note that the low-rank matrix recovery problem usually involves finding the SVD of matrices, and its computational complexity is $O(mn\min\{m, n\})$. For large-scale problems, the cost of the traditional SVD is very expensive. In the past ten years, randomized algorithms have been more and more widely used in low-rank matrix approximations [16, 36, 43, 51]. Compared with classic algorithms, randomized algorithms involve fewer floating-point operations (flops) and are more effective for large-scale problems. Our strategy for choosing the regularization value can also be easily incorporated into the randomized algorithms and the cost can be reduced by 18 times when compared with SURE methods.

The outline of this paper is organized as follows. In Section 2, we consider the solution of the regularization model (1.3) when the penalty function is the nuclear norm or its variants. In Section 3, we discuss the relationship between the regularized model and the constrained model. The estimation of the bound η for the residual is also given. In Section 4, we solve (1.4) for the regularization parameter λ in terms of the bound η. Then we consider the solution of the constrained model (1.1) for various nuclear-norm-type problems. In Section 5, numerical results are given to

demonstrate the effectiveness of the proposed method. Finally, a short conclusion is given in Section 6.

2. Regularized Minimization Model. In this section, we introduce some basic notation and consider the solution for the regularization minimization model in (1.3).

2.1. Notation. We first introduce the singular value decomposition (SVD). Any $m \times n$ ($m \leq n$) matrix Y can be factorized as follows $Y = U\Sigma V^T$ where $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$ and $V = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m] \in \mathbb{R}^{n \times m}$ are orthogonal, and $\Sigma \in \mathbb{R}^{m \times m}$ is a diagonal matrix. The column vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are respectively called the i-th left and right singular vectors, and the diagonal entry $\Sigma_{i,i} = \sigma_i$ is called the i-th singular value of the matrix Y. We assume $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m \geq 0$.

The SVD can also be reformulated by the outer product form $Y = \sum_{i=1}^{m} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. A k-truncated matrix of Y is defined as $Y_k \equiv \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. By applying Eckart-Young-Mirsky theorem [14, 40], we know that

$$Y_k = \underset{\text{rank}(X) \leq k}{\text{argmin}} \|X - Y\|_F^2.$$

We remark that the k-truncated matrix of Y is not unique if $\sigma_{k+1} = \sigma_k$.

The main aim in (1.1) is to recover an unknown matrix X from the noisy observation Y. One common assumption is that the matrix X has low-rank structure. It is natural to choose the objective function $f(X) = \text{rank}(X)$ and consider the following problem

$$\min_X \text{rank}(X) + \frac{1}{2\lambda}\|X - Y\|_F^2. \tag{2.1}$$

The rank of a matrix can be defined as the nonzero numbers of its singular values, i.e., the ℓ_0-norm of the singular values. The solution to (2.1) is given by the hard thresholding operator $X = \sum_{i=1}^{m} \text{HT}_\lambda(\sigma_{Y,i})\mathbf{u}_i\mathbf{v}_i^T$, where $\text{HT}_\lambda(x)$ is a map defined by

$$\text{HT}_\lambda(x) = \begin{cases} x, & x > \sqrt{2\lambda}, \\ 0 \text{ or } x, & x = \sqrt{2\lambda}, \\ 0, & x < \sqrt{2\lambda}. \end{cases}$$

2.2. Nuclear Norm. In compressed sensing, in order to characterize the sparsity of the signal, the ℓ_1-norm is usually used to approximate the ℓ_0-norm. Similarly, in the rank minimization problem (2.1), the rank of a matrix can be approximately replaced by the ℓ_1-norm of the singular values, which is the nuclear norm of the matrix.

DEFINITION 2.1. *Given a matrix $X \in \mathbb{R}^{m \times n}$ with $\sigma_{X,i}$ being its i-th singular values, the nuclear norm (NN) $\|X\|_*$ of X is defined as the sum of its singular values, i.e., $\|X\|_* \equiv \sum_{i=1}^{m} \sigma_{X,i}$.*

The nuclear norm regularized minimization problem is given by

$$\min_X \|X\|_* + \frac{1}{2\lambda}\|X - Y\|_F^2. \tag{2.2}$$

The following theorem states that the optimal minimizer of (2.2) is a function with respect to the regularization parameter λ.

THEOREM 2.2. *[6] Let Y be an $m \times n$ matrix, $\mathbf{u}_i$ and $\mathbf{v}_i$ be the i-th left and right singular vectors of Y, and $\sigma_{Y,i}$ be the i-th singular value. Then the solution of (2.2) is*

given by $\widehat{X}(\lambda) = \mathrm{SVT}_\lambda(Y)$, *where* $\mathrm{SVT}_\lambda(Y)$ *is the singular value thresholding (SVT) operator defined as*

$$\widehat{X}(\lambda) = \mathrm{SVT}_\lambda(Y) \equiv \sum_{i=1}^m (\sigma_{Y,i} - \lambda)_+ \mathbf{u}_i \mathbf{v}_i^T. \tag{2.3}$$

Here $(x)_+ \equiv \max(x, 0)$.

2.3. Truncated Nuclear Norm. The nuclear norm shrinks all singular values equally and the large singular values are usually over-penalized. To overcome the shortcomings of nuclear norm, Hu *et al.* [27] used the truncated nuclear norm instead of the standard nuclear norm in order to keep the largest r singular values unchanged while shrinking the other singular values.

DEFINITION 2.3. *Given a matrix* $X \in \mathbb{R}^{m \times n}$ *with* $\sigma_{X,i}$ *being its i-th singular values, the truncated nuclear norm (TNN)* $\|X\|_r$ *of* X *is defined as the partial sum of its singular values, i.e.,* $\|X\|_r \equiv \sum_{i=r+1}^m \sigma_{X,i}$, *where* $r < m$ *is an integer.*

The TNN regularized minimization problem is given by

$$\min_X \|X\|_r + \frac{1}{2\lambda} \|Y - X\|_F^2. \tag{2.4}$$

The TNN has been proposed in low-rank matrix recovery and matrix completion problem [27, 31, 33]. The TNN is non-convex and the solution of TNN regularized minimization problem is generally non-trivial. In [27], the Von Neumann's trace inequality [41] was utilized to handle the non-convexity. The idea is to rewrite TNN into an equivalent form

$$\|X\|_r = \|X\|_* - \min_{(U,V) \in \mathcal{A}} \mathrm{trace}(UXV^T)$$

where $\mathcal{A} = \{(U, V) : UU^T = I, VV^T = I, U \in \mathbb{R}^{r \times m}, V \in \mathbb{R}^{r \times n}\}$.

According to Von Neumann's lemma, the inner product of two matrices is always bounded by the sum of the products of their corresponding singular values, while the maximum of the inner product of two matrices can only be achieved when these two matrices have the same left and right singular vectors. Based on these facts, the optimal solution of (2.4) can be expressed by the partial singular value thresholding operator [44].

THEOREM 2.4. *For any* $\lambda > 0$, *a global solution of (2.4) with a target rank* $r < m$ *is given by the partial singular value thresholding (PSVT)*

$$\mathrm{PSVT}_{r,\lambda}(Y) \equiv \sum_{i=1}^r \sigma_{Y,i} \mathbf{u}_i \mathbf{v}_i^T + \sum_{i=r+1}^m (\sigma_{Y,i} - \lambda)_+ \mathbf{u}_i \mathbf{v}_i^T. \tag{2.5}$$

We note that the PSVT operator is comprised of two terms: the first term is the r-truncated matrix $Y_r \equiv \sum_{i=1}^r \sigma_{Y,i} \mathbf{u}_i \mathbf{v}_i^T$, and the second term is the singular value thresholding operator of $Z = Y - Y_r$, i.e., $\mathrm{SVT}_\lambda(Z) = \sum_{i=r+1}^m (\sigma_{Y,i} - \lambda)_+ \mathbf{u}_i \mathbf{v}_i^T$. Hence we have

$$\mathrm{PSVT}_{r,\lambda}(Y) = Y_r + \mathrm{SVT}_\lambda(Y - Y_r),$$

which implies that the truncated nuclear norm minimization problem can be solved through the nuclear norm minimization problem. For the constrained minimization

problem, we have

$$\underset{\|Y-X\|_F^2 \leq \eta}{\operatorname{argmin}} \|X\|_r \equiv Y_r + \underset{\|Z-X\|_F^2 \leq \eta}{\operatorname{argmin}} \|X\|_*, \tag{2.6}$$

where $Z = Y - Y_r$. Thus the constrained minimization problem for the truncated nuclear norm can be reformulated as that of the nuclear norm.

2.4. Weighted Nuclear Norm. To improve the flexibility of the nuclear norm, Gu *et al.* [22] proposed a weighted nuclear norm to approximate the rank function, which has good performance in image denoising problem.

DEFINITION 2.5. *Given a matrix $X \in \mathbb{R}^{m \times n}$ with $\sigma_{X,i}$ being its i-th singular values and the non-negative weight vector $\mathbf{w} = (w_1, w_2, \ldots, w_m)$, the weighted nuclear norm (WNN) $\|X\|_{\mathbf{w},*}$ of X is defined by $\|X\|_{\mathbf{w},*} \equiv \sum_{i=1}^m w_i \sigma_{X,i}$.*

We remark that the standard nuclear norm and the truncated nuclear norm can be regarded as the special version of the weighted nuclear norm, and the weight vectors are $\mathbf{w} = (1, \ldots, 1)$ for the standard nuclear norm and $\mathbf{w} = (0, \ldots, 0, 1, \ldots, 1)$ for the truncated nuclear norm.

We emphasize that the choice of the weights is very important. A reasonable weight vector should guarantee that the resulting singular values are in a non-increasing order. When the weights are arranged in non-decreasing order, larger singular values should be less penalized. In [22], the weights are suggested to be

$$w_i = (\sigma_{X,i} + \epsilon)^{-1}, \quad i = 1, ..., m,$$

where the parameter ϵ is a sufficiently small positive number in order to avoid dividing by zero and should be set slightly smaller than the expected nonzero singular value of X. In this case, $\|X\|_{\mathbf{w},*}$ approximates the number of the nonzero singular values of X, i.e., the rank of X. In [28], Huang *et al.* generalized the weights to

$$w_i = (\sigma_{X,i} + \epsilon)^{p-1}, \quad 0 \leq p < 1. \tag{2.7}$$

In this way, $\|X\|_{\mathbf{w},*}$ approximates the Schatten p-norm of X, which is defined by $(\sum_{i=1}^m \sigma_{X,i}^p)^{1/p}$. Because the singular values $\sigma_{X,i}$ are unknown, an iterative approach is applied to estimate them. It is proved in [28] that the optimal solution of the reweighted nuclear norm minimization problem is also a solution of the Schatten p-norm minimization problem.

In this paper, the weight vector defined in [28] (i.e., (2.7)) is adopted, and we call the corresponding norm the generalized weighted nuclear norm (GWNN). Although the weighted nuclear norm is non-convex, its global optimal solution can still be obtained. Similar to Theorem 2.2, the closed form solution for the reweighted problem (2.8) can be represented by weighted singular value thresholding.

THEOREM 2.6. *[23] For any $\lambda > 0$, $Y \in \mathbb{R}^{m \times n}$, if the weights $\{w_i\}_{i=1}^m$ satisfy $0 \leq w_1 \leq w_2 \leq ... \leq w_m$, then a global solution of the following minimization problem*

$$\min_X \|X\|_{\mathbf{w},*} + \frac{1}{2\lambda} \|Y - X\|_F^2 \tag{2.8}$$

is given by the weighted singular value thresholding (WSVT) operator

$$\widehat{X}_{\mathbf{w}}(\lambda) = \mathrm{WSVT}_\lambda(Y) \equiv \sum_{i=1}^m (\sigma_{Y,i} - \lambda w_i)_+ \mathbf{u}_i \mathbf{v}_i^T. \tag{2.9}$$

If there are zero entries in the weight vector $\mathbf{w}$, i.e., $\mathbf{w} = (0, ..., 0, w_{r+1}, ..., w_m)$ with $0 < w_{r+1} \leq w_{r+2} \leq ... \leq w_m$, we have

$$\text{WSVT}_\lambda(Y) = Y_r + \sum_{i=r+1}^{m} (\sigma_{Y,i} - \lambda w_i)_+ \mathbf{u}_i \mathbf{v}_i^T. \tag{2.10}$$

This is because the WSVT operator is comprised of the r-truncated matrix and the WSVT operator with nonzero weights. Therefore, without loss of generalization, in the following when we talk about weighted nuclear norm, we only focus on the case where all the weights are positive.

3. Constrained model and regularized model. Because the minimization problem of the truncated nuclear norm or weighted nuclear norm with zero weights can be reformulated as that of the nuclear norm or the weighted nuclear norm with positive weights respectively (see (2.6) or (2.10)), in this section, we assume that the objective function $f(X)$ is either the nuclear norm or weighted nuclear norm with positive weights only.

We discuss the relationship between the constrained model (1.1) and the regularized model (1.3). Mathematically, the two minimization models are equivalent in the sense that given a parameter λ, there exists a corresponding η such that the solution of (1.3) is also a solution of (1.1), and vice versa. The following lemmas give the detailed explanations.

LEMMA 3.1. *If $X^\dagger$ is a global minimizer of the unconstrained regularized problem (1.3) with $\lambda > 0$, then there exists an $\eta \geq 0$ such that $X^\dagger$ is also a global minimizer of (1.1).*

Proof. Define $\eta \equiv \|Y - X^\dagger\|_F^2 \geq 0$. For any X satisfying $\|Y - X\|_F^2 \leq \eta$, applying the fact that $X^\dagger$ is a global minimizer of (1.3), we have

$$\frac{\eta}{2\lambda} + f(X^\dagger) = \frac{1}{2\lambda}\|Y - X^\dagger\|_F^2 + f(X^\dagger) \leq \frac{1}{2\lambda}\|Y - X\|_F^2 + f(X) \leq \frac{\eta}{2\lambda} + f(X).$$

Hence $f(X^\dagger) \leq f(X)$, which implies that $X^\dagger$ is also a global minimizer of (1.1). □

In order to exclude a trivial solution $X = 0$ for the constrained minimization problem in (1.1), we have assumed that $\|Y\|_F^2 > \eta$, see (1.2). We show that a minimizer of the constrained minimization problem (1.1) always lies on the boundary of the constraint if $\|Y\|_F^2 > \eta$.

LEMMA 3.2. *If $X^\dagger$ is a global minimizer of (1.1) with $\|Y\|_F^2 > \eta > 0$, then we have $\left\|Y - X^\dagger\right\|_F^2 = \eta$.*

Proof. Let $\sigma_{X^\dagger,i}$ and $\sigma_{Y,i}$ be the singular values of $X^\dagger$ and Y respectively. If, by contradiction, $X^\dagger$ is an interior point of the constrained set, then we have

$$\sum_{i=1}^{m}(\sigma_{Y,i} - \sigma_{X^\dagger,i})^2 \leq \|Y - X^\dagger\|_F^2 < \eta,$$

where the first inequality is obtained by applying Von Neumann's trace inequality [41]. Let $\sigma_{X^\dagger,k}$ be the smallest positive singular value of X. The above inequality implies that there exists a positive scalar δ with $0 < \delta < \sigma_{X^\dagger,k}$ such that

$$\sum_{i=1}^{k-1}\left(\sigma_{Y,i} - \sigma_{X^\dagger,i}\right)^2 + \left(\sigma_{Y,k} - \sigma_{X^\dagger,k} + \delta\right)^2 + \sum_{i=k+1}^{m}\left(\sigma_{Y,i}\right)^2 \leq \eta. \tag{3.1}$$

Let $\Sigma_{\widehat{X}}$ be the diagonal matrix with the diagonal entries $\sigma_{\widehat{X},k} = \sigma_{X^\dagger,k} - \delta$ and $\sigma_{\widehat{X},i} = \sigma_{X^\dagger,i}$ for $i \neq k$. Clearly, the diagonal entries of $\Sigma_{\widehat{X}}$ still keep the non-increasing ordering.

Let the SVD of Y be $Y = U\Sigma_Y V^T$ and we set $\widehat{X} = U\Sigma_{\widehat{X}} V^T$. By (3.1), $\|Y - \widehat{X}\|_F^2 < \eta$, i.e., $\widehat{X}$ is a feasible solution of (1.1). However, we have

$$\sum_{i=1}^{m} w_i \sigma_{\widehat{X},i} = f(\widehat{X}) < f(X^\dagger) = \sum_{i=1}^{m} w_i \sigma_{X^\dagger,i},$$

(for nuclear norm, all $w_i = 1$). This contradicts the assumption that $X^\dagger$ is a global minimizer of (1.1). Therefore, we must have $\|Y - X^\dagger\|_F^2 = \eta$. $\square$

LEMMA 3.3. *Let $\|Y\|_F^2 > \eta > 0$ and the objective function $f(X)$ in (1.1) be the nuclear norm or the weighted nuclear norm with weights $0 < w_1 \leq w_2 \leq ... \leq w_m$. If $X^\dagger$ is a global minimizer of (1.1), then there exists a parameter $\lambda > 0$ such that $X^\dagger$ is also a global minimizer of (1.3).*

Proof. First, we show that $X^\dagger$ and Y have the same sets of left and right singular vectors. For if otherwise, then by Von Neumann's trace inequality [41], we have

$$\sum_{i=1}^{m} (\sigma_{Y,i} - \sigma_{X^\dagger,i})^2 < \|Y - X^\dagger\|_F^2 = \eta,$$

where the last equality is by Lemma 3.2. Similar to the proof in Lemma 3.2, we can then construct a matrix $\widehat{X}$ such that $f(\widehat{X}) < f(X^\dagger)$ while $\|Y - \widehat{X}\|_F^2 \leq \eta$. This contradicts the assumption that $X^\dagger$ is a global minimizer of (1.1).

Next, we show that the singular values of $X^\dagger$ can be obtained by the Lagrangian multiplier method. Let $Y = U\Sigma_Y V^T$ be the SVD of Y. By the above paragraph, the solution of (1.1) is given by $X^\dagger = U\Sigma_{X^\dagger} V^T$. Here $\Sigma_{X^\dagger}$ can be obtained by solving the minimization problem

$$\min_{\|\Sigma_X - \Sigma_Y\|_F^2 \leq \eta} f(\Sigma_X). \tag{3.2}$$

We note that the constraint $\|\Sigma_X - \Sigma_Y\|_F^2 \leq \eta$ must be an active constraint. For if not, then $X = 0$ is a minimizer and we obtain $\|\Sigma_Y\|_F^2 = \|Y\|_F^2 \leq \eta$, a contradiction.

Without loss of generality, we consider $f(\Sigma_X) = \sum_i w_i \sigma_{X,i}$ (for nuclear norm, all $w_i = 1$). Then (3.2) can be reformulated into

$$\min_{\sigma_X \in \mathcal{S}} \sum_i w_i \sigma_{X,i}, \tag{3.3}$$

where

$$\mathcal{S} = \{\sigma_X = (\sigma_{X,1}, \ldots, \sigma_{X,m})^T : \|\sigma_X - \sigma_Y\|_2^2 \leq \eta, \sigma_{X,i} \geq \sigma_{X,2} \geq \cdots \geq \sigma_{X,m} \geq 0\}.$$

Since $\sigma_Y \in \mathcal{S}$ is a strictly feasible point, the minimization problem satisfies Slater's condition [2, Proposition 3.3.9]. Therefore, there must exist a Lagrangian multiplier $\alpha > 0$ such that the solution of (3.3) (which is the same as the solution of (1.1)) can be obtained by minimizing the Lagrangian function

$$\min_{\sigma_X \geq 0} \sum_{i=1}^{m} w_i \sigma_{X,i} + \frac{\alpha}{2} \left(\|\sigma_X - \sigma_Y\|_2^2 - \eta \right).$$

The objective function is separable in each i, hence we obtain

$$\sigma_{X^\dagger,i} = (\sigma_{Y,i} - w_i/\alpha)_+ \quad \forall i = 1, \ldots, m. \tag{3.4}$$

It is easy to check that we have $\sigma_{X^\dagger,i} \geq \sigma_{X^\dagger,i+1}$ due to the ordering of the singular values $\sigma_{Y,i}$ and the weights ω_i.

Finally, we show that the lemma holds. It is obvious that $\sigma_{X^\dagger,i}$ given by (3.4) are precisely those singular values in the SVT operator (see (2.3)) or a WSVT operator (see (2.9)) with the thresholding parameter $\lambda = 1/\alpha$. Therefore, by Theorem 2.2 or Theorem 2.8, we have

$$f(X^\dagger) + \frac{\alpha}{2}\|X^\dagger - Y\|_F^2 \leq f(X) + \frac{\alpha}{2}\|X - Y\|_F^2,$$

for all $X \in \mathbb{R}^{m \times n}$. $\square$

Based on the above lemma, we can find a suitable regularization parameter λ when the bound η is given such that a solution of (1.3) is one of the solution of (1.1).

3.1. Choice of Upper bound. It is very important to choose a suitable regularization parameter λ in the regularized minimization problem (1.3) because the quality of the recovered matrix highly depends on this parameter. If λ is too large, the given data cannot be fitted correctly, and if it is too small, the shrinkage is insufficient. The Morozov discrepancy principle [15, 42] is one method to choose λ. This principle selects λ such that the residual norm is bounded, i.e., a good regularized solution $X(\lambda)$ should lie in a set $\{X : \|Y - X(\lambda)\|_F^2 \leq \eta\}$, where η is an upper bound of the discrepancy depending on the noise level. Now we consider how to choose the bound η.

According to the discussion in Section 2, the minimizer of the regularized problem (1.3) has a closed form solution and is a function of the parameter λ. We will use $\widehat{X}(\lambda)$ to denote the minimizer. Notice that the residual $Y - \widehat{X}(\lambda)$ can be rewritten as $Y - \widehat{X}(\lambda) = X_\dagger + W - \widehat{X}(\lambda)$, where $X_\dagger$ is the true solution and each entry of the error matrix W is the Gaussian noise with zero mean and variance τ^2. If $\widehat{X}(\lambda)$ is a good estimation of $X_\dagger$, then the residual should be dominated by the Gaussian noise W. We introduce the following theorems which can be found in [24, 38, 39].

THEOREM 3.4. *Assume that all the entries of W are of normal independent distribution with mean 0 and variance τ^2. Then $\|Y - X_\dagger\|_F^2$ is χ^2-distributed with variance τ^2 and mn degree of freedom, i.e., $\frac{1}{\tau^2}\|W\|_F^2 \sim \chi_{mn}^2$. Moreover, we have*

$$\mathbb{E}\|W\|_F^2 = mn\tau^2.$$

THEOREM 3.5. *Suppose that g is a Lipschitz function on matrices:*

$$|g(X) - g(Y)| \leq L\|X - Y\|_F, \quad \text{for all } X, Y \in \mathbb{R}^{m \times n}.$$

Given a Gaussian matrix $W \sim \mathcal{N}(0, \tau^2 I)$, then

$$\mathbb{P}\{g(W) \geq \mathbb{E}g(W) + L\tau t\} \leq e^{-t^2/2}.$$

Since $\|\cdot\|_F$ is Lipschitz continuous with $L = 1$, we obtain

$$\mathbb{P}\{\|W\|_F \leq \mathbb{E}\|W\|_F + t\tau\} \geq 1 - e^{-t^2/2}.$$

We can change t to get an upper bound of the Gaussian noise $\|W\|_F$ with high probability (e.g. when $t \geq 4$, the noise is bounded with probability greater than 0.9996).

According to Theorems 3.4 and 3.5, it is natural to set the upper bound of $\|W\|_F^2$ to $\eta = mn\tau^2$. However, $\widehat{X}(\lambda)$ is obtained by applying the SVT operator to Y, hence $\widehat{X}(\lambda)$ is dependent on the noise W. It is reasonable to modify the upper bound to

$$\eta = cmn\tau^2, \tag{3.5}$$

where $c \simeq 1$ can be adjusted appropriately to suit the applications [4, 17]. We see that the parameter η depends on the estimation of the noise level τ. If the noise level is not given, τ can be estimated by using the median rule [13, 35], i.e.,

$$\tau = \text{median}(|\widehat{Y}_{HH}|)/0.6745, \tag{3.6}$$

where $\widehat{Y}_{HH}$ is the high-high coefficients of Y at the finest wavelet transform level.

Once the upper bound η is determined, we are in the position to find λ in terms of η. Because the regularization parameter λ can be regarded as the Lagrange multiplier associated with the active constraint, the complementarity condition states

$$\|Y - \widehat{X}(\lambda)\|_F^2 = \eta. \tag{3.7}$$

It means that the optimal regularization parameter λ is the root of the nonlinear equation (3.7).

4. Solutions for the Constrained and Unconstrained problems. In this section, we first consider finding an analytical expression of λ for the unconstrained problem (1.3) when the bound η for the constrained problem (1.1) is given (e.g. as in (3.5)). Since the regularization parameter λ is the same as the thresholding parameter used in the singular value soft-thresholding operator, we can hence solve the unconstrained problem (1.3) easily. By the equivalence of the two minimization problems, the constrained problem with the bound η is also then solved.

4.1. Nuclear Norm. Let $f(X) = \|X\|_*$, we know that the minimizer of the nuclear norm regularized minimization problem (2.2) is given by the SVT operator, see Theorem 2.2. We first show that $\|Y - \widehat{X}(\lambda)\|_F^2$ is a positive and strictly increasing function of λ.

THEOREM 4.1. *Let $\sigma_{Y,i}, i = 1, ..., m$ be the singular values of $Y \in \mathbb{R}^{m \times n}$ with $\sigma_{Y,1} \geq \sigma_{Y,2} \geq \ldots \geq \sigma_{Y,m} \geq 0$ and $\widehat{X}(\lambda)$ be the solution of (2.2) obtained by the SVT operator acting on Y. Then $\|Y - \widehat{X}(\lambda)\|_F^2$ is a positive and strictly increasing function of λ in the interval $[0, \sigma_{Y,1}]$.*

Proof. According to Theorem 2.2, we have

$$Y - \widehat{X}(\lambda) = \sum_{i=1}^{m} \sigma_{Y,i} \mathbf{u}_i \mathbf{v}_i^T - \sum_{i=1}^{m} \max(\sigma_{Y,i} - \lambda, 0) \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^{m} \min(\sigma_{Y,i}, \lambda) \mathbf{u}_i \mathbf{v}_i^T.$$

Hence we have

$$\phi(\lambda) \equiv \|Y - \widehat{X}(\lambda)\|_F^2 = \sum_{i=1}^{m} \min(\sigma_{Y,i}^2, \lambda^2). \tag{4.1}$$

10

Without loss of generality, we assume $\sigma_{Y,1} > 0$ and define $\sigma_{Y,m+1} = 0$. For any λ satisfying $0 = \sigma_{Y,m+1} < \lambda \leq \sigma_{Y,1}$, there must exist an index k such that $\lambda \in (\sigma_{Y,k+1}, \sigma_{Y,k}]$. Then we have

$$\phi(\lambda) = k\lambda^2 + \sum_{i=k+1}^{m} \sigma_{Y,i}^2, \tag{4.2}$$

where the first term is a quadratic function with respect to λ and the second term is independent of λ. Thus $\phi(\lambda)$ is a strictly increasing function in $(\sigma_{Y,k+1}, \sigma_{Y,k}]$. One can easily check that

$$\lim_{\lambda \to \sigma_{Y,k+1}^-} \phi(\lambda) = \phi(\sigma_{Y,k+1}) = k\sigma_{Y,k+1}^2 + \sum_{i=k+1}^{m} \sigma_{Y,i}^2,$$

which implies that $\phi(\lambda)$ is a continuous function. Thus we can deduce that $\phi(\lambda)$ is a positive and strictly increasing function in the interval $[0, \sigma_{Y,1}]$. □

Based on Theorem 4.1, We know $\phi(\lambda) = 0$ for $\lambda = 0$ and $\phi(\lambda) = \sum_{i=1}^{m} \sigma_{Y,i}^2 = \|Y\|_F^2$ for $\lambda \geq \sigma_{Y,1}$. Thus we obtain $0 \leq \phi(\lambda) \leq \|Y\|_F^2$ and we have the following theorems.

THEOREM 4.2. *Let $\widehat{X}(\lambda)$ be the solution of* (2.2) *obtained by the SVT operator acting on Y and let $\eta > 0$ be such that $\|Y\|_F^2 > \eta$. Then the nonlinear equation $\|Y - \widehat{X}(\lambda)\|_F^2 = \eta$ has a unique solution $\lambda > 0$.*

Proof. We define a new sequence

$$b_j = j\sigma_{Y,j}^2 + \sum_{i=j+1}^{m} \sigma_{Y,i}^2. \tag{4.3}$$

It is obvious that the sequence b_j is non-increasing, that is $b_1 \geq b_2 \geq \ldots \geq b_m$. Define $b_{m+1} = 0$. Since $b_1 > \eta > 0$, we choose the subscript k such that $b_{k+1} < \eta \leq b_k$. Let

$$\lambda = \sqrt{\left(\eta - \sum_{i=k+1}^{m} \sigma_{Y,i}^2\right)/k}. \tag{4.4}$$

Since $\eta > b_{k+1}$, we have $\lambda > 0$. By (4.2), we have $\phi(\lambda) = \eta$. □

4.2. Generalized Weighted Nuclear Norm. Now we consider the monotonicity of the discrepancy $\|Y - \widehat{X}_{\mathbf{w}}(\lambda)\|_F^2$ with respective to the parameter λ. For any $0 < \lambda \leq \sigma_{Y,1}$, there exists an index k such that $\lambda \in (\sigma_{Y,k+1}, \sigma_{Y,k}]$. Suppose $\widehat{X}_{\mathbf{w}}(\lambda)$ is the solution of (2.8) with non-descending weights $\{w_i\}, i = 1, ..., m$, then using (2.9) and after some manipulations similar to those in Theorem 4.1, we have

$$\psi(\lambda) \equiv \|Y - \widehat{X}_{\mathbf{w}}(\lambda)\|_F^2 = \left(\sum_{i=1}^{k} w_i^2\right)\lambda^2 + \sum_{i=k+1}^{m} \sigma_{Y,i}^2. \tag{4.5}$$

Comparing to the function $\phi(\lambda)$ in (4.2), the coefficient k of λ^2 is replaced by the sum of the weighted square $\sum_{i=1}^{k} w_i^2$ for the function $\psi(\lambda)$. Obviously, we can obtain the following result similar to Theorem 4.1.

THEOREM 4.3. *Let $\widehat{X}_{\mathbf{w}}(\lambda)$ be the solution of* (2.8) *obtained by applying the WSVT operator* (2.9) *acting on Y. Then the function $\psi(\lambda)$ is a positive and strictly increasing function of λ in the interval $[0, \sigma_{Y,1}]$.*

11

The result for Theorem 4.2 can be easily extend to the case for the weighted nuclear norm minimization problem. Notice that the WSVT operator in (2.9) can be reformulated into

$$\text{WSVT}_\lambda(Y) = \sum_{i=1}^{m} w_i \left(\frac{\sigma_{Y,i}}{w_i} - \lambda \right)_+ \mathbf{u}_i \mathbf{v}_i^T.$$

Hence the weighted singular values $\sigma_{Y,i}/w_i$ which are less than the threshold parameter λ are set to zero. Similar to (4.3), we define the sequence

$$b_j = (\frac{\sigma_{Y,j}}{w_j})^2 \sum_{i=1}^{j} w_i^2 + \sum_{i=j+1}^{m} \sigma_{Y,j}^2, \quad j = 1, ..., m, \tag{4.6}$$

with $b_{m+1} = 0$. Then $\{b_j\}_{j=1}^{m+1}$ is a non-increasing sequence. Since by Theorem 4.3, the function $\psi(\lambda)$ defined by (4.5) is a positive and strictly increasing function of the parameter λ, we see that there exists an index k such that $b_{k+1} < \eta \leq b_k$. Hence we have the following theorem.

THEOREM 4.4. *Let* $\widehat{X}_\mathbf{w}(\lambda)$ *be the solution of* (2.8) *obtained by applying the WSVT operator* (2.9) *on* Y. *Then the nonlinear equation* $\psi(\lambda) \equiv \|Y - \widehat{X}_\mathbf{w}(\lambda)\|_F^2 = \eta$ *has a unique solution*

$$\lambda = \sqrt{(\eta - \sum_{i=k+1}^{m} \sigma_{Y,i}^2)/\sum_{i=1}^{k} w_i^2}. \tag{4.7}$$

for any η *satisfying* $0 < \eta < \|Y\|_F^2$.

Based on the above Theorems, there exists a parameter λ such that the minimizer of (1.1) is a SVT operator or WSVT operator with respect to λ. We can find a suitable regularization parameter λ when the bound η is given such that a solution of (1.1) is one of the solution of (1.3). Since the regularization parameter λ is the same as the thresholding parameter used in the singular value soft-thresholding operator, we can hence solve the unconstrained problem (1.3) easily. By the equivalence of the two minimization problems, the constrained problem with the bound η is also then solved. We summarize the algorithm for solving the nuclear norm minimization problem (1.1) in Algorithm 1.

Algorithm 1 Low-rank matrix recovery for constrained minimization problem
---
Input: Observation matrix Y and η
Output: Estimated matrix X
  1: **if** $\|Y\|_F^2 \leq \eta$ **then**
  2:     $X = \mathbf{0}$
  3: **else**
  4:     $Y = U\Sigma_Y V^T, \Sigma_Y = \text{diag}(\sigma_{Y,1}, ..., \sigma_{Y,m})$
  5:     Compute $b_j, j = 1, ..., m$ by (4.3) or (4.6)
  6:     Find k such that $b_k > \eta \geq b_{k+1}$
  7:     Compute λ by (4.4) or (4.7)
  8: **end if**
  9: **return** $X = \text{SVT}_\lambda(Y)$ or $X = \text{WSVT}_\lambda(Y)$

4.3. Randomized Algorithm for Large-scale Problems. When the singular value thresholding operator is applied to recover the low-rank matrix, we need to calculate all the singular values of the observed matrix $Y \in \mathbb{R}^{m \times n}$. The computation complexity is $O(m^2 n)$ (here $m \leq n$). For large-scale matrices, singular values calculation is very time-consuming, which makes it infeasible in practical applications such as those in data sciences and image processing. In order to avoid directly computing all the singular values, one can use the randomized SVD (RSVD) approach [24, 37, 43]. The key is to extract a small core matrix by finding an orthonormal matrix with the unitary invariant property.

Let $\text{rank}(X) = s$ and $s < \ell \leq m$. There exists an orthogonal matrix $Q_\ell \in \mathbb{R}^{n \times \ell}$ with $Q_\ell^T Q_\ell = I_\ell$ and a matrix $A \in \mathbb{R}^{m \times \ell}$ such that X has the factorization $X = A Q_\ell^T$. Hence we have

$$\|Y - X\|_F^2 = \|Y Q_\ell - A\|_F^2 + a. \tag{4.8}$$

Here $a = \|Y\|_F^2 - \|Y Q_\ell\|_F^2$. Obviously, $Y Q_\ell \in \mathbb{R}^{m \times \ell}$ is a matrix with the size less than Y. Notice that $f(A) = f(X)$ for $\text{rank}(X) \leq \ell$, where $f(X)$ denotes the nuclear norm, the truncated nuclear norm or the weighted nuclear norm of X. Therefore, instead of solving the regularized minimization problem (1.3), by changing the variable, we consider the following reduced minimization problem

$$\min_A f(A) + \frac{1}{2\lambda} \|Y Q_\ell - A\|_F^2. \tag{4.9}$$

The size of the matrix $Y Q_\ell$ is smaller than that of Y.

Here we consider the solution of the reduced problem and once we obtain the optimal solution for A, the matrix X can be recovered by $X = A Q_\ell^T$. Therefore the computation complexity can be significantly improved when $\ell \ll n$. Similarly, for the constrained minimization problem, the constraint can be reformulated as $\|Y Q_\ell - A\|_F^2 \leq \eta - a$ according to (4.8). Thus we consider the following equivalent reduced problem

$$\min_{\|Y Q_\ell - A\|_F^2 \leq \eta - a} f(A). \tag{4.10}$$

During the procedure to construct the reduced problem, we have assumed that Q_ℓ is known in advance and the rank of X is less than the column number of Q_ℓ. In some applications, however, the rank of X is unknown. We can estimate the rank of X according to the dominant components of the singular values of the noisy observation matrix Y. One way to obtain the dominant components of the singular values of Y is to construct an orthogonal matrix Q_ℓ such that the columns of Q_ℓ are the ℓ dominant right singular vectors of Y. The randomized method can be applied to efficiently compute Q_ℓ. An ideal Q_ℓ should satisfy the condition that the number of columns of the orthogonal matrix Q_ℓ is as small as possible such that the error $\|Y Q_\ell Q_\ell^T - Y\|$ is less than some desired tolerance. The details for the randomized method can be found in [24].

Clearly, once Q_ℓ and η are given, we can apply the same ideas developed in Subsection 4.1 and Subsection 4.2 to obtain the regularization parameter λ and hence solve the reduced problems (4.9) and (4.10).

Now we analyze the computational complexity of using the standard SVD in Algorithm 1. The SVD costs $O(m^2 n)$. Computing the sequences b_j and λ require $O(m^2)$ and $O(m)$, respectively. Then forming the optimal X needs $O(m^2 n)$. Therefore, the

total computational complexity of Algorithm 1 is $O(2m^2n+m^2+m)$. If we use the randomized SVD, the total computational complexity is reduced to $O(2\ell^2m+\ell^2+\ell)$. Since $\ell \ll m$, the computation speed can be significantly improved. We note that before calculating the projection problem (4.10), the orthogonal matrix Q_ℓ must be calculated. The cost of obtaining Q_ℓ in q iterations of the power method is $O(2qmn\ell+(m+n)\ell^2q)$, see [24].

5. Applications. The discrepancy principle is a classical method to choose the regularization parameter λ for the regularized model (1.3) provided that an upper bound of the constraint $\|X - Y\|_F^2 \leq \eta$ is given. In Section 3, we have developed the closed form formulas of λ for the NN case and the GWNN case. Once the parameter λ is determined, the regularized solution of (1.3) can be reconstructed by the soft-thresholding formulas (2.3) and (2.9) directly. These solutions are also the solutions for the unconstrained problem (1.1) for the given η. The ideas can easily be extended to cover the TNN case (which is a shifted NN case, see Subsection 2.3) and the randomized case, see Subsection 4.3.

We illustrate the efficiency of solving the problems (1.1) and (1.3) using our approach in the following experiments. We denote the results obtained by the nuclear norm minimization problem (2.2) using our proposed regularization parameter λ as "NN-DP" (see (4.4)). When the truncated nuclear norm or the generalized weighted nuclear norm is used, we denote our method as "TNN-DP" or "GWNN-DP" respectively. In the following tests, we set $r = 1$ for TNN, and the weights $w_i = (\sigma_{Y,i}+\epsilon)^{p-1}$ with $p = 0.7$ and $\epsilon = 10^{-6}$ for GWNN.

We compare the proposed algorithms with the following two approaches:

- **SURE [9]:** The optimal soft thresholding parameter can be estimated by minimizing

$$\text{SURE}\left(\text{SVT}_\lambda(Y)\right) = -mn\tau^2 + \sum_{i=1}^{m} \min\left(\lambda^2, \sigma_{Y,i}^2\right) + 2\tau^2 \,\text{div}\left(\text{SVT}_\lambda(Y)\right), \quad (5.1)$$

  where 'div' is the divergence of SVT_λ with respect to Y. The solution is then given by

$$\widehat{X}_{\text{SURE}} = \sum_{i=1}^{m}(\sigma_{Y,i} - \lambda^\dagger)_+ \mathbf{u}_i\mathbf{v}_i^T,$$

  where $\lambda^\dagger$ minimizes the SURE function. As suggested in [9], $\lambda^\dagger$ is obtained by trial-and-error on 101 logarithmically equally spaced points between 10^{-1} and 10^7. The SURE approach can be easily generalized to the TNN and GWNN too. We denote the resulting methods by "NN-SURE", "TNN-SURE" and "GWNN-SURE".
- **HardT [18]:** The optimal hard threshold is chosen as $\lambda = (4/\sqrt{3})\sqrt{n}\tau$ when the noise level τ is known or $0.2858\sigma_{Y,med}$ when τ is unknown. Here $\sigma_{Y,med}$ denotes the median value of the singular values of Y. The recovered matrix is given by

$$\widehat{X}_{\text{HardT}} = \sum_{i=1}^{m} \sigma_{Y,i}\mathbf{1}_{(\sigma_{Y,i}>\lambda)}\mathbf{u}_i\mathbf{v}_i^T.$$

14
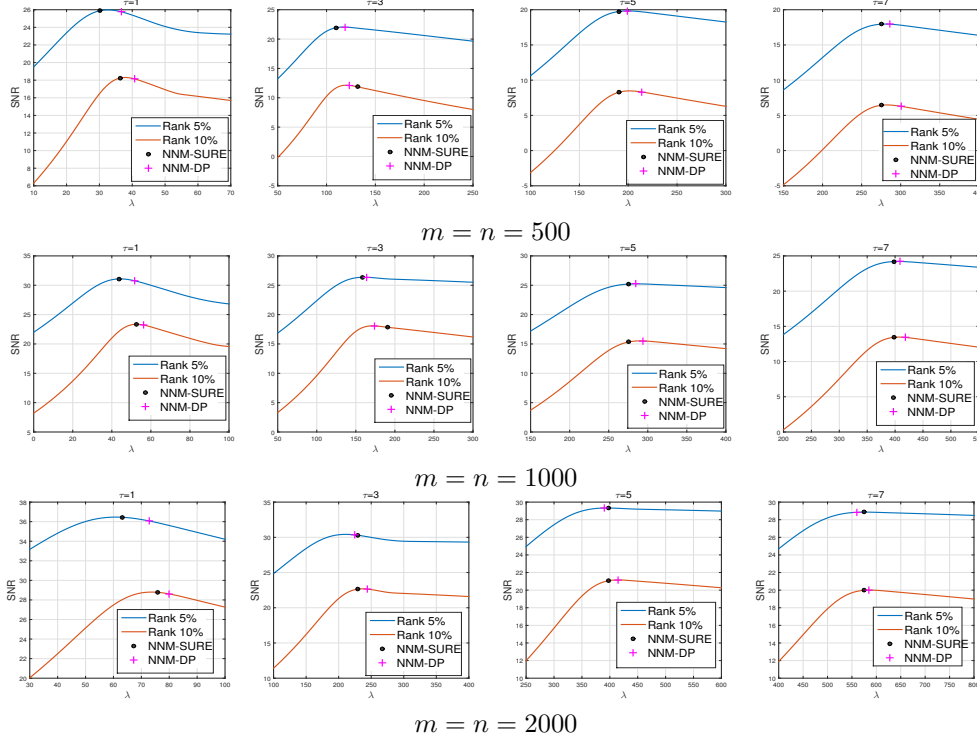
Figure 5.1. *The threshold parameter λ versus SNR for the different rank rations, different noise levels, and different matrix size. The parameters determined by "SURE" and "NNM-DP" are represented by black "●", and magenta "+", respectively.*

5.1. Simulations on Synthetic Data. In the first set of experiments, we consider removing noise from synthetical matrices with size $m \times n$. The rank of the true low-rank matrix X is determined by the rank ratio ρ and is generated by the product of two randomly sampling matrices $M \in \mathbb{R}^{m \times s}$ and $N \in \mathbb{R}^{n \times s}$ from the standard uniform distribution, i.e., $X_{\dagger} = MN^T$ where $s = \text{round}(\rho \max(m, n))$. The observed matrix Y is obtained by $Y = X_{\dagger} + W$ where $W \in \mathbb{R}^{m \times n}$ is a matrix of zero-mean Gaussian white noise with standard deviation τ. We recover the matrix $X_{\dagger}$ from its noisy observation matrix Y by solving the constrained minimization problem (1.1). The bound η is determined by (3.5) with $c = 1$. The Signal-to-Noise Ratio (SNR)[1] is used to measure the quality of the recovered matrix.

5.1.1. Regularization Parameter Selection. We first show that the selected regularization parameter by our method is a good one. We consider the standard nuclear norm as the penalty function. Rank ratio is set to $\rho = 5\%$ and 10% and the noise standard derivation level is set to $\tau = 1, 3, 5, 7$ respectively. The size of the matrix is set to $m = n = 500, 1000, 2000$ respectively.

The plots in Figure 5.1 show the curves of the SNR against the regularization parameter λ for different rank ratios and different noise levels. The SNR obtained by our approach (i.e., solving the minimization problem (1.3) using Algorithm 1) and by

---

[1]SNR is defined by $\text{SNR} = 10 \log \left(\frac{\|X\|_F^2}{\|X - \widehat{X}\|_F^2} \right)$, where $\widehat{X}$ is the estimated matrix.

the "SURE" are marked by black "•" and magenta "+" respectively. We observe in the figures that even if the noise level is the same, the optimal regularization parameter varies with the size of the matrix. We also observe that the SNRs obtained by "SURE" are close to the maximum of the curve in most cases because the parameter obtained by "SURE" has the minimum MSE within a given range of the parameter.

The regularization parameter obtained by our method tend to be slightly larger than the optimal one. This is consistent with the observations that the selection of discrepancy principle with $c = 1$ usually yields an over smoothed result [11, 17]. In order to obtain the optimal SNRs, one can choose a small value for c (i.e., $c < 1$) or decrease the regularization parameter.

We remark that when the size of the matrices is big and the rank ratio is small (see the last row in the figure), our method gives better λ. We also note that "SURE" obtains the regularization parameters by solving the minimization problem (5.1) by trial-and-error, which is time-consuming, while our method provides a closed form formula for λ directly (see (4.4) and (4.7)). We will see below that the CPU running time of our method is indeed much faster than that of "SURE".

5.1.2. Comparisons of Denoising Results. In this subsection, we compare the different methods in denoising. Here two tasks are performed. The first one is based on the full observation matrix, and the second one is based on the randomized projected matrix after dimensionality reduction. In the randomized case, the size ℓ of the projection matrix Q_ℓ (see Subsection 4.3) is specified in advance for the best rank-ℓ approximation. Here we take $\ell = s + 5$, where $s = \mathrm{round}(\rho \max(m, n))$. According to (4.8), the bound $\hat{\eta}$ for the matrix $\widehat{Y}_\ell = Y Q_\ell$ is given by

$$\hat{\eta} = \eta - \|Y\|_F^2 + \|\widehat{Y}_\ell\|_F^2.$$

Note that "HardT" and "SURE" need the noise level of the low-dimensional matrix $\widehat{Y}_\ell$ as an input argument. Its noise level can be estimated by $\hat{\tau} = (cmn\tau^2 - \|Y\|_F^2 + \|\widehat{Y}_\ell\|_F^2)/(n\ell)$ if the noise level τ of Y is given. Otherwise, one needs to estimate it either implicitly or explicitly from $\widehat{Y}_\ell$.

We set the matrix size $m = n = 500, 1000, 2000, 4000$, and the standard deviation $\tau = 3, 5$ respectively. The rank ratio ρ is set to $1\%, 5\%, 10\%, 20\%, 40\%$, and the rank of the clean matrix $X_\dagger$ is fixed to $\rho \max(m, n)$. We consider four different types of objective functions: $\mathrm{rank}(X)$, $\|X\|_*$, $\|X\|_r$ (with $r = 1$) and $\|X\|_{\mathbf{w},*}$. In fact, the penalty functions $\|X\|_*$, $\|X\|_r$ and $\|X\|_{\mathbf{w},*}$ are relaxations of the function $\mathrm{rank}(X)$. Here the SNRs and CPU running times are computed by running the experiments 10 times and taking the average of the 10 tests. We use boldface to mark the best method for a given matrix amongst all the different regularization norms and methods; and we use the italic font to mark the best method for the given regularization norm.

The SNRs obtained by different methods are shown in Table 5.1 for the standard non-randomized algorithms. From the table, we see that the SNRs obtained by the rank function are larger than those obtained by the nuclear norm and its variants when X is of small rank and small size. With the increase of the rank and the size, the SNRs obtained by the rank function become smaller than those obtained by other functions. When comparing the parameter strategy of "DP" (i.e. ours) and "SURE" in NN model, we observe that the SNRs are approximately equal, while the SNR obtained by our "DP" is slightly larger than that by "SURE" in many of the cases. We also observe that the SNRs obtained by TNN and GWNN are larger than those obtained by NN, which indicates their superiority in characterizing low-rank properties.

		$f(X)$	Rank(X)	$\|X\|_*$		$\|X\|_r$		$\|X\|_{\mathbf{w},*}$	
$m=n$	τ	ρ	HardT	NN-SURE	NN-DP	TNN-SURE	TNN-DP	GWNN-SURE	GWNN-DP
500	3	1%	14.47	11.78	*11.96*	**14.47**	**14.47**	*13.49*	13.05
		5%	**22.99**	21.84	*21.94*	**22.99**	**22.99**	*22.82*	*22.82*
		10%	26.25	25.69	25.72	26.25	**26.29**	26.22	26.21
		20%	29.43	*29.30*	29.30	**29.59**	**29.59**	29.51	*29.54*
		40%	32.48	32.72	*32.75*	32.84	**32.89**	*32.84*	*32.84*
	5	1%	**11.35**	*8.53*	8.51	**11.35**	**11.35**	*9.96*	9.31
		5%	**21.80**	19.71	*19.75*	**21.80**	21.79	*21.30*	21.26
		10%	**25.57**	24.16	*24.20*	**25.57**	25.53	*25.34*	25.32
		20%	**29.06**	28.26	28.27	**29.06**	29.03	28.97	28.94
		40%	32.29	*31.92*	31.91	**32.31**	**32.31**	32.27	32.26
Average SNR			24.57	23.39	*23.43*	**24.62**	**24.62**	*24.27*	24.16
Average time (sec)			0.054	1.124	0.059	1.178	0.060	1.144	0.058
1000	3	1%	19.22	17.69	*17.94*	19.22	**19.25**	18.90	*18.97*
		5%	26.46	*26.34*	26.34	26.64	**26.65**	26.61	26.62
		10%	29.51	29.71	29.72	29.87	**29.88**	29.84	29.85
		20%	32.54	*33.14*	33.10	**33.20**	33.18	*33.16*	33.15
		40%	35.39	*36.67*	36.59	**36.71**	36.63	*36.65*	36.59
	5	1%	**17.76**	15.34	*15.48*	**17.76**	**17.76**	*16.98*	16.90
		5%	**26.08**	25.22	*25.25*	**26.08**	26.05	*25.97*	25.95
		10%	**29.32**	28.89	28.88	**29.32**	29.31	29.30	29.26
		20%	32.44	32.29	32.29	**32.52**	32.51	*32.47*	32.46
		40%	35.51	35.57	*35.63*	35.68	**35.73**	35.67	*35.69*
Average SNR			28.42	28.09	*28.12*	**28.70**	**28.70**	28.56	28.54
Average time (sec)			0.240	5.646	0.261	5.992	0.267	5.895	0.249
2000	3	1%	22.64	22.62	*22.63*	23.02	**23.09**	23.04	*23.08*
		5%	29.54	30.27	*30.31*	30.37	*30.40*	**30.50**	30.44
		10%	32.48	*33.79*	33.67	33.83	33.71	**33.80**	33.74
		20%	35.64	*37.27*	37.11	37.28	37.13	**37.33**	37.14
		40%	39.12	*41.06*	40.74	**41.07**	40.74	*41.05*	40.74
	5	1%	**22.17**	21.07	*21.15*	**22.17**	22.16	21.99	*22.00*
		5%	29.46	*29.30*	29.29	**29.53**	**29.53**	29.46	*29.49*
		10%	32.52	32.54	*32.59*	32.67	**32.70**	*32.67*	*32.67*
		20%	35.55	35.86	*35.87*	35.91	**35.92**	35.85	*35.89*
		40%	38.55	*39.26*	39.25	39.29	**39.27**	*39.26*	39.23
Average SNR			31.77	*32.30*	32.26	**32.51**	32.47	*32.50*	32.44
Average time (sec)			1.572	28.956	1.571	31.651	1.583	30.140	1.562
4000	3	1%	26.06	*27.19*	27.12	*27.35*	27.27	**27.58**	27.48
		5%	33.37	*34.85*	34.66	*34.88*	34.68	**35.05**	34.85
		10%	36.70	*38.23*	38.03	*38.25*	38.04	**38.42**	38.16
		20%	40.21	*41.95*	41.54	*41.96*	41.55	**42.05**	41.62
		40%	43.99	**45.90**	45.28	*45.90*	45.29	**45.92**	45.32
	5	1%	25.57	*25.44*	*25.44*	**25.75**	**25.75**	25.72	*25.73*
		5%	32.55	32.87	*32.97*	32.92	*33.03*	33.00	**33.04**
		10%	35.56	*36.31*	36.25	**36.33**	36.28	*36.32*	36.28
		20%	38.44	*39.69*	39.63	**39.70**	39.64	*39.70*	39.64
		40%	41.61	*43.38*	43.17	**43.39**	43.18	*43.32*	43.16
Average SNR			35.41	*36.58*	36.41	*36.64*	36.47	**36.71**	36.53
Average time (sec)			14.232	159.368	*14.313*	161.864	*14.355*	160.932	14.334

In terms of CPU running time, our "DP" method is almost as fast as the "HardT" method as we only need one substitution to get λ (see (4.4) and (4.7)) and then obtain the solution by the thresholding operators similar to the "HardT" method. In contrast, "SURE" method requires a minimization of the SURE function (5.1) by trial-and-error. According to the CPU running time, our "DP" method is at least 11 times faster than the "SURE" method.

Now we compare the performance of different methods for the randomized algorithms, and we list the SNRs in Table 5.2. Similar conclusions can be observed among different methods. Note that for the nuclear-norm problem, our "DP" approach out-
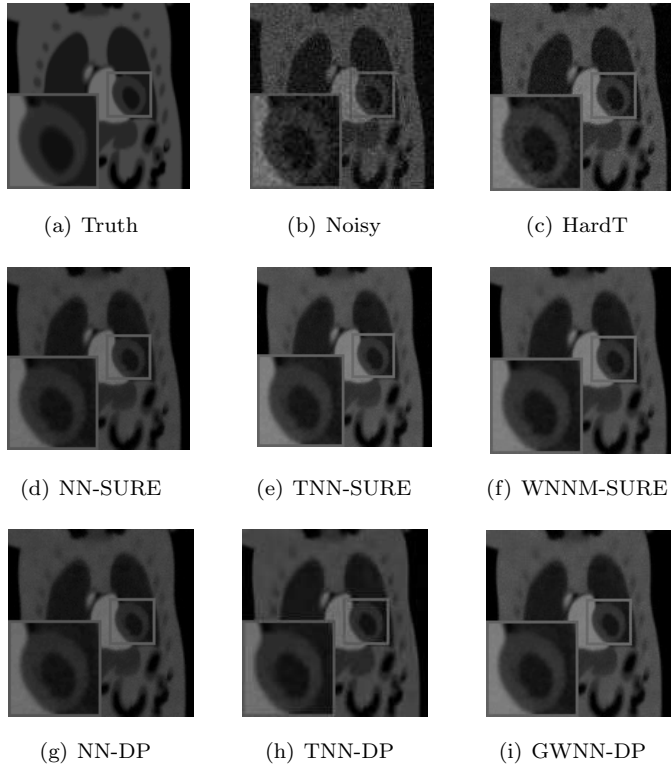
Table 5.2

Comparison of SNR(dB) for different methods with randomized algorithm.

		f(X)	Rank(X)	‖X‖∗		‖X‖r		‖X‖w,∗	
$m=n$	τ	ρ	HardT	NN-SURE	NN-DP	TNN-SURE	TNN-DP	GWNN-SURE	GWNN-DP
500	3	1%	**14.47**	11.78	*12.28*	***14.47***	***14.47***	*13.62*	13.15
		5%	**22.99**	21.75	*22.02*	***22.99***	22.97	22.75	*22.81*
		10%	**26.25**	25.52	*25.72*	***26.25***	**26.25**	26.15	*26.18*
		20%	29.43	29.05	*29.29*	29.43	**29.57**	29.39	*29.52*
		40%	32.48	32.19	*32.74*	32.48	**32.88**	32.47	*32.84*
	5	1%	**11.35**	8.19	*8.79*	***11.35***	***11.35***	*9.97*	9.45
		5%	**21.80**	19.39	*19.90*	***21.80***	21.79	*21.30*	21.29
		10%	**25.57**	23.90	*24.25*	***25.57***	25.52	*25.34*	25.32
		20%	**29.06**	28.04	*28.27*	***29.06***	29.03	28.93	*28.94*
		40%	32.29	31.72	*31.91*	32.29	**32.31**	32.25	*32.26*
Average SNR			24.57	23.15	*23.52*	24.57	***24.61***	*24.22*	24.18
Average time (sec)			0.005	0.116	0.006	0.123	0.007	0.117	0.006
1000	3	1%	**19.22**	17.71	*18.11*	***19.22***	19.18	18.90	*18.95*
		5%	26.46	26.00	*26.25*	26.46	**26.51**	26.43	*26.47*
		10%	29.51	29.32	*29.64*	29.51	**29.78**	29.50	*29.75*
		20%	32.54	32.53	*33.07*	32.54	**33.14**	32.54	*33.12*
		40%	35.56	35.76	*36.59*	35.82	**36.62**	35.89	*36.59*
	5	1%	**17.76**	15.56	*15.80*	***17.76***	17.75	*17.19*	16.99
		5%	**26.08**	25.04	*25.30*	***26.08***	26.01	*25.97*	25.92
		10%	**29.32**	28.72	*28.88*	***29.32***	29.28	*29.27*	29.23
		20%	32.44	32.12	*32.28*	32.44	**32.49**	32.41	*32.45*
		40%	35.51	35.35	*35.63*	35.51	**35.73**	35.51	*35.68*
Average SNR			28.44	27.81	*28.16*	28.47	***28.65***	28.36	*28.52*
Average time (sec)			0.020	0.366	0.022	0.394	0.023	0.393	0.021
2000	3	1%	22.64	22.19	*22.25*	22.49	***22.55***	22.32	*22.52*
		5%	29.50	29.69	*30.01*	29.73	*30.08*	29.49	***30.09***
		10%	32.54	33.33	*33.51*	33.36	*33.55*	33.15	***33.56***
		20%	35.64	36.83	*37.06*	36.84	*37.08*	37.02	***37.09***
		40%	38.95	40.72	*40.73*	40.73	***40.74***	40.54	*40.73*
	5	1%	**22.17**	21.25	*21.27*	***22.17***	22.09	*21.99*	21.96
		5%	**29.46**	29.12	*29.23*	***29.46***	29.43	*29.45*	29.39
		10%	32.52	32.35	*32.52*	32.52	**32.63**	32.51	*32.59*
		20%	35.55	35.46	*35.84*	35.55	**35.89**	35.54	*35.86*
		40%	38.57	38.56	*39.24*	38.57	**39.27**	38.57	*39.23*
Average SNR			31.75	31.95	*32.17*	32.14	***32.33***	32.06	*32.30*
Average time (sec)			0.089	1.810	0.095	2.040	0.094	2.016	0.092
4000	3	1%	23.82	25.39	*26.14*	25.49	*26.21*	25.51	***26.22***
		5%	32.88	33.98	*34.34*	33.99	*34.36*	33.94	***34.40***
		10%	36.60	37.71	*37.89*	37.72	*37.90*	37.68	***37.96***
		20%	40.11	*41.62*	41.51	***41.62***	41.51	*41.60*	41.56
		40%	43.70	*45.66*	45.28	***45.66***	45.28	*45.64*	45.31
	5	1%	25.57	24.69	*25.16*	25.02	**25.39**	*24.92*	25.37
		5%	32.55	32.33	*32.73*	32.36	**32.78**	32.48	*32.77*
		10%	35.56	35.87	*36.11*	35.88	**36.14**	35.43	*36.13*
		20%	38.58	39.18	*39.59*	39.19	**39.60**	39.29	*39.59*
		40%	41.62	42.95	*43.17*	42.95	**43.18**	42.98	*43.15*
Average SNR			35.10	35.94	*36.19*	35.99	*36.24*	35.95	***36.25***
Average time (sec)			**0.500**	9.363	*0.509*	9.532	*0.522*	9.578	0.516

perform the "SURE" approach in 38 out of the 40 different cases we tested. We also observe that in most cases, the SNRs obtained by the randomized algorithm are e-qual to or slightly lower than these by the standard algorithm. We remark that the difference of the SNRs comes from the difference between $YQ_\ell Q_\ell^T$ and Y. The speed advantage of the randomization algorithm is more obvious with the increase of matrix dimension. When $m = n = 4000$, the random algorithm is 28 times faster than the standard algorithm.

τ	Noisy	HardT	NN-SURE	NN-DP	TNN-SURE	TNN-DP	GWNN-SURE	GWNN-DP
5	27.50	33.19	*33.76*	33.71	34.34	*34.42*	***35.03***	35.02
10	21.48	27.60	*28.79*	28.73	29.64	*29.69*	***30.13***	**30.13**
15	17.96	24.33	*25.90*	25.85	26.89	*27.01*	***27.33***	27.31
20	15.46	22.01	*23.88*	23.83	25.09	*25.14*	***25.34***	25.31
25	13.52	20.21	22.23	*22.26*	*23.73*	23.70	***23.81***	23.76
30	11.94	18.74	*21.02*	20.98	**22.54**	22.53	***22.54***	22.48
35	10.60	17.49	19.84	*19.90*	**21.59**	21.55	*21.46*	21.40
40	9.44	16.41	*18.97*	18.96	**20.77**	20.70	*20.52*	20.46
45	8.42	15.45	18.11	*18.13*	**19.99**	19.94	*19.68*	19.62
50	7.50	14.59	17.29	*17.39*	**19.34**	19.27	*18.92*	18.88
Average	14.38	21.00	*22.98*	22.97	24.39	*24.40*	***24.48***	24.44
Average Time	–	19.01	733.34	20.12	761.91	20.06	754.24	20.81

(a) Truth   (b) Noisy   (c) HardT

(d) NN-SURE   (e) TNN-SURE   (f) WNNM-SURE

(g) NN-DP   (h) TNN-DP   (i) GWNN-DP

Figure 5.2. *The truth image with an enlarged portion cropped out from the image for $T = 10$, its noisy image by $\tau = 30$, and the denoised images by different methods.*

5.2. Simulations on PINCAT Numerical Phantom. In the following, we compare the performance of different methods in dynamic MRI-denoising. We use the same data set and parameter settings presented in [9]. There are 50 images of size 128×128 taken at 50 time-steps. The physiologically-improved NCAT (PINCAT) numerical phantom [47] simulates a first-pass myocardial perfusion real-time MRI series. The free-breathing model is available in the kt-SLR software package [32].

(a) NN-SURE	(b) TNN-SURE	(c) GWNN-SURE
(d) NN-DP	(e) TNN-DP	(f) GWNN-DP
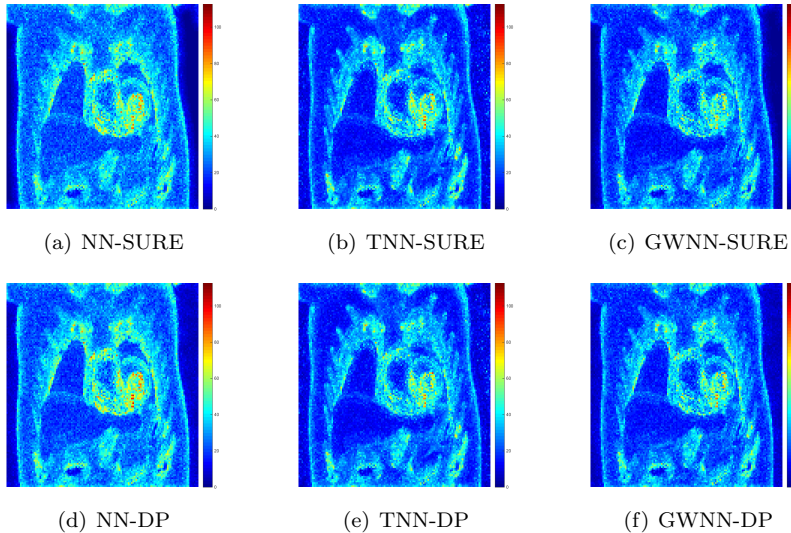
Figure 5.3. *The worst-case error through time of different methods for the the PINCAT Numerical Phantom.*

The noise is mainly caused by thermal noise and physiological noise, which follows the Gaussian distribution. Complex identical and independent Gaussian noise were added to the image data.

The locally low-rank recovery (LLR) method [48] is applied to remove the noise. LLR is the generalization of low-rank matrix reconstruction, which spatially partitions the image sequence into small blocks by a fixed sliding window with stride 1. In the experiment, the sliding window size is set to 7×7. We reshape each partitioned block into a 49×50 Casorati matrix (a matrix whose columns comprise vectorized patch of the image sequence). Because the changes of each frame in the sequence are small, the Casoratic matrix constructed from a clean image sequence is of low-rank. We solve the low-rank matrix minimization problem by using each of these Casoratic matrices as an observed matrix, and there are $128^2 = 16,384$ of such minimization problems to solve.

The SNR value is used to evaluate quantitatively the performance of different methods, which are listed in Table 5.3. The best results among all methods are shown in boldface and the best results for the given regularization norm are marked in italic font. We see from the average SNR values that our method is only 0.01dB lower than SURE for "NN" and "TNN" and 0.04dB lower for "GWNN". However, since we have to perform denoising on $16,384$ minimization problems of size 49×50 each, the "SURE" method is very time-consuming. Our "DP" method is at least 35 times faster than the "SURE" method.

Figure 5.2 shows the truth image with an enlarged portion cropped out from the image at the 10th time-step, the noisy image with $\tau = 30$, and the denoised images by different methods, respectively. We observe that there is still residual noise in the denoised image obtained by "HardT", which implies that the parameter chosen as in [18] is not suitable for LLR. This is because the threshold selected is related to the noise level only, and therefore the same threshold is selected for each block.

We also compare the worst-case absolute error[2] for our "DP" and "SURE" methods, which is shown in Figure 5.3. From the figures, we observe that "NN" exhibits higher residual error than "TNN" and "GWNN". Once again, the figures verify that the results obtained by the truncated nuclear norm or weighted nuclear norm are better than those by the standard nuclear norm or the hard thresholding scheme.

6. Conclusion. The constrained model and the regularized model for the low-rank matrix recovery were considered in this paper. We have derived a formula for the regularization parameter when a bound of the residual norm is given. The results were used to select the regularization parameter automatically using the discrepancy principle and to solve the constrained problem when the bound of the constraint is given. Experimental results showed that the proposed approach is competitive to the SURE methods in terms of noise removal and adaptive parameter selection, and is much faster in terms of CPU time.

REFERENCES

[1] S. W. Anzengruber and R. Ramlau, *Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators*, Inverse Problems, 26 (2009), 025001.
[2] D. P. Bertsekas, *Nonlinear programming*, Journal of the Operational Research Society, 48 (1997), pp. 332–334.
[3] J. Bigot, C. Deledalle, and D. Feral, *Generalized SURE for optimal shrinkage of singular values in low-rank matrix denoising*, The Journal of Machine Learning Research, 18 (2017), pp. 4991–5040.
[4] P. Blomgren and T. F. Chan, *Modular solvers for image restoration problems using the discrepancy principle*, Numerical Linear Algebra with Applications, 9 (2002), pp. 347–358.
[5] T. Bonesky, *Morozov's discrepancy principle and Tikhonov-type functionals*, Inverse Problems, 25 (2008), p.015015.
[6] J. Cai, E. J. Candès, and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM Journal on Optimization, 20 (2010), pp. 1956–1982.
[7] L. Cambier and P.-A. Absil, *Robust low-rank matrix completion by Riemannian optimization*, SIAM Journal on Scientific Computing, 38 (2016), pp. S440–S460.
[8] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 9 (2009), pp. 717–772.
[9] E. J. Candès, C. Singlong, and J. D. Trzasko, *Unbiased risk estimates for singular value thresholding and spectral estimators*, IEEE Transactions on Signal Processing, 61 (2013), pp. 4643–4657.
[10] C. Deledalle, S. Vaiter, J. Fadili, and G. Peyre, *Stein unbiased gradient estimator of the risk (SUGAR) for multiple parameter selection*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 2448–2487.
[11] G. Demoment, *Image reconstruction and restoration: Overview of common estimation structures and problems*, IEEE Transactions on Acoustics Speech and Signal Processing, 37 (1989), pp. 2024–2036.
[12] W. Dong, G. Shi, and X. Li, *Nonlocal image restoration with bilateral variance estimation: A low-rank approach*, IEEE Transactions on Image Processing, 22 (2013), pp. 700–711.
[13] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.
[14] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.
[15] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problem*, Kluwer Academic Publishers, 1996.
[16] X. Feng, W. Yu, and Y. Li, *Faster matrix completion using randomized SVD*, in IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), 2018, pp. 608–615.

---

[2]The worst-case absolute error is computed by $\max_t |X_{ij}(t) - \widehat{X}_{ij}(t)|$, where $X(t)$ and $\widehat{X}(t)$ are the truth image and the recovered image at time t.

[17] N. P. Galatsanos and A. K. Katsaggelos, *Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation*, IEEE Transactions on Image Processing, 1 (1992), p. 322.

[18] M. Gavish and D. L. Donoho, *The optimal hard threshold for singular values is $4/\sqrt{3}$*, IEEE Transactions on Information Theory, 60 (2014), pp. 5040–5053.

[19] M. Gavish and D. L. Donoho, *Optimal shrinkage of singular values*, IEEE Transactions on Information Theory, 63 (2017), pp. 2137–2152.

[20] G. H. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.

[21] E. W. Grever, *Using the L-curve for determining optimal regularization parameters*, Numerische Mathematik, 69 (1994), pp. 25–31.

[22] S. Gu, L. Zhang, W. Zuo, and X. Feng, *Weighted nuclear norm minimization with application to image denoising*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2014, pp. 2862–2869.

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

[24] N. Halko, P. G. Martinsson, and J. A. Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Review, 53 (2010), pp. 217–288.

[25] P. Hansen and D. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM Journal on Scientific Computing, 14 (1993), pp. 1487–1503.

[26] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM review, 34 (1992), pp. 561–580.

[27] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, *Fast and accurate matrix completion via truncated nuclear norm regularization*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (2012), pp. 2117–2130.

[28] Y. M. Huang, H. Y. Yan, Y. W. Wen, and X. Yang, *Rank minimization with applications to image noise removal*, Information Sciences, 429 (2018), pp. 147–163.

[29] A. N. Iusem, *On the convergence properties of the projected gradient method for convex optimization*, Computational and Applied Mathematics, 22 (2003), pp. 37–52.

[30] J. Josse and S. Sardy, *Adaptive shrinkage of singular values*, Statistics and Computing, 26 (2016), pp. 715–724.

[31] C. Lee and E. Y. Lam, *Computationally efficient truncated nuclear norm minimization for high dynamic range imaging*, IEEE Transactions on Image Processing, 25 (2016), pp. 4145–4157.

[32] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, *Accelerated dynamic MRI exploiting sparsity and low-rank structure: kt-SLR*, IEEE Transactions on Medical Imaging, 30 (2011), pp. 1042–1054.

[33] Q. Liu, Z. Lai, Z. Zhou, F. Kuang, and Z. Jin, *A truncated nuclear norm regularization method based on weighted residual error for matrix completion*, IEEE Transactions on Image Processing, 25 (2015), pp. 316–330.

[34] S. Ma, D. Goldfarb, and L. Chen, *Fixed point and bregman iterative methods for matrix rank minimization*, Mathematical Programming, 128 (2011), pp. 321–353.

[35] S. Mallat, *A Wavelet Tour of Signal Processing*, Elsevier, 1999.

[36] P. G. Martinsson, V. Rokhlin, and M. Tygert, *A randomized algorithm for the decomposition of matrices*, Applied and Computational Harmonic Analysis, 30 (2011), pp. 47–68.

[37] P. G. Martinsson and S. Voronin, *A randomized blocked algorithm for efficiently computing rank-revealing factorizations of matrices*, SIAM Journal on Scientific Computing, 38 (2016), pp. S485–S507.

[38] J. Mead, *χ^2 test for total variation regularization parameter selection*, Inverse Problems and Imaging, 14 (2020), pp. 401–421.

[39] J. L. Mead and C. C. Hammerquist, *χ^2 tests for the choice of the regularization parameter in nonlinear inverse problems*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 1213–1230.

[40] L. Mirsky, *Symmetric gauge functions and unitarily invariant norms*, The Quarterly Journal of Mathematics, 11 (1960), pp. 50–59.

[41] L. Mirsky, *A trace inequality of John Von Neumann*, Monatshefte für Mathematik, 79 (1975), pp.303306.

[42] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, 1984.

[43] T. H. Oh, Y. Matsushita, Y. W. Tai, and I. So Kweon, *Fast randomized singular value thresholding for nuclear norm minimization*, in Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition(CVPR), 2015, pp. 4484–4493.

[44] T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon, *Partial sum minimization of singular values in robust PCA: Algorithm and applications*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38 (2016), pp. 744–758.

[45] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, *Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (2012), pp. 2233–2246.

[46] J. Qi-Nian, *Applications of the modified discrepancy principle to Tikhonov regularization of nonlinear ill-posed problems*, SIAM Journal on Numerical Analysis, 36 (1999), pp. 475–490.

[47] B. Sharif and Y. Bresler, *Physiologically improved NCAT phantom (PINCAT) enables in-silico study of the effects of beat-to-beat variability on cardiac MR*, in Proceedings of the 15th Annual Meeting of ISMRM, vol. 3418, 2007.

[48] J. Trzasko, A. Manduca, and E. Borisch, *Local versus global low-rank promotion in dynamic MRI series reconstruction*, Proc. Int. Symp. Magn. Reson. Med., 19 (2011), p. 4371.

[49] Y. W. Wen, R. H. Chan, *Parameter selection for total-variation-based image restoration using discrepancy principle*, IEEE Transactions on Image Processing, 21 (2012), pp.1770-1781.

[50] J. Wright, A. Ganesh, S. R. Rao, Y. Peng, and Y. Ma, *Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization*, Neural Information Processing Systems, (2009), pp. 2080–2088.

[51] H. Xiang and J. Zou, *Regularization with randomized SVD for large-scale discrete inverse problems*, Inverse Problems, 29 (2013), 085008.

[52] S. K. Yadav, R. Sinha, and P. K. Bora, *A fast data driven shrinkage of singular values for arbitrary rank signal matrix denoising*, preprint, arXiv:1701.05223, (2017).

[53] H. Yan, L. Guisheng, X. Yijian, Z. Lei, and L. Jie, *Low-rank approximation via generalized reweighted iterative nuclear and frobenius norms*, IEEE Transactions on Image Processing, 29 (2019), pp. 2244–2257.