Weighted Schatten $p$-Norm Minimization for Image Denoising with Local and Nonlocal Regularization

Yuan Xie

Technical Report CSM-WSNM-STV

ISSN 1460-9673

January 2015
Weighted Schatten $p$-Norm Minimization for Image Denoising with Local and Nonlocal Regularization

Yuan Xie

Department of Computing, The Hong Kong Polytechnic University & Key Lab. of Complex Systems and Intelligence Science
Institute of Automation
Beijing, China
Telephone +86 137 1620 6758
Email yuan.xie@ia.ac.cn

Technical Report CSM-WSNM-STV

ISSN 1460-9673

January 2015
Abstract

This paper presents a patch-wise low-rank based image denoising method with constrained variational model involving local and nonlocal regularization. On one hand, recent patch-wise methods can be represented as a low-rank matrix approximation problem whose convex relaxation usually depends on nuclear norm minimization (NNM). Here, we extend the NNM to the nonconvex schatten p-norm minimization with additional weights assigned to different singular values, which is referred to as the Weighted Schatten p-Norm Minimization (WSNM). An efficient algorithm is also proposed to solve the WSNM problem. The proposed WSNM not only gives better approximation to the original low-rank assumption, but also considers physical meanings of different data components. On the other hand, due to the naive aggregation schema which integrates all the denoised patches into a whole image, current patch-wise denoising methods always produce various degree of artifacts in denoised results. Therefore, to further reduce artifacts, a data-driven regularizer called Steering Total Variation (STV) combined with nonlocal TV is derived for a variational model, which imposes local and nonlocal consistency constraints on the patch-wise denoised image. A highly simple but efficient algorithm is proposed to solve this variational model with convergence guarantee. Both WSNM and local & nonlocal consistent regularization are integrated into an iterative restoration framework to produce final results. Extensive experimental testing shows, both qualitatively and quantitatively, that the proposed method can effectively remove noise, as well as reduce artifacts compared with state-of-the-art methods.
1 Introduction

It is known that digital images captured by modern cameras may often get corrupted by noise during the process of acquisition and transmission [1], rendering those images unsuitable for vision application such as surveillance, object recognition, and so on. The major challenge of image denoising is to suppress noise while keeping important image structures and details. Generally, the degradation model for denoising problem can be described as:

\[
y = x + n
\]

where column vectors \( x \) and \( y \) represent the vectorized underlying latent image and its noisy observation, respectively. \( n \) denotes the noise which is assumed to be additive white Gaussian noise (AWGN) in this paper.

As a primary low-level image processing procedure, image denoising has been extensively studied and many approaches have been proposed [2–13]. The success of recent denoising methods partly stem from nonlocal technique, which exploits nonlocal self-similarities in natural images. The Nonlocal Means (NLM) [9] is the first filter that makes use of the nonlocal self-similarity in the whole image, and obtains a denoised patch by weighted averaging all the similar patches. BM3D [10], a representative benchmark method, builds on the concept of NLM by finding similar patches and grouping them together, and then performs collaborative filtering in the transform domain. In [12], the low rank decomposition is introduced into denoising approach toward modeling nonlocal similarity in images. The derived method, which is referred to as Sparsely-Adaptive Iterative Singular-value Thresholding (SAIST), extends the soft-thresholding from local wavelet-based models to nonlocal SVD-based (singular value decomposition) models. A similar observation is exploited in [19], which proposes the weighted nuclear norm minimization (WNNM) to impose low rank on nonlocal similar patches. Compared with traditional nuclear norm minimization (NNM, a common regularizer used in low rank matrix approximation (LRMA)), the proposed WNNM gets a natural way to assign different weights to different singular values such that the values of soft thresholds become more reasonable.

Another class of the denoising methods utilize the sparse and redundant representations over trained dictionary to produce a high-quality denoised image [15–18]. K-SVD proposed in [15] takes advantage of this assumption to learn an effective overcomplete dictionary for the noisy image patches, such that each noisy patch can be represented as a linear combination of only a few atoms among the dictionary. Instead of learning a single overcomplete dictionary for the entire image, the authors of the K-LLD [16] first perform a clustering step on patches by using local weight function from [24], and then denoise the patches from each cluster separately by finding best dictionary for each cluster. The motivation is that similar patches should share similar subdictionaries. Similarly, by introducing the idea of the NLM, Mairal et al. [17] explicitly exploits self-similarities in natural images combining with sparse coding to improve the performance of the dictionary based denoising method. The proposed method, called learned simultaneous sparse coding (LSSC), can impose that similar patches share the same dictionary elements in their sparse decomposition.

As patch-based model, most of the aforementioned approaches, either nonlocal based or dictionary based methods always produce various degree of artifacts in denoised results, due to the naive aggregation schema (averaging or weighted averaging) which integrates all the denoised patches into a final image. Fortunately, there exist lots of advanced filters [9, 24, 25] or regularizers [26–29] that can help to remove possible artifacts by enforcing local or nonlocal constraint. Total Variation (TV) [26–27] tends to smooth the homogenous regions while preserving edges sharpness. Based on the spirit of TV and the bilateral filter [25], Farsiu et al. [28] introduce a robust regularizer called Bilateral Total Variation (BTV) to impose local constraint on reconstructed image, which considers a larger neighborhood than TV with much less computational complexity. More recently, [29] extends NLM to nonlocal total variation (NLTV) by including variational method using functionals with nonlocal regularization. By sharing the similar idea, our previous work [35] proposes a locally adaptive regularizer called steering total variation (STV) based on steering kernel regression (SKR) [24], and apply it to single image super resolution successfully.

In this paper, we present a patch-wise low-rank based image denoising method with image-wise variational constraint involving local and nonlocal regularization. Traditional LRMA is usually induced by nuclear norm — the tightest relaxation of original rank minimization problem [38, 39]. However, this kind of relaxation may suffer from two major drawbacks. First, it suppresses the low rank components and shrinks the reconstructed data as pointed out in details in experimental section [6, 1]. Second, it treats all low rank components equal, ignoring their physical meanings in many practical problems, then gets suboptimal estimation of the low-rank components [42]. Even though different weights have been considered, the effect of the WNNM can still be degraded by the first drawback. To overcome the aforementioned problem, inspired by schatten \( p \)-norm minimization [40–43] and WNNM [19], we propose a new low rank regularizer namely \textbf{W}eighted \textbf{S}chatten \textbf{p}-\textbf{N}orm \textbf{M}inimization (WSNM), and analyze its optimal solution in LRMA. It not only introduces unequal treatments for different rank components, leading to flexibility in dealing with many practical problem; but also gives better approximation to the original low
rank assumption, getting better results theoretically and practically. As it can be seen later, the WSNM includes the WNNM into a unified rank based framework, meeting that the WNNM is a special case of the proposed WSNM.

After performing low-rank-based patch-wise denoising, the reconstructed image is further refined by local and nonlocal regularizer that combines both NLTV and STV. Such a regularizer, capturing nonlocal repetitive structures and local data-driven weights distribution of an image simultaneously, can reduce artifacts effectively. To the best of our knowledge, this kind of regularizer has never been considered in image denoising before. Moreover, we propose a simple but highly efficient optimization algorithm to solve this regularization problem with convergence guarantee. Both WSNM and local & nonlocal consistent regularizer are integrated into an iterative restoration framework to produce final output. The extensive experimental results demonstrate that our proposed method outperforms many state-of-the-art denoising methods both visually and quantitatively.

The contributions of this paper are summarized as follows:

- We propose a new and reasonable object function for low rank matrix approximation, namely Weighted Schatten $p$-Norm Minimization (WSNM). We also analyze and provide the optimal solution of the WSNM.
- A strong prior term that combines the STV and NLTV is proposed to be used as image-wise regularizer, leading to a further improvement in denoised result. Moreover, we design a simple optimization algorithm to solve this regularization, which is supported by a rigorous proof of convergence.
- We adopt the proposed WSNM with local and nonlocal regularization to image denoising to achieve excellent performance, which demonstrates their great potentials in low level vision application.

The remainder of this paper is organized as follows. Section 2 gives a briefly review of the NLTV and SKR. In Section 3 we describe our proposed WSNM in details and give an optimization algorithm to solve it. In Section 4 we motivate the proposed local and nonlocal consistent regularizers, and develop a new optimization algorithm to solve it with convergence guarantee. In Section 5 the WSNM and local & nonlocal regularizers are used to construct an iterative image denoising algorithm. The experimental results are demonstrated in Section 6 and we conclude the proposed methods and discuss the future direction in Section 7.

2 Related Work

2.1 Non-Local Total Variation

Suppose a digital image model by a graph $(\Omega, E)$, where $\Omega$ is a finite set of $N$ nodes (pixels), $E$ is the set of edges. The notation $x \sim y$ is used to denote the edge between the nodes $x$ and $y$. An image $u$ is a function defined on $\Omega$, which can be represented by a column vector, then the value at node $x$ can be denoted by $u(x)$. In the following, we consider a weight function $w(x, y)$ for the edge $x \sim y \in E$. The weight function is symmetric and can be set to 0 if two nodes $x$ and $y$ are not connected. In this case, unlike classical total variation, nodes of NLTV may directly interact with nodes that are not neighbors, which means “nonlocal”.

For a given image $u(x)$ defined on $\Omega$, the weight graph gradient $\nabla_w u(x)$ is defined as the vector of all directional derivatives (or edge derivative) $\nabla_wu(x, \cdot)$ at $x$:

$$\nabla_w u(x) := (\nabla_w u(x, y))_{y \in \Omega}$$

where

$$\nabla_w u(x, y) := \sqrt{(u(y) - u(x))^2 w_n(x, y)}, \forall y \in \Omega$$

The directional derivatives apply to all the nodes $y$ since the weight $w(x, y)$ is extended to the whole domain $\Omega \times \Omega$. Here, the similarity weight is given by:

$$w_n(x, y) = \exp \left(-\frac{||N_x - N_y||_G^2}{h_n^2}\right)$$

where $N_x$ and $N_y$ represent the column vectors by expanding the pixels of the image patches centered at the locations of $x$ and $y$ in lexicographic ordering, respectively. $h_n$ is the bandwidth of the weighting computation, and parameter $G$ is a Gaussian kernel that assigns a larger weight to the pixels close to the center one. Additionally, a graph divergence $div_w$ of a vector $p : \Omega \times \Omega \rightarrow \mathbb{R}$ can be defined as follows:

$$div_w p(x) = \sum_{y \in \Omega} (p(x, y) - p(y, x)) \sqrt{w_n(x, y)}$$

2
Then, the graph Laplacian is defined by:

$$\Delta_w u(x) := \frac{1}{2} div_w(\nabla_w u(x)) = \sum_{y \in \Omega} (u(y) - u(x))w_n(x, y)$$

Using the notations above, the nonlocal total variation can be defined as follows:

$$J_n(u) = J_{NLTV}(u) := \frac{1}{2} \sum_{x \in \Omega} |\nabla_w u(x)| = \frac{1}{2} \sum_{x \in \Omega} \left( \sum_{y \in \Omega} (u(x) - u(y))^2 w_n(x, y) \right)$$  \hspace{1cm} (2.1)$$

For simplicity, let $x_i$ and $x_j$ denote the $i$-th and $j$-th pixels in image $x$ respectively, then the NLTV can be described more concisely:

$$|\nabla_n(x)| = \sum_{i \in \Omega} \sqrt{\sum_{j \in P(x_i)} (x_j - x_i)^2 w_n(i, j)}$$  \hspace{1cm} (2.2)$$

where $P(x_i)$ represents the nonlocal searching region of pixel $x_i$.

### 2.2 Steering Kernel Regression

Steering kernel regression, which is one of the most representative data-adaptive kernel regression filters, has been successfully used for image restoration [24]. Under SKR framework, the image model can be represented by

$$y_i = z(x_i) + \varepsilon_i, \quad i = 1, \ldots, W \times W$$  \hspace{1cm} (2.3)$$

where $y_i$ is a noisy sample at $x_i$, $z(\cdot)$ is the regression function to be estimated, $\varepsilon_i$ represents an i.i.d. zero mean noise, $W \times W$ is the size of local analysis window around a pixel position $x$ of interest. The kernel regression framework provides a rich mechanism for computing point-wise estimates of the regression function with minimal assumptions about global signal or noise models. As a local method, the SKR uses 2nd order Taylor series to estimate the value at position $x$:

$$\hat{z}(x_i) = \arg\min_{z_i} \sum_{i=1}^{W \times W} |y_i - z(x_i)|^2$$  \hspace{1cm} (2.4)$$

$$\approx \arg\min_{z_i} \sum_{i=1}^{W \times W} |y_i - \beta_0 - \beta_1^T (x_i - x) - \beta_2^T \text{vech}((x_i - x)(x_i - x)^T)|^2 K_{H^k}(x_i - x)$$  \hspace{1cm} (2.5)$$

where $\text{vech}(\cdot)$ stands for an operation to stack the lower triangular part of a matrix into a column vector, and $\beta_0$, $\beta_1$ and $\beta_2$ are as follows:

$$\beta_0 = z(x), \quad \beta_1 = \nabla z(x) = \begin{bmatrix} \frac{\partial z(x)}{\partial x_1}, \frac{\partial z(x)}{\partial x_2} \end{bmatrix}^T$$  \hspace{1cm} (2.6)$$

$$\beta_2 = \frac{1}{2} \begin{bmatrix} \frac{\partial^2 z(x)}{\partial x_1^2}, \frac{\partial^2 z(x)}{\partial x_1 \partial x_2}, \frac{\partial^2 z(x)}{\partial x_2^2} \end{bmatrix}^T$$  \hspace{1cm} (2.7)$$

where $x_1$ and $x_2$ are 2D coordinates of the pixel $x$, $K_{H^k}(\cdot)$ is actually a data-adapted steering kernel function, which relies on not only the sample location and density but also on the radiometric properties of these samples. The steering kernel is mathematically represented as:

$$K_{H^k}(x_i - x) = \sqrt{\frac{\det(C_1)}{2\pi h_k^2}} \exp\left(-\frac{(x_i - x)^T C_1 (x_i - x)}{2h_k^2}\right)$$  \hspace{1cm} (2.8)$$

where $C_1$ is a symmetric gradient covariance matrices at $x_i$ in the vertical and horizontal directions, and $h_k$ is a smoothing parameter to control the supporting range of the steering kernel. Transforming the Eqn. (2.5) into matrix form, we can get a weighted least-squares optimization problem:

$$\hat{b} = \arg\min_b \|y - X_b b\|_W^2$$

$$= \arg\min_b (y - X_b b)^T W_x (y - X_b b)$$  \hspace{1cm} (2.9)$$
where

\[ y = [y_1, y_2, \ldots, y_p]^T, \quad b = [\beta_0, \beta_2^T, \ldots, \beta_N^T] \]

\[ W_x = \text{diag}[K_{H_1}(x_1 - x), K_{H_1}(x_2 - x), \ldots, K_{H_1}(x_p - x)] \]

\[ X_x = \begin{bmatrix}
1 & (x_1 - x)^T & \text{vech}^T \{ (x_1 - x)(x_1 - x)^T \} \\
1 & (x_2 - x)^T & \text{vech}^T \{ (x_2 - x)(x_2 - x)^T \} \\
\vdots & \vdots & \vdots \\
1 & (x_1 - x)^T & \text{vech}^T \{ (x_1 - x)(x_1 - x)^T \}
\end{bmatrix} \]

where “diag” defines a diagonal matrix, and \( P = W \times W \). Since our primary interest is to estimate the \( z(x) \) which is actually \( \beta_0 \), it corresponds to the first element in the vector \( \hat{\alpha} \)

\[ \hat{\alpha} = (X_x^T W_x X_x)^{-1} X_x^T W_x y \]

where \( (X_x^T W_x X_x)^{-1} X_x^T W_x \) can be reformulated as the so-called equivalent kernel. For more details, please refer to the literature [24].

### 3 Low-Rank Matrix Approximation with Weighted Schatten \( p \)-Norm

Generally, the problem of reconstructing a matrix \( X \) based on a few observed entries is ill-posed. There are infinite number of matrices which agree with the observed entries of \( X \) perfectly. Therefore, without additional assumptions, it is difficult to prefer some matrices over others as candidates for \( \hat{X} \). Hence, many approaches assume that the values in the data matrix are always correlated and the rank of the data matrix is low, which is the well known Low-Rank assumption [20, 21]. However, the original rank minimization problem (RMP) is NP-hard in general and NP-hard to approximate. The nuclear norm, the tightest convex relaxation of original RMP, is commonly used:

\[ ||X||_* = \sum_{i=1}^{\min\{n,m\}} \sigma_i(X) \]

where \( \sigma_i \) is the \( i \)-th singular value of \( X \). Cai et al. [22] have proved that the nuclear norm based LRMA problem with F-norm data fidelity can be easily solved by a soft-thresholding operation on the singular values of observation matrix. But the nuclear norm suppresses the rank components and ignores the imbalance among different data components. To overcome those shortcomings, in the following subsections, we firstly define the proposed WSNM, then develop an algorithm to solve the WSNM based LRMA model.

#### 3.1 Definitions of Weighted Schatten \( p \)-Norm

The Schatten \( p \)-norm (\( 0 < p \leq 1 \)) of a matrix \( X \in \mathbb{R}^{n \times m} \) is defined as

\[ ||X||_{S_p} = \left( \sum_{i=1}^{\min\{n,m\}} \sigma_i^p \right)^{\frac{1}{p}} = \left( \text{Tr}((X^T X)^\frac{p}{2}) \right)^{\frac{2}{p}} \]

where \( \sigma_i \) is the \( i \)-th singular value of \( X \). The widely used trace norm (or nuclear norm) can be represented by choosing \( p = 1 \):

\[ ||X||_{S_1} = \left( \sum_{i=1}^{\min\{n,m\}} \sigma_i \right) = \left( \text{Tr}((X^T X)^\frac{1}{2}) \right). \]

As \( p \) becomes small, Eqn. (3.2) is closer to the rank constraint:

\[ \lim_{p \to 0} \text{Tr}(X^T X)^\frac{p}{2} = \lim_{p \to 0} \sum_k \sigma_k^p = \text{rank}(X) \]

here, we define \( 0^0 = 0 \). Having the definition above, we introduce the proposed weighted schatten \( p \)-norm which is defined as

\[ ||X||_{w,S_p} = \left( \sum_{i=1}^{\min\{n,m\}} w_i \sigma_i^p \right)^{\frac{1}{p}} \]
where \( w = [w_1, \ldots, w_{\min(n,m)}] \) is a non-negative vector, introducing the unequal treatment for different rank components. Then the weighted schatten \( p \)-norm of a matrix \( X \) to the power \( p \) is

\[
\|X\|_{w,S_p}^p = \min_{\min(n,m)} \sum_{i=1}^{\min(n,m)} w_i \sigma_i^p = \text{tr}(W \Delta^p)
\]  
(3.6)

where both \( W \) and \( \Delta \) are diagonal matrices whose diagonal entries are comprised of \( w_i \) and \( \sigma_i \), respectively.

### 3.2 LRMA Problem and Optimization

Given a matrix \( Y \), our proposed LRMA model aims to find a matrix \( X \), which is as close to \( Y \) as possible under F-norm data fidelity and WSNM:

\[
\hat{X} = \arg \min_X \|X - Y\|_F^2 + \|X\|_{w,S_p}^p.
\]  
(3.7)

We will solve the weighted schatten \( p \)-norm minimization problem in this subsection. Having discussed in [19], the convexity property of the optimization problem can not be preserved because of adding the weights to NNM. Furthermore, the nonconvex relaxation brought by the schatten \( p \)-norm will make the above LRMA much more difficult to optimize. To solve (3.7), we give the following theorem before analyzing the optimization of the proposed WSNM:

**Theorem 3.1.** (Von-Neumann) For any \( m \times n \) matrices \( A \) and \( B \), \( \sigma(A) = [\sigma_1(A), \ldots, \sigma_r(A)]^T \) and \( \sigma(B) = [\sigma_1(B), \ldots, \sigma_r(B)]^T \) where \( r = \min(m,n) \), are the singular values of \( A \) and \( B \) respectively, then \( \text{tr}(A^T B) \leq \text{tr}(\sigma(A)^T \sigma(B)) \). The case of equality occurs if and only if it is possible to find unitaries \( U \) and \( V \) that simultaneously singular value decompose \( A \) and \( B \) in the sense that

\[
A = U \Sigma_A V^T, \text{ and } B = U \Sigma_B V^T
\]  
(3.8)

where \( \Sigma_A \) and \( \Sigma_B \) denote ordered eigenvalue matrices with the singular values \( \sigma(A) \) and \( \sigma(B) \) along the diagonal with the same order, respectively.

**Lemma 3.2.** Suppose all the singular values are in the same order (ascending or descending), let the SVD of \( Y \) be \( Y = U \Sigma V^T \) with \( \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r) \), then the optimal solution of (3.7) be \( X = U \Delta V^T \) with \( \Delta = \text{diag}(\delta_1, \ldots, \delta_r) \), where \( \delta_i \) is given by solving the problem below:

\[
\min_{\delta_1, \ldots, \delta_r} \sum_{i=1}^r \left[(\delta_i - \sigma_i)^2 + w_i \delta_i^p\right], \text{s.t. } \delta_i \geq 0, i = 1, \ldots, r.
\]  
(3.9)

**Proof.** Let the optimal solution of the problem (3.7) have the compact SVD \( X = Q \Delta R^T \), and the SVD of the matrix \( Y \) is \( Y = U \Sigma V^T \), both \( \Delta \) and \( \Sigma \) are ordered eigenvalue matrices with the same order. Using the theorem (3.1), we have

\[
\|X - Y\|_F^2 = \text{tr}(\Delta^T \Delta) + \text{tr}(\Sigma^T \Sigma) - 2\text{tr}(X^TY) \\
\geq \text{tr}(\Delta^T \Delta) + \text{tr}(\Sigma^T \Sigma) - 2\text{tr}(\Delta^T \Sigma) = \|\Delta - \Sigma\|_F^2.
\]  
(3.10)

This implies

\[
\|X - Y\|_F^2 + \text{tr}(W \Delta^p) \geq \|\Delta - \Sigma\|_F^2 + \text{tr}(W \Delta^p).
\]  
(3.11)

Note that the equality holds if and only if \( Q = U \) and \( R = V \) according to (3.8). Therefore, minimizing the (3.7) can be reduced to minimizing Eqn. (3.9).

According to Lemma 3.2 solving the problem (3.9) is much easier than the original problem (3.7). Obviously, the optimization of (3.9) can be decoupled into \( r \) independent subproblem for each \( \delta_i \)

\[
J(\delta) = \min_{\delta_i} (\delta_i - \sigma_i)^2 + w_i \delta_i^p, \text{s.t. } \delta_i \geq 0,
\]  
(3.12)

which is a quadratic equation in one variable, then it has two local optima: \( \delta_i^* > 0 \) and \( \delta_i^2 = 0 \). In practice, since its quadratic convergence, we can use the standard Newton’s method to compute \( \delta_i^* \) with the update rule:

\[
\delta = \delta - \frac{J'(\delta)}{J''(\delta)}
\]  
(3.13)
where

\[ J'(\delta) = 2(\delta - \sigma_i) + w_i p \delta^{p-1} \]  
(3.14)

\[ J''(\delta) = 2 + w_i (p - 1) \delta^{p-1} \]  
(3.15)

Then, choose the \( \delta \in \{ \delta_1^*, \delta_2^* \} \) which minimizes the \( J(\delta) \) as the final solution. When \( p = 1 \), the optimal solution can be obtained by setting the first derivation of Eqn. (3.12) equal to zero, then we get

\[ \delta_1^* = \left( \sigma_i - \frac{w_i}{2} \right) \]  
(3.16)

where \( x_+ = \max\{x, 0\} \). Therefore, it can be easily verified that the optimal solution of (3.12) is

\[ \delta_1^* = \max\left( \sigma_i - \frac{w_i}{2}, 0 \right) \]  
(3.17)

which is exactly the same as in WNNM [19, 23]. Hence, the WNNM is a special case of the proposed WSNM.

Note that in [19, 23], authors have discussed the solution of the WNNM under three different configurations of the weights: non-ascending order, non-descending order and arbitrary order. Only in case of non-ascending order, can the WNNM problem reach a closed-form global optimal solution, while other cases merely have local optimum. However, we can conclude from the Lemma 3.2 that, unlike the complicated solution in WNNM, the proposed WSNM provides a unified solution for any possible permutation of the weights.

4 Consistency Constraint via Local and Nonlocal Regularization

Applying the WSNM to image denoising actually performs the patch level noise removal, and then the denoised result is obtained by aggregating all the denoised patches into a whole one. Such aggregation is usually realized by using weighted averaging neighboring patches in a certain region, leading to artifacts easily. Therefore, we need a regularization to form a reconstruction constraint to remove possible artifacts, and promote the restored image more consistent and natural. In the following subsections, we describe the steering total variation and the proposed variational model, then develop a fast algorithm to solve it efficiently.

4.1 Steering Total Variation

NLTV has been used as a popular and effective image prior model in regularization-based imaging problem [30, 31], due to its reduction of undesired staircase effects presented in Total Variation results. Actually, the choice of directions characterized by traditional local gradients, i.e. gradients along horizontal and vertical, is regarded as the main drawback of TV prior. To circumvent this limitation, NLTV generalizes the traditional gradients to non-local gradients by using image-driven directions, as shown in (2.1), which is defined on a similarity graph. Therefore, NLTV can explore repetitive structures to preserve important details of an image. While resorting to non-local similar patterns, NLTV ignores the local consistency since the neighborhood always provides important information for recovering a certain pixel. However, those neighboring pixels don’t make equal contributions for recovery. For example, suppose a pixel is located near an edge, then its nearby pixels which place on the same side of the edge will have much stronger influence than the pixels on the opposite side. Therefore, even in a local region, the weights of neighbors should be treated unequally in estimating latent value of a certain pixel.

Fortunately, [24] provides an elegant tool namely steering kernel, which could capture local weight distribution for estimating latent value of a certain pixel. Inspired by NLTV and steering kernel, our previous work [35] propose a novel regularization term called Steering Total Variation (STV) to exploit the image local correlation:

\[ |\nabla_s(x)| = \sum_{i \in \Omega} \sqrt{\sum_{j \in N(x_i)} (x_j - x_i)^2} w_s(i, j) \]  
(4.1)

where \( w_s(i, j) \) is the element of the equivalent kernel indicated by Eqn. (2.13) in \( W \times W \) local region.

4.2 Local and Nonlocal Constraint via the Extended-ROF Model

Both NLTV and STV are defined under an unified variational framework to impose local and nonlocal consistent constraint:

\[ \min_u \frac{1}{2} \|u - v\|^2_2 + \mu_1 |\nabla_n u|_1 + \mu_2 |\nabla_s u|_2 \]  
(4.2)
where $u$ denotes the image to be retrieved, $v$ is assigned by the patch-wise restored image. Unlike TV regularizer based Rudin-Osher-Fatemi (ROF) model [26], we refer to the problem (4.2) as the Extended-ROF model. Nevertheless, the minimization of this model remains as a difficult optimization problem due to the computation complexity and the non-differentiability. As suggestion in [30] [35], the split Bregman can be used to solve the above optimization problem (4.2). The key idea of the split Bregman is that it “de-couple” the $l_1$ and $l_2$ portions of the energy in (4.2). By using split Bregman, we can achieve an iterative optimization algorithm as follows (the detailed derivation in Appendix B in supplemental material):

$$u^{k+1} = (I - 2\lambda_2 \Delta_n - 2\lambda_2 \Delta_s)^{-1} (v + \lambda_3 \text{div}_n(b^k_n - d^k_n) + \lambda_3 \text{div}_s(b^k_s - d^k_s))$$  \hspace{1cm} (4.3)

$$d^{k+1}_n = \text{shrink}\left(\nabla_n u^{k+1} + b^k_n, \frac{\mu_1}{\lambda_1}\right)$$ \hspace{1cm} (4.4)

$$d^{k+1}_s = \text{shrink}\left(\nabla_s u^{k+1} + b^k_s, \frac{\mu_2}{\lambda_2}\right)$$ \hspace{1cm} (4.5)

$$b^{k+1}_n = \nabla_n u^{k+1} + b^k_n - d^{k+1}_n$$ \hspace{1cm} (4.6)

$$b^{k+1}_s = \nabla_s u^{k+1} + b^k_s - d^{k+1}_s$$ \hspace{1cm} (4.7)

where $\nabla_n$ and $\nabla_s$ denote non-local gradient operator and local steering gradient operator respectively, and $\Delta_n$ and $\Delta_s$ present non-local and local steering graph Laplacian operator respectively, div$_n$ and div$_s$ are graph divergence operators (See the mathematical definitions of all those operators in Appendix A). Moreover, the function $\text{shrink}(\cdot, \cdot)$ is defined as:

$$\text{shrink}(x, \gamma) = \frac{x}{|x|} \cdot \max(|x| - \gamma, 0)$$ \hspace{1cm} (4.8)

Commonly, using the Split Bregman method, as shown in Eqn. (4.3), still requires solving a partial difference equation in each iteration, therefore the resulting algorithms are very complicated. Even though the Gauss-Seidel method has been used [30] [35], the above algorithm still needs to calculate many Laplacian operators, which introduce extra computational time. Accordingly, in the next subsection, a fast algorithm without solving PDEs is proposed to overcome this drawback.

### 4.3 Finding an Efficient Solution for the Extended-ROF Model

Here, we extend the method proposed in [34], which only handles the standard ROF model, to the general extended-ROF case, and provide a very simple but highly efficient and effective optimization algorithm without involving any PDE calculation.

The following new iteration schema can find the unique solution $u^*$ efficiently for the minimization problem (4.2):

Let $b_n^0 = 0, b_s^0 = 0$ and $u^1 = v$, for $k = 1, 2, \ldots$, the iteration is as follows:

$$b_n^k = \text{cut}(\nabla_n u^k + b_n^{k-1}, \frac{\mu_1}{\lambda_1})$$ \hspace{1cm} (4.9)

$$b_s^k = \text{cut}(\nabla_s u^k + b_s^{k-1}, \frac{\mu_2}{\lambda_2})$$ \hspace{1cm} (4.10)

$$u^{k+1} = v + \frac{\lambda_1}{\mu_1} \text{div}_n b_n^k + \frac{\lambda_2}{\mu_2} \text{div}_s b_s^k$$ \hspace{1cm} (4.11)

where $\text{cut}(x, \gamma) = x - \frac{x}{|x|} \cdot \max(|x| - \gamma, 0)$.

In the above equations, $\nabla_n$ and $\nabla_s$ denote non-local gradient operator and local steering gradient operator respectively, div$_n$ and div$_s$ are the non-local and local graph divergence operators respectively (see the detailed mathematical definitions of all these operators in the Appendix). We prove the convergence of the proposed fast algorithm in the following theorem:

**Theorem 4.1.** For $k = 1, 2, \ldots$, let $b_n^k, b_s^k$ and $u^{k+1}$ be given by the iteration (4.9) to (4.11). If $0 < 20\lambda_1 + 25\lambda_2 < 1$, then $\lim_{k \to \infty} u^k = u^*$.

**Proof.** See the proof in Section [A.1]
5 Iterative Image Denoising

In this section, we explain how to integrate the proposed WSNM and the local & nonlocal regularization into an iterative framework to construct an effective denoising method. For a local \( n_p \times n_p \) patch \( y_i \) in degraded image \( y \), we search its nonlocal similar patches \( \{\hat{y}_j\}_{j=1}^{n_s} \) by the block matching method proposed in [10]. Then, \( \{\hat{y}_j\}_{j=1}^{n_s} \) can be stacked into a matrix \( Y_i \), whose columns are composed of the vectorized patches \( y_i \), \( (i = 1, \ldots, n_t) \). According to degradation model (1.1), we have \( Y_i = X_i + N_i \), where \( X_i \) and \( N_i \) are the patch matrices of original image and noise, respectively. Under the assumption of low rank, the matrix \( X_i \) can be estimated from \( Y_i \) by using the low rank matrix approximation methods. Hence, we apply the proposed WSNM model to estimate \( X_i \), and its corresponding optimization problem can be defined as

\[
\hat{X}_i = \arg \min_{X_i} \frac{1}{\sigma_n} \|Y_i - X_i\|_F^2 + \|X_i\|_{W,S_p}^p \tag{5.1}
\]

where \( \sigma_n^2 \) denotes the noise variance, the first term of (5.1) represents the F-norm data fidelity term, and the second term plays the role as weighted schatten term for low rank approximation purpose. Usually, \( \sigma_n \) should be inversely proportional to \( \delta \) since assumption that the empirical distribution of singular values can also be modeled by an i.i.d. Laplacian distribution.

Moreover, we analysis how to make influence on the quality of the denoised image with the changing values of \( \gamma \). Similarly, \( \delta_j (X_i) \), the \( j \)-th optimal solution of (5.1), owns the same property such that the larger the value of \( \delta_j (X_i) \), the less it should be shrunk. Therefore, an intuitive way for setting weights is that the weight should be inversely proportional to \( \delta_j (X_i) \), then we have

\[
w_j = c\sqrt{n} / (\delta_j (X_i) + \varepsilon) \tag{5.2}
\]

where \( n \) is the number of similar patches in \( Y_i \), \( \varepsilon \) is set to \( 10^{-16} \) to avoid dividing by zero, and \( c = 2\sqrt{2}\sigma_n^2 \), in which \( \sigma_n^2 \) denotes noise variance. The setting of coefficient \( c \) is according to the thumb rule used in [12] under the assumption that the empirical distribution of singular values can also be modeled by an i.i.d. Laplacian distribution. Since \( \delta_j (X_i) \) is unavailable, it can be initialized by

\[
\delta_j (X_i) = \sqrt{\max \{\sigma_j^2 (Y_i) - n\sigma_n^2, 0\}} \tag{5.3}
\]

To restore noisy image iteratively, we adopt the iterative regularization scheme in [12], which is to add filtered noise back to the denoised image as follows:

\[
y^{(k)} = \hat{x}^{(k-1)} + \alpha (y - \hat{x}^{(k-1)}) \tag{5.4}
\]

where \( k \) denotes the iteration number and \( \alpha \) is a relaxation parameter. Inspired by this iterative scheme, the estimation of noise variance can be updated by

\[
\sigma_n^{(k+1)} = \gamma \sqrt{\sigma_n^2 - \|y - y^{(k)}\|^2} \tag{5.5}
\]

where \( \gamma \) is a scaling factor controlling the re-estimation of noise variance. Then, in next iteration, \( \delta_j^{(k+1)} (X_i) \) can be initialized as:

\[
\delta_j^{(k+1)} (X_i) = \sqrt{\max \{(\sigma_j^{(k)} (Y_i))^2 - n(\sigma_n^{(k+1)})^2, 0\}} \tag{5.6}
\]

Given the initial value of \( \delta_j^{(k+1)} (X_i) \), the corresponding weight \( w_j \) can be obtained by using Eqn. (5.2), then the optimal value of \( \delta_j^{(k+1)} (X_i) \) can be achieved by solving problem (5.1). By applying the above procedures to each patch and aggregating all filtered patches together, the reconstructed image \( x \) can be further regularized by using optimization model (4.2) to produce final output. The whole denoising algorithm is summarized in Algorithm 1.

6 Experimental Results and Analysis

This section presents extensive experimental validation of the proposed image denoising method. We first point out the suppression problem of the WNNM, and show that the proposed WSNM can remedy this to some extent. Moreover, we analysis how to make influence on the quality of the denoised image with the changing values of power \( p \) under different noise levels. Next, to illustrate the advantages of the local and nonlocal regularization, we compare the results optimized using the proposed variational model with those without using local and nonlocal constraint. Additionally, the theoretical computational complexity, convergence rate and the practical CPU time
of the proposed fast solver are given and analyzed. Finally, both qualitative and quantitative methods are used to evaluate the performance of the proposed method in comparison with several state-of-the-art methods.

Several parameters need to be set in the proposed algorithm. For WSNM, according to the analysis of the power $p$ (discussed in Section 6.1), we choose $p = \{0.7, 0.6, 0.85, 1\}$ for $\sigma_n \leq 20$, $20 < \sigma_n \leq 40$, $40 < \sigma_n \leq 60$ and $60 < \sigma_n$, respectively. Also, we empirically set patch size to $6 \times 6, 7 \times 7, 8 \times 8,$ and $9 \times 9$ for the aforementioned four cases, respectively. In practical, the iteration number $K$ is set to 12, 15, 17, and 19 respectively, on these noise levels. Throughout all the experiments, for NLTV and STV, the local patch size is set to $5 \times 5$ (the region for local similarity measurement) and the search window size is set to $21 \times 21$ (the region for searching the similar patches), and the number of best neighbors is set to 10 (the accepted similar pixels in the search window). Furthermore, in local and nonlocal regularization, the parameters $\lambda_1$ and $\lambda_2$ should satisfy the condition defined in Lemma A.1 so we choose $\lambda_1 = 0.01$ and $\lambda_2 = 0.02$. Considering the tradeoff between nonlocal regularization (NLTV) and local regularization (STV), $\mu_1$ and $\mu_2$ are chosen to be equal: $\mu_1 = \mu_2 = 0.02$. The iteration number $M$ (step 13 in Algorithm 5) is not manually assigned but depends on the difference (in the sense of MSE) between the current result and the latest one. The solver of variational model keeps optimizing until the difference is smaller than a pre-determined threshold ($\theta = 0.001$). The proposed method is implemented in Matlab with MEX, and all the experiments are performed on a standard Intel Core i7 2.8 GHz computer.

We compare the proposed method with six representative algorithms (the abbreviation of each method is in the brackets): block-matching 3D filtering [10] (BM3D), patch-based near-optimal image denoising [11] (PBNO), spatially adaptive iterative singular-value thresholding [12] (SAIST), expected patch log likelihood for image denoising [13] (EPLL), global image denoising [14] (GID), and weighted nuclear norm minimization [19] (WNNM). It is worth to note that those methods, especially the SAIST and WNNM, are the schemes in the open literature whose performance has shown convincing improvements over BM3D. Therefore, it is significative to compare with those algorithms. The denoising results of all methods are generated from the source codes or executables provided by their authors, and we keep the parameters setting mentioned in their papers for all the test images. The code and data of the proposed method is available on the website https://sites.google.com/site/yuanxiehomepage/.

### 6.1 Advantages of the Schatten $p$-norm Minimization

This subsection illustrates the advantages of the proposed weighted schatten $p$-norm minimization. Here, we use a test to point out that, weighted nuclear norm minimization suffers from a problem: the obtained singular values are suppressed which leads to shrinkage of the reconstructed data. As Fig.1 shown, both WSNM and WNNM are used to reconstruct the two patches (marked by the red and green boxes) randomly cropped from the noisy image (in Fig.1(b)). Let $\{\delta_i\}$ be the singular values of the similar patch group (step 7 in Algorithm 1), and $\{\delta_i^{(p)}\}$ be the optimal solution $X_i^{*}$ of the model (5.1) with respect power $p$. We will show solution $\{\delta_i^{(p)}\}$ in Fig.1(c) and (d) along with $\{\delta_i\}$ for patches #1 and #2, respectively. From Fig.1(c), We can see that the $\{\delta_i^{(p=1)}\}$ (i.e. the singular

| Algorithm 1: Image Denoising by WSNM and local/nonlocal consistency constraint |
|---|
| **Input:** Noisy image $y$ |
| **Output:** Denoised image $\hat{x}^K$ |
| **Initialization:** |
| 1. Initialize $\hat{x}^0 = y, \hat{y}^0 = y$ |
| **for** $k = 1 : K$ **do** |
| 2. Iterative regularization $y^{(k)} = \hat{x}^{(k-1)} + \alpha(y - x^{(k-1)})$ |
| **for each patch $y^{(k)}_j$** do |
| 3. Find similar patch group $Y_j$ |
| 4. Estimate weight vector $w$ by Eqn. 10 Singular value decomposition $[U, \Sigma, V] = SVD(Y_j)$ |
| 5. Calculate $\Delta$ by using Eqn. (4.12) |
| 6. Get the estimation: $X_j = U\Delta V^T$ |
| **end** |
| 7. Aggregate $X_j$ to form the denoised image $\hat{x}^k$ |
| 8. Set $v = \hat{x}^k$, and initial $b_n^0 = 0, b_n^0 = 0$ and $u^1 = v$ |
| 9. Iterative update $b_n^m, b_n^m, u^m$ for $m = 1, \ldots, M$ by using Eqn. (4.9) to Eqn. (4.11) |
| 10. Get the clean image $\hat{x}^k = u^M$ with local and nonlocal regularization |
| **end** |
| **Return** The final denoised image $\hat{x}^K$; |
values, denoted by cyan line) are deviated far from \( \{ \delta_i \} \) (denoted by magenta line), meaning that the suppression problem is serious. As \( p \) decreases, more high rank part of \( \{ \delta_i^p \} \) becomes zeros, while the low rank part of \( \{ \delta_i^p \} \) are more and more closer to \( \{ \delta_i \} \). For example, in Fig.1(c), for solution \( X^*_p \) \( \{ \delta_i^p \}_{i=0}^{20} \), the high rank part \( k = 14 \) is equal to zero compared with that \( k = 16 \) of \( X^*_p \) \( \{ \delta_i \}_{i=0}^{20} \), meanwhile the low rank part \( k = 1 \) is higher than that of \( X^*_p \) \( \{ \delta_i \}_{i=0}^{13} \). Therefore, it is demonstrated that the weighted nuclear norm relaxation might deviate the solution away from the real solution of original rank minimization problem. Moreover, the suppression of singular values can lead to the shrinkage effect in reconstructed data, which is illustrated in Fig.1(e) and (f). Suppose \( \{ \sigma_i \} \) is the singular values of original patch (#1 or #2), and \( \{ \sigma_i^{(p)} \} \) represents the singular values of the reconstructed patch with respect to power \( p \). As shown in Fig.1(e) and (f), the singular values \( \{ \sigma_i^{(p=1)} \} \) (denoted by cyan line) are lowest among all \( \{ \sigma_i^{(p)} \} \) with changing \( p \). Obviously, this reflects the shrinkage effect.

Additionally, it is necessary for us to analyze, at each noise level \( \sigma_n \), the influence of the changing power \( p \) upon the quality of denoised image. This is deserved for notice since it not only shows the advantages of the proposed WSNM in improving denoising quality, but also can provide a guidance for choosing optimal \( p \) for different noise levels. The PSNR results in Fig.2 show the influence on two test images House and Monarch. In this test, five noise levels \( \sigma_n = \{ 10, 20, 30, 40, 50 \} \) are used, and the power \( p \) are changing from 0.05 to 1 with interval 0.05 under each noise level. When handling these low and medium noise levels, as all the sub-figure shown in Fig.2, the optimal values for \( p \) are usually in \([0.5, 1]\). Note that the rightmost point in each sub-figure indicates the PSNR of denoised result obtained by WNNM, so we can conclude that the nonconvex relaxation of the rank minimization is superior to the traditional tightest convex relaxation. Furthermore, we also analyze the performance of different \( p \) for handling the strong noise levels \( \sigma_n = \{ 60, 70, 80, 90, 100 \} \). Fig.3 shows the PSNR results on House image, and it suggests that one should choose large \( p \) (e.g. \( p = 1 \)) under strong noise levels, due to the monotonically increasing property of all the curves.

Here, we provide a reasonable explanation for this phenomena: as pointed out in literature [12], singular values \( \{ \delta_i \} \) can be interpreted as the results of nonlocal variances estimation. With the noise level becoming stronger, the nonlocal variances of similar patches group will be increasing. However, under the low rank assumption, those nonlocal variances need to be reduced. Accordingly, the most suppressed \( \delta_i^{(p=1)} \) (e.g. the cyan lines in Fig.1(c) and (d)) indicates that nonlocal variances of similar patches group are reduced significantly, which leads to shrinkage on reconstruction of group of similar patches. Meanwhile, the stronger the noise level is, the more principal components of patch are contaminated. Therefore, the shrinkage effect can remove most of noise in those principal components when fixing the noisy bases \( U \) and \( V \) (step 9 in Algorithm [1]), although some image details might be lost.
Figure 2: The influence of changing $p$ upon denoised results under weak and medium noise levels (House and Monarch).

6.2 Analysis of Steering Total Variation

Table 1: Comparison of CPU time (in seconds) for the split Bregman and the proposed fast algorithm for differently sized images

| Image Size | 256 × 256 | 512 × 512 |
|------------|------------|------------|
| Split Bregman [30] | 6.461s | 30.112s |
| Proposed | 4.042s | 22.354s |
| Reduction | 37.44% | 25.76% |

This subsection will illustrate the advantages of local and nonlocal regularization both in improvement of visual quality and reduction of computational complexity. To this end, firstly, the visual quality of the denoised image generated by local and nonlocal constrained WSNM (namely CWSNM) will be compared with that of image recovered by WSNM alone. Fig. 4 shows the comparison of the denoised results by two methods mentioned above at two noise levels: 30 and 100. In noise level $\sigma_n = 30$, without local and nonlocal regularization, the WSNM produces more artifacts than CWSNM (Fig. 4(c) and (d)). Nevertheless, due to the image being affected by only relative weak noise, the observed improvement between two denoised images is small. However, in the case of
Figure 3: The influence of changing $p$ upon denoised results under relative strong noise levels (House).

Figure 4: Comparison of the reference images constructed by three methods

Figure 5: Comparison of PSNR and Convergence Rate
strong noise where $\sigma_n = 100$, the improvement on visual quality becomes striking as illustrated in Fig. 4(f) and (g). The stronger the noise is, the more obvious effect of artifact removal becomes. Another merit can be seen from the Fig. 5(a), which plots PSNR of all iterative steps for WSNM and CWSNM on Monarch at noise level $\sigma_n = 100$. Compare the two curves, and one can find that, CWSNM reaches a certain PSNR value more quickly. In other words, for a fixed number of iteration, we should prefer CWSNM to achieve better PSNR results.

Additionally, we analyze the computational complexity, convergence rate and CPU time of the proposed fast algorithm and the split Bregman method (the iterative optimization algorithm is presented in Section 4.1 (from Eqn. 4.3) to (4.7)). For a fixed number of iterations, both the split Bregman method and our approach are linear in $N$ (the number of pixels), since each step only contains addition and scalar multiplication operations. Therefore, we compare them in a different way by accounting for the number of atom operators in the key steps of each method, which in this case are Eqn. (4.9) to (4.11), respectively, because the complexity difference between $\text{cut}()$ and $\text{shrink}()$ is a constant. Considering one pixel access as an atom operator (ao), we can compare the complexity of the two methods in detail. $\nabla_n$, $\Delta_n$ and $\text{div}_n$ are 20 aos if we choose ten similar patterns in NLTV, similarly, $\nabla_s$, $\Delta_s$ and $\text{div}_s$ are 50 aos if the local analysis window for steering kernel is set to $5 \times 5$ pixels. Therefore, solving Eqn. (4.3) requires $212N$ aos, while Eqn. (4.11) only requires $71N$ aos. So, for a large image or a large number of optimized iterations, the proposed algorithm can significantly reduce the computational time.

Moreover, to show the improvement in reducing computational complexity, we further compare the proposed extended-ROF model solver with traditional split Bregman iteration in terms of the convergence rate. To perform comparison, we apply each algorithm to denoised image ($256 \times 256$) achieved by WSNM until the difference (based on the energy) to the last solution is below a certain threshold $\epsilon$ ($\epsilon = 0.0001$ in our test). The convergence of both methods are presented in Figure 5(b). It illustrates that the proposed method converges quickly compared with traditional Bregman iteration. Actually, our approach only use nine steps to reach convergence. Also, Table 1 compares their practical CPU seconds. The absence of PDEs leads to at least a one-quarter reduction in running time, and the proposed fast algorithm is therefore highly efficient.

6.3 Comparison with State-Of-The-Art Methods

![Figure 6: The 20 test images for image denoising.](image)

In this subsection, we evaluate the proposed method with its six counterparts on 20 widely used test images to illustrate the performance comparison. All these images are displayed in Fig. 6. Zero mean additive white Gaussian noises (with variance $\sigma_n = 20, 30, 50, 100$) are added to those test images to generate the noisy observations. Due to the limitation of space, we can only show partial denoising results in the paper, and the results of other more noise levels can be found in supplementary material.

The PSNR performance of seven competing denoising algorithms are reported in Table 2 (the highest PSNR values are marked in bold to facilitate the comparison). An overall impression observed from Table 2 is that the proposed CWSNM achieves the highest PSNR in almost all cases. In the relative low noise levels ($\sigma_n = 20, 30$), the improvement is noticeable but not striking, e.g. 0.06dB and 0.08dB improvements over the second best method WNNM in average (see the AVG row of each noise level), respectively. Then, as the noise levels are increased to 50 and 100, the improvements on WNNM become notable values 0.1dB and 0.12dB in average respectively (up to 0.24dB on image Rice with noise $\sigma_n = 50$ and 0.32dB on image Truck with $\sigma_n = 100$). It is also observed that the proposed CWSNM outperforms the benchmark BM3D method by 0.32dB~0.52dB. Such a gain in PSNR is remarkable because only a few methods can exceed BM3D more than 0.3dB in average [36, 37]. To sum up, on the average, our proposed CWSNM outperforms all other six benchmark methods at all noise levels, and the improvement becomes more significant as the noise level increases.
|                  | BM3D   | PBN0   | EPLL   | GID    | SAIST   | WNNM   | CWSNM   |
|------------------|--------|--------|--------|--------|---------|--------|---------|
| $\sigma_n = 20$  |        |        |        |        |         |        |         |
| C.Man            | 30.48  | 29.58  | 32.98  | 32.81  | 33.75   | 34.01  | 34.05   |
| House            | 33.77  | 33.56  | 32.45  | 30.75  | 30.75   | 30.75  | 30.80   |
| Peppers          | 31.29  | 30.55  | 31.17  | 31.00  | 31.53   | 31.62  | 31.03   |
| Monarch          | 30.35  | 29.55  | 30.48  | 30.16  | 30.76   | 31.13  | 28.36   |
| Airplane         | 32.53  | 32.06  | 32.41  | 32.39  | 32.82   | 32.84  | 27.56   |
| Barbara          | 31.77  | 31.06  | 29.76  | 30.21  | 30.19   | 30.19  | 30.22   |
| Boat             | 30.88  | 30.39  | 30.66  | 29.53  | 30.84   | 31.00  | 31.02   |
| Couple           | 27.26  | 27.70  | 27.49  | 26.49  | 27.31   | 27.42  | 27.54   |
| Eprint           | 30.76  | 30.22  | 30.54  | 29.28  | 30.66   | 30.82  | 30.84   |
| Fstones          | 28.80  | 27.76  | 28.28  | 27.95  | 28.99   | 29.02  | 29.09   |
| Girl             | 31.48  | 31.57  | 31.24  | 30.81  | 31.40   | 31.44  | 31.44   |
| Hill             | 30.72  | 30.32  | 30.49  | 29.59  | 30.85   | 30.83  | 30.83   |
| J.Bean           | 35.64  | 35.22  | 35.13  | 34.48  | 36.01   | 36.18  | 36.29   |
| Lena             | 33.05  | 32.75  | 32.61  | 31.74  | 33.08   | 33.12  | 33.13   |
| Man              | 30.59  | 30.15  | 30.63  | 29.59  | 30.54   | 30.74  | 30.77   |
| Parrot           | 29.96  | 29.22  | 29.97  | 28.96  | 29.97   | 30.19  | 30.21   |
| Rice             | 34.60  | 34.49  | 33.59  | 34.62  | 34.73   | 35.25  | 35.38   |
| Straw            | 27.07  | 25.86  | 26.92  | 26.63  | 27.23   | 27.44  | 27.68   |
| Truck            | 30.95  | 30.77  | 30.97  | 29.87  | 30.77   | 31.03  | 31.06   |
| AVG.             | 31.08  | 30.54  | 30.74  | 30.12  | 31.12   | 31.34  | 31.40   |

|                  | BM3D   | PBN0   | EPLL   | GID    | SAIST   | WNNM   | CWSNM   |
|------------------|--------|--------|--------|--------|---------|--------|---------|
| $\sigma_n = 100$ |        |        |        |        |         |        |         |
| C.Man            | 26.13  | 25.71  | 26.02  | 25.48  | 26.42   | 26.44  | 23.08   |
| House            | 29.69  | 29.44  | 28.76  | 27.62  | 29.99   | 30.23  | 30.36   |
| Peppers          | 26.68  | 26.46  | 26.62  | 25.60  | 26.60   | 26.81  | 26.94   |
| Monarch          | 25.81  | 25.53  | 25.77  | 24.97  | 26.09   | 26.18  | 26.30   |
| Airplane         | 25.10  | 27.77  | 27.88  | 26.91  | 28.25   | 28.44  | 28.49   |
| Barbara          | 27.22  | 26.95  | 24.82  | 25.17  | 27.49   | 27.79  | 27.83   |
| Boat             | 26.78  | 26.67  | 26.65  | 25.59  | 26.63   | 26.97  | 27.01   |
| Couple           | 25.37  | 24.39  | 23.69  | 22.88  | 23.49   | 23.73  | 23.77   |
| Eprint           | 24.52  | 24.29  | 23.59  | 23.09  | 24.54   | 24.67  | 24.73   |
| Fstones          | 25.10  | 24.86  | 24.89  | 24.26  | 25.41   | 25.44  | 25.63   |
| Girl             | 28.94  | 28.72  | 28.52  | 28.20  | 28.82   | 28.95  | 29.00   |
| Hill             | 27.19  | 27.02  | 26.95  | 25.93  | 27.04   | 27.34  | 27.36   |
| J.Bean           | 30.66  | 30.32  | 29.92  | 30.01  | 29.70   | 30.78  | 31.08   |
| Lena             | 29.05  | 28.81  | 28.42  | 27.69  | 29.01   | 29.24  | 29.28   |
| Man              | 26.80  | 26.72  | 26.72  | 25.83  | 26.67   | 26.93  | 26.98   |
| Parrot           | 25.89  | 25.37  | 25.83  | 25.76  | 26.00   | 26.10  | 26.22   |
| Rice             | 29.18  | 28.64  | 28.03  | 28.49  | 29.43   | 29.65  | 29.89   |
| Straw            | 22.40  | 22.81  | 22.00  | 21.98  | 22.65   | 22.74  | 22.93   |
| Truck            | 27.82  | 27.51  | 27.63  | 26.85  | 27.52   | 27.85  | 27.92   |
| AVG.             | 26.75  | 26.67  | 26.45  | 25.83  | 26.86   | 27.14  | 27.24   |
Figure 7: Denoising performance comparison for Barbara image (cropped patch) at noise level $\sigma_n = 50$. (a) Ground Truth. (b) Noisy Image. (c) BM3D [10], PSNR = 27.22dB. (d) PBNO [11], PSNR = 26.95dB. (e) GID [13], PSNR = 25.17dB. (f) SAIST [12], PSNR = 27.49dB. (g) WNNM [19], PSNR = 27.79dB. (h) CWSNM, PSNR = 27.83dB.

In terms of visual quality, as shown in Fig. 7, 8 and 9, our method also outperforms other state-of-the-art denoising algorithms. To facilitate the visual comparison, we display the denoised results of noisy image degraded by medium and strong levels ($\sigma_n = 50, 100$, more visual comparison results can be found in the supplementary material). In Fig. 7, we have compared the cropped portions of denoised results for test image Barbara at noise level $\sigma_n = 50$. None of the methods except CWSNM and WNNM can recover the texture behind women’s head, meanwhile WNNM produces more artifacts than the proposed method on the forehead of the women. Note that, we only select the results of the best six methods to display by using PSNR as selection criteria, and the same hereinafter. Similar results are shown in Fig. 8. Compared with CWSNM and WNNM, methods PBNO, EPLL and SAIST achieve over-smooth results in the sands area of image Boat (see Fig. 8 (d), (e) and (f)). In the highlighted red window, due to relative strong noise, the texture of brick wall (Fig. 8 (a)) cannot be recovered. The proposed CWSNM tends to get a smooth region, but WNNM generates more artifacts in the same place. All those visual improvements might be attribute to the approximation of the original low rank assumption by introducing the schatten $p$-norm. Then, as we increase the noise level to 100, it can be seen in the zoomed-in window (in Fig. 9) that, the proposed CWSNM can well construct wing veins of the butterfly, while many artifacts are produced in the reconstructed images by other methods, especially the PBNO, SAIST and WNNM. This phenomenon illustrates that the NLTV and STV based regularization can provide useful local and nonlocal constraint in the procedure of reconstruction by exploiting locally and nonlocally structural information of image. In summary, CWSNM presents strong denoising capability, producing promising denoising outputs with excellent visual quality while keeping higher PSNR indices.

7 Conclusions

In this paper, a weighted schatten $p$-norm based low rank matrix approximation (LRMA) approach has been proposed for image denoising with local and nonlocal regularization. The derived LRMA model, namely WSNM, has two major merits. On one hand, it owns a flexible property that can fit into practical applications by providing different treatments for different rank components. On the other hand, the schatten $p$-norm promotes the constructed LRMA more closer to the original low rank assumption. To further remove or at least reduce artifacts, the aggregated image, which is formed by getting all denoised patches together, is iteratively optimized using a variational model containing both local and nonlocal regularization. The proposed optimization algorithm efficiently solves this model with convergence guarantee. The final denoised output is produced by integrating WSNM and consistent constraint into an iterative restoration framework. The experimental results demonstrate that the proposed method makes impressively quantitative improvements over state-of-the-art methods, as well as preserves much
Figure 8: Denoising performance comparison for Boat image at noise level $\sigma_n = 50$. (a) Ground Truth. (b) Noisy Image. (c) BM3D [10], PSNR = 26.78dB. (d) PBNO [11], PSNR = 26.67dB. (e) EPLL [13], PSNR = 26.65dB. (f) SAIST [12], PSNR = 26.63dB. (g) WNNM [19], PSNR = 26.97dB. (h) CWSNM, PSNR = 27.01dB.

better image local structures and generates much less artifacts on visual quality.

Additionally, as it is known to all, a patch can be naturally represented as a matrix. However, current LRMA methods always need vectorized operation to change an image patch to a vector, discarding the neighboring relationship inside the patch. In contrast, tensor based representation can preserve the spatial information much better. Therefore, as a future direction, it is reasonable for us to extend the proposed WSNM to the case of low-rank tensor recovery (LRTR), although further adaptations are likely. We can also expect that there are plenty of computer vision applications (e.g., background modeling, structure from motion, MRI data recovery) that could benefit from the proposed WSNM and local-nonlocal regularization.

References

[1] G. E. Healey, R. Kondepudy. Radiometric CCD camera calibration and noise estimation. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 16, no. 3, pp. 267-276, 1994.

[2] S. G. Chang, B. Yu, M. Vetterli. Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Trans. on Image Processing, vol. 9, no. 9, pp. 1522-1531, 2000.

[3] A. Pizurica, W. Philips, I. Lamachieu, M. Acheroy. A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Trans. on Image Processing, vol. 11, no. 7, pp. 545-557, 2002.

[4] L. Zhang, B. Paul, X. Wu. Hybrid inter- and intra wavelet scale image restoration. Pattern Recognition, vol. 36, no. 8, pp. 1737-1746, 2003.

[5] J. Portilla, V. Strela, M. J. Wainwright, E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. on Image Processing, vol. 12, no. 11, pp. 1338-1351, 2003.

[6] L. Zhang, W. Dong, D. Zhang, and G. Shi. Two-stage image denoising by principal component analysis with local pixel grouping. Pattern Recognition, vol. 43, no. 4, pp. 1531-1549, 2010.

[7] P. Chatterjee, P. Milanfar. Is denoising dead, IEEE Trans. on Image Processing, vol. 19, no. 4, pp. 895-911, 2010.

[8] C. Kervrann, J. Boulanger. Optimal spatial adaptation for patch-based image denoising. IEEE Trans. on Image Processing, vol. 15, no. 10, pp. 2866-2878, 2006.
Figure 9: Denoising performance comparison for Monarch image at noise level $\sigma_n = 100$. (a) Ground Truth. (b) Noisy Image. (c) BM3D [10], PSNR = 22.51dB. (d) PBNO [11], PSNR = 22.19dB. (e) EPLL [13], PSNR = 22.23dB. (f) SAIST [12], PSNR = 22.63dB. (g) WNNM [19], PSNR = 22.95dB. (h) CWSNM, PSNR = 23.00dB.

[9] A. Buades, B. Coll, J. Morel. A non-local algorithm for image denoising. Proceedings of IEEE Computer Vision and Pattern Recognition, 2005.

[10] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. on Image Processing, vol. 16, no. 8, pp. 2080-2095, 2007.

[11] P. Chatterjee, P. Milanfar. Patch-based near-optimal image denoising. IEEE Trans. on Image Processing, vol. 21, no. 4, pp. 1635-1649, 2012.

[12] W. Dong, G. Shi, X. Li. Nonlocal image restoration with bilateral variance estimation: a low-rank approach. IEEE Trans. on Image Processing, vol. 22, no. 2, pp. 700-711, 2013.

[13] D. Zoran, Y. Weiss. From learning models of natural image patches to whole image restoration. IEEE International Conference on Computer Vision, 2011.

[14] H. Talebi, P. Milanfar. Global image denoising. IEEE Trans. on Image Processing, vol. 23, no. 2, pp. 755-768, 2014.

[15] M. Elad, M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. on Image Processing, vol. 15, no. 12, pp. 3736-3745, 2006.

[16] P. Chatterjee, P. Milanfar. Clustering-based denoising with locally learned dictionaries. IEEE Trans. on Image Processing, vol. 18, no. 7, pp. 1438-1451, 2009.

[17] J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman. Non-local sparse models for image restoration. IEEE International Conference on Computer Vision, pp. 2272-2279, 2009.
[18] W. Dong, L. Zhang, G. Shi, X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE Trans. on Image Processing*, vol. 22, no. 4, pp. 1620-1630, 2013.

[19] S. Gu, L. Zhang, W. Zuo, X. Feng. Weighted nuclear norm minimization with application to image denoising. *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2014.

[20] E. J. Candes, B. Recht. Exact matrix completion via convex optimization, *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717-772, 2009.

[21] E. J. Candes, Y. Plan. Matrix completion with noise, *In Proceedings of IEEE*, 2010.

[22] J. Cai, E. J. Candes, Z. Shen. A singular value thresholding algorithm for matrix completion, *SIAM Journal of Optimization*, vol. 20, no. 4, pp. 1956-1982, 2010.

[23] Q. Xie, D. Meng, S. Gu, L. Zhang, W. Zuo, X. Feng, and Z. Xu. On the optimal solution of weighted nuclear norm minimization. arXiv:1405.6012v1, 2014.

[24] H. Takeda, S. Farsiu, P. Milanfar. Kernel Regression for Image Processing and Reconstruction. *IEEE Trans. on Image Processing*, vol. 16, no. 2, pp. 349-366, 2007.

[25] C. Tomasi, R. Manduchi. Bilateral filtering for gray and color images. *International Conference on Computer Vision*, pp. 839-846, 1998.

[26] L. I. Rudin, S. Osher, E. Fatemi. Nonlinear total variation based noise removal algorithm. *Journal of Physics D: Applied Physics*, pp. 259–268, 1992.

[27] L. Rudin, S. Osher. Total variation based image restoration with free local constraints. *IEEE International Conference on Image Processing*, pp. 31-35, 1994.

[28] S. Farsiu, M. Robinson, M. Elad, P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. on Image Processing*, vol. 13, no. 10, pp. 1327-1344, 2004.

[29] G. Gilboa, S. Osher. Nonlocal operators with applications to image processing. *SIAM Journal on Multiscale Modeling and Simulation*, pp. 1005-1028, vol. 7, no. 3, 2008.

[30] X. Zhang, M. Burger, X. Bresson, S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, *SIAM Journal of Imaging Sciences*, vol. 3, no. 3, pp. 253-276, 2010.

[31] Y. Mao, J. Gilles. Nonrigid geometric distortions correction - Application to atmospheric turbulence stabilization. *Inverse Problems and Imaging*, vol. 6, no. 3, pp. 531-546, 2012.

[32] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, *SIAM J. Multiscale Modeling and Simulation*(2005)460–489.

[33] T. Goldstein, S. Osher, The split Bregman method for L1-regularized problems, *SIAM J. Imaging Sciences*(2009)323–343.

[34] R. Jia, H. Zhao. A fast algorithm for the total variation model of image denoising. *Advances in Computational Mathematics*, 231-241(33), 2010.

[35] L. Li, Y. Xie, W. Hu, W. Zhang. Single image super-resolution using combined total variation regularization by split bregman iteration. *Neurocomputing*, 551-560(142), 2014.

[36] A. Levin, B. Nadler. Natural image denoising: Optimality and inherent bounds. *IEEE International Conference on Computer Vision*, 2011.

[37] A. Levin, B. Nadler, F. Durand, W. T. Freeman. Patch complexity, finite pixel correlations and optimal denoising. *European Conference on Computer Vision*, 2012.

[38] M. Fazel. Matrix rank minimization with applications. Ph.D. dissertation, Stanford University, 2002.

[39] B. Recht, M. Fazel, P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Review*, vol. 52, no. 3, pp. 471-501, 2010.

[40] F. Nie, H. Huang, C. Ding. Low-rank matrix recovery via efficient schatten p-norm minimization. *AAAI Conference on Artificial Intelligence*, 2012.
A Appendix

A.1 Proof of Theorem 1

We would demonstrate that the algorithm given by (4.9), (4.10), and (4.11) has the equivalent formulation described as follows:

Let \( d_n^k = d_s^k = d_y^k = 0, b_n^k = b_s^k = b_y^k = 0, u^1 = v \). For \( k = 1, 2, \ldots \), let

\[
\begin{align*}
&d_n^k = \text{argmin}_{d_n} \{ \mu_1 \| d_n \|_1 + \frac{\lambda_1}{2} \| d_n - \nabla_n u^k - b_n^{k-1} \| \} \quad \text{(A.1)} \\
&d_s^k = \text{argmin}_{d_s} \{ \mu_2 \| d_s \|_1 + \frac{\lambda_2}{2} \| d_s - \nabla_s u^k - b_s^{k-1} \| \} \quad \text{(A.2)} \\
&b_n^k = \nabla_n u^k + b_n^{k-1} - d_n^k \quad \text{(A.3)} \\
&b_s^k = \nabla_s u^k + b_s^{k-1} - d_s^k \quad \text{(A.4)} \\
&u^{k+1} = \text{argmin}_u \left\{ \frac{1}{2} \| B(u - v) \|_2^2 - \langle B^2(u^k - v), u - u^k \rangle + \frac{\lambda_1}{2\mu_1} \| d_n^k - \nabla_n u \|_2^2 + \frac{\lambda_2}{2\mu_2} \| d_s^k - \nabla_s u \|_2^2 \right\} \quad \text{(A.5)}
\end{align*}
\]

where \( B = I + 2\lambda_1 \Delta_n + 2\lambda_2 \Delta_s \). The detailed proof of the theorem is given below.

Lemma A.1. Given \( 0 < 20\lambda_1 + 25\lambda_2 < 1 \), the real symmetric linear operator \( I + 2\lambda_1 \Delta_n + 2\lambda_2 \Delta_s \) is positive definite.

Lemma A.2. For \( k = 1, 2, \ldots \), let \( d_n^k, d_s^k, b_n^k, b_s^k \) and \( u^{k+1} \) be given by the iteration (A.1) to (A.5). Then \( \lim_{k \to \infty} (u^{k+1} - u^k) = 0 \), and

\[
b_n^k = \text{cut}(\nabla_n u^k + b_n^{k-1}, \frac{\mu_1}{\lambda_1}) \\
b_s^k = \text{cut}(\nabla_s u^k + b_s^{k-1}, \frac{\mu_2}{\lambda_2})
\]

Lemma A.3. For \( k = 1, 2, \ldots \), let \( d_n^k, d_s^k, b_n^k, b_s^k \) and \( u^{k+1} \) be given by the iteration (A.1) to (A.5). Then all the sequences \( (d_n^k)_{k=1,2,\ldots}, (d_s^k)_{k=1,2,\ldots}, (b_n^k)_{k=1,2,\ldots}, (b_s^k)_{k=1,2,\ldots} \) and \( (u^k)_{k=1,2,\ldots} \) are bounded. Moreover,

\[
u^{k+1} = v + \frac{\lambda_1}{\mu_1} \text{div}_n b_n^k + \frac{\lambda_2}{\mu_2} \text{div}_s b_s^k
\]

Lemma A.4. For \( k = 1, 2, \ldots \), let \( d_n^k, d_s^k, b_n^k, b_s^k \) and \( u^{k+1} \) be given by the iteration (A.1) to (A.5). Then

\[
\lim_{k \to \infty} (d_n^k - \nabla_n u^k) = 0 \\
\lim_{k \to \infty} (d_s^k - \nabla_s u^k) = 0
\]

The proofs of the above lemmas are given in Appendix in supplemental material.

Theorem 1. For \( k = 1, 2, \ldots \), let \( d_n^k, d_s^k, b_n^k, b_s^k \) and \( u^{k+1} \) be given by the iteration (A.1) to (A.5). If \( 0 < 20\lambda_1 + 25\lambda_2 < 1 \), then \( \lim_{k \to \infty} u^k = u^* \).

Proof. Let \( F(u) := (1/2) \| u - v \|_2^2 \), then \( \partial F(u) = u - v \). For \( w \in \mathbb{R}^N \) we can get

\[
F(u^{k+1} + w) - F(u^{k+1}) - (u^{k+1} - v) \geq 0
\]
From Lemma A.3, we have 

\[-(u^{k+1} - v) = -\frac{\lambda_1}{\mu_1} \text{div}_n b_n^k - \frac{\lambda_2}{\mu_2} \text{div}_s b_s^k, \text{ and moreover } \langle \text{div}_n p, q \rangle = -\langle p, \nabla_n q \rangle \text{ (similar to STV).} \]

Then,

\[F(u^{k+1} + w) - F(u^{k+1}) + \langle \lambda_1 b_n^k, \nabla_n w \rangle + \langle \lambda_2 b_s^k, \nabla_s w \rangle \geq 0 \quad (A.7)\]

Recall that \(G(d) = \|d\|_1, \frac{\lambda_1}{\mu_1} b_n^k \in \partial G(d_n^k), \frac{\lambda_2}{\mu_2} b_s^k \in \partial G(d_s^k)\), so,

\[\|d_n^k + \nabla_n w\|_1 - \|d_n^k\|_1 - \langle \frac{\lambda_1}{\mu_1} b_n^k, \nabla_n w \rangle \geq 0 \quad (A.8)\]

\[\|d_s^k + \nabla_s w\|_1 - \|d_s^k\|_1 - \langle \frac{\lambda_2}{\mu_2} b_s^k, \nabla_s w \rangle \geq 0 \quad (A.9)\]

Adding \((A.7) \sim (A.9)\) gives

\[\|d_n^k\|_1 + \|d_s^k\|_1 + F(u^{k+1}) \leq \|d_n^k + \nabla_n w\|_1 + \|d_s^k + \nabla_s w\|_1 + F(u^{k+1} + w) \quad (A.10)\]

Suppose that \((k_j)_{j=1,2,...}\) is an increasing sequence of positive integers such that the sequence \((u^{k_j})_{j=1,2,...}\) converges to the limit \(\tilde{u}\). By Lemma A.2, we have \(\lim_{k \to \infty} (u^{k+1} - u^k) = 0\). Therefore, \(\lim_{j \to \infty} u^{k_j+1} = \tilde{u}\). Moreover, we have the following via the Lemma A.4

\[\lim_{j \to \infty} d_n^k = \lim_{j \to \infty} [\langle d_n^k - \nabla_n u^k \rangle + \nabla_n u^k] = \nabla_n \tilde{u} \quad (A.11)\]

also have

\[\lim_{j \to \infty} d_s^k = \lim_{j \to \infty} [\langle d_s^k - \nabla_s u^k \rangle + \nabla_s u^k] = \nabla_s \tilde{u} \quad (A.12)\]

Replacing \(k\) by \(k_j\) in \((A.10)\) and let \(j \to \infty\), we have:

\[\|\nabla_n \tilde{u}\|_1 + \|\nabla_s \tilde{u}\|_1 + F(\tilde{u}) \leq \|\nabla_n (\tilde{u} + w)\|_1 + \|\nabla_s (\tilde{u} + w)\|_1 + F(\tilde{u} + w) \]

The above equations hold for all the \(w \in \mathbb{R}^N\). On the other hand, \(u^*\) is the unique solution to the minimization problem \((4.2)\). Therefore, we must have \(\tilde{u} = u^*\). Since \((u^k)_{k=1,2,...}\) is a bounded sequence, we have

\[\lim_{k \to \infty} u^k = u^* \quad (A.13)\]

This completes the proof of the Main Theorem 4.1.