Stable Sparse Model with Non-Tight Frame

Min Zhang, Yunhui Shi, Na Qi and Baocai Yin

1 Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; lengyuewuyan@yeah.net (M.Z.); qina@bjut.edu.cn (N.Q.); ybc@bjut.edu.cn (B.Y.)
2 Computer Science and Technology, Dalian University of Technology, Dalian 116023, China

Received: 17 January 2020; Accepted: 25 February 2020; Published: 4 March 2020

Abstract: Overcomplete representation is attracting interest in image restoration due to its potential to generate sparse representations of signals. However, the problem of seeking sparse representation must be unstable in the presence of noise. Restricted Isometry Property (RIP), playing a crucial role in providing stable sparse representation, has been ignored in the existing sparse models as it is hard to integrate into the conventional sparse models as a regularizer. In this paper, we propose a stable sparse model with non-tight frame (SSM-NTF) via applying the corresponding frame condition to approximate RIP. Our SSM-NTF model takes into account the advantage of the traditional sparse model, and meanwhile contains RIP and closed-form expression of sparse coefficients which ensure stable recovery. Moreover, benefitting from the pair-wise of the non-tight frame (the original frame and its dual frame), our SSM-NTF model combines a synthesis sparse system and an analysis sparse system. By enforcing the frame bounds and applying a second-order truncated series to approximate the inverse frame operator, we formulate a dictionary pair (frame pair) learning model along with a two-phase iterative algorithm. Extensive experimental results on image restoration tasks such as denoising, super resolution and inpainting show that our proposed SSM-NTF achieves superior recovery performance in terms of both subjective and objective quality.

Keywords: sparse dictionary; stable recovery; frame; RIP

1. Introduction

Sparse representation of signals in dictionary domains has been widely studied and has provided promising performance in numerous signal processing tasks such as image denoising [1–5], super resolution [6–8], inpainting [9,10] and compression [11,12]. It is well known that images are represented by a linear combination of certain atoms of a dictionary. Overcomplete sparse representation is the overcomplete system with a sparse constraint. Common overcomplete systems differ from the traditional bases, such as DCT, DFT and Wavelet, because they offer a wider range of generating elements; potentially, this wider range allows more flexibility and effectiveness in signal sparse representation. However, it is a severely under-constrained illposed problem to find the underlying overcomplete representation due to the redundancy of the systems. When the underlying representation is sparse and the overcomplete systems have stable properties, the ill-posedness will disappear [13]. Sparse models are generally classified into two categories: Synthesis sparse models and analysis sparse model [14]. The commonly referred to sparse models are synthesis sparse models. The analysis ones characterize the signal by multiplying it with an analysis overcomplete dictionary, leading to a sparse outcome. A variety of effective sparse models have been investigated and established such as the classical synthesis sparse model [9,15], the classical analysis sparse model [14], the nonlocal sparse model [16,17] and the 2D sparse model [18]. Unfortunately, these models ignore the stability...
recovery property which claims that once a sufficient sparse solution is found, all alternative solutions necessarily reside very close to it [9]. Recently, the stable recovery of sparse representation has drawn attention in signal processing theory. Generally speaking, stable recovery can be guaranteed by two properties: Sufficient sparsity and a favorable structure of the dictionary [19]. Donoho defines the concept of mutual incoherence of the dictionary and applies it to prove some possibility of stable recovery [19]. The authors of [20] propose a sparsity-based orthogonal dictionary learning method to minimize the mutual incoherence. The authors of [21] propose an incoherent dictionary learning scheme by integrating a low rank gram matrix of the dictionary into the dictionary learning model.

A more powerful stable recovery guarantee developed by Candes and Tao, termed Restricted Isometry Property (RIP), makes consequent analysis easy [22]. A matrix $\Phi$ is said to satisfy the RIP of order $k$ if there exists a constant $\delta_k \in (0, 1)$ such that

$$
(1 - \delta_k)\|y\|_2^2 \leq \|\Phi y\|_2^2 \leq (1 + \delta_k)\|y\|_2^2
$$

(1)

holds for all $k$-sparse vectors $y$. $\delta_k$ is defined as the smallest constant which satisfies the above inequalities and is called the restricted isometry constant of $\Phi$.

Most RIP research substantially investigates applying RIP as a stability analysis instrument [17,23,24] or finding optimal RIP constant [25,26] which are all theoretical analyses rather than practical applications. According to the research of [21], the intrinsic property of a dictionary has a direct influence on its performance. All familiar algorithms are staggeringly unstable with a coherent or degenerate dictionary [19]. Recognizing the gap between theoretical analyses and practical applications of RIP, this paper aims to build a stable sparse model satisfying the RIP.

Recently, the frame as a stable overcomplete system has drawn some attention in signal processing as the given signal can be represented by its canonical expansion in a manner similar to conventional bases under the frame. Some data-driven approaches are proposed in [1,27–30]. The authors of [27,29,30] utilize redundant tight frame in compressed sensing and [28] applies tight frame to few-view image reconstruction. Study [1] presents a data-driven method that the dictionary atoms associated with the tight frame are generated by filters. These approaches achieve much better image processing performance than previous methods, and meanwhile the tight frame condition which requires the frame almost-orthogonality will limit the flexibility in sparse representation. Study [31] derives stable recovery result for $l_1$-analysis minimization in redundant, possibly non-tight frames. Inspired by this result and the relationship between RIP and frame, we aim to establish a stable sparse model with RIP based on non-tight frame.

We call a sequence $\{\phi_i\}_{i=1}^M \in \mathbf{H}$ a frame if and only if there exist two positive numbers $A$ and $B$ such that

$$
A\|x\|_2^2 \leq \sum_{i=1}^M |\langle x, \phi_i \rangle|^2 \leq B\|x\|_2^2 \quad \forall x \in \mathbf{H}^N
$$

(2)

Here, $A$ and $B$ are called the bounds of the frame. We find that every submatrix $\Phi_k$ satisfied RIP is a non-tight frame with $(1 - \delta_k)$ and $(1 + \delta_k)$ as its frame bounds with a given $k$. Obviously, there is an essential connection between the non-tight frame and the RIP.

In this paper we focus on a stable sparse model and more specifically on the development of an algorithm that would learn a pair of non-tight frame based dictionaries from a set of signal examples. We propose a stable sparse model via applying the non-tight frame condition to approximate the RIP. This model shares the favorite overcomplete structures with the common sparse models, and meanwhile it contains RIP and closed-form sparse coefficient expression which ensure stable recovery. Recognizing that the optimal framebounds are essentially the maximum and minimum singular values of the frame, RIP is actually enforced on the dictionary pair (the frame and its dual frame) by constraining the singular values of them. We also formulate a dictionary pair learning model via applying the second-order truncated Taylor series to approximate the inverse frame operator.
Then we present an efficient algorithm to learn the dictionary pair via a two-phase iterative approach. To summarize, this paper makes the following contributions:

1. We propose a stable sparse model along with a dictionary pair learning model. Non-tight frame condition is utilized to develop a relaxation of RIP to guarantee stable recovery of sparse representation. Moreover, the sparse coefficients are also modeled, which leads to a more stable recovery especially for seriously noisy image.

2. It is nearly impossible to solve the dictionary pair learning model in a straightforward way since the inverse frame operator is involved. We provide an effective way to modify the model via applying a second-order truncated Taylor series to approximate the inverse frame operator, and provide an efficient algorithm for the modified one.

3. We present the stability analysis of the proposed model and demonstrate it on natural and synthetic image denoising, super resolution and image inpainting. The denoising results show that the proposed approach outperforms synthesis models such as the KSVD and the data-driven tight frame based methods for natural image case in terms of average PSNR. Moreover, it also gains comparable performance to the Analysis KSVD for a piecewise-constant (PWC) image in terms of average PSNR. The meaningful structures in the trained dictionary pairs for natural images and a PWC image are observed. The super resolution results show that the SSM-NTF produces better performance than the Bicubic interpolation method and the method in [32]. The inpainting results show that our model is able to eliminate text of fonts completely.

This paper is organized as follows: Section 2 reviews the related work on frame, synthesis sparse model and analysis sparse model. Section 3 presents our stable sparse model with non-tight frame SSM-NTF along with a dictionary pair learning model. Section 4 proposes the corresponding dictionary pair learning algorithm. Section 5 proposes the image restoration method of our proposed SSM-NTF model. In Section 6 we analyze the computational complexity of our proposed algorithm. In Section 7, we demonstrate the the effectiveness of our SSM-NTF model by analyzing the convergence of the corresponding algorithm, denoising natural and piecewise constant images, super resolution and image inpainting. Finally, Section 8 concludes this paper.

2. Related Work

In this section, we briefly review the related work on frame, synthesis sparse model and analysis sparse model.

Frame: A frame $\Phi$ is called a tight frame if the frame bounds are equal in the Equation (2) [32]. There are two associated operators can be defined between the Hilbert space $H^N$ and Square Integrable Space $L^2_M(\cdot)$ once a frame is defined. One is the analysis operator $T$ defined by

$$ (Tx)_i = \langle x, \phi_i \rangle, \quad \forall x \in H^N $$ (3)

and the other is its adjoint operator $T^*$ which is called the synthesis operator

$$ T^* c = \sum_{i=1}^{M} c_i \phi_i \quad \forall c = (c_i)_{i \in J} \in L^2_M(T) $$ (4)

then, the frame operator can be defined as the following canonical expansion

$$ Fx = T^* Tx = \sum_{i=1}^{M} \langle x, \phi_i \rangle \phi_i $$ (5)
In Euclidean space, a given frame $\Phi$ can be represented in manner of matrix with its columns of it as the frame elements. Then one of its adjoint operator can be represented as $\psi_i = F^{-1} \phi_i$ [32]. Let $x \in \mathbb{R}^N$ be an arbitrary vector, a reconstruction function can be expressed as the following form

$$x = \sum_{i=1}^{M} < x, \psi_i > \phi_i$$

(6)

Synthesis sparse model: The conventional synthesis sparse model represents a vector $x$ by the linear combination of a few atoms from a large dictionary $\Phi$, denoted as $x = \Phi y$, $\|y\|_0 \leq L$, where $L$ is the sparsity of $y$. The computational techniques for approximating sparse coefficient $y$ under a given dictionary $\Phi$ and $x$ includes greedy pursuit (e.g., OMP [9]) and convex relaxation optimization, such as Lasso [33] and FISTA [8]. In order to improve the performance of sparse representation, some modified models such as the nonlocal sparse model [16], the frame based sparse model [21], and the MD sparse model [18] are also investigated.

Analysis sparse model: The analysis sparse model is defined as: $y = \Omega x$, $\|y\|_0 = p - l$ where $\Omega \in \mathbb{R}^{p \times d}$ is a linear operator (also called as a dictionary), and $l$ denotes the co-sparsity of the signal $x$. The analysis representative vector $y$ is sparse with $l$ zeros. The zeros in $y$ denote the low-dimensional subspace to which the signal $x$ belongs. The analysis sparse coding [14] and dictionary learning [34] approach are also been proposed.

However, all these models ignore the stability recovery property which provides stable reconstruction of the signals in presence of noise.

Dictionary learning methods: The dictionaries include analytical dictionaries, such as DCT, DWT, curvelets and contourlets and learned dictionaries. Some dictionary learning method are proposed, such as the classical KSVD [9] algorithm, the efficient sparse coding which convert the original dictionary learning problem to two least squares problem by applying the Lagrange dual [3], the non-local sparse model [16] which learns a set of PCA sub-dictionaries by cluster the samples into K clusters using image nonlocal self-similarity prior and its improved version which using the $l_q$-norm to instead the $l_2$-norm in order to handle different image contents. With the realization of stability, some mutual-coherence based methods are proposed. In [20] a sparsity-based orthogonal dictionary learning method is proposed to minimize the mutual-coherence of the dictionary. The authors of [21] propose an in coherent dictionary learning scheme by integrating a low rank gram matrix of the dictionary into the dictionary learning model. However, these methods only concern the capability of the dictionary without modeling the sparse coefficients which still has some probability of instability.

3. The Proposed SSM-NTF

In this section, we present the stable sparse model with non-tight frame, (Section 3.1), the stability analysis of the proposed model, (Section 3.2) and the dictionary pair (the frame pair) learning model, (Section 3.3).

3.1. Stable Sparse Model with Non-Tight Frame

In this section, we derive our stable sparse model with non-tight frame where the non-tight frame condition serves as an approximation to the RIP.

According to [35], a $k$-th RIP constant can be express as

$$\delta_k(\Phi) = \frac{\Gamma_k(\Phi) - 1}{\Gamma_k(\Phi) + 1}$$

(7)

where

$$\Gamma_k(\Phi) = \frac{\rho_{\text{max}}}{\rho_{\text{min}}}$$

(8)
\[ \theta^k_{\max} = \max_{\|y\|_0 = k} \frac{\|\Phi y\|_2^2}{\|y\|_2^2} \quad \theta^k_{\min} = \min_{\|y\|_0 = k} \frac{\|\Phi y\|_2^2}{\|y\|_2^2} \] (9)

The Equation (7) provides a new perspective in integrating the RIP to sparse model via applying \( \theta^k_{\max} \) and \( \theta^k_{\min} \) instead of the RIP constant \( \delta_k(\Phi) \). The difficulty in building a stable sparse model decreases. However, the sparsity \( k \) varies with the noise level, and also, in a feasible numerical calculation method, it is impossible to sweep through all the samples satisfying \( |x|_0 = k \) to pursue an unknown dictionary \( \Phi \).

Let \( x \) be a signal vector, the frame reconstruction function can be formulated as \( x = \Phi \Psi^T x \) where \( \Psi \) is a dual frame of \( \Phi \). Adding a reasonable sparsity prior to the signal \( x \) over \( \Psi \) domain, we can derive

\[ \frac{\|\Phi(\Psi^T x)\|_2^2}{\|\Psi^T x\|_2^2} = \frac{\|x\|_2^2}{\|\Psi^T x\|_2^2} = \frac{1}{\|\Psi^T x\|_2^2} \] (10)

Denoting the optimal frame bounds of \( \Phi \) as \( A \) and \( B \), the frame condition of \( \Psi \) can be formulated as \( \frac{1}{A} \leq \frac{\|\Psi^T x\|_2^2}{\|x\|_2^2} \leq \frac{1}{B} \). Then a pair of bounds for Equation (10) can be obtained as \( A \leq \frac{\|\Phi(\Psi^T x)\|_2^2}{\|\Psi^T x\|_2^2} \leq B \). A formula similar to Equation (9) is derived as \( B = \max_{x \in \mathcal{J}} \frac{\|\Phi(\Psi^T x)\|_2^2}{\|\Psi^T x\|_2^2}, \quad A = \min_{x \in \mathcal{J}} \frac{\|\Phi(\Psi^T x)\|_2^2}{\|\Psi^T x\|_2^2} \) where \( \mathcal{J} \) is the data set. Imitating Equation (7), we can obtain a RIP-like constant expression

\[ \delta(\Phi) = \frac{\hat{\Gamma}(\Phi) - 1}{\hat{\Gamma}(\Phi) + 1} \] (11)

where \( \hat{\Gamma}(\Phi) = \frac{B}{A} \). Obviously, \( \delta(\Phi) \) can be regarded as an approximation of the RIP constant which benefits the computation due to the ignorance on sparsity degree. In a word, the RIP constraint can be satisfied by constraining the frame bounds. Thus, a stable overcomplete system with a sparsity prior can be established.

Now we discuss the characteristic of the frame bounds \( A \) and \( B \). The Frame Condition (2) has a more compact form \( \sqrt{A} \leq \eta_0 \leq \sqrt{B} \) where \( \eta_0 \) denotes any singular value of \( \Phi \). More specifically, \( \sqrt{A} = \eta_{\min}, \sqrt{B} = \eta_{\max} \) where \( \eta_{\max} \) and \( \eta_{\min} \) denote the maximum and minimum singular values of \( \Phi \), respectively. Then, we can obtain \( \eta_{\min} \geq \theta^k_{\min}, \eta_{\max} \leq \theta^k_{\max} \). It is easy to know that \( \delta(\Phi) \geq \delta_k(\Phi) \). Obviously, \( \delta(\Phi) \) is a reasonable relaxation of \( \delta_k(\Phi) \) as \( \delta(\Phi) \) is slightly exceed \( \delta_k(\Phi) \) but resides very close to it as long as the data is not seriously degraded. Therefore, the RIP constraint can be enforced on the frames by limiting the maximum and minimum singular values.

In this paper, we integrate non-frame to traditional sparse model to establish a stable sparse model with RIP. Let \( x \) be a signal vector. Under the assumption of the sparsity prior of \( \Psi^T x \), we apply a soft thresholding operator \( S_\lambda(\cdot) \) (which shall be defined in the next subsection) on it such that

\[ x = \Phi S_\lambda(\Psi^T x) \] (12)

where \( \lambda \) is a vector with elements \( \lambda_i \) corresponding to \( \psi_i, \ i = 1, 2, \ldots, M \). Therefore, we propose the stable sparse model with non-tight frame (SSM-NTF) as follows

\[ y = S_\lambda(\Psi^T x), \quad x = \Phi y, \] (13)

s.t. \( \|y\|_0 \leq s \).

Here, the correlation between the frame \( \Phi \) and its dual frame \( \Psi \) is formulated as \( \Psi = \Phi^{-1} \Phi \). The frame operator \( F \) is formulated as \( \Phi \Phi^T \) which is indeed a gram matrix of \( \Phi \). The singular values of \( \Phi \) are constrained by \( \sqrt{A} \leq \eta_0 \leq \sqrt{B} \) to satisfy the RIP. Actually, by constraining the singular values of \( \Phi \), the elements of the gram matrix are also bounded which meets the theory of mutual coherence.

In order to be consistent with the traditional sparse models, we refer to the frame \( \Phi \) and its dual frame \( \Psi \) as dictionary and its dual dictionary.
3.2. The Stability Analysis of the Proposed Model

In sparse representation problem, a given noiseless signal \( x \), can be formulated as

\[
(P_0) : \min_x \| y \|_0 \quad \text{s.t.} \quad x = \Phi y
\]

(14)

where \( \Phi \) is the sparse representation dictionary and \( y \) is the sparse coefficients. While \( x = \Phi y \) is an underdetermined linear system, the problem \( (P_0) \) has the unique solution \( y_0 \) as soon as it satisfies the uniqueness property which is formulated as

\[
\| y \|_0 < \frac{1}{2}(1 + 1/\mu)
\]

(15)

where \( \mu \) is the mutual-coherence of \( \Phi \) [9]. However, the signals are usually acquired with noise, then the problem \( (P_0) \) should be relaxed to the problem \( (P_\epsilon) \) which is expressed as

\[
(P_\epsilon) : \min_x \| y \|_0 \quad \text{s.t.} \quad \| x - \Phi y \|_F^2 \leq \epsilon
\]

(16)

where \( \epsilon \) is an error-tolerant which exists due to the noise. The problem \( (P_\epsilon) \) will no longer maintain the uniqueness of solution as \( x = \Phi y + \epsilon \) is an inequality system. Thus, the notion of Uniqueness Property (15) is replaced by the notion of stability which claims that all the alternative solutions reside very close to the ideal solution. Under the stable guarantee, we can yet ensure that the recovery results of our methods produce meaningful solutions. Assume that \( y_0 \) is the ideal solution to the model \( (P_\epsilon) \) and \( \hat{y} \) is the candidate one, the traditional sparse model has a stability claim of the form [9]

\[
\| \hat{y} - y_0 \|_2^2 \leq \frac{4\epsilon^2}{1 - (2s_0 - 1)\mu}
\]

(17)

where \( \mu \) is the mutual coherence which is formulated as \( \mu = \max_{i \neq j} | \langle \phi_i, \phi_j \rangle |, i, j = 1, 2, \cdots, M \). Apparently, the error bound of Equation (17) can only be determined with given sparsity \( s_0 \) and the mutual coherence \( \mu \). However, the mutual coherence of an unknown dictionary is very difficult to calculate which lead to a result that we can not ensure the stability in the dictionary learning case. In contrast, we derive a similar stability claim of our proposed SSM-NTF model.

Defining \( d = \hat{y} - y_0 \) with \( y_0 \) as the ideal solution to the model, we have that \( \| \Phi \hat{y} - \Phi y_0 \|_2 = \| \Phi d \|_2 \leq 2\epsilon \). From the previous subsection, we have know that the frame \( \Phi \) satisfies the RIP with the corresponding parameter \( \delta(\Phi) \). Thus, using this property and exploiting the lower-bound part in Equation (1), we get

\[
(1 - \delta(\Phi)) \| d \|_2^2 \leq \| \Phi d \|_2^2 \leq 4\epsilon^2
\]

(18)

where \( \delta(\Phi) = \gamma(\Phi) - 1 = \frac{2s_0 - 1}{2 + 1} \). Thus, we get a stability claim of the form

\[
\| d \|_2^2 = \| \hat{y} - y_0 \|_2^2 \leq \frac{4\epsilon^2}{1 - \frac{2s_0 - 1}{2 + 1}}
\]

(19)

Obviously, the error bound of the SSM-NTF is determined by \( \frac{2}{2} \), the ratio of the upper bound to the lower bound of the frame, rather than the specific values of \( A \) and \( B \). Thus, for the convenience of numerical experiments, we usually set \( A \) to a fixed value. A main advantage of standard orthogonal transformations is that they maintain the energy of the signals in the transform domain as its frame bounds \( A \) and \( B \) are equal to 1. However, the standard orthogonal basis is non-redundant that limits its performance in sparse representation. In order to make a trade off between the represent accuracy and the degree of redundant, we usually set the lower frame bound \( A \) to a value a little smaller than 1 but not over-small as \( A \) is the minimum singular value of \( \Phi \) which determines the condition
number of $\Phi$. Thus, once the tolerance error is given, the value of $B$ can be easily calculated. Further, a pair of dictionaries conform to the given error can be obtained using the proposed SSM-NTF model. On the other hand, if the value of $B$ is given by experience, the error bound of our model can be measured.

### 3.3. Learning Model of Dictionary Pair

Assuming $X \in \mathbb{R}^{N \times L}$ is the training data with signal vectors $x_i \in \mathbb{R}^N$, $i = 1, 2, \ldots, L$, as its columns. The dictionary pair learning model can be written as

$$\min_{\Phi, \Psi, \lambda, Y} \|X - \Phi Y\|_F^2 + \gamma_1 \|Y - S_\lambda (\Psi^T X)\|_F^2 + \gamma_2 \|Y\|_0 + \gamma_3 \|Y - F^{-1} \Phi\|_F^2$$

s.t. $\sqrt{A} \leq \eta_\Phi \leq \sqrt{B}$

(20)

However, the Problem (20) is difficult to solve. First, the inverse of the frame operator $F$ has no closed-form explicit expression. Secondly, the thresholding operator is a highly nonlinear operator which makes the optimization with respect to $\lambda$ hard to optimize.

Apparently, the Problem (20) is difficult to solve as the existence of the inverse of $F$. Fortunately, the matrix $F^{-1}$ can be expressed as a convergent series [36] which is formulated as

$$F^{-1} = \frac{2}{A + B} \sum_{k=0}^{\infty} (I - \frac{2F}{A + B})^k$$

(21)

Here, we truncated the series at $k = 1$ to make a tradeoff between computational complexity and approximation accuracy. It is formulated as

$$F^{-1} \approx \frac{2}{A + B} + \frac{2}{A + B} (I - \frac{2F}{A + B}) = \frac{2}{A + B} (2I - \frac{2F}{A + B})$$

(22)

In this way, once the frame bounds are given, the inverse of $F$ can be calculated easily. Then the optimization problem for training RIP-dictionary pair is formulated as

$$\min_{\Phi, \Psi, \lambda, Y} \|X - \Phi Y\|_F^2 + \gamma_1 \|Y - S_\lambda (\Psi^T X)\|_F^2 + \gamma_2 \|Y\|_0 + \gamma_3 \|Y - \frac{2}{A + B} (2I - \frac{2F}{A + B}) \Phi\|_F^2$$

s.t. $\sqrt{A} \leq \sigma_\Phi \leq \sqrt{B}$

(23)

where $S_\lambda (\cdot)$ is the elementwise thresholding operator. There are two basic thresholding methods: The hard thresholding method whose thresholding operator defines as $S_\lambda (\cdot) \rightarrow \max(\cdot - \lambda, 0)$ and the soft thresholding whose operator is defined as $S_\lambda (\cdot) \rightarrow \text{sgn}(\cdot) \max(\cdot - \lambda, 0)$. Both of the two operator are are non-convex and highly discontinuous which lead to big challenges to solve Problem (23). The mean reason is the fact that the update of the thresholding values $\lambda$ causing non-smooth changes to the cost function. To solve this difficulty, we design an alternative direction method via global search and least square that will be introduce in Section 4.1.

### 4. Dictionary Pair Learning Algorithm

In this subsection, we propose the two-phase iterative algorithm for dictionary pair learning by dividing Problem (23) into two subproblems: The sparse coding phase which updates the sparse coefficients $Y$ and thresholding values $\lambda$, and the dictionary pair update phase which computes $\Phi$ and $\Psi$.

#### 4.1. Sparse coding phase

In this subsection, we discuss how to calculate the sparse coefficients $Y$ and the threshold values $\lambda$ with given $\Phi$ and $\Psi$ under our SSM-NTF model.
Given a pair of dictionaries \( \Phi \) and \( \Psi \), calculating \( Y \) and \( \lambda \) from \( X \) is formulated as:

\[
\{ \hat{Y}, \hat{\lambda} \} = \min_{Y, \lambda} \|X - \Phi Y\|_F^2 + \gamma_1 \|Y - S_\lambda(\Psi^T X)\|_F^2 + \gamma_2 \|Y\|_0
\]  
(24)

We pursue the two variables alternatively. Firstly, with fixed \( \lambda \), we obtain the sparse coefficients \( Y \) by solving Problem (24) through OMP [9] as it can be easily convert to the classical synthesis sparse expression \( \min \|Z - DY\|_F^2 \) s.t. \( \|Y\|_0 \) where \( Z = [X \sqrt{\tau_1} S_\lambda(\Psi^T X)] \) and \( D = [\Phi \sqrt{\tau_1} I] \).

Secondly, the pursuit of \( \lambda \) is equivalent to solving the following problem

\[
\hat{\lambda} = \arg \min_\lambda \|Y - S_\lambda(\Psi^T X)\|_F^2
\]  
(25)

which can be decomposed into \( M \) individual optimization problems

\[
\hat{\lambda}_i = \arg \min_{\lambda_i} \|Y_i - S_{\lambda_i}(\psi^T_i X)\|_2^2, \quad i = 1, \ldots, M.
\]  
(26)

where \( \psi_i \) is the column of \( \Psi \). From the definition of soft thresholding operator, we can know that the function of Problem (26) is discrete. By denoting the data indices set that remains intact after the thresholding as \( J \), we split the data \( X \) into two parts: \( X_{\bar{J}} \) and \( X_{\tilde{J}} \) such that

\[
S_{\lambda_i}(\psi_i^T X_{\bar{J}}) = \psi_i^T X_{\bar{J}} - \text{sgn}(\psi_i^T X_{\tilde{J}})\lambda_i
\]  
(27)

\[
S_{\lambda_i}(\psi_i^T X_{\tilde{J}}) = 0.
\]  
(28)

where \( \tilde{J} \) is a supplementary to the intact indices \( J \) which turn the all elements to zero. It is clear to know that the variables \( J_i \) and \( \tilde{J}_i \) are both functions of \( \lambda_i \) without explicit expressions which leads to a large challenge in optimization.

In order to solve Problem (26), an intermediate variable \( \mu_i \) is necessarily to introduced to separate the whole problem into two parts: The update of the indices \( J_i \) and \( \tilde{J}_i \) (determined by \( \mu_i \)) and the update of the explicit thresholding value \( \lambda_i \). Then Problem (26) can be transformed to another optimization problem:

\[
\{ \hat{\lambda}_i, \mu_i \} = \arg \min_{\lambda_i, \mu_i} \|Y_i^{\bar{J}}\|_2^2 + \|Y_i^{\tilde{J}} - (\psi_i^T X_{\tilde{J}} - \text{sgn}(\psi_i^T X_{\tilde{J}})\lambda_i)\|_2^2 + \tau/2 \|\lambda_i - \mu_i\|_2^2
\]  
(29)

where \( \bar{J}_i \) and \( \tilde{J}_i \) are two functions of the intermediate variable \( \mu_i \).

At the \( k \)-th step, to obtain \( \mu_i \), we solve Problem (29) with \( \lambda_i \) fixed and denote the functions as \( f(\mu_i) + g(\mu_i) + l(\mu_i) \) where \( f(\mu_i) = \|Y_i^{\bar{J}}\|_2^2 \), \( g(\mu_i) = \|Y_i^{\tilde{J}} - (\psi_i^T X_{\tilde{J}} - \text{sgn}(\psi_i^T X_{\tilde{J}})\lambda_i)\|_2^2 \), \( l(\mu_i) = \tau/2 \|\lambda_i - \mu_i\|_2^2 \). Optimizing this expression is obviously non-trivial as the target function is non-convex and highly discontinuous. Actually, with \( \lambda_i \) fixed, the minimization of \( f(\mu_i) + g(\mu_i) \) can be globally solved due to its discrete finite nature. In another word, if a series of candidate terms of \( \mu_i \) are given, the global search is guaranteed to succeed.

Once a \( \lambda_i \) is given, the \( f(\mu_i) + g(\mu_i) \) will be a piecewise constant function. It means that the function values remain unchanged within a series of intervals which are determined by \( |\psi_i^T x_i|, i = i_1, i_2, \ldots, i_L \). Therefore, \( |\psi_i^T x_i|, i = 1, 2, \ldots, L \) can be taken as a portion of candidate terms of \( \mu_i \). For the function \( l(\mu_i) \), it is clear that it minimizes at \( \mu_i = \lambda_i \) and monotonically increases with the increasing distance between \( l(\mu_i) \) and the given \( l(\lambda_i) \). So, to minimize \( l(\mu_i) \), we only need to choose the closest point in the feasible region.
Without loss of generality, we assume that all the $|\psi_i^T x_i|, i = 1, 2, \ldots, L$ are ascending ordered and the corresponding signals are in the same order. We compute all the possible values of $f(\mu_i) + g(\mu_i)$ by

\[
\begin{align*}
f(\psi_i^T X_{j+1}) &= f(\psi_i^T x_i) + y_i^2(j+1) \\
g(\psi_i^T X_{j+1}) &= ||y_i - \psi_i^T x + sgn(\psi_i^T x)\mu_i||^2 - \beta_{k+1}
\end{align*}
\]  

(30)

where $\beta_{k+1} = \beta_k + [y_i(k+1) - \psi_i^T x_{k+1} + sgn(\psi_i^T x_{k+1})\mu_i]^2$, $\beta_1 = [y_i - \psi_i^T x_1 + sgn(\psi_i^T x_1)\mu_i]^2$.

Sort $|\psi_i^T x_i|, i = 1, 2, \ldots, L$ in descending order of $f(\mu_i) + g(\mu_i)$, and every two adjacent values form an interval on which the function value remains unchanged. In another word, the objective function $f(\mu_i) + g(\mu_i) + l(\mu_i)$ is minimized at the point closest to $\lambda_i$ in the interval. Thus, compute all the minimizer values on every interval and the minimum must be the optimal result.

With $\mu_i$ fixed, we solve the following problem in order to pursue $\lambda_i$:

\[
\{\lambda_i\} = \arg\min_{\lambda_i} ||y_j - [\psi_i^T x_j - sgn(\psi_i^T x_j)\lambda_i]||^2_2 + \tau/2||\lambda_i - \mu_i||^2_2
\]  

(31)

This is a standard continuous convex function that can be easily solved by least square.

We summarize our sparse coding method in Algorithm 1.

---

**Algorithm 1** Sparse coding algorithm

**Input and Initialization:**

- Training data $X \in \mathbb{R}^{N \times L}$, iteration number $r$, initial value $\lambda_i = 0$.

**Output:**

- Sparse coefficients $y$, and thresholding values $\lambda_i$.

1. Compute the sparse coefficients $y$ via Problem (24) according to the OMP algorithm.
2. Sort the columns of $X$ and $y$ in increasing order of $|\psi_i^T x|$.  
3. For $p = 1 : r$
   
   For $j = 1 : L$
   
   Compute all the possible values for $f(\mu_i) + g(\mu_i)$ by

   \[
   \begin{align*}
f(\psi_i^T X_{j+1}) &= f(\psi_i^T x_i) + y_i^2(j+1) \\
g(\psi_i^T X_{j+1}) &= ||y_i - \psi_i^T x + sgn(\psi_i^T x)\mu_i||^2 - \beta_{k+1}
\end{align*}
\]  

where $\beta_{k+1} = \beta_k + [y_i(k+1) - \psi_i^T x_{k+1} + sgn(\psi_i^T x_{k+1})\mu_i]^2$, $\beta_1 = [y_i - \psi_i^T x_1 + sgn(\psi_i^T x_1)\mu_i]^2$.

Denote them as a vector $v$.

End for
4. Sort the elements of $|\psi_i^T x|$ in descending order of $v$. Denote the intervals bounded as $\xi_{q, q} = 1, 2, L - 1$.
5. compute every $\nu_q + l(\hat{\mu}_i)$ where $\hat{\mu}_i$ is the point closest to $\lambda_i$ in $\xi_{j-1}$.
6. $\hat{\mu}_i = \arg\min_{\mu_i} \nu_q + l(\mu_i)$.
7. Compute $\lambda_i$ via Problem (31).

End for
4.2. Dictionary Pair Update Phase

To obtain $\Psi$, we solve the following problem with all other variables fixed:

$$\hat{\Psi} = \arg \min_{\Psi} \| Y - S_{\lambda} (\Psi^T X) \|^2_F + \frac{\gamma_3}{\gamma_1} \| \Psi - F^{-1} \Phi \|^2_F$$  

(32)

Such problem is a highly nonlinear optimization due to the definition of $S_{\lambda}$. Here we solve $\Psi$ columnwisely by updating each column of $\Psi$.

For each $\psi_i$, we solve the following subproblem:

$$\hat{\psi}_i = \arg \min_{\psi_i} \| y_i - S_{\lambda} (\psi_i^T X) \|^2_F + \frac{\gamma_3}{\gamma_1} \| \psi_i - F^{-1} \phi_i \|^2_F$$  

(33)

We denote $\hat{\psi}_i$ and $\hat{\gamma}_i$ as the indices set as before. Set the elements of $\hat{\psi}_i$ corresponding to the indices $\hat{\gamma}_i$ to be zeros and denote the new vector as $\tilde{z}_i$. This operation leads to a consequence that $\psi_i^T X \tilde{z}_i \approx 0$. Then we solve the following quadratic optimization problem that is easy to solve with least squares.

$$\hat{\phi}_i = \arg \min_{\phi_i} \| z_i - \psi_i^T X \|^2_F + \frac{\gamma_3}{\gamma_1} \| \phi_i - F^{-1} \phi_i \|^2_F$$  

(34)

The optimization problem to pursue $\Phi$ is formulated as

$$\Phi = \arg \min_{\phi_i} \| X - \Phi Y \|^2_F + \gamma_3 \| \Psi - F^{-1} \Phi \|^2_F$$  

(35)

s.t. $\sqrt{A} \leq \delta_{\Phi} \leq \sqrt{B}$

(36)

where the frame operator $F$ is given by $\Phi \Phi^T$ and $F^{-1}$ is defined as Equation (22). The target function then becomes

$$\| X - \Phi Y \|^2_F + \eta_3 \| \Psi - \frac{2}{A + B} (2I - 2\Phi \Phi^T) \Phi \|^2_F$$  

(37)

which is denoted by $h(\Phi)$. We apply the gradient descent method to unconstraint version of Problem (35) and then project the solution to the feasible space. The gradient is given by a very complicated form as follows

$$\nabla h(\Phi) = (X - \Phi Y) Y^T - \gamma_3 \left\{ \frac{4}{a} h(\Phi) + \frac{4}{a^2} [\Phi \Phi^T h(\Phi) + \Phi h(\Phi)^T \Phi + h(\Phi) \Phi^T \Phi] \right\}.$$  

(38)

In order to reduce the complexity, the gradient can also be computed with the fixed $F$ calculated in the previous step of the ADM. Then at the $k$-th iteration, the gradient can be written as

$$\nabla h(\Phi^k) = (X - \Phi Y) Y^T - \gamma_3 (\Phi^k - F^{-1} \Phi^k)$$  

(39)

where $F = \Phi^{k-1} \Phi^{(k-1)^T}$. The descent step length can be obtained by optimizing the problem $\min_{\phi} h(\Phi + \theta \nabla h(\Phi))$ with fixed $F$, which is given by

$$\hat{\theta} = \frac{\langle a, b \rangle + \gamma_3 \langle c, d \rangle}{\|a\|^2_F + \gamma_3 \|c\|^2_F}$$  

(40)

where $a = \nabla h(\Phi) Y, b = X - \Phi Y, c = F^{-1} \nabla h(\Phi), d = \Psi - F^{-1} \Phi$. For the frame condition $\sqrt{A} \leq \delta_{\Phi} \leq \sqrt{B}$, we apply a SVD decomposition $\Phi = U \Sigma \Phi^T$ and map the singular values to the interval $[\sqrt{A}, \sqrt{B}]$ linearly. We denote the mapped singular matrix as $\hat{\Sigma}$ and reconstruct $\Phi$ by $\Phi = U \hat{\Sigma} \Phi^T$.

We summarize our algorithm in Algorithm 2.
Algorithm 2 Dictionary pair learning algorithm

Input and Initialization:
Training data $X$, frame bound $A,B$, iteration $num$, gradient descent iterations $r$.

Build frame $\Phi \in \mathbb{R}^{N \times M}$ and $\Psi \in \mathbb{R}^{N \times M}$, either by using random entries, or using $M$ randomly chosen data.

Output:
Frame $\Phi, \Psi$, Sparse coefficients $Y$, and thresholding values $\lambda$.

For $l=1-num$

Sparse Coding Step:
1: Compute the sparse coefficients $Y$ and the thresholding values $\lambda$ via Algorithm 1.

Frame Update Step:
2: Update $\Psi$ in column-wise. Compute $W = S(\Psi^T X)$.
   
   For $i = 1 : M$
   Denote $\hat{J}_i$ as the indices of zeros in the $i$-th column of $W$. Set $\psi_i^T X \hat{J}_i = 0$. Compute $\psi_i$ via Equation (34).
   End For
3: Update $\Phi$ via Gradient Descent and singular value map.
   
   For $k=1:r$
   Calculate Gradient via Equation (39) and calculate the descent step via Equation (40).
   End for
4: Apply SVD decomposition $\Phi = U\Sigma V^T$, map $\Sigma$ to obtain $\hat{\Sigma}$ and reconstruct $\Phi = U\hat{\Sigma}V^T$.
   End for

5. Restoration

The image restoration aims to reconstruct a high-quality image $\hat{I}$ from its degraded (e.g., noisy, blurred and/or downsampled) version $L$, denoted by $L = SHI + n$, where $H$ represents a blurring filter, $S$ the downsampling operator, and $n$ is a noisy signal. For the signal satisfies the SSM-NTF, the restoration model based on SSM-NTF is formulated as

$$
\min_{I,y,\lambda} \| L - SHI \|_F^2 + \gamma \sum_i \| R_i I - \Phi y_i \|_F^2 + \gamma I \sum_i \| y_i - S_i(\Psi^T R_i I)\|_F^2 + \gamma_2 \sum_i \| y_i \|_0
$$

(41)

where $R_i$ is an operator that extracts the $i$-th patch of the image $I$ and $y_i$ is the $i$-th column of $Y$. $\lambda$ denotes a vector $[\lambda_1, \lambda_2, \cdots, \lambda_M]$ with $\lambda_j$ operating on the $j$-th element of $\Psi^T R_i I$. On the right side of Equation (41), the first term is the global force that demands the proximity between the degraded image $L$, and its high-quality version $I$. The rest terms are the local constraints to make sure every patch at location $i$ satisfies the SSM-NTF.

To solve Problem (41), we apply Algorithm 1 to obtain the sparse coefficients $Y$ and the threshold values $\lambda$. We mainly state the iterative method to obtain $\hat{I}$. Assume the sign of $\Psi^T R_i I$ will not change much between two steps, we set it in the $k$-th step by $c_k = sign(\Psi^T R_i I_{k-1})$. where sign is the sign function. Denote $d_k^k = \Psi^T R_i I_{k-1}$. We set $O_k$ as an index set that satisfies $|d_k| \leq \lambda_i, i \in O_k$.

Set $u^k \in \mathbb{R}^M$ as a vector with elements $u_i = \begin{cases} \lambda_i & I \in O_k \\ 0 & otherwise \end{cases}$. Then the non-convex and non-smooth
threshold can be removed with the substitution that \( y_i - S_{\lambda}(\Psi^T R_i I^k) \approx y_i + c^k \odot u - \Psi^T R_i I^k \). Thus, in the \( k \)-th step, the problem needs to be solved is expressed as
\[
\{ \hat{I}^k \} = \min_{\hat{I}^k} \| L - SHT^{k-1} \|^2_F + \gamma \sum_{i} \| R_i I^{k-1} - \Phi y_i \|^2_F + \gamma_1 \sum_{i} \| y_i + c^k \odot u^k - \Psi^T R_i I^{k-1} \|^2_F
\]
(42)
where \( \odot \) is point multiplication. This convex problem can be easily solved by gradient descent algorithm.

We summarized the restoration algorithm in Algorithm 3.

**Algorithm 3** Restoration algorithm

**Input**

Training dictionaries \( \Phi, \Psi \), iteration number \( r \), a degraded image \( L \), set \( I_0 = L \).

**Output**

The high quality image \( \hat{I} \)

1: Compute \( Y \) and \( \lambda \) via the method in Algorithm 1.

**For** \( k=1:r \)

2: compute \( d^k = \Psi^T R_i I^{k-1} \). Set \( c^k = \text{sign}(d^k) \). set \( O^k \) as an index set that satisfies \( |d^k_i| \leq \lambda_i, i \in O^k \).

Set \( u_i^k = \begin{cases} 
\lambda_i & i \in O^k, \\
0 & \text{otherwise.}
\end{cases} \)

3: Solving the Problem (42) via gradient descent algorithm.

**End for**

6. Complexity Analysis

In this section, we discuss the computational complexity of our sparse coding and dictionary pair learning algorithms with regard to those of conventional sparse model counterparts.

We first analyze complexities of the main components of the sparse coding (SC) and dictionary updating (DU) algorithms. In terms of SC, given a set of training samples, \( X \in \mathbb{R}^{N \times L} \), the complexity of BOMP of calculating \( \hat{Y} = \min_{Y} \| X - \Phi Y \|^2_F + \gamma_1 \| Y - S_{\lambda}(\Psi^T X) \|^2_F + \gamma_2 \| Y \|_0 \) is \( O(K^2ML) \) where \( K \) is the target sparsity and the complexity of threshold of calculating \( \lambda = \min_{\lambda} \| Y - S_{\lambda}(\Psi^T X) \|^2_F \) is \( O(NML) \), which cost most of time in SC step at each iteration. The sparse coefficients \( Y \in N \times L \) and the threshold values \( \lambda \) are computed with fixed dictionaries \( \Phi \in \mathbb{R}^{N \times M} \) and \( \Psi \in \mathbb{R}^{N \times M} \). Correspondingly, the traditional sparse coefficients \( B \in N \times L \) is sparse approximated by dictionary \( D \in \mathbb{R}^{N \times M} \) and the computational complexity is \( O(K^2ML) \).

In terms of DU, with the given training samples \( X \in \mathbb{R}^{N \times L} \), we learn a pair of dictionary \( \Phi \in \mathbb{R}^{N \times M} \) and \( \Psi \in \mathbb{R}^{N \times M} \). We update \( \Psi \) via Problem (34) with a computational complexity of \( O(N^2L) \). In order to update \( \Phi \), we need to calculate the gradient via Equation (39) with a computational complexity of \( O(NML) \) and the step size via Equation (40) with a computational complexity of \( O(rNML) \) where \( r \) is the iteration number of the gradient descent. For the traditional dictionary learning, the corresponding training set is \( X \in \mathbb{R}^{N \times L} \) and the dictionary \( D \in \mathbb{R}^{N \times M} \) is updated by SVD decomposition of rank-1 with a computational complexity of \( O(KML) \).

7. Experimental Results

We demonstrate the effectiveness of our SSM-NTF model by first discussing the convergence of our dictionary pair learning algorithm and then evaluating the performance on natural and piecewise constant image denoising, super resolution and image inpainting.
7.1. Convergence Analysis

The convergence of the presented dictionary pair learning algorithm is evaluated in Figure 1. Here, we train a pair of dictionaries $\Phi$ and $\Psi$ of size $100 \times 200$ from 65,000 patches, which are of size $10 \times 10$ randomly sampled from six natural images. We apply the frame reconstruction function $\Phi S_\lambda(\Psi^T x)$ to reconstruct the patches. The convergence of the presented dictionary pair learning algorithm is evaluated in Figure 1. The dictionary pair is illustrated in Figure 2. They exhibit that our dictionary pair learning method is able to capture the feature of the image along with the convergence property.

Figure 1. Convergence analysis. The X-labels are the iteration number. The Y-labels are the is the objective function of System (20) (left) and the restoration result (measured by ‘PSNR’) (right). It is shown that our dictionary pair learning algorithm is a convergence one.

Figure 2. The exemplified dictionary pair ($\Phi$, $\Psi$) in our stable sparse model with non-tight frame (SSM-NTF) model training by natural images.

7.2. Image Denoising

In this subsection, we evaluate the performance of our proposed SSM-NTF model on image denoising. Benefitting from the concept of non-tight frame, the proposed SSM-NTF model contains a pair of dictionaries: The frame and its dual frame. As a result, our proposed SSM-NTF model contains an analysis system and a synthesis system. The analysis-like system is denoted as

$$y = S_\lambda(\Psi^T x), \ \|y\| \leq s$$  (43)
which analyzes the signals in $\Psi$ domain. The synthesis system is denoted as
\[ x = \Phi y, \quad \|y\| \leq s \] (44)
which reconstructs the analyzed signals. The two systems share the same sparse coefficients $Y$.

Therefore, we compare our proposed SSM-NTF with synthesis and analysis models, respectively. It is well known that the synthesis sparse model has advantage in dealing with the natural image while the analysis sparse model is mostly used to address the piecewise constant image. Therefore, we respectively, perform the denoising experiments on natural images and piecewise images comparing with the most related approaches.

7.2.1. Natural Image Denoising

We now turn to present experimental results on six classical natural images named 'Barbara', 'Boat', 'Couple', 'Hill', 'Lena' and 'Man' which are shown in [1], to evaluate the performance of the training algorithm. The denoising problem which has been widely studied in sparse representation is used as the target application. We add Gaussian white noise to these images at different noise levels $\sigma = 20, 30, 40, 50, 60, 70, 80, 90, 100$. Then we use the learned dictionary pair to denoise the natural images, with overlap of 1 pixel between adjacent patches of size $10 \times 10$. The patch denoising stage is followed by weighted averaging the overlapping patch recoveries to obtain the final clean image. The parameters in our scheme are $\gamma_1 = 1.1$ and $\gamma_3 = 1.2 \left(\frac{L}{M}\right)^2$ where $L$ and $M$ are the sample and dictionary size, respectively. We have stated in Section 3.2 that we usually set $A$ to a positive number around but smaller than 1. In fact, we set $A$ from 0.6 to 1 by a step of 0.03 to test the denoising performance to determine the specific value of it. Then, with fixed $A$, we set $B$ from 1 to 4 with a step length of 0.3 to run experiments on every noise level to determine the values of $B$. The values of frame bounds $A$ and $B$ are shown in Table 1. For example, when the noise level $\sigma = 40$, $A$ and $B$ are set to be 0.8 and 1.8, respectively.

| $\sigma$ | 20  | 30  | 40  | 50  | 60  | 70  | 80  | 90  | 100 |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A       | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| B       | 1.8 | 1.8 | 1.8 | 1.8 | 2.4 | 3.0 | 3.3 | 3.3 | 3.6 |

Table 1 shows the comparison results in terms of PSNR. There are three related image denoising methods involved, including the classical dictionary learning algorithm KSVD [9], the data-driven tight frame based denoising method [1] and the incoherent dictionary learning based method [21]. The patch size of KSVD [9] and the method in [21] are $8 \times 8$ with stripe 1 and the dictionaries are of size $64 \times 256$ at their optimal state according to the previous work. We point out that [1] works on filters of size $16 \times 16$ instead of image patches and initialized by 64 3-level Harr wavelet filters in size $16 \times 16$. All the three compared methods can achieve their best performance with 50 iterations.

Table 2 shows that the incoherent dictionary learning method [21] outperforms the KSVD [9] in average as the mutual incoherent of dictionary can provide stable recovery. That [1] outperforms [21] implies that the tight frame is a more stable system. Then our stable sparse model based method outperforms [1] in average suggests that applying non-tight frame to approximate RIP can provide even better and more stable reconstruction results. Figure 3 shows two exemplified visual results on images 'Man' and 'Couple' at noise levels $\sigma = 50$ and $\sigma = 40$, respectively. The proposed method shows much clearer and better visual results than the other competing methods.
Figure 3. Visual comparison of reconstruction results by different methods on ‘Man’ ($\sigma = 50$) and ‘Couple’ ($\sigma = 40$). From left to right: original image, noise image, KSVD [9], method of [1], method of [21] and our proposed method.

Table 2. PSNR (dB) for nature image denoising results.

| $\sigma$ | Image | Barbara | Boat | Couple | Hill | Lena | Man | Average |
|---------|-------|---------|------|--------|------|------|-----|---------|
| 20      | KSVD [9] | 31.01   | 30.50 | 30.15 | 30.27 | 32.51 | 30.26 | 30.78   |
|         | [21]   | 31.03   | 30.45 | 30.16 | 30.30 | 32.52 | 30.30 | 30.79   |
|         | [1]    | 31.07   | 30.35 | 30.20 | 30.31 | 32.56 | 30.07 | 30.76   |
|         | SSM-NTF| 31.06   | 30.56 | 30.32 | 30.42 | 32.60 | 30.33 | 30.88   |
| 30      | KSVD [9] | 28.75   | 28.60 | 28.07 | 28.51 | 30.59 | 28.43 | 28.83   |
|         | [21]   | 28.78   | 28.63 | 28.06 | 28.53 | 30.60 | 28.47 | 28.85   |
|         | [1]    | 29.07   | 28.48 | 28.22 | 28.64 | 30.60 | 28.26 | 28.88   |
|         | SSM-NTF| 29.03   | 28.71 | 28.32 | 28.74 | 30.82 | 28.59 | 29.04   |
| 40      | KSVD [9] | 27.03   | 27.23 | 26.54 | 27.23 | 29.13 | 27.17 | 27.39   |
|         | [21]   | 27.05   | 27.18 | 26.59 | 27.21 | 29.10 | 27.12 | 27.38   |
|         | [1]    | 27.38   | 27.20 | 26.87 | 27.49 | 29.25 | 26.99 | 27.56   |
|         | SSM-NTF| 27.52   | 27.49 | 27.07 | 27.65 | 29.43 | 27.41 | 27.76   |
| 50      | KSVD [9] | 25.71   | 26.05 | 25.42 | 26.29 | 27.92 | 26.18 | 26.26   |
|         | [21]   | 25.77   | 26.08 | 25.40 | 26.31 | 27.87 | 26.23 | 26.28   |
|         | [1]    | 25.64   | 25.15 | 25.84 | 26.63 | 28.15 | 26.09 | 26.55   |
|         | SSM-NTF| 25.40   | 26.41 | 26.04 | 26.88 | 26.61 | 26.49 | 26.79   |
| 60      | KSVD [9] | 24.45   | 25.18 | 24.57 | 25.69 | 27.01 | 25.40 | 25.38   |
|         | [21]   | 24.45   | 25.20 | 24.50 | 25.65 | 27.03 | 25.42 | 25.38   |
|         | [1]    | 25.64   | 25.33 | 25.04 | 25.91 | 27.22 | 25.38 | 25.75   |
|         | SSM-NTF| 25.52   | 25.55 | 25.22 | 26.11 | 27.41 | 25.74 | 25.93   |
| 70      | KSVD [9] | 23.40   | 24.46 | 23.90 | 25.10 | 26.18 | 24.77 | 24.63   |
|         | [21]   | 23.46   | 24.46 | 23.93 | 25.15 | 26.19 | 24.79 | 24.66   |
|         | [1]    | 24.88   | 24.67 | 24.36 | 25.30 | 26.44 | 24.78 | 25.07   |
|         | SSM-NTF| 24.74   | 25.05 | 24.60 | 25.50 | 26.61 | 25.06 | 25.26   |
Table 2. Cont.

|    | KSVD [9] | 80      | 22.80 | 23.85 | 23.44 | 24.69 | 25.34 | 24.19 | 24.05 |
|----|----------|---------|-------|-------|-------|-------|-------|-------|-------|
|    | [21]     |         | 22.87 | 23.93 | 23.47 | 24.76 | 25.40 | 24.27 | 24.09 |
|    | [1]      |         | 24.18 | 24.09 | 23.79 | 24.77 | 25.74 | 24.25 | 24.47 |
|    | SSM-NTF  |         | 24.11 | 24.42 | 24.02 | 24.91 | 26.02 | 24.58 | 24.68 |
|    | ksvd [9] | 90      | 22.31 | 23.34 | 22.98 | 24.31 | 24.96 | 23.78 | 23.61 |
|    | [21]     |         | 22.31 | 23.38 | 23.01 | 24.36 | 24.97 | 23.84 | 23.64 |
|    | [1]      |         | 23.50 | 23.56 | 23.27 | 24.30 | 25.14 | 23.79 | 23.93 |
|    | SSM-NTF  |         | 23.41 | 23.84 | 23.47 | 24.58 | 25.43 | 24.13 | 24.14 |
|    | ksvd [9] | 100     | 21.88 | 22.92 | 22.64 | 23.96 | 24.44 | 23.39 | 23.21 |
|    | [21]     |         | 21.89 | 22.95 | 22.67 | 23.99 | 24.47 | 23.43 | 23.23 |
|    | [1]      |         | 23.01 | 23.11 | 22.83 | 23.86 | 24.59 | 23.38 | 23.46 |
|    | SSM-NTF  |         | 22.90 | 23.46 | 23.05 | 24.16 | 24.90 | 23.68 | 23.69 |

7.2.2. Piecewise Constant Image Denoising

In this subsection, we demonstrate the analytical property of our SSM-NTF model using a synthetic image. The denoising problem which has been widely studied in sparse representation is used as the target application. We start with a piecewise constant of size $256 \times 256$ contaminated by Gaussian white noise with noise level $\sigma = 5$ and extract all possible $5 \times 5$ image paths. For the denoising we apply the dictionary pair learning algorithm with the parameters $\gamma_1 = 1.5$, $\gamma_3 = 1.2 \frac{L}{M}$, $A = 0.8$ and $B = 1.8$ in parallel with patch denosing with the synthesis KSVD [9] and the analysis KSVD [14]. We apply 100 iterations of the our dictionary learning method on this training set, and learning dictionary pair of size $25 \times 50$. The experimental set of the synthesis KSVD [9] and the analysis KSVD [14] are at their optimal state according to the previous work.

The learned dictionary pair $\Phi$ which exhibits much like the synthesis dictionary and $\Psi$ which exhibits a high resemblance to the analysis dictionary are illustrated in Figure 4. The resulting PSNRs of the denoised images are 45.32 dB for Analysis KSVD, 43.60 dB for Synthesis KSVD, and 45.17 dB for our proposed algorithm. The figure shows that our dictionary pair learning method is able to capture the features of the piecewise constant image. Figure 5 shows the absolute difference images for each of the three methods. Note that these images are displayed in the dynamic range $[0, 20]$ to make the differences more pronounced. Our proposed approach leads to a much better denoising result than the synthesis KSVD and is comparable with the analysis KSVD.

Figure 4. The exemplified dictionaries training by the piecewise constant. From left to right: Synthesis KSVD ($25 \times 100$), our proposed dictionary pair ($25 \times 50$) and analysis KSVD ($25 \times 50$).
7.3. Super Resolution

We evaluate our SSM-NTF in comparison with two exemplar-based scheme [37] for image Super Resolution (SR) Problem (41) with a bicubic filter. Figure 6 shows the 15 test natural images [18] with both rich texture and structure. All the schemes are applied to the illumination channel, where the scale factor is 3, we always use $3 \times 3$ low-resolution patches with overlap of 1 pixel between adjacent patches, corresponding to $9 \times 9$ patches with overlap of 3 pixels for the high-resolution patches. In these experiments, we have used the following parameters: $A = 0.8$, $B = 1.8$, $\gamma_1 = 1.1$ and $\gamma_3 = 1.2 \frac{L}{M}$ where $L$ and $M$ are the sample and dictionary size, respectively. In our scheme, dictionary learning is performed between HR and middle-level (MR) images which are the first-, and second-order derivatives of the upsampled version of one LR image by a factor of 2. The four 1D filters used to extract the derivatives are:

$$f_1 = [-1, 0, 1], \quad f_2 = f_1^T$$
$$f_3 = [1, 0, -2, 0, 1], \quad f_4 = f_3^T$$

(45)

We train two pairs of HR/LR dictionaries $\{\Phi_h, \Psi_h\}$ and $\{\Phi_l, \Psi_l\}$ from 100,000 HR/LR patch pairs $[X_h, X_l]$ randomly sampled from the collected natural images which are also used in [37] where $X_h$ is sampled from the HR images and $X_l$ is sampled from the four feature images. The feature images are obtained by applying the four filters to the upsampled LR image. Given $\Phi$ and $\Psi$ and the four MR feature images, the sparse coefficients $Y$ and threshold value $\lambda$ can be calculated by Algorithm 1. With the theory in [37], the HR image can be recovered via Algorithm 3. In the experiment, our HR dictionary pair are of size $81 \times 450$ and MR ones are of size $144 \times 450$. The dictionary size of [37] is $81 \times 1024$ (HR) and $144 \times 1024$ (MR) at its best performance as stated in the paper. Thus, the dictionary size of [37] is larger than the sum of our dictionaries. Table 3 shows the objective evaluation results of our proposed SSM-NTF compared with bicubic interpolation and [37]. On average, our SSM-NTF presents best in PSNR. Figure 7 presents the corresponding visual comparison of the illumination SR results of Image 12. We can observe that the result of bicubic interpolation is too smooth and the result of [37] suffers from obvious ringing artifact and noises. The HR reconstruction of our SSM-NTF method provides more clear details.
Figure 7. Visual quality comparison of SR results for Image 12 corresponding to Table 3. From left to right: Original image, result of bicubic interpolation ($\text{PSNR} = 30.28$), ($\text{PSNR} = 30.62$) and our SSM-NTF method ($\text{PSNR} = 30.99$), respectively.

Table 3. PSNR (dB) for $3 \times$ SR reconstructions results.

| Image | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| Bicubic [37] | 32.69 | 26.32 | 26.04 | 28.39 | 28.70 | 25.81 | 37.21 | 24.40 |
| SC [37] | 32.89 | 26.53 | 26.16 | 28.51 | 29.09 | 26.25 | 37.59 | 24.61 |
| SSM – NTF | 33.10 | 26.75 | 26.41 | 28.82 | 29.31 | 26.40 | 37.49 | 24.77 |

| Image | 9   | 10  | 11  | 12  | 13  | 14  | 15  |
|-------|-----|-----|-----|-----|-----|-----|-----|
| Bicubic [37] | 31.40 | 28.72 | 33.24 | 30.28 | 26.49 | 28.11 | 31.90 | 29.31 |
| SC [37] | 31.73 | 29.35 | 33.50 | 30.62 | 26.75 | 28.74 | 32.13 | 29.63 |
| SSM – NTF | 31.90 | 29.49 | 33.66 | 30.99 | 26.95 | 29.02 | 32.19 | 29.82 |

7.4. Image Inpainting

To illustrate the potential applicability of our proposed SSM-NTF model on image inpainting, we apply it to the applications of text removal. In these experiments, we have used the following parameters: $A = 0.8, B = 1.8, \gamma_1 = 1.1$ and $\gamma_3 = 1.2 \frac{L}{M}$ where $L$ and $M$ are the sample and dictionary size, respectively. We operate on the image ‘Adar’, ‘Lena’, ‘Couple’, ‘Hill’ with super-imposed text of various fonts.

In this experiment, we applied our SSM-NTF model to image inpainting in a way similar to the non-blind KSVD inpainting algorithm [9], which requires the knowledge of which pixels are corrupted and required inpainting. Actually, only the non-corrupted pixels are used to training the dictionary pair and inpainting the images. We operate our method on paths of size $10 \times 10$ that extract from the images with overlap of 1 pixel between adjacent. The trained dictionary pair are of size $100 \times 200$. The KSVD algorithm in this experiment is dealing with patches of size $8 \times 8$ that extract from the images with overlap of 1 pixel between adjacent. The dictionary size is $64 \times 256$ at its best performance according to [9]. The patch inpainting stage is followed solving Problem (41). Table 4 shows the objective evaluation results of our proposed SSM-NTF compared with DCT and KSVD [9]. The visual comparisons are shown in Figures 8 and 9. We find that the proposed SSM-NTF method is able to eliminate text of fonts completely while the KSVD is dull. Our SSM-NTF method achieves better performance in terms of both subjective and objective quality.
Figure 8. Visual quality comparison of text image inpainting results. From left to right: Original image, text image, result of DCT ($PSNR = 27.11$), KSVD ($PSNR = 28.01$) and our SSM-NTF method ($PSNR = 28.95$), respectively.

Figure 9. Visual quality comparison of scratch image inpainting results. From left to right: Original image, scratch image, result of DCT ($PSNR = 30.81$), KSVD ($PSNR = 31.69$), and our SSM-NTF method ($PSNR = 32.02$), respectively.

Table 4. PSNR (dB) for image inpainting results.

| Image   | Adar | Lena | Hill | Couple | Average |
|---------|------|------|------|--------|---------|
| DCT [9] | 27.11| 31.30| 28.53| 29.21  | 29.04   |
| KSVD [9]| 28.01| 31.69| 28.90| 29.50  | 29.53   |
| SSM−NTF | **28.95**| **32.02**| **29.31**| **29.87**| **30.04**|

8. Conclusions

In this paper, we propose a stable sparse model with non-tight frame (SSM-NTF) and further formulate a dictionary pair learning model to stably recover the signals. We theoretically analyze the rationality of the approximation for RIP with the non-tight frame condition. The proposed SSM-NTF has RIP and the closed-form expression of the sparse coefficients that ensure the stable recovery especially for seriously noise images. The proposed SSM-NTF contains both a synthesis sparse and an analysis system which share the common sparse coefficients without taking into account the thresholding. We also propose an efficient dictionary pair learning algorithm via developing an explicit analytical expression of the inherent relation between the dictionary pair. The proposed algorithm is capable of approximating structures of signals via a pair of adaptive dictionaries. The effectiveness of our proposed SSM-NTF and its corresponding algorithms are demonstrated in image denoising, image super-resolution and image inpainting. The results of numerical experiments show that the proposed SSM-NTF achieves superior to the compared methods in objective and subjective quality on most of the cases.

On the other hand, our proposed SSM-NTF is actually a 1D sparse model. The 1D sparse model suffers from high memory as well as high computational costs especially when handling high dimensional data. MD frame can be expressed as the kronecker product of a series of 1D frames. Benefitting from this good characteristic, in future work, we will extend our stable sparse model to propose a MD stable sparse model. Moreover, the proposed SSM-NTF is not effective enough to remove other kinds of noise (e.g., salt and pepper noise) as the loss function of SSM-NTF is gaussian. We would like to improve the performance of our model by changing the loss function.
Author Contributions: M.Z. derived the theory, analyzed the data, performed the performance and wrote the original draft; Y.S. and N.Q. researched the relevant theory, participated in discussions of the work and revised the manuscript; B.Y. supervised the project. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by NSFC (No.61672066,61976011, U18111463, U19B2039, 61632006, 61906008, 61906009), Beijing municipal science and technology commission (No.Z171100004417023) and Common Program of Beijing Municipal Commission of Education (KM202010005018).

Acknowledgments: This work was supported by Beijing Advanced Innovation Center for Future Internet Technology, Beijing Key Laboratory of Multimedia and Intelligent Software Technology. We deeply appreciate the organizations mentioned above.

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

References
1. Cai, J.F.; Ji, H.; Shen, Z.; Ye, G.B. Data-driven tight frame construction and image denoising. *Appl. Comput. Harmon. Anal.* 2014, 37, 89–105. [CrossRef]
2. Xie, J.; Feris, R.S.; Yu, S.S.; Sun, M.T. Joint super resolution and denoising from a single depth image. *IEEE Trans. Multimed.* 2015, 17, 1525–1537. [CrossRef]
3. Rubinstein, R.; Zibulevsky, M.; Elad, M. Efficient Implementation of the KSVD Algorithm and the Batch-OMP Method. Tech. Rep.; Department of Computer Science, Technion: Haifa, Israel, 2008.
4. Fadili, M.J.; Starck, J.-L.; Murtagh, F. Inpainting and zooming using sparse representations. *Comput. J.* 2009, 52, 64–79.
5. Liu, Y.; Zhai, G.; Gu, K.; Liu, X.; Zhao, D.; Gao, G. Reduced-reference image quality assessment in free-energy principle and sparse representation. *IEEE Trans. Multimed.* 2018, 20, 379–391. [CrossRef]
6. Zhao, J.; Hu, H.; Cao, F. Image super-resolution via adaptive sparse representation. *Knowl. Based Syst.* 2017, 124, 23–33. [CrossRef]
7. Zhu, Z.; Guo, F.; Yu, H.; Chen, C. Fast single image super-resolution via self-example learning and sparse representation. *IEEE Trans. Multimed.* 2014, 16, 2178–2190. [CrossRef]
8. Beck, A.; Teboulle, M. A fast iterative shrinkage thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* 2009, 2, 183–202. [CrossRef]
9. Elad, M. Sparse and Redundant Representations. In *From Theory to Applications in Signal and Image Processing*; Springer: Berlin, Germany, 2010.
10. Zhang, L.; Bioucas-Dias, J.M. Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 2018, 11, 730–742. [CrossRef]
11. Kalluri, M.; Jiang, M.; Ling, N.; Zheng, J.; Zhang, P. Adaptive RD optimal sparse coding with quantization for image compression. *IEEE Trans. Multimed.* 2019, 21, 39–50. [CrossRef]
12. Liu, Y.; Dimitris A. Pados, Compressed-sensed-domain l1-pca video surveillance. *IEEE Trans. Multimed.* 2016, 18, 351–363. [CrossRef]
13. Babaie-Zadeh, M.; Jutten, C. Corrections to on the stable recovery of the sparsest overcomplete representations in presence of noise [oct 10 5396–5400]. *IEEE Trans. Signal Process.* 2011, 59, 1913. [CrossRef]
14. Rubinstein, R.; Peleg, T.; Elad, M. Analysis KSVD: A dictionary-learning algorithm for the analysis sparse model. *IEEE Trans. Signal Process.* 2012, 61, 661–677. [CrossRef]
15. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process* 2006, 15, 3736–3745. [CrossRef] [PubMed]
16. Dong, W.; Zhang, L.; Shi, G.; Li, X. Nonlocally centralized sparse representation for image restoration. *IEEE Trans. Image Process.* 2013, 22, 1620–1630. [CrossRef] [PubMed]
17. Song, X.; Peng, X.; Xu, J.; Shi, G.; Wu, F. Distributed compressive sensing for cloud-based wireless image Transmission. *IEEE Trans. Multimed.* 2017, 19, 1351–1364. [CrossRef]
18. Qi, N.; Shi, Y.; Sun, X.; Yin, B. Tensr: Multi-dimensional tensor sparse representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, Las Vegas, NV, USA, 27–30 June 2016; pp. 5916–5925.
19. Donoho, D.L.; Elad, M.; Vladimir, N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* 2006, 52, 6–18. [CrossRef]
20. Bao, C.; Cai, J.-F.; Ji, H. Fast sparsity-based orthogonal dictionary learning for image restoration. In Proceedings of the IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, 1–8 December 2013; pp. 3384–3391.

21. Wang, J.; Cai, J.-F.; Shi, Y.; Yin, B. Incoherent dictionary learning for sparse representation based image denoising. In Proceedings of the IEEE International Conference on Image Processing, ICIP 2014, Paris, France, 27–30 October 2014; pp. 4582–4586.

22. Candès, E.J.; Tao, T. Decoding by linear programming. IEEE Trans. Inf. Theory 2005, 51, 4203–4215. [CrossRef]

23. Wang, J.; Li, G.; Rencker, L.; Wang, W.; Gu, Y. An rip-based performance guarantee of covariance-assisted matching pursuit. IEEE Signal Process. Lett. 2018, 25, 828–832. [CrossRef]

24. Akbari, A.; Trocan, M.; Granado, B. Sparse recovery-based error concealment. IEEE Trans. Multimed. 2017, 19, 1339–1350. [CrossRef]

25. Zhang, R.; Li, S. A proof of conjecture on restricted isometry property constants $\delta_t$ ($0 < t < 4/3$). IEEE Trans. Inf. Theory 2018, 64, 1699–1705.

26. Lin, J.; Li, S.; Shen, Y. New bounds for restricted isometry constants with coherent tight frames. IEEE Trans. Signal Process. 2013, 61, 611–621. [CrossRef]

27. Cao, C.; Gao, X. Compressed sensing image restoration based on data-driven multi-scale tight frame. J. Comput. Appl. Math. 2017, 309, 622–629. [CrossRef]

28. Li, J.; Zhang, W.; Zhang, H.; Li, L.; Yan, B. Data driven tight frames regularization for few-view image reconstruction. In Proceedings of the 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2017, Guilin, China, 29–31 July 2017; pp. 815–820.

29. Liu, Y.; Zhan, Z.; Cai, J.-F.; Guo, D.; Chen, Z.; Qu, X. Projected iterative softthresholding algorithm for tight frames in compressed sensing magnetic resonance imaging. IEEE Trans. Med. Imaging 2016, 35, 2130–2140. [CrossRef]

30. Bai, H.; Li, S.; He, X. Sensing matrix optimization based on equiangular tight frames with consideration of sparse representation error. IEEE Trans. Multimed. 2016, 18, 2040–2053. [CrossRef]

31. Haltmeier, M. Stable signal reconstruction via $l_1$-minimization in redundant, non-tight frames. IEEE Trans. Signal Process. 2013, 61, 420–426. [CrossRef]

32. Li, D.F.; Sun, W.C. Expansion of frames to tight frames. Acta Mathematica Sinica Eng. Ser. 2009, 25, 287–292. [CrossRef]

33. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. J. R. Statist. Soc. Ser. B 1996, 58, 267–288. [CrossRef]

34. Nam, S.; Davies, M.E.; Elad, M.; Gribonval, R. The cosparse analysis model and algorithms. Appl. Comput. Harmon. Anal. 2013, 34, 205–229. [CrossRef]

35. Zhang, Y. Theory of compressive sensing via $l_1$-minimization: A non-rip analysis and extensions. J. Opt. Res. Soc. China 2013, 1, 79–105. [CrossRef]

36. Christensen, O. Finite dimensional approximation of the inverse frame operator and applications to wavelet frames and Gabor frames. J. Fourier Anal. Appl. 2000, 1, 79–90. [CrossRef]

37. Yang, J.; Member, J.W.; Huang, T.S.; Ma, Y. Image Super-Resolution Via Sparse Representation. IEEE Trans. Image Process. 2010, 19, 2861–2873. [CrossRef]