Image Restoration: A Data-Driven Perspective

Bin Dong, Zuowei Shen

Abstract. We are living in the era of big data. The discovery, interpretation and usage of the information, knowledge and resources hidden in all sorts of data to benefit human beings and to improve everyone's day to day life is a challenge to all of us. The huge amount of data we collect nowadays is so complicated, and yet what we expect from it is so much. It is hard to imagine that one can characterize these complicated data sets and solve real life problems by solving merely a few mathematical equations. However, generic mathematical models can be used to obtain a coarse level approximation (or low accuracy solution) to the answers we are seeking. The first question is how to use generic prior knowledge of the underlying solutions of the problem in hand and to set up a proper model for a good low level approximation? The second question is whether we are able to use the knowledge and information from the approximate solution derived from the given data to further improve the model itself so that more accurate solutions can be obtained? That is: how to engage an interactive data-driven approach to solve complex problems?

As images are one of the most useful and commonly used types of data, in this article, we review the development of the wavelet frame (or more general redundant system) based approach for image restoration from a data-driven perspective. We will observe that a good system for approximating any function, including images, should be capable of effectively capturing both global patterns and local features of the function. A wavelet frame is one of the examples of such system. We will show how algorithms of wavelet frame based image restoration are developed via the generic knowledge of images. Then, we will show how specific information of a given image can be used to further improve the models and algorithms. Through this process, we shall reveal some insights and understandings of the wavelet frame based approach for image restoration. We hope that this also leads to new ideas on how to analyze more complex data sets generated from other real life problems.

Mathematics Subject Classification (2000). 35A15, 42C40, 45Q05, 65K10, 68U10.

Keywords. Wavelet frames, data-driven, sparse approximation, image restoration.

1. Introduction

As we are living in a digital world now, the creation, distribution, integration, interpretation and manipulation of data have become an important part of our society. Digital images are no doubt one of the most important components of data. This is not only because image is a powerful and widely used medium of communication, but also because it is an easy, compact, and widespread way to represent the physical world. Advances in computer technology have made it possible to apply some of the most sophisticated developments in mathematics and sciences to the design and implementation of fast algorithms running on a large number of processors to process image data. As a result, image processing and analysis techniques are now widely applied to natural sciences, technical disciplines and social medias; and digital images have come into everyone's life.

Image restoration, including image denoising, deblurring, inpainting, medical imaging, etc., is one of the most important areas in imaging science. Its major purpose is to obtain high quality reconstructions of images that are corrupted in various ways during the process of imaging, acquisition and communication, and enable us to see crucial but subtle objects that reside in the images. Mathematics has been playing an important role in image restoration from the very beginning. In fact, it has been one of the driving forces of the modern development of image restoration. Conversely, image restoration brings to mathematics a host of challenging new problems and fascinating applications that gave birth to many new mathematical tools whose application has even gone beyond the scope of image restoration.

Most of the existing models and algorithms for image restoration are transformation based. A good transformation for image restoration should be capable of capturing both global patterns and local features of images. The global patterns are smooth image components that provide a global view of images, while the local features are "sharp" image components that characterize local singularities (or fine details) of images such as edges and hidden edges (jump discontinuities after 1st order differentiations). For example, when the transform deduced by convolution of a certain filter bank is used, the global patterns are normally encoded by dense coefficients obtained from low-pass filtering; and the local features are often encoded by coefficients obtained from high-pass filtering whose magnitudes are concentrated at zero with only a small portion of large (in magnitude) coefficients which are mostly located near image singularities. Such property is known as "sparsity". A good example of such filter bank based transform is the tight wavelet frame transform.

Proceedings of the ICIAM,
Beijing, China, 2015

One of the earliest transform used is the Fourier transform, which is effective on signals that are smooth and sinusoidal like. However, Fourier transform can only well capture global patterns due to its poor spatial localization. This makes Fourier transform much less effective on signals with multiple localized frequency components. Windowed Fourier transforms [1] were introduced to overcome the poor spatial localization of the Fourier transform. A window Fourier transform provides a "global patterns plus local features" type of decomposition of signals. However, the high frequency coefficients in the transform domain are not ideally sparse for images due to the fixed time-frequency resolution of windowed Fourier transforms. This is why wavelets and wavelet frames are much more effective for images than Fourier or windowed Fourier transforms, because of their varied time-frequency resolution which enables them to provide better sparse approximation to local image features. This leads to a very successful application of orthonormal or biorthogonal wavelets in image compression (see e.g. [2]). In this paper, we shall present a story of image restoration by redundant systems, especially wavelet frames that can provide a good sparse approximation to local image features, while at the same time, capture global patterns of images as well.

Images can be regarded as a certain discrete realization of piecewise smooth functions (see (3.4) for the definition of piecewise smooth function space). Image restoration problems can be formulated as the following linear inverse problem

$$\boldsymbol{f} = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{\eta} \tag{1.1}$$

where the matrix $\boldsymbol{A}$ is some linear operator (not invertible in general) and $\boldsymbol{\eta}$ denotes a perturbation caused by the additive noise in the observed image, which is typically assumed to be white Gaussian noise. Different image restoration problem corresponds to a different type of $\boldsymbol{A}$, e.g., the identity operator for image denoising, a restriction operator for inpainting, a convolution operator for image deconvolution, a partial collection of line integrations for CT imaging, a partial Fourier transform for MR Imaging, etc.

The problem (1.1) is usually ill-posed, which makes solving (1.1) non-trivial. A naive inversion of $\boldsymbol{A}$, such as pseudo-inversion or via Tikhonov regularization [3], may result in a recovered image with amplified noise and smeared-out edges. A good image restoration method should be capable of smoothing the image so that noise is suppressed to the greatest extend, while at the same time, preserving important image features such as edges, ridges, corners, etc. This is a rather challenging task since smoothing and preservation of features are often contradictory to each other. Therefore, the key to the success of any transformation based image restoration method is to find a transform that can identify local features from the given image, or in other words, to separate singularities and smooth image components.

Wavelet frames represent images as an addition of global patterns, i.e. smooth image components, and local features, i.e. image singularities. In wavelet frame domain, global patterns are represented by densely distributed coefficients obtained from low-pass filtering, while local features are represented by sparse coefficients obtained from high-pass filtering. This makes it easy for us to differentiate between smooth and sharp image components. Therefore, wavelet frames can effectively separate smooth image components and image features, which is the key to their success in image restoration. In addition to provide sparse approximation to local image features, the large coefficients from high-pass filtering can also be used to accurately detect the locations and estimate the types of image singularities [2, 4, 5]. In other words, the coefficients from high-pass filtering also provide reliable analysis and classifications of image features in the transform domain. For simplicity, we shall refer to the wavelet frame coefficients obtained from low-pass filtering as dense coefficients and those obtained from high-pass filtering as sparse coefficients.

Most of the transformation based image restoration methods conduct processing in the transform domain. For wavelet frame transforms, the dense coefficients are normally untouched since the regularity of the smooth image components they approximate is automatically granted by the low-pass filtering. On the other hand, since the sparse coefficients of wavelet frame transforms are highly concentrated, it is natural to set small coefficients to zero since they are most likely to be noise; and sharpen the large coefficients so that image features are enhanced. Such processing is known as shrinkage, whose main purpose is to enhance image features while maintaining smoothness in smooth image regions. Simplest examples of shrinkage operators include the well-known soft- and hard-thresholding that are widely used in transformation based image restoration. However, there are more sophisticated shrinkage operators used in the literature, such as the adaptive wavelet frame shrinkage [6], which works better than soft-thresholding. The soft-thresholding operator was later shown to be equivalent as the minimization of an ℓ_1-norm based optimization model [7]. The hard-thresholding operator is equivalent as the minimization of an ℓ_0-"norm" based optimization model [8, 9]. These discoveries laid the foundation for the recent development of transformation based image restoration models and algorithms.

Redundant systems such as wavelet frames have been implemented with excellent results in both classical and some more challenging image restoration problems. Their applications in classical image restoration problems include image inpainting [10], super-resolution [11], deblurring [12–15], demosaicing [16] and enhancement [17]. Wavelet frames are also applied to more challenging image restoration problems such as blind deblurring [18, 19], blind inpainting [20], and denoising with unknown noise type [21]. Wavelet frame related algorithms have been developed to solve medical and biological image

processing problems as well, e.g. X-ray computer tomography (CT) image reconstruction [22, 23], and protein molecule 3D reconstruction from electron microscopy images [24]. Frames provide large flexibility in designing adaptive and non-local filters with improved performance in applications. For example, the filters used for image restoration in [25–27] are learned from the image, resulting in filters that capture certain features of the image and lead to a transform that gives better sparse approximations for local feature. In [28], Gabor frame filter banks are designed to achieve high orientation selectivity that adapts to the geometry of image edges for sparse image approximation. The application of wavelet frames has gone beyond image restoration. They have been successfully used in video processing [29], 4D CT image reconstruction [30, 31], image segmentation [32, 33] and classifications [34, 35]. More recently, wavelet frames are constructed on non-flat domains such as surfaces [36, 37] and graphs [38–41] with applications to denoising [36, 37, 41] and graph clustering [41].

Another class of methods for image restoration that have been developed through a rather different path is the PDE based approach [42–44], which includes variational and (nonlinear) PDE based methods. The basic idea of variational methods is to characterize images as functions living in a certain function space, such as BV space [45] (space of functions with bounded variations), and an energy functional is designed according to the function space assumption. While wavelet frame based approach also has a function space interpretation of the underlying solutions, i.e. piecewise smooth function space (3.4), variational methods can be understood as a transformation based methods as well, where the transformations are the differential operators involved in the variational models. PDE models, on the other hand, do not seem to be transformation based at the first glance. Most of the design of (nonlinear) PDEs for image restoration is based on geometric or physical properties of the images to be recovered. It is not clear what transformation is essentially being used in these PDE models. However, based on the recent findings by [6], PDE models can be understood as transformation based methods as well, where the transformation is the discretization of the differential operators involved in the PDEs.

In recent work by [46, 47], fundamental connections between wavelet frame based approach and variational methods were established. In particular, connections to the total variation model [45] was established in [46], and to the Mumford-Shah model [48] was established in [47]. Furthermore, in [6], the authors established a generic connection between iterative wavelet frame shrinkage and general nonlinear evolution PDEs. The authors showed that wavelet frame shrinkage algorithms are discretizations of a generic type of nonlinear evolution PDEs that includes the Perona-Malik equation [49] and the shock-filters [50]. The series of three papers [6, 46, 47] showed that wavelet frame transforms are discretization of differential operators in both variational and PDE frameworks, and such discretization is superior to some of the traditional finite difference schemes for image restoration. This new understanding essentially merged the two seemingly unrelated areas: wavelet frame base approach and PDE based approach. It also gave birth to many innovative and more effective image restoration models and algorithms.

In this paper, we focus on a review of some classical and recent wavelet frame based image restoration methods from a data-driven point of view. In fact, one of the earliest papers on wavelet frame based image restoration [11] followed a data-driven intuition, which inspired many later research on the subject. Many algorithms developed afterwards can be understood as having the same algorithmic structure and obeying the same image processing philosophy as [11].

The algorithm of [11], as well as some of the later wavelet frame based image restoration methods, uses a generic prior knowledge of images, i.e. local image features can be sparsely approximated by wavelet frames. However, such generic representation and the associated shrinkage based algorithm may not be ideal for a specifically given image. How can we utilize the prior knowledge on a given image to improve the classical wavelet frame methods that were designed based on generic prior knowledge on images?

Many recent successful frame based methods are adaptive to a specifically given image. These methods respect the features of the observed image and exploit those features in the modeling or algorithmic design to achieve better image restoration results. In this paper, we will discuss three approaches to improve classical wavelet frame based methods by making them adaptive to any specifically given image. One approach is to design tight frames for the specifically given image, instead of using generic tight frames. We shall review two types of tight frames learned from a given image so that the image can be more effectively decomposed to its global patterns plus local features. They are the adaptive tight frames of [26] and the data-driven non-local tight frames of [25]. The second approach is the piecewise smooth image restoration models by [47, 51], whose idea is to actively estimate the locations of local image features so that the underlying wavelet frame system can have a better approximation to global patterns and sparser approximation to local features specifically for a given image. Moreover, the models by [47, 51] even went beyond the scope of low level image processing and opened a door to the unified low and high level modeling for image restoration, object identification and image classification. The third data-driven approach is to design shrinkage operators that are adaptive to the local features of the given image. We will review the adaptive wavelet frame shrinkage algorithm proposed in [6] which also has generic and natural link to nonlinear evolution PDEs.

The paper is organized as follows. In Section 2, we start with a brief review of some basics of wavelet

frames followed by a discussion on how functions and images are approximated by wavelet frames. Then, we discuss wavelet frame based image restoration algorithms and models based on generic data-driven designs. Finally, we show how the same idea of algorithmic design can be applied to problems beyond image restoration, such as matrix completion. In Section 3, we discuss how one can create algorithms and models that are adaptive to a specific image. We will start with the constructoin of adaptive tight frames and data-driven non-local tight frames, followed by the piecewise smooth image restoration models and the adaptive wavelet frame shrinkage algorithm.

2. Generic Data-Driven Algorithms and Models

The key to image restoration is to preserve smooth image components and enhance image features which are normally image singularities. The generic prior knowledge we assume on images is that images have rich local features with some global patterns. The advantage of using wavelet frames for image restoration is their capability of effectively decomposing images to their "global patterns" plus their "local features", or in other words, densely approximating smooth image components and sparsely approximating image features. In this section, we start with a review of some basic concepts of wavelet frames, followed by a detailed description of the "global patterns plus local features" decomposition of images. Then, we describe the basic idea of iterative frame wavelet shrinkage for image denoising. We discuss how a similar idea can be applied to wavelet frame shrinkage for generic image restoration problems or even some data processing problems. These methods are of data-driven nature based on some key observations of the class of images or data to be processed. Therefore, we shall refer to these methods as generic data-driven methods.

2.1. Short Review of Wavelet Frames. In this section, we briefly introduce the concept of wavelet frames. The interested readers should consult [52–55] for theories of frames and wavelet frames, [56] for a short survey on the theory and applications of frames, and [57] for a more detailed survey.

Frame theory and its applications, notably the Gabor frames (see e.g. [1, 2, 54]) and generic wavelet frames (see e.g. [2,54]), were developed long before the discovery of the multiresolution analysis (MRA) of [58,59] and the systematic construction of the MRA-based compactly supported orthonormal wavelets of [60]. The concept of frame can be traced back to [61]. The rich literature of Gabor and wavelet frames provides a wide range of applications including time frequency analysis for signal processing, coherent state in quantum mechanics, filter bank design in electrical engineering, edge and singularity detection in image processing, and etc.

Comprehensive characterization and construction of MRA-based wavelet frames, especially tight wavelet frames, started with the analysis of shift-invariant systems and generalized shift-invariant systems of functions by [52, 53, 62–64]. The duality principle for Gabor frames [63] and the unitary and mixed extension principles for wavelet frames [52, 53] are two central results that follow from this analysis. Wavelet (or affine) systems are not shift-invariant systems and hence the theory of shift-invariant systems by [64] cannot be directly applied. It was shown by [52] that a wavelet system is a tight frame if and only if its corresponding quasi-affine system generated by the same set of functions is a tight frame. This striking fact not only led to a characterization of tight frames without MRA, but also led to the unitary extension principle of [52] which provided a generic characterization of MRA-based tight wavelet frames and made construction of tight wavelet frames painless [55, 65–69]. Further theoretical developments on MRA-based wavelet frames can be found in e.g. [55, 70, 71] and the references therein. More recently in [72], connection between the unitary extension principle and the duality principle was established, which led to a simple construction scheme for MRA-based multivariate tight wavelet frames.

A set $X = \{g_j : j \in \mathbb{Z}\} \subset L_2(\mathbb{R}^d)$, with $d \in \mathbb{N}$, is called a frame of $L_2(\mathbb{R}^d)$ if

$$A\|f\|^2_{L_2(\mathbb{R}^d)} \le \sum_{j\in\mathbb{Z}} |\langle f, g_j\rangle|^2 \le B\|f\|^2_{L_2(\mathbb{R}^d)}, \quad \forall f \in L_2(\mathbb{R}^d),$$

where $\langle\cdot,\cdot\rangle$ is the inner product of $L_2(\mathbb{R}^d)$. We call X a tight frame if it is a frame with $A = B = 1$. For any given frame X of $L_2(\mathbb{R}^d)$, there exists another frame $\widetilde{X} = \{\widetilde{g}_j : j \in \mathbb{Z}\}$ of $L_2(\mathbb{R}^d)$ such that

$$f = \sum_{j\in\mathbb{Z}} \langle f, g_j\rangle \widetilde{g}_j \quad \forall f \in L_2(\mathbb{R}^d).$$

We call $\widetilde{X}$ a dual frame of X. We shall call the pair $(X, \widetilde{X})$ bi-frames. When X is a tight frame, we have

$$f = \sum_{j\in\mathbb{Z}} \langle f, g_j\rangle g_j \quad \forall f \in L_2(\mathbb{R}^d).$$

For given $\Psi = \{\psi_1, \dots, \psi_r\} \subset L_2(\mathbb{R}^d)$, the corresponding quasi-affine system $X(\Psi)$ generated by Ψ is defined by the collection of the dilations and the shifts of Ψ as

$$X(\Psi) = \{\psi_{\ell,n,\boldsymbol{k}} : \ 1 \le \ell \le r; n \in \mathbb{Z}, \boldsymbol{k} \in \mathbb{Z}^d\}, \tag{2.1}$$

where $\psi_{\ell,n,\boldsymbol{k}}$ is defined by

$$\psi_{\ell,n,\boldsymbol{k}} = \begin{cases} 2^{\frac{nd}{2}}\psi_\ell(2^n \cdot -\boldsymbol{k}), & n \ge 0; \\ 2^{nd}\psi_\ell(2^n \cdot -2^{n-J}\boldsymbol{k}), & n < 0. \end{cases} \tag{2.2}$$

When $X(\Psi)$ forms a (tight) frame of $L_2(\mathbb{R}^d)$, each function ψ_ℓ, $\ell = 1, \dots, r$, is called a (tight) framelet and the whole system $X(\Psi)$ is called a (tight) wavelet frame system. Note that in the literature, the affine system is commonly used, which corresponds to the decimated wavelet (frame) transforms. The quasi-affine system, which corresponds to the so-called undecimated wavelet (frame) transforms, was first introduced and analyzed by [52]. Here, we only discuss the quasi-affine system (2.2), since it works better in image restoration and its connection to variational models and PDEs is more natural than the affine system [6, 46, 47].

The constructions of framelets Ψ, which are desirably (anti)symmetric and compactly supported functions, are usually based on a multiresolution analysis (MRA) that is generated by some refinable function ϕ with refinement mask $\boldsymbol{p}$ and its dual MRA generated by $\widetilde{\phi}$ with refinement mask $\widetilde{\boldsymbol{p}}$ satisfying

$$\phi = 2^d \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \boldsymbol{p}[\boldsymbol{k}]\phi(2\cdot -\boldsymbol{k}) \quad \text{and} \quad \widetilde{\phi} = 2^d \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \widetilde{\boldsymbol{p}}[\boldsymbol{k}]\widetilde{\phi}(2\cdot -\boldsymbol{k}).$$

The idea of an MRA-based construction of bi-framelets $\Psi = \{\psi_1, \dots, \psi_r\}$ and $\widetilde{\Psi} = \{\widetilde{\psi}_1, \dots, \widetilde{\psi}_r\}$ is to find masks $\boldsymbol{q}^{(\ell)}$ and $\tilde{\boldsymbol{q}}^{(\ell)}$, which are finite sequences, such that, for $\ell = 1, 2, \dots, r$,

$$\psi_\ell = 2^d \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \boldsymbol{q}^{(\ell)}[\boldsymbol{k}]\widetilde{\phi}(2\cdot -\boldsymbol{k}) \quad \text{and} \quad \widetilde{\psi}_\ell = 2^d \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{k}]\phi(2\cdot -\boldsymbol{k}). \tag{2.3}$$

For a sequence $\{\boldsymbol{p}[\boldsymbol{k}]\}_{\boldsymbol{k}}$ of real numbers, we use $\widehat{\boldsymbol{p}}(\boldsymbol{\omega})$ to denote its Fourier series: $\widehat{\boldsymbol{p}}(\boldsymbol{\omega}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \boldsymbol{p}[\boldsymbol{k}]e^{-i\boldsymbol{k}\cdot\boldsymbol{\omega}}$.

The mixed extension principle (MEP) of [53] provides a general theory for the construction of MRA-based wavelet bi-frames. Given two sets of finitely supported masks $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(r)}\}$ and $\{\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}}_1, \dots, \tilde{\boldsymbol{q}}_r\}$, the MEP says that as long as we have

$$\widehat{\boldsymbol{p}}(\boldsymbol{\xi})\overline{\widehat{\tilde{\boldsymbol{p}}}(\boldsymbol{\xi})} + \sum_{\ell=1}^r \widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi})\overline{\widehat{\tilde{\boldsymbol{q}}}^{(\ell)}(\boldsymbol{\xi})} = 1 \quad \text{and} \quad \widehat{\boldsymbol{p}}(\boldsymbol{\xi})\overline{\widehat{\tilde{\boldsymbol{p}}}(\boldsymbol{\xi}+\boldsymbol{\nu})} + \sum_{\ell=1}^r \widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi})\overline{\widehat{\tilde{\boldsymbol{q}}}^{(\ell)}(\boldsymbol{\xi}+\boldsymbol{\nu})} = 0, \tag{2.4}$$

for all $\boldsymbol{\nu} \in \{0, \pi\}^d \setminus \{\boldsymbol{0}\}$ and $\boldsymbol{\xi} \in [-\pi, \pi]^d$, the quasi-affine systems $X(\Psi)$ and $X(\widetilde{\Psi})$ with Ψ and $\widetilde{\Psi}$ given by (2.3) forms a pair of *bi-frames* in $L_2(\mathbb{R}^d)$. In particular, when $\boldsymbol{p} = \widetilde{\boldsymbol{p}}$ and $\boldsymbol{q}^{(\ell)} = \widetilde{\boldsymbol{q}}^{(\ell)}$ for $\ell = 1, \dots, r$, the MEP (2.4) become the following unitary extension principle (UEP) discovered in [52]:

$$|\widehat{\boldsymbol{p}}(\boldsymbol{\xi})|^2 + \sum_{\ell=1}^r |\widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi})|^2 = 1 \quad \text{and} \quad \widehat{\boldsymbol{p}}(\boldsymbol{\xi})\overline{\widehat{\boldsymbol{p}}(\boldsymbol{\xi}+\boldsymbol{\nu})} + \sum_{\ell=1}^r \widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi})\overline{\widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi}+\boldsymbol{\nu})} = 0, \tag{2.5}$$

and the system $X(\Psi)$ is a *tight frame* of $L_2(\mathbb{R}^d)$. Here, $\boldsymbol{p}$ and $\widetilde{\boldsymbol{p}}$ are lowpass filters and $\boldsymbol{q}^{(\ell)}, \widetilde{\boldsymbol{q}}^{(\ell)}$ are highpass filters. These filters generate discrete bi-frame (or tight frame if UEP is satisfied) system for the sequence space $\ell_2(\mathbb{Z}^d)$. Note that some of the filters involved in the discussions of later sections may only satisfy the first identity of (2.4) or (2.5), and the system generated by these filters shall be called *discrete bi-frame (tight frame) system*. In this case the system generated by the functions associated to these filters does not form a frame or tight frame for $L_2(\mathbb{R}^d)$. However, these filters do form frames or tight frames (undecimated) for $\ell_2(\mathbb{Z}^d)$. Since images are elements in $\ell_2(\mathbb{Z}^d)$, (tight) frames in the sequence space can also be used to efficiently represent images. Furthermore, it was shown in [73] that whenever the first condition of (2.4) is satisfied by the filters, the translation-invariant system, i.e. the system with dyadic dilations and continuous translations, generated by the corresponding wavelet functions forma a frame for $L_2(\mathbb{R}^d)$. Therefore, we shall consider both types of filter banks and refer to them all as bi-frames or tight frames.

Now, we show two simple but useful examples of univariate tight framelets. The framelet given in Example 2.1 is known as the Haar wavelet. Since the quasi-affine system that the Haar wavelet generates is a tight frame of $L_2(\mathbb{R})$, we shall refer to ψ_1 in Example 2.1 as the "Haar framelet". The tight framelets given by Example 2.2 are constructed from piecewise linear B-spline first given by [52]. We shall refer to ψ_1 and ψ_2 in Example 2.2 as "piecewise linear framelets". The framelets constructed by B-splines, especially the piecewise linear framelets, are widely used in wavelet frame based image restoration. In this paper, we shall refer the tight wavelet frame system constructed by [52] as the

B-spline tight wavelet frame system in general.

Example 2.1. Let $\boldsymbol{p} = \frac{1}{2}[1,1]$ be the refinement mask of the piecewise constant B-spline $B_1(x) = 1$ for $x \in [0,1]$ and 0 otherwise. Define $\boldsymbol{q}_1 = \frac{1}{2}[1,-1]$. Then $\boldsymbol{p}$ and $\boldsymbol{q}^{(1)}$ satisfy both identities of (2.5). Hence, the system $X(\psi_1)$ defined in (2.1) is a tight frame of $L_2(\mathbb{R})$.

Example 2.2. [52]. Let $\boldsymbol{p} = \frac{1}{4}[1,2,1]$ be the refinement mask of the piecewise linear B-spline $B_2(x) = \max(1-|x|,0)$. Define $\boldsymbol{q}^{(1)} = \frac{\sqrt{2}}{4}[1,0,-1]$ and $\boldsymbol{q}_2 = \frac{1}{4}[-1,2,-1]$. Then $\boldsymbol{p}$, $\boldsymbol{q}^{(1)}$ and $\boldsymbol{q}^{(2)}$ satisfy both identities of (2.5). Hence, the system $X(\Psi)$ where $\Psi = \{\psi_1, \psi_2\}$ defined in (2.1) is a tight frame of $L_2(\mathbb{R})$.

Notice that all the high-pass filters in Example 2.1 and 2.2 are of the same length as the corresponding low-pass filter. Thus, the wavelet frame functions in the tight frame system have the same support as the corresponding B-splines. This is true for all B-spline tight wavelet frame systems constructed by [52]. Filters having short supports is highly desirable in applications since it means low computation costs and sparse approximation to local features. In fact, for any given refinable box spline whose mask satisfying the 0th-order sum rule, one can always construct a tight frame system with each generator having support at most with the same size as the refinable box spline [72]. More general, for any given refinable function whose mask satisfying the 0th-order sum rule, one can always find a bi-frame system with elements having supports no larger than that of the refinable function [73].

In the discrete setting, let an image $\boldsymbol{f}$ be a d-dimensional array. We denote by $\mathcal{I}_d = \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_d}$ the set of all d-dimensional images. We denote the d-dimensional fast $(L+1)$-level wavelet frame transform/decomposition with filters $\{\boldsymbol{q}^{(0)} = \boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(r)}\}$ (see, e.g., [57]) as

$$\boldsymbol{W}\boldsymbol{u} = \{\boldsymbol{W}_{\ell,l}\boldsymbol{u} : (\ell,l) \in \mathbb{B}\}, \quad \boldsymbol{u} \in \mathcal{I}_d, \tag{2.6}$$

where

$$\mathbb{B} = \{(\ell,l) : \ 1 \le \ell \le r, 0 \le l \le L\} \cup \{(0,L)\}.$$

The wavelet frame coefficients of $\boldsymbol{u}$ are computed by $\boldsymbol{W}_{\ell,l}\boldsymbol{u} = \boldsymbol{q}_{\ell,l}[-\cdot] \circledast \boldsymbol{u}$, where $\circledast$ denotes the convolution operator with a certain boundary condition, e.g., periodic boundary condition, and $\boldsymbol{q}_{\ell,l}$ is defined as

$$\boldsymbol{q}_{\ell,l} = \check{\boldsymbol{q}}_{\ell,l} \circledast \check{\boldsymbol{q}}_{l-1,0} \circledast \ldots \circledast \check{\boldsymbol{q}}_{0,0} \quad \text{with} \quad \check{\boldsymbol{q}}_{\ell,l}[\boldsymbol{k}] = \begin{cases} \boldsymbol{q}^{(\ell)}[2^{-l}\boldsymbol{k}], & \boldsymbol{k} \in 2^l\mathbb{Z}^d; \\ 0, & \boldsymbol{k} \notin 2^l\mathbb{Z}^d. \end{cases} \tag{2.7}$$

Similarly, we can define $\widetilde{\boldsymbol{W}}\boldsymbol{u}$ and $\widetilde{\boldsymbol{W}}_{\ell,l}\boldsymbol{u}$ given a set of dual filters $\{\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}}^{(1)}, \ldots, \tilde{\boldsymbol{q}}^{(r)}\}$. We denote the inverse wavelet frame transform (or wavelet frame reconstruction) as $\widetilde{\boldsymbol{W}}^\top$, which is the adjoint operator of $\widetilde{\boldsymbol{W}}$, and by the MEP, we have the perfect reconstruction formula

$$\boldsymbol{u} = \widetilde{\boldsymbol{W}}^\top \boldsymbol{W}\boldsymbol{u}, \quad \text{for all } \boldsymbol{u} \in \mathcal{I}_d.$$

In particular when $\boldsymbol{W}$ is the transform for a tight frame system, the UEP gives us

$$\boldsymbol{u} = \boldsymbol{W}^\top \boldsymbol{W}\boldsymbol{u}, \quad \text{for all } \boldsymbol{u} \in \mathcal{I}_d. \tag{2.8}$$

For simplicity, we will mostly focus our discussions on the case $d = 2$, i.e. for 2-dimensional images.

2.2. Approximation.

Ideally speaking, the system we use to approximate images should be able to effectively capture both global patterns and local features of images. We need to have a system, or transform, $\boldsymbol{W}$ that has two subsystems, $\boldsymbol{W}_{\mathrm{P}}$, and $\boldsymbol{W}_{\mathrm{F}}$. The subsystem $\boldsymbol{W}_{\mathrm{P}}$ is specialized to approximate the global patterns. To well capture the global patterns, it is desirable for the atoms of subsystem $\boldsymbol{W}_{\mathrm{P}}$ to have global supports. However, having globally supported atoms is not computationally efficient. When $\boldsymbol{W}_{\mathrm{P}}$ is locally supported instead, the global features can only be well characterized by dense representations of $\boldsymbol{W}_{\mathrm{P}}$. In order to well approximate local features and to distinguish them from global patterns, atoms in $\boldsymbol{W}_{\mathrm{F}}$ should be locally supported, and the local features can be sparsely approximated by $\boldsymbol{W}_{\mathrm{F}}$. Moreover, in order to reduce the artifacts of the approximation, the atoms in the systems $\boldsymbol{W}_{\mathrm{P}}$ and $\boldsymbol{W}_{\mathrm{F}}$ should not be entirely independent from each other. The corresponding spaces they approximate need to be partially overlapped. Therefore, most transformations used in image restoration are redundant. The simplest type of redundant systems is tight frames. In general, for given data $\boldsymbol{u}$ in some class, we would like to have

$$\boldsymbol{W}\boldsymbol{u} = \begin{pmatrix} \boldsymbol{W}_{\mathrm{P}}\boldsymbol{u} \\ \boldsymbol{W}_{\mathrm{F}}\boldsymbol{u} \end{pmatrix} = \begin{pmatrix} \text{Dense} \\ \text{Sparse} \end{pmatrix}.$$

When we restore images in transform domain, it is often required to be able to recover $\boldsymbol{u}$ from $\boldsymbol{W}\boldsymbol{u}$. To make reconstruction of $\boldsymbol{u}$ simple, it is good to have the system $\boldsymbol{W}$ to be a tight frame, i.e. $\boldsymbol{W}^T\boldsymbol{W} = I$.

There are many such systems and tight wavelet frame system is one of the good ones.

Now, we discuss how wavelet frames approximate functions in the fashion of "global patterns plus local features". For simplicity, we restrict our discussions on quasi-affine tight wavelet frame systems constructed from the UEP (2.5). Similar arguments apply to bi-frames constructed from MEP (2.4) as well.

We start with the approximation of function $f \in L_2(\mathbb{R}^d)$. Let $\Psi = \{\psi_\ell : 1 \le \ell \le r\} \subset L_2(\mathbb{R}^d)$ be the set of compactly supported tight framelets constructed from the UEP (2.5) with $\phi \in L_2(\mathbb{R}^d)$ the corresponding compactly supported refinable function. Then, for any given integer $N \in \mathbb{Z}$, the system

$$X^*(\phi, \Psi; N) = \{\phi_{N,\boldsymbol{k}}, \psi_{\ell,n,\boldsymbol{k}} : 1 \le \ell \le r, n \ge N, \boldsymbol{k} \in \mathbb{Z}^d\}$$

forms a tight frame of $L_2(\mathbb{R}^d)$, i.e. for any $f \in L_2(\mathbb{R}^d)$,

$$f = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle f, \phi_{N,\boldsymbol{k}}\rangle \phi_{N,\boldsymbol{k}} + \sum_{\ell=1}^{r} \sum_{n\ge N} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle f, \psi_{\ell,n,\boldsymbol{k}}\rangle \psi_{\ell,n,\boldsymbol{k}}, \tag{2.9}$$

where $\phi_{n,\boldsymbol{k}}$ and $\psi_{\ell,n,\boldsymbol{k}}$ are defined by (2.2) (see e.g. [57] for details). For convenience, let

$$\mathcal{L}_N f = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle f, \phi_{N,\boldsymbol{k}}\rangle \phi_{N,\boldsymbol{k}} \quad \text{and} \quad \mathcal{H}_{\ell,n} f = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \langle f, \psi_{\ell,n,\boldsymbol{k}}\rangle \psi_{\ell,n,\boldsymbol{k}}.$$

Then, (2.9) can be written as

$$f = \mathcal{L}_N f + \sum_{\ell=1}^{r} \sum_{n\ge N} \mathcal{H}_{\ell,n} f. \tag{2.10}$$

The term $\mathcal{L}_N f$ on the right hand side of (2.10) is the quasi-interpolatory projection of f into the shift-invariant space generated by $\phi_{N,\boldsymbol{0}}$. In other words, the term $\mathcal{L}_N f$ a smooth approximation of f at scale N, which represents the glocal patterns of function f. The term $\mathcal{H}_{\ell,n} f$ is a sparse approximation of f in band ℓ and at scale n. For different band ℓ, $\mathcal{H}_{\ell,n} f$ represents different type of local features (or singularities) of f, such as jump discontinuities and hidden jump discontinuities (see Section 3.2 and [6] for more details). For different scale n, $\mathcal{H}_{\ell,n} f$ represents features of f at different scales. Therefore, a tight wavelet frame system constructed from UEP (or more generally wavelet bi-frame systems constructed from MEP) provides the following decomposition for a given function $f \in L_2(\mathbb{R}^d)$:

$$f = \text{ Global Patterns } + \text{ Local Features }.$$

In discrete setting, we consider $\boldsymbol{u} \in \mathcal{I}_d$. Let $\boldsymbol{W}$ be the tight wavelet frame transform defined by (2.6). Define $\boldsymbol{W}_{\mathcal{L}} = \boldsymbol{W}_{0,L}$ and

$$\boldsymbol{W}_{\mathcal{H}}^\top = \left(\boldsymbol{W}_{1,0}^\top, \cdots, \boldsymbol{W}_{r,0}^\top, \boldsymbol{W}_{1,1}^\top, \cdots, \boldsymbol{W} r, L^\top\right)^\top.$$

For wavelet frame systems, the global patterns are captured by $\boldsymbol{W}_{\mathcal{L}}$, i.e. $\boldsymbol{W}_{\mathrm{P}} = \boldsymbol{W}_{\mathcal{L}}$; and the local features are captured by $\boldsymbol{W}_{\mathcal{H}}$, i.e. $\boldsymbol{W}_{\mathrm{F}} = \boldsymbol{W}_{\mathcal{H}}$ (see Figure 2).

The perfect reconstruction formula (2.8) can be written equivalently as

$$\boldsymbol{u} = \boldsymbol{W}_{\mathcal{L}}^\top \boldsymbol{W}_{\mathcal{L}} \boldsymbol{u} + \boldsymbol{W}_{\mathcal{H}}^\top \boldsymbol{W}_{\mathcal{H}} \boldsymbol{u} = \boldsymbol{W}_{0,L}^\top \boldsymbol{W}_{0,L} \boldsymbol{u} + \sum_{\ell=1}^{r} \sum_{l=0}^{L} \boldsymbol{W}_{\ell,l}^\top \boldsymbol{W}_{\ell,l} \boldsymbol{u}. \tag{2.11}$$

The decomposition (2.11) is the discrete analogue to (2.10), where $\boldsymbol{W}_{\mathcal{L}}^\top \boldsymbol{W}_{\mathcal{L}} \boldsymbol{u}$ is the smooth approximation of $\boldsymbol{u}$ while $\boldsymbol{W}_{\mathcal{H}}^\top \boldsymbol{W}_{\mathcal{H}} \boldsymbol{u}$ is the sparse component of $\boldsymbol{u}$ and $\boldsymbol{W}_{\ell,l}^\top \boldsymbol{W}_{\ell,l} \boldsymbol{u}$ represents the local features of $\boldsymbol{u}$ in band ℓ and at scale l. Therefore, (2.11) shows the same "global patterns plus local features" type of decomposition as (2.10). To illustrate the decomposition (2.11), we present a numerical simulation in Figure 1, where the tight wavelet frame system used is the piecewise linear B-spline tight wavelet frame system with filters given by Example 2.2.

Given an image $\boldsymbol{u}$, its global patterns are represented by $\boldsymbol{W}_{\mathcal{L}}^\top \boldsymbol{W}_{\mathcal{L}} \boldsymbol{u}$. Global patterns are normally slow varying image components that mainly occupy the lower frequency bands in Fourier domain. They can be regarded as a smooth approximation of the original image that provide us with a global view of the image. The local features, of $\boldsymbol{u}$ is represented by $\boldsymbol{W}_{\mathcal{H}}^\top \boldsymbol{W}_{\mathcal{H}} \boldsymbol{u}$. Local image features are normally image singularities which typically include edges, which are jump discontinuities, and hidden edges, which are jump discontinuities after 1st order differentiations. For images that are discrete realizations of piecewise smooth functions and when compactly supported wavelet frame systems are used, the coefficients $\boldsymbol{W}_{\mathcal{L}} \boldsymbol{u}$ typically form a dense array, while the coefficients $\boldsymbol{W}_{\mathcal{H}} \boldsymbol{u}$ typically form a sparse array. The large (in magnitude) coefficients $\boldsymbol{W}_{\mathcal{H}} \boldsymbol{u}$ only appear near image features when the supports of the associated high-pass filters intersect with these image features. Therefore, the "global patterns plus

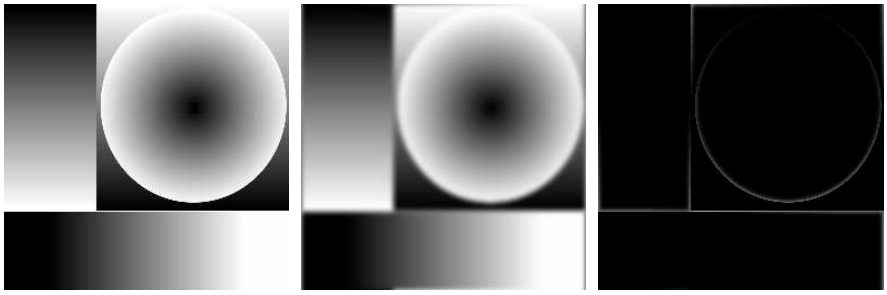


Figure 1. This figure shows the decomposition: $\boldsymbol{u} = \boldsymbol{W}_{\mathcal{L}}^{\top}\boldsymbol{W}_{\mathcal{L}}\boldsymbol{u} + \boldsymbol{W}_{\mathcal{H}}^{\top}\boldsymbol{W}_{\mathcal{H}}\boldsymbol{u}$.

local features" type of decomposition can also be stated in transform domain as the "dense plus sparse approximation" of images. The global image patterns can be densely approximated by the coefficients $\boldsymbol{W}_{\mathcal{L}}\boldsymbol{u}$, while the local image features can be sparsely approximated by the coefficients $\boldsymbol{W}_{\mathcal{H}}\boldsymbol{u}$. To illustrate such decomposition, we present a 1-level wavelet frame decomposition of a simple piecewise smooth image in Figure 2. The tight wavelet frame system used is the piecewise linear B-spline tight wavelet frame system with filters given by Example 2.2.

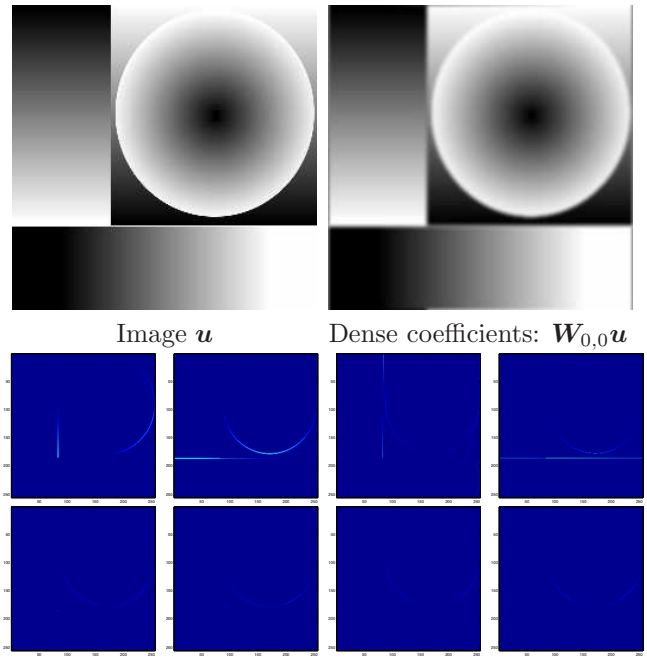


Figure 2. Demonstration of a 1-level tight wavelet frame transform.

It is natural to require a transformation to be capable of decomposing an image to its global patterns plus its local features, or in other words, dense plus sparse coefficients in transform domain. Transformations that can only well capture either global patterns or local features may not be as effective in image restoration as the transformations that can capture both. Most of the recent designs of transforms or systems for image restoration have only been emphasizing on sparsity (or sparse approximation) due to the success and significant impact of compressed sensing [74–77]. Although sparsity, or capturing local features, is important, we should not completely ignore the ability of the underlying system to capture global features. After all, there are two components to make a successful image restoration model: smoothness of global patterns and sharpness of local features. The piecewise smooth image restoration model [47, 51] is a good example, where it was shown that a better characterization of global patterns does make a difference in practice.

Other than wavelets and wavelet frames, windowed Fourier transforms provide a "global patterns plus local features" type of decomposition of images as well. However, the sparse coefficients in the

transform domain are not ideally sparse for images due to their fixed time-frequency resolution. Orthonormal or biorthogonal wavelets provide better sparse approximation to local features than windowed Fourier transforms, which is crucial to their success in image compression. However, for image restoration, redundant systems such as wavelet frames can provide better sparse approximation to local image features than (bi)orthogonal systems. Wavelet frame systems can better balance between smoothness and sparsity so that artifacts generated by the Gibbs phenomenon can be further reduced, which in turn leads to better image restoration results. Also, the sparse approximation to the local image features provided by $\boldsymbol{W}_{\mathcal{H}}^{\top}\boldsymbol{W}_{\mathcal{H}}\boldsymbol{u}$ will be more effective if the associated high-pass filters have varied orders of vanishing moments specializing in capturing image features of different types, with different orientations and at different scales. Therefore, a system with good sparse approximation to local image features has to be redundant.

Note that "local" features are not limited to localness in spatial domain. Image features can be local in similarity domain as well. For example, two pixels that are spatially farer apart may have similar patches, i.e. they are close in similarity domain. One transform that uses such generalized concept of localness is the data-driven non-local tight frames constructed by [25], which will be reviewed later in Section 3.1. Also, since only the wavelet frame coefficients $\boldsymbol{W}_{\mathcal{H}}\boldsymbol{u}$ are sparse, the shrinkage operation normally only applies to $\boldsymbol{W}_{\mathcal{H}}\boldsymbol{u}$ while $\boldsymbol{W}_{\mathcal{L}}\boldsymbol{u}$ is left untouched. However, we may receive benefits from processing $\boldsymbol{W}_{\mathcal{L}}\boldsymbol{u}$ for image segmentation as demonstrated in [33], where the authors showed that properly defined shrinkage on the dense coefficients can significantly speed up the algorithm.

2.3. Image Denoising. Image denoising is one of the most classical image restoration problems whose corresponding model takes the form of (1.1) with $\boldsymbol{A} = \boldsymbol{I}$. Its main objective is to regularize the given image and preserve sharp image features (e.g. edges) at the same time. The challenge, however, is that it is not straightforward on how to separate smooth and sharp regions of a given image, especially at the presence of noise, so that suitable treatments can be applied to them separately. Therefore, wavelet frame is the right tool for image denoising since in transform domain image features are well separated from smooth image components.

One of the earliest successful image denoising methods is the shift-invariant wavelet soft-thresholding algorithm by [78]. The shift-invariant wavelet systems were later shown in [52] to be tight wavelet frame systems for the function space $L_2(\mathbb{R}^d)$. In fact, they are special tight wavelet frame systems since they are quasi-affine systems (see (2.1) and (2.2)) generated by orthonormal wavelets. More general tight wavelet frames can be constructed by the UEP of [52] such as the B-spline tight wavelet frame systems. The main idea of the algorithm by [78] is to first take wavelet frame transform of the given noisy image, conduct thresholding to the sparse coefficients in the transform domain, and transform the data back to image domain by taking the inverse wavelet transform.

Let $\boldsymbol{f}$ be the observed noisy image. The generic wavelet frame thresholding algorithm for image denoising can be written as

$$\boldsymbol{u}^{\star} = \widetilde{\boldsymbol{W}}^{\top}\mathcal{T}_{\boldsymbol{\lambda}}(\boldsymbol{W}\boldsymbol{f}), \tag{2.12}$$

where $\boldsymbol{W}$ is the wavelet frame transform and $\widetilde{\boldsymbol{W}}$ is the dual wavelet frame transform such that $\widetilde{\boldsymbol{W}}^{\top}\boldsymbol{W} = \boldsymbol{I}$. Examples of the thresholding operator can be the soft-thresholding operator [79] defined componentwise on the coefficients $\boldsymbol{\alpha} = \{\alpha_{\ell,l,\boldsymbol{k}} = (\boldsymbol{W}_{\ell,l}\boldsymbol{u})[\boldsymbol{k}] : \ (\ell,l) \in \mathbb{B}, \boldsymbol{k} \in \boldsymbol{\Omega}\}$ as

$$\mathcal{T}_{\boldsymbol{\lambda}}^{s}(\boldsymbol{\alpha}) = \left\{ \mathcal{T}_{\lambda_{\ell,l,\boldsymbol{k}}}^{s}(\alpha_{\ell,l,\boldsymbol{k}}) = \frac{\alpha_{\ell,l,\boldsymbol{k}}}{|\alpha_{\ell,l,\boldsymbol{k}}|}\max\{|\alpha_{\ell,l,\boldsymbol{k}}| - \lambda_{\ell,l,\boldsymbol{k}}, 0\} : \ \boldsymbol{k} \in \boldsymbol{\Omega} \right\}. \tag{2.13}$$

It can also be the hard-thresholding operator [8, 9, 14, 15, 79–81] defined as

$$\mathcal{T}_{\boldsymbol{\lambda}}^{h}(\boldsymbol{\alpha}) = \left\{ \mathcal{T}_{\lambda_{\ell,l,\boldsymbol{k}}}^{h}(\alpha_{\ell,l,\boldsymbol{k}}) : \ (\ell,l) \in \mathbb{B}, \boldsymbol{k} \in \boldsymbol{\Omega} \right\},$$

where

$$\mathcal{T}_{\lambda_{\ell,l,\boldsymbol{k}}}^{h}(\alpha_{\ell,l,\boldsymbol{k}}) = \begin{cases} \alpha_{\ell,l,\boldsymbol{k}} & \text{if } |\alpha_{\ell,l,\boldsymbol{k}}| > \lambda_{\ell,l,\boldsymbol{k}}, \\ \{\alpha_{\ell,l,\boldsymbol{k}}, 0\} & \text{if } |\alpha_{\ell,l,\boldsymbol{k}}| = \lambda_{\ell,l,\boldsymbol{k}}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.14}$$

Since thresholding is to promote sparsity of the sparse coefficients, we choose $\lambda_{0,L,\boldsymbol{k}} = 0$ as convention, i.e. the dense coefficients are not processed by the thresholding operator. Note that the thresholding operator $\mathcal{T}_{\boldsymbol{\lambda}}$ can be more sophisticated than soft- or hard-thresholding. For example, it can be an adaptive multiplicative shrinkage proposed by [6], or an adaptive soft-thresholding which will be presented in Section 3.3.

The key that makes the thresholding algorithm 2.12 work well is the fact that wavelet frame transforms effectively decompose images to their global patterns plus local features, or in other words, they can densely approximate smooth image components and sparsely approximate image features. Then after thresholding, large sparse coefficients that encode crucial image features are well preserved, while the rest of the sparse coefficients are set to zero without introducing much error since most of the small

nonzero sparse coefficients correspond to noise instead of signal. The dense coefficients are left untouched since the regularity of the smooth image components is automatically granted by the low-pass filtering. More precisely speaking, the approximated global patterns of the recovered image is given by $\boldsymbol{W}_{\mathcal{L}}^{\top}\boldsymbol{W}_{\mathcal{L}}\boldsymbol{f}$. Another way to understand this is that by skipping processing of the dense coefficients, we implicitly let the global patterns of the recovered image directly follow the data $\boldsymbol{f}$. Since $\boldsymbol{W}_{\mathcal{L}}^{\top}\boldsymbol{W}_{\mathcal{L}}\boldsymbol{f}$ is generally a good approximation to global patterns, removing noise from sparse coefficients is enough.

The performance of the above thresholding algorithm highly depends on the quality of the sparse approximation to local image features provided by the underlying wavelet frame system. Although orthogonal or biorthogonal wavelets provide sparse approximation to local image features, redundant systems such as wavelet frames can provide better sparse approximation which in turn leads to better image restoration results. One advantage of wavelet frames (e.g. the B-spline tight wavelet frame systems of [52]) over (bi)orthogonal wavelets, in terms of sparse approximation, is the presence of wavelet functions in the system with varied vanishing moments and generally shorter supports. The varied orders of vanishing moment enables the wavelet frame system to posses subsystems specialized in the sparse approximation of different types of singularity, such as jump discontinuities, jumps after first order differentiation (ridges). The shorter supports of wavelet frames lead to more concentrated coefficients in the transform domain. Another advantage of wavelet frames over (bi)orthogonal wavelets are their robustness to errors, thanks to the redundancy of these systems. After thresholding in the transform domain, errors are inevitably introduced no matter how careful the thresholding operator is designed. However, if a redundant system is used, there is a good chance that these errors are canceled out after transforming back to image domain. From the point of view of information recovery, if some of the coefficients of a redundant system is damaged or missing, due to redundancy, that same piece of information may exist in other coefficients as well. However, if the underlying system is orthogonal or biorthogonal, each coefficient carries a unique piece of information. Thus, the lost information from a coefficient cannot be recovered from other coefficients.

The original design of image denoising method of [78], as well as the generic image restoration algorithm by [11] that will be discussed in the next subsection, is based on wavelet frame thresholding (2.12). The intuition behind these algorithms is the effective dense and sparse approximation of global patterns and local features in transform domain. Then, thresholding, such as soft- or hard-thresholding, of sparse coefficients can enhance image features while removing noise at the same time. It was shown later in the literature that the algorithm (2.12) can be written equivalently as certain optimization models. It was shown by [26] that when $\boldsymbol{W}$ corresponds to a tight wavelet frame system, i.e. $\boldsymbol{W}^{\top}\boldsymbol{W}=\boldsymbol{I}$, and soft-thresholding is used, we have

$$\boldsymbol{u}^{\star}=\boldsymbol{W}^{\top}\left(\arg\min_{\boldsymbol{\alpha}}\frac{1}{2}\|\boldsymbol{W}^{\top}\boldsymbol{\alpha}-\boldsymbol{f}\|_2^2+\frac{1}{2}\|(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{W}^{\top})\boldsymbol{\alpha}\|_2^2+\|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_1\right), \tag{2.15}$$

where $\boldsymbol{u}^{\star}$ is given by (2.12). When hard-thresholding is used instead, we have [35]

$$\boldsymbol{u}^{\star}=\boldsymbol{W}^{\top}\left(\arg\min_{\boldsymbol{\alpha}}\frac{1}{2}\|\boldsymbol{W}^{\top}\boldsymbol{\alpha}-\boldsymbol{f}\|_2^2+\frac{1}{2}\|(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{W}^{\top})\boldsymbol{\alpha}\|_2^2+\frac{1}{2}\|\boldsymbol{\lambda}^2\cdot\boldsymbol{\alpha}\|_0\right), \tag{2.16}$$

In particular, when $\boldsymbol{W}$ is the corresponding transformation of an orthonormal wavelet system, i.e. $\boldsymbol{W}\boldsymbol{W}^{\top}=\boldsymbol{W}^{\top}\boldsymbol{W}=\boldsymbol{I}$, the identities (2.15) and (2.16) reduce to

$$\boldsymbol{u}^{\star}=\arg\min_{\boldsymbol{u}}\frac{1}{2}\|\boldsymbol{u}-\boldsymbol{f}\|_2^2+\|\boldsymbol{\lambda}\cdot\boldsymbol{W}\boldsymbol{u}\|_1, \tag{2.17}$$

and

$$\boldsymbol{u}^{\star}=\arg\min_{\boldsymbol{u}}\frac{1}{2}\|\boldsymbol{u}-\boldsymbol{f}\|_2^2+\frac{1}{2}\|\boldsymbol{\lambda}^2\cdot\boldsymbol{W}\boldsymbol{u}\|_0. \tag{2.18}$$

The term $\|(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{W}^{\top})\boldsymbol{\alpha}\|_2^2$ in (2.15) and (2.16) tries to minimize the distance between $\boldsymbol{\alpha}$ and the range of $\boldsymbol{W}$, which essentially controls the regularity of the reconstructed image. In other words, this term balances between regularity of $\boldsymbol{u}$ and sparsity of the wavelet frame coefficients $\boldsymbol{\alpha}$ to achieve a desirable reconstruction. Hence, the optimization model in (2.15) and (2.16) is called the balanced model and is used for generic image restoration problems as well [10, 82–84].

If we introduce a tuning parameter κ to the middle term of (2.15) and (2.16), we have the following slightly more general form of the balanced model

$$\min_{\boldsymbol{\alpha}}\frac{1}{2}\|\boldsymbol{W}^{\top}\boldsymbol{\alpha}-\boldsymbol{f}\|_2^2+\frac{\kappa}{2}\|(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{W}^{\top})\boldsymbol{\alpha}\|_2^2+\frac{1}{2}\|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_p\qquad p=0,1. \tag{2.19}$$

When $\kappa=0$ in (2.19), we have the following synthesis model

$$\min_{\boldsymbol{\alpha}}\frac{1}{2}\|\boldsymbol{W}^{\top}\boldsymbol{\alpha}-\boldsymbol{f}\|_2^2+\frac{1}{2}\|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_p\qquad p=0,1. \tag{2.20}$$

When $\kappa = \infty$ in (2.19), we have the following analysis model

$$\min_{\boldsymbol{u}} \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{f}\|_2^2 + \frac{1}{2}\|\boldsymbol{\lambda} \cdot \boldsymbol{W}\boldsymbol{u}\|_p \qquad p = 0, 1. \tag{2.21}$$

Although (2.20) and (2.21) look simpler than (2.19), they do not have a closed-form solution (unless $\boldsymbol{W}$ is orthonormal), or in other words, their solutions need to be approximated through an iterative algorithm. Therefore, the balanced model (2.19) is more natural than synthesis or analysis model in the sense that its closed-form solution takes the form of the thresholding scheme (2.12) which comes from a data-driven approach. Note that when $\boldsymbol{W}$ is orthonormal, model (2.19), (2.20) and (2.21) are all equivalent and take (2.12) as the closed-form solution. Therefore, when orthonormal wavelets are used, all three models are natural. We will resume the discussions on the three models in the next section for generic image restoration problem.

We note that if a general shrinkage operator $\mathcal{S}_{\boldsymbol{\lambda}}$ is used in the shrinkage algorithm (2.12), the algorithm may not correspond to any optimization model. However, finding an optimization model for wavelet frame based image denoising or more general image restoration problems has never been the starting point of wavelet (frame) based approach. What is crucial for wavelet (frame) based approach is their capability to separate image features from smooth image components which leads to dense approximation of global image patterns and sparse approximation of local images features. Then, properly designed shrinkage, not necessarily the soft- or hard-thresholding, in the wavelet (frame) domain is able to recover image features and remove noise simultaneously.

The thresholding scheme (2.12) provides a good low level approximation of the solution to image denoising based on generic knowledge of the underlying solution, i.e. local image features can be sparsely approximated by wavelet frames. When some special shrinkage operators are used, the algorithm led to optimization models that were later generalized to solve generic image restoration problems. To move from low level approximation to more accurate approximations or even high level understanding of images, we need to make the shrinkage scheme adaptive to a specifically given image and automatically readjust the scheme according to the approximate solution at each step to gradually increase accuracy of the approximate solution. The algorithmic structure of (2.12), as well as the more general 3-steps procedure in Section 2.4, is a perfect vessel to realize such data-driven idea.

2.4. General Image Restoration Problems. For a generic image restoration problem, i.e. when $\boldsymbol{A} \neq \boldsymbol{I}$, the simple procedure of thresholding of sparse coefficients in the wavelet frame domain and transforming back to image domain does not work since we now have a non-trivial linear system $\boldsymbol{A}$ to solve. However, the same idea of thresholding in the transform domain using a good representation of images still applies. Wavelet frame based method for more general image restoration problems started from [11], where high-resolution image reconstruction was studied. By viewing the high-resolution image reconstruction as an inpainting problem in the wavelet frame domain, an iterative algorithm by applying thresholding to sparse wavelet frame coefficients at each iteration to preserve sharp edges and to remove noises in the image was proposed. What's even more important of the pioneer work of [11], as well as [79], is the image restoration philosophy they convey. That is finding good "dense plus sparse" approximation at each iterate for approximate solution by applying suitable shrinkage operators in the transform domain that removes noise and enhances image features.

Finding approximated solutions to linear system (1.1) has a much longer history than wavelet frames and has been fairly well-developed. Many of the existing iterative solvers mainly contain two components in the algorithms, i.e. finding an approximate solution to the linear system, and update residual so that more accurate solution can be found in the next iteration. It is generally easy to find an approximate solution to (1.1). However, due to ill-pose nature of the problem (1.1), recovered image from a linear system solver may either have image features, such as edges, smeared out or noise magnified. It is generally difficult to preserve image features and remove noise at the same time by only relying on linear system solvers. We have to use our prior knowledge on images, i.e. wavelet frames provide dense approximation to global image patterns and sparse approximation to local images features. This is the key to the success of the algorithm proposed by [11], where thresholding of sparse coefficients in the wavelet frame domain is embedded to a linear system solver at each iteration. Therefore, the linear system is inverted while the quality of the recovered image is also guaranteed since important images features are sharpened while noise is removed by iterative thresholding in wavelet frame domain.

The main idea of the iterative wavelet frame shrinkage algorithm of [11] boils down to the following three recursive steps:

3-Steps Procedure:

1. Find an approximate solution $\boldsymbol{u}^1$ to the linear system (1.1);
2. Shrinkage in wavelet frame domain: $\boldsymbol{u}^2 = \widetilde{\boldsymbol{W}}^\top \mathcal{S}_{\boldsymbol{\lambda}}(\boldsymbol{W}\boldsymbol{u}^1)$;
3. Update the residual and iterate.

For step 1, One may use any approximated solver for the linear system (1.1). Solution $\boldsymbol{u}^1$ from step 1 may not be a desirably reconstructed image due to ill-posed nature of the system $\boldsymbol{A}$ and the presence of noise $\boldsymbol{\eta}$. For example, one may observe blurred edges and oscillations in smooth regions in $\boldsymbol{u}^1$. Step 2 is the key to the 3-steps procedure which is designed to suppress noise and recover sharp edges and other image singularities by performing shrinkage operations on sparse coefficients in wavelet frame domain. The key point here is the data-driven shrinkage design in transform domain. For example, if the image has a sparse approximation in transform domain, then, the simplest shrinkage is thresholding. Step 3 is to update residual which usually takes the form of updating an auxiliary variable which will be fed back to Step 1 in the next iteration. After edge sharpened and noise removed, the new solution $\boldsymbol{u}^2$ may no longer be a good approximate solution of (1.1). Therefore, we return to step 1 and find a new approximate solution based on the current solution $\boldsymbol{u}^2$. We keep iterating the three steps until convergence.

Note that the order of the three steps may vary for different algorithms, and some time step 3 may be merged with step 1. The transforms $\boldsymbol{W}$ and $\widetilde{\boldsymbol{W}}$ are not only limited to wavelet frame transforms. They can be whatever transform that provides sparse approximation to the local features of the data we want to recover, e.g. the singular value decomposition for low rank matrix completion [85] and the data-driven transforms that will be reviewed in Section 3.1. Different types of shrinkage operator $\mathcal{S}_{\boldsymbol{\lambda}}$ can be used in the above procedure, such as the soft- and hard-thresholding which were both used in [11]. Different shrinkage operator may significantly affect the quality of image restoration (see e.g. [11,14,15]). However, the shrinkage operator $\mathcal{S}_{\boldsymbol{\lambda}}$ is not only limited to soft- or hard-thresholding operators. It may be multiplicative or/and adaptive to the observed image $\boldsymbol{f}$ and the current approximate solution $\boldsymbol{u}^1$ (see Section 3.3). Furthermore, the 3-steps procedure may also work for nonlinear inverse problems as long as one has a reliable nonlinear system solver and combines it with shrinkage in transform domain. There may be other variants to the current form of the 3-steps procedure. However, what is crucial to the algorithm is the step of shrinkage in transform domain where the local features be recovered can be sparsely approximated.

Many wavelet frame based image restoration algorithms can be understood as having the same algorithmic structure as the 3-steps procedure. We start with the algorithm for image inpainting. The mathematical model for image inpainting can be stated as follows (see e.g. [86, 87]). Let the original image $\boldsymbol{u} \in \mathbb{R}^{m\times n}$ be defined on the domain $\boldsymbol{\Omega} = \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$ and the nonempty set $\boldsymbol{\Lambda} \subsetneq \boldsymbol{\Omega}$ be the given observed region. Then the observed (incomplete) image $\boldsymbol{f}$ is

$$\boldsymbol{f}[\boldsymbol{k}] = \begin{cases} \boldsymbol{u}[\boldsymbol{k}] + \boldsymbol{\eta}[\boldsymbol{k}], & \boldsymbol{k} \in \boldsymbol{\Lambda}, \\ \text{arbitrary}, & \boldsymbol{k} \in \boldsymbol{\Omega} \setminus \boldsymbol{\Lambda}, \end{cases} \tag{2.22}$$

The goal is to find $\boldsymbol{u}$ from the its partially observed noisy pixels $\boldsymbol{f}$. The corresponding linear system $\boldsymbol{A}$ for image inpainting is simply a restriction operator restricting $\boldsymbol{u} \in \mathbb{R}^{m\times n}$ to its subindex set $\boldsymbol{\Lambda} \subsetneq \boldsymbol{\Omega}$. We denote the linear system for inpainting as $\boldsymbol{A}_{\boldsymbol{\Lambda}}$.

The simple idea of the wavelet frame based image inpainting algorithm comes as follows [11]: one may use any simple interpolation scheme to interpolate the given data that leads to an inpainted image. Edges might be blurred and noise is still present in the inpainted image. One of the simplest ways to sharpen the image and remove noise at the same time is to set small sparse coefficients to zero. When we reconstruct the image using the modified wavelet frame coefficients, it will no longer interpolate the data, and the simplest way to make it interpolate the given data is to correct the inpainted image using observed data. Then, we iterate this procedure till convergence. The algorithm reads as follows, where we assume tight wavelet frame systems are used, i.e. the transform $\boldsymbol{W}$ satisfies $\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}$.

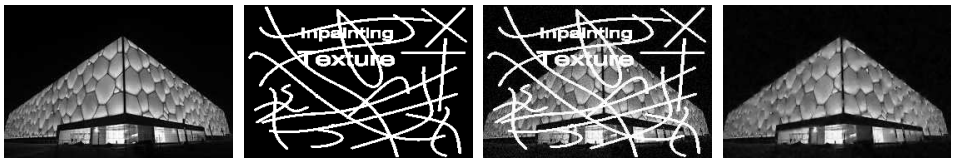
Algorithm (Wavelet Frame Based Image Inpainting): Initialize $\boldsymbol{u}^0 = \boldsymbol{0}$. Iterate the following steps for $k = 1, 2, \ldots$ until convergence:

$$\begin{cases} \boldsymbol{u}^k = (\boldsymbol{I} - \mu \boldsymbol{A}_{\boldsymbol{\Lambda}}^\top \boldsymbol{A}_{\boldsymbol{\Lambda}}) \boldsymbol{u}^{k-1} + \mu \boldsymbol{A}_{\boldsymbol{\Lambda}}^\top \boldsymbol{f}, \\ \boldsymbol{u}^{k+1} = \boldsymbol{W}^\top \mathcal{T}_{\boldsymbol{\lambda}}^s (\boldsymbol{W} \boldsymbol{u}^k). \end{cases} \tag{2.23}$$

Recall that $\mathcal{T}_{\boldsymbol{\lambda}}^s(\boldsymbol{\alpha})$ is the soft-thresholding operator [79] defined by (2.13). Algorithm (2.23) takes the form of the 3-steps procedure. The first step of (2.23) consists of both the step of update residual, i.e. $\boldsymbol{r}^k = \boldsymbol{A}_{\boldsymbol{\Lambda}} \boldsymbol{u}^{k-1} - \boldsymbol{f}$, and finding an approximate solution for (1.1), i.e. $\boldsymbol{u}^k = \boldsymbol{u}^{k-1} - \mu \boldsymbol{A}_{\boldsymbol{\Lambda}}^\top \boldsymbol{r}^k$. Note that the inpainting algorithm (2.23) applies to generic image restoration problem (1.1) as well. One simply needs to replace $\boldsymbol{A}_{\boldsymbol{\Lambda}}$ by the associated linear system to a given image restoration problem. Also, the algorithm can be significantly speeded up by using a more sophisticated update of the approximate solution $\boldsymbol{u}^k$ in step 1 of (2.23) (see [88, 89]).

Here, we present one numerical simulation of the algorithm (2.23) on an inpainting problem. The unknown original image is shown in Figure 3(a). The inpainting domain $\boldsymbol{\Lambda}$ where the pixel values are missing is shown in Figure 3(b). The observed image $\boldsymbol{f}$ is shown in Figure 3(c), where minor noise $\boldsymbol{\eta}$ is added in region $\boldsymbol{\Lambda}^c$. The reconstruction result using algorithm (2.23) is presented in Figure 3(d).

The convergence of algorithm (2.23) was analyzed by [10]. It turns out that if we let $\boldsymbol{\alpha}^k = \mathcal{T}_{\boldsymbol{\lambda}}^s(\boldsymbol{W}\boldsymbol{u}^k)$,



(a) (b) (c) (d)

Figure 3. Images from left to right are: original image, inpainting domain $\boldsymbol{\Lambda}$, observed image $\boldsymbol{f}$ and reconstructed image using algorithm (2.23).

the sequence $\boldsymbol{\alpha}^k$ generated from the above algorithm coincides with the sequence generated form the proximal forward-backward splitting algorithm [90–95] solving the balanced model [10, 82–84], which takes the following form:

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{W}^\top\boldsymbol{\alpha}-\boldsymbol{f}\|_2^2+\frac{\kappa}{2}\|(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{W}^\top)\boldsymbol{\alpha}\|_2^2+\|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_1. \tag{2.24}$$

One of the key observations to such connection is the relation between soft-thresholding and ℓ_1-norm based optimization. The balanced model takes the synthesis based model [96–100] and analysis based model [12, 101, 102] as special cases. Indeed, when $\kappa = 0$, we obtain the synthesis based model

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{W}^\top\boldsymbol{\alpha}-\boldsymbol{f}\|_2^2+\|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_1. \tag{2.25}$$

When $\kappa = \infty$, which forces $\boldsymbol{\alpha} = \boldsymbol{W}\boldsymbol{W}^\top\boldsymbol{\alpha}$, we obtain the analysis based model by letting $\boldsymbol{u} = \boldsymbol{W}^\top\boldsymbol{\alpha}$:

$$\min_{\boldsymbol{u}} \ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u}-\boldsymbol{f}\|_2^2+\|\boldsymbol{\lambda}\cdot\boldsymbol{W}\boldsymbol{u}\|_1. \tag{2.26}$$

Note that, to make solving the analysis based model (2.26) easier, it is normally treated as a constrained optimization problem, which takes the following equivalent form

$$\min_{\boldsymbol{u},\boldsymbol{W}\boldsymbol{u}=\boldsymbol{\alpha}} \ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u}-\boldsymbol{f}\|_2^2+\|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_1. \tag{2.27}$$

The split Bregman algorithm [12, 103] that we will discuss next was designed to solve (2.27) and hence the analysis based model (2.26).

The advantage of the 3-steps procedure lies in its generality and data-driven nature which makes it easier to design algorithms that are adaptive to the observed image. For example, one may change the wavelet frame system by any other systems that can provide a better "dense plus sparse" approximation to a specifically given image. Such systems can be adaptive to the observed image and be updated along with each iteration (see e.g. [26, 104]). The shrinkage operator can also be chosen adaptively to the given image and change at each iteration (see [6]). These modifications based on the 3-steps procedure can significantly improve the quality of the image restoration results. However, there may not be an associated optimization model to any of these adaptive thresholding algorithms. Therefore, from a practical point of view, the 3-step procedure is more flexible and easier to be made data-driven than the approaches based on designing optimization/variational models.

2.5. Other Algorithms for Image Restoration. Algorithms for finding sparse solutions of (1.1) can be developed from different approaches, e.g. ℓ_1/ℓ_0 minimization. However, most of them can be understood by the 3-steps procedure. Different algorithms may have a different way to compute an approximate solution $\boldsymbol{u}^1$ or update the residual; and different types of regularization result in different types of shrinkage operator $\mathcal{S}_{\boldsymbol{\lambda}}$, such as the soft-thresholding for ℓ_1-minimization models and hard-thresholding for ℓ_0-minimization models. This may lead to different optimization models, but the key is to applying shrinkage in the sparse domain (step 2), which is also the key to the data-driven approaches for seeking sparse solutions of the underlying problems. Here, we give two examples.

The first example is the linearized Bregman algorithm was originally introduced by [105, 106] for compressed sensing. It was later used in wavelet frame based image deblurring [13] and blind image deblurring [18, 19]. We recall the linearized Bregman algorithm as follows:

Algorithm (Linearized Bregman): Given any initial value for $\boldsymbol{\beta}^0$, iterate the following steps for

$k = 0, 1, \ldots$ until convergence:

$$\begin{aligned} \boldsymbol{\alpha}^{k+1} &= \frac{1}{\kappa}\mathcal{T}^s_{\boldsymbol{\lambda}}(\boldsymbol{\beta}^k) \\ \boldsymbol{\beta}^{k+1} &= \boldsymbol{\beta}^k + \boldsymbol{W}\boldsymbol{A}^\top(\boldsymbol{f} - \boldsymbol{A}\boldsymbol{W}^\top\boldsymbol{\alpha}^{k+1}). \end{aligned} \tag{2.28}$$

Let $\boldsymbol{\alpha}^\star$ be the resulting sequence after convergence. The reconstructed image for the linear inverse problem (1.1) is $\boldsymbol{W}^\top\boldsymbol{\alpha}^\star$. The first step of (2.28) is shrinkage in wavelet frame domain to recover image features and remove noise. The second step is a combination of finding approximate solution and update residual for the linear system $\boldsymbol{A}\boldsymbol{W}^\top\boldsymbol{\alpha} = \boldsymbol{f}$ in wavelet frame domain. Therefore, the linearized Bregman algorithm can also be understood as taking the form of the 3-steps procedure.

When the linearized Bregman algorithm (2.28) was first introduced [105, 106], the convergence analysis was missing because it was not clear what is the corresponding optimization problem the algorithm is solving. Later in [13], the authors discovered that the associated optimization problem to the linearized Bregman algorithm takes the following form

$$\min_{\boldsymbol{\alpha}}\{\frac{\kappa}{2}\|\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_1 : \boldsymbol{A}\boldsymbol{W}^\top\boldsymbol{\alpha} = \boldsymbol{f}\}. \tag{2.29}$$

Such observation led to the convergence analysis of the linearized Bregman algorithm [13].

The development of the linearized Bregman algorithm tells us the benefit of having an corresponding optimization model for a given algorithm, i.e. it makes convergence analysis of the algorithm easier. However, algorithms should be designed to achieve best image restoration quality by fully utilizing the prior knowledge of images; instead of trying to make it an algorithm associated to an optimization model by sacrificing quality of the recovered images. In general, the data-driven shrinkage operator and the linear system solver embedded in the 3-steps procedure may not be chosen such that there is an associated optimization model to the algorithm, although it may have better recovery of image features,

Now, we present the second example, the split Bregman algorithm [12,103], that solves the analysis based model (2.26):

Algorithm (Split Bregman): Let $\boldsymbol{b}^0 = \boldsymbol{d}^0 = \boldsymbol{0}$, $\boldsymbol{u}^0 = \boldsymbol{0}$. Iterate the following steps for $k = 0, 1, \ldots$ until convergence:

$$\begin{cases} \boldsymbol{u}^{k+1} = \left(\mu\boldsymbol{W}^\top\boldsymbol{W} + \boldsymbol{A}^\top\boldsymbol{A}\right)^{-1}\left[\mu\boldsymbol{W}^\top(\boldsymbol{d}^k - \boldsymbol{b}^k) + \boldsymbol{A}^\top\boldsymbol{f}\right], \\ \boldsymbol{d}^{k+1} = \mathcal{T}^s_{\boldsymbol{\lambda}/\mu}(\boldsymbol{W}\boldsymbol{u}^{k+1} + \boldsymbol{b}^k), \\ \boldsymbol{b}^{k+1} = \boldsymbol{b}^k + \delta(\boldsymbol{W}\boldsymbol{u}^{k+1} - \boldsymbol{d}^{k+1}). \end{cases} \tag{2.30}$$

The first step of (2.30) is to find an approximate solution; the second step is thresholding in the wavelet frame domain to recover image features; and the third step is to update residual. Therefore, the split Bregman algorithm can be understood as having the same algorithmic structure as the 3-steps procedure. Note that the complication of the Split Bregman algorithm comparing to the inpainting algorithm (2.23) is due to the difference of the objective function these algorithms attempt to solve. However, both of the algorithms share the same philosophy as the 3-steps procedure.

We present a numerical simulation of the split Bregman (**SBreg**) algorithm (2.30) for an image deblurring problem. The original blur and noise free image is shown by the left image of Figure 4 and the blurry and noisy observed image $\boldsymbol{f}$ is shown in the middle, where the kernel of the blur is generated in MATLAB by "fspecial('gaussian',11,1.0)" and the standard deviation of noise $\boldsymbol{\eta}$ is 3. The recovered image is shown by the right image in Figure 4.

Figure 4. Images from left to right are: the original image, the observed blurry and noisy image $\boldsymbol{f}$ and the reconstructed image using the split Bregman algorithm (2.30).

Both the linearized and split Bregman algorithm are based on the Bregman iterative algorithm [105,107,108]. The Bregman iterative algorithm was originally designed based on the Bregman distance

functional [109]. Its development is motivated by the needs of discovering effective algorithms in image restoration and compressed sensing. All these Bregman algorithms are discovered independently from available algorithms in optimization, although it was later realized by the community that the split Bregman algorithm is equivalent to the alternating direction method of multipliers (ADMM) [110–112] applied to the augmented Lagrangian [113–115] of the problem (2.27) (see e.g. [116]) and the linearized Bregman algorithm can be derived using the Uzawa's algorithm [117] (see [85]).

2.6. Beyond Image Restoration. The idea of the algorithmic design of the 3-steps procedure goes even beyond wavelet frame based image restoration. In fact, it promotes a general data processing philosophy based on the sparsity of the data. That is to first find an approximate solution based on given observations of the data; then conduct appropriate shrinkage operation in whatever domain where the local features of the data can be sparsely approximated; and iterative until convergence. Taking low rank matrix completion as an example [118], the observed data is the partial entries of the matrix to be recovered, and this matrix is known to have a low rank relative to its dimensions. The corresponding linear inverse problem to the matrix completion problem is not difficult to solve. However, finding a solution with low rank at the same time is highly non-trivial.

The linear system, denoted as $\boldsymbol{A}_{\boldsymbol{\Lambda}}$, associated to the matrix completion problem is an operator that extracts entries from a subindex set $\boldsymbol{\Lambda} \subsetneq \{1,2,\dots,m\} \times \{1,2,\dots,n\}$ of a matrix $\boldsymbol{u} \in \mathbb{R}^{m\times n}$. To be more precise, $\boldsymbol{A}_{\boldsymbol{\Lambda}}\boldsymbol{u} = \boldsymbol{u}$ on $\boldsymbol{\Lambda}$ and is set to 0 elsewhere. To recover the matrix $\boldsymbol{u} \in \mathbb{R}^{m\times n}$ from its noisy submatrix $\boldsymbol{f}$ is generally impossible. However, as shown by [118] that if the matrix to be recovered has low rank, a robust recovery of $\boldsymbol{u}$ from $\boldsymbol{f}$ is possible by minimizing the nuclear norm.

The earliest efficient matrix completion algorithm that works on large matrices was first proposed by [85], which is called singular value thresholding (SVT). The key observation on the low rank matrices is that they are sparse in the singular value domain, or in other words, it is sparse after applying the singular value decomposition (SVD). Hence, applying thresholding in sparse domain leads to the singular value thresholding (SVT) algorithm of [85]. The SVT algorithm can be understood as taking the form of the 3-steps procedure. In fact, a matrix completion problem can be viewed as an inpainting problem as well. The only difference is the sparsifying transformation used for matrix completion is the singular value decomposition (SVD) and the tresholding is performed in the singular value domain. Here, we recall the SVT algorithm of [85] as follows:

Algorithm (Singular Value Thresholding): Initialize $\boldsymbol{u}^0 = \boldsymbol{f}$ and $\boldsymbol{v}^0 = \mathcal{D}_\lambda(\boldsymbol{f})$. Iterate the following steps for $k = 1, 2, \dots$ until convergence:

$$\begin{cases} \boldsymbol{v}^k = \boldsymbol{v}^{k-1} + \delta \boldsymbol{A}_{\boldsymbol{\Lambda}}(\boldsymbol{f} - \boldsymbol{u}^{k-1}); \\ \boldsymbol{u}^k = \mathcal{D}_\lambda(\boldsymbol{v}^k). \end{cases}$$

Here, $\mathcal{D}_\lambda(\boldsymbol{u})$ is the singular value thresholding operator defined by

$$\mathcal{D}_\lambda(\boldsymbol{u}) = \boldsymbol{U}\mathcal{T}_\lambda^s(\boldsymbol{\Sigma})\boldsymbol{V}^\top, \tag{2.31}$$

where we have $\boldsymbol{u} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ and $\mathcal{T}_\lambda^s(\boldsymbol{\Sigma})$ is the soft-thresholding operator applied to the diagonal entries of $\boldsymbol{\Sigma}$. The first step of the SVT algorithm is the step of updating residual, while the second step is shrinkage in sparse domain of low rank matrices. Putting both steps together, we have $\boldsymbol{u}^k$ as the approximated solution of the associated linear system and of low rank at the same time. Therefore, the SVT algorithm can also be understood as one example of the 3-step procedure.

It was shown in [85] that the SVT algorithm (2.31) is in fact the Uzawa's algorithm solving the following constrained optimization model

$$\min_{\boldsymbol{u}\in\mathbb{R}^{m\times n}} \{\frac{1}{2}\|\boldsymbol{u}\|_F^2 + \lambda\|\boldsymbol{u}\|_* : \boldsymbol{A}_{\boldsymbol{\Lambda}}\boldsymbol{u} = \boldsymbol{f}\}. \tag{2.32}$$

Here, $\|\boldsymbol{u}\|_F$ is the Frobenius norm defined by $\|\boldsymbol{u}\|_F^2 = \boldsymbol{tr}(\boldsymbol{u}^\top\boldsymbol{u})$, and $\|\boldsymbol{u}\|_*$ is the nuclear norm of $\boldsymbol{u}$ which is equal to the sum of the singular values of $\boldsymbol{u}$. The SVT algorithm of [85] was the earliest efficient algorithm for large scale matrix completion problems. After the introduction of the SVT algorithm, many other algorithms were introduced in the literature which have improved efficiency and reconstruction quality. In addition, when the matrix to be recovered satisfies additional properties such as having sparse entries [119–121], model (2.32) can be properly modified and algorithms can be designed to accommodate these additional prior knowledge on data. However, none of these newly developed algorithms goes beyond the crucial idea used by the SVT algorithm (2.31) or more generally the 3-steps procedure, i.e. to conduct proper shrinkage in the singular value domain, as the low rank is the key property to preserve, while whether it is close to the minimizer of a certain optimization model is of secondary priority in low rank matrix completion.

The ideas of low rank matrix completion and wavelet frame based image restoration were combined and successfully applied to video restoration problems [29, 122–124] and 4D medical imaging [30, 31]. The key to video data processing is to separate background, which are in fact global patterns of the

video, from fast moving objects in the video, which are essentially local features. The low rank matrices extracted properly from the video together with the wavelet frame form a "global patterns plus local features" type of decomposition of the video data. The low rank component effectively approximates the background or other slow varying video contents, and the sparse component provided by wavelet frames can sparsely approximate moving objects. This is the main reason why these models are successful in processing videos or more generally, time-dependent image data such as 4D medical images.

3. Specific Data-Driven Approaches

The iterative wavelet frame shrinkage algorithm of [11], as well as most of existing wavelet frame based methods, used a generic data-driven prior on images, i.e. local image features can be sparsely approximated by wavelet frames. However, such generic prior knowledge may not be well satisfied by a specifically given image. For example, the representation by wavelet frames for a specifically given image may not be ideal, since traditional wavelet frames are generic systems that are not adaptive to any specifically given data. The main objective of this section is to discuss how we can design specific data-driven frame based algorithms and models that are more effective than the classical generic data-driven approach.

3.1. Data-Driven Transformations.

For most image restoration problem, the quality of the reconstruction highly depends on the choice of the transformation. In this section, we discuss how we can create data-driven transformations that provide better sparse approximation of local features and global patterns of a specifically given image.

3.1.1. Data-Driven Adaptive Tight Frames.

Finding a good representation of images is crucial in image restoration. It is known that wavelet frames provide good dense plus sparse approximation for piecewise smooth functions, such as images. However, for a given image, generic wavelet frame systems may not be ideal. In other words, using data-driven sparse approximation that derive a transform adaptive to the observed image may be more desirable.

The concept of data-driven representations has been explored in recent years by the so-called learning approaches [125–130]. These approaches learn a redundant dictionary from the input image itself to achieve better sparsity of the input image over the learned dictionary. These redundant dictionaries are very effective on sparsely approximating images with rich textures. Despite the success of these adaptively learning methods, the resulting dictionaries lack several properties desired for image restoration. For instance, finding an optimal redundant system often leads to a severely under-constrained ill-posed problem. It also remains a challenging task to develop fast and stable numerical methods for estimating an optimal redundant system. Furthermore, the dictionary learnt is redundant, but can be under-complete, that makes the reconstruction step even harder.

In [26], the authors developed a new approach to construct discrete tight frames that are adaptive to input images. The proposed adaptively learned tight frame is more likely to give a highly sparse approximation of the local features of the input image than existing tight wavelet frames such as the B-spline tight wavelet frames. In addition, the minimization problems arising in the construction of tight frames are better conditioned than those of generic redundant dictionaries, thanks to the Parseval's identity satisfied by tight frames. Therefore, by considering a class of tight frames with certain special properties, a fast and stable numerical method is available to construct data-driven tight frames.

An adaptive tight frame denoising algorithm based on a data-driven tight frame construction scheme was also proposed in [26]. The experiments show that the adaptive tight frame denoising technique significantly outperforms standard wavelet thresholding approaches on images of rich textures. Also, it is much faster than the K-SVD method [131], because the size of the problem is much smaller than that of the K-SVD method, while it has comparable performance as the K-SVD method. Another shortcoming of the K-SVD method is that the learned dictionary is not guaranteed to be complete in the underlying Euclidean space.

Let $\boldsymbol{f}$ be a given observed image. Let $\{\boldsymbol{q}_\ell : \ell = 0, \dots, r\}$ be the set of filters to be learned from the given image $\boldsymbol{f}$, and $\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}$ be the associated tight frame transform defined by (2.7). The set of filters $\{\boldsymbol{q}_\ell\}$ can be learned from $\boldsymbol{f}$ by solving the following optimization problem [26]:

$$\begin{aligned} \min_{\boldsymbol{v},\{\boldsymbol{q}_\ell\}} \ & \frac{1}{2}\left\|\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}^\top \boldsymbol{v} - \boldsymbol{f}\right\|_2^2 + \frac{1}{2}\left\|\left(\boldsymbol{I} - \boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}^\top\right)\boldsymbol{v}\right\|_2^2 + \nu\|\boldsymbol{v}\|_0 \\ \text{s.t.} \ & \boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}^\top \boldsymbol{W}_{\{\boldsymbol{q}_\ell\}} = \boldsymbol{I}. \end{aligned} \tag{3.1}$$

Here, $\|\boldsymbol{v}\|_0$ is the ℓ_0-"norm" of $\boldsymbol{v}$ that outputs the number of nonzero entries of $\boldsymbol{v}$. The penalization of the ℓ_0-"norm" of $\boldsymbol{v}$ makes sure that the representation $\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}$ provides a highly sparse approximation for the given image $\boldsymbol{f}$. The constraint $\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}^\top \boldsymbol{W}_{\{\boldsymbol{q}_\ell\}} = \boldsymbol{I}$ makes sure that $\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}$ is a tight frame. The

ℓ_0-minimization on $\boldsymbol{v}$ of (3.1) ensures the model to find a system with best sparse approximation to local features of the given image $\boldsymbol{f}$. The constraint $\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}^\top \boldsymbol{W}_{\{\boldsymbol{q}_\ell\}} = \boldsymbol{I}$ enables the system learned from (3.1) provide dense approximation to global patterns of the given image $\boldsymbol{f}$, which enables the problem (3.1) to be solved stably by a fast numerical algorithm. Since the resulting system $\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}$ from (3.1) is learned from the observed image $\boldsymbol{f}$, it will be referred to as a data-driven *adaptive tight frame*, and the associated transform will be denoted simply as $\boldsymbol{W}^{(\boldsymbol{f})}$.

The optimization problem (3.1) can be solved rather efficiently. An alternative optimization approach was adopted by [26]: $\boldsymbol{v}$ is solved while $\{\boldsymbol{q}_\ell\}$ is fixed and then $\{\boldsymbol{q}_\ell\}$ is solved while $\boldsymbol{v}$ is fixed. Each of the subproblem has a closed form solution: solution for $\boldsymbol{v}$ is obtained by hard-thresholding while $\{\boldsymbol{q}_\ell\}$ is solved through SVD. This makes the algorithm computationally efficient. The convergence analysis of the alternative optimization algorithm was later provided by [27], where the authors showed that there exists at least one subsequence that converges to a stationary point of (3.1). Moreover, it was shown that convergence of the entire sequence can be achieved by adding an proximal term to each of the subproblem [27]. The modified algorithm was later used for face recognition [35].

Note that we can generalize the model (3.1), which was proposed for image denoising, to a model for generic image restoration problem. The generalized model reads as follows:

$$\begin{aligned} \min_{\boldsymbol{v},\{\boldsymbol{q}_\ell\}} \ & \frac{1}{2}\left\|\boldsymbol{A}\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}^\top \boldsymbol{v} - \boldsymbol{f}\right\|_2^2 + \frac{1}{2}\left\|\left(\boldsymbol{I} - \boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}\boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}^\top\right)\boldsymbol{v}\right\|_2^2 + \nu\|\boldsymbol{v}\|_0 \\ \text{s.t.} \quad & \boldsymbol{W}_{\{\boldsymbol{q}_\ell\}}^\top \boldsymbol{W}_{\{\boldsymbol{q}_\ell\}} = \boldsymbol{I}. \end{aligned} \tag{3.2}$$

Model (3.2) is obviously more complicated to solve than (3.1). Alternative optimization approach can be adopted, whereas each subproblem no longer has a closed form solution and has to be solved iteratively.

More recently, adaptive tight frame and bi-frame construction method was proposed by [132], where an ℓ_1-norm based analysis type of model was proposed. Connections between multi-level frame transforms and deep convolution networks were studied.

3.1.2. Data-Driven Non-Local Tight Frames. Most of the classical wavelet frames used in image restoration are locally supported in spatial domain. Therefore, the sparsity prior of images under an existing wavelet frame system only refers to the sparse nature of local features of images in spatial domain. Such local approach works well on regions of the image with slow variations, while it may be less effective in textural regions (i.e. regions with high frequency fluctuation). An alternative approach is the non-local methods based on the assumption that image structures of small image regions tend to repeat themselves in spatial domain. Such approach is of non-local nature in the sense that image pixels that are far apart may be related to each other. There are two types of non-local schemes proposed to use such a non-local prior. One is the non-local mean [133] and non-local operators based approach [104,134–136]. Another approach is the patch-based method which groups the similar patches together followed by a collaboratively filtering [129, 137–139]. One way to state the main difference between local and non-local approaches is that local approaches calculate local variations in spatial domain, while non-local approaches calculate local variations in similarity domain.

The local approaches and the non-local approaches have their own advantages and disadvantages. Most local approaches are not very effectively for images with relatively complex texture regions, as the local variations of textures are normally not sparse. The non-local mean based approaches overcomes such weakness by regrouping pixels according to the similarities of the associated patches, so that local variations calculated in the similarity domain are still sparse. However, the performance of the non-local mean based approaches is still not comparable to that of some patch-based non-local approaches such as the BM3D method. By applying 3D filters on stacks image patches grouped based on their similarities, the BM3D method considered both local image sparsity and global self-recursive image structures. However, it lacks an automatic mechanism to deal with images without much self-recursive structures, such as medical images. Thus, the BM3D method may produce poor restoration results for this type of images by wrongly imposing a global similarity prior.

In a recent paper [25], the authors proposed a scheme of constructing data-driven non-local tight frames adaptive to the input image. Such non-local frame systems combine the merits of both local and non-local approaches. The associated representation of images provides a good sparse approximation to local intensity variations, self-recursive prior of local image structures, and self-recursive prior of local image structures across different scales. The non-local tight frames still decompose images to their global patterns and local features. However, what makes them different from classical wavelet frames is that the localness not only includes spatial localness, but also localness in similarity domain where two nearby pixels could be far apart spatially or in scales. Image restoration algorithm based on the non-local tight frames was also proposed in [25], where it was shown that the denoising results of the non-local tight frames are much better than those of the classical wavelet frame based approach, variational and PDE based approach. It is also much faster than the BM3D algorithm with comparable denoising quality.

The data-driven *non-local tight frames* proposed by [25] reads as follows:

$$\boldsymbol{V}^{(\boldsymbol{f})} = \frac{1}{\sqrt{2}} \begin{pmatrix} \boldsymbol{W} \\ \boldsymbol{J}^{(\boldsymbol{f})}\boldsymbol{W} \end{pmatrix}, \tag{3.3}$$

where $\boldsymbol{W}$ is the transformation associated with some existing multi-level wavelet frame system (e.g. the B-spline wavelet frame systems), and $\boldsymbol{J}^{(\boldsymbol{f})}$ is the system that encodes the self-recursive property of wavelet frame coefficients $\boldsymbol{W}\boldsymbol{f}$ in spatial domain and cross multiple scales. By construction, we have $\left(\boldsymbol{V}^{(\boldsymbol{f})}\right)^{\top}\boldsymbol{V}^{(\boldsymbol{f})} = \boldsymbol{I}$, i.e. it is indeed a discrete tight frame. The non-local tight frame $\boldsymbol{V}^{(\boldsymbol{f})}$ is composed of two systems: one is some existing discrete tight wavelet frame $\boldsymbol{W}$ which extracts global patterns and (spatially) local features, and the other is the non-local version of $\boldsymbol{W}$ generated by the multiplication of a linear system $\boldsymbol{J}^{\boldsymbol{f}}$ which extracts local features in similarity domain. The system $\boldsymbol{J}^{\boldsymbol{f}}$ is learned adaptively from the given image $\boldsymbol{f}$, and is used for relating the wavelet frame coefficients corresponding to the same image structure, which are often spread out over the whole image. Thus the system $\boldsymbol{J}^{\boldsymbol{f}}$ can be viewed as a non-local operator to relate wavelet frame coefficients that may be spatially far away from each other. We shall skip details on the construction of $\boldsymbol{J}^{\boldsymbol{f}}$, as well as the application of $\boldsymbol{V}^{(\boldsymbol{f})}$ in image restoration, and refer the interested reader to [25].

3.2. Piecewise-Smooth Image Model. In PDE based approach, such as the well-known Rudin-Osher-Fatemi (ROF) model [45], images are often described by functions of bounded variations (BV). For wavelet or wavelet frame based methods (see e.g. [12, 96, 97]) with the regularization term being the ℓ_1-norm of the wavelet frame coefficients, images are essentially approximated by functions in the homogenous Besov space $B^1_{1,1}$. The BV space is a fairly large function space. It is known (see e.g. [140]) that a BV function can always be decomposed to the sum of an absolute continuous part with respect to the Lebesgue measure, a jump part (e.g. edges), and a Cantor measure. However, since images always have a limited resolution, we normally cannot observe the Cantor measure in images. On the other hand, the Besov space $B^1_{1,1}$ sometimes is not large enough to include some types of images.

The most suitable way of describing images is "piecewise smooth", or in other words, images are smooth away from their singularity sets. Such concept has already been used for a long time in wavelet and wavelet frame based approach for image restoration. However, there has not been a clear definition of piecewise smooth function space. Moreover, existing methods do not actively detect and restore image singularities in the underlying models or algorithms. Instead, the concept of "piecewise smooth" has been used as a generic image prior knowledge. In the recent work by [47, 51], piecewise smooth image restoration models are introduced in a data-driven fashion, where a precise definition of piecewise smooth function space was given by [47]. These models estimate both the image to be restored and its singularity set simultaneously, so that singularities can be well protected and enough regularization in smooth regions is properly enforced at the same time. The models themselves are adaptive to the specifically given image, and they combine the merits of the PDE based approach [42–45, 48, 49] and the traditional wavelet frame based approach [11, 12, 57].

We define the space of piecewise smooth functions (with jumps and hidden jumps) as

$$\mathcal{H}^{1,s}(\{\Omega_{j,\tilde{j}}\}) = \{f \in L_2(\Omega) : \|f\|_{\mathcal{H}^{1,s}(\{\Omega_{j,\tilde{j}}\})} < \infty\}, \tag{3.4}$$

where

$$\|f\|_{\mathcal{H}^{1,s}(\{\Omega_{j,\tilde{j}}\})} = \sum_{j=1}^{m}\left[\|f\|_{H^1(\Omega_j)} + \sum_{\tilde{j}=1}^{m_j}\|f\|_{H^{s_{j,\tilde{j}}}(\Omega_{j,\tilde{j}})}\right],$$

where $s = \min\{s_{j,\tilde{j}}\}$ and $s_{j,\tilde{j}} \geq 2$. Here, H^s denotes the Sobolev space equipped with the norm $\|f\|_{H^s(\Omega)} = \sum_{0\leq|\boldsymbol{i}|\leq s}\|D_{\boldsymbol{i}}f\|_{L_2(\Omega)}$, where $D_{\boldsymbol{i}}(f(x,y)) = \frac{\partial^{|\boldsymbol{i}|}f}{\partial x^{i_1}\partial y^{i_2}}$. The first order weak derivatives of $f \in \mathcal{H}^{1,s}(\{\Omega_{j,\tilde{j}}\})$ are integrable in sub-domains $\{\Omega_j\} \subset \Omega$ with jump discontinuities at the boundaries of these sub-domains, which is the jump set. Second order or higher weak derivatives of $f \in \mathcal{H}^{1,s}(\{\Omega_{j,\tilde{j}}\})$ are integrable in the sub-domains $\{\Omega_{j,\tilde{j}}\} \subset \Omega_j$ and the first order derivatives have jump discontinuities at the boundaries of $\Omega_{j,\tilde{j}}$, which is the hidden jump set. We shall refer to the union of jump and hidden jump sets simply as the jump set of f. Note that the piecewise smooth functions with higher order hidden jump sets, i.e. jump discontinuities of second or higher order derivatives, can be similarly defined. Moreover, what makes the piecewise smooth function space (3.4) different from the SBV space [141] is the characterization of hidden jump sets.

The wavelet frame based piecewise smooth image restoration model by [47] reads as follows:

$$\inf_{\boldsymbol{u},\ \boldsymbol{\Gamma}} \|\boldsymbol{\lambda}\cdot \boldsymbol{W}_{\boldsymbol{\Gamma}^c}\boldsymbol{u}\|_2^2 + \|\boldsymbol{\gamma}\cdot \boldsymbol{W}_{\boldsymbol{\Gamma}}\boldsymbol{u}\|_1 + \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u}-\boldsymbol{f}\|_2^2. \tag{3.5}$$

where

$$\|\boldsymbol{\lambda}\cdot \boldsymbol{W}_{\boldsymbol{\Gamma}^c}\boldsymbol{u}\|_2^2 = \sum_{\boldsymbol{k}\in\boldsymbol{\Gamma}^c}\sum_{l=0}^{L}\sum_{\ell=1}^{r}\lambda_{\ell,l,\boldsymbol{k}}\left|(\boldsymbol{W}_{\ell,l}\boldsymbol{u})[\boldsymbol{k}]\right|^2$$

and

$$\|\boldsymbol{\gamma}\cdot \boldsymbol{W}_{\boldsymbol{\Gamma}}\boldsymbol{u}\|_1 = \sum_{\boldsymbol{k}\in\boldsymbol{\Gamma}}\left[\sum_{l=0}^{L}\left(\sum_{\ell=1}^{r}\gamma_{\ell,l,\boldsymbol{k}}\left|(\boldsymbol{W}_{\ell,l}\boldsymbol{u})[\boldsymbol{k}]\right|^2\right)^{\frac{1}{2}}\right].$$

In [47], an alternative optimization algorithm was also proposed to solve (3.5):

Algorithm (Piecewise Smooth): Let $\boldsymbol{u}^0$ and $\boldsymbol{\Gamma}^0$ be some initial data. For $k=1,2,\ldots$. Iterate

1. Given $\boldsymbol{\Gamma}^{k-1}$, compute $\boldsymbol{u}^k$ by

$$\boldsymbol{u}^k = \arg\min_{\boldsymbol{u}\in\mathcal{I}_2}\left\|\boldsymbol{\lambda}\cdot \boldsymbol{W}_{(\boldsymbol{\Gamma}^{k-1})^c}\boldsymbol{u}\right\|_2^2 + \|\boldsymbol{\gamma}\cdot \boldsymbol{W}_{\boldsymbol{\Gamma}^{k-1}}\boldsymbol{u}\|_1 + \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u}-\boldsymbol{f}\|_2^2.$$

   Note that $\boldsymbol{u}^k$ can be solve by the split Bregman algorithm.

2. Given $\boldsymbol{u}^k$, estimate the jump set $\boldsymbol{\Gamma}^k$ by

$$\boldsymbol{\Gamma}^k = \left\{\boldsymbol{p}:\ \sum_{l=0}^{L}\left(\sum_{\ell=1}^{r}\gamma_{\ell,l}[\boldsymbol{p}]\left|(\boldsymbol{W}_{\ell,l}\boldsymbol{u}^k)[\boldsymbol{p}]\right|^2\right)^{\frac{1}{2}} \le \sum_{l=0}^{L}\sum_{\ell=1}^{r}\lambda_{\ell,l}[\boldsymbol{p}]\left|(\boldsymbol{W}_{\ell,l}\boldsymbol{u}^k)[\boldsymbol{p}]\right|^2\right\}. \tag{3.6}$$

The step (3.6) of the above algorithm is the crucial component of the piecewise smooth image restoration algorithm. The jump set is determined by going through each of the pixel location in wavelet frame domain comparing a grouped ℓ_1-norm and ℓ_2-norm of the coefficient across all levels l and bands ℓ. Not only jump discontinuities can be detected by (3.6), singularities of different types (e.g. hidden jumps) can also be detected. This is because different types of singularities are naturally extracted by wavelet frames with different orders of vanishing moment. Also, the multiscale structure of wavelet frames grants robustness to the estimation of $\boldsymbol{\Gamma}$ at presence of noise.

Here, we present a numerical simulation of the piecewise smooth (**PS**) algorithm for the same image deblurring scenario we considered for SBreg and LBreg algorithms. The original blur and noise free image is shown by the left image of Figure 5 and the blurry and noisy observed image $\boldsymbol{f}$ is shown in the middle. The recovered image and the estimated jump set $\boldsymbol{\Gamma}$ from the PS algorithm is shown by the right image in Figure 5.

Figure 5. Images from left to right are: the original image, the observed blurry and noisy image $\boldsymbol{f}$, the reconstructed image using the PS algorithm, and the estimated jump set which is enclosed in the red curves.

The piecewise smooth image restoration model by [51] reads as follows:

$$\inf_{\boldsymbol{u},\ \boldsymbol{\Gamma}}\left\{\|\boldsymbol{\lambda}\cdot \boldsymbol{W}_{\boldsymbol{\Gamma}^c}\boldsymbol{u}\|_2^2 + \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u}-\boldsymbol{f}\|_2^2:\ |\boldsymbol{\Gamma}|\le\gamma\right\}. \tag{3.7}$$

The major difference between (3.5) and (3.7) is the regularization of the jump set $\boldsymbol{\Gamma}$ is formulated as a constrain in (3.7) instead of as a penalty term as in (3.5). The associated fast numerical algorithms solving (3.7) can be found in [51].

The ℓ_2-regularization in subsystem $\boldsymbol{W}_{\boldsymbol{\Gamma}^c}$ of wavelet frame domain is to keep the global patterns and smooth components of images smooth. The ℓ_1-regularization of (3.5) on $\boldsymbol{W}_{\boldsymbol{\Gamma}}$ is to protect image singularities (i.e. local features) and to introduce proper regularity restriction on the geometry of the

jump set $\mathbf{\Gamma}$. The constraint $|\mathbf{\Gamma}| \leq \gamma$ of the model (3.7) has a similar effect as the ℓ_1-term of (3.5), whereas the former has a more direct control of the regularity of the jump set which may lead to better image reconstruction results (see [51]). The advantage of model (3.5) is that it is not sensitive to the estimation of $\mathbf{\Gamma}$. Indeed, when $\mathbf{\Gamma}^c = \emptyset$, it reduces to the analysis based model (2.26), which has been proven in the literature to be a good image restoration model. Intuitively, the reconstruction results of (3.5) cannot be worse than those of the analysis based model (2.26).

Both model (3.5) and (3.7) are data-driven, as the singularity set Γ is derived iteratively from the given image and the approximate solution at each step. Comparing with wavelet frame based analysis model (2.26), instead of using the subsystem $\boldsymbol{W}_{\mathcal{H}}$ for local features and $\boldsymbol{W}_{\mathcal{L}}$ for global patterns, we use the subsystem $\boldsymbol{W}_{\Gamma}$ to achieve sparser approximation of local features, and use both the subsystem $\boldsymbol{W}_{\mathcal{L}}$ and $\boldsymbol{W}_{\mathbf{\Gamma}^c}$ to capture global patterns. In other words, the subsystem $\boldsymbol{W}_{\mathbf{\Gamma}^c}$ further refines the partition of the whole system $\boldsymbol{W}$ for the given image data $\boldsymbol{f}$ by including a portion of the original sparse components $\boldsymbol{W}_{\mathcal{H}}$ in $\boldsymbol{W}_{\mathcal{L}}$. Numerical examples in [47,51] show that such data-driven refinements of $\boldsymbol{W}_{\mathcal{L}}$ and $\boldsymbol{W}_{\mathcal{H}}$ by the models (3.5) and (3.7) indeed leads to significantly improved image restoration results.

Furthermore, the model (3.5) was shown in [47] to be a discrete realization of a brand new variational model. To be more precise, the authors showed that when image resolution goes to infinite, the energy function of (3.5) Gamma-converges (see e.g. [142] for an introduction of Gamma-convergence) to a brand new variational functional under the topology of the piecewise smooth function space (3.4). This variational functional is related to, but more effective than, the well-known Mumford-Shah functional [48]. The Mumford-Shah functional also promotes an estimate of both the smooth image component and singularity set. However, the specific formulation of the Mumford-Shah functional makes it difficult to solve in practice. In addition, due to the presence of wavelet frame functions with varied orders of vanishing moment, the types of singularities characterized by the new variational functional is far richer than the Mumford-Shah functional. This is a rather desirable property for image restoration considering the fact that images do contain singularities of different types. We refer the interested readers to [47] for further details on the analysis.

The significance of the piecewise smooth image restoration models (3.5) and (3.7) is even beyond the scope of image restoration, which has been categorized as low level image processing. In the literature of imaging science, image restoration has been treated as a data preprocessing step for higher level image analysis tasks, and how the images being processed will be used at a higher level stage is generally ignored. However, the risk of it is that key features for high level image analysis may not be well recovered or even removed during image restoration process. The models (3.5) and (3.7) can be regarded as a link between lower level image processing and high level image understanding. Indeed, the singularity set $\mathbf{\Gamma}$ obtained from these models encode valuable information of the geometries of the objects reside in the image, which can be used for object recognition, quantification, image classification, etc. Therefore, the two piecewise smooth image restoration models opened a door towards the unified modeling for both low level image processing and high level image analysis.

3.3. Data-Driven Shrinkage. The quality of an image restoration method is not only determined by the choice of "dense plus sparse" approximation of images. The shrinkage operator plays a rather important role as well. This section is focused on the choice of data-driven shrinkage operators that are adaptive to the given image to obtain better image restoration results.

As stated earlier, the key to the success of image restoration methods is to preserve or enhance image features (i.e. singularities) and maintain smooth image components. Most transformation based methods use certain transforms to separate image features and smooth image component, and do processing in the transform domain. One of the most popular type of processing in the transform domain is thresholding, such as soft- and hard-thresholding, with relatively uniform threshold values. However, such processing does not have to be limited to thresholding. It can be a generic shrinkage operator that takes soft- and hard-thresholding as special cases. Moreover, the amount of shrinkage for each coefficient in transform domain does not have to be uniform. One fixed shrinkage value cannot possibly be ideal for the entire image domain. For example, the sparse wavelet frame coefficients of an image are generally small in smooth regions and large near jump discontinuities or other type of singularities. Therefore, we should apply different shrinkage levels at different locations of the given image, in other word, a data-driven design of shrinkage.

In [6], a generic iterative wavelet frame shrinkage algorithm was studied. The algorithm takes the following form

$$\boldsymbol{u}^{k+1} = \widetilde{\boldsymbol{W}}^{\top} \mathcal{S}_{\boldsymbol{\lambda}^k}(\boldsymbol{W}\boldsymbol{u}^k), \qquad k = 0, 1, \ldots, \tag{3.8}$$

Here, $\boldsymbol{\lambda}^k = \{\lambda_{\ell,l,\boldsymbol{k}}(\boldsymbol{\alpha}^k) : (\ell, l) \in \mathbb{B}, \boldsymbol{k} \in \mathbb{Z}^2\}$ with $\boldsymbol{\alpha}^k = \boldsymbol{W}\boldsymbol{u}^k$. The shrinkage operator $\mathcal{S}_{\boldsymbol{\lambda}}(\boldsymbol{\alpha})$ is given by

$$\mathcal{S}_{\boldsymbol{\lambda}}(\boldsymbol{\alpha}) = \{S_{\lambda_{\ell,l,\boldsymbol{k}}(\boldsymbol{\alpha})}(\alpha_{\ell,l}[\boldsymbol{k}]) = \alpha_{\ell,l}[\boldsymbol{k}](1 - \lambda_{\ell,l,\boldsymbol{k}}(\boldsymbol{\alpha})) : \ (\ell, l) \in \mathbb{B}, \boldsymbol{k} \in \mathbb{Z}^2\}. \tag{3.9}$$

In (3.9), the shrinkage value $\boldsymbol{\lambda}$ varies at different level of wavelet frame decomposition l, wavelet frame band ℓ and location $\boldsymbol{k}$. The shrinkage value $\lambda_{\ell,l,\boldsymbol{k}}(\boldsymbol{\alpha})$ also depends on the values of the wavelet frame coefficients $\boldsymbol{\alpha}$. Therefore, the shrinkage operator given by (3.9) is a rather general operator, which in

fact takes the soft- and hard-thresholding as special cases (see [6]). It is also adaptive in nature since the shrinkage value depends on the wavelet frame coefficients of the current approximation of the solution $\boldsymbol{u}^k$. The flexibility and the adaptive nature of the shrinkage operator $\mathcal{S}_{\boldsymbol{\lambda}}$ enable us to design data-driven shrinkage operators to achieve higher quality image reconstruction than classical thresholding operators. Yet, it is still unclear on how one should choose the shrinkage level $\boldsymbol{\lambda}$ properly for a given image.

One of the major discoveries of [6] is the connections between the iterative wavelet frame shrinkage algorithm (3.8) and the following generic nonlinear evolution PDE

$$u_t = \sum_{\ell=1}^{r} \frac{\partial^{\boldsymbol{\alpha}_\ell}}{\partial x^{\boldsymbol{\alpha}_\ell}} \Phi_\ell(\boldsymbol{D}u, u), \quad \text{with } \boldsymbol{D} = \Big(\frac{\partial^{\boldsymbol{\beta}_1}}{\partial x^{\boldsymbol{\beta}_1}}, \dots, \frac{\partial^{\boldsymbol{\beta}_r}}{\partial x^{\boldsymbol{\beta}_r}}\Big), \tag{3.10}$$

where $|\boldsymbol{\alpha}_\ell|, |\boldsymbol{\beta}_\ell| \geq 0$ for all $1 \leq \ell \leq r$. Note that the PDE (3.10) includes nonlinear diffusions (e.g. the Perona-Malik equation [49]) and nonlinear hyperbolic equations (e.g. the shock-filters [50]) as special cases. It was shown in [6] that for any given evolution PDE taking the form (3.10), one can always select the bi-frame transformations $\boldsymbol{W}$ and $\widetilde{\boldsymbol{W}}$, and the shrinkage values $\boldsymbol{\lambda}(\boldsymbol{\alpha}) = \{\lambda_{\ell,l,\boldsymbol{k}}(\boldsymbol{\alpha})\}$ properly such that the iterative shrinkage algorithm (3.8) is a discretization of the PDE (3.10). For nonlinear diffusions satisfying certain assumptions, the discrete solution indeed converges to that of the PDE as meshsize goes to infinity (see [6] for details).

Such connections between wavelet frame shrinkage and nonlinear evolution PDEs provide new and inspiring interpretations of both approaches. For example, some of the wavelet frame shrinkage algorithms that are commonly used in image restoration, such as the iterative soft-thresholding algorithms, lead to new types of nonlinear PDEs that have not been considered in the literature. Also, the accelerated iterative wavelet frame shrinkage algorithm of [24, 89] led to a new time dependent nonlinear PDE of both parabolic and hyperbolic type. On the other hand, through such connections, the PDE based approach grants geometric insights to the wavelet frame shrinkage algorithms, which enables us to choose desirable shrinkage levels adaptively to given image. Furthermore, the discretization provided by wavelet frame transforms is superior to some standard finite difference discretization of differential operators for image restoration problems. The reason is that the discrete transformation of a wavelet frame system, such as B-spline tight wavelet frame systems, typically contains filters with varied orders of vanishing moments specializing in capturing image features of different types, with different orientations and at different scales. Hence, wavelet frames provides better sparse approximation to local image features than the system corresponding to a finite difference discretization. On the other hand, wavelet frames also provides dense approximation to global image patterns which is missing from the system associated to classical finite difference discretizations. Therefore, wavelet frame transforms are more desirable than finite difference discretizations in approximating differential operators for image restoration problems.

The significance of these findings is beyond what it may appear. The analysis and discussions in [6] indicates that wavelet frame based approach is a new way of solving PDEs in general. Together with the earlier work [46, 47], wavelet frame based approach can be regarded as a brand new tool for numerical analysis to discretize and solve variational and PDE models that arise from different disciplines in science and engineering.

Through the generic connections between (3.8) and (3.10), we can borrow the ideas from PDE based approach to design an adaptive shrinkage strategy. The specific PDE we consider is the well-known Perona-Malik (PM) equation used for image denoising and enhancement [49]. First, we recall the PM equation:

$$u_t = \operatorname{div}\big(g(|\nabla u|^2)\nabla u\big),$$

where g is a function satisfying

$$\begin{cases} g : [0,\infty) \mapsto (0,\infty) \text{ decreasing;} \\ g(0) = 1;\ g(x) \to 0 \text{ as } x \to \infty; \\ g(x) + 2xg'(x) > 0 \text{ for } x \leq K; \quad g(x) + 2xg'(x) < 0 \text{ for } x > K. \end{cases} \tag{3.11}$$

One example of g is $g(x) = \frac{1}{1+x^p/K}$ for some constant $K > 0$ and $p > 1/2$. The diffusion coefficient of the Perona-Malik equation is $g(|\nabla u|^2)$ which controls the amount of diffusion at each location. By the specific assumption on g, we can see that at smooth regions ($|\nabla u|$ is small), $g(|\nabla u|^2)$ is large which means more diffusion is allowed; while near singularities ($|\nabla u|$ is large), $g(|\nabla u|^2)$ is small meaning less diffusion is allowed.

Thanks to the link of (3.8) to (3.10), we can select the shrinkage values $\boldsymbol{\lambda}(\boldsymbol{\alpha}) = \{\lambda_{\ell,l,\boldsymbol{k}}(\boldsymbol{\alpha})\}$ similarly as the selection of function g of the PM equation so that the shrinkage is adaptive to local image geometry. For example, one may select $\mathcal{S}_{\boldsymbol{\lambda}}$ so that (3.8) is exactly a discretization of the PM equation. However, we have much more flexibility in selecting the shrinkage operator than merely trying to match it with an existing PDE.

For generic image restoration problem, we need to have the underlying image restoration problem properly embedded in the iterative wavelet frame shrinkage algorithm (3.8), because otherwise all image

features will eventually be smeared out as $k \to \infty$. In fact, the 3-steps procedure tells us exactly how one should embed image restoration problem in the algorithm. The following tight wavelet frame based adaptive thresholding algorithm is one of many examples:

Algorithm (Adaptive Soft-Thresholding): Initialize $\boldsymbol{u}^0 = \boldsymbol{0}$. Iterate the following steps for $k = 1, 2, \ldots$ until convergence:

$$\begin{cases} \boldsymbol{u}^{k+1/2} = \boldsymbol{W}^\top \mathcal{T}^s_{\boldsymbol{\lambda}(\boldsymbol{\alpha}^k)}(\boldsymbol{W}(\boldsymbol{u}^k)), \\ \boldsymbol{u}^{k+1} = (I - \mu \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{u}^{k+1/2} + \mu \boldsymbol{A}^\top \boldsymbol{f}. \end{cases} \tag{3.12}$$

where $\mathcal{T}^s$ is the soft-thresholding operator given by (2.13). The adaptive threshold level $\boldsymbol{\lambda}(\boldsymbol{\alpha})$, which depends on the magnitudes of the wavelet frame coefficients $\boldsymbol{\alpha} = \{\alpha_{\ell,l}[\boldsymbol{k}] : (\ell, l) \in \mathbb{B}, \boldsymbol{k} \in \mathbb{Z}^2\}$, is given by $\boldsymbol{\lambda}(\boldsymbol{\alpha}) = \{\lambda_{\ell,l}[\boldsymbol{k}] : \ \ell, l, \boldsymbol{k}\}$ with

$$\lambda_{\ell,l}[\boldsymbol{k}] = \begin{cases} 0 & \text{if } \ell = 0 \\ \nu g\left(\sum_\ell |\alpha_{\ell,l}[\boldsymbol{k}]|^2\right) & \text{for } \ell \neq 0. \end{cases}$$

Here, ν is some fixed constant. The scalar function g can be chosen similarly as the PM equation [49].

Here, we present a numerical simulation of the adaptive soft-thresholding (**AST**) algorithm for the same image deblurring scenario we considered for SBreg, and PS algorithms. The original blur and noise free image is shown by the left image of Figure 6 and the blurry and noisy observed image $\boldsymbol{f}$ is shown in the middle image. The recovered image is shown by the right image in Figure 6.

Figure 6. Images from left to right are: the original image, the observed blurry and noisy image $\boldsymbol{f}$ and the reconstructed image using the AST algorithm.

Note that the adaptive soft-thresholding algorithm (3.12) is only one of many possible adaptive wavelet frame shrinkage algorithms. More algorithms can be found in [6] with more numerical simulations and comparisons with non-adaptive algorithms. For each of the adaptive algorithms, one can find its corresponding nonlinear evolution PDE based on the generic connections between (3.8) and (3.10). Also, all of the these algorithms can be understood as taking the form of the 3-steps procedure, which essentially indicates that the corresponding PDE based approach follows the 3-steps procedure as well in the discrete setting. Interested readers should consult [6] for details.

References

[1] K. Gröchenig, *Foundations of time-frequency analysis*. Birkhauser, 2001.

[2] S. Mallat, *A wavelet tour of signal processing*, vol. 2nd ed. New York: Academic. Academic press, 1999.

[3] A. Tikhonov, V. Arsenin, and F. John, *Solutions of ill-posed problems*. VH Winston Washington, DC, 1977.

[4] S. Jaffard, "Pointwise smoothness, two-microlocalization and wavelet coefficients," in *Publicacions Matematiques*, vol. 35, pp. 155–168, 1991.

[5] S. Jaffard, Y. Meyer, and R. D. Ryan, *Wavelets: tools for science and technology*. SIAM, 2001.

[6] B. Dong, Q. Jiang, and Z. Shen, "Image restoration: Wavelet frame shrinkage, nonlinear evolution pdes, and beyond," *Preprint*, 2013.

[7] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, "Maximum entropy and the nearly black object," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 41–81, 1992.

[8] J. Fan, "Comments on "Wavelets in statistics: A review" by A. Antoniadis," *Journal of the Italian Statistical Society*, vol. 6, no. 2, pp. 131–138, 1997.

[9] A. Antoniadis and J. Fan, "Regularization of wavelet approximations," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 939–967, 2001.

[10] J. Cai, R. Chan, and Z. Shen, "A framelet-based image inpainting algorithm," *Applied and Computational Harmonic Analysis*, vol. 24, no. 2, pp. 131–149, 2008.

[11] R. Chan, T. Chan, L. Shen, and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," *SIAM Journal on Scientific Computing*, vol. 24, no. 4, pp. 1408–1432, 2003.

[12] J. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, vol. 8, no. 2, pp. 337–369, 2009.

[13] J. Cai, S. Osher, and Z. Shen, "Linearized Bregman iterations for frame-based image deblurring," *SIAM J. Imaging Sci*, vol. 2, no. 1, pp. 226–252, 2009.

[14] Y. Zhang, B. Dong, and Z. Lu, "ℓ_0 minimization of wavelet frame based image restoration," *Mathematics of Computation*, vol. 82, pp. 995–1015, 2013.

[15] B. Dong and Y. Zhang, "An efficient algorithm for ℓ_0 minimization in wavelet frame based image restoration," *Journal of Scientific Computing*, vol. 54 (2-3), pp. 350–368, 2013.

[16] J. Liang, J. Li, Z. Shen, and X. Zhang, "Wavelet frame based color image demosaicing," *Inverse Problems and Imaging*, vol. 7, no. 3, pp. 777–794, 2013.

[17] L. Hou, H. Ji, and Z. Shen, "Recovering over-/underexposed regions in photographs.," *SIAM J. Imaging Sciences*, vol. 6, no. 4, pp. 2213–2235, 2013.

[18] J. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring using multiple images," *Journal of Computational Physics*, vol. 228, no. 14, pp. 5057–5071, 2009.

[19] J. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 104–111, IEEE, 2009.

[20] B. Dong, H. Ji, J. Li, Z. Shen, and Y. Xu, "Wavelet frame based blind image inpainting," *accepted by Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 268–279, 2011.

[21] Z. Gong, Z. Shen, and K.-C. Toh, "Image restoration with mixed or unknown noises," *Multiscale Modeling & Simulation*, vol. 12, no. 2, pp. 458–487, 2014.

[22] X. Jia, B. Dong, Y. Lou, and S. Jiang, "GPU-based iterative cone-beam CT reconstruction using tight frame regularization," *Physics in Medicine and Biology*, vol. 56, pp. 3787–3807, 2011.

[23] B. Dong, J. Li, and Z. Shen, "X-ray CT image reconstruction via wavelet frame based regularization and Radon domain inpainting," *Journal of Scientific Computing*, vol. 54 (2-3), pp. 333–349, 2013.

[24] M. Li, Z. Fan, H. Ji, and Z. Shen, "Wavelet frame based algorithm for 3d reconstruction in electron microscopy," *SIAM Journal on Scientific Computing*, vol. 36, no. 1, pp. B45–B69, 2014.

[25] Y. Quan, H. Ji, and Z. Shen, "Data-driven multi-scale non-local wavelet frame construction and image recovery," *Journal of Scientific Computing*, pp. 1–23, 2014.

[26] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 89–105, 2014.

[27] C. Bao, H. Ji, and Z. Shen, "Convergence analysis for iterative data-driven tight frame construction scheme," *Applied and Computational Harmonic Analysis*, 2014.

[28] H. Ji, Z. Shen, and Y. Zhao, "Directional frames for image recovery: multi-scale finite discrete gabor frames," *Preprint*, 2014.

[29] H. Ji, S. Huang, Z. Shen, and Y. Xu, "Robust video restoration by joint sparse and low rank matrix approximation," *SIAM Journal on Imaging Sciences*, vol. 4, no. 4, pp. 1122–1142, 2011.

[30] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao, "Robust principal component analysis-based four-dimensional computed tomography," *Physics in medicine and biology*, vol. 56, no. 11, p. 3181, 2011.

[31] J. Cai, X. Jia, H. Gao, S. Jiang, Z. Shen, and H. Zhao, "Cine cone beam ct reconstruction using low-rank matrix factorization: Algorithm and a proof-of-principle study.," *IEEE transactions on medical imaging*, vol. 33, no. 8, pp. 1581–1591, 2014.

[32] B. Dong, A. Chien, and Z. Shen, "Frame based segmentation for medical images," *Communications in Mathematical Sciences*, vol. 9(2), pp. 551–559, 2010.

[33] C. Tai, X. Zhang, and Z. Shen, "Wavelet frame based multiphase image segmentation," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2521–2546, 2013.

[34] H. Wendt, P. Abry, S. Jaffard, H. Ji, and Z. Shen, "Wavelet leader multifractal analysis for texture classification," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 3829–3832, IEEE, 2009.

[35] C. Bao, H. Ji, Y. Quan, and Z. Shen, "L0 norm based dictionary learning by proximal methods with global convergence," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 3858–3865, IEEE, 2014.

[36] Q. Jiang and D. K. Pounds, "Highly symmetric bi-frames for triangle surface multiresolution processing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 370–391, 2011.

[37] B. Dong, Q. Jiang, C. Liu, and Z. Shen, "Multiscale representation of surfaces by tight wavelet frames with applications to denoising," *accepted by Applied and Computational Harmonic Analysis*, 2015.

[38] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[39] M. Gavish, B. Nadler, and R. R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 367–374, 2010.

[40] N. Leonardi and D. Van De Ville, "Tight wavelet frames on multislice graphs," *Signal Processing, IEEE Transactions on*, vol. 61, no. 13, pp. 3357–3367, 2013.

[41] B. Dong, "Sparse representation on graphs by tight wavelet frames and applications," *Preprint*, 2014.

[42] G. Sapiro, *Geometric partial differential equations and image analysis*. Cambridge University Press, 2001.

[43] S. Osher and R. Fedkiw, *Level set methods and dynamic implicit surfaces*. Springer, 2003.

[44] T. Chan and J. Shen, *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. Society for Industrial Mathematics, 2005.

[45] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.

[46] J. Cai, B. Dong, S. Osher, and Z. Shen, "Image restorations: total variation, wavelet frames and beyond," *Journal of American Mathematical Society*, vol. 25(4), pp. 1033–1089, 2012.

[47] J. Cai, B. Dong, and Z. Shen, "Image restorations: a wavelet frame based model for piecewise smooth functions and beyond," *Preprint*, 2014.

[48] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on pure and applied mathematics*, vol. 42, no. 5, pp. 577–685, 1989.

[49] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[50] S. Osher and L. Rudin, "Feature-oriented image enhancement using shock filters," *SIAM Journal on Numerical Analysis*, vol. 27, pp. 919–940, Aug 1990.

[51] H. Ji, Y. Luo, and Z. Shen, "Image recovery via geometrically structured approximation," *Preprint*, 2015.

[52] A. Ron and Z. Shen, "Affine systems in $L_2(\mathbb{R}^d)$: The analysis of the analysis operator," *Journal of Functional Analysis*, vol. 148, no. 2, pp. 408–447, 1997.

[53] A. Ron and Z. Shen, "Affine systems in $L_2(\mathbb{R}^d)$ ii: dual systems," *Journal of Fourier Analysis and Applications*, vol. 3, no. 5, pp. 617–638, 1997.

[54] I. Daubechies, *Ten lectures on wavelets*, vol. CBMS-NSF Lecture Notes, SIAM, nr. 61. Society for Industrial and Applied Mathematics, 1992.

[55] I. Daubechies, B. Han, A. Ron, and Z. Shen, "Framelets: MRA-based constructions of wavelet frames," *Applied and Computational Harmonic Analysis*, vol. 14, pp. 1–46, Jan 2003.

[56] Z. Shen, "Wavelet frames and image restorations," in *Proceedings of the International Congress of Mathematicians*, vol. 4, pp. 2834–2863, 2010.

[57] B. Dong and Z. Shen, "MRA-Based Wavelet Frames and Applications," *IAS Lecture Notes Series, Summer Program on "The Mathematics of Image Processing", Park City Mathematics Institute*, 2010.

[58] S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of L 2 (R)," *Transactions of the American Mathematical Society*, vol. 315, no. 1, pp. 69–87, 1989.

[59] Y. Meyer, *Wavelets and operators. Translated by DH Salinger*. Cambridge Studies in Advanced Mathematics, 1992.

[60] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, 1988.

[61] R. Duffin and A. Schaeffer, "A class of nonharmonic Fourier series," *Transactions of the American Mathematical Society*, vol. 72, no. 2, pp. 341–366, 1952.

[62] A. Ron and Z. Shen, "Generalized shift-invariant systems," *Constructive Approximation*, vol. 22, no. 1, pp. 1–45, 2005.

[63] A. Ron and Z. Shen, "Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$," *Duke Mathematical Journal*, vol. 89, no. 2, pp. 237–282, 1997.

[64] A. Ron and Z. Shen, "Frames and Stable Bases for Shift-Invariant Subspaces of $L_2(\mathbb{R}^d)$," *Canadian Journal of Mathematics*, vol. 47, no. 5, pp. 1051–1094, 1995.

[65] A. Ron and Z. Shen, "Compactly supported tight affine spline frames in $L_2(\mathbb{R}^d)$," *Mathematics of Computation*, vol. 67, no. 221, pp. 191–207, 1998.

[66] B. Dong and Z. Shen, "Pseudo-splines, wavelets and framelets," *Applied and Computational Harmonic Analysis*, vol. 22, no. 1, pp. 78–104, 2007.

[67] B. Han, "Matrix splitting with symmetry and symmetric tight framelet filter banks with two high-pass filters," *Applied and Computational Harmonic Analysis*, vol. 35, no. 2, pp. 200–227, 2013.

[68] B. Han, "Symmetric tight framelet filter banks with three high-pass filters," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 140–161, 2014.

[69] Q. Jiang and Z. Shen, "Tight wavelet frames in low dimensions with canonical filters," *accepted by Journal of Approximation Theory*, 2015.

[70] C. Chui, W. He, and J. Stöckler, "Compactly supported tight and sibling frames with maximum vanishing moments," *Applied and computational harmonic analysis*, vol. 13, no. 3, pp. 224–262, 2002.

[71] B. Han and Z. Shen, "Dual wavelet frames and Riesz bases in Sobolev spaces," *Constructive Approximation*, vol. 29, no. 3, pp. 369–406, 2009.

[72] Z. Fan, H. Ji, and Z. Shen, "Dual gramian analysis: Duality principle and unitary extension principle," *Preprint*, 2013.

[73] Z. Fan, A. Heinecke, and Z. Shen, "Duality for frames," *Journal of Fourier Analysis and Applications*, vol. to appear, 2015.

[74] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[75] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[76] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[77] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, 2006.

[78] R. Coifman and D. Donoho, "Translation-invariant de-noising," *Wavelets and statistics*, vol. 103, p. 125, 1995.

[79] D. Donoho and I. Johnstone, "Threshold selection for wavelet shrinkage of noisy data," in *Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE*, pp. A24–A25, IEEE, 1994.

[80] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[81] Z. Lu and Y. Zhang, "Penalty decomposition methods for l_0-norm minimization," *Technical report, Department of Mathematics, Simon Fraser University, Canada*, 2010.

[82] J. Cai, R. Chan, L. Shen, and Z. Shen, "Convergence analysis of tight framelet approach for missing data recovery," *Advances in Computational Mathematics*, vol. 31, no. 1, pp. 87–113, 2009.

[83] J. Cai, R. Chan, and Z. Shen, "Simultaneous cartoon and texture inpainting," *Inverse Problems and Imaging (IPI)*, vol. 4, no. 3, pp. 379–395, 2010.

[84] J. Cai and Z. Shen, "Framelet based deconvolution," *J. Comp. Math*, vol. 28, no. 3, pp. 289–308, 2010.

[85] J. Cai, E. Candès, and Z. Shen, "A Singular Value Thresholding Algorithm for Matrix Completion," *SIAM Journal on Optimization*, vol. 20(4), pp. 1956–1982, 2010.

[86] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pp. 417–424, ACM Press/Addison-Wesley Publishing Co., 2000.

[87] T. Chan and J. Shen, "Variational image inpainting," *Commun. Pure Appl. Math*, vol. 58, pp. 579–619, 2005.

[88] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[89] Z. Shen, K.-C. Toh, and S. Yun, "An accelerated proximal gradient algorithm for frame-based image restoration via the balanced approach," *SIAM Journal on Imaging Sciences*, vol. 4, no. 2, pp. 573–596, 2011.

[90] P. Tseng, "Applications of a splitting algorithm to decomposition in convex programming and variational inequalities," *SIAM Journal on Control and Optimization*, vol. 29, no. 1, pp. 119–138, 1991.

[91] G. H. Chen and R. Rockafellar, "Convergence rates in forward–backward splitting," *SIAM Journal on Optimization*, vol. 7, no. 2, pp. 421–444, 1997.

[92] P. L. Combettes*, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization*, vol. 53, no. 5-6, pp. 475–504, 2004.

[93] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2006.

[94] E. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for ℓ_1-regularization with application to compressed sensing," *CAAM Technical Report TR, Rice University, Houston, TX*, pp. 07–07, 2007. CAAM Technical Report TR07-07, Rice University, Houston, TX.

[95] K. Bredies, "A forward–backward splitting algorithm for the minimization of non-smooth convex functionals in banach space," *Inverse Problems*, vol. 25, no. 1, p. 015005, 2009.

[96] I. Daubechies, G. Teschke, and L. Vese, "Iteratively solving linear inverse problems under general convex constraints," *Inverse Problems and Imaging*, vol. 1, no. 1, p. 29, 2007.

[97] M. Fadili and J. Starck, "Sparse representations and Bayesian image inpainting," *Proc. SPARS*, vol. 5, 2005.

[98] M. Fadili, J. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal*, vol. 52, no. 1, p. 64, 2009.

[99] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.

[100] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2, pp. II–782, IEEE, 2005.

[101] M. Elad, J. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[102] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE transactions on image processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[103] T. Goldstein and S. Osher, "The split Bregman algorithm for L1 regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[104] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM Journal on Imaging Sciences*, vol. 3, pp. 253–276, 2010.

[105] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM J. Imaging Sci*, vol. 1, no. 1, pp. 143–168, 2008.

[106] S. Osher, Y. Mao, B. Dong, and W. Yin, "Fast linearized bregman iteration for compressive sensing and sparse denoising," *Communications in Mathematical Sciences*, vol. 8, no. 2, pp. 93–111, 2010.

[107] T.-C. Chang, L. He, and T. Fang, "Mr image reconstruction from sparse radial samples using bregman iteration," in *Proceedings of the 13th Annual Meeting of ISMRM, Seattle*, p. 696, 2006.

[108] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation based image restoration," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 460–489, 2005.

[109] L. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 3, pp. 200–217, 1967.

[110] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[111] D. Bertsekas and J. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice-Hall, Inc., 1989.

[112] J. Eckstein and D. Bertsekas, "On the douglasrachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, 1992.

[113] M. Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969.

[114] M. Powell, "A method for non-linear constraints in minimization problems," *Optimization, Ed. R. Fletcher (Academic Press, New York)*, pp. 283–298, 1969.

[115] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*. Society for Industrial and Applied Mathematics, 1989.

[116] E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split Bregman," *CAM Report*, vol. 9-31, 2009.

[117] I. Ekeland and R. Temam, *Convex analysis and variational problems*. Classics in Applied Mathematics, Society for Industrial Mathematics, 1999.

[118] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *preprint*, Jan 2008.

[119] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of ACM*, vol. 58(1), pp. 1–37, 2009.

[120] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma, "Stable principal component pursuit," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 1518–1522, IEEE, 2010.

[121] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[122] H. Jiang, W. Deng, and Z. Shen, "Surveillance video processing using compressive sensing," *Inverse Problems and Imaging*, vol. 6, no. 2, pp. 201–214, 2012.

[123] F. Yang, H. Jiang, Z. Shen, W. Deng, and D. Metaxas, "Adaptive low rank and sparse decomposition of video using compressive sensing," in *IEEE International Conference on Image Processing (ICIP)*, 2013.

[124] H. Jiang, S. Zhao, Z. Shen, W. Deng, P. A. Wilford, and R. Haimi-Cohen, "Surveillance video analysis using compressive sensing with low latency," *Bell Labs Technical Journal*, vol. 18, no. 4, pp. 63–74, 2014.

[125] M. Lewicki and T. Sejnowski, "Learning overcomplete representations," *Neural computation*, vol. 12, no. 2, pp. 337–365, 2000.

[126] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural computation*, vol. 15, no. 2, pp. 349–396, 2003.

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

[128] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *Image Processing, IEEE Transactions on*, vol. 17, no. 1, pp. 53–69, 2008.

[129] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 2272–2279, IEEE, 2009.

[130] W. Dong, X. Li, D. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 457–464, IEEE, 2011.

[131] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, vol. 54, no. 11, p. 4311, 2006.

[132] C. Tai and W. E, "Multiscale adaptive representation of signals: I. the basic framework," *Preprint*, 2015.

[133] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 60–65, IEEE, 2005.

[134] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model Sim*, vol. 7, pp. 1005–1028, Jan 2008.

[135] G. Peyré, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *Computer Vision–ECCV 2008*, pp. 57–68, Springer, 2008.

[136] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, "Image recovery via nonlocal operators," *Journal of Scientific Computing*, vol. 42, no. 2, pp. 185–197, 2010.

[137] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007.

[138] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image restoration by sparse 3d transform-domain collaborative filtering," in *Electronic Imaging 2008*, pp. 681207–681207, International Society for Optics and Photonics, 2008.

[139] A. Danielyan, V. Katkovnik, and K. Egiazarian, "Bm3d frames and variational image deblurring," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1715–1728, 2012.

[140] G. Aubert and P. Kornprobst, *Mathematical problems in image processing: partial differential equations and the calculus of variations*. Springer, 2006.

[141] A. Chambolle, "Finite-differences discretizations of the mumford-shah functional," *ESAIM: Mathematical Modelling and Numerical Analysis*, vol. 33, no. 02, pp. 261–288, 1999.

[142] G. Dal Maso, *Introduction to* Γ*-convergence*. Birkhauser, 1993.

Beijing International Center for Mathematical Research (BICMR), Peking University, 5 Yiheyuan Road, Beijing, China, 100871.
E-mail: dongbin@math.pku.edu.cn

Department of Mathematics, National University of Singapore, Block S17, 10 Lower Kent Ridge Road, Singapore, 119076.
E-mail: matzuows@nus.edu.sg