

# IMAGE RESTORATION: TOTAL VARIATION, WAVELET FRAMES, AND BEYOND

JIAN-FENG CAI, BIN DONG, STANLEY OSHER, AND ZUOWEI SHEN

ABSTRACT. The variational techniques (e.g., the total variation based method [1]) are well established and effective for image restoration, as well as many other applications, while wavelet frame based approach is relatively new and came from a different school. This paper is designed to establish a connection between these two major approaches for image restoration. The main result of this paper shows that when spline wavelet frames of [2] are used, a special model of a wavelet frame method, called the analysis based approach, can be viewed as a discrete approximation at a given resolution to variational methods. A convergence analysis as image resolution increases is given in terms of objective functionals and their approximate minimizers. This analysis goes beyond the establishment of the connections between these two approaches, since it leads to new understandings for the both approaches. First, it provides geometric interpretations to the wavelet frame based approach as well as its solutions. On the other hand, for any given variational model, wavelet frame based approaches provide various and flexible discretizations which immediately lead to fast numerical algorithms for both the wavelet frame based approaches and the corresponding variational model. Furthermore, the built-in multiresolution structure of wavelet frames can be utilized to adaptively choose proper differential operators in different regions of a given image according to the order of the singularity of the underlying solutions. This is important when multiple orders of differential operators are used in various models that generalize the total variation based method. These observations will enable us to design new methods according to the problems at hand, hence, lead to wider applications of both the variational and wavelet frame based approaches. Links of wavelet frame based approaches to some more general variational methods developed recently will also be discussed.

## 1. Introduction

From the beginning of science, visual observations have been playing important roles. Advances in computer technology have made it possible to apply some of the most sophisticated developments in mathematics and the 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 applied to virtually all natural sciences and technical disciplines ranging from computer sciences and electronic engineering to biology and medical sciences; and digital images have come into everyone's life. Image restoration, including image denoising, deblurring, inpainting, computed tomography, etc., is one of the most

---

2010 *Mathematics Subject Classification.* 35A15, 41A25, 42C40, 45Q05, 65K10, 68U10.

*Key words and phrases.* Image restoration, total variation, variational method, (tight) wavelet frames, framelets, split Bregman, pointwise convergence,  $\Gamma$ -convergence.

Received by the editors April 3, 2011; in revised form, March 26, 2012.

important areas in image processing and analysis. Its major purpose is to enhance the quality of a given image that is corrupted in various ways during the process of imaging, acquisition and communication, and enables us to see crucial but subtle objects reside in the image. Therefore, image restoration is an important step to take towards the accurate interpretations of the physical world and making the optimal decisions.

Mathematics has been playing an important role in image and signal processing from the very beginning; for example, Fourier analysis is one of the main tools in signal and image analysis, processing, and restoration. In fact, mathematics has been one of the driving forces of the modern development of image analysis, processing and restorations. At the same time, the interesting and challenging problems in imaging science also gave birth to new mathematical theories, techniques and methods. The variational methods (e.g., total variation based methods) and wavelets and wavelet frame based methods developed in the last few decades for image and signal processing are two successful recent examples among many. This paper is designed to establish connections between these two major image restoration approaches: variational methods and wavelet frame based methods. Such connections provide new interpretations and understanding of both approaches, and hence, lead to new applications for both approaches.

We start with an introduction of both the variational and wavelet frame based methods. The basic linear image restoration model used for variational methods is usually given as

$$(1.1) \quad f = Au + \eta,$$

where  $A$  is some linear bounded operator (not invertible in general) mapping one function space to another, e.g., the identity operator for image denoising, or a convolution operator for image deconvolution, and  $\eta$  denotes a perturbation caused by the additive noise in the observed image (or measurements), which is typically assumed to be a white Gaussian noise.

In order to recover the unknown image  $u$  from the equation (1.1), a typical variational approach takes the following form:

$$(1.2) \quad \inf_u \nu \|Du\|_* + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2,$$

where  $D$  is a vector of some (weighted) differential operators,  $\|\cdot\|_*$  is some properly chosen norm,  $\Omega$  is some domain in  $\mathbb{R}^2$ , and  $\nu$  is some scalar parameter. For example, when  $D = \nabla$  and  $\|\nabla u\|_* := \int_{\Omega} |\nabla u| dx dy$ , then (1.2) is the well-known Rudin-Osher-Fatemi (ROF) model [1]:

$$(1.3) \quad \inf_u \nu \int_{\Omega} |\nabla u| dx dy + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2.$$

the ROF model works exceptionally well in terms of preserving edges while suppressing noise. After the ROF model was proposed, there were many variational methods developed in the literature. We shall provide a short review of variational methods in the next section.

In general, the infimum of a convex objective functional, such as the ones in (1.2) and (1.3), may not be attainable. Some additional conditions, such as coercivity [3], are needed in order to ensure that. However, practically, we only need to deal with approximate minimizers, or more precisely,  $\epsilon$ -optimal solutions (see (3.7) for the definition) when solving image restoration problems, which always exist

and are numerically computable for all objective functionals considered in this paper. Therefore, we will not enforce additional conditions on  $\mathbf{D}$  and  $\mathbf{A}$  to ensure attainability of the infimum.

The image restoration model (1.1) views images as functions defined on a continuum, i.e. analog signals. What we observe in practice, however, are digital images which are discrete versions of their continuous counterparts. Given an analog data  $f$ , its digital/discrete version  $\mathbf{f}$  can be obtained by various sampling schemes. However, digital images are never given in a function form. Hence, when model (1.1) is adopted, one needs to obtain an approximation of an observed function from the given digital image. Since digital images are discrete sequences and the restoration of a digital image is to restore a sequence from an observed sequence, it is more natural to view a digital image as a sequence  $\mathbf{f}$  and establish a restoration model in a sequence space instead of a function space. This is what the wavelet frame based approach is for. In fact, as we will see, the wavelet frame based approach naturally fits the discrete setting of digital images.

The digital image restoration model is the discrete version of (1.1), which is to find the unknown true digital image  $\mathbf{u}$  from an observed image (or measurements)  $\mathbf{f}$  defined by

$$(1.4) \quad \mathbf{f} = \mathbf{A}\mathbf{u} + \boldsymbol{\eta},$$

where  $\mathbf{A}$  is a linear operator and  $\boldsymbol{\eta}$  is a Gaussian noise. Here the linear operator  $\mathbf{A}$  is the identity operator for image denoising, a convolution operator for image deconvolution, and a projection operator for image inpainting.

The problem (1.4) is a typical linear inverse problem. The frame based image restoration model that we shall focus on in this paper is the analysis based approach, which is to solve

$$(1.5) \quad \inf_{\mathbf{u}} \nu \|\boldsymbol{\lambda} \cdot \mathbf{W}\mathbf{u}\|_* + \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2,$$

where  $\mathbf{W}$  is the discrete wavelet frame transform using filters of some tight wavelet frame system, and  $\|\cdot\|_*$  is some properly chosen norm that reflects the regularity or sparse properties of the underlying solutions, e.g., the weighted  $\ell_1$ -norm is commonly used. We will provide details on wavelet frame transforms and the definition of  $\|\cdot\|_*$  in the next section. In order to have a desired solution of (1.4) via solving (1.5), the underlying solution should have a good sparse approximation under the wavelet frame transform  $\mathbf{W}$ . One of the advantages of using tight wavelet frame systems is that they provide reasonably good sparse approximations to piecewise smooth functions, which form a large class of functions that images belong to.

When the wavelet frame based approach (1.5) is used, the digital image data is given in a sequence form and the minimization is also done in sequence space. Hence the solution is naturally a sequence. The underlying function from which the digital image is sampled does not appear explicitly. It appears implicitly when the regularity of the underlying solution is mentioned, which is stated in terms of the decay of the wavelet frame coefficients  $\mathbf{W}\mathbf{u}$ . Furthermore, one can obtain an approximate solution in function form whenever it is needed. Indeed, since wavelet frame based approaches normally generate coefficients of a wavelet frame system (or coefficients of shifts of the corresponding refinable function that generates the wavelet frame system), it is easy to reconstruct a function from these coefficients.

In fact, as we will see, when wavelet frames are used, we can interpret the digital image  $\mathbf{f}$  as  $\mathbf{f}[\mathbf{k}] = \langle f, \phi(\cdot - \mathbf{k}) \rangle$  for some function  $\phi$  to link it to function

space. The operator  $\mathbf{A}$  can be viewed as a discrete version of the continuous linear operator  $A$  of (1.1) through certain discretization. Furthermore, when special wavelet frames are chosen, the wavelet frame transform  $\boldsymbol{\lambda} \cdot \mathbf{W}\mathbf{u}$  can be regarded as certain discretizations of  $\mathbf{D}u$ . This motivates us to explore whether the model (1.5) approximates (1.2) and whether the approximation becomes more accurate when image resolution increases. One of the major contributions of this paper is to establish a connection between the wavelet frame based image variational approach (1.5) and the differential operator based variational model (1.2), as well as their corresponding approximate minimizers. An approximate minimizer is the one on which the value of the objective functional is close to the infimum.

In order to briefly summarize our results, we need to introduce some notation. The details can be found in later sections. First, define the objective function

$$F_n(\mathbf{u}) := \nu \|\boldsymbol{\lambda} \cdot \mathbf{W}\mathbf{u}\|_* + \frac{1}{2} \|\mathbf{A}_n \mathbf{u} - \mathbf{f}\|_2^2,$$

where  $\mathbf{u}$  and  $\mathbf{f}$  are  $N \times N$  with  $N = 2^n + 1$ . The operator  $\mathbf{A}_n$  should be understood as a proper discretization of some operator  $A$  defined on a function space. We will first show that the approximate minimizers of  $F_n$  can be put into correspondence with those of the following objective functional:

$$E_n(u) := \nu \|\boldsymbol{\lambda} \cdot \mathbf{W}_n \mathbf{T}_n u\|_* + \frac{1}{2} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2^2,$$

where  $\mathbf{T}_n$  is some discretization operator associated to the given wavelet frame system. In particular, if  $\mathbf{u}_n^*$  is an (approximate) minimizer of  $F_n$ , then we can explicitly construct a  $u_n^*$  such that  $u_n^*$  is an (approximate) minimizer of  $E_n$ . The converse argument is also true. Note that  $E_n$  is a functional defined on some function space, as well as the objective functional

$$E(u) := \nu \|\mathbf{D}u\|_* + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2.$$

For clarity of the presentation here, we postpone the detailed definition of the function space of  $u$  and the domains of the operators  $\mathbf{D}$ ,  $\mathbf{W}$  and  $\mathbf{T}_n$  to later sections.

While  $E_n$  is essentially a reformulation of  $F_n$ , it can also be regarded as an approximation of  $E$ . For this, we first show that  $E_n$  pointwise converges to  $E$ . Then, we prove that  $E_n$   $\Gamma$ -converges to  $E$  (see Definition 3.1 for the definition of  $\Gamma$ -convergence; also see, e.g., [4] for an introduction of  $\Gamma$ -convergence). In fact, we shall prove a convergence result that is stronger than the  $\Gamma$ -convergence. The analysis is based on approximation of functions through multiresolution structures generated by B-splines and their associated spline wavelet frames constructed from [2]. In numerical computation for both the variational and wavelet frame based approaches, the computed solutions are often those whose values of the corresponding objective functional  $E$  or  $E_n$  are  $\epsilon$  close to the infimum. We refer to such solutions as  $\epsilon$ -optimal solutions to  $E$  or  $E_n$  (see (3.7) for the definition). Denote by  $u_n^*$  an  $\epsilon$ -optimal solution to  $E_n$ , which can be constructed by the  $\epsilon$ -optimal solution to  $F_n$ , denoted as  $\mathbf{u}_n^*$ . As a result of  $\Gamma$ -convergence, we are able to connect the functional  $E_n$  with  $E$ , hence  $F_n$  with  $E$ , in the following sense:

$$\limsup_{n \rightarrow \infty} F_n(\mathbf{u}_n^*) = \limsup_{n \rightarrow \infty} E_n(u_n^*) \leq \inf_u E(u) + \epsilon.$$

The above inequality shows that  $\{u_n^*\}$  is a sequence that enables  $E_n$  to approximate the optimal value of  $E$  from below. Furthermore, any cluster point of  $\{u_n^*\}$  is an  $\epsilon$ -optimal solution to  $E$ .

The discussions of the similarity and difference between the total variation method and wavelet method for image denoising have already been given through the discussions of the space of bounded variation functions (BV space) and the Besov space (e.g.,  $B_1^{1,1}$  space) in [5], which provides a fundamental understanding of the two approaches for image denoising. The results given here not only consider more general image restoration problems, but also reveal close connections between the solutions of wavelet frame based image restoration, which is used in numerical computing, and those of the total variation method. Furthermore, since the number of levels used in wavelet decomposition is fixed independent of the resolution (though the resolution can be higher when the give data is denser) and since the parameters for the low pass filter are always set to be zero in numerical implementation of (1.5), the weighted  $\ell_1$ -norm of wavelet coefficients here is not equivalent to the Besov norm as discussed in [5]. All these indicate that the study here takes a different approach and leads to a different set of results which will extend our understanding for both the wavelet frame based and variational approaches. Our conclusion goes beyond the theoretic justifications of the linkage of the two approaches. Since the total variation approach has a strong geometric interpretation, this connection gives geometric interpretations to the wavelet frame based approach (1.5) as well as its minimizers, as it can be understood as the discrete form of (1.2). This also leads to even wider applications of the wavelet frame based approach, e.g., image segmentation [6] and 3D surface reconstruction from unorganized point sets [7]. On the other hand, for any given variational model (1.2), (1.5) provides various and flexible discretizations, as well as fast numerical algorithms. Using a wavelet frame based approach, we can approximate various differential operators for model (1.2) by choosing a proper wavelet frame transform and parameters  $\lambda$ . Furthermore, by putting (1.2) into a wavelet frame setting, one can use a multiresolution structure to adaptively choose proper differential operators in different regions of a given image according to the orders of the singularities of underlying solutions. It should be pointed out here that if one wants to use a more general differential operator in (1.2), the ability of applying different differential operators according to where various singularities are located is the key to make such a generalization successful. The wavelet frame based approach has a built-in adaptive mechanism via the multiresolution analysis that provides a natural tool for this very purpose. Finally, we can use more general wavelet frame based approaches, e.g., the wavelet frame based balanced approach, or using two-system models to solve various generalizations of (1.2), as will be shown in a later section.

The rest of this paper is organized as follows. In Section 2, we will first review some of the classical as well as some most recent variational approaches. Then we will review some basic notions of wavelet frames, fast framelet transforms, and wavelet frame based image restoration approaches. Finally, we will motivate the readers by showing the link between the wavelet frame based approach with Haar framelet and the ROF model with  $A = I$ . Theoretical justifications of the link between (1.2) and (1.5) will be given in Section 3 with some technical details given later in Section 4. In particular, in Section 3, we will prove that  $E_n$  pointwise converges to  $E$ , which indicates that  $E_n$  can be used to approximate  $E$ . In order to

understand further connections between  $E_n$  and  $E$ , another convergence result of  $E_n$  to  $E$  is also proven. This result implies that  $E_n$   $\Gamma$ -converges to  $E$ . It also leads to the fact that the approximate minimizers of  $E$  can be bounded below in general by  $\limsup_{n \rightarrow \infty} (\inf E_n)$  up to some small  $\epsilon$ . Furthermore, approximated minimizers of  $E$  can be directly approximated by those of  $E_n$  for many cases. All these provide a fundamental link between the optimization problems (1.2) and (1.5), and a link between their approximate minimizers. Numerical algorithms that solve (1.5) as well as some simulations will be provided in Section 5. Finally, in Section 6, we will present some further extensions of (1.5) and their connections to some recently proposed variational methods.

## 2. Models and Motivations

In this section, we first recall some variational methods, as well as their history and recent developments. Then we review some basic notation of wavelet frames and wavelet frame based image restoration approaches. At the end, we shall give an example to illustrate that these two approaches are closely related.

### 2.1. Total Variation Based Models and Their Generalizations.

2.1.1. *Total Variation Based Models.* The trend of variational methods and partial differential equation (PDE) based image processing started with the refined Rudin-Osher-Fatemi (ROF) model [1]:

$$(2.1) \quad \inf_u \nu \int_{\Omega} |\nabla u| dx dy + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2.$$

The ROF model aims at finding a desirable solution of (1.1). The choice of regularization  $\int_{\Omega} |\nabla u| dx dy$  is the total variation (TV) of  $u$ , which is much more effective than the classical choice  $\int_{\Omega} |\nabla u|^2 dx dy$  in terms of preserving sharp edges, which are key features for images. Many of the current PDE based methods for image denoising and decomposition utilize TV regularization for its beneficial edge-preserving property (see, e.g., [5, 8, 9]). The ROF model is especially effective on restoring images that are piecewise constant, e.g., binary images.

To solve the ROF model (2.1), one can solve the corresponding Euler-Lagrange equation:

$$(2.2) \quad -\nu \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) + A^{\top}(Au - f) = 0.$$

To solve the Euler-Lagrange equation (2.2), it was proposed in [1] to use a time evolution PDE which is essentially the steepest descent equation for the objective functional (2.1). This time evolution PDE was later modified by [10] to improve the Courant-Friedrichs-Lewy (CFL) condition when solving the PDE using explicit finite difference schemes. To completely remove the CFL restriction, a fixed point iteration scheme which solves the Euler-Lagrange equation (2.2) directly was proposed in [11].

All the numerical methods in [1, 10, 11] aimed at solving the Euler-Lagrange equation (2.2), which becomes singular when  $|\nabla u| = 0$ . To avoid this, [1, 10, 11] used a regularized parabolic term for (2.2) instead, i.e.

$$-\nu \nabla \cdot \left( \frac{\nabla u}{|\nabla u|_{\epsilon}} \right) + A^{\top}(Au - f) = 0, \quad \text{where } |\nabla u|_{\epsilon} := \sqrt{|\nabla u|^2 + \epsilon^2}.$$

Although such regularization prevents the coefficients of the parabolic term of (2.2) from getting arbitrarily large, a large  $\epsilon$  also reduces the ability of the ROF model to preserve edges. On the other hand, when  $\epsilon$  is small, the efficiency of these algorithms will also be significantly reduced. To overcome this dilemma, a duality based method was proposed by [12, 13] exploiting the dual formula of the ROF model. Recent developments of the duality based method can be found in [14, 15, 16, 17, 18] and the reference therein.

The more recently developed efficient algorithm for the ROF model (2.1) is the split Bregman algorithm proposed by [19] that is based on the Bregman distance [20]. In fact, the split Bregman algorithm of [19] solves a wavelet frame based approach implicitly when a proper discretization of the differential operator  $\nabla$  is chosen. For example, one may choose the following discretizations:

$$\partial_x u \approx \frac{\mathbf{u}[i+1, j] - \mathbf{u}[i, j]}{h} \quad \text{and} \quad \partial_y u \approx \frac{\mathbf{u}[i, j+1] - \mathbf{u}[i, j]}{h}.$$

Then, with a proper choice and organization of the above difference operators and parameter  $\lambda$  (see, e.g., Section 2.3), the split Bregman algorithm solving the ROF model (2.1) can be viewed as solving the following optimization problem:

$$(2.3) \quad \inf_{\mathbf{u}} \nu \|\lambda \mathbf{H} \mathbf{u}\|_{1,2} + \frac{1}{2} \|\mathbf{A} \mathbf{u} - \mathbf{f}\|_2^2,$$

with  $\mathbf{H}$  being the Haar wavelet transform. Model (2.3) is a discretization of the ROF denoising model (2.1) (see (2.19) for the definition of  $\|\cdot\|_{1,2}$ ). In other words, instead of solving the ROF model (2.1), the split Bregman algorithm of [19] solves the wavelet frame based approach (2.3) if a proper discretization is used (see Section 2.3 for details). This shows that when the split Bregman algorithm of [19] is used to obtain efficient algorithms solving the ROF model (2.1), one implicitly solves a wavelet frame based model (1.5) when a proper discretization is used. The efforts made in seeking efficient algorithms for the ROF model (2.1) came across with wavelet frame based approaches at this point. Convergence analysis of the split Bregman algorithm for frame based image restoration was given by [21]. The connections of the split Bregman algorithm to some earlier algorithms were later discovered by [22, 23]. We shall review split Bregman algorithm in Section 5.

*2.1.2. Generalizations of Total Variation Based Models.* As we mentioned earlier, the ROF model is especially effective on restoring images that are piecewise constant, e.g., binary images. However, it will also create the staircase effect for images that are not piecewise constant [24, 25].

To reduce the staircase effect of the ROF model, [26] understood the image  $u$  as an addition of two image layers, i.e.  $u = u_1 + u_2$  and proposed the inf-convolution model:

$$(2.4) \quad \inf_{u_1, u_2} \int_{\Omega} \nu_1 |\nabla u_1| + \nu_2 |\nabla^2 u_2| dx dy + \frac{1}{2} \|A(u_1 + u_2) - f\|_{L_2(\Omega)}^2,$$

where  $|\nabla^2 u_2| := \sqrt{|\partial_{xx} u_2|^2 + |\partial_{yy} u_2|^2 + 2|\partial_{xy} u_2|^2}$ . Following the idea of the inf-convolution model, [27, 28] proposed the following model where the Hessian  $\nabla^2$  used in the inf-convolution model is replaced by the Laplacian  $\Delta$ :

$$\inf_{u_1, u_2} \int_{\Omega} \nu_1 |\nabla u_1| + \nu_2 |\Delta u_2|^p dx dy + \frac{1}{2} \|A(u_1 + u_2) - f\|_{L_2(\Omega)}^2, \quad p = 1, 2.$$

Another approach reducing the staircase effect without multi-layer assumptions of the images is adding a higher order regularization term to the original ROF model by [29]. Thanks to the higher order term, the solution prefers “ramps” over “staircases” (see [29] for details).

More recently, a more general variational approach was proposed by [30], where the authors introduced total generalized variation (TGV) as the regularization term. Here we will recall a modified version of the general TGV model (we shall refer to it as the TGV model in this paper):

$$(2.5) \quad \inf_{u,v} \nu_1 \|\nabla u - v\|_{L_1(\Omega)} + \nu_2 \|\nabla \cdot v\|_{L_1(\Omega)} + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2,$$

where the variable  $v = (v_1, v_2)$  varies in the space of all continuously differential 2-tensors and

$$\|\nabla \cdot v\|_{L_1(\Omega)} := \int_{\Omega} \sqrt{(\partial_x v_1)^2 + (\partial_y v_1)^2 + (\partial_x v_2)^2 + (\partial_y v_2)^2} dx dy.$$

The TGV model generalizes the inf-convolution model in the sense that it coincides with the inf-convolution model when  $v$  only varies in the range of  $\nabla$ . The TGV model improves the results for image denoising over the ROF model and inf-convolution model. The interested reader should consult [30] for details.

Similar to total variation based methods, we will establish the connections between the inf-convolution model (2.4) and the TGV model (2.5) with some wavelet frame based approach. Both the inf-convolution model (2.4) and the TGV model (2.5) can be approximated by some tight wavelet frame based approach, while it is more natural and efficient to have a frame based approach for these generalizations as well. We shall postpone the detailed discussions to Section 6.

**2.2. Wavelet Frame Based Models.** This section is devoted to recalling the wavelet frame based image restoration approaches. We start with the concept of a tight frame and a tight wavelet frame, and then introduce the analysis based image restoration approach and other more general wavelet frame based approaches.

**2.2.1. MRA-Based Tight Frames.** In this subsection, we briefly introduce the concept of tight frames and tight wavelet frames. The interested readers should consult [2, 31, 32] for theories of frames and wavelet frames, [33] for a short survey on the theory and applications of frames, and [34] for a more detailed survey.

A countable set  $X \subset L_2(\mathbb{R}^d)$ , with  $d \in \mathbb{Z}^+$ , is called a tight frame of  $L_2(\mathbb{R}^d)$  if

$$(2.6) \quad f = \sum_{g \in X} \langle f, g \rangle g \quad \forall f \in L_2(\mathbb{R}^d),$$

where  $\langle \cdot, \cdot \rangle$  is the inner product of  $L_2(\mathbb{R}^d)$ .

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

$$(2.7) \quad X^J(\Psi) = \{\psi_{\ell, n, \mathbf{k}} : 1 \leq \ell \leq q; n \in \mathbb{Z}, \mathbf{k} \in \mathbb{Z}^d\},$$

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

$$(2.8) \quad \psi_{\ell, n, \mathbf{k}} := \begin{cases} 2^{\frac{n d}{2}} \psi_{\ell}(2^n \cdot - \mathbf{k}), & n \geq J; \\ 2^{(n - \frac{J}{2})d} \psi_{\ell}(2^n \cdot - 2^{n-J} \mathbf{k}), & n < J. \end{cases}$$

When  $X^J(\Psi)$  forms a tight frame of  $L_2(\mathbb{R}^d)$ , each function  $\psi_\ell$ ,  $\ell = 1, \dots, q$ , is called a (tight) framelet and the whole system  $X^J(\Psi)$  is called a tight wavelet frame. In this section, we shall focus on the case  $J = 0$ , where we simply denote  $X(\Psi) := X^0(\Psi)$ . Similar arguments hold for general  $J$  (see [34]). Note that here we only discuss the quasi-affine system (2.8), since we only use undecimated wavelet frames which implicitly use quasi-affine frames. The interested reader can find further details on the affine wavelet frame systems and its connections to the quasi-affine frames [2, 34, 35].

The constructions of framelets  $\Psi$ , which are desirably (anti)symmetric and compactly supported functions, are usually based on a multiresolution analysis (MRA) that is generated by some refinable function  $\phi$  with refinement mask  $\mathbf{a}_0$  satisfying

$$(2.9) \quad \phi = 2^d \sum_{\mathbf{k} \in \mathbb{Z}^d} \mathbf{a}_0[\mathbf{k}] \phi(2 \cdot -\mathbf{k}).$$

The idea of an MRA-based construction of framelets  $\Psi = \{\psi_1, \dots, \psi_q\} \subset L_2(\mathbb{R}^d)$  is to find masks  $\mathbf{a}_\ell$ , which are finite sequences, such that

$$(2.10) \quad \psi_\ell = 2^d \sum_{\mathbf{k} \in \mathbb{Z}^d} \mathbf{a}_\ell[\mathbf{k}] \phi(2 \cdot -\mathbf{k}), \quad \ell = 1, 2, \dots, q.$$

The sequences  $\mathbf{a}_1, \dots, \mathbf{a}_q$  are called wavelet frame masks, or the high pass filters of the system, and the refinement mask  $\mathbf{a}_0$  is also known as the low pass filter.

The unitary extension principle (UEP) of [2] provides a general theory of the construction of MRA-based tight wavelet frames. Roughly speaking, as long as  $\{\mathbf{a}_1, \dots, \mathbf{a}_q\}$  are finitely supported and their Fourier series satisfy

$$(2.11) \quad \sum_{\ell=0}^q |\widehat{\mathbf{a}}_\ell(\xi)|^2 = 1 \quad \text{and} \quad \sum_{\ell=0}^q \widehat{\mathbf{a}}_\ell(\xi) \overline{\widehat{\mathbf{a}}_\ell(\xi + \nu)} = 0,$$

for all  $\nu \in \{0, \pi\}^d \setminus \{\mathbf{0}\}$  and  $\xi \in [-\pi, \pi]^d$ , the quasi-affine system  $X(\Psi)$  (as well as the traditional wavelet system) with  $\Psi = \{\psi_1, \dots, \psi_q\}$  defined by (2.10) forms a tight frame in  $L_2(\mathbb{R}^d)$ .

We now show three simple but useful examples of univariate framelets. The framelet given in Example 2.1 is known as the Haar wavelet. When one uses a wavelet (affine) system, it generates an orthonormal basis of  $L_2(\mathbb{R})$ . The quasi-affine system that the Haar wavelet generates, however, is not an orthonormal basis, but a tight frame of  $L_2(\mathbb{R})$  instead. We shall refer to  $\psi_1$  in Example 2.1 as the ‘‘Haar framelet’’. The framelets given by Example 2.2 and Example 2.3 are constructed from piecewise linear and cubic B-splines respectively first given in [2] (see [2] for details). The masks of those framelets are exactly discrete difference operators up to a scaling. These framelets are widely used in frame based image restoration problems because they provide sparse approximations to piecewise smooth, especially piecewise linear, functions such as images (see, e.g., [35, 36, 37, 38, 39, 40, 41, 42, 43, 21]). We shall refer to  $\psi_1$  and  $\psi_2$  in Example 2.2 as ‘‘piecewise linear framelets’’, and  $\psi_1$ ,  $\psi_2$  and  $\psi_3$  in Example 2.3 as ‘‘piecewise cubic framelets’’.

Notice that the framelet masks shown by the following examples correspond to standard difference operators up to some proper scaling, which is also true for framelets constructed by higher order B-splines [2]. This is a crucial observation that made us believe that a link does exist between the variational and wavelet frame based methods.

In this paper, the B-splines are the centered B-splines defined as follows. The centered B-spline of order  $m$ , denoted as  $\phi$ , is defined by its Fourier transform  $\widehat{\phi}$  as

$$\widehat{\phi}(\omega) = e^{-\frac{i\omega j_m}{2}} \left( \frac{\sin(\frac{\omega}{2})}{\frac{\omega}{2}} \right)^m,$$

where  $j_m := 0$  when  $m$  is even and  $j_m := 1$  when  $m$  is odd.

**Example 2.1.** Let  $\mathbf{a}_0 = \frac{1}{2}[1, 1]$  be the refinement mask of the piecewise constant B-spline  $B_1(x) = 1$  for  $x \in [0, 1]$  and 0 otherwise. Define  $\mathbf{a}_1 = \frac{1}{2}[1, -1]$ . Then  $\mathbf{a}_0$  and  $\mathbf{a}_1$  satisfy (2.11). Hence, the system  $X(\psi_1)$  defined in (2.7) is a tight frame of  $L_2(\mathbb{R})$ . The mask  $\mathbf{a}_1$  corresponds to a first order difference operator up to a scaling.

**Example 2.2.** Let  $\mathbf{a}_0 = \frac{1}{4}[1, 2, 1]$  be the refinement mask of the piecewise linear B-spline  $B_2(x) = \max(1 - |x|, 0)$ . Define  $\mathbf{a}_1 = \frac{\sqrt{2}}{4}[1, 0, -1]$  and  $\mathbf{a}_2 = \frac{1}{4}[-1, 2, -1]$ . Then  $\mathbf{a}_0$ ,  $\mathbf{a}_1$  and  $\mathbf{a}_2$  satisfy (2.11). Hence, the system  $X(\Psi)$  where  $\Psi = \{\psi_1, \psi_2\}$  defined in (2.7) is a tight frame of  $L_2(\mathbb{R})$ . The masks  $\mathbf{a}_1$  and  $\mathbf{a}_2$  correspond to the first order and second order difference operators respectively up to a scaling.

**Example 2.3.** Let  $\mathbf{a}_0 = \frac{1}{16}[1, 4, 6, 4, 1]$  be the refinement mask of the piecewise cubic B-spline  $B_4$ . Define  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$  as follows:

$$\begin{aligned} \mathbf{a}_1 &= \frac{1}{8}[-1, -2, 0, 2, 1], & \mathbf{a}_2 &= \frac{\sqrt{6}}{16}[1, 0, -2, 0, 1], \\ \mathbf{a}_3 &= \frac{1}{8}[-1, 2, 0, -2, 1], & \mathbf{a}_4 &= \frac{1}{16}[1, -4, 6, -4, 1]. \end{aligned}$$

Then  $\{\mathbf{a}_\ell : 0 \leq \ell \leq 4\}$  satisfy (2.11) and hence the system  $X(\Psi)$ , where  $\Psi = \{\psi_\ell : 1 \leq \ell \leq 4\}$  defined in (2.7) is a tight frame of  $L_2(\mathbb{R})$ . The filters  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ , and  $\mathbf{a}_4$  correspond to the first to fourth order difference operators respectively up to a scaling.

For practical concerns, we need to consider tight frames of  $L_2(\mathbb{R}^d)$  with  $d = 2$  or 3, since a typical image is a discrete function with its domain in 2 or 3 dimensional space. One way to construct tight frames for  $L_2(\mathbb{R}^d)$  is by taking *tensor products* of univariate tight frames. For simplicity of notation, we will consider the 2D case, i.e.  $d = 2$ . Arguments for  $d = 3$  or higher dimensions are similar.

Given a set of univariate masks  $\{\mathbf{a}_\ell : \ell = 0, 1, \dots, r\}$ , define the 2D masks  $\mathbf{a}_i[\mathbf{k}]$ , with  $\mathbf{i} := (i_1, i_2)$  and  $\mathbf{k} := (k_1, k_2)$ , as

$$(2.12) \quad \mathbf{a}_i[\mathbf{k}] := \mathbf{a}_{i_1}[k_1]\mathbf{a}_{i_2}[k_2], \quad 0 \leq i_1, i_2 \leq r; (k_1, k_2) \in \mathbb{Z}^2.$$

Then the corresponding 2D refinable function and framelets are defined by

$$\psi_{\mathbf{i}}(x, y) = \psi_{i_1}(x)\psi_{i_2}(y), \quad 0 \leq i_1, i_2 \leq r; (x, y) \in \mathbb{R}^2,$$

where we have let  $\psi_0 := \phi$  for convenience. We denote

$$\Psi_2 := \{\psi_{\mathbf{i}}; 0 \leq i_1, i_2 \leq r; \mathbf{i} \neq (0, 0)\}.$$

If the univariate masks  $\{\mathbf{a}_\ell\}$  are constructed from UEP, then it is easy to verify that  $\{\mathbf{a}_i\}$  satisfies (2.11) and thus  $X(\Psi_2)$  is a tight frame for  $L_2(\mathbb{R}^2)$ .

There are  $m$  framelets with vanishing moment from  $1, \dots, m$  constructed from the B-spline of order  $m$  by the UEP in [2]. Recall that the vanishing moment of a function is the order of the zero of its Fourier transform at the origin. We shall order the indices of the framelets according to their orders of vanishing moments. Note that the corresponding index for any B-spline is 0, since it has no vanishing moment. Then, for the tensor product framelet  $\psi_{\mathbf{i}}$  with  $\mathbf{i} = (i_1, i_2)$ , if there is a

differential operator associate with it, the differential operator should be  $D_{\mathbf{i}}$ , i.e. applying the  $i_1$  derivative at the first variable and the  $i_2$  derivative at the second variable.

We now provide the masks of 2-dimensional Haar and piecewise linear framelets constructed by tensor product in the following example.

**Example 2.4.**

- (1) The tensor-product 2-dimensional Haar tight frame system has filters

$$\begin{aligned} \mathbf{a}_{0,0} &= \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{a}_{0,1} = \frac{1}{4} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \\ \mathbf{a}_{1,0} &= \frac{1}{4} \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \mathbf{a}_{1,1} = \frac{1}{4} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \end{aligned}$$

- (2) The tensor-product 2-dimensional piecewise linear B-spline tight frame system has filters

$$\begin{aligned} \mathbf{a}_{0,0} &= \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}, \mathbf{a}_{0,1} = \frac{\sqrt{2}}{16} \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix}, \mathbf{a}_{0,2} = \frac{1}{16} \begin{pmatrix} -1 & 2 & -1 \\ -2 & 4 & -2 \\ -1 & 2 & -1 \end{pmatrix}, \\ \mathbf{a}_{1,0} &= \frac{\sqrt{2}}{16} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}, \mathbf{a}_{1,1} = \frac{1}{8} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \mathbf{a}_{1,2} = \frac{\sqrt{2}}{16} \begin{pmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & -2 & 1 \end{pmatrix}, \\ \mathbf{a}_{2,0} &= \frac{1}{16} \begin{pmatrix} -1 & -2 & -1 \\ 2 & 4 & 2 \\ -1 & -2 & -1 \end{pmatrix}, \mathbf{a}_{2,1} = \frac{\sqrt{2}}{16} \begin{pmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \\ -1 & 0 & 1 \end{pmatrix}, \mathbf{a}_{2,2} = \frac{1}{16} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}. \end{aligned}$$

In the discrete setting, let an image  $\mathbf{f}$  be a  $d$ -dimensional array. We denote by

$$\mathcal{I}_d := \mathbb{R}^{N_1 \times N_2 \times \dots \times N_d}$$

the set of all  $d$ -dimensional images. We will further assume that all images are square images, i.e.  $N_1 = N_2 = \dots = N_d = N$  and they all have supports in the open unit  $d$ -dimensional cube  $\Omega = (0, 1)^d$ . Note that these assumptions are not essential, and all arguments and results in this paper can be easily extended to general cases.

For simplicity, we will focus on  $d = 2$  throughout the rest of this paper. We denote the 2-dimensional fast (discrete) framelet transform (see, e.g., [34]) with levels of decomposition  $L$  as

$$(2.13) \quad \mathbf{W}\mathbf{u} = \{\mathbf{W}_{l,i}\mathbf{u} : 0 \leq l \leq L-1, 0 \leq i_1, i_2 \leq r\}, \quad \mathbf{u} \in \mathcal{I}_2.$$

The fast framelet transform  $\mathbf{W}$  is a linear operator with  $\mathbf{W}_{l,i}\mathbf{u} \in \mathcal{I}_2$  denoting the frame coefficients of  $\mathbf{u}$  at level  $l$  and band  $\mathbf{i}$ . Furthermore, we have

$$\mathbf{W}_{l,i}\mathbf{u} := \mathbf{a}_{l,i}[-\cdot] \circledast \mathbf{u},$$

where  $\circledast$  denotes the convolution operator with a certain boundary condition, e.g., periodic boundary condition, and  $\mathbf{a}_{l,i}$  is defined as

$$(2.14) \quad \mathbf{a}_{l,i} = \tilde{\mathbf{a}}_{l,i} \circledast \tilde{\mathbf{a}}_{l-1,0} \circledast \dots \circledast \tilde{\mathbf{a}}_{0,0} \quad \text{with} \quad \tilde{\mathbf{a}}_{l,i}[\mathbf{k}] = \begin{cases} \mathbf{a}_i[2^{-l}\mathbf{k}], & \mathbf{k} \in 2^l\mathbb{Z}^2; \\ 0, & \mathbf{k} \notin 2^l\mathbb{Z}^2. \end{cases}$$

Notice that  $\mathbf{a}_{0,i} = \mathbf{a}_i$ .

We denote the inverse framelet transform as  $\mathbf{W}^\top$ , which is the adjoint operator of  $\mathbf{W}$ , and we will have the perfect reconstruction formula

$$\mathbf{u} = \mathbf{W}^\top \mathbf{W}\mathbf{u}, \quad \text{for all } \mathbf{u} \in \mathcal{I}_2.$$

We note that we will also denote the fast framelet decomposition and reconstruction as  $\mathbf{W}_n$  and  $\mathbf{W}_n^\top$ , whenever the image resolution level  $n$  becomes relevant. We will also use  $\mathbf{H}$  and  $\mathbf{H}^\top$  to denote the decomposition and reconstruction of Haar framelets.

It is well known that large wavelet frame coefficients occur wherever there are singularities. The order of the singularity can be observed by the order of vanishing moments of the framelets whose corresponding coefficients are large. Recall that the order of the vanishing moments of a function is the order of zeros of its Fourier transform at the origin (see, e.g., [31, 44, 34]). When, for example, piecewise linear framelets are used, different framelet has a different order of vanishing moment. Therefore, the distributions of large wavelet frame coefficients indicates the distributions of the singularities and their orders of the underlying functions. Since the wavelet frame based approach essentially keeps large wavelet frame coefficients, it preserves singularities, i.e. features such as edges, of the underlying solutions.

*2.2.2. Wavelet Frame Based Image Restoration Models.* As briefly discussed in Section 1, the objective of image restoration is to find the unknown true image  $\mathbf{u} \in \mathcal{I}_2$  from an observed image (or measurements)  $\mathbf{f} \in \mathcal{I}_2$  defined by

$$(2.15) \quad \mathbf{f} = \mathbf{A}\mathbf{u} + \boldsymbol{\eta},$$

where  $\mathbf{A}$  is some linear operator and  $\boldsymbol{\eta}$  is the Gaussian noise. One can solve  $\mathbf{u}$  from (2.15) by considering the following least squares problem:

$$\min_{\mathbf{u}} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2.$$

This is, however, not a good idea in general. Taking the image deblurring problem as an example, since the matrix  $\mathbf{A}$  is ill-conditioned, the noise  $\boldsymbol{\eta}$  possessed by  $\mathbf{f}$  will be amplified after solving the above least squares problem.

Therefore, in order to suppress the effect of noise and also preserve key features of the image, e.g., edges, various regularization based optimization models were proposed in the literature. In this section, we recall some basics of the wavelet frame based approaches, which are all based on the fact that images, especially natural images, can be regarded as piecewise smooth functions. It is known that wavelet frames can usually provide good sparse approximations to piecewise smooth functions (thus the small wavelet frame coefficients for such functions can be ignored) due to their short supports and varied orders of vanishing moments. This motivates the research on wavelet frame based image restoration [35, 36, 37, 38, 39, 40, 41, 42, 43, 21].

Since wavelet frame systems are redundant systems, the mapping from the image  $\mathbf{u}$  to its coefficients is not one-to-one, i.e., the representation of  $\mathbf{u}$  in the wavelet frame domain is not unique. Therefore, there are mainly three formulations utilizing the sparseness of the wavelet frame coefficients, namely analysis based approach, synthesis based approach, and balanced approach. Although the analysis based, synthesis based and balanced approaches are developed independently in the literature, the balanced approach can be viewed as a way to balance the analysis and synthesis based approaches. Detailed and integrated descriptions of the three approaches can be found in [34].

The wavelet frame based image processing started from [36, 37] for high-resolution image reconstructions. By viewing the high-resolution image reconstruction as an inpainting problem in the wavelet frame domain, an iterative algorithm by applying

thresholding to wavelet frame coefficients at each iteration to preserve sharp edges of images was proposed in [36, 37]. The algorithm works in the image domain and frame domain alternatively. This simple algorithm converges as shown in [39] in the wavelet frame domain to a solution of

$$(2.16) \quad \inf_{\alpha} \frac{1}{2} \|\mathbf{A}\mathbf{W}^{\top}\alpha - \mathbf{f}\|_2^2 + \frac{1}{2} \|(\mathbf{I} - \mathbf{W}\mathbf{W}^{\top})\alpha\|_2^2 + \nu \|\boldsymbol{\lambda} \cdot \alpha\|_1.$$

The recovered image is derived by  $\mathbf{u} := \mathbf{W}^{\top}\alpha$ . The first term is a data fidelity term. The last term is a weighted  $\ell_1$ -norm of wavelet frame coefficients that reflects the sparsity of the frame coefficients  $\alpha$ . The second term penalizes the distance of the wavelet frame coefficient to the range of the analysis operator, and therefore, it forces  $\alpha$  to be close to the canonical coefficient. It is well known that the weighted  $\ell_1$ -norm of the canonical coefficient is related to some function norm of the corresponding function  $u$  (see, e.g., [45]). Thus, the second and third terms together balance the sparsity of the wavelet frame coefficient and the smoothness of the underlying image. The formulation (2.16) is applied to various applications in [38, 40, 35, 43]. In order to gain more flexibility, in [41, 42], we introduced a weighting before the second term in (2.16):

$$(2.17) \quad \inf_{\alpha} \frac{1}{2} \|\mathbf{A}\mathbf{W}^{\top}\alpha - \mathbf{f}\|_2^2 + \frac{\gamma}{2} \|(\mathbf{I} - \mathbf{W}\mathbf{W}^{\top})\alpha\|_2^2 + \nu \|\boldsymbol{\lambda} \cdot \alpha\|_1.$$

This formulation is referred to as the balanced approach since it balances the sparsity of the wavelet frame coefficient and the smoothness of the image. When  $\gamma = 0$ , the sparsity of the frame coefficient is penalized. This is called the synthesis based approach, as the image is synthesized by the sparsest coefficient; see [46, 47, 48, 49, 50]. When  $\gamma = +\infty$ , the smoothness of the image and the sparsity of canonical wavelet frame coefficients are penalized. It is called the analysis based approach, as the coefficient is in the range of the analysis operator; see [21, 51, 52]. The three approaches become the same approach when  $\mathbf{W}^{\top}\mathbf{W} = \mathbf{W}\mathbf{W}^{\top} = \mathbf{I}$ . In this paper, we will focus on the analysis based approach because it provides a direct link to the local geometry of  $\mathbf{u}$ . Since the coefficients being sought are in the range of the analysis operator, the analysis based approach [21] can be rewritten as

$$(2.18) \quad \inf_{\mathbf{u}} \nu \|\boldsymbol{\lambda} \cdot \mathbf{W}\mathbf{u}\|_{1,p} + \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2,$$

with  $p = 1$  or  $2$ . Here, we extend the  $\ell_1$ -norm in (2.17) to a generalized  $\ell_1$ -norm defined as

$$(2.19) \quad \|\boldsymbol{\lambda} \cdot \mathbf{W}\mathbf{u}\|_{1,p} := \left\| \sum_{l=0}^{L-1} \left( \sum_{i=(0,0)}^{(r,r)} \lambda_{l,i} |\mathbf{W}_{l,i}\mathbf{u}|^p \right)^{\frac{1}{p}} \right\|_1$$

where  $|\cdot|^p$  and  $(\cdot)^{\frac{1}{p}}$  are entrywise operations and  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm of  $\mathcal{I}_2$ . For  $l = 0$ , we simply denote  $\lambda_i := \lambda_{0,i}$  and  $\mathbf{W}_{0,i} := \mathbf{W}_i$ . We shall refer to the norm  $\|\cdot\|_{1,p}$  as the  $\ell_{1,p}$ -norm.

For  $p = 1$ ,  $\|\cdot\|_{1,1}$  is the usual  $\ell_1$ -norm used for frame and wavelet based image restoration problems [35, 36, 37, 38, 39, 40, 41, 42, 43, 21, 46, 47, 48, 49, 50]. For  $p = 2$ , the  $\ell_{1,2}$ -norm can be understood as an isotropic  $\ell_1$ -norm of the frame coefficients which uses an  $\ell_2$ -norm to combine different frame bands at a given location and decomposition level, while the  $\ell_{1,1}$ -norm can be understood as an anisotropic

$\ell_1$ -norm of the frame coefficients. By using the  $\ell_{1,2}$ -norm, the model becomes unbiased to horizontal and vertical directions. We further note that the isotropic and anisotropic  $\ell_1$ -norm of the wavelet frame coefficients resembles the definition of isotropic TV  $\int_{\Omega} \sqrt{|u_x|^2 + |u_y|^2} dx dy$  and the definition of the anisotropic TV  $\int_{\Omega} (|u_x| + |u_y|) dx dy$ . As a special case of the general results we shall establish in later sections, when  $\boldsymbol{\lambda}$  and  $\mathbf{W}$  are properly chosen, the analysis based approach (2.18) approximates the ROF model (2.1) with isotropic TV for  $p = 2$ , and anisotropic TV for  $p = 1$ .

Model (2.18) preserves large wavelet frame coefficients of the underlying solution. It will generate a good solution if the underlying solution has a good sparse representation under the wavelet frame transform  $\mathbf{W}$ ; namely, a good approximation of the underlying solution is achieved by only keeping the large wavelet frame coefficients. Since the large wavelet frame coefficients reflect the singularities of the underlying solution such as edges, (2.19) preserves edges of the underlying solution. The sparsity is also used in the total variation based method. That the total variation based approach (2.1) can well preserve edges, especially for piecewise constant images, is simply because piecewise constant functions can be sparsely approximated by the gradient transform.

When large coefficients of a particular framelet are kept, it means that the difference operator is applied to the underlying image at the location where singularity occurs. Since different wavelet frame masks reflect different orders of difference operators, the wavelet frame based approach applies difference operators adaptively according to the singularities of the underlying solutions. Hence, it can well preserve various types of edges simultaneously. This indicates why wavelet frame based approaches can outperform (1.2) for some applications when the order of singularities of images varies in different regions. When spline wavelet frames are used, each wavelet frame mask can be viewed as a standard difference operator that can be understood as a discretization of the corresponding differential operator. Hence, if using wavelet frame based approaches to approximate (1.2), one implicitly applies differential operators adaptive to the singularities. In fact, the theory developed here can be modified to apply to images that are piecewise smooth. Furthermore, wavelet frame based approaches can automatically adapt to pieces with different regularities easily in numerical implications.

We further note that for the analysis based approach (2.18), the inverse wavelet frame transform  $\mathbf{W}^{\top}$  is not used. Therefore, instead of using the wavelet frame transform, one can replace it by any other linear transforms, e.g., a general frame transform (not necessary a tight wavelet frame, or not even a wavelet frame). In particular, when the split Bregman algorithm is applied to (2.1), it actually solves (2.18) with  $\mathbf{W}$  being some transform derived from whatever discretization scheme one may choose. Since  $\mathbf{W}$  can be extended to an onto mapping without changing (2.18) by properly adjusting  $\boldsymbol{\lambda}$ , the split Bregman based method solving (2.1) is essentially a wavelet frame based method with the corresponding wavelet frame depending on the choice of discretization.

As we will see in later sections, a special case of the analysis based approach of the wavelet frame method (2.18), with only one level of decomposition, is used to link the analysis based approach to various differential operator based approach (1.2). Furthermore, it can also be viewed as one of the efficient ways to solve (1.2).

It should be noted that the wavelet frame based approaches solve image restoration problems in digital domain directly, and they give a wide variety of models, (e.g., the synthesis based, analysis based and balanced approaches, and approaches with multiple frame systems), as well as the associated efficient algorithms which utilize the multilevel nature of wavelet frame transforms to achieve better sparse approximation of the underlying solutions. The interested readers should consult, for example, the survey articles [33, 34] for further details.

**2.3. Motivation by an Example.** As stated, we are aiming at establishing the connections between the wavelet frame based approach and differential operator based variational approaches. The connection is done by using the analysis based approach (2.18) to approximate variational models when a proper  $\lambda$  is chosen. In this section, we use the ROF denoising model, i.e.,

$$(2.20) \quad \min_u E(u) := \nu \int_{\Omega} |\nabla u| dx dy + \frac{1}{2} \|u - f\|_{L_2(\Omega)}^2$$

as an example. Note that “min” is used here because the minimal value of  $E$  defined above is attainable since  $E$  is coercive [3].

A complete analysis of more general cases is given in Section 3. The emphasis here is to use a simple example to motivate us and ready ourselves for more complicated ones. Some of the technical details may be left out and a complete analysis will be given in Sections 3 and 4.

For simplicity, we start with a  $u \in L_2(\Omega)$  that is sufficiently smooth. Let  $\mathbf{u}_1$  denote the restriction of  $u$  on the discrete mesh with meshsize  $h = 1/(N - 1)$ , i.e.,

$$(\mathbf{u}_1)[i, j] = u(x_i, y_j), \quad \text{with } (x_i, y_j) = (ih, jh) \text{ and } 0 \leq i, j \leq N - 1.$$

For the data fidelity term, it is straightforward to use  $\frac{1}{2} \|\mathbf{u}_1 - \mathbf{f}_1\|_2^2$ . Here the  $\ell_2$ -norm, as well as other norms involved, are scaled by taking the meshsize  $h$  into account. In particular, we define  $\|\mathbf{x}\|_p^p := h^2 \sum_{i,j=0}^{N-1} |\mathbf{x}[i, j]|^p$ . It can be verified that

$$(2.21) \quad \frac{1}{2} \|\mathbf{u}_1 - \mathbf{f}_1\|_2^2 \rightarrow \frac{1}{2} \|u - f\|_{L_2(\Omega)}^2, \quad \text{as } h \rightarrow 0.$$

For the regularization term, our motivation is that the filters of Haar framelets are discrete difference operators, as shown in Example 2.4. We will illustrate our observations as follows. Since  $u$  is smooth enough,  $\int_{\Omega} |\nabla u| dx dy = \int_{\Omega} \sqrt{u_x^2 + u_y^2} dx dy$ . Then by Taylor’s expansion on  $u$  at  $(x_i, y_i)$ , we have

$$\begin{aligned} \frac{2}{h} (\mathbf{a}_{0,1}[-\cdot] \otimes \mathbf{u}_1) [i, j] &= \frac{1}{2h} (u(x_i, y_j) - u(x_i - h, y_j)) + \frac{1}{2h} (u(x_i, y_j - h) - u(x_i - h, y_j - h)) \\ &\rightarrow u_x(x_i, y_j), \quad \text{as } h \rightarrow 0. \end{aligned}$$

Similarly, we can easily calculate that

$$\begin{aligned} \frac{2}{h} (\mathbf{a}_{1,0}[-\cdot] \otimes \mathbf{u}_1) [i, j] &= \frac{1}{2h} (u(x_i, y_j) - u(x_i, y_j - h)) + \frac{1}{2h} (u(x_i - h, y_j) - u(x_i - h, y_j - h)) \\ &\rightarrow u_y(x_i, y_j), \quad \text{as } h \rightarrow 0. \end{aligned}$$

Therefore, we have that

(2.22)

$$\nu h^2 \sum_{i,j} \left( \left( \frac{2}{h} \right)^2 \left( |(\mathbf{a}_{1,0}[-\cdot] \otimes \mathbf{u}_\perp)[i,j]|^2 + |(\mathbf{a}_{0,1}[-\cdot] \otimes \mathbf{u}_\perp)[i,j]|^2 \right) \right)^{1/2} \rightarrow \nu \int_{\Omega} \sqrt{u_x^2 + u_y^2} dx dy,$$

as  $h \rightarrow 0$ . If we choose

$$(2.23) \quad \boldsymbol{\lambda} = \begin{pmatrix} \lambda_{0,0} & \lambda_{0,1} \\ \lambda_{1,0} & \lambda_{1,1} \end{pmatrix} = \frac{4\nu^2}{h^2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then the left hand side of (2.22) is exactly  $\|\boldsymbol{\lambda} \cdot \mathbf{H}\mathbf{u}_\perp\|_{1,2}$ , i.e.,

$$(2.24) \quad \|\boldsymbol{\lambda} \cdot \mathbf{H}\mathbf{u}_\perp\|_{1,2} \rightarrow \nu \int_{\Omega} \sqrt{u_x^2 + u_y^2} dx dy, \quad \text{as } h \rightarrow 0.$$

Combining (2.21) and (2.24) together, we get that, by choosing  $\boldsymbol{\lambda}$  as in (2.23),

$$\|\boldsymbol{\lambda} \cdot \mathbf{H}\mathbf{u}_\perp\|_{1,2} + \frac{1}{2} \|\mathbf{u}_\perp - \mathbf{f}\|_2^2 \rightarrow E(u), \quad \text{as } h \rightarrow 0,$$

i.e., the objective functional of the analysis based approach converges to that of the ROF denoising model.

The discretization of  $u$  by its restriction  $\mathbf{u}_\perp$  requires a strong assumption on the smoothness of  $u$ , and it is used only for motivational purposes. When a wavelet frame is used, the discretization of  $u$  is naturally given by the corresponding refinable function, and then the smoothness requirement for  $u$  is reduced. Precisely speaking, we assume that  $u \in W_1^1(\Omega)$  and we use

$$(2.25) \quad \mathbf{T}_n u = \{2^n \langle u, \phi_{n,\mathbf{k}} \rangle : \mathbf{k} = (k_1, k_2), 0 \leq k_1, k_2 \leq N-1\}$$

to discretize  $u$  and  $f$ . Here we assume that  $N = 2^n + 1$  and  $h = 2^{-n}$ , for some  $n \geq 0$ . Then, we use

$$(2.26) \quad E_n(u) := \|\boldsymbol{\lambda} \cdot \mathbf{H}\mathbf{T}_n u\|_{1,2} + \frac{1}{2} \|\mathbf{T}_n u - \mathbf{T}_n f\|_2^2,$$

where  $\boldsymbol{\lambda}$  is given by (2.23), to approximate the ROF denoising model  $E$ . Since  $\mathbf{T}_n u$  is some sampling of  $u$ , it is expected that

$$(2.27) \quad E_n(u) \rightarrow E(u), \quad \text{as } n \rightarrow \infty.$$

Indeed, our main result in the next section confirms it. By applying Theorem 3.1 in Section 3.2, we have the following corollary.

**Corollary 2.1.** *Assume that  $u \in W_1^1(\Omega)$ , and  $\boldsymbol{\lambda}$  is in (2.23). Then (2.27) holds.*

Furthermore,  $E_n$   $\Gamma$ -converges to  $E$ . This further indicates the relations between approximate minimizers, more precisely, the  $\epsilon$ -optimal solutions, of (2.26) and those of the ROF model (2.20), as given in the following corollary of Theorem 3.2. (The definition of  $\epsilon$ -optimal solution is given in (3.7).)

**Corollary 2.2.** *Let  $E$  and  $E_n$  be energy functionals defined on the space  $W_1^1(\Omega)$ , and let  $\boldsymbol{\lambda}$  be given by (2.23). Let  $u_n^*$  be an  $\epsilon$ -optimal solution to  $E_n$ . Then we have*

$$\limsup_{n \rightarrow \infty} E_n(u_n^*) \leq \inf_u E(u) + \epsilon.$$

*Furthermore, any cluster point of  $\{u_n^*\}$  is an  $\epsilon$ -optimal solution of  $E$ .*

In Section 3, we will show that the above approaches work in a more general setting; i.e., they work not only for the ROF model, but also for a wide variety of regularizations; not only for the denoising problem, but also for the inpainting and deblurring problems; not only for Haar framelets, but also for general B-spline framelets that lead to a richer family of differential operator based variational methods; not only a specific choice of  $\lambda$ , but also a family of choices of  $\lambda$ .

### 3. Connections Between Models

This section is devoted to establishing a connection between the wavelet frame based image restoration approach, especially the analysis based approach, and variational methods. A complete analysis will be given. The first part of this section provides some preliminaries for our analysis; the second part gives the analysis of the approximation by the objective functional of the wavelet frame based approach to that of variational methods; finally, the relations among approximate minimizers of different approaches are investigated. Some technical details and proofs are left to Section 4.

**3.1. Notation and Basics.** In this subsection, we will introduce some basic notation and concepts that will be used in the rest of this paper. We will also demonstrate how to discretize functions defined on a square domain  $\Omega \subset \mathbb{R}^2$ .

For convenience to be referred later, we introduce some symbols and notation that will be used throughout the rest of the paper.

**Notation 3.1.** We focus our analysis on  $\mathbb{R}^2$ , i.e., the 2-dimensional cases. All the 2-dimensional refinable functions and framelets are constructed by tensor products of univariate B-splines and the associated framelets obtained from the UEP (2.11).

- (1) We assume all functions we consider are defined on the open unit square  $\Omega := (0,1)^2 \subset \mathbb{R}^2$ , and that their discrete versions, i.e. digital images, are defined on an  $N \times N$  Cartesian grid on  $\bar{\Omega}$  with  $N = 2^n + 1$  for  $n \geq 0$ . We denote by  $h = 2^{-n}$  the meshsize of the  $N \times N$  grid.
- (2) We use bold face letters  $\mathbf{i}$ ,  $\mathbf{j}$ , and  $\mathbf{k}$  to denote double indices in  $\mathbb{Z}^2$ . We denote by  $\mathbb{O}^2 := \{0, 1, \dots, N-1\}^2$  as the set of indices of the  $N \times N$  Cartesian grid.
- (3) For 2-dimensional cases,  $\phi_{n,\mathbf{k}}$  (also  $\tilde{\phi}_{n,\mathbf{k}}$ ,  $\varphi_{n,\mathbf{k}}$ ,  $\tilde{\varphi}_{n,\mathbf{k}}$ ,  $\psi_{n,\mathbf{k}}$ , etc.) takes the form  $\phi_{n,\mathbf{k}} = 2^n \phi(2^n \cdot -\mathbf{k})$ . Given a wavelet frame system and its corresponding refinable function  $\phi$ , we denote  $\mathbb{M}^2$  the set of indices  $\mathbf{k} \in \mathbb{O}^2$  such that the support of  $\phi_{n,\mathbf{k}}$ , denoted as  $\Lambda_{n,\mathbf{k}}$ , is completely supported in  $\bar{\Omega}$ . Then obviously,  $\mathbb{M}^2 \subset \mathbb{O}^2$ . When needed, we will also use  $\Lambda_{n,\mathbf{k}}$  to denote the support of  $\varphi_{n,\mathbf{k}}$ , and use  $\tilde{\Lambda}_{n,\mathbf{k}}$  to denote the support of  $\tilde{\phi}_{n,\mathbf{k}}$  or  $\tilde{\varphi}_{n,\mathbf{k}}$ .
- (4) In order to properly handle boundary conditions, we further restrict the domain of the wavelet frame transform  $\mathbf{W}$  to the set of discrete sequences defined on the grid  $\mathbb{M}^2 \subset \mathbb{O}^2$ . We denote such a set of sequences as  $\mathbb{R}^{M^2}$  with  $M^2$  the cardinality of  $\mathbb{M}^2$ .
- (5) For simplicity, we assume that the level of wavelet frame decomposition is 1, i.e.  $L = 1$ , while our analysis can be easily extended to the general cases with  $L > 1$ . If  $L = 1$ , we have

$$\mathbf{W}\mathbf{u} = \{\mathbf{W}_i\mathbf{u} : 0 \leq i_1, i_2 \leq r\}, \quad \mathbf{W}_i\mathbf{u} := \mathbf{a}_i[-\cdot] \otimes \mathbf{u}, \quad \text{with } \mathbf{u} \in \mathbb{R}^{M^2}.$$

- (6) We denote  $\mathbb{K}^2 \subset \mathbb{M}^2$  as the index set when the boundary condition of  $\mathbf{a}_i[-\cdot] \circledast \mathbf{u}$  is inactive for all  $\mathbf{i}$ , or in other words,  $\mathbf{a}_i[-\cdot] * \mathbf{u}$  is well defined for all  $\mathbf{i}$ , where “ $*$ ” is the standard discrete convolution operator. Let  $K^2$  be the cardinality of  $\mathbb{K}^2$ . Then  $\mathbf{W}_i : \mathbb{R}^{M^2} \mapsto \mathbb{R}^{K^2}$  for each  $(0,0) \leq \mathbf{i} \leq (r,r)$ . Note that the index sets  $\mathbb{O}^2$ ,  $\mathbb{M}^2$  and  $\mathbb{K}^2$  all depend on the image resolution  $n$ .
- (7) In order to link the continuous and discrete settings, we need to take resolution into account. Therefore, for any array  $\mathbf{v} \in \mathbb{R}^{M^2}$ , the discrete  $\ell_p$ -norm we are using is defined as

$$(3.1) \quad \|\mathbf{v}\|_p^p := \sum_{\mathbf{i} \in \mathbb{M}^2} |\mathbf{v}[\mathbf{i}]|^p h^2.$$

Then, we have

$$\|\boldsymbol{\lambda} \cdot \mathbf{W}\mathbf{u}\|_{1,p} := \left\| \left( \sum_{\mathbf{i}=(0,0)}^{(r,r)} \lambda_{\mathbf{i}} |\mathbf{W}_{\mathbf{i}}\mathbf{u}|^p \right)^{\frac{1}{p}} \right\|_1 = h^2 \sum_{\mathbf{k} \in \mathbb{K}^2} \left( \sum_{\mathbf{i}=(0,0)}^{(r,r)} \lambda_{\mathbf{i}}[\mathbf{k}] \left| (\mathbf{W}_{\mathbf{i}}\mathbf{u})[\mathbf{k}] \right|^p \right)^{\frac{1}{p}}.$$

We use  $W_p^r(\Omega)$  to denote the Sobolev space in which the  $r$ -th weak derivative is in  $L_p(\Omega)$ , and equipped with norm  $\|f\|_{W_p^r(\Omega)} := \sum_{|\mathbf{k}| \leq r} \|D_{\mathbf{k}}f\|_p$ , where  $D_{\mathbf{k}}$  is the standard multi-index notation of differential operators. Here  $v \in L_p(\Omega)$  is said to be a weak derivative of  $u \in L_p(\Omega)$  if, for all functions  $w \in C_c^\infty(\Omega)$ , we have

$$\langle v, w \rangle = (-1)^{|\mathbf{k}|} \langle u, D_{\mathbf{k}}w \rangle,$$

where

$$\langle v, w \rangle := \int_{\Omega} v w d\mathbf{x}.$$

For functions  $f, g \in L_2(\mathbb{R}^2)$ , we denote their inner product as

$$\langle f, g \rangle_{L_2(\mathbb{R}^2)} := \int_{\mathbb{R}^2} f g d\mathbf{x}.$$

For any given function  $u \in L_2(\Omega)$ , we discretize it in the finite dimensional space

$$(3.2) \quad \mathcal{V}_n = \text{span}\{\phi_{n,\mathbf{k}} : \mathbf{k} \in \mathbb{M}^2\} \subset L_2(\Omega).$$

In particular, the discrete version of  $u \in L_2(\Omega)$  is defined by

$$(3.3) \quad \mathbf{T}_n u = \{2^n \langle u, \phi_{n,\mathbf{k}} \rangle : \mathbf{k} \in \mathbb{M}^2\} \in \mathbb{R}^{M^2}.$$

Throughout the rest of this paper, we assume that  $\phi \in L_2(\mathbb{R}^2)$  is compactly supported in  $\Omega$ . Then for  $u \in L_2(\Omega)$ , the inner product  $\langle u, \phi_{n,\mathbf{k}} \rangle = \int_{\Omega} u \phi_{n,\mathbf{k}} d\mathbf{x}$  is well-defined. Note that the factor  $2^n$  in the definition of  $\mathbf{T}_n$  in (3.3) is needed to cancel out the factor  $h^2$  in the definition of the  $\ell_p$ -norm (3.1), so that we can have  $\|\mathbf{T}_n u\|_2^2 = \sum_{\mathbf{k} \in \mathbb{M}^2} |\langle u, \phi_{n,\mathbf{k}} \rangle|^2$ , which is consistent with the notation used in the literature of wavelets.

With the notation above, we can show the following lemma. The proof of  $\{\phi_{n,\mathbf{k}} : \mathbf{k} \in \mathbb{M}^2\}$  being a Riesz basis for  $\mathcal{V}_n$  is straightforward and can be found in, e.g., [53]. We shall only present the proof of (2) which follows from Lemma 4.1.

**Lemma 3.1.** *Let  $\phi \in L_2(\Omega)$  be constructed by tensor product from a univariate B-spline function and the corresponding operator  $\mathbf{T}_n$  be defined by (3.3). Then  $\{\phi_{n,\mathbf{k}} : \mathbf{k} \in \mathbb{M}^2\}$  forms a Riesz basis for  $\mathcal{V}_n$  with Riesz bounds independent of  $n$ . In particular, there exists  $C$  independent of  $n$  such that:*

(1) For all  $u \in L_2(\Omega)$ , we have

$$\|\mathbf{T}_n u\|_2^2 \leq C \|u\|_{L_2(\Omega)}^2.$$

In addition:

(2) For all  $u \in L_2(\Omega)$ , we have

$$\lim_{n \rightarrow \infty} \|\mathbf{T}_n u\|_2^2 = \|u\|_{L_2(\Omega)}^2.$$

*Proof.* We prove part (2). Notice that

$$\|\mathbf{T}_n u\|_2^2 = \left\langle \sum_{\mathbf{k} \in \mathbb{M}^2} \langle u, \phi_{n,\mathbf{k}} \rangle \phi_{n,\mathbf{k}}, u \right\rangle.$$

By Lemma 4.1 with  $p = 2$ , we have

$$\lim_{n \rightarrow \infty} \left\| \sum_{\mathbf{k} \in \mathbb{M}^2} \langle u, \phi_{n,\mathbf{k}} \rangle \phi_{n,\mathbf{k}} - u \right\|_{L_2(\Omega)} = 0.$$

Then we have

$$\lim_{n \rightarrow \infty} \|\mathbf{T}_n u\|_2^2 = \left\langle \lim_{n \rightarrow \infty} \sum_{\mathbf{k} \in \mathbb{M}^2} \langle u, \phi_{n,\mathbf{k}} \rangle \phi_{n,\mathbf{k}}, u \right\rangle = \|u\|_{L_2(\Omega)}^2.$$

□

**3.2. Connections of Two Models via Approximation.** The major task of this section is to establish the link of the analysis based approach (2.18) to various differential operator based variational methods. Precisely speaking, we will show that when one chooses various  $\boldsymbol{\lambda}$ , the analysis based approach (2.18) can be understood as certain discretizations of various types of differential operator based variational methods, e.g., the variational approaches reviewed in the introduction. In general, we shall focus on the following variational approach:

$$(3.4) \quad \inf_{u \in W_1^s(\Omega)} \nu \|Du\|_{1,p} + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2,$$

where  $D = (D_i)$ ,  $s = \max_i s_i \geq 1$  with  $s_i = |\mathbf{i}|$ , and

$$\|Du\|_{1,p} := \left\| \left( \sum_i |D_i u|^p \right)^{\frac{1}{p}} \right\|_{L_1(\Omega)}.$$

Here the differential operator  $D_i$  with  $\mathbf{i} = (i_1, i_2)$  is defined conventionally as  $D_i(u(x, y)) = \frac{\partial^{|\mathbf{i}|} u}{\partial x^{i_1} \partial y^{i_2}}$ . To make the second term of (3.4) meaningful, we assume that  $f \in L_2(\Omega)$  and  $A : L_2(\Omega) \mapsto L_2(\Omega)$  is a bounded linear operator. Note from the Sobolev imbedding theorem (see, e.g., [54, Theorem 4.12]), we have  $W_1^s(\Omega) \subset L_2(\Omega)$ . Therefore, the second term of (3.4) is well-defined.

**3.2.1. Reformulation.** We can write the analysis based approach at level  $n$  according to (2.18) as

$$(3.5) \quad \inf_{\mathbf{u}_n \in \mathbb{R}^{M^2}} \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{u}_n\|_{1,p} + \frac{1}{2} \|\mathbf{A}_n \mathbf{u}_n - \mathbf{f}_n\|_2^2.$$

Here we added the subscript  $n$  to emphasize the dependence of each variable and operator on  $n$ . Note that  $\mathbf{W}_n$  is defined in the exact same way as in (5) and (6) of Notation 3.1, and the index  $n$  merely indicates that the domain and range of  $\mathbf{W}_n$  depend on  $n$ . Equation (3.5) is an optimization problem with respect to the

discrete array  $\mathbf{u}_n \in \mathbb{R}^{M^2}$ . As we have mentioned before, the discrete array  $\mathbf{u}_n$  is understood as a discretization of some function  $u \in L_2(\Omega)$  through operator  $\mathbf{T}_n$ . Then, by replacing  $\mathbf{u}_n$  in (3.5) by  $\mathbf{T}_n u$ , we can obtain a minimization at level  $n$  with respect to function  $u$ . In addition, since we will prove connections between the frame based approach (3.5) with the variational method (3.4), it is reasonable to assume that the minimization is conducted for all  $u$  in the Sobolev space  $W_1^s(\Omega)$ . Therefore, we shall consider the following optimization problem:

$$(3.6) \quad \inf_{u \in W_1^s(\Omega)} \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} + \frac{1}{2} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2^2.$$

In the following proposition we will show that the optimization problem (3.5) is equivalent to (3.6).

For notational convenience, we denote the objective functional in (3.5) as

$$F_n(\mathbf{u}) := \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{u}_n\|_{1,p} + \frac{1}{2} \|\mathbf{A}_n \mathbf{u}_n - \mathbf{f}_n\|_2^2.$$

We denote the objective functional in (3.6) as

$$E_n(u) := \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} + \frac{1}{2} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2^2.$$

In numerical computations, the task is to find an approximate minimizer, i.e., the one on which the value of the corresponding objective functional is close to its infimum. We say that  $u^*$  is an  $\epsilon$ -optimal solution to a given objective functional  $E$  if

$$(3.7) \quad E(u^*) \leq \inf_u E(u) + \epsilon, \quad \text{for some } \epsilon > 0.$$

We say that  $u^*$  is a minimizer of  $E$  if  $E(u^*) = \inf_u E(u)$ . It is clear that  $\epsilon$ -optimal solutions of  $E$ ,  $E_n$ , and  $F_n$  always exist because each of them has an infimum. In this paper, whenever we say that  $u^*$  is an approximate minimizer of  $E$  we mean that  $u^*$  is an  $\epsilon$ -optimal solution to  $E$  with some sufficiently small  $\epsilon$ . In other words,  $E(u^*)$  is very close to  $\inf E$ .

**Proposition 3.1.** *Let the refinable function  $\phi \in L_2(\Omega)$  be constructed by tensor product from a univariate B-spline function, and the corresponding operator  $\mathbf{T}_n$  be defined by (2.25). Then  $F_n$  and  $E_n$  have the same infimum. Furthermore, for any given minimizer (or  $\epsilon$ -optimal solution) of  $\mathbf{u}_n^*$  of  $F_n$ , we can construct a minimizer  $u_n^*$  (or  $\epsilon$ -optimal solution) of  $E_n$  such that  $\mathbf{T}_n u_n^* = \mathbf{u}_n^*$ . Conversely, for any given minimizer  $u_n^*$  (or  $\epsilon$ -optimal solution) of  $E_n$ ,  $\mathbf{T}_n u_n^*$  is a minimizer (or  $\epsilon$ -optimal solution) of  $F_n$ . In other words, (3.5) and (3.6) are equivalent for image restoration.*

*Proof.* We first note that for any given univariate B-spline function, we can construct a compactly supported dual refinable function with any prescribed regularity such that the shifts of the B-spline and its dual form a biorthogonal system (see, e.g., [55]). This is still true in 2D for a  $\phi$  that is constructed by tensor product from a univariate B-spline function. In particular, for any given  $s \in \mathbb{Z}^+$ , there exists a compactly supported refinable function  $\tilde{\phi} \in W_1^s(\mathbb{R}^2)$ , such that for any  $n \in \mathbb{Z}$  and  $\mathbf{i}, \mathbf{j} \in \mathbb{Z}^2$ ,

$$(3.8) \quad \langle \phi_{n,\mathbf{i}}, \tilde{\phi}_{n,\mathbf{j}} \rangle_{L_2(\mathbb{R}^2)} = \begin{cases} 1 & \text{when } \mathbf{i} = \mathbf{j}; \\ 0 & \text{otherwise.} \end{cases}$$

Let  $\mathbf{u}_n \in \mathbb{R}^{M^2}$  be any given point. We construct a function  $u_n \in W_1^s(\Omega)$  as  $u_n := 2^{-n} \sum_{\mathbf{k} \in \mathbb{M}^2} \mathbf{u}_n[\mathbf{k}] \tilde{\phi}_{n,\mathbf{k}} \chi_\Omega$ . By the duality of  $\phi$  and  $\tilde{\phi}$  given by (3.8), it is easy to see that  $\mathbf{T}_n u_n = \mathbf{u}_n$  and  $E_n(u_n) = F_n(\mathbf{u}_n)$ . Thus, the infimum of  $E_n$  is less than that of  $F_n$ . On the other hand, the infimum of  $F_n$  is less than that of  $E_n$ . Indeed, let  $\mathbf{u}_n^*$  be any given point. We can simply take  $\mathbf{u}_n = \mathbf{T}_n \mathbf{u}_n^*$ , which is well-defined since  $u_n \in W_1^s(\Omega) \subset L_2(\Omega)$ . This concludes that the infimum of  $F_n$  is the same as that of  $E_n$ . The rest of the proposition follows directly from the above arguments.  $\square$

*Remark 3.1.* For a given approximate minimizer  $\mathbf{u}_n^*$  of (3.5), the approximate minimizer  $u_n^*$  of (3.6) is constructed from the dual refinable function  $\tilde{\phi}$ . In general  $u_n^*$  does not belong to  $\mathcal{V}_n$ . However, when  $\mathcal{V}_n \subset W_1^s(\Omega)$ , then we can find the dual basis  $\tilde{\phi}_{n,\mathbf{k}}$  as a linear combination of  $\{\phi_{n,\mathbf{k}} : \mathbf{k} \in \mathbb{M}^2\}$  such that  $\tilde{\phi}_{n,\mathbf{k}} \in \mathcal{V}_n$ , and hence,  $u_n^* \in \mathcal{V}_n$ . The condition  $\mathcal{V}_n \subset W_1^s(\Omega)$  will be satisfied whenever  $\phi$  is a B-spline of order  $s+1$  or higher. For image processing, in order to preserve sharpness of features, we usually prefer the order of smoothness of  $\tilde{\phi}$  to be as small as possible in order not to smear out the edges, although to carry out the analysis here, we need  $\tilde{\phi} \in W_1^s(\Omega)$ . In Section 3.3, we will focus on the relation between  $u_n^*$  and an approximate minimizer  $u^*$  of (3.4).

Since Proposition 3.1 clarifies the relations between approximate minimizers of (3.5) and (3.6), throughout the rest of this paper, we will focus on the connections of (3.6) with (3.4). Rewrite  $E_n$  as

$$E_n(u) = E_n^{(1)}(u) + E_n^{(2)}(u),$$

where

$$(3.9) \quad E_n^{(1)}(u) := \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} \quad \text{and} \quad E_n^{(2)}(u) := \frac{1}{2} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2^2.$$

Denote the objective functional of (3.4) as

$$E(u) := \nu \|\mathbf{D}u\|_{1,p} + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2.$$

We will also split  $E$  into

$$E(u) = E^{(1)}(u) + E^{(2)}(u),$$

where

$$E^{(1)}(u) = \nu \|\mathbf{D}u\|_{1,p} \quad \text{and} \quad E^{(2)}(u) = \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2.$$

For convenience, we shall take  $\nu = 1$  and then  $E^{(1)}(u) = \|\mathbf{D}u\|_{1,p}$ .

**3.2.2. Assumptions.** We shall impose some assumptions under which we can prove that  $E_n$  converges to  $E$ . We will show that all assumptions given here are satisfied in our choice of wavelet frame and discretization scheme. Hence, they are properties satisfied by our choice of wavelet frame and discretization scheme, rather than imposed assumptions. However, when other wavelet frames and discretization schemes are chosen, the proof given here still works, as long as these assumptions are satisfied.

The first condition is consistency between the bounded operator  $A$  and its discretization  $\mathbf{A}_n$ , i.e., the bounded linear operator  $A$  should be consistent with its discretization  $\mathbf{A}_n$ . More precisely, we assume that  $A$  and  $\mathbf{A}_n$  satisfy the following assumption.

(A1)  $A : L_2(\Omega) \rightarrow L_2(\Omega)$  is a continuous linear operator and its discretization  $\mathbf{A}_n$  satisfies

$$(3.10) \quad \lim_{n \rightarrow \infty} \|\mathbf{T}_n A u - \mathbf{A}_n \mathbf{T}_n u\|_2 = 0, \quad \forall u \in L_2(\Omega).$$

The above assumption says that the discretization  $\mathbf{T}_n$  and the linear operator  $A$  should commute asymptotically. It is clear that when the operator  $A$  is the identity (e.g., for the image denoising), the obviously choice of  $\mathbf{A}_n$  is also the identity. In this case, the consistent assumption condition (A1) is satisfied automatically. In Section 4.2, we will demonstrate how to properly discretize  $A$  such that Assumption (A1) is indeed satisfied for image inpainting and image deblurring.

We also need to impose further assumptions on the tight wavelet frame systems and the parameter  $\lambda_n$ . We shall focus our analysis on the case with  $L = 1$ , i.e. only one level of decomposition, for simplicity. However, the proofs for multilevel decomposition are similar. When  $L = 1$ , we can rewrite  $E_n^{(1)}(u)$  according to (7) of Notation 3.1 as

$$E_n^{(1)}(u) = \|\lambda_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} = \left\| \left( \sum_{\mathbf{i}=(0,0)}^{(r,r)} \lambda_{\mathbf{i}} |\mathbf{W}_{\mathbf{i}} u|^p \right)^{\frac{1}{p}} \right\|_1.$$

Let

$$\mathbb{B} = \{\mathbf{i} : 0 \leq i_1, i_2 \leq r\} \setminus \{(0,0)\}$$

denote all the B-spline wavelet frame bands constructed from the tensor product of the univariate B-spline of order  $r$  (see [2]). Without loss of generality, we assume that the set of univariate B-spline framelets  $\{\psi_i : 0 \leq i \leq r\}$  is arranged according to the ascend order of vanishing moment of  $\psi_i$ , i.e. the order of vanishing moment of  $\psi_i = i$  (see Examples 2.1-2.3). Note that for each framelet  $\psi_i$  that is constructed by the tensor product of univariate B-spline framelets, the only differential operator  $D_{\mathbf{i}}$  that can be associated with it should have the same tensor product structure. On the other hand, for a given vector of differential operator  $\mathbf{D}$  with order  $s$ , we choose a B-spline wavelet frame system with the  $r$  given in (5) of Notation 3.1 satisfying  $2r \geq s$ . Such a choice always exists since we can construct a tight wavelet frame system for a B-spline of any order using the UEP [2]. Denote the index sets  $\mathbb{I}$  and  $\mathbb{J}$  as

$$\mathbb{I} := \{\mathbf{i} : D_{\mathbf{i}} \text{ is in } \mathbf{D}\} \quad \text{and} \quad \mathbb{J} := \mathbb{B} \setminus \mathbb{I}.$$

Under these settings, we will consider  $u \in W_1^s(\Omega)$  with  $s = \max_{\mathbf{i} \in \mathbb{I}} s_{\mathbf{i}}$  and  $s_{\mathbf{i}} = |\mathbf{i}|$ . The set  $\mathbb{I}$  corresponds to the active wavelet frame bands, on which we expect that  $E_n^{(1)}(u) \rightarrow E^{(1)}(u)$  for each  $u$ . The set  $\mathbb{J}$  corresponds to the inactive wavelet frame bands, since for  $\mathbf{j} \in \mathbb{J}$ ,  $D_{\mathbf{j}}$  does not appear in  $\mathbf{D}$ . Thus, we need the effect of the terms in  $E_n^{(1)}$  that correspond to  $\mathbb{J}$  to go to 0 as  $n \rightarrow \infty$ . Consequently, proper choices for  $\lambda$  at different bands are needed.

Now, we list the crucial assumptions of the set of framelets  $\Psi$  and the parameters  $\lambda$  as follows:

(A2) For each  $\mathbf{i} \in \mathbb{B}$ , there exists an  $s_{\mathbf{i}}$ -differentiable function  $\varphi_{\mathbf{i}}$  with  $c_{\mathbf{i}} = \int_{\mathbb{R}^2} \varphi_{\mathbf{i}} d\mathbf{x} \neq 0$  such that  $D_{\mathbf{i}} \varphi_{\mathbf{i}} = \psi_{\mathbf{i}}$  and  $\varphi_{\mathbf{i}}$  has the same support as  $\psi_{\mathbf{i}}$ .

- (A3) We choose  $\lambda_{0,0} = 0$ . For every  $\mathbf{i} \in \mathbb{I}$ , we set  $\lambda_{\mathbf{i}} = \left(\frac{1}{c_{\mathbf{i}}} 2^{s_{\mathbf{i}}(n-1)}\right)^p$ , where  $c_{\mathbf{i}}$  is given by (A2). For every  $\mathbf{j} \in \mathbb{J}$ , we set  $0 \leq \lambda_{\mathbf{j}} \leq O(2^{s_{\mathbf{j}}p(n-1)})$  for some  $\mathbf{j}' \in \mathbb{B} \cup \{\mathbf{0}\}$  such that  $\mathbf{0} \leq \mathbf{j}' < \mathbf{j}$  and  $s_{\mathbf{j}'} \leq s$ .

*Remark 3.2.*

- (1) The sets  $\mathbb{I}$  and  $\mathbb{J}$  form a partition of all the wavelet frame bands  $\mathbb{B}$ . For the indices  $\mathbf{j} \in \mathbb{J}$ , one can choose  $\lambda_{\mathbf{j}}$  to be either zero or a nonzero value bounded by  $O(2^{ps_{\mathbf{j}}(n-1)})$ . For both cases, one can show that (see the proof of Lemma 3.3) the effect of  $\lambda_{\mathbf{j}}|\mathbf{W}_{\mathbf{j}}\mathbf{u}|^p$  will vanish when  $n \rightarrow \infty$ . This means that for both cases,  $E_n^{(1)}(u)$  will be approximating the same  $E^{(1)}(u)$ , which only depends on the choice of  $\lambda_{\mathbf{i}}$  for  $\mathbf{i} \in \mathbb{I}$ . However, for different choices of  $\lambda_{\mathbf{j}}$ ,  $\mathbf{j} \in \mathbb{J}$ , the term  $E_n^{(1)}(u)$  corresponds to a different type of discretization of the same  $E^{(1)}(u)$ . Choosing a nonzero  $\lambda_{\mathbf{j}}$  is to obtain a better discretization for various purposes. Taking Haar framelets as an example, if we use the following choice of  $\boldsymbol{\lambda}$ ,

$$(3.11) \quad \boldsymbol{\lambda} = \frac{4\nu^2}{h^2} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

then in addition to the two terms we had in (2.22), we also have the term  $\mathbf{a}_{1,1}[-\cdot] \otimes \mathbf{u}$ . Then using (3.11) we will have the following discretization of  $|\nabla u|^2$ :

$$(3.12) \quad |\nabla u|^2 \approx \frac{1}{4} \left[ \left( \frac{u[x_i, y_j] - u[x_i - h, y_j]}{h} \right)^2 + \left( \frac{u[x_i, y_j - h] - u[x_i - h, y_j - h]}{h} \right)^2 \right. \\ \left. + \left( \frac{u[x_i, y_j] - u[x_i, y_j - h]}{h} \right)^2 + \left( \frac{u[x_i - h, y_j] - u[x_i - h, y_j - h]}{h} \right)^2 \right] \\ + \frac{1}{2} \left[ \left( \frac{u[x_i, y_j] - u[x_i - h, y_j - h]}{\sqrt{2}h} \right)^2 + \left( \frac{u[x_i - h, y_j] - u[x_i, y_j - h]}{\sqrt{2}h} \right)^2 \right],$$

which is better than the corresponding discretization using  $\boldsymbol{\lambda}$  as in (2.23) because it takes the rotation invariance of  $|\nabla u|$  into account (the last two terms of (3.12) utilize the rotation invariance of  $|\nabla u|$  by 45 degrees).

- (2) Note that the Assumption (A2) is indeed satisfied by our choice of wavelet frame systems. Since we use the tensor-product B-spline framelets of [2], we only need to consider the univariate case for a given set of framelets. When the univariate framelet system generated by certain orders of the B-spline is given, to obtain functions  $\varphi$ , one can simply apply the anti-derivative  $s$  times to  $\psi$ , where  $s$  is the order of vanishing moments of  $\psi$ . It can be checked easily that  $\varphi$  has the same support as  $\psi$  and  $\int_{\mathbb{R}^2} \varphi d\mathbf{x} \neq 0$ . In Section 4.3, we will provide details on the quantities defined above using Haar and piecewise linear framelets as examples. The existence of  $\varphi_{\mathbf{i}}$  is guaranteed by the work [56]. In fact, a uniform form of such a function  $\varphi_{\mathbf{i}}$  for an arbitrary B-spline framelet is given by [56].
- (3) When multi-level framelet decomposition is used, i.e.  $L > 1$ , we can follow the same proof as the case  $L = 1$  to obtain an approximation of the differential operators in the bands  $\mathbf{i} \in \mathbb{I}$  and the effects of the bands  $\mathbf{j} \in \mathbb{J}$  vanishing as  $n \rightarrow \infty$  for multiple levels. However, the wavelet frame transform by

multiple levels may get a better approximation to differential operators. Since this paper is to emphasize on the connection between two approaches of (3.4) and (3.5) for image restorations instead of using approach (3.5) to find high order approximation to the variational method (3.4), we will only focus on the case  $L = 1$ .

**3.2.3. Pointwise Convergence.** We show that the objective functional  $E_n$  converges pointwise to  $E$  in this section. To be precise, we prove the following theorem which is one of the two main results of this paper.

**Theorem 3.1** (Pointwise Convergence). *Assume that Assumptions (A1)–(A3) are satisfied. Then, for any  $u \in W_1^s(\Omega)$ , we have*

$$(3.13) \quad \lim_{n \rightarrow \infty} E_n(u) = E(u).$$

Notice that here we imposed the smoothness of  $u$ , i.e.,  $u \in W_1^s(\Omega)$ , where  $s$  is the maximum order of the differential operators involved in  $E^{(1)}(u)$ . This smoothness assumption is quite natural, as approximate minimizers of  $E$  have to be in  $W_1^s(\Omega)$  due to the term  $\|\mathbf{D}u\|_{1,p}$ . To prove the theorem, we need Lemmas 3.2 and 3.3 where we will show that  $\lim_{n \rightarrow \infty} E_n^{(1)}(u) = E^{(1)}(u)$  and  $\lim_{n \rightarrow \infty} E_n^{(2)}(u) = E^{(2)}(u)$  respectively. Then Theorem 3.1 is obtained straightforwardly.

We now prove Lemma 3.2 under Assumption (A1) as follows.

**Lemma 3.2.** *Assume that Assumption (A1) is satisfied. Then for every  $u \in W_1^s(\Omega)$ , we have  $\lim_{n \rightarrow +\infty} E_n^{(2)}(u) = E^{(2)}(u)$ , i.e.,*

$$\lim_{n \rightarrow \infty} \frac{1}{2} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2^2 = \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2.$$

*Proof.* Note from the Sobolev Imbedding theorem, we have  $u \in L_2(\Omega)$ . Then

$$\begin{aligned} \|\mathbf{T}_n(Au - f)\|_2 - \|(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)u\|_2 &\leq \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2 \\ &\leq \|\mathbf{T}_n(Au - f)\|_2 + \|(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)u\|_2. \end{aligned}$$

By letting  $n \rightarrow +\infty$  and assumption (A1), we get

$$(3.14) \quad \lim_{n \rightarrow \infty} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2^2 = \lim_{n \rightarrow \infty} \|\mathbf{T}_n(Au - f)\|_2^2.$$

By item (2) of Lemma 3.1, we have

$$\lim_{n \rightarrow \infty} \|\mathbf{T}_n(Au - f)\|_2^2 = \|Au - f\|_{L_2(\Omega)}^2.$$

Therefore,  $\lim_{n \rightarrow \infty} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2^2 = \|Au - f\|_{L_2(\Omega)}^2$ .  $\square$

Lemma 3.3, which states that  $\lim_{n \rightarrow +\infty} E_n^{(1)}(u) = E^{(1)}(u)$ , will be true under Assumptions (A2) and (A3). The proof of it is left to Section 4.4.

**Lemma 3.3.** *Assume that Assumptions (A2) and (A3) are satisfied. Then, for any  $u \in W_1^s(\Omega)$ , we have*

$$(3.15) \quad \lim_{n \rightarrow +\infty} \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} = \|\mathbf{D}u\|_{1,p}.$$

**3.3. Connections of Two Models via  $\Gamma$ -Convergence.** The analysis in the previous section, especially the pointwise convergence of  $E_n$  to  $E$ , indicates that the analysis based approach (3.5) with various choices of parameters  $\lambda$  and wavelet frame systems can be used to approximate to and be regarded as different types of discretizations of various variational methods. Since numerical solutions of our image restoration problems are approximate minimizers to the objective functionals  $E_n$  and  $E$ , it is more important to discover the relations of  $E_n$  and  $E$  as optimization problems. For example, we want to know whether the infimum value of  $E$  can be approximated and bounded by the infimum value of  $E_n$  from below. However, such relations need more than pointwise convergence of  $E_n$  to  $E$ , and we need the  $\Gamma$ -convergence. In fact, we shall prove a convergence result that is stronger than the  $\Gamma$ -convergence.

To state the second main theorem of this paper, we first present a proposition indicating that for each fixed  $u$ ,  $E_n$  is continuous uniformly in  $n$ . The proof is postponed to Section 4.5.

**Proposition 3.2.** *Suppose that Assumptions (A1)–(A3) are satisfied. Given an arbitrary  $u \in W_1^s(\Omega)$ , for any  $\epsilon > 0$ , there exist an integer  $\mathcal{N}$  and  $\delta > 0$  independent of  $n$  such that for all  $v \in W_1^s(\Omega)$  satisfying  $\|u - v\|_{W_1^s(\Omega)} < \delta$  and  $n > \mathcal{N}$ , we have  $|E_n(v) - E_n(u)| < \epsilon$ .*

Now, we present the second main result of this paper. The first main result of this paper is Theorem 3.1. For this, we give the definition of  $\Gamma$ -convergence.

**Definition 3.1.** Given  $E_n(u) : W_1^s(\Omega) \mapsto \bar{\mathbb{R}}$  and  $E(u) : W_1^s(\Omega) \mapsto \bar{\mathbb{R}}$ , we say that  $E_n$   $\Gamma$ -converges to  $E$  if:

- (i) for every sequence  $u_n \rightarrow u$  in  $W_1^s(\Omega)$ ,  $E(u) \leq \liminf_{n \rightarrow \infty} E_n(u_n)$ ;
- (ii) for every  $u \in W_1^s(\Omega)$ , there is a sequence  $u_n \rightarrow u$  in  $W_1^s(\Omega)$ , such that  $E(u) \geq \limsup_{n \rightarrow \infty} E_n(u_n)$ .

**Theorem 3.2.** *Suppose that Assumptions (A1)–(A3) are satisfied. Then, for every sequence  $u_n \rightarrow u$  in  $W_1^s(\Omega)$ , we have  $\lim_{n \rightarrow +\infty} E_n(u_n) = E(u)$ . Consequently,  $E_n$   $\Gamma$ -converges to  $E$  in  $W_1^s(\Omega)$ .*

*Proof.* By Theorem 3.1 and Proposition 3.2, we have that, for an arbitrary given  $u \in W_1^s(\Omega)$ , and  $\epsilon > 0$ ,

- (a)  $\lim_{n \rightarrow +\infty} |E_n(u) - E(u)| = 0$ ;
- (b) there exist an integer  $\mathcal{N}$  and  $\delta > 0$  satisfying  $|E_n(v) - E_n(u)| < \epsilon$  whenever  $\|u - v\|_{W_1^s(\Omega)} < \delta$  and  $n > \mathcal{N}$ .

Note that for arbitrary  $u_n \in W_1^s(\Omega)$ , we have

$$|E_n(u_n) - E(u)| \leq |E_n(u) - E(u)| + |E_n(u) - E_n(u_n)|.$$

Let the sequence  $u_n \rightarrow u$  in  $W_1^s(\Omega)$ , and let  $\epsilon > 0$  be a given arbitrary number. On the one hand, by (a), there exists an  $\mathcal{N}_1$  such that  $|E_n(u) - E(u)| < \epsilon/2$  whenever  $n > \mathcal{N}_1$ . On the other hand, by (b), there exist  $\mathcal{N}$  and  $\delta$  such that  $|E_n(v) - E_n(u)| < \epsilon/2$  whenever  $\|u - v\|_{W_1^s(\Omega)} < \delta$  and  $n > \mathcal{N}$ ; since  $u_n \rightarrow u$ , there exists  $\mathcal{N}_2$  such that  $\|u - u_n\|_{W_1^s(\Omega)} < \delta$  whenever  $n > \mathcal{N}_2$ ; letting  $v = u_n$  leads to  $|E_n(u_n) - E_n(u)| < \epsilon/2$  whenever  $n > \max\{\mathcal{N}, \mathcal{N}_2\}$ . Combining all together, we have

$$|E_n(u_n) - E(u)| \leq \epsilon$$

whenever  $n > \max\{\mathcal{N}, \mathcal{N}_1, \mathcal{N}_2\}$ . This shows that  $\lim_{n \rightarrow +\infty} E_n(u_n) = E(u)$ , and hence both conditions given in Definition 3.1 are satisfied. Therefore,  $E_n$  is  $\Gamma$ -convergent to  $E$ .  $\square$

In both the variational and wavelet frame based approaches (i.e. (1.2) and (1.5)) for image restorations, the key is to find a sparse approximate solution of the linear inverse problem in some transform domain, for example the gradient domain for the TV model and the wavelet frame domain for wavelet frame based models, since all models (explicitly or implicitly) assume that the underlying solutions have a good sparse approximate solution in the transform domains. The main idea to compute an accurate sparse numerical solution is to threshold each numerical approximate solution in the transform domain to make it sparse, and then update the residues and iterate the process. Most successful algorithms follow such a procedure. At the same time, in most of those algorithms, the value of the objective functional converges to its infimum as the iteration proceeds. The small value of the objective functional at the approximate solutions reflects that these solutions are closed to the sparse solution of the original equation to some extent depending on the transform and the blurring operator. The main task for both the variational and wavelet frame based approaches is to find sparse approximate solutions in transform domains which are usually those approximate solutions with small values of the underlying objective functionals. The minimizers for either variational or wavelet frame based models are not the focus here, although they may or may not exist. Even when they do exist, they are commonly not unique, and it is usually difficult and, in fact, unnecessary to find them numerically. Since the objective functionals for both the variational and wavelet frame based models are convex, there is the infimum for each given model. Therefore, instead of minimizers, we commonly seek approximate minimizers, more precisely,  $\epsilon$ -optimal solutions. Recall that  $u^*$  is an  $\epsilon$ -optimal solution to  $E$  if

$$E(u^*) \leq \inf_u E(u) + \epsilon, \quad \text{for some } \epsilon > 0.$$

Since  $E_n$  pointwise converges to  $E$  by Theorem 3.1, it is natural to use  $E_n$  approximating  $E$ . Since we are interested in  $\epsilon$ -optimal solutions to  $E$ , it is natural to ask whether it is bounded below in some way by  $\epsilon$ -optimal solutions to  $E_n$  or even can be approximated by them for some cases. These all follow from the  $\Gamma$ -convergence of  $E_n$  to  $E$  as shown in the following lemma:

**Corollary 3.1.** *Suppose that Assumptions (A1)–(A3) are satisfied. Let  $u_n^*$  be an  $\epsilon$ -optimal solution of  $E_n$  for a given  $\epsilon > 0$  and for all  $n$ .*

(1) *We have*

$$(3.16) \quad \limsup_{n \rightarrow \infty} E_n(u_n^*) \leq \inf_u E(u) + \epsilon.$$

*In particular, when  $u_n^*$  is a minimizer of  $E_n$ , we have*

$$\limsup_{n \rightarrow \infty} E_n(u_n^*) \leq \inf_u E(u).$$

(2) *If, in addition, the set  $\{u_n^*\}$  has a cluster point  $u^*$ , then  $u^*$  is an  $\epsilon$ -optimal solution to  $E$ . In particular, when  $u_n^*$  is a minimizer of  $E_n$  and  $u^*$  a cluster point of the set  $\{u_n^*\}$ , then*

$$E(u^*) = \limsup_{n \rightarrow \infty} (E_n(u_n^*)) = \inf_u E(u)$$

*and  $u^*$  is a minimizer of  $E$ .*

*Proof.* Part (1): For any  $u \in W_1^s(\Omega)$ , let  $\{u_n\}$  be the sequence as given in item (ii) of the definition of  $\Gamma$ -convergence. Then, we have

$$E(u) \geq \limsup_{n \rightarrow \infty} E_n(u_n) \geq \limsup_{n \rightarrow \infty} \left( \inf_u E_n(u) \right) \geq \limsup_{n \rightarrow \infty} E_n(u_n^*) - \epsilon,$$

which implies (3.16).

Part (2): If  $u^*$  is a cluster point of  $\{u_n^*\}$ , let  $\{u_{n_k}^*\}$  be a subsequence of  $\{u_n^*\}$  such that  $u_{n_k}^* \rightarrow u^*$  as  $k \rightarrow \infty$ . Then by item (i) of the definition of  $\Gamma$ -convergence, we have

$$E(u^*) \leq \liminf_{k \rightarrow \infty} E_{n_k}(u_{n_k}^*) \leq \limsup_{k \rightarrow \infty} E_{n_k}(u_{n_k}^*) \leq \limsup_{n \rightarrow \infty} E_n(u_n^*) \leq \inf_u E(u) + \epsilon,$$

where the last inequality follows from (3.16). This shows that  $u^*$  is an  $\epsilon$ -optimal solution to  $E$ .  $\square$

While it is tricky to construct, or compute numerically, a sequence of  $\epsilon$ -optimal solutions, so that it has at least one cluster point, it should be pointed out that the emphasis here is to establish relations between (1.2) and (1.5) (through the connections between (3.4) and (3.6)) at a conceptual level. Such a connection not only provides a better understanding of the variational model (1.2) and the wavelet frame based approach (1.5), but also brings opportunities for new applications. It gives geometric explanations of the wavelet frame based approach by viewing it as a discrete form of the variational model (1.2). It also provides the legitimacy of solving (1.2) via solving the corresponding discrete optimization problem (3.6). Although the wavelet frame based approach (1.5) can be used to approximate the variational model (1.2), finding a good approximation of (1.2) through solving (1.5) is not a focus here. For image restorations, the wavelet frame based model has been proven to be a very efficient method, and for many cases it is even better than variational models. Hence, it is sufficient to adopt the wide range of wavelet frame based approaches in many applications. Furthermore, when a solution in function form is needed, one can obtain it easily as suggested in Remark 3.1. The choice of the dual function  $\tilde{\phi}$  depends on the balance between the preservation of edges and approximation orders.

#### 4. Technical Details

In this section, we present the technical details that are temporarily avoided in Section 3. More precisely, we will first discuss the discretization of the operator  $A$  and the differential operator  $\mathbf{D}$ . Then the proofs of Lemma 3.3 and Proposition 3.2 will be provided. We now start this section with a preliminary lemma from approximation theory.

**4.1. A Lemma in Approximation Theory.** We present a lemma in approximation theory in this section. The lemma will be used later in the discretization and the proofs.

**Lemma 4.1.** *Let  $\varphi$  and  $\tilde{\varphi}$  be two compactly supported bounded functions satisfying  $\int_{\mathbb{R}^2} \varphi d\mathbf{x} = 1$  and  $\int_{\mathbb{R}^2} \tilde{\varphi} d\mathbf{x} = 1$ . In addition, we assume that  $\varphi$  satisfies the partition of unity, i.e.  $\sum_{\mathbf{j} \in \mathbb{Z}^2} \varphi(\cdot + \mathbf{j}) = 1$ . Let  $\mathbb{M}^2$  (resp.  $\tilde{\mathbb{M}}^2$ ) be the index set for  $\mathbf{k}$  such that  $\varphi_{n,\mathbf{k}}$  (resp.  $\tilde{\varphi}_{n,\mathbf{k}}$ ) is completely supported in  $\bar{\Omega}$ . In addition, we assume that*

the support of  $\varphi$  is contained within the support of  $\tilde{\varphi}$ . Then, for any  $u \in L_p(\Omega)$ , for  $1 \leq p \leq \infty$ , we have

$$(4.1) \quad \lim_{n \rightarrow \infty} \left\| u - \sum_{\mathbf{k} \in \mathbb{M}^2} \langle u, \tilde{\varphi}_{n,\mathbf{k}} \rangle \varphi_{n,\mathbf{k}} \right\|_{L_p(\Omega)} = 0.$$

*Proof.* We first establish (4.1) for all  $u \in C_c^1(\Omega)$ , the space of compactly supported and continuously differentiable functions. We note that since  $\varphi$  and  $\tilde{\varphi}$  are bounded and have compact supports, hence they both belong to  $L_p(\mathbb{R}^2)$  for  $1 \leq p \leq \infty$ .

Let  $u_e$  be an extension of  $u \in L_p(\Omega)$  to  $L_p(\mathbb{R}^2)$  such that  $u_e = u$  on  $\Omega$  and  $u_e = 0$  elsewhere. Since we have  $(\tilde{\varphi}\tilde{\varphi})(0) = 1$  and  $\varphi$  satisfies the partition of unity, then it was shown by [57, Theorem 3.2] that, for any  $v \in L_p(\mathbb{R}^2)$ ,

$$(4.2) \quad \lim_{n \rightarrow \infty} \left\| v - \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} \right\|_{L_p(\mathbb{R}^2)} = 0.$$

Denote by  $\tilde{\Lambda}_{n,\mathbf{k}}$  the support of  $\tilde{\varphi}_{n,\mathbf{k}}$  and by  $\mathfrak{L}(\cdot)$  the Lebesgue measure. Note that  $\tilde{\Lambda}_{n,\mathbf{k}} = 2^{-n}(\tilde{\Lambda}_{0,\mathbf{0}} + \mathbf{k})$ . Define the index set  $I_n \subset \mathbb{Z}^2$  as

$$I_n := \{\mathbf{k} \in \mathbb{Z}^2 : \mathfrak{L}(\tilde{\Lambda}_{n,\mathbf{k}} \cap \Omega) > 0\} \setminus \tilde{\mathbb{M}}^2.$$

Similarly, define the index set  $\tilde{I}_n \subset \mathbb{Z}^2$  as

$$\tilde{I}_n := \{\mathbf{k} \in \mathbb{Z}^2 : \mathfrak{L}(\tilde{\Lambda}_{n,\mathbf{k}} \cap \Omega) > 0\} \setminus \mathbb{M}^2.$$

By the assumption that the support of  $\varphi$  is contained in the support of  $\tilde{\varphi}$ , we have  $\tilde{\mathbb{M}}^2 \subset \mathbb{M}^2$ , and hence  $\tilde{I}_n \subset I_n$ . Then, noting that  $\langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} = 0$  for  $\mathbf{k} \in \mathbb{Z}^2 \setminus (\mathbb{M}^2 \cup \tilde{I}_n)$ , we have

$$\begin{aligned} & \left\| u_e - \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} \right\|_{L_p(\mathbb{R}^2)} \\ &= \left\| u_e - \left( \sum_{\mathbf{k} \in \mathbb{M}^2} \langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} + \sum_{\mathbf{k} \in \tilde{I}_n} \langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} \right) \right\|_{L_p(\mathbb{R}^2)} \\ &\geq \left\| u_e - \sum_{\mathbf{k} \in \mathbb{M}^2} \langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} \right\|_{L_p(\mathbb{R}^2)} - \left\| \sum_{\mathbf{k} \in \tilde{I}_n} \langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} \right\|_{L_p(\mathbb{R}^2)} \\ &= \left\| u - \sum_{\mathbf{k} \in \mathbb{M}^2} \langle u, \tilde{\varphi}_{n,\mathbf{k}} \rangle \varphi_{n,\mathbf{k}} \right\|_{L_p(\Omega)} - \left\| \sum_{\mathbf{k} \in \tilde{I}_n} \langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} \right\|_{L_p(\mathbb{R}^2)}. \end{aligned}$$

Therefore, substituting  $v = u_e$  in (4.2), it suffices to show that

$$(4.3) \quad \left\| \sum_{\mathbf{k} \in \tilde{I}_n} \langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} \right\|_{L_p(\mathbb{R}^2)} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Let  $\mathfrak{S}_n = \cup_{\mathbf{k} \in I_n} \tilde{\Lambda}_{n,\mathbf{k}}$ . Since  $|I_n| \leq C2^n$  where  $C$  depends only on  $\Lambda_{0,\mathbf{0}}$ , we have  $\mathfrak{L}(\mathfrak{S}_n) \leq C12^{-n}$ . Since  $u_e$  is bounded and  $\tilde{I}_n \subset I_n$ , we have

$$\begin{aligned} \left\| \sum_{\mathbf{k} \in \tilde{I}_n} \langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n,\mathbf{k}} \right\|_{L_p(\mathbb{R}^2)} &\leq \left\| \sum_{\mathbf{k} \in \tilde{I}_n} |\langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)}| |\varphi_{n,\mathbf{k}}| \right\|_{L_p(\mathbb{R}^2)} \\ &\leq \left\| \sum_{\mathbf{k} \in I_n} |\langle u_e, \tilde{\varphi}_{n,\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)}| |\varphi_{n,\mathbf{k}}| \right\|_{L_p(\mathbb{R}^2)} \\ &\leq \|u\|_{L_\infty(\mathbb{R}^2)} \left\| \sum_{\mathbf{k} \in I_n} \|\tilde{\varphi}_{n,\mathbf{k}}\|_{L_1(\mathbb{R}^2)} |\varphi_{n,\mathbf{k}}| \right\|_{L_p(\mathbb{R}^2)}. \end{aligned}$$

Notice that

$$\|\tilde{\varphi}_{n,\mathbf{k}}\|_{L_1(\mathbb{R}^2)} = \int_{\mathbb{R}^2} 2^n |\tilde{\varphi}(2^n \mathbf{x} - \mathbf{k})| d\mathbf{x} = 2^{-n} \|\tilde{\varphi}\|_{L_1(\mathbb{R}^2)}.$$

Thus

$$\begin{aligned} \|u\|_{L_\infty(\mathbb{R}^2)} \left\| \sum_{\mathbf{k} \in I_n} \|\tilde{\varphi}_{n,\mathbf{k}}\|_{L_1(\mathbb{R}^2)} |\varphi_{n,\mathbf{k}}| \right\|_{L_p(\mathbb{R}^2)} &= 2^{-n} \|u\|_{L_\infty(\mathbb{R}^2)} \|\tilde{\varphi}\|_{L_1(\mathbb{R}^2)} \left\| \sum_{\mathbf{k} \in I_n} |\varphi_{n,\mathbf{k}}| \right\|_{L_p(\mathbb{R}^2)} \\ &= 2^{-n} \|u\|_{L_\infty(\mathbb{R}^2)} \|\tilde{\varphi}\|_{L_1(\mathbb{R}^2)} \left\| \sum_{\mathbf{k} \in I_n} |\varphi_{n,\mathbf{k}}| \chi_{\tilde{\Lambda}_{n,\mathbf{k}}} \right\|_{L_p(\mathbb{R}^2)} \\ &\leq 2^{-n} \|u\|_{L_\infty(\mathbb{R}^2)} \|\tilde{\varphi}\|_{L_1(\mathbb{R}^2)} \left\| 2^n \|\varphi\|_{L_\infty(\mathbb{R}^2)} \sum_{\mathbf{k} \in I_n} \chi_{\tilde{\Lambda}_{n,\mathbf{k}}} \right\|_{L_p(\mathbb{R}^2)} \\ &= \|u\|_{L_\infty(\mathbb{R}^2)} \|\varphi\|_{L_\infty(\mathbb{R}^2)} \|\tilde{\varphi}\|_{L_1(\mathbb{R}^2)} \left\| \sum_{\mathbf{k} \in I_n} \chi_{\tilde{\Lambda}_{n,\mathbf{k}}} \right\|_{L_p(\mathbb{R}^2)}. \end{aligned}$$

Observe that  $\sum_{\mathbf{k} \in I_n} \chi_{\tilde{\Lambda}_{n,\mathbf{k}}} \leq C\chi_{\mathfrak{S}_n}$  with  $C$  only depending on  $\tilde{\Lambda}_{0,\mathbf{0}}$ . We then have

$$\left\| \sum_{\mathbf{k} \in I_n} \chi_{\tilde{\Lambda}_{n,\mathbf{k}}} \right\|_{L_p(\mathbb{R}^2)} \leq C \|\chi_{\mathfrak{S}_n}\|_{L_p(\mathbb{R}^2)} \rightarrow 0,$$

as  $n \rightarrow \infty$ . We have now established (4.3) and hence (4.1) follows for all  $u \in C_c^1(\Omega)$ .

We now extend the result to the entire  $L_p(\Omega)$ . Note from [58, Theorems 2.1 and 3.1] that the linear operator  $\sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v, \tilde{\varphi}_{0,\mathbf{k}} \rangle \varphi_{0,\mathbf{k}}$  is bounded on  $L_p(\mathbb{R}^2)$  since  $\varphi$  and

$\tilde{\varphi}$  are compactly supported and bounded. For  $v \in L_p(\mathbb{R}^2)$ , we have that

$$\begin{aligned} \left\| \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v, \tilde{\varphi}_{n, \mathbf{k}} \rangle \varphi_{n, \mathbf{k}} \right\|_{L_p(\mathbb{R}^2)} &= \left\| \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v, 2^n \tilde{\varphi}(2^n \cdot -\mathbf{k}) \rangle 2^n \varphi(2^n \cdot -\mathbf{k}) \right\|_{L_p(\mathbb{R}^2)} \\ &= \left\| \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v(2^{-n} \cdot), \tilde{\varphi}_{0, \mathbf{k}} \rangle \varphi(2^n \cdot -\mathbf{k}) \right\|_{L_p(\mathbb{R}^2)} \\ &= 2^{-2n/p} \left\| \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v(2^{-n} \cdot), \tilde{\varphi}_{0, \mathbf{k}} \rangle \varphi_{0, \mathbf{k}} \right\|_{L_p(\mathbb{R}^2)}, \end{aligned}$$

and  $\|v(2^{-n} \cdot)\|_{L_p(\mathbb{R}^2)} = 2^{2n/p} \|v\|_{L_p(\mathbb{R}^2)}$ . Therefore, for all  $v \in L_p(\mathbb{R}^2)$ , we have

$$\frac{\left\| \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v, \tilde{\varphi}_{n, \mathbf{k}} \rangle \varphi_{n, \mathbf{k}} \right\|_{L_p(\mathbb{R}^2)}}{\|v\|_{L_p(\mathbb{R}^2)}} = \frac{\left\| \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v(2^{-n} \cdot), \tilde{\varphi}_{0, \mathbf{k}} \rangle \varphi_{0, \mathbf{k}} \right\|_{L_p(\mathbb{R}^2)}}{\|v(2^{-n} \cdot)\|_{L_p(\mathbb{R}^2)}} \leq C.$$

This shows that  $\sum_{\mathbf{k} \in \mathbb{Z}^2} \langle v, \tilde{\varphi}_{n, \mathbf{k}} \rangle \varphi_{n, \mathbf{k}}$  is bounded on  $L_p(\mathbb{R}^2)$  with the bound independent of  $n$ .

Now, for  $v \in L_p(\mathbb{R}^2)$ , we have

$$\begin{aligned} \left\| \sum_{\mathbf{k} \in \mathbb{M}^2} \langle v, \tilde{\varphi}_{n, \mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \varphi_{n, \mathbf{k}} \right\|_{L_p(\mathbb{R}^2)} &\leq \left\| \sum_{\mathbf{k} \in \mathbb{M}^2} \langle |v|, |\tilde{\varphi}_{n, \mathbf{k}}| \rangle_{L_2(\mathbb{R}^2)} |\varphi_{n, \mathbf{k}}| \right\|_{L_p(\mathbb{R}^2)} \\ &\leq \left\| \sum_{\mathbf{k} \in \mathbb{Z}^2} \langle |v|, |\tilde{\varphi}_{n, \mathbf{k}}| \rangle_{L_2(\mathbb{R}^2)} |\varphi_{n, \mathbf{k}}| \right\|_{L_p(\mathbb{R}^2)} \\ &\leq C \|v\|_{L_p(\mathbb{R}^2)}. \end{aligned}$$

Then by letting  $v = u_e$ , we can see that the operator  $\sum_{\mathbf{k} \in \mathbb{M}^2} \langle u, \tilde{\varphi}_{n, \mathbf{k}} \rangle \varphi_{n, \mathbf{k}}$  is bounded on  $L_p(\Omega)$  with bound independent of  $n$ . Since  $C_c^1(\Omega)$  is dense in  $L_p(\Omega)$ , the rest of the proof follows from the standard density arguments.  $\square$

*Remark 4.1.* We forgo optimizing the assumptions in this lemma, since the present form of this lemma is sufficient for proving our main results. For example, the assumption that the support of  $\tilde{\varphi}$  contains that of  $\varphi$  can be easily removed. Although we have used the relation  $\tilde{I}_n \subset I_n$  in the proof, without such an assumption, the cardinality of  $\tilde{I}_n$  is still of order  $2^n$  with the underlying constant depending only on the supports of  $\tilde{\varphi}$  and  $\varphi$ .

**4.2. Discretization of Linear Operator  $A$ .** We assume that  $A$  is a bounded linear operator that maps  $L_2(\Omega)$  into  $L_2(\Omega)$ . In this subsection, we demonstrate how to discretize  $A$  to get a matrix  $\mathbf{A}_n$ . As stated in Section 3.2.2, we require that  $A$  and  $\mathbf{A}_n$  satisfy Assumption (A1); i.e., the discretization  $\mathbf{A}_n$  should be consistent with  $A$  asymptotically. Obviously, not all discretizations of  $A$  would satisfy the assumption (A1). In this subsection, we demonstrate how to properly discretize  $A$  such that Assumption (A1) is indeed satisfied. We shall focus on image denoising, image inpainting, and image deblurring problems.

**Image denoising.** In this case,  $A$  is the identity operator and we choose  $\mathbf{A}_n$  to be the identity matrix. Then Assumption (A1) is obviously true.

**Image inpainting.** In this case,  $Au = \chi_\Lambda u$ , where  $\Lambda \subset \Omega$  is the closed domain on which the image is known. We assume that  $\Lambda$  contains finitely many disconnected components whose boundaries are piecewise  $C^2$  with finite lengths. We define  $\mathbf{A}_n$  to be the  $M^2 \times M^2$  diagonal matrix with diagonal entries

$$\mathbf{A}_n[\mathbf{k}, \mathbf{k}] = \begin{cases} 1 & \text{if } \mathfrak{L}(\Lambda_{n,\mathbf{k}} \cap \Lambda) > 0, \\ 0 & \text{otherwise,} \end{cases} \quad \forall \mathbf{k} \in \mathbb{M}^2,$$

where  $\Lambda_{n,\mathbf{k}}$  denotes the support of  $\phi_{n,\mathbf{k}}$ . We now verify the consistency condition (3.10).

Define

$$\mathfrak{S}_n = \cup_{\{\mathbf{k}: \mathbf{A}_n[\mathbf{k}, \mathbf{k}] = 1, \mathbf{k} \in \mathbb{M}^2\}} \Lambda_{n,\mathbf{k}} \quad \text{and} \quad I_n := \{\mathbf{k} : \mathbf{A}_n[\mathbf{k}, \mathbf{k}] = 1\} \setminus \{\mathbf{k} : \Lambda_{n,\mathbf{k}} \subset \Lambda\}.$$

Then, for any given  $u \in L_2(\Omega)$ ,

$$\begin{aligned} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n A u\|_2^2 &= \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n \chi_\Lambda u\|_2^2 = \sum_{\mathbf{k} \in I_n} |\langle u - \chi_\Lambda u, \phi_{n,\mathbf{k}} \rangle|^2 \\ &= \sum_{\mathbf{k} \in I_n} |\langle (\chi_{\mathfrak{S}_n} - \chi_\Lambda) u, \phi_{n,\mathbf{k}} \rangle|^2 \leq \|\mathbf{T}_n (\chi_{\mathfrak{S}_n} - \chi_\Lambda) u\|_2^2 \\ &\leq C_1 \|(\chi_{\mathfrak{S}_n} - \chi_\Lambda) u\|_2^2 = C_1 \int_{\mathfrak{S}_n \setminus \Lambda} |u|^2 d\mathbf{x}. \end{aligned}$$

The last “ $\leq$ ” sign above follows from item (1) of Lemma 3.1. Here, we also used the fact that  $\Lambda \subset \mathfrak{S}_n$  for all  $n$  large enough, which is indeed true because  $\Lambda$  is closed and is strictly contained in  $\Omega$ , an open domain. Following a similar argument as in the proof of Lemma 4.1, we can show that

$$(4.4) \quad \lim_{n \rightarrow \infty} \mathfrak{L}(\mathfrak{S}_n \setminus \Lambda) = 0,$$

where  $\mathfrak{L}(\cdot)$  is the Lebesgue measure. Then

$$\lim_{n \rightarrow \infty} \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n A u\|_2^2 \leq C_1 \lim_{n \rightarrow \infty} \int_{\mathfrak{S}_n \setminus \Lambda} |u|^2 d\mathbf{x} = 0,$$

which verifies the consistency condition (3.10).

**Image deblurring.** In this case,  $Au = a * u$ , where  $*$  is the convolution and  $a \in L_\infty(\mathbb{R}^2) \cap L_2(\mathbb{R}^2)$  is some compactly supported kernel function. We define  $a * u \in L_2(\Omega)$  as

$$(a * u)(\mathbf{x}) = \langle a(\mathbf{x} - \cdot), u \rangle, \quad \mathbf{x} \in \Omega.$$

We choose  $\mathbf{A}_n$  to be the matrix whose  $(\mathbf{i}, \mathbf{j})$ -entry is

$$\mathbf{A}_n[\mathbf{i}, \mathbf{j}] = 2^{-n} \langle a(\cdot - \cdot), \phi_{n,\mathbf{j}-\mathbf{i}} \rangle_{L_2(\mathbb{R}^2)}, \quad \forall \mathbf{i}, \mathbf{j} \in \mathbb{M}^2.$$

Here  $\phi$  is the corresponding refinable function (B-spline) for the given tight wavelet frame system and is assumed to be bounded, compactly supported and satisfy the partition of unity. Now we check Assumption (A1). For  $\phi_{n,\mathbf{k}}$  supported on  $\Omega$ , we have

$$\mathbf{T}_n A u[\mathbf{k}] = 2^n \langle a * u, \phi_{n,\mathbf{k}} \rangle = \langle u, 2^n a(\cdot - \cdot) * \phi_{n,\mathbf{k}} \rangle$$

and

$$\begin{aligned} \mathbf{A}_n \mathbf{T}_n u[\mathbf{k}] &= 2^n \sum_{\mathbf{j} \in \mathbb{M}^2} \mathbf{A}_n[\mathbf{k}, \mathbf{j}] \langle u, \phi_{n,\mathbf{j}} \rangle = \sum_{\mathbf{j} \in \mathbb{M}^2} \langle a(\cdot), \phi_{n,\mathbf{j}-\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \langle u, \phi_{n,\mathbf{j}} \rangle \\ &= \langle u, \sum_{\mathbf{j} \in \mathbb{M}^2} \langle a(\cdot), \phi_{n,\mathbf{j}-\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n,\mathbf{j}} \rangle. \end{aligned}$$

Therefore,

$$\begin{aligned} (4.5) \quad |\mathbf{T}_n A u[\mathbf{k}] - \mathbf{A}_n \mathbf{T}_n u[\mathbf{k}]| &= \left| \langle u, 2^n a(\cdot) * \phi_{n,\mathbf{k}} \rangle - \langle u, \sum_{\mathbf{j} \in \mathbb{M}^2} \langle a(\cdot), \phi_{n,\mathbf{j}-\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n,\mathbf{j}} \rangle \right| \\ &= \left| \langle u, 2^n a(\cdot) * \phi_{n,\mathbf{k}} - a(\cdot + \mathbf{k}/2^n) \rangle \right. \\ &\quad \left. - \langle u, \sum_{\mathbf{j} \in \mathbb{M}^2} \langle a(\cdot), \phi_{n,\mathbf{j}-\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n,\mathbf{j}} - a(\cdot + \mathbf{k}/2^n) \rangle \right| \\ &\leq \|u\|_{L_2(\Omega)} \left( \|2^n a(\cdot) * \phi_{n,\mathbf{k}} - a(\cdot + \mathbf{k}/2^n)\|_{L_2(\Omega)} \right. \\ &\quad \left. + \left\| \sum_{\mathbf{j} \in \mathbb{M}^2} \langle a(\cdot), \phi_{n,\mathbf{j}-\mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n,\mathbf{j}} - a(\cdot + \mathbf{k}/2^n) \right\|_{L_2(\Omega)} \right). \end{aligned}$$

To estimate the first summand of (4.5), denote by “ $\bar{*}$ ” the convolution on the entire  $\mathbb{R}^2$ . We have

$$\begin{aligned} (4.6) \quad \|2^n a(\cdot) * \phi_{n,\mathbf{k}} - a(\cdot + \mathbf{k}/2^n)\|_{L_2(\Omega)} &\leq \|2^n a(\cdot) \bar{*} \phi_{n,\mathbf{k}} - a(\cdot + \mathbf{k}/2^n)\|_{L_2(\mathbb{R}^2)} \\ &= \|a(\cdot) \bar{*} (2^n \phi_{n,\mathbf{0}}) - a(\cdot)\|_{L_2(\mathbb{R}^2)}. \end{aligned}$$

To estimate the second summand of (4.5), let  $\Lambda_{n,\mathbf{k}}$  be the support of  $\phi_{n,\mathbf{k}}$  and define

$$I_n := \{\mathbf{k} \in \mathbb{Z}^2 : \mathfrak{L}(\Lambda_{n,\mathbf{k}} \cap \Omega) > 0\} \setminus \mathbb{M}^2.$$

Let  $\mathfrak{S}_n = \cup_{\mathbf{k} \in I_n} \Lambda_{n,\mathbf{k}}$ . A similar argument of Lemma 4.1 leads to

$$(4.7) \quad \lim_{n \rightarrow \infty} \|\chi_{\mathfrak{S}_n}\|_{L_2(\Omega)} = 0.$$

Then,

$$\begin{aligned}
 & \left\| \sum_{\mathbf{j} \in \mathbb{M}^2} \langle a(\cdot), \phi_{n, \mathbf{j} - \mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} - a(\cdot + \mathbf{k}/2^n) \right\|_{L_2(\Omega)} \\
 & \leq \left\| \sum_{\mathbf{j} \in \mathbb{Z}^2} \langle a(\cdot), \phi_{n, \mathbf{j} - \mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} - a(\cdot + \mathbf{k}/2^n) \right\|_{L_2(\Omega)} \\
 & \quad + \left\| \sum_{\mathbf{j} \in \mathbb{Z}^2 \setminus \mathbb{M}^2} \langle a(\cdot), \phi_{n, \mathbf{j} - \mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} \right\|_{L_2(\Omega)} \\
 & \leq \left\| \sum_{\mathbf{j} \in \mathbb{Z}^2} \langle a(\cdot), \phi_{n, \mathbf{j} - \mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} - a(\cdot + \mathbf{k}/2^n) \right\|_{L_2(\mathbb{R}^2)} \\
 (4.8) \quad & \quad + \left\| \sum_{\mathbf{j} \in I_n} \langle a(\cdot), \phi_{n, \mathbf{j} - \mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} \right\|_{L_2(\Omega)} \\
 & = \left\| \sum_{\mathbf{j} \in \mathbb{Z}^2} \langle a(\cdot), \phi_{n, \mathbf{j} - \mathbf{k}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j} - \mathbf{k}} - a(\cdot) \right\|_{L_2(\mathbb{R}^2)} \\
 & \quad + \left\| \sum_{\mathbf{j} \in I_n} \langle a(\cdot + \mathbf{k}/2^n), \phi_{n, \mathbf{j}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} \right\|_{L_2(\Omega)} \\
 & = \left\| \sum_{\mathbf{j} \in \mathbb{Z}^2} \langle a(\cdot), \phi_{n, \mathbf{j}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} - a(\cdot) \right\|_{L_2(\mathbb{R}^2)} \\
 & \quad + \left\| \sum_{\mathbf{j} \in I_n} \langle a(\cdot + \mathbf{k}/2^n), \phi_{n, \mathbf{j}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} \right\|_{L_2(\Omega)}.
 \end{aligned}$$

Since  $\phi$  is a bounded compactly supported function satisfying the partition of unity, following a similar proof as in Lemma 4.1, we have

$$\begin{aligned}
 (4.9) \quad & \left\| \sum_{\mathbf{j} \in I_n} \langle a(\cdot + \mathbf{k}/2^n), \phi_{n, \mathbf{j}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} \right\|_{L_2(\Omega)} \leq \left\| \sum_{\mathbf{j} \in I_n} \left| \langle a(\cdot + \mathbf{k}/2^n), \phi_{n, \mathbf{j}} \rangle_{L_2(\mathbb{R}^2)} \right| |\phi_{n, \mathbf{j}}| \right\|_{L_2(\Omega)} \\
 & \leq 2^{-n} \|a\|_{L_\infty(\mathbb{R}^2)} \|\phi\|_{L_1(\mathbb{R}^2)} \left\| \sum_{\mathbf{j} \in I_n} |\phi_{n, \mathbf{j}}| \right\|_{L_2(\Omega)} \\
 & = 2^{-n} \|a\|_{L_\infty(\mathbb{R}^2)} \|\phi\|_{L_1(\mathbb{R}^2)} \left\| \sum_{\mathbf{j} \in I_n} |\phi_{n, \mathbf{j}}| \chi_{\Lambda_{n, \mathbf{j}}} \right\|_{L_2(\Omega)} \\
 & \leq C \|a\|_{L_\infty(\mathbb{R}^2)} \|\phi\|_{L_1(\mathbb{R}^2)} \|\phi\|_{L_\infty(\mathbb{R}^2)} \|\chi_{\mathfrak{S}_n}\|_{L_2(\Omega)}.
 \end{aligned}$$

Therefore, by (4.5), (4.6), (4.8), and (4.9), we obtain

$$\begin{aligned}
 & \left\| \mathbf{T}_n A u - \mathbf{A}_n \mathbf{T}_n u \right\|_2^2 \\
 & \leq \|u\|_{L_2(\Omega)}^2 \left\{ \left\| a(\cdot) \bar{*} (2^n \phi_{n, \mathbf{0}}) - a(\cdot) \right\|_{L_2(\mathbb{R}^2)} \right. \\
 (4.10) \quad & \quad + \left\| \sum_{\mathbf{j} \in \mathbb{Z}^2} \langle a(\cdot), \phi_{n, \mathbf{j}} \rangle_{L_2(\mathbb{R}^2)} \phi_{n, \mathbf{j}} - a(\cdot) \right\|_{L_2(\mathbb{R}^2)} \\
 & \quad \left. + C \|a\|_{L_\infty(\mathbb{R}^2)} \|\phi\|_{L_1(\mathbb{R}^2)} \|\phi\|_{L_\infty(\mathbb{R}^2)} \|\chi_{\mathfrak{S}_n}\|_{L_2(\Omega)} \right\}^2.
 \end{aligned}$$

Since  $\phi$  is compactly supported and  $\int_{\mathbb{R}^2} \phi d\mathbf{x} = 1$ , the function  $2^n \phi_{n, \mathbf{0}}$  approximates the delta function. From the standard results in approximation of the identity (see, e.g., [59]), we have

$$(4.11) \quad \lim_{n \rightarrow \infty} \|a(\cdot) \bar{*} (2^n \phi_{n, \mathbf{0}}) - a(\cdot)\|_{L_2(\mathbb{R}^2)} = 0.$$

This, together with (4.2) and (4.7), implies that the right-hand side of (4.10) tends to 0 as  $n \rightarrow \infty$ , i.e., (3.10) is verified.

**4.3. Discretization of Differential Operator  $D$ .** We imposed Assumptions (A2) and (A3) in Section 3.2.2 for the discretization of the differential operators. In this section, we verify that all B-spline tight frame systems constructed by [2] do satisfy these assumptions.

Since we will regard the digital image  $\mathbf{u}$  as the discretization of some function  $u$  through  $\mathbf{T}_n$ , i.e.  $\mathbf{u} = \mathbf{T}_n u$ , the corresponding tight quasi-affine frame systems we shall use are  $X^n(\Psi)$ . We recall the definition of a quasi-affine system (given by (2.8)) for 2D. Given  $\Psi \subset L_2(\mathbb{R}^2)$ , the quasi-affine system  $X^J(\Psi)$  is defined by

$$(4.12) \quad X^J(\Psi) = \{\psi_{\ell,j,\mathbf{k}} : 1 \leq \ell \leq r; j \in \mathbb{Z}, \mathbf{k} \in \mathbb{Z}^2\},$$

where  $\psi_{\ell,j,\mathbf{k}}$  is defined by

$$(4.13) \quad \psi_{\ell,j,\mathbf{k}} := \begin{cases} 2^j \psi_{\ell}(2^j \cdot -\mathbf{k}), & j \geq J; \\ 2^{2j-J} \psi_{\ell}(2^j \cdot -2^{j-J} \mathbf{k}), & j < J. \end{cases}$$

Note from (4.13) that, under the quasi-affine system  $X^n(\Psi)$ , we will have

$$\psi_{\ell,n,\mathbf{k}} = 2^n \psi_{\ell}(2^n \cdot -\mathbf{k}) \quad \text{and} \quad \psi_{\ell,n-1,\mathbf{k}} = 2^{n-2} \psi_{\ell}(2^{n-1} \cdot -\mathbf{k}/2).$$

Since our analysis only focuses on one level of the wavelet frame decomposition, we will be repeatedly using the above two formulas for the dilation levels  $n$  and  $n-1$ .

As we have seen in Section 2.3, our starting point is the fact that the filters in the B-spline framelets are proportional to finite difference operators. More precisely, let  $\{\mathbf{a}_i\}$  denote the collection of filters used in the B-spline tight wavelet frame systems. Then each  $\mathbf{a}_i$  is proportional to some discrete finite difference operator. Since  $\mathbf{T}_n u$  is a discretization of  $u \in L_2(\Omega)$ , we use the filter  $\mathbf{a}_i$  to pass through  $\mathbf{T}_n u$  to approximate a discretization of  $D_i u$ . In particular, we use

$$\lambda_i \mathbf{a}_i[-\cdot] * \mathbf{T}_n u$$

as a discretization of  $D_i u$ . Here  $*$  is the discrete convolution operator. Since  $\Omega$  is an open domain, we do not use any boundary extension. When the convolution needs to use data out of the boundary, we simply neglect it and the convolution is not defined. Let  $S$  be the support of the filters  $\mathbf{a}_i$  and recall that  $\mathbb{K}^2 \subset \mathbb{M}^2$ , the index set where  $\mathbf{a}_i[-\cdot] * \mathbf{T}_n u$  is defined for all  $\mathbf{i}$  (see (6) of Notation 3.1). Then, we have, for any  $\mathbf{k} \in \mathbb{K}^2$ ,

$$\begin{aligned} \lambda_i (\mathbf{a}_i[-\cdot] * \mathbf{T}_n u) [\mathbf{k}] &= \lambda_i \sum_{j \in S+\mathbf{k}} \mathbf{a}_i[j-\mathbf{k}] (\mathbf{T}_n u) [j] = 2^n \lambda_i \sum_{j \in S+\mathbf{k}} \mathbf{a}_i[j-\mathbf{k}] \langle u, \phi_{n,j} \rangle \\ &= 2^n \lambda_i \langle u, \sum_{j \in S+\mathbf{k}} \mathbf{a}_i[j-\mathbf{k}] \phi_{n,j} \rangle = 2^n \lambda_i \langle u, \psi_{i,n-1,\mathbf{k}} \rangle. \end{aligned}$$

According to the definition of the weak derivative and noticing

$$D_i \varphi_{i,n-1,\mathbf{k}} = 2^{s_i(n-1)} \psi_{i,n-1,\mathbf{k}},$$

we get

$$\lambda_i |(\mathbf{a}_i[-\cdot] * \mathbf{T}_n u) [\mathbf{k}]|^p = 2^{np} \lambda_i |\langle u, \psi_{i,n-1,\mathbf{k}} \rangle|^p = 2^{s_i p(1-n)} 2^{np} \lambda_i |\langle D_i u, \varphi_{i,n-1,\mathbf{k}} \rangle|^p.$$

Recall that  $\varphi_i$  is defined in Assumption (A2) as  $D_i \varphi_i = \psi_i$ . Therefore, if  $\lambda_i$  is properly chosen, then  $\lambda_i (\mathbf{a}_i[-\cdot] * \mathbf{T}_n u)$  can be seen as a sampling of  $D_i u$ . Different from the sampling of  $u$  via  $\mathbf{T}_n$  where  $\phi$  is used, the sampling here of  $D_i u$  uses  $\varphi_i$ .

In order for the sampling of  $D_{\mathbf{i}}u$  using  $\varphi_{\mathbf{i}}$  to work, we have to impose Assumption (A2) on  $\varphi_{\mathbf{i}}$  and Assumption (A3) on  $\lambda_{\mathbf{i}}$  as well.

We used  $\mathbb{I}$  to denote the set of double indices where the differential operator  $D_{\mathbf{i}}$  is involved in  $\|\mathbf{D}u\|_{1,p}$ . It might also happen that the differential operator  $D_{\mathbf{j}}$  is not involved, while the corresponding filter  $\mathbf{a}_{\mathbf{j}}$  appears in  $\|\mathbf{W}u\|_{1,p}$ . The set of all such double indices is denoted by  $\mathbb{J}$ . For all  $\mathbf{j} \in \mathbb{J}$ , we need to choose  $\lambda_{\mathbf{j}}$  properly such that, for any  $u \in W_1^s(\Omega)$ , we have that  $(\lambda_{\mathbf{j}})^{1/p} \|\mathbf{a}_{\mathbf{j}}[\cdot] * \mathbf{T}_n u\|_1$  tends to zero, as  $n$  goes to infinity. The choice of  $\lambda_{\mathbf{j}}$  for  $\mathbf{j} \in \mathbb{J}$  in Assumption (A3) is designed for this very purpose.

We note that all B-spline tight wavelet frames constructed by [2] satisfy Assumptions (A2). Here, we provide the details for Haar and piecewise linear framelets as follows. For the uniform form of such a function  $\varphi_{\mathbf{i}}$  for an arbitrary B-spline wavelet frame, we refer the readers to [56] for further details.

- (1) For the Haar framelet, the wavelet functions and corresponding difference operators are listed in the following table, where

$$\psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise,} \end{cases} \quad \varphi(x) = \begin{cases} x & 0 \leq x < \frac{1}{2} \\ \frac{1}{2} - x & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

	$\varphi_{\mathbf{i}}$	$c_{\mathbf{i}}$	$D_{\mathbf{i}}$	$s_{\mathbf{i}}$
$\psi_{1,0}(x, y) = \psi(x)\phi(y)$	$\varphi_{1,0}(x, y) = \varphi(x)\phi(y)$	$c_{1,0} = \frac{1}{4}$	$D_{1,0} = \frac{\partial}{\partial x}$	$s_{1,0} = 1$
$\psi_{0,1}(x, y) = \phi(x)\psi(y)$	$\varphi_{0,1}(x, y) = \phi(x)\varphi(y)$	$c_{0,1} = \frac{1}{4}$	$D_{0,1} = \frac{\partial}{\partial y}$	$s_{0,1} = 1$
$\psi_{1,1}(x, y) = \psi(x)\psi(y)$	$\varphi_{1,1}(x, y) = \varphi(x)\varphi(y)$	$c_{1,1} = \frac{1}{16}$	$D_{1,1} = \frac{\partial^2}{\partial x \partial y}$	$s_{1,1} = 2$

If we want to use the Haar framelet transform to approximate TV, we can choose  $\mathbb{I} = \{(0, 1), (1, 0)\}$  and  $\mathbb{J} = \{(1, 1)\}$ . For this case,  $s = 1$ ,  $s_{\mathbf{i}} = 1$  for  $\mathbf{i} \in \mathbb{I}$  and  $s_{\mathbf{j}} = 2$  for  $\mathbf{j} \in \mathbb{J}$ . Then Assumption (A2) is obviously satisfied by the above table.

- (2) For the piecewise linear framelet, the wavelet functions and corresponding difference operators are listed in the following table, where

$$\psi_1(x) = \begin{cases} \sqrt{2}(x+1) & -1 \leq x < -\frac{1}{2} \\ -\sqrt{2}x & -\frac{1}{2} \leq x < \frac{1}{2} \\ \sqrt{2}(x-1) & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise,} \end{cases} \quad \psi_2(x) = \begin{cases} -1-x & -1 \leq x < -\frac{1}{2} \\ 1+3x & -\frac{1}{2} \leq x < 0 \\ 1-3x & 0 \leq x < \frac{1}{2} \\ x-1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$\varphi_1(x) = \begin{cases} \frac{\sqrt{2}}{2}(x+1)^2 & -1 \leq x < -\frac{1}{2} \\ \frac{\sqrt{2}}{2}\left(\frac{1}{2} - x^2\right) & -\frac{1}{2} \leq x < \frac{1}{2} \\ \frac{\sqrt{2}}{2}(x-1)^2 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$\varphi_2(x) = \begin{cases} -\frac{1}{6}(1+x)^3 & -1 \leq x < -\frac{1}{2} \\ \frac{1}{54}(1+3x)^3 - \frac{1}{6}x - \frac{11}{108} & -\frac{1}{2} \leq x < 0 \\ \frac{1}{6}x + \frac{1}{54}(1-3x)^3 - \frac{11}{108} & 0 \leq x < \frac{1}{2} \\ \frac{1}{6}(x-1)^3 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

	$\varphi_i$	$c_i$	$D_i$	$s_i$
$\psi_{1,0}(x, y) = \psi_1(x)\phi(y)$	$\varphi_{1,0}(x, y) = \varphi_1(x)\phi(y)$	$c_{1,0} = \frac{\sqrt{2}}{4}$	$D_{1,0} = \frac{\partial}{\partial x}$	$s_{1,0} = 1$
$\psi_{0,1}(x, y) = \phi(x)\psi_1(y)$	$\varphi_{0,1}(x, y) = \phi(x)\varphi_1(y)$	$c_{0,1} = \frac{\sqrt{2}}{4}$	$D_{0,1} = \frac{\partial}{\partial y}$	$s_{0,1} = 1$
$\psi_{1,1}(x, y) = \psi_1(x)\psi_1(y)$	$\varphi_{1,1}(x, y) = \varphi_1(x)\varphi_1(y)$	$c_{1,1} = \frac{1}{8}$	$D_{1,1} = \frac{\partial^2}{\partial x \partial y}$	$s_{1,1} = 2$
$\psi_{2,0}(x, y) = \psi_2(x)\phi(y)$	$\varphi_{2,0}(x, y) = \varphi_2(x)\phi(y)$	$c_{2,0} = -\frac{1}{16}$	$D_{2,0} = \frac{\partial^2}{\partial x^2}$	$s_{2,0} = 2$
$\psi_{0,2}(x, y) = \phi(x)\psi_2(y)$	$\varphi_{0,2}(x, y) = \phi(x)\varphi_2(y)$	$c_{0,2} = -\frac{1}{16}$	$D_{0,2} = \frac{\partial^2}{\partial y^2}$	$s_{0,2} = 2$
$\psi_{2,1}(x, y) = \psi_2(x)\psi_1(y)$	$\varphi_{2,1}(x, y) = \varphi_2(x)\varphi_1(y)$	$c_{2,1} = -\frac{\sqrt{2}}{64}$	$D_{2,1} = \frac{\partial^3}{\partial x^2 \partial y}$	$s_{2,1} = 3$
$\psi_{1,2}(x, y) = \psi_1(x)\psi_2(y)$	$\varphi_{1,2}(x, y) = \varphi_1(x)\varphi_2(y)$	$c_{1,2} = -\frac{\sqrt{2}}{64}$	$D_{1,2} = \frac{\partial^3}{\partial x \partial y^2}$	$s_{1,2} = 3$
$\psi_{2,2}(x, y) = \psi_2(x)\psi_2(y)$	$\varphi_{2,2}(x, y) = \varphi_2(x)\varphi_2(y)$	$c_{2,2} = \frac{1}{256}$	$D_{2,2} = \frac{\partial^4}{\partial x^2 \partial y^2}$	$s_{2,2} = 4$

**4.4. Proof of Lemma 3.3.** In this section, we give the proof of Lemma 3.3, i.e.  $\lim_{n \rightarrow \infty} E_n^{(1)}(u) = E^{(1)}(u)$  for  $u \in W_1^s(\Omega)$ .

*Proof of Lemma 3.3.* For simplicity, we only prove this lemma for

$$\|\mathbf{D}u\|_{1,p} = \int_{\Omega} \left( \sum_{\mathbf{i} \in \mathbb{I}} |D_{\mathbf{i}}u|^p \right)^{1/p} d\mathbf{x}$$

and, correspondingly,

$$\|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} = h^2 \sum_{\mathbf{k} \in \mathbb{K}^2} \left( \sum_{\mathbf{i} \in \mathbb{I}} \lambda_{\mathbf{i}} |(\mathbf{a}_{\mathbf{i}}[-\cdot] * \mathbf{T}_n u)[\mathbf{k}]|^p + \sum_{\mathbf{j} \in \mathbb{J}} \lambda_{\mathbf{j}} |(\mathbf{a}_{\mathbf{j}}[-\cdot] * \mathbf{T}_n u)[\mathbf{k}]|^p \right)^{1/p}.$$

Other cases can be shown similarly. Recall that the sets  $\mathbb{I}$  and  $\mathbb{J}$  are defined as

$$\mathbb{I} := \{\mathbf{i} : D_{\mathbf{i}} \text{ is in } \mathbf{D}\} \quad \text{and} \quad \mathbb{J} := \mathbb{B} \setminus \mathbb{I},$$

where  $\mathbb{B} = \{\mathbf{i} : 0 \leq i_1, i_2 \leq r\} \setminus \{(0,0)\}$ . The parameters  $\lambda_{\mathbf{i}}$  and  $\lambda_{\mathbf{j}}$  are properly chosen so that Assumption (A3) is satisfied.

Let us first show the lemma when  $\mathbb{J} = \emptyset$ . Note that  $D_{\mathbf{i}}\varphi_{\mathbf{i}} = \psi_{\mathbf{i}}$  implies

$$D_{\mathbf{i}}\varphi_{\mathbf{i},n-1,\mathbf{k}} = 2^{(s_{\mathbf{i}}+1)(n-1)-1} \psi_{\mathbf{i}}(2^{n-1} \cdot -\mathbf{k}/2) = 2^{s_{\mathbf{i}}(n-1)} \psi_{\mathbf{i},n-1,\mathbf{k}}.$$

Since  $\varphi_{\mathbf{i}}$  is smooth and compactly supported and both  $\psi_{\mathbf{i},n-1,\mathbf{k}}$  and  $\varphi_{\mathbf{i},n-1,\mathbf{k}}$  are supported on  $\Omega$ , according to the definition of weak derivatives, we have

$$\langle D_{\mathbf{i}}u, \varphi_{\mathbf{i},n-1,\mathbf{k}} \rangle = (-1)^{s_{\mathbf{i}}} \langle u, D_{\mathbf{i}}\varphi_{\mathbf{i},n-1,\mathbf{k}} \rangle = (-1)^{s_{\mathbf{i}}} 2^{s_{\mathbf{i}}(n-1)} \langle u, \psi_{\mathbf{i},n-1,\mathbf{k}} \rangle.$$

On the other hand,

$$\lambda_{\mathbf{i}} |(\mathbf{a}_{\mathbf{i}}[-\cdot] * \mathbf{T}_n u)[\mathbf{k}]|^p = 2^{np} \left( \frac{1}{c_{\mathbf{i}}} 2^{s_{\mathbf{i}}(n-1)} \right)^p |\langle u, \psi_{\mathbf{i},n-1,\mathbf{k}} \rangle|^p,$$

and hence

$$\begin{aligned} \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} &= h^2 \sum_{\mathbf{k} \in \mathbb{K}^2} \left( \sum_{\mathbf{i} \in \mathbb{I}} \lambda_{\mathbf{i}} |(\mathbf{a}_{\mathbf{i}}[-\cdot] * \mathbf{T}_n u)[\mathbf{k}]|^p \right)^{\frac{1}{p}} \\ &= 2^{-n} \sum_{\mathbf{k} \in \mathbb{K}^2} \left( \sum_{\mathbf{i} \in \mathbb{I}} |\langle D_{\mathbf{i}}u, \frac{1}{c_{\mathbf{i}}} \varphi_{\mathbf{i},n-1,\mathbf{k}} \rangle|^p \right)^{\frac{1}{p}}. \end{aligned}$$

Let  $\square_{\mathbf{k}}$  be the rectangular domain  $[\frac{k_1}{2^n}, \frac{k_1+1}{2^n}] \times [\frac{k_2}{2^n}, \frac{k_2+1}{2^n}]$  where  $\mathbf{k} = (k_1, k_2)$ , and recall that  $\mathbb{O} = \{0, 1, \dots, N-1\}^2$  and  $\Omega = (0, 1)^2$ . Then the above relations imply

that

$$\begin{aligned}
 (4.14) \quad & \left| \|Du\|_{1,p} - \|\lambda_n \cdot W_n T_n u\|_{1,p} \right| \\
 &= \left| \sum_{\mathbf{k} \in \mathbb{O}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} d\mathbf{x} - 2^{-n} \sum_{\mathbf{k} \in \mathbb{K}^2} \left( \sum_{i \in \mathbb{I}} \left| \langle D_i u, \frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} \rangle \right|^p \right)^{1/p} \right| \\
 &\leq \sum_{\mathbf{k} \in \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left| \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} - 2^n \left( \sum_{i \in \mathbb{I}} \left| \langle D_i u, \frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} \rangle \right|^p \right)^{1/p} \right| d\mathbf{x} + \sum_{\mathbf{k} \in \mathbb{O}^2 \setminus \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} d\mathbf{x} \\
 &\leq \sum_{\mathbf{k} \in \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u - 2^n \langle D_i u, \frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} \rangle|^p \right)^{1/p} d\mathbf{x} + \sum_{\mathbf{k} \in \mathbb{O}^2 \setminus \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} d\mathbf{x} \\
 &\leq \sum_{\mathbf{k} \in \mathbb{O}^2} \int_{\square_{\mathbf{k}}} \sum_{i \in \mathbb{I}} |D_i u - 2^n \langle D_i u, \frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} \rangle| d\mathbf{x} + \sum_{\mathbf{k} \in \mathbb{O}^2 \setminus \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} d\mathbf{x} \\
 &= \sum_{i \in \mathbb{I}} \sum_{\mathbf{k} \in \mathbb{O}^2} \int_{\square_{\mathbf{k}}} |D_i u - 2^n \langle D_i u, \frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} \rangle| d\mathbf{x} + \sum_{\mathbf{k} \in \mathbb{O}^2 \setminus \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} d\mathbf{x} \\
 &= \sum_{i \in \mathbb{I}} \int_{\Omega} |D_i u - \sum_{\mathbf{k} \in \mathbb{O}^2} 2^n \langle D_i u, \frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} \rangle \chi_{\square_{\mathbf{k}}} | d\mathbf{x} + \sum_{\mathbf{k} \in \mathbb{O}^2 \setminus \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} d\mathbf{x} \\
 &= \sum_{i \in \mathbb{I}} \|D_i u - \sum_{\mathbf{k} \in \mathbb{O}^2} 2^n \langle D_i u, \frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} \rangle \chi_{\square_{\mathbf{k}}}\|_{L_1(\Omega)} + \sum_{\mathbf{k} \in \mathbb{O}^2 \setminus \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} d\mathbf{x}.
 \end{aligned}$$

The second to the last identity follows from the fact that  $\cup_{\mathbf{k} \in \mathbb{O}^2} \square_{\mathbf{k}} = \overline{\Omega}$  and  $\mathfrak{L}(\square_{\mathbf{k}} \cap \square_{\mathbf{j}}) = 0$  for  $\mathbf{k} \neq \mathbf{j}$ . It remains to show that

$$(4.15) \quad \lim_{n \rightarrow \infty} \sum_{\mathbf{k} \in \mathbb{O}^2 \setminus \mathbb{K}^2} \int_{\square_{\mathbf{k}}} \left( \sum_{i \in \mathbb{I}} |D_i u|^p \right)^{1/p} d\mathbf{x} = 0$$

and

$$(4.16) \quad \lim_{n \rightarrow \infty} \|D_i u - \sum_{\mathbf{k} \in \mathbb{O}^2} 2^n \langle D_i u, \frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} \rangle \chi_{\square_{\mathbf{k}}}\|_{L_1(\Omega)} = 0, \quad \text{for } i \in \mathbb{I}.$$

For (4.15), it is easy to show that  $\lim_{n \rightarrow \infty} \mathfrak{L}(\cup_{\mathbf{k} \in \mathbb{O}^2 \setminus \mathbb{K}^2} \square_{\mathbf{k}}) = 0$ . Therefore, by the Lebesgue dominated convergence theorem, we get (4.15). For (4.16), notice that  $2^n \chi_{\square_{\mathbf{k}}} = \phi_{n,\mathbf{k}}^{(H)}$ , where  $\phi^{(H)}$  is the characteristic function on the unit square, i.e., the tensor product of the piecewise constant B-spline which satisfies the partition of unity property. Furthermore, by the definition of quasi-affine systems (4.13), we have

$$\frac{1}{c_i} \varphi_{i,n-1,\mathbf{k}} = \frac{1}{4c_i} 2^n \varphi_i(2^{n-1}x - \mathbf{k}/2) = \left( \frac{\varphi_i(\cdot/2)}{4c_i} \right)_{n,\mathbf{k}}$$

and  $\int_{\Omega} \frac{\varphi_i(\cdot/2)}{4c_i} d\mathbf{x} = 1$ . We also note that the support of  $\varphi_i(\cdot/2)$  contains the support of  $\phi^{(H)}$ . Together with  $D_i u \in L_1(\Omega)$ , we establish (4.16) by Lemma 4.1.

In the case of  $\mathbb{J} \neq \emptyset$ , if we can show that, for all  $\mathbf{j} \in \mathbb{J}$ ,

$$(4.17) \quad \lim_{n \rightarrow \infty} (\lambda_{\mathbf{j}})^{1/p} \|\mathbf{a}_{\mathbf{j}}[-] * \mathbf{T}_n u\|_1 = 0,$$

then we get (3.15). Indeed, define

$$E_I = h^2 \sum_{\mathbf{k} \in \mathbb{K}^2} \left( \sum_{\mathbf{i} \in \mathbb{I}} \lambda_{\mathbf{i}} |(\mathbf{a}_{\mathbf{i}}[-\cdot] * \mathbf{T}_n u)[\mathbf{k}]|^p \right)^{1/p}.$$

Then we have

$$E_I \leq \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} \leq E_I + \sum_{j \in \mathbb{J}} (\lambda_j)^{1/p} \|\mathbf{a}_j[-\cdot] * \mathbf{T}_n u\|_1.$$

Taking the limit of the above inequality and noticing (4.17) leads to

$$\lim_{n \rightarrow \infty} E_I = \lim_{n \rightarrow \infty} \|\boldsymbol{\lambda}_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p}.$$

It now remains to show (4.17).

By assumption (A2), there exist  $\varphi_j$  and  $\varphi_{j'}$  such that  $D_j \varphi_j = \psi_j$  and  $D_{j'} \varphi_{j'} = \psi_{j'}$ . Choose  $j'$  satisfying  $\mathbf{0} \leq j' < j$  and  $s_{j'} \leq s$ , as suggested in assumption A(3). Note that such a  $j'$  always exists, since, for example, one may pick  $j' = \mathbf{0}$ . Let  $\bar{\psi}_j = D_{j-j'} \varphi_j$ . Then obviously we have  $D_{j'} \bar{\psi}_j = \psi_j$ , due to the tensor product structure of  $\varphi_j$  that ensures  $D_{j'} D_{j-j'} \varphi_j = D_j \varphi_j$ . For any real number  $t \geq 0$ , the function

$$\tilde{\varphi}_t := \frac{1}{c_{j'}} \varphi_{j'} + t \bar{\psi}_j,$$

with  $c_{j'} = \int_{\mathbb{R}^2} \varphi_{j'} d\mathbf{x}$ , is smooth, compactly supported, and of integral 1 (since obviously  $\int_{\mathbb{R}^2} \psi_j d\mathbf{x} = 0$ ). This together with  $D_{j'} \tilde{\varphi}_t = \frac{1}{c_{j'}} \psi_{j'} + t \psi_j$  leads to, for  $u \in W_1^s(\Omega)$ ,

$$\langle D_{j'} u, \tilde{\varphi}_{t,n-1,k} \rangle = (-1)^{s_{j'}} 2^{s_{j'}(n-1)} \langle u, \frac{1}{c_{j'}} \psi_{j',n-1,k} + t \psi_{j,n-1,k} \rangle.$$

Therefore,

$$\|2^{s_{j'}(n-1)} (\frac{1}{c_{j'}} \mathbf{a}_{j'}[-\cdot] + t \mathbf{a}_j[-\cdot]) * \mathbf{T}_n u\|_1 = 2^{-n} \sum_{\mathbf{k} \in \mathbb{K}^2} |\langle D_{j'} u, \tilde{\varphi}_{t,n-1,k} \rangle|.$$

Following the exact same steps as in (4.14), by removing the summation with respect to  $\mathbf{i} \in \mathbb{I}$ , replacing  $\varphi$  by  $\tilde{\varphi}_t$ ,  $D_{\mathbf{i}}$  by  $D_{j'}$  and  $c_{\mathbf{i}}$  by  $c_{j'}$ , we have

$$\|D_{j'} u\|_1 = \lim_{n \rightarrow \infty} \|2^{s_{j'}(n-1)} (\frac{1}{c_{j'}} \mathbf{a}_{j'}[-\cdot] + t \mathbf{a}_j[-\cdot]) * \mathbf{T}_n u\|_1.$$

In particular, when  $t = 0$ , we have

$$\|D_{j'} u\|_1 = \lim_{n \rightarrow \infty} \|2^{s_{j'}(n-1)} \frac{1}{c_{j'}} \mathbf{a}_{j'}[-\cdot] * \mathbf{T}_n u\|_1.$$

These two equations imply that

$$t \limsup_{n \rightarrow \infty} \|2^{s_{j'}(n-1)} \mathbf{a}_j[-\cdot] * \mathbf{T}_n u\|_1 \leq 2 \|D_{j'} u\|_1.$$

Since  $t$  is arbitrary, we must have

$$\lim_{n \rightarrow \infty} \|2^{s_{j'}(n-1)} \mathbf{a}_j[-\cdot] * \mathbf{T}_n u\|_1 = 0.$$

In view of  $0 \leq \lambda_j \leq C(2^{s_{j'}(n-1)})^p$ , we get (4.17). This finishes the proof.  $\square$

**4.5. Proof of Proposition 3.2.** In this section, we provide the proof of Proposition 3.2, which states that  $E_n$  is continuous uniformly in  $n$ .

*Proof of Proposition 3.2.* We consider  $E_n^{(1)}$  and  $E_n^{(2)}$  separately as the property in the proposition is additive.

We first prove the uniform continuity of  $E_n^{(1)}$ . Define the space  $\ell_{1,p}^*(\mathbb{Z}^2) := \{\mathbf{b} : \|\mathbf{b}\|_{1,p}^* < +\infty\}$  with

$$\|\mathbf{b}\|_{1,p}^* = \sum_{\mathbf{j} \in \mathbb{Z}^2} \left( \sum_{\mathbf{i}=(0,0)}^{(r,r)} |\mathbf{b}_{\mathbf{i}[\mathbf{j}]}|^p \right)^{\frac{1}{p}},$$

which can be regarded as a finite tensor of the space of all absolute summable sequences on  $\mathbb{Z}^2$ . We also recall the  $\ell_{1,p}$ -norm that we defined in (7) of Notation 3.1 (with  $\lambda = \mathbf{1}$ ), as follows:

$$\|\mathbf{b}\|_{1,p} = \sum_{\mathbf{j} \in \mathbb{K}^2} \left( \sum_{\mathbf{i}=(0,0)}^{(r,r)} |\mathbf{b}_{\mathbf{i}[\mathbf{j}]}|^p \right)^{\frac{1}{p}} 2^{-2n}.$$

Since for any given  $n$  and  $v \in W_1^s(\Omega)$ , we have  $\lambda_n \cdot \mathbf{W}_n \mathbf{T}_n v \in \ell_{1,p}^*(\mathbb{Z}^2)$ ,

$$\|2^{-2n} \lambda_n \cdot \mathbf{W}_n \mathbf{T}_n v\|_{1,p}^* = \|\lambda_n \cdot \mathbf{W}_n \mathbf{T}_n v\|_{1,p}.$$

Since  $\mathbf{T}_n$  is a bounded linear operator on  $L_2(\Omega)$  and  $\mathbf{W}_n$  is a linear operator on a finite dimensional space, then we have

$$\|2^{-2n} \lambda_n \cdot \mathbf{W}_n \mathbf{T}_n v\|_{1,p}^* \leq C_n \|v\|_{L_2(\Omega)} \leq \tilde{C}_n \|v\|_{W_1^s(\Omega)},$$

where the last inequality follows from Sobolev imbedding theorems (see, e.g., [54]). This shows that

$$2^{-2n} \lambda_n \cdot \mathbf{W}_n \mathbf{T}_n : W_1^s(\Omega) \mapsto \ell_{1,p}^*(\mathbb{Z}^2)$$

is a bounded linear operator. In addition, for any fixed  $v \in W_1^s(\Omega)$ , we

$$\lim_{n \rightarrow \infty} \|\lambda_n \cdot \mathbf{W}_n \mathbf{T}_n v\|_{1,p} = \|\mathbf{D}v\|_{1,p}$$

by Lemma 3.3. Therefore,

$$\sup_n \|2^{-2n} \lambda_n \cdot \mathbf{W}_n \mathbf{T}_n v\|_{1,p}^* = \sup_n \|\lambda_n \cdot \mathbf{W}_n \mathbf{T}_n v\|_{1,p} < +\infty$$

for every  $v \in W_1^s(\Omega)$ . By applying the uniform boundedness principle, we get

$$\sup_n \|2^{-2n} \lambda_n \cdot \mathbf{W}_n \mathbf{T}_n\|_{op} \leq C,$$

where  $\|\cdot\|_{op}$  stands for the operator norm and  $C$  is a constant independent of  $n$ . Then,

$$\begin{aligned} |E_n^{(1)}(u) - E_n^{(1)}(v)| &= \|\lambda_n \cdot \mathbf{W}_n \mathbf{T}_n u\|_{1,p} - \|\lambda_n \cdot \mathbf{W}_n \mathbf{T}_n v\|_{1,p} \\ &\leq \|\lambda_n \cdot \mathbf{W}_n \mathbf{T}_n(u - v)\|_{1,p} = \|2^{-2n} \lambda_n \cdot \mathbf{W}_n \mathbf{T}_n(u - v)\|_{1,p}^* \\ &\leq C \|u - v\|_{W_1^s(\Omega)}, \end{aligned}$$

which implies that the first term is Lipschitz continuous with Lipschitz constant independent of  $n$ . Therefore, by choosing  $N = 1$  and  $\delta = \epsilon/C$ , the proposition is proved for  $E_n^{(1)}$ .

Next, we prove the uniform continuity of  $E_n^{(2)}(u)$  with respect to  $n$  at each fixed  $u$ . Define  $x(u) = \|Au - f\|_{L_2(\Omega)}$  for any fixed  $u \in L_2(\Omega)$ . For any given  $\epsilon$ , it is obvious that there exists  $\delta'$  such that

$$(4.18) \quad \left| \frac{1}{2}x^2(u) - \frac{1}{2}y^2 \right| < \epsilon/2, \quad \forall |x(u) - y| < \delta'.$$

With this  $\delta'$ , by Lemma 3.2, there exists an  $\mathcal{N}$  such that

$$(4.19) \quad \|\mathbf{A}_n \mathbf{T}_n u - \mathbf{T}_n f\|_2 - x(u) < \delta'/2, \quad \forall n > \mathcal{N},$$

and thus by (4.18) we get

$$(4.20) \quad \left| \frac{1}{2}x^2(u) - E_n^{(2)}(u) \right| < \epsilon/2, \quad \forall n > \mathcal{N}.$$

Similarly as before, we define the space  $\ell_2^*(\mathbb{Z}^2) := \{\mathbf{b} : \|\mathbf{b}\|_2^* < \infty\}$  with  $\|\mathbf{b}\|_2^* := \left(\sum_{j \in \mathbb{Z}^2} |\mathbf{b}[j]|^2\right)^{\frac{1}{2}}$ . Then, for any given  $n$ , the linear operator

$$2^{-n}(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A) : W_1^s(\Omega) \mapsto \ell_2^*(\mathbb{Z}^2)$$

is bounded. Note from the definition of the discrete  $\ell_2$ -norm defined by (3.1), we have

$$\|(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)v\|_2 = \|2^{-n}(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)v\|_2^*.$$

By Assumption (A1),  $\|(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)v\|_2$ , hence  $\|2^{-n}(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)v\|_2^*$  as well, converges to 0 for any given  $v \in W_1^s(\Omega)$ . By the uniform boundedness principle and following a similar argument as before, we can show that  $2^{-n}(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)$  is uniformly bounded on  $W_1^s(\Omega)$ , i.e.,  $\|2^{-n}(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)\|_{op} \leq B$ , where  $\|\cdot\|_{op}$  stands for the operator norm and  $B$  is a positive number independent of  $n$ . Recall the Sobolev imbedding theorem [54] once again; i.e., for an arbitrary  $v \in W_1^s(\Omega)$ ,  $\|v\|_{L_2(\Omega)} \leq \tilde{C}\|v\|_{W_1^s(\Omega)}$ . Then, for arbitrary  $v, w \in W_1^s(\Omega) \subset L_2(\Omega)$ ,

$$\begin{aligned} & \left| \|\mathbf{A}_n \mathbf{T}_n v - \mathbf{T}_n f\|_2 - \|\mathbf{A}_n \mathbf{T}_n w - \mathbf{T}_n f\|_2 \right| \\ & \leq \|\mathbf{A}_n \mathbf{T}_n(v - w)\|_2 \leq \|\mathbf{T}_n A(v - w)\|_2 + \|\mathbf{A}_n \mathbf{T}_n(v - w) - \mathbf{T}_n A(v - w)\|_2 \\ & \leq C\|A(v - w)\|_{L_2(\Omega)} + \|(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)(v - w)\|_2 \\ & = C\|A(v - w)\|_{L_2(\Omega)} + \|2^{-n}(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)(v - w)\|_2^* \\ & \leq (\tilde{C}\|A\|_{op} + \|2^{-n}(\mathbf{A}_n \mathbf{T}_n - \mathbf{T}_n A)\|_{op})\|v - w\|_{W_1^s(\Omega)} \\ & \leq (\tilde{C}\|A\|_{op} + B)\|v - w\|_{W_1^s(\Omega)}. \end{aligned}$$

In other words,  $\|\mathbf{A}_n \mathbf{T}_n v - \mathbf{T}_n f\|_2$  is Lipschitz continuous and the Lipschitz constant is independent of  $n$ . So, if we choose  $\delta = \frac{\delta'}{2(\tilde{C}\|A\|_{op} + B)}$ , then we have

$$\left| \|\mathbf{A}_n \mathbf{T}_n w - \mathbf{T}_n f\|_2 - \|\mathbf{A}_n \mathbf{T}_n v - \mathbf{T}_n f\|_2 \right| \leq \frac{\delta'}{2}, \quad \forall \|w - v\|_{W_1^s(\Omega)} < \delta.$$

For given  $u$ , this together with (4.19) leads to

$$\left| x(u) - \|\mathbf{A}_n \mathbf{T}_n v - \mathbf{T}_n f\|_2 \right| \leq \delta', \quad \forall \|u - v\|_{W_1^s(\Omega)} < \delta, \quad n > \mathcal{N}.$$

By (4.20) and letting  $y = \|\mathbf{A}_n \mathbf{T}_n v - \mathbf{T}_n f\|_2 = E_n^{(2)}(v)$  in (4.18) we obtain

$$|E_n^{(2)}(u) - E_n^{(2)}(v)| \leq \left| \frac{1}{2}x^2(u) - E_n^{(2)}(u) \right| + \left| \frac{1}{2}x^2(u) - E_n^{(2)}(v) \right| < \epsilon,$$

for given  $u$  and all  $v$ , such that  $\|u - v\|_{W_1^s(\Omega)} < \delta$ ,  $n > \mathcal{N}$ .  $\square$

## 5. Algorithm and Experiments

In this section, we will present an efficient numerical algorithm that solves the proposed frame based image restoration approach (2.18) and then conduct some numerical experiments to show the effects of using the  $\ell_{1,2}$ -norm and its advantages over the standard  $\ell_{1,1}$ -norm that has been used in the literature. The efficient algorithms for the standard  $\ell_{1,1}$ -norm have been developed and used for various wavelet frame based approaches. The interested reader should consult [33] and [34] for details. The emphasis here is to present an efficient numerical algorithm that solves the proposed frame based image restoration approach (2.18) with the  $\ell_{1,2}$ -norm and to compare it with the standard  $\ell_{1,1}$ -norm for the analysis based approach of (2.18).

In our numerical simulations, we choose the conventional periodic boundary condition for the convolutions of the fast wavelet transforms. Unpleasant artifacts can be reduced by using proper boundary conditions. Comparing with what was used in the theoretical analysis given in previous sections, other boundary conditions may put back some or all (modified) boundary elements, as well as the corresponding sample data, in the wavelet system whose supports overlap with the boundary of  $\bar{\Omega}$ . Since  $\Omega$  is an open set, we may assume that the boundary elements are bounded and of the same smoothness as other elements in the system on the open set  $\Omega$ . Furthermore, the boundary elements at different dilation levels can usually be made to be a dilation of the boundary elements at the ground level. Hence the contributions of boundary elements will vanish as the resolution level goes to infinity, since the measure of the total supports of the boundary elements will go to zero as the resolution level goes to infinity. Therefore, as long as proper boundary elements are chosen, some careful modifications can be made so that the results given in the previous sections are still valid for some properly chosen boundary conditions.

One of the major difficulties of solving (2.18) is that the  $\ell_{1,p}$ -norm is not smooth. This prevents us from using optimization methods designed for smooth functions. Another difficulty is that the term  $\|\boldsymbol{\lambda} \cdot \mathbf{W}\mathbf{u}\|_{1,p}$  is not separable since  $\mathbf{W}$  couples the entries of  $\mathbf{u}$  together. Therefore, one cannot simply use the soft thresholding as one normally does for the synthesis based approach [46, 47, 48, 49, 50]. To overcome this, we can convert (2.18) to a problem involving only separable nonsmooth terms by a simple change of variables, and then iteratively enforce the constraints on the change of variables via Bregman iterations [60, 61]. This is the main idea of the recently proposed split Bregman algorithm in [19] which shows great efficiency in solving  $\ell_1$ -norm related optimization.

**5.1. Split Bregman Algorithm.** The split Bregman algorithm was first proposed in [19] and was shown to be powerful in [19, 62] when it is applied to various PDE based image restoration approaches, e.g., ROF and nonlocal variational models. Convergence analysis of the split Bregman and its application to the standard  $\ell_{1,1}$ -norm for the analysis based approach of (2.18) were given in [21].

The idea of the split Bregman algorithm is as follows. One first replaces the term  $\mathbf{W}\mathbf{u}$  in (2.18) by a new variable  $\mathbf{d}$  and then adds a new constraint  $\mathbf{d} = \mathbf{W}\mathbf{u}$  into (2.18). Hence, (2.18) is now equivalent to

$$(5.1) \quad \inf_{\mathbf{u}, \mathbf{d}} \|\boldsymbol{\lambda} \cdot \mathbf{d}\|_{1,p} + \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2 \quad \text{subject to} \quad \mathbf{d} = \mathbf{W}\mathbf{u}.$$

In order to solve (5.1), an iterative algorithm based on the Bregman distance [20, 60] with an inexact solver was proposed in [19]. This leads to the alternating split Bregman algorithm for (5.1).

The derivation of the split Bregman algorithm in [19, 21] is based on Bregman distance. It was recently shown (see, e.g., [23, 63]) that the split Bregman algorithm can also be derived by applying the augmented Lagrangian method (see, e.g., [64]) on (5.1). The connection between the split Bregman algorithm and the Douglas Rachford splitting was addressed by [22]. We shall skip the detailed derivations and recall the split Bregman algorithm that solves (2.18) as follows:

$$(5.2) \quad \begin{cases} \mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2 + \frac{\mu}{2} \|\mathbf{W}\mathbf{u} - \mathbf{d}^k + \mathbf{b}^k\|_2^2, \\ \mathbf{d}^{k+1} = \arg \min_{\mathbf{d}} \|\boldsymbol{\lambda} \cdot \mathbf{d}\|_{1,p} + \frac{\mu}{2} \|\mathbf{d} - \mathbf{W}\mathbf{u}^{k+1} - \mathbf{b}^k\|_2^2, \\ \mathbf{b}^{k+1} = \mathbf{b}^k + \delta(\mathbf{W}\mathbf{u}^{k+1} - \mathbf{d}^{k+1}). \end{cases}$$

Note that the minimal values of the two subproblems of (5.2) are indeed attainable. Therefore, we can use “argmin” here, as well as many other places for the rest of this and the next section.

The first subproblem of (5.2) can be solved easily as follows:

$$(5.3) \quad \mathbf{u}^{k+1} = (\mathbf{A}^\top \mathbf{A} + \mu \mathbf{I})^{-1} (\mathbf{A}^\top \mathbf{f} + \mu \mathbf{W}^\top (\mathbf{d}^k - \mathbf{b}^k)),$$

where the inverse can be computed efficiently when  $\mathbf{A}$  is simply a projection operator for inpainting problems or a convolution operator for deconvolution problems.

The second subproblem of (5.2) can also be solved rather efficiently. For the case  $p = 1$ ,  $\mathbf{d}^{k+1}$  can be obtained by soft-thresholding as follows (see, e.g., [65, 66]):

$$(5.4) \quad \mathbf{d}^{k+1} = \mathcal{T}_{\boldsymbol{\lambda}/\mu}^1(\mathbf{W}\mathbf{u}^{k+1} + \mathbf{b}^k),$$

where the soft-thresholding operator  $\mathcal{T}_\tau^1(\mathbf{v})$  is an entry-wise operation defined by

$$\mathcal{T}_\tau^1(\mathbf{v}) := \frac{\mathbf{v}}{|\mathbf{v}|} \max\{|\mathbf{v}| - \tau, 0\}.$$

The superscript 1 of  $\mathcal{T}_\tau^1$  emphasizes that it is the shrinkage operator corresponding to the  $\ell_{1,1}$ -norm. We shall call  $\mathcal{T}_\tau^1$  *anisotropic shrinkage*.

For  $p = 2$ , we will need some proper assumptions on the parameter  $\boldsymbol{\lambda}$  in order to obtain a neat analytical formula for  $\mathbf{d}^{k+1}$ . Let  $J_l$ , for each  $0 \leq l \leq L - 1$ , be a subset of the index set  $\{\mathbf{i} := (i_1, i_2) : 0 \leq i_1, i_2 \leq r\}$  and denote  $J_l^c := \{\mathbf{i} : 0 \leq i_1, i_2 \leq r\} \setminus J_l$ . Assume that

$$(5.5) \quad \lambda_{l,\mathbf{i}} = \begin{cases} 0, & \mathbf{i} \in J_l \\ \tau_l, & \mathbf{i} \in J_l^c, \end{cases}$$

for each  $0 \leq l \leq L - 1$  and  $\tau_l \in \mathcal{I}_2$ . Note that by virtue of the structure of the framelet decomposition  $\mathbf{W}\mathbf{u}$ ,  $(0, 0) \notin J_l$  unless for  $l = L - 1$ . Since the band  $\mathbf{i} = (0, 0)$  of  $\mathbf{W}\mathbf{u}$  corresponds to the low frequency components of  $u$ , which is not sparse in general, we should not penalize the  $\ell_1$ -norm of it and assume that  $(0, 0) \in J_{L-1}$ , i.e.  $\lambda_{L-1,0,0} = 0$ . Under this assumption on  $\boldsymbol{\lambda}$ , we then have

$$(5.6) \quad \mathbf{w}^* = \arg \min_{\mathbf{w}} \|\boldsymbol{\lambda} \cdot \mathbf{w}\|_{1,2} + \frac{\mu}{2} \|\mathbf{w} - \mathbf{v}\|_2^2 = \mathcal{T}_{\boldsymbol{\lambda}/\mu}^2(\mathbf{v}),$$

where

$$(5.7) \quad \left( \mathcal{T}_{\boldsymbol{\lambda}/\mu}^2(\mathbf{v}) \right)_{l,\mathbf{i}} = \begin{cases} v_{l,\mathbf{i}}, & \mathbf{i} \in J_l \\ \frac{v_{l,\mathbf{i}}}{R_l} \max\{R_l - \tau_l/\mu, 0\}, & \mathbf{i} \in J_l^c, \end{cases}$$

with  $R_l := \left(\sum_{i \in J_l^c} |v_{l,i}|^2\right)^{1/2}$  and  $\tau_l$  defined as in (5.5). In this paper, we shall call  $\mathcal{T}_\tau^2$  *isotropic shrinkage*. The isotropic shrinkage (5.7) was first considered by [67] for the Haar framelet. The optimality property of the shrinkage (5.7), i.e. the validity of the second equality in (5.6), is given by [68].

Now combining (5.3), (5.4) and (5.6) with (5.2), we have the following algorithm solving the analysis based approach (2.18) for the case  $p = 1, 2$ . Since the proof of convergence of the split Bregman algorithm given by [21] is very general, it directly implies the convergence of the following Algorithm 5.1.

**Algorithm 5.1.** (*Split Bregman*)

- (i) Set initial guesses  $\mathbf{d}_0$  and  $\mathbf{b}_0$ . Choose an appropriate set of parameters  $(\boldsymbol{\lambda}, \mu, \rho)$ .
- (ii) For  $k = 0, 1, \dots$ , perform the following iterations until convergence:

$$\begin{aligned}
 \mathbf{u}^{k+1} &= (\mathbf{A}^\top \mathbf{A} + \mu \mathbf{I})^{-1} (\mathbf{A}^\top \mathbf{f} + \mu \mathbf{W}^\top (\mathbf{d}^k - \mathbf{b}^k)), \\
 \mathbf{d}^{k+1} &= \mathcal{T}_{\boldsymbol{\lambda}/\mu}^p (\mathbf{W} \mathbf{u}^{k+1} + \mathbf{b}^k), \\
 \mathbf{b}^{k+1} &= \mathbf{b}^k + \rho (\mathbf{W} \mathbf{u}^{k+1} - \mathbf{d}^{k+1}).
 \end{aligned}
 \tag{5.8}$$

*Remark 5.1.* For the case  $p = 2$  and  $\boldsymbol{\lambda}$  not satisfying assumption (5.5), one can still derive an analytic form, theoretically at least, for the optimization problem (5.6). However, the analytic form is rather complicated because at some point, we need to select one of the four roots of a quadratic polynomial. Therefore, using numerical optimization techniques to solve (5.6) for general  $\boldsymbol{\lambda}$  seems more realistic. In this paper, the  $\boldsymbol{\lambda}$  we use will satisfy (5.5), so we will not go into details of solving (5.6) for general  $\boldsymbol{\lambda}$ .

The theory of [21] can be applied here, leading to the following convergence result: let  $\mathbf{u}^k$  be the sequence generated by Algorithm 5.1 for the given objective function

$$F(\mathbf{u}) = \|\boldsymbol{\lambda} \cdot \mathbf{W} \mathbf{u}\|_1 + \frac{1}{2} \|\mathbf{A} \mathbf{u} - \mathbf{f}\|_2^2;$$

then

$$\lim_{k \rightarrow \infty} F(\mathbf{u}^k) = \inf_{\mathbf{u}} F(\mathbf{u}),$$

as long as there exists a minimizer of  $F$ . Hence, the split Bregman algorithm used here generates a minimizing sequence  $F(\mathbf{u}^k)$ , while the sequence  $\{\mathbf{u}^k\}$  itself may or may not converge. Even when the sequence  $\{\mathbf{u}^k\}$  itself converges to a minimizer (for instance when there exists a unique minimizer of  $F$  as shown in [21]), we can only stop at a finite number of iterations during actual computations. In other words, an  $\epsilon$ -optimal solution to  $F$  is what one obtains in practice, rather than an actual minimizer of  $F$ . Furthermore, since it is hard (and unnecessary) to find an actual minimizer of  $F$  numerically, an  $\epsilon$ -optimal solution to  $F$  is what we commonly seek when solving image restoration problems.

Next, we take a closer look at (5.8) in Algorithm 5.1 from a different angle. The first step can be viewed as finding an approximate solution of equation  $\mathbf{A} \mathbf{u} = \mathbf{f}$  with updated residues from the previous step, and the third step can be viewed as an update of residues. The second step can be viewed as a step to make the approximate solution obtained in the first step sparse by applying thresholding in the transform domain where the underlying solution has a sparse representation. The

step of thresholding is very important, since it preserves sparsity of the approximate solution of  $\mathbf{A}\mathbf{u} = \mathbf{f}$  at each iteration. We further remark that the key point of wavelet frame based image restorations is to find a good approximate solution of the equation  $\mathbf{A}\mathbf{u} = \mathbf{f}$  that is sparse in the transform domain, rather than a minimizer of an objective functional. The iterative procedure of (5.8) generates a minimizing sequence which leads us to an  $\epsilon$ -optimal solution to the objective functional. These are the central ideas that initiated the iterative algorithms with a built-in thresholding step at the beginning of the wavelet frame based image restoration (see, e.g., [36, 69]).

Furthermore, since the second step of (5.8) in Algorithm 5.1 keeps the large wavelet frame coefficients of the approximate solution of the current iteration and large wavelet frame coefficients reflect the singularities of the approximate solution, the step of thresholding sharpens the edges of the approximate solution by removing small coefficients. When large coefficients of a particular framelet are kept, it means that the corresponding difference operator applies to the approximate solution locally. Since different wavelet frame masks reflect different orders of difference operators, this iterative algorithm applies difference operators adaptively according to the singularities of the approximate solution. Hence, it can well preserve different types of edges at each iteration.

**5.2. Numerical Experiments.** In this section, we will perform some experiments using the analysis based approach (2.18) with Haar and piecewise linear framelets. The image restoration scenarios will be inpainting and deblurring. We will use Algorithm 5.1 to solve (2.18).

For the image inpainting problem, the operator  $A$  in (2.15) is a projection operator defined as

$$(\mathbf{A}\mathbf{u})[i, j] = \begin{cases} \mathbf{u}[i, j], & (i, j) \in \Lambda \\ 0, & \text{otherwise,} \end{cases}$$

where  $\Lambda$  is the domain where information of the image  $\mathbf{u}$  is known. We will assume that the observed image is not corrupted by noise for simplicity (i.e.,  $\eta = 0$  in (2.15)). However, the analysis based approach using framelets is very robust to noise as shown by various previous work (see, e.g., [34, 21]). Since there is not any noise in  $f$ , after we obtain a solution  $\mathbf{u}^*$  using Algorithm 5.1, we will replace the values of  $\mathbf{u}^*$  in  $\Lambda$  by those of  $\mathbf{f}$ , i.e., we let  $\mathbf{u}^*[i, j] = \mathbf{f}[i, j]$  for  $(i, j) \in \Lambda$ .

For the image deblurring problem, the operator  $\mathbf{A}$  in (2.15) is a convolution operator with kernel generated by some Gaussian function. In MATLAB, this kernel function  $\mathbf{g}$  is obtained by “ $\mathbf{g} = \text{fspecial}(\text{'gaussian'}, 15, 2)$ ”. Furthermore, the observed image  $\mathbf{f}$  is also corrupted by Gaussian white noise with  $\sigma = 3$ .

As we see from previous sections with various choices of the parameter  $\boldsymbol{\lambda}$ , the analysis based approach (2.18) corresponds to various variational methods. To be specific, we will focus on the following choices of  $\boldsymbol{\lambda}$  and levels of decomposition  $L$ :

- (I) Inpainting:  $L = 1$ ,  $p = 1$  or  $2$  and
  - (i) Haar framelet

$$(5.9) \quad \boldsymbol{\lambda} := \left(\frac{2\nu}{h}\right)^p \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad E_n^{(1)}(u) \approx \nu \int_{\Omega} (|u_x|^p + |u_y|^p)^{\frac{1}{p}} dx dy;$$

(ii) Piecewise linear framelet

$$(5.10) \quad \boldsymbol{\lambda} := \left( \frac{\sqrt{2}\nu}{h} \right)^p \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad E_n^{(1)}(u) \approx \nu \int_{\Omega} (|u_x|^p + |u_y|^p)^{\frac{1}{p}} dx dy;$$

$$(5.11) \quad \boldsymbol{\lambda} := \left( \frac{4\nu}{h^2} \right)^p \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad E_n^{(1)}(u) \approx \nu \int_{\Omega} \sqrt{|u_{xx}|^2 + |u_{yy}|^2} dx dy;$$

(II) Deblurring:  $L = 4$ ,  $p = 1$  or  $2$  and

(i) Haar framelet

$$(5.12) \quad (\lambda_{0,i})_i := \left( \frac{2\nu}{h} \right)^p \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad \lambda_{l,i} = \frac{1}{2^{lp}} \lambda_{0,i}, \quad E_n^{(1)}(u) \approx \nu \int_{\Omega} (|u_x|^p + |u_y|^p)^{\frac{1}{p}} dx dy;$$

(ii) Piecewise linear framelet

$$(5.13) \quad (\lambda_{0,i})_i := \left( \frac{\sqrt{2}\nu}{h} \right)^p \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \lambda_{l,i} = \frac{1}{2^{lp}} \lambda_{0,i}, \quad E_n^{(1)}(u) \approx \nu \int_{\Omega} (|u_x|^p + |u_y|^p)^{\frac{1}{p}} dx dy.$$

*Remark 5.2.* Here we note from (5.9) and (5.10) (as well as (5.12) and (5.13)) that by using the Haar framelet and the piecewise linear framelet and choosing a constant  $\boldsymbol{\lambda}$  across framelet bands, the regularization term  $E_n^{(1)}(u)$  always approximates the total variation of  $u$ . However, when different wavelet frame systems are used,  $E_n^{(1)}(u)$  corresponds to a rather different type of discretization of  $\int_{\Omega} |\nabla u| dx dy$ , which makes a big difference in terms of the qualities of the image restoration. On the other hand, for a fixed wavelet frame system (i.e. Haar or piecewise linear) and taking the constant  $\boldsymbol{\lambda}$  across both framelet bands and levels, when one chooses a different decomposition level,  $E_n^{(1)}(u)$  will also correspond to different type of discretization of  $\int_{\Omega} |\nabla u| dx dy$ , which will also improve the results of image recovery. According to our experience,  $L = 4$  is usually desirable for image inpainting and deblurring. This also shows that more flexibility is given when one takes a wavelet frame based approach directly. One can even go further to the general balanced approach if it is needed.

5.2.1. *Inpainting.* For image inpainting, we adopt the following stopping criteria:

$$\frac{\|\mathbf{A}\mathbf{u}^k - \mathbf{f}\|_2}{\|\mathbf{f}\|_2} < \epsilon,$$

with  $\epsilon = 5 \times 10^{-5}$  for  $\boldsymbol{\lambda}$  given by (5.9) and (5.10), and  $\epsilon = 8 \times 10^{-6}$  for  $\boldsymbol{\lambda}$  given by (5.11).

Figure 1 shows a comparison between  $p = 1$  (anisotropic) and  $p = 2$  (isotropic) for  $\boldsymbol{\lambda}$  given as in (5.9). It is clear that  $p = 2$  recovers the missing regions better than  $p = 1$  in terms of the geometry of the edges. Figure 2 and Figure 3 show the inpainting results using  $\boldsymbol{\lambda}$  given by (5.9), (5.10) and (5.11). As one can see, piecewise linear framelets perform better than the Haar framelet, and  $p = 2$  performs better than  $p = 1$  (except for the piecewise linear framelet where the two are comparable). It is also worth noticing that the algorithm converges faster for  $p = 2$  than for  $p = 1$ . This indicates that using  $p = 2$  is generally better than  $p = 1$  when one

considers both quality and efficiency. More interestingly, when one uses  $\lambda$  given in (5.11), the inpainting result is better than all other cases. For example, the regions pointed at by the dotted white arrows indicate that using  $\lambda$  given in (5.11), the corresponding model, which corresponds to a second order variational model, does a good job in filling information into large gaps.

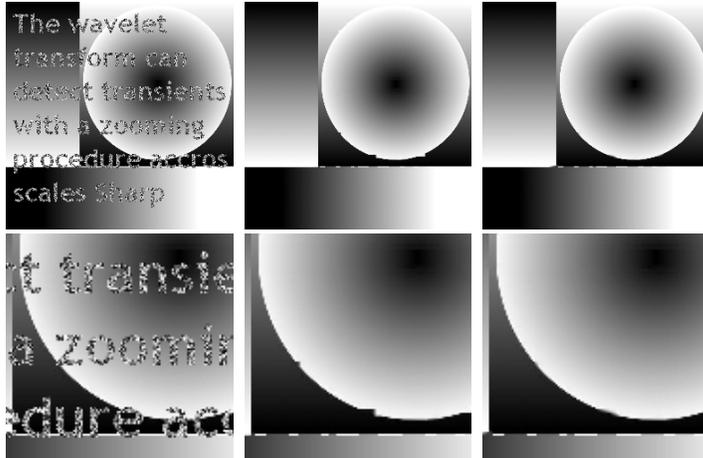


FIGURE 1. First row, from left to right, shows the observed image, inpainted image using Haar with  $p = 1$  (PSNR = 32.6267, Iterations = 535) and inpainted image using Haar with  $p = 2$  (PSNR = 34.534, Iterations = 395). The second row shows close-up views of the corresponding images in the first row.

5.2.2. *Deblurring.* For image deblurring, we adopt the following stopping criteria:

$$\frac{\|d^k - \mathbf{W}u^k\|_2}{\|f\|_2} < 10^{-4}.$$

Figure 4 shows comparisons between  $p = 1$  (anisotropic) and  $p = 2$  (isotropic) with  $\lambda$  given by (5.12) and (5.13). As one can see, the performance of the piecewise linear framelets with  $p = 2$  is better than all others in terms of both speed and quality. The result of the Haar framelet with  $p = 1$  shows clear block effect. Although such an effect was greatly reduced when we use  $p = 2$ , the recovered image still looks piecewise constant. The quality of the recovery is clearly better when piecewise linear framelets are used, and  $p = 2$  produces a sharper recovery than  $p = 1$ .

## 6. Extensions

The previous sections laid the foundations of the links between the wavelet frame based approach (2.18) and the differential operator based variational method (3.4). In this section, we will present some of the extensions of (2.18) and discuss their connections to some variational methods.

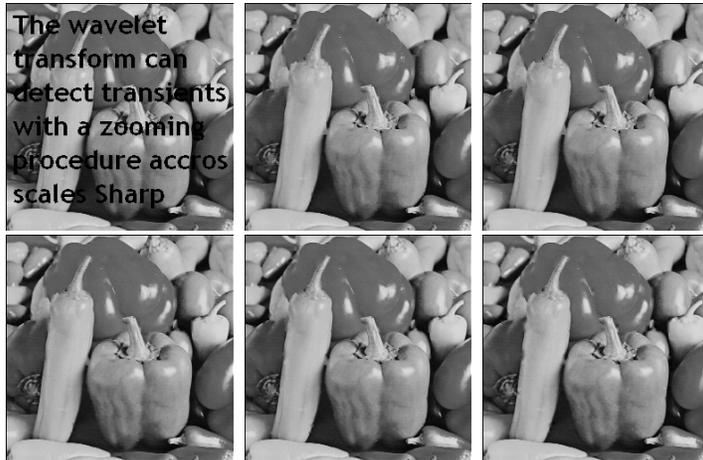


FIGURE 2. Image from left to right are: observed image, inpainted image using Haar with  $p = 1$  (PSNR = 32.6311, Iterations = 734), inpainted image using Haar with  $p = 2$  (PSNR = 32.766, Iterations = 502), inpainted image using piecewise linear with  $p = 1$  (PSNR = 33.8631, Iterations = 771), inpainted image using piecewise linear with  $p = 2$  (PSNR = 33.6978, Iterations = 370), and inpainted image using partial bands of piecewise linear with  $p = 2$  (PSNR = 34.3902, Iterations = 688).

**6.1. Two-System Wavelet Frame Based Models.** Real images can usually be regarded as being formed by two layers, namely cartoons (the piecewise smooth part of the image) and textures (the oscillating components of the image). Different layers usually have sparse approximations under different wavelet frame systems. Therefore, these two different layers should be considered separately. One natural idea is to use two wavelet frame systems that can sparsely represent cartoons and textures separately. The corresponding image restoration problem can be formulated as the following problem [21, 51, 52, 34]:

$$(6.1) \quad \inf_{\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{I}_2} \|\lambda_1 \cdot \mathbf{W}_1 \mathbf{u}_1\|_{1,p_1} + \|\lambda_2 \cdot \mathbf{W}_2 \mathbf{u}_2\|_{1,p_2} + \frac{1}{2} \|\mathbf{A}(\mathbf{u}_1 + \mathbf{u}_2) - \mathbf{f}\|_2^2,$$

where  $p_1, p_2$  are either 1 or 2,  $\mathbf{W}_1$  and  $\mathbf{W}_2$  denote two, possibly different, wavelet frame transforms, and  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are the two layers of image  $\mathbf{u}$  satisfying  $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ . The two-system model (6.1) was proposed and well studied by [21], which was shown to have excellent performance in restoration images with both cartoon and texture components.

The optimization problem (6.1) can be solved by the split Bregman algorithm. Similarly as in Section 5, we let  $\mathbf{d}_1 = \mathbf{W}_1 \mathbf{u}_1$  and  $\mathbf{d}_2 = \mathbf{W}_2 \mathbf{u}_2$ . Then the problem (6.1) is equivalent to

$$\inf_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{d}_1 = \mathbf{W}_1 \mathbf{u}_1, \mathbf{d}_2 = \mathbf{W}_2 \mathbf{u}_2} \|\lambda_1 \cdot \mathbf{d}_1\|_{1,p_1} + \|\lambda_2 \cdot \mathbf{d}_2\|_{1,p_2} + \frac{1}{2} \|\mathbf{A}(\mathbf{u}_1 + \mathbf{u}_2) - \mathbf{f}\|_2^2.$$

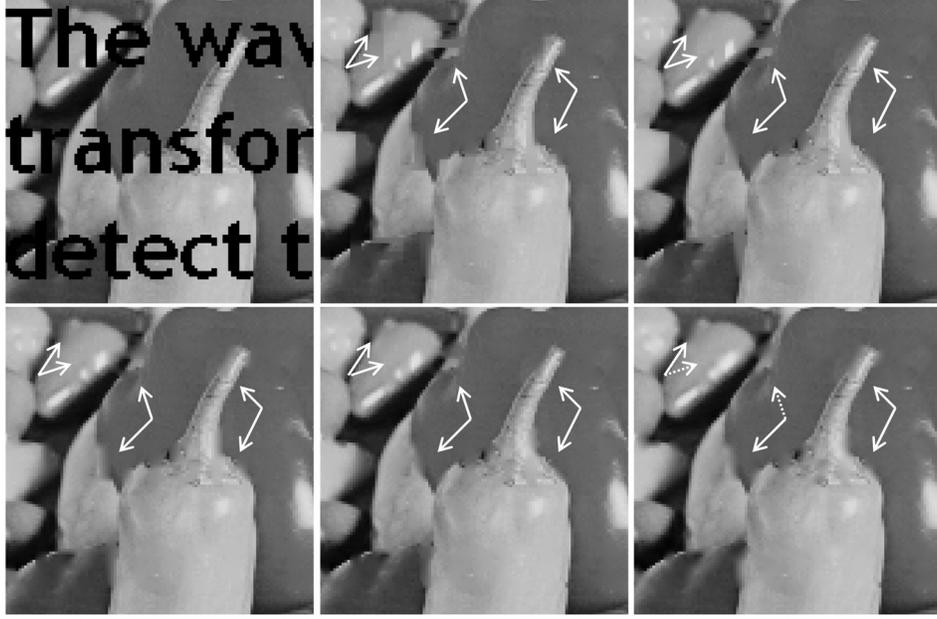


FIGURE 3. Images are the close-up views of those shown in Figure 2. The white arrows indicate the regions worth noticing, and the dotted white arrows indicate the regions where the partial bands of piecewise linear model with  $p = 2$  work exceptionally better than the others.

Then the corresponding Bregman iteration that solves the above problem can be written as

$$(6.2) \quad \begin{cases} \mathbf{u}_1^{k+1} = \arg \min_{\mathbf{u}_1} \frac{1}{2} \|\mathbf{A}(\mathbf{u}_1 + \mathbf{u}_2^k) - \mathbf{f}\|_2^2 + \frac{\mu_1}{2} \|\mathbf{W}_1 \mathbf{u}_1 - \mathbf{d}_1^k + \mathbf{b}_1^k\|_2^2, \\ \mathbf{u}_2^{k+1} = \arg \min_{\mathbf{u}_2} \frac{1}{2} \|\mathbf{A}(\mathbf{u}_1^{k+1} + \mathbf{u}_2) - \mathbf{f}\|_2^2 + \frac{\mu_2}{2} \|\mathbf{W}_2 \mathbf{u}_2 - \mathbf{d}_2^k + \mathbf{b}_2^k\|_2^2, \\ \mathbf{d}_1^{k+1} = \arg \min_{\mathbf{d}_1} \|\lambda_1 \cdot \mathbf{d}_1\|_{1,p_1} + \frac{\mu_1}{2} \|\mathbf{d}_1 - \mathbf{W}_1 \mathbf{u}_1^{k+1} - \mathbf{b}_1^k\|_2^2, \\ \mathbf{d}_2^{k+1} = \arg \min_{\mathbf{d}_2} \|\lambda_2 \cdot \mathbf{d}_2\|_{1,p_2} + \frac{\mu_2}{2} \|\mathbf{d}_2 - \mathbf{W}_2 \mathbf{u}_2^{k+1} - \mathbf{b}_2^k\|_2^2, \\ \mathbf{b}_1^{k+1} = \mathbf{b}_1^k + \delta_1 (\mathbf{W}_1 \mathbf{u}_1^{k+1} - \mathbf{d}_1^{k+1}), \\ \mathbf{b}_2^{k+1} = \mathbf{b}_2^k + \delta_2 (\mathbf{W}_2 \mathbf{u}_2^{k+1} - \mathbf{d}_2^{k+1}). \end{cases}$$

Note that the subproblems for  $u_1$  and  $u_2$  can be solved similarly as in (5.2) by inverting two linear systems, and the subproblems for  $d_1$  and  $d_2$  can be solved by anisotropic/isotropic shrinkage.

**6.2. Connections to the TGV model.** We start with the inf-convolution model (2.4) by rewriting  $\nabla^2$  as  $\nabla(\nabla)$ :

$$(6.3) \quad \inf_{u_1, u_2} \int_{\Omega} \nu_1 |\nabla u_1| + \nu_2 |\nabla(\nabla u_2)| dx dy + \frac{1}{2} \|A(u_1 + u_2) - f\|_{L_2(\Omega)}^2.$$

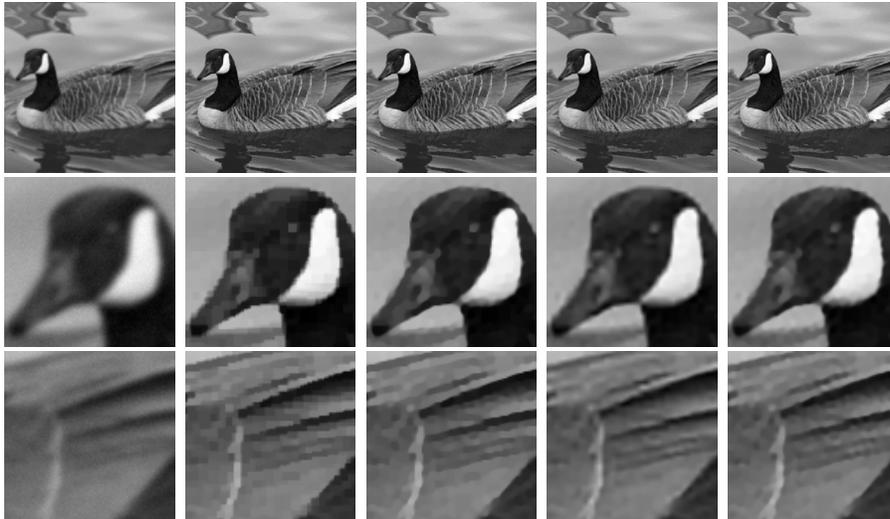


FIGURE 4. The first row shows the observed image, deblurred image using Haar with  $p = 1$  (PSNR = 28.9845, Iterations = 117), deblurred image using Haar with  $p = 2$  (PSNR = 29.4147, Iterations = 54), deblurred image using piecewise linear with  $p = 1$  (PSNR = 29.5616, Iterations = 237), and deblurred image using piecewise linear with  $p = 2$  (PSNR = 29.6375, Iterations = 38). The second and third rows show close-up views of the corresponding images in the first row.

If certain discretization is used for (6.3), then the discrete version of (6.3) coincides with the following Haar framelet based approach:

$$\inf_{\mathbf{u}_1, \mathbf{u}_2} \|\lambda_1 \cdot \mathbf{H}\mathbf{u}_1\|_{1,2} + \|\lambda_2 \cdot \mathbf{H}(\mathbf{H}\mathbf{u}_2)\|_{1,2} + \frac{1}{2} \|\mathbf{A}(\mathbf{u}_1 + \mathbf{u}_2) - \mathbf{f}\|_2^2.$$

Note that the transform  $\mathbf{H}(\mathbf{H}\mathbf{u})$  can be understood as a wavelet packet transform (without decimation) [70, 71, 72].

More generally, we can use the following wavelet frame based approach to approximate the inf-convolution model:

$$(6.4) \quad \inf_{\mathbf{u}_1, \mathbf{u}_2} \|\lambda_1 \cdot \mathbf{W}\mathbf{u}_1\|_{1,2} + \|\lambda_2 \cdot \mathbf{W}(\mathbf{W}\mathbf{u}_2)\|_{1,2} + \frac{1}{2} \|\mathbf{A}(\mathbf{u}_1 + \mathbf{u}_2) - \mathbf{f}\|_2^2.$$

With different choices of  $\lambda_i$  and  $\mathbf{W}$ , the wavelet frame based approach (6.4) provides various ways of solving the inf-convolution model (6.3) in the discrete setting. Furthermore, (6.4) is in fact a special case of the two-system model (6.1) in the previous section. Indeed, if we take  $\mathbf{W}_1 = \mathbf{W}$ ,  $\mathbf{W}_2 = \mathbf{W}(\mathbf{W})$  and  $p_1 = p_2 = 2$  for the two-system model (6.1), we obtain (6.4).

Therefore, with certain discretization for the inf-convolution model, it can be regarded as a special case of the two-system model (6.1). For example, if we choose  $\mathbf{W}_1$  to be one level of Haar framelet decomposition,  $\mathbf{W}_2$  to be one level of the

piecewise linear framelet decomposition and take

$$\boldsymbol{\lambda}_1 = \left(\frac{2\nu_1}{h}\right)^2 \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{\lambda}_2 = \left(\frac{4\nu_2}{h^2}\right)^2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & \gamma & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

then we will have that

$$\|\boldsymbol{\lambda}_1 \cdot \mathbf{W}_1 \mathbf{u}_1\|_{1,2} + \|\boldsymbol{\lambda}_2 \cdot \mathbf{W}_2 \mathbf{u}_2\|_{1,2}$$

approximate

$$(6.5) \quad \nu_1 \int_{\Omega} |\nabla u_1| dx dy + \nu_2 \int_{\Omega} |\nabla^2 u_2|_{\gamma} dx dy,$$

where  $\nabla^2 u_2$  denotes the Hessian of  $u_2$  and

$$|\nabla^2 u_2|_{\gamma} := \sqrt{|\partial_{xx} u_2|^2 + |\partial_{yy} u_2|^2 + \gamma |\partial_{xy} u_2|^2}.$$

The functional (6.5) with  $\gamma = 2$  is precisely the regularization term of the inf-convolution functional (2.4).

As we mentioned in Section 2.1 that the TGV model generalizes the inf-convolution model (6.3). Indeed, if we let  $w = u_2$  and  $u_1 = u - w$  in (6.3), the inf-convolution model can be rewritten as

$$\inf_{u,w} \int_{\Omega} \nu_1 |\nabla u - \nabla w| + \nu_2 |\nabla(\nabla w)| dx dy + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2.$$

Let  $v = \nabla w$ ; then the above problem is equivalent to

$$\inf_{v \in \text{Range}(\nabla), u} \int_{\Omega} \nu_1 |\nabla u - v| + \nu_2 |\nabla \cdot v| dx dy + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2.$$

If we relax the restriction  $v \in \text{Range}(\nabla)$  by letting  $v$  vary in the space of continuously differentiable 2-tensors, then we obtain the TGV model (2.5).

Following a similar idea, we shall derive a new wavelet frame based model from the two-system model (6.1). This model includes the TGV model (2.5) as a special case in the discrete setting. Letting  $\mathbf{v} = \mathbf{u}_2$  and  $\mathbf{u}_1 = \mathbf{u} - \mathbf{v}$ , we can rewrite (6.1) as

$$(6.6) \quad \inf_{\mathbf{u}, \mathbf{v} \in \mathcal{I}_2} \|\boldsymbol{\lambda}_1 \cdot \mathbf{W}_1(\mathbf{u} - \mathbf{v})\|_{1,p_1} + \|\boldsymbol{\lambda}_2 \cdot \mathbf{W}_2 \mathbf{v}\|_{1,p_2} + \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2.$$

Let  $\boldsymbol{\alpha} := \mathbf{W}_1 \mathbf{v}$  and hence  $\mathbf{v} = \mathbf{W}_1^{\top} \boldsymbol{\alpha}$ . Then (6.6) can be further written as

$$(6.7) \quad \inf_{\mathbf{u} \in \mathcal{I}_2, \boldsymbol{\alpha} \in \text{Range}(\mathbf{W}_1)} \|\boldsymbol{\lambda}_1 \cdot (\mathbf{W}_1 \mathbf{u} - \boldsymbol{\alpha})\|_{1,p_1} + \|\boldsymbol{\lambda}_2 \cdot \mathbf{W}_2 \mathbf{W}_1^{\top} \boldsymbol{\alpha}\|_{1,p_2} + \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2.$$

Note that (6.7) is still equivalent to (6.1). If we relax the restriction that  $\boldsymbol{\alpha} \in \text{Range}(\mathbf{W}_1)$  and replace  $\mathbf{W}_2 \mathbf{W}_1^{\top}$  in (6.7) by some general transformation, we then have the following optimization problem:

$$(6.8) \quad \inf_{\mathbf{u} \in \mathcal{I}_2, \boldsymbol{\alpha} \in \mathcal{S}} \|\boldsymbol{\lambda}_1 \cdot (\mathbf{W}\mathbf{u} - \boldsymbol{\alpha})\|_{1,p_1} + \|\boldsymbol{\lambda}_2 \cdot \mathbf{T}\boldsymbol{\alpha}\|_{*,1,p_2} + \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2.$$

The space  $\mathcal{S}$  that the variable  $\boldsymbol{\alpha}$  lives in has the same structure as  $\mathbf{W}\mathbf{u}$ . Therefore, the definition of  $\mathcal{S}$  will depend on the structure of  $\mathbf{W}$ . We now define the space  $\mathcal{S}$  that is associated to the transform  $\mathbf{W}$  as

$$\mathcal{S} := \mathcal{I}_2 \times \mathbb{R}^{J \times L} = \mathbb{R}^{N \times N \times J \times L},$$

where  $J$  is the total number of bands of  $\mathbf{W}$  for each level and  $L$  is the total level of decomposition of  $\mathbf{W}$ . The space  $\mathcal{S}$  is a collection of all 4D arrays  $\alpha$  with  $\alpha[\cdot, \cdot, j, l] \in \mathcal{I}_2$  for each  $1 \leq j \leq J$  and  $1 \leq l \leq L$ . The operator  $\mathbf{T}$  is defined by

$$\mathbf{T}\alpha := \{\mathbf{T}_{j,l}\alpha[\cdot, \cdot, j, l] : 1 \leq j \leq J, 1 \leq l \leq L\},$$

where  $\mathbf{T}_{j,l}$  corresponds to the framelet decomposition associated to a certain wavelet frame system which is possibly different for different  $j$  and  $l$ . The norm  $\|\lambda \cdot \mathbf{T}\alpha\|_{*,1,p}$  is defined as

$$\|\lambda \cdot \mathbf{T}\alpha\|_{*,1,p} := \sum_{l=1}^L \left( \sum_{j=1}^J \|\lambda_{j,l} \cdot \mathbf{T}_{j,l}\alpha[\cdot, \cdot, j, l]\|_{1,p}^p \right)^{1/p}.$$

Clearly  $\mathbf{T}$  itself is the wavelet frame decomposition corresponding to some wavelet frame system, and we have  $\mathbf{T}^\top \mathbf{T} = I$ . In fact,  $\mathbf{T}$  can also be interpreted as the wavelet frame packet decomposition [72], which resembles the well known wavelet packet decomposition [70, 71]. Furthermore, the new model (6.8) is generally different from (6.1) because the range of  $\mathbf{W}$  is strictly contained in  $\mathcal{S}$  unless  $\mathbf{W}$  is unitary (i.e., the corresponding wavelet system is orthonormal).

In particular, if one takes  $\mathbf{T}_j = \mathbf{W}$  and  $\mathbf{W}$  to be the Haar framelet decomposition in (6.8), and if  $\lambda_1$  and  $\lambda_2$  are properly chosen, then the regularization term of (6.8), i.e.,

$$F(\mathbf{u}) := \min_{\alpha \in \mathcal{S}} \|\lambda_1 \cdot (\mathbf{W}\mathbf{u} - \alpha)\|_{1,p_1} + \|\lambda_2 \cdot \mathbf{T}\alpha\|_{*,1,p_2}$$

approximates

$$\inf_v \nu_1 \|\nabla u - v\|_{L_1} + \nu_2 \|\nabla \cdot v\|_{L_1},$$

where the variable  $v = (v_1, v_2)$  varies in the space of all continuously differential 2-tensors. Therefore, with properly chosen parameters  $\lambda_1$  and  $\lambda_2$ , model (6.8) approximates the TGV model (2.5).

The optimization problem (6.8) can be solved by the split Bregman algorithm. We let  $\mathbf{d}_1 = \mathbf{W}\mathbf{u} - \alpha$  and  $\mathbf{d}_2 = \mathbf{T}\alpha$ , and it can then be written as

$$\inf_{\mathbf{u}, \alpha, \mathbf{d}_1 = \mathbf{W}\mathbf{u} - \alpha, \mathbf{d}_2 = \mathbf{T}\alpha} \|\lambda_1 \cdot \mathbf{d}_1\|_{1,p_1} + \|\lambda_2 \cdot \mathbf{d}_2\|_{*,1,p_2} + \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2.$$

Then the corresponding Bregman iteration that solves the above problem can be written as

$$(6.9) \quad \begin{cases} \mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{f}\|_2^2 + \frac{\mu_1}{2} \|\mathbf{W}\mathbf{u} - \alpha^k - \mathbf{d}_1^k + \mathbf{b}_1^k\|_2^2, \\ \alpha^{k+1} = \arg \min_{\alpha} \frac{\mu_1}{2} \|\mathbf{W}\mathbf{u}^{k+1} - \alpha - \mathbf{d}_1^k + \mathbf{b}_1^k\|_2^2 + \frac{\mu_2}{2} \|\mathbf{T}\alpha - \mathbf{d}_2^k + \mathbf{b}_2^k\|_2^2, \\ \mathbf{d}_1^{k+1} = \arg \min_{\mathbf{d}_1} \|\lambda_1 \cdot \mathbf{d}_1\|_{1,p_1} + \frac{\mu_1}{2} \|\mathbf{W}\mathbf{u}^{k+1} - \alpha^{k+1} - \mathbf{d}_1 + \mathbf{b}_1^k\|_2^2, \\ \mathbf{d}_2^{k+1} = \arg \min_{\mathbf{d}_2} \|\lambda_2 \cdot \mathbf{d}_2\|_{*,1,p_2} + \frac{\mu_2}{2} \|\mathbf{T}\alpha^{k+1} - \mathbf{d}_2 + \mathbf{b}_2^k\|_2^2, \\ \mathbf{b}_1^{k+1} = \mathbf{b}_1^k + \delta_1 (\mathbf{W}\mathbf{u}^{k+1} - \alpha^{k+1} - \mathbf{d}_1^{k+1}), \\ \mathbf{b}_2^{k+1} = \mathbf{b}_2^k + \delta_2 (\mathbf{T}\alpha^{k+1} - \mathbf{d}_2^{k+1}). \end{cases}$$

Similar to (6.2), the subproblems for  $\mathbf{u}$  and  $\alpha$  in (6.9) can be solved by inverting two linear systems, and the subproblems for  $\mathbf{d}_1$  and  $\mathbf{d}_2$  can be solved by anisotropic/isotropic shrinkage.

**6.3. Numerical Experiments.** We now present some numerical results of the two wavelet frame based approaches (6.1) and (6.8) using algorithms (6.2) and (6.9). Since (6.4) is a special case of (6.1), we forgo the numerical simulations for this case.

We apply the algorithms (6.2) to the same deblurring problem as given in Section 5.2.2 with the same blurring kernel and noise level. We shall take  $p_i = 2$  for  $i = 1, 2$  and adopt the following stopping criteria:

$$\frac{\sqrt{\|d_1^k - \mathbf{W}_1 \mathbf{u}_1^k\|_2^2 + \|d_2^k - \mathbf{W}_2 \mathbf{u}_2^k\|_2^2}}{\|\mathbf{f}\|_2} < 10^{-4}.$$

We take the following two combinations of  $\mathbf{W}_1$  and  $\mathbf{W}_2$  where 4 levels of decomposition are conducted for both  $\mathbf{W}_1$  and  $\mathbf{W}_2$ :

- (1) Haar framelet for  $\mathbf{W}_1$ , and piecewise linear framelets for  $\mathbf{W}_2$  (denoted as “Haar-Linear”);
- (2) Haar framelet for  $\mathbf{W}_1$ , and piecewise cubic framelets for  $\mathbf{W}_2$  (denoted as “Haar-Cubic”).

The reason that we are using Haar for  $\mathbf{W}_1$  is that we expect  $\mathbf{u}_1$  to contain only the cartoon component of the image. We use higher order framelets for  $\mathbf{W}_2$  in order to properly model texture components  $\mathbf{u}_2$ . The results are shown in Figure 5 (Haar-Linear) and Figure 6 (Haar-Cubic). It is worth noticing that both Haar-Linear and Haar-Cubic (where the latter produces a slightly better recovery than the former) produce better results than all the results of using only the single wavelet frame system given in Figure 4.

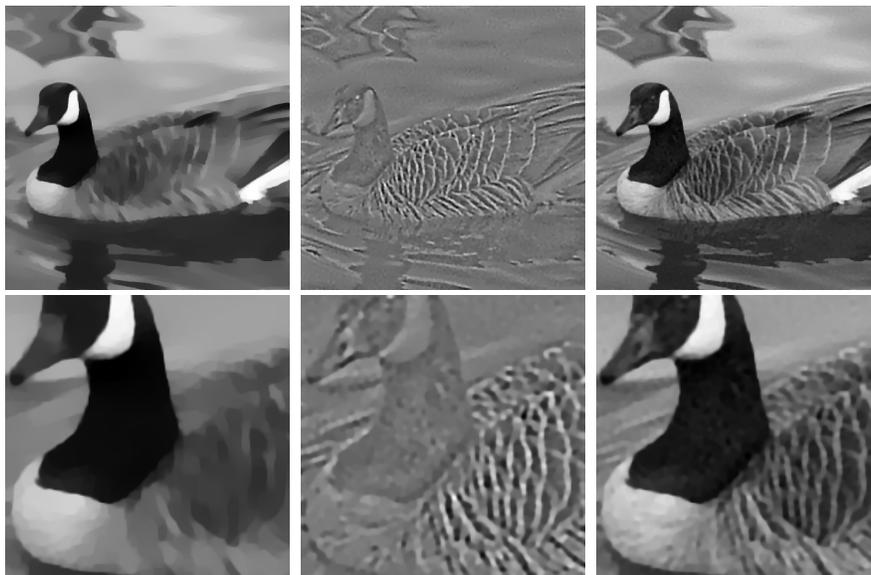


FIGURE 5. Results for Haar-Linear, where the images from left to right are the cartoon component  $\mathbf{u}_1$ , the texture component  $\mathbf{u}_2$  and the recovered image  $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ ; the images in the second row are zoom-in views of the those in the first row. PSNR= 29.7082 and total number of iterations is 37.

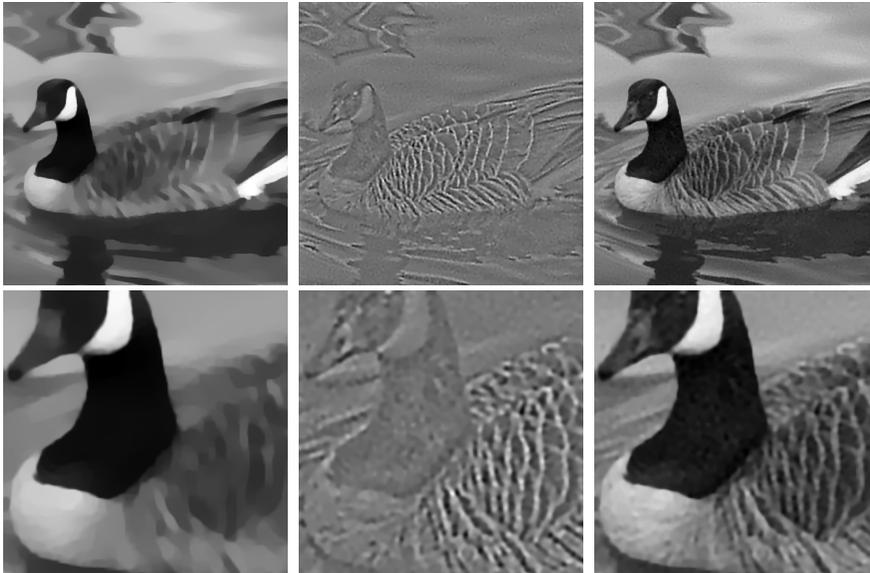


FIGURE 6. Results for Haar-Cubic, where the images from left to right are the cartoon component  $\mathbf{u}_1$ , the texture component  $\mathbf{u}_2$  and the recovered image  $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ ; the images in the second row are zoom-in views of the those in the first row. PSNR= 29.7218 and total number of iterations is 31.

We now apply the algorithms (6.9) to the same example as we did for algorithm (6.2). We pick  $\mathbf{W}$  to be piecewise linear framelet decomposition (4 levels) and  $\mathbf{T}_j$  to be Haar framelet decomposition (1 level). We shall take  $p_i = 2$  for  $i = 1, 2$  and adopt the following stopping criteria:

$$\frac{\sqrt{\|\mathbf{d}_1^k - \mathbf{W}\mathbf{u}^k\|_2^2 + \|\mathbf{d}_2^k - \mathbf{T}\boldsymbol{\alpha}^k\|_2^2}}{\|\mathbf{f}\|_2} < 2 \times 10^{-5}.$$

A numerical result is given in Figure 7.

Judging from the PSNR values given by Figures 4, 5, 6 and 7, the two-system model (6.1) gives the best performances for image deblurring problems. Also note that (6.8) performs better than the analysis based approach (2.18), while not as good as the two-systems model (6.1).

REFERENCES

[1] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Phys. D*, vol. 60, pp. 259–268, 1992.  
 [2] 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.  
 [3] I. Ekeland and R. Temam, *Convex analysis and variational problems*. Classics in Applied Mathematics, Society for Industrial Mathematics, 1999.  
 [4] G. Dal Maso, *Introduction to  $\Gamma$ -convergence*. Birkhauser, 1993.  
 [5] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*. American Mathematical Society, 2001.  
 [6] B. Dong, A. Chien, and Z. Shen, “Frame based segmentation for medical images,” *Communications in Mathematical Sciences*, vol. 9(2), pp. 551–559, 2010.



FIGURE 7. Results of algorithm (6.9). The first row shows the recovered image  $\mathbf{u}^*$  and  $\mathbf{W}^\top \boldsymbol{\alpha}^*$ , where  $\mathbf{u}^*$  and  $\boldsymbol{\alpha}^*$  are the stationary points of algorithm (6.9). The second row shows zoom-in views of the first row. PSNR= 29.6419 and total number of iterations is 38.

- [7] B. Dong and Z. Shen, “Frame based surface reconstruction from unorganized points,” *Journal of Computational Physics*, vol. 230, pp. 8247–8255, 2011.
- [8] G. Sapiro, *Geometric partial differential equations and image analysis*. Cambridge University Press, 2001.
- [9] S. Osher and R. Fedkiw, *Level set methods and dynamic implicit surfaces*. Springer, 2003.
- [10] A. Marquina and S. Osher, “Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal,” *SIAM Journal on Scientific Computing*, vol. 22, no. 4, pp. 387–405, 2000.
- [11] C. Vogel and M. Oman, “Iterative methods for total variation denoising,” *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, 1996.
- [12] T. Chan, G. Golub, and P. Mulet, “A nonlinear primal-dual method for total variation-based image restoration,” *SIAM Journal on Scientific Computing*, vol. 20, no. 6, pp. 1964–1977, 1999.
- [13] A. Chambolle, “An algorithm for total variation minimization and applications,” *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 89–97, 2004.
- [14] M. Zhu and T. Chan, “An efficient primal-dual hybrid gradient algorithm for total variation image restoration,” *Mathematics Department, UCLA, CAM Report*, pp. 08–34, 2007.
- [15] M. Zhu, S. Wright, and T. Chan, “Duality-based algorithms for total variation image restoration,” *Mathematics Department, UCLA, CAM Report*, pp. 08–33, 2008.
- [16] M. Zhu, S. Wright, and T. Chan, “Duality-based algorithms for total-variation-regularized image restoration,” *Computational Optimization and Applications*, pp. 1–24, 2008.
- [17] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, “An algorithm for minimizing the Mumford-Shah functional,” in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 1133–1140, IEEE, 2010.
- [18] E. Esser, X. Zhang, and T. Chan, “A general framework for a class of first order primal-dual algorithms for TV minimization,” *UCLA CAM Report*, pp. 09–67, 2009.
- [19] 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.

- [20] 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.
- [21] 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.
- [22] S. Setzer, "Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage," *Scale Space and Variational Methods in Computer Vision*, pp. 464–476, 2009.
- [23] E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split Bregman," *CAM Report*, vol. 9-31, 2009.
- [24] W. Ring, "Structural properties of solutions to total variation regularization problems," *Mathematical modelling and numerical analysis*, vol. 34, no. 4, pp. 799–810, 2000.
- [25] M. Nikolova, "Local strong homogeneity of a regularized estimator," *SIAM Journal on Applied Mathematics*, vol. 61, no. 2, pp. 633–658, 2000.
- [26] A. Chambolle and P. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.
- [27] T. Chan, S. Esedoglu, and F. Park, "Image decomposition combining staircase reduction and texture extraction," *Journal of Visual Communication and Image Representation*, vol. 18, no. 6, pp. 464–486, 2007.
- [28] T. Chan, S. Esedoglu, and F. Park, "A fourth order dual method for staircase reduction in texture extraction and image restoration problems," *UCLA CAM Report*, pp. 05–28, 2005.
- [29] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 503–516, 2001.
- [30] K. Bredies, K. Kunisch, and T. Pock, "Total Generalized Variation," *SIAM Journal on Imaging Sciences*, vol. 3, p. 492, 2010.
- [31] I. Daubechies, *Ten lectures on wavelets*, vol. CBMS-NSF Lecture Notes, SIAM, nr. 61. Society for Industrial and Applied Mathematics, 1992.
- [32] 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.
- [33] Z. Shen, "Wavelet frames and image restorations," *Proceedings of the International Congress of Mathematicians, Hyderabad, India*, 2010.
- [34] 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.
- [35] A. Chai and Z. Shen, "Deconvolution: A wavelet frame approach," *Numerische Mathematik*, vol. 106, no. 4, pp. 529–587, 2007.
- [36] 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.
- [37] R. Chan, S. Riemenschneider, L. Shen, and Z. Shen, "Tight frame: an efficient way for high-resolution image reconstruction," *Applied and Computational Harmonic Analysis*, vol. 17, no. 1, pp. 91–115, 2004.
- [38] J. Cai, R. Chan, L. Shen, and Z. Shen, "Restoration of chopped and noded images by framelets," *SIAM J. Sci. Comput.*, vol. 30, no. 3, pp. 1205–1227, 2008.
- [39] J. Cai, R. Chan, L. Shen, and Z. Shen, "Convergence analysis of tight framelet approach for missing data recovery," *Advances in Computational Mathematics*, pp. 1–27, 2008.
- [40] 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.
- [41] 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.
- [42] J. Cai and Z. Shen, "Framelet based deconvolution," *J. Comp. Math*, vol. 28, no. 3, pp. 289–308, 2010.
- [43] R. Chan, Z. Shen, and T. Xia, "A framelet algorithm for enhancing video stills," *Applied and Computational Harmonic Analysis*, vol. 23, no. 2, pp. 153–170, 2007.
- [44] S. Mallat, *A wavelet tour of signal processing*, vol. 2nd ed. New York: Academic. Academic press, 1999.
- [45] B. Han and Z. Shen, "Dual wavelet frames and Riesz bases in Sobolev spaces," *Constructive Approximation*, vol. 29, no. 3, pp. 369–406, 2009.

- [46] 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.
- [47] M. Fadili and J. Starck, "Sparse representations and Bayesian image inpainting," *Proc. SPARS*, vol. 5, 2005.
- [48] M. Fadili, J. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal*, vol. 52, no. 1, p. 64, 2009.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] 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.
- [53] R. Jia, "Spline wavelets on the interval with homogeneous boundary conditions," *Advances in Computational Mathematics*, vol. 30, no. 2, pp. 177–200, 2009.
- [54] R. Adams, *Sobolev Spaces*. Academic Press, New York, 1975.
- [55] A. Cohen, I. Daubechies, and J. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 45, no. 5, pp. 485–560, 1992.
- [56] Z. Shen and Z. Xu, "On b-spline framelets derived from the unitary extension principle," *Arxiv preprint arXiv:1112.5771*, 2011.
- [57] R. Jia, "Approximation with scaled shift-invariant spaces by means of quasi-projection operators," *Journal of Approximation Theory*, vol. 131, no. 1, pp. 30–46, 2004.
- [58] R. Jia and C. Micchelli, "Using the refinement equations for the construction of pre-wavelets II: Powers of two," *Curves and surfaces*, pp. 209–246, 1991.
- [59] G. Folland, *Real analysis*. Wiley New York, 1984.
- [60] 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.
- [61] 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.
- [62] 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.
- [63] X. Tai and C. Wu, "Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model," *Scale Space and Variational Methods in Computer Vision*, pp. 502–513, 2009.
- [64] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*. Society for Industrial and Applied Mathematics, 1989.
- [65] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [66] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2006.
- [67] P. Mrázek and J. Weickert, "Rotationally invariant wavelet shrinkage," *Pattern Recognition*, pp. 156–163, 2003.
- [68] Y. Wang, W. Yin, and Y. Zhang, "A fast algorithm for image deblurring with total variation regularization," *Rice University CAAM Technical Report TR07-10*, 2007.
- [69] R. Chan, S. Riemenschneider, L. Shen, and Z. Shen, "High-resolution image reconstruction with displacement errors: A framelet approach," *International Journal of Imaging Systems and Technology*, vol. 14, no. 3, pp. 91–104, 2004.
- [70] R. Coifman, Y. Meyer, and V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and Their Applications*, p. 153, Jones & Bartlett, Boston, MA, 1992.

- [71] R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 713–718, 2002.
- [72] S. Pan, "Tight wavelet frame packets," *Ph.D. thesis at National University of Singapore*, 2011.

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF IOWA, IOWA CITY, IA, 52242-1419.  
*E-mail address:* `jianfeng-cai@uiowa.edu`

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF ARIZONA, 617 NORTH SANTA RITA AVENUE, TUCSON, AZ, USA, 85721-0089.  
*E-mail address:* `dongbin@math.arizona.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, 405 HILGARD AVENUE, LOS ANGELES, CA, 90095-1555.  
*E-mail address:* `sjo@math.ucla.edu`

DEPARTMENT OF MATHEMATICS, NATIONAL UNIVERSITY OF SINGAPORE, BLOCK S17, 10 LOWER KENT RIDGE ROAD, SINGAPORE, 119076.  
*E-mail address:* `matzuows@nus.edu.sg`