Image Restoration: Wavelet Frame Shrinkage, Nonlinear Evolution PDEs, and Beyond

Bin Dong[a], Qingtang Jiang[b], and Zuowei Shen[c]

[a] *Department of Mathematics, University of Arizona, Tucson, AZ 85721, U.S.A.*
[b] *Department of Mathematics and Computer Science,*
University of Missouri–St. Louis, St. Louis, MO 63121, U.S.A.
[c] *Department of Mathematics, National University of Singapore*
10 Lower Kent Ridge Road, Singapore, 119076

Abstract

In the past few decades, mathematics based approaches have been widely adopted in various image restoration problems, among which the partial differential equation (PDE) based approach (e.g. the total variation model [56] and its generalizations, nonlinear diffusions [15, 52], etc.), and wavelet frame based approach are some of successful examples. These approaches were developed through different paths and generally provided understandings from different angles of the same problem. As shown in numerical simulations, implementations of wavelet frame based approach and PDE based approach quite often end up with solving a similar numerical problem with similar numerical behaviors, even though different approaches have advantages in different applications. Since wavelet frame based and PDE based approaches have all been modeling the same type of problems with success, it is natural to ask whether wavelet frame based approach is fundamentally connected with PDE based approach when we trace all the way back to their roots. A fundamental connection of a wavelet frame based approach with total variation model and its generalizations were established in [8]. This connection gives wavelet frame based approach a geometric explanation and, at the same time, it equips a PDE based approach with a time frequency analysis. It was shown in [8] that a special type of wavelet frame model using generic wavelet frame systems can be regarded as an approximation of a generic variational model (with the total variation model as a special case) in the discrete setting. A systematic convergence analysis, as the resolution of the image goes to the infinity, which is the key step to link the two approaches, is also given in [8]. Motivated by [8] and [47], this paper is to establish a fundamental connection between wavelet frame based approach and nonlinear evolution PDEs, provide interpretations and analytical studies of such connections, and propose new algorithms for image restoration based on the new understandings. Together with the results in [8], we now have a better picture of how wavelet frame based approach can be used to interpret general PDE based approach (e.g. the variational models or nonlinear evolution PDEs) and can be used as a new and useful tool in numerical analysis to discretize and solve various variational and PDE models. To be more precise on our contributions, we shall establish that: (1) The connections between wavelet frame shrinkage and nonlinear evolution PDEs provide new and inspiring interpretations of both approaches that enable us to derive new PDE models and (better) wavelet frame shrinkage algorithms for image restoration. (2) A generic nonlinear evolution PDEs (of parabolic or hyperbolic type) can be approximated by wavelet frame shrinkage with properly chosen wavelet frame systems and carefully designed shrinkage functions. (3) The main idea of this work is beyond the scope of image restoration. Our analysis and discussions indicate that wavelet frame shrinkage is a new way of solving PDEs in general, which will provide a new insight that will enrich the existing theory and applications of numerical PDEs, as well as, those of wavelet frames.

Keywords and phrases. *Image restoration, nonlinear diffusion, wavelet frames, wavelet frame shrinkage.*

1 Introduction

Image restoration, including image denoising, deblurring, inpainting, computed tomography, etc., is one of the most important areas in imaging science. 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 enable us to see crucial but subtle objects that reside in the image. Mathematics has become one of the main driving forces of the modern development of image restoration. There are several mathematics based approaches, the partial differential equation (PDE) based approach (e.g. the total variation method, nonlinear evolution PDEs), and wavelet frame based approach developed in the last few decades are successful examples among many.

One of the most commonly used nonlinear PDE based approach for image restoration, especially the classical problem of image denoising, is the nonlinear diffusion. Since the introduction of the 2nd-order nonlinear diffusion by Perona and Malik (PM) in 1990 [52], a variety of 2nd-order nonlinear diffusions have been proposed (see e.g. [15, 25, 67] and the references therein). The fourth order nonlinear diffusion was proposed in [68, 69] to resolve the blocky effects that PM diffusion and its variants tend to produce in image denoising. Later, the fourth order nonlinear diffusion has also been studied in [49], and high order diffusion with an edge enhancing functional was proposed in [62]. The theoretical properties of high order diffusion have been studied in [26]. Other than nonlinear diffusions, nonlinear hyperbolic equations, such as shock filters [51], were also used for image restoration. What these PDE models for image restoration have in common is the seek of a good balance between the two seemingly contradictory objectives: smoothness at locations where noise or other artifacts have been removed; and preservation or even enhancement of the sharpness of edges, corners, etc., which are singularities.

Wavelet frame based methods are generally considered as a different approach from the nonlinear PDE based methods, and they were developed along a fairly different path. The wavelet frame based image processing started from [19, 20] for high-resolution image reconstructions, where an iterative algorithm by applying thresholding to wavelet frame coefficients at each iteration to preserve sharp edges of images was proposed. In order to gain more flexibility, in [10, 12], the authors introduced an additional weighting and proposed the model now known as the balanced model. It was shown by [7] that the algorithm of [19, 20] converges to a solution of a special case of the balanced model. The balance algorithm has been applied to various applications in [11, 14, 16, 17]. The model proposed by [10, 12] is called the balanced model, since it balances the sparsity of the wavelet frame coefficient and the smoothness of the restored image. This includes two other wavelet frame based models as special cases. One is known as the synthesis based model [24, 36, 37, 39, 40], where the sparsity term in the balanced model is emphasized. The other is known as the analysis based model, [13, 34, 60], where the smoothness of the restored image is emphasized. The three approaches are different from each other, unless the underlying wavelet frame systems is in fact orthonormal/biorthogonal. However, what they have in common, is the penalization of the sparsity of the wavelet frame coefficients of the image to be restored.

Ever since the early 90's, numerous image restoration models and algorithms based on PDEs and wavelets (wavelet frames) were proposed and studied in the literature. Many of them are rather successful in accurate modeling of given image restoration problems. These approaches were developed through different paths and generally provided understandings from different angles of the same problem. As shown in many numerical simulations, implementations of wavelet based approach and PDE based approach quite often end up with solving a similar numerical problem and their numerical behaviors are often comparable, although different approaches have advantages in different applications. Since all these different approaches are modeling the same type of problem with success, it is natural to ask whether wavelet frame based approach is fundamentally connected with PDE methods when we trace all the way back to their roots. A fundamental connection between a wavelet frame based approach and a general variation model (with total variation model [56] as a special case) were established in [8]. It was shown in [8] that a special type of wavelet frame model using generic wavelet frame systems can be regarded as an approximation of the variational model in the discrete setting.

Motivated by [8] and [47], this paper is to establish a fundamental connection between wavelet frame shrinkage and nonlinear evolution PDEs, provide interpretations and analytical studies of such connection, and propose new algorithms for image restoration based on the new understandings. This connection automatically gives a wavelet frame approach a geometric explanation through nonlinear PDEs and, at the time, it equips the PDE based approach with a time frequency analysis by the nature of the two approaches. Together with the results in [8], we now have a better picture of how wavelet frame based approach can be used to interpret general PDE based approach and how can it be used as a new and useful tool in numerical analysis to discretize and solve various variational and PDE models.

Some earlier results in [64–66] showed the correspondence between Haar wavelet shrinkage and the 2nd-order nonlinear diffusions. This work was recently generalized to the wavelet frame shrinkage and higher order nonlinear diffusions for the 1-dimensional case by [47]. However, it is not clear how can the 2-D wavelet frame shrinkage be related to nonlinear PDEs? Can we theoretically justify such connection? Furthermore, can we observe something new and something we could not see without establishing a relation between them? These important questions are yet to be answered. As we will see that this fundamental connection between wavelet based shrinkage and nonlinear evolution PDEs provides answers to all these questions.

The key idea of the wavelet frame based approach is to apply shrinkage (especially soft-thresholding) to wavelet frame coefficients iteratively, so that it converges to an optimal solution of some objective functional. The bridge between the wavelet frame based approach and PDE based approach is established when we see that a proper choice of numerical scheme of solving nonlinear evolution PDEs can be viewed as applying a properly chosen shrinkage operator on wavelet frame coefficients iteratively. Corresponding optimal properties are also discussed. Furthermore, the asymptotic analysis, i.e. when the image resolution goes to infinity, will be given for a few important cases. The connection between the wavelet frame based and PDE based approaches become clear once we have shown the followings in the paper.

Firstly, we show that we can approximate a generic nonlinear evolution PDE (of parabolic or hyperbolic type) using iterative (discrete) wavelet frame shrinkage by properly choosing the underlying wavelet frame systems and carefully designing the associated shrinkage operator. Such nonlinear evolution PDE includes the nonlinear parabolic equation called Perona-Malik equation [52], the nonlinear hyperbolic equation known as the Osher-Rudin's shock filter [51], and many others as examples. Our key observations are the connections between the (discrete) wavelet frame decomposition and differential operators (that were observed in [8]); and between the (discrete) wavelet frame reconstruction and divergence operators. The approximations by wavelet frame transforms is fundamentally different from the widely used finite difference approximation and the well-known wavelet Galerkin methods, in that the underlying solution and its derivatives are sampled differently in different function spaces. We will elaborate what exactly are the samplings corresponding to wavelet frame transforms, how they are related to the existing methods and why they are superior. Our arguments are also supported by our numerical experiments for image restoration, in which the advantage of approximating nonlinear diffusion equations using wavelet frames shrinkage over some standard finite difference discretization is presented. Various optimal properties in wavelet frame domain for these numerical schemes are discussed. Furthermore, we provide a rigorous convergence analysis of the discretization by wavelet frame shrinkage. We prove that, for a certain quasilinear parabolic equation, the associated iterative wavelet frame shrinkage algorithm does converge to the solution of the PDE as meshsize goes to zero. The given convergence analysis can be generalized to other well-posed nonlinear evolution PDEs under suitable conditions.

Secondly and more importantly, the connections between wavelet frame shrinkage and nonlinear evolution PDEs provide new and inspiring interpretations of both approaches. On one hand, the optimality property of the wavelet frame shrinkage shed lights on that of the PDEs'. In addition, some of the wavelet frame shrinkage algorithms that are commonly used in image restoration, such as the iterative soft-thresholding algorithms, lead to new types of nonlinear PDEs that have not been considered in the literature. In particular, one of these PDEs can be regarded as a regularized version of the well-known mean curvature flow. On the other hand, the nonlinear PDE based approach also provides new insights into the desirable choices of adaptive thresholds for wavelet frame shrinkage, which has not yet been systematically studied. In particular, the idea of anisotropy of the Perona-Malik equation can be used to create an adaptive wavelet frame shrinkage algorithm that outperform the traditional wavelet frame shrinkage algorithms. As shown by the numerical experiments of this paper, the performance of some of the new iterative wavelet frame shrinkage algorithms inspired by our theoretical studies are generally better than some existing iterative wavelet frame shrinkage algorithm that is currently widely used in image restoration.

Finally, although we will mostly focus on PDE models and wavelet frame shrinkages for image restoration, the significance of our findings is beyond what it may appear. In fact, our analysis and discussions in this paper already indicate that wavelet frame based approach is a new way of solving PDEs in general. We believe that the advantage of wavelet frame based approach over existing methods is not limited to image restoration. PDE is one of the most powerful tools modeling the physical world. Finding numerical solutions of PDEs has always been in the heart of numerical analysis. For different types of PDEs arise in different applications, the quality measures of the solutions may be different, and it is very hard to predict if using wavelet frame based approach can outperform conventional methods. However, given the vast collection of wavelet frame systems with a variety of desirable properties suitable to approximate functions living in different function spaces, we think wavelet frame based approach will at least provide a new school of thoughts as complement to the existing theory and applications of numerical PDEs.

The rest of the paper is organized as follows. We start with a brief introduction of wavelet frames and fast wavelet frame transforms in Section 2. In the same section, we also collect several examples of wavelet frame filters that will be used in later sections. In addition, we will introduce a general formula for iterative wavelet frame shrinkage and discuss its optimality properties when different types of shrinkage are used. Then, we discuss how wavelet frame shrinkage algorithms can be regarded as a discrete approximation to nonlinear evolution PDEs in a rather general setting. Differences between our approach and some existing numerical methods for PDEs, such as finite difference methods and wavelet Galerkin methods, are given at the end of Section 2. In Section 3, we start with generic derivations for the correspondence of wavelet frame shrinkage to nonlinear diffusion equations. We will show how the commonly used PDE models for image restoration can be derived from iterative shrinkage of wavelet frame coefficients. In Section 4, we present some new high-order diffusion equations that correspond to the B-spline tight wavelet frame systems which are commonly used in image restoration. In the same section, we also study the rotation-invariant high-order diffusion equation and its associated frame filter banks. In Section 5, we show that some of the iterative wavelet frame shrinkage algorithms, i.e. iterative soft-thresholding, commonly used in image restoration lead to new nonlinear diffusion equations. We also discuss how we can borrow the idea of anisotropy of the PM nonlinear diffusion to design new iterative wavelet frame shrinkage algorithms which are adaptive to local image features. In Section 6, we prove the convergence of iterative wavelet frame shrinkage algorithm to the solution of a 2nd-order nonlinear diffusion equation as meshsize goes to zero. We also address stability and convergence of generic iterative wavelet frame shrinkage algorithms. Finally, numerical experiments are presented in Section 7.

2 Preliminaries and Main Ideas

This section starts with an overview of wavelet frames, including some basic concepts of wavelet frames such as vanishing moments, and generic iterative wavelet frame shrinkage algorithms. Then, we provide some of the general ideas of how wavelet frame shrinkage algorithms can be regarded as discrete approximations to nonlinear evolution PDEs. Finally, we point out what are the fundamental differences of wavelet frame based approach in solving nonlinear evolution PDEs from the existing methods such as finite difference methods and wavelet Galerkin methods, especially, in the content of image restorations.

2.1 Review of Wavelet Frames

In this subsection, we first briefly introduce the concept of wavelet frames. The interested readers should consult [22, 23, 54, 55] for theories of frames and wavelet frames, [57] for a short survey on the theory and applications of frames, and [32] for a more detailed survey.

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

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

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

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

We call $\widetilde{X}$ a dual frame of X. We shall call the pair $(X, \widetilde{X})$ bi-frames. In general, for a given frame X, its dual frame is not unique. However, when X is a tight frame, it is self-dual, i.e. $\widetilde{X} = X$.

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

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

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

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

When $X(\Psi)$ forms a (tight) frame of $L_2(\mathbb{R}^d)$, each function ψ_ℓ, $\ell=1,\ldots,L$, is called a (tight) framelet and the whole system $X(\Psi)$ is called a (tight) wavelet frame system. Note that in the literature, the affine (or wavelet) system is commonly used, which corresponds to the decimated wavelet (frame) transforms. The quasi-affine system, which corresponds to the so-called undecimated wavelet (frame) transforms, was first introduced and analyzed by [54]. Here, we only discuss the quasi-affine system (2.2), since it works better in image restoration and its connection to PDEs is more natural than the affine system. The interested reader can find further details on the affine wavelet frame systems and its relation to the quasi-affine frames in [16, 32, 54].

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

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

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

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

For a sequence $\{\boldsymbol{p}[\boldsymbol{k}]\}_{\boldsymbol{k}\in\mathbb{Z}^2}$ of real numbers, we use $\widehat{\boldsymbol{p}}(\boldsymbol{\omega})$ to denote its (two-scale) symbol (it is also called a filter here):

$$\widehat{\boldsymbol{p}}(\boldsymbol{\omega}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^2}\boldsymbol{p}[\boldsymbol{k}]e^{-i\boldsymbol{k}\boldsymbol{\omega}}.$$

When $\boldsymbol{p}$ is a sequence with finitely many nonzero terms, its corresponding two-scale symbols $\widehat{\boldsymbol{p}}(\boldsymbol{\omega})$ is a trigonometric polynomial, and we shall call it a finite impulse response (FIR) filter.

The mixed extension principle (MEP) of [55] provides a general theory of the construction of MRA-based wavelet bi-frames. Given two sets of FIR filters $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\ldots,\boldsymbol{q}^{(L)}\}$ and $\{\tilde{\boldsymbol{p}},\tilde{\boldsymbol{q}}_1,\ldots,\tilde{\boldsymbol{q}}_L\}$, the MEP says that as long as we have

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

for all $\boldsymbol{\nu}\in\{0,\pi\}^d\setminus\{\boldsymbol{0}\}$ and $\boldsymbol{\xi}\in[-\pi,\pi]^d$, the quasi-affine systems $X(\Psi)$ and $X(\widetilde{\Psi})$ with Ψ and $\widetilde{\Psi}$ given by (2.3) forms a pair of *bi-frames* in $L_2(\mathbb{R}^d)$. In particular, when $\boldsymbol{p}=\widetilde{\boldsymbol{p}}$ and $\boldsymbol{q}^{(\ell)}=\widetilde{\boldsymbol{q}}^{(\ell)}$ for $\ell=1,\ldots,L$, the MEP (2.4) become the following unitary extension principle (UEP) discovered in [54]:

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

and the system $X(\Psi)$ is a *tight frame* of $L_2(\mathbb{R}^d)$. We call $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}},\widetilde{\boldsymbol{q}}^{(1)},\cdots,\widetilde{\boldsymbol{q}}^{(L)}\}$ a pair of *bi-frame filter banks* if they satisfy (2.4). $\boldsymbol{p}$ is called *lowpass* filter and $\boldsymbol{q}^{(\ell)},\widetilde{\boldsymbol{q}}^{(\ell)}$ are called *highpass* filters. If $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(L)}\}$ satisfies (2.5), then it is called a *tight frame filter bank.* Note that some of the filter banks we use in later sections only satisfy the first identity of (2.4) or (2.5), and they shall be called *undecimated bi-frame filter banks* or *undecimated tight frame filter banks.* In this case the system generated by the functions associated to these filters does not form a frame or tight frame for $L_2(\mathbb{R}^d)$. However, these filters do form frames or tight frames (undecimated) for sequence space $\ell_2(\mathbb{Z}^d)$. Since image data are elements in $\ell_2(\mathbb{Z}^d)$, undecimated (tight) frames in the sequence space can also be used to efficiently represent images. Therefore, we shall consider both types of filter banks and refer to them all as bi-frame or tight frame filter banks.

Now, we show two simple but useful examples of univariate tight framelets. The framelet given in Example 2.1 is known as the Haar wavelet. 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 ψ_1 in Example 2.1 as the "Haar framelet". The tight framelets given by Example 2.2 are constructed from piecewise linear B-spline first given by [54]. We shall refer to ψ_1 and ψ_2 in Example 2.2 as "piecewise linear framelets". The framelets constructed by B-splines, especially the piecewise linear framelets, are widely used in frame based image restoration problems because they provide sparse approximations to piecewise smooth, especially piecewise linear, functions such as images (see, e.g., [7, 11–13, 16, 18–20, 27–29, 46, 70]). In this paper, we shall refer the tight wavelet frame system constructed by Ron and Shen in [54] as the B-spline tight wavelet frame system in general.

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 [54]. This is a crucial observation in [8] indicating that a link does exist between the variational and wavelet frame based approach. We shall further extend such an observation to framelet masks that are not standard finite difference operators. In fact, as we will see in the next subsection, the order of the finite difference operator corresponding to a given frame highpass filter is closely related to the vanishing moment of the associated framelet.

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

Example 2.2. *[54]. Let $\boldsymbol{p} = \frac{1}{4}[1, 2, 1]$ be the refinement mask of the piecewise linear B-spline $B_2(x) = \max(1 - |x|, 0)$. Define $\boldsymbol{q}_1 = \frac{\sqrt{2}}{4}[1, 0, -1]$ and $\boldsymbol{q}_2 = \frac{1}{4}[-1, 2, -1]$. Then $\boldsymbol{p}$, $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ satisfy both identities of (2.5). Hence, the system $X(\Psi)$ where $\Psi = \{\psi_1, \psi_2\}$ defined in (2.1) is a tight frame of $L_2(\mathbb{R})$. The masks $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ correspond to the first order and second order difference operators respectively up to a scaling.*

For practical concerns, we need to consider 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 frames for $L_2(\mathbb{R}^d)$ is by taking *tensor products* of univariate frames. For simplicity of notation, we will consider the 2-D case, i.e. $d = 2$. Arguments for $d = 3$ or higher dimensions are similar.

Given a set of univariate masks $\{\boldsymbol{q}_\ell : \ell = 0, 1, \ldots, r\}$ (here we let $\boldsymbol{q}_0 = \boldsymbol{p}$ for convenience), define the 2-D masks $\boldsymbol{q}_{\boldsymbol{i}}[\boldsymbol{k}]$, with $\boldsymbol{i} := (i_1, i_2)$ and $\boldsymbol{k} := (k_1, k_2)$, as

$$\boldsymbol{q}_{\boldsymbol{i}}[\boldsymbol{k}] := \boldsymbol{q}_{i_1}[k_1]\boldsymbol{q}_{i_2}[k_2], \quad 0 \le i_1, i_2 \le r;\ (k_1, k_2) \in \mathbb{Z}^2. \tag{2.6}$$

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

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

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

$$\Psi_2 := \{\psi_{\boldsymbol{i}};\ 0 \le i_1, i_2 \le r;\ \boldsymbol{i} \ne (0, 0)\}.$$

Similarly, we can obtain $\widetilde{\boldsymbol{q}}_\ell$ and $\widetilde{\Psi}_2$. If the pair of univariate masks $\{\boldsymbol{q}_\ell\}$ and $\{\widetilde{\boldsymbol{q}}_\ell\}$ are constructed from the MEP (2.4), then it is easy to verify that $\{\boldsymbol{q}_{\boldsymbol{i}}\}$ and $\{\widetilde{\boldsymbol{q}}_{\boldsymbol{i}}\}$ satisfies the MEP conditions as well, and thus $(X(\Psi_2), X(\widetilde{\Psi}_2))$ is a pair of wavelet bi-frames for $L_2(\mathbb{R}^2)$.

In the discrete setting, let an image $\boldsymbol{f}$ be a d-dimensional array. We denote by $\mathcal{I}_d := \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_d}$ the set of all d-dimensional images. We will further assume that all images are square images, i.e. $N_1 = N_2 = \cdots = N_d = N$ and they all have supports in the open unit d-dimensional cube $\Omega = (0, 1)^d$. For simplicity, we will focus on $d = 2$ throughout the rest of this paper. We denote the 2-dimensional fast Lev-level wavelet frame transform/decomposition with $\{q^{(0)}, q^{(1)}, \cdots, q^{(L)}\}$ (see, e.g., [32]) as

$$\boldsymbol{W}\boldsymbol{u} = \{\boldsymbol{W}_{l,\ell}\boldsymbol{u} : 0 \le l \le \mathrm{Lev} - 1, 0 \le \ell \le L\}, \qquad \boldsymbol{u} \in \mathcal{I}_2. \tag{2.7}$$

We will mostly consider the case Lev $=1$ in this paper and, in that case, $\boldsymbol{W}_\ell = \boldsymbol{W}_{0,\ell}$. However, all results we have can be generalized to Lev >1 without much difficulty. In some of our analytical results and numerical experiments, we also use multi-level wavelet frame decomposition for better image restoration quality.

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

$$\boldsymbol{W}_{l,\ell}\boldsymbol{u} := \boldsymbol{q}_{l,\ell}[-\cdot] \circledast \boldsymbol{u},$$

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

$$\boldsymbol{q}_{l,\ell} = \check{\boldsymbol{q}}_{l,\ell} \circledast \check{\boldsymbol{q}}_{l-1,0} \circledast \ldots \circledast \check{\boldsymbol{q}}_{0,0} \quad \text{with} \quad \check{\boldsymbol{q}}_{l,\ell}[\boldsymbol{k}] = \begin{cases} \boldsymbol{q}_\ell[2^{-l}\boldsymbol{k}], & \boldsymbol{k} \in 2^l\mathbb{Z}^2; \\ 0, & \boldsymbol{k} \notin 2^l\mathbb{Z}^2. \end{cases} \tag{2.8}$$

Notice that $\boldsymbol{q}_{0,\ell} = \boldsymbol{q}_\ell$, and we let $\boldsymbol{q}_0 = \boldsymbol{p}$ for convenience. Similarly, we can define $\widetilde{\boldsymbol{W}}\boldsymbol{u}$ and $\widetilde{\boldsymbol{W}}_{l,\ell}\boldsymbol{u}$. We denote the inverse wavelet frame transform (or wavelet frame reconstruction) as $\widetilde{\boldsymbol{W}}^\top$, which is the adjoint operator of $\widetilde{\boldsymbol{W}}$, and by the MEP, we have the perfect reconstruction formula

$$\boldsymbol{u} = \widetilde{\boldsymbol{W}}^\top \boldsymbol{W}\boldsymbol{u}, \quad \text{for all } \boldsymbol{u} \in \mathcal{I}_2. \tag{2.9}$$

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

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

2.2 Vanishing Moments

The concept of vanishing moments of wavelet frames and their associated FIR filters is closely related to the orders of differential operators and their corresponding finite difference operators. The correspondence between the vanishing moments of wavelet frames and the orders of differential operators was crucial to the analysis of [8]. In this paper, the key observation, which is given in Lemma 2.1, is the connection between the vanishing moments of FIR filters and the order of finite difference operators (and the orders of approximation as well).

For an FIR highpass filter $\boldsymbol{q}$, let $\widehat{\boldsymbol{q}}(\boldsymbol{\omega}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}]e^{-i\boldsymbol{k}\boldsymbol{\omega}}$ be its two-scale symbol. Throughout this paper, for a multi-index $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{Z}_+^2$ and $\boldsymbol{\omega} \in \mathbb{R}^2$, denote

$$\boldsymbol{\alpha}! = \alpha_1!\alpha_2!, \ |\boldsymbol{\alpha}| = \alpha_1 + \alpha_2, \ \frac{\partial^{\boldsymbol{\alpha}}}{\partial \boldsymbol{\omega}^{\boldsymbol{\alpha}}} = \frac{\partial^{\alpha_1+\alpha_2}}{\partial\omega_2^{\alpha_2}\partial\omega_1^{\alpha_1}}.$$

We say $\boldsymbol{q}$ (and $\widehat{\boldsymbol{q}}(\boldsymbol{\omega})$) to have *vanishing moments of order* $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, where $\boldsymbol{\alpha} \in \mathbb{Z}_+^2$, provided that

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{k}^{\boldsymbol{\beta}} \boldsymbol{q}[\boldsymbol{k}] = i^{|\boldsymbol{\beta}|} \frac{\partial^{\boldsymbol{\beta}}}{\partial \boldsymbol{\omega}^{\boldsymbol{\beta}}} \widehat{\boldsymbol{q}}(\boldsymbol{\omega})\Big|_{\boldsymbol{\omega}=\boldsymbol{0}} = 0$$

for all $\boldsymbol{\beta} \in \mathbb{Z}_+^2$ with $|\boldsymbol{\beta}| < |\boldsymbol{\alpha}|$ and for all $\boldsymbol{\beta} \in \mathbb{Z}_+^2$ with $|\boldsymbol{\beta}| = |\boldsymbol{\alpha}|$ but $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$. By convention, we say that $\boldsymbol{q}$ has the vanishing moment of order $(0,0)$ if $\sum_{\boldsymbol{k}} \boldsymbol{q}[\boldsymbol{k}] \neq 0$. We also say $\boldsymbol{q}$ to have *total vanishing moments of order* K with $K \in \mathbb{Z}_+$, if

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{k}^{\boldsymbol{\beta}} \boldsymbol{q}[\boldsymbol{k}] = i^{|\boldsymbol{\beta}|} \frac{\partial^{\boldsymbol{\beta}}}{\partial \boldsymbol{\omega}^{\boldsymbol{\beta}}} \widehat{\boldsymbol{q}}(\boldsymbol{\omega})\Big|_{\boldsymbol{\omega}=\boldsymbol{0}} = 0 \quad \text{for all } \boldsymbol{\beta} \in \mathbb{Z}_+^2 \text{ with } |\boldsymbol{\beta}| < K. \tag{2.10}$$

Suppose $K \geq 1$. If (2.10) holds for all $\boldsymbol{\beta} \in \mathbb{Z}_+^2$ with $|\boldsymbol{\beta}| < K$ except for $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ with certain $\boldsymbol{\beta}_0 \in \mathbb{Z}_+^2$ and $|\boldsymbol{\beta}_0| = J < K$, then we say $\boldsymbol{q}$ to have *total vanishing moments of order* $K\backslash\{J+1\}$.

Clearly, if $\boldsymbol{q}$ has vanishing moments of order $\boldsymbol{\alpha}$, then it has total vanishing moments of order at least $|\boldsymbol{\alpha}|$, and it has total vanishing moments of order $K\backslash\{|\boldsymbol{\alpha}|+1\}$ with $K \geq |\boldsymbol{\alpha}|+1$. It is obvious that if $\sum_{k_1\in\mathbb{Z}} k_1^{\beta_1} \boldsymbol{q}[k_1,k_2] = 0$ for all $0 \leq \beta_1 < \alpha_1, k_2 \in \mathbb{Z}$ and $\sum_{k_2\in\mathbb{Z}} k_2^{\beta_2} \boldsymbol{q}[k_1,k_2] = 0$ for all $0 \leq \beta_2 < \alpha_2, k_1 \in \mathbb{Z}$, then $\boldsymbol{q}$ has vanishing moments of order $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$. To have a better understanding of the concept of vanishing moments, let us look at two examples.

Let $\widehat{\boldsymbol{q}}_1(\boldsymbol{\omega}) = e^{i\omega_1} - e^{-i\omega_1}$. Then

$$\widehat{\boldsymbol{q}}_1(\mathbf{0}) = 0, \ \frac{\partial}{\partial \omega_1}\widehat{\boldsymbol{q}}_1(\mathbf{0}) = 2i \neq 0, \ \frac{\partial}{\partial \omega_2}\widehat{\boldsymbol{q}}_1(\mathbf{0}) = 0.$$

Thus $\widehat{\boldsymbol{q}}_1(\boldsymbol{\omega})$ has vanishing moments of order $(1,0)$. In addition, we have

$$\frac{\partial^2}{\partial \omega_1^2}\widehat{\boldsymbol{q}}_1(\mathbf{0}) = 0, \ \frac{\partial^2}{\partial \omega_1 \partial \omega_2}\widehat{\boldsymbol{q}}_1(\mathbf{0}) = 0, \ \frac{\partial^2}{\partial \omega_2^2}\widehat{\boldsymbol{q}}_1(\mathbf{0}) = 0.$$

Therefore, $\boldsymbol{q}_1$ has total vanishing moments of order $3\backslash\{|(1,0)|+1\}$, or $3\backslash\{2\}$ (it does not have total vanishing moments of order $4\backslash\{2\}$ since $\frac{\partial^3}{\partial \omega_1^3}\widehat{\boldsymbol{q}}_1(\mathbf{0}) = -2i \neq 0$).

Let $\widehat{\boldsymbol{q}}_2(\boldsymbol{\omega}) = (e^{i\omega_1} - e^{-i\omega_1})(1 - e^{-i\omega_2})^2$. Then

$$\frac{\partial^{\boldsymbol{\beta}}}{\partial \omega^{\boldsymbol{\beta}}}\widehat{\boldsymbol{q}}_2(\mathbf{0}) = 0 \quad \text{for } |\boldsymbol{\beta}| < 3 \text{ and } \boldsymbol{\beta} = (3,0), (2,1), (0,3),$$

and $\frac{\partial^3}{\partial \omega_1 \partial \omega_2^2}\widehat{\boldsymbol{q}}_2(\mathbf{0}) = -4i \neq 0$. Thus $\boldsymbol{q}_2$ has vanishing moments of order $(1,2)$. Observe that $\frac{\partial^4}{\partial \omega_1 \partial \omega_2^3}\widehat{\boldsymbol{q}}_2(\mathbf{0}) = -4 \neq 0$. Therefore, $\boldsymbol{q}_2$ has total vanishing moments of order $4\backslash\{4\}$ (instead of $5 \setminus \{4\}$).

Lemma 2.1. *Let $\boldsymbol{q}$ be an FIR highpass filter with vanishing moments of order $\boldsymbol{\alpha} \in \mathbb{Z}_+^2$. Then for a smooth function $F(\boldsymbol{x})$ on $\mathbb{R}^2$, we have*

$$\frac{1}{\varepsilon^{|\boldsymbol{\alpha}|}} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}] F(\boldsymbol{x} + \varepsilon \boldsymbol{k}) = C_{\boldsymbol{\alpha}} \frac{\partial^{\boldsymbol{\alpha}}}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} F(\boldsymbol{x}) + O(\varepsilon), \ \text{as } \varepsilon \to 0, \tag{2.11}$$

where $C_{\boldsymbol{\alpha}}$ is the constant defined by

$$C_{\boldsymbol{\alpha}} = \frac{1}{\boldsymbol{\alpha}!} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{k}^{\boldsymbol{\alpha}} \boldsymbol{q}[\boldsymbol{k}] = \frac{i^{|\boldsymbol{\alpha}|}}{\boldsymbol{\alpha}!} \frac{\partial^{\boldsymbol{\alpha}}}{\partial \boldsymbol{\omega}^{\boldsymbol{\alpha}}} \widehat{\boldsymbol{q}}(\boldsymbol{\omega}) \Big|_{\boldsymbol{\omega}=\mathbf{0}}. \tag{2.12}$$

If, in addition, $\boldsymbol{q}$ has total vanishing moments of order $K\backslash\{|\boldsymbol{\alpha}|+1\}$ for some $K > |\boldsymbol{\alpha}|$, then

$$\frac{1}{\varepsilon^{|\boldsymbol{\alpha}|}} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}] F(\boldsymbol{x} + \varepsilon \boldsymbol{k}) = C_{\boldsymbol{\alpha}} \frac{\partial^{\boldsymbol{\alpha}}}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} F(\boldsymbol{x}) + O(\varepsilon^{K-|\boldsymbol{\alpha}|}), \ \text{as } \varepsilon \to 0. \tag{2.13}$$

Proof. Straightforward calculations based on Taylor's expansion. ☐

2.3 Wavelet Frame Filter Banks

For any nonlinear evolution PDE considered in this paper, we can simply use the filter bank of one of the tensor-product B-spline wavelet frame system constructed in [54], as long as the highest order of vanishing moments of the highpass filters is no lower than half of the order of the PDE. All we need to do is to choose appropriate parameters for each of the highpass filter such that the ones that are inactive in the given PDE, i.e. the filters whose associated differential operators do not appear in the PDE, converge to zero asymptotically. One may also simply set those parameters to zero. However, choosing a parameter that asymptotically goes to zero some time leads to better image restoration results. This idea was first appeared in [8]. We will also present some specific choices of the parameters associated to the inactive highpass filters in Corollary 3.1 and 3.2. Numerical examples showing the benefit of having inactive highpass filters in the filter bank are given in Section 7.

However, using B-spline type filter banks for all nonlinear evolution PDEs may not always be efficient, especially when we have too many inactive filters. We need to compute the decomposition transform associated to those inactive filters in any case, otherwise the first identity of (2.4) is violated and we will not have the perfect reconstruction (2.9). In some applications, when image quality is less of a concern than computation efficiency, it is desirable to specifically construct filter bank for a given PDE that has as few filters as possible. Therefore, we list some of the FIR filter banks that will be used in later sections. As we will elaborate that these filters can be used to discretize different type of nonlinear evolution PDEs, especially nonlinear diffusion equations.

Let $\{\boldsymbol{a}, \boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}\}$ be piecewise linear B-spline tight frame filter bank of [54] (given by Example 2.2)

$$\begin{aligned}
&\widehat{\boldsymbol{a}}(\omega) = \frac{1}{2} + \frac{1}{4}(e^{-i\omega} + e^{i\omega}) = \frac{1}{4}e^{i\omega}(1 + e^{-i\omega})^2,\\
&\widehat{\boldsymbol{b}}^{(1)}(\omega) = \frac{\sqrt{2}}{4}(e^{i\omega} - e^{-i\omega}), \widehat{\boldsymbol{b}}^{(2)}(\omega) = \frac{1}{2} - \frac{1}{4}(e^{-i\omega} + e^{i\omega}) = \frac{1}{4}e^{i\omega}(1 - e^{-i\omega})^2.
\end{aligned} \tag{2.14}$$

By tensor product, we can construct the separable 2-D piecewise linear B-spline tight frame filter bank from (2.14), which is commonly used in image denoising, image inpainting and other areas. We note that all filter banks in this subsection except for (2.15) are undecimated bi-frame or tight frame filter banks, which mean they only satisfy the first equation of (2.4) or (2.5).

Example 2.3. *Let $\boldsymbol{a}$, $\boldsymbol{b}^{(1)}$ and $\boldsymbol{b}^{(2)}$ be given in* (2.14)*. The separable 2-D piecewise linear B-spline tight frame filter bank $\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(8)}$, with $\boldsymbol{\omega} = (\omega_1, \omega_2)$, are given by*

$$\begin{aligned}
&\widehat{\boldsymbol{p}}(\boldsymbol{\omega}) = \widehat{\boldsymbol{a}}(\omega_1)\widehat{\boldsymbol{a}}(\omega_2),\ \widehat{\boldsymbol{q}}^{(1)}(\boldsymbol{\omega}) = \widehat{\boldsymbol{b}}^{(1)}(\omega_1)\widehat{\boldsymbol{a}}(\omega_2),\ \widehat{\boldsymbol{q}}^{(2)}(\boldsymbol{\omega}) = \widehat{\boldsymbol{a}}(\omega_1)\widehat{\boldsymbol{b}}^{(1)}(\omega_2),\\
&\widehat{\boldsymbol{q}}^{(3)}(\boldsymbol{\omega}) = \widehat{\boldsymbol{b}}^{(2)}(\omega_1)\widehat{\boldsymbol{a}}(\omega_2),\ \widehat{\boldsymbol{q}}^{(4)}(\boldsymbol{\omega}) = \widehat{\boldsymbol{b}}^{(1)}(\omega_1)\widehat{\boldsymbol{b}}^{(1)}(\omega_2),\ \widehat{\boldsymbol{q}}^{(5)}(\boldsymbol{\omega}) = \widehat{\boldsymbol{a}}(\omega_1)\widehat{\boldsymbol{b}}^{(2)}(\omega_2),\\
&\widehat{\boldsymbol{q}}^{(6)}(\boldsymbol{\omega}) = \widehat{\boldsymbol{b}}^{(2)}(\omega_1)\widehat{\boldsymbol{b}}^{(1)}(\omega_2),\ \widehat{\boldsymbol{q}}^{(7)}(\boldsymbol{\omega}) = \widehat{\boldsymbol{b}}^{(1)}(\omega_1)\widehat{\boldsymbol{b}}^{(2)}(\omega_2),\ \widehat{\boldsymbol{q}}^{(8)}(\boldsymbol{\omega}) = \widehat{\boldsymbol{b}}^{(2)}(\omega_1)\widehat{\boldsymbol{b}}^{(2)}(\omega_2).
\end{aligned} \tag{2.15}$$

It is straightforward to obtain the following orders of vanishing moments of $\boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(8)}$:

$$\boldsymbol{\beta}_1 = (1,0),\ \boldsymbol{\beta}_2 = (0,1),\ \boldsymbol{\beta}_3 = (2,0),\ \boldsymbol{\beta}_4 = (1,1), \boldsymbol{\beta}_5 = (0,2),\ \boldsymbol{\beta}_6 = (2,1),\ \boldsymbol{\beta}_7 = (1,2),\ \boldsymbol{\beta}_8 = (2,2).$$

These filters will be used to discretize nonlinear diffusion equation (4.1)*.*

We can construct a similar type of tight frame filter bank as (2.15) with fewer highpass filters.

Example 2.4. *Let $\boldsymbol{a}$ be given in* (2.14)*. A separable 2-D piecewise linear B-spline tight frame filter bank with fewer highpass filters is given by*

$$\begin{aligned}
&\widehat{\boldsymbol{p}}(\boldsymbol{\omega}) = \widehat{\boldsymbol{a}}(\omega_1)\widehat{\boldsymbol{a}}(\omega_2), \widehat{\boldsymbol{q}}^{(1)}(\boldsymbol{\omega}) = \frac{\sqrt{2}}{16}\left(e^{i\omega_1} - e^{-i\omega_1}\right)\left(1 + e^{-i\omega_2}\right)^2 e^{i\omega_2},\\
&\widehat{\boldsymbol{q}}^{(2)}(\boldsymbol{\omega}) = \frac{\sqrt{2}}{8}\left(e^{i\omega_1} + e^{-i\omega_1}\right)\left(e^{i\omega_2} - e^{-i\omega_2}\right), \widehat{\boldsymbol{q}}^{(3)}(\boldsymbol{\omega}) = \frac{1}{16}\left(1 - e^{-i\omega_1}\right)^2\left(1 + e^{-i\omega_2}\right)^2 e^{i\omega_1}e^{i\omega_2},\\
&\widehat{\boldsymbol{q}}^{(4)}(\boldsymbol{\omega}) = \frac{\sqrt{2}}{8}\left(e^{i\omega_1} - e^{-i\omega_1}\right)\left(e^{i\omega_2} - e^{-i\omega_2}\right), \widehat{\boldsymbol{q}}^{(5)}(\boldsymbol{\omega}) = \frac{1}{4}\left(1 - e^{-i\omega_2}\right)^2 e^{i\omega_2}.
\end{aligned} \tag{2.16}$$

These filters will be used to discretize nonlinear diffusion equations (4.4) *and* (4.5)*.*

Furthermore, we can construct tight frame filter banks which result in the rotation invariant diffusion equations of arbitrary orders.

Example 2.5. *The lowpass and highpass filters are given respectively by*

$$\widehat{\boldsymbol{p}}(\boldsymbol{\omega}) = \frac{1}{2^{2m}}\left(1 + e^{-i\omega_1}\right)^m\left(1 + e^{-i\omega_2}\right)^m e^{i[m/2](\omega_1+\omega_2)}, \quad m \geq 1 \tag{2.17}$$

and

$$\begin{aligned}
&\widehat{\boldsymbol{q}}^{(s,k)}(\boldsymbol{\omega}) = \\
&\frac{1}{2^{2m-k}}\sqrt{\binom{m}{s}\binom{s}{k}}e^{i[\frac{m-k}{2}]\omega_1}e^{i[m/2]\omega_2}\left(1 + e^{-i\omega_1}\right)^{m-s}\left(1 + e^{-i\omega_2}\right)^{m-k}\left(1 - e^{-i\omega_1}\right)^{s-k}\left(1 - e^{-i\omega_2}\right)^k.
\end{aligned} \tag{2.18}$$

These filters will be used to discretize the rotation invariant nonlinear diffusion equation (4.10)*.*

Now, we present filter banks that will be used for the diffusion equations of Perona-Malik's and TV models, image inpainting diffusion, and shock filtering.

Example 2.6. *The following bi-frame filter bank is constructed for the Perona-Malik diffusion* (3.20)*:*

$$\begin{cases}\widehat{\boldsymbol{p}}(\boldsymbol{\omega})=\widehat{\widetilde{\boldsymbol{p}}}(\boldsymbol{\omega})=\frac{1}{4}\big(1+e^{-i\omega_1}\big)\big(1+e^{-i\omega_2}\big),\\ \widehat{\boldsymbol{q}}^{(1)}(\boldsymbol{\omega})=\frac{1}{2}\big(1-e^{-i\omega_1}\big),\ \widehat{\widetilde{\boldsymbol{q}}}^{(1)}(\boldsymbol{\omega})=\frac{1}{16}\big(1-e^{-i\omega_1}\big)\big(6+e^{i\omega_2}+e^{-i\omega_2}\big),\\ \widehat{\boldsymbol{q}}^{(2)}(\boldsymbol{\omega})=\frac{1}{2}\big(1-e^{-i\omega_2}\big),\ \widehat{\widetilde{\boldsymbol{q}}}^{(2)}(\boldsymbol{\omega})=\frac{1}{16}\big(1-e^{-i\omega_2}\big)\big(6+e^{i\omega_1}+e^{-i\omega_1}\big).\end{cases}\tag{2.19}$$

Example 2.7. *The following bi-frame filter bank is constructed for the image inpainting diffusion* (3.23)*:*

$$\begin{cases}\widehat{\boldsymbol{p}}(\boldsymbol{\omega})=\widehat{\widetilde{\boldsymbol{p}}}(\boldsymbol{\omega})=\frac{1}{4}(1+e^{-i\omega_1})(1+e^{-i\omega_2}),\widehat{\boldsymbol{q}}^{(1)}(\boldsymbol{\omega})=\frac{1}{2}(1-e^{-i\omega_1}),\ \widehat{\boldsymbol{q}}^{(2)}(\boldsymbol{\omega})=-\frac{1}{2}(1-e^{-i\omega_2}),\\ \widehat{\boldsymbol{q}}^{(3)}(\boldsymbol{\omega})=-\frac{1}{32}(1-e^{-i\omega_1})^2e^{i\omega_1}(6+e^{-i\omega_2}+e^{i\omega_2}),\widehat{\boldsymbol{q}}^{(4)}(\boldsymbol{\omega})=-\frac{1}{32}(1-e^{-i\omega_2})^2e^{i\omega_2}(6+e^{-i\omega_1}+e^{i\omega_1}),\\ \widehat{\widetilde{\boldsymbol{q}}}^{(1)}(\boldsymbol{\omega})=\frac{1}{2}(e^{i\omega_2}-1),\ \widehat{\widetilde{\boldsymbol{q}}}^{(2)}(\boldsymbol{\omega})=\frac{1}{2}(e^{i\omega_1}-1),\widehat{\widetilde{\boldsymbol{q}}}^{(3)}(\boldsymbol{\omega})=\widehat{\widetilde{\boldsymbol{q}}}^{(4)}(\boldsymbol{\omega})=1.\end{cases}\tag{2.20}$$

Example 2.8. *The following bi-frame filter bank is constructed for the nonlinear hyperbolic equation of shock filters* (3.27)*:*

$$\begin{cases}\widehat{\boldsymbol{p}}(\boldsymbol{\omega})=\widehat{\widetilde{\boldsymbol{p}}}(\boldsymbol{\omega})=\frac{1}{16}(1+e^{-i\omega_1})^2(1+e^{-i\omega_2})^2e^{i\omega_1}e^{i\omega_2},\\ \widehat{\boldsymbol{q}}^{(1)}(\boldsymbol{\omega})=\frac{1}{8}(e^{i\omega_1}-e^{-i\omega_1})(1+e^{-i\omega_2})^2e^{i\omega_2},\widehat{\boldsymbol{q}}^{(2)}(\boldsymbol{\omega})=\frac{1}{8}(e^{i\omega_2}-e^{-i\omega_2})(1+e^{-i\omega_1})^2e^{i\omega_1},\\ \widehat{\boldsymbol{q}}^{(3)}(\boldsymbol{\omega})=-\frac{1}{32}(1-e^{-i\omega_1})^2e^{i\omega_1}(6+e^{-i\omega_2}+e^{i\omega_2}),\widehat{\boldsymbol{q}}^{(4)}(\boldsymbol{\omega})=-\frac{1}{32}(1-e^{-i\omega_2})^2e^{i\omega_2}(6+e^{-i\omega_1}+e^{i\omega_1}),\\ \widehat{\widetilde{\boldsymbol{q}}}^{(1)}(\boldsymbol{\omega})=\frac{1}{64}(e^{i\omega_1}-e^{-i\omega_1})(6+e^{-i\omega_2}+e^{i\omega_2}),\widehat{\widetilde{\boldsymbol{q}}}^{(2)}(\boldsymbol{\omega})=\frac{1}{64}(e^{i\omega_2}-e^{-i\omega_2})(6+e^{-i\omega_1}+e^{i\omega_1}),\\ \widehat{\widetilde{\boldsymbol{q}}}^{(3)}(\boldsymbol{\omega})=\widehat{\widetilde{\boldsymbol{q}}}^{(4)}(\boldsymbol{\omega})=1.\end{cases}\tag{2.21}$$

Finally, we note that the filter bank that can be used to discretize a given PDE is not uniquely determined, even we do not allow having inactive highpass filters in the filter bank. For example, other than (2.20), we can use the following filter bank for the image inpainting diffusion (3.23):

$$\begin{aligned}&\widehat{\boldsymbol{p}}(\boldsymbol{\omega})=\widehat{\widetilde{\boldsymbol{p}}}(\boldsymbol{\omega})=\frac{1}{16}(1+e^{-i\omega_1})^2(1+e^{-i\omega_2})^2e^{i\omega_1}e^{i\omega_2},\\ &\widehat{\boldsymbol{q}}^{(1)}(\boldsymbol{\omega})=\frac{1}{2}(e^{i\omega_1}-e^{-i\omega_1}),\ \widehat{\boldsymbol{q}}^{(2)}(\boldsymbol{\omega})=-\frac{1}{2}(e^{i\omega_2}-e^{-i\omega_2}),\\ &\widehat{\boldsymbol{q}}^{(3)}(\boldsymbol{\omega})=-\frac{1}{512}(1-e^{-i\omega_1})^2e^{i\omega_1}(6+e^{-i\omega_1}+e^{i\omega_1})(22+4e^{-i\omega_2}+4e^{i\omega_2}+e^{-i2\omega_2}+e^{i2\omega_2}),\\ &\widehat{\boldsymbol{q}}^{(4)}(\boldsymbol{\omega})=-\frac{1}{512}(1-e^{-i\omega_2})^2e^{i\omega_2}(6+e^{-i\omega_2}+e^{i\omega_2})(22+4e^{-i\omega_1}+4e^{i\omega_1}+e^{-i2\omega_1}+e^{i2\omega_1}),\\ &\widehat{\widetilde{\boldsymbol{q}}}^{(1)}(\boldsymbol{\omega})=\frac{1}{2}(e^{i\omega_2}-e^{-i\omega_2}),\ \widehat{\widetilde{\boldsymbol{q}}}^{(2)}(\boldsymbol{\omega})=\frac{1}{2}(e^{i\omega_1}-e^{-i\omega_1}),\widehat{\widetilde{\boldsymbol{q}}}^{(3)}(\boldsymbol{\omega})=\widehat{\widetilde{\boldsymbol{q}}}^{(4)}(\boldsymbol{\omega})=1.\end{aligned}$$

However, constructing multiple filter banks for a given PDE is not the focus of this paper. Therefore, we shall only use the filter banks presented in this section as examples.

2.4 Iterative Wavelet Frame Shrinkage

As iterative wavelet frame shrinkage (especially the soft- and hard-thresholding) is the key ingredient of wavelet frame based approach for image restoration in [11,14,16,17,19,20,29,70]. In this section, we review the formula for more general iterative wavelet frame shrinkage, which goes beyond the soft- and hard-thresholding, rewrite it into a compact form that will be repeatedly used in later sections.

Let $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}},\widetilde{\boldsymbol{q}}^{(1)},\cdots,\widetilde{\boldsymbol{q}}^{(L)}\}$ be a pair of FIR filters that satisfy the first equation of (2.4). Let $\{\boldsymbol{u}^0_{\boldsymbol{j}}\}_{\boldsymbol{j}\in\mathbb{Z}^2}$ be the initial data. The (1-level) wavelet frame transform based denoising consists of the following processes:

$$\begin{aligned}&L_{\boldsymbol{n}}=\sum_{\boldsymbol{j}\in\mathbb{Z}^2}\boldsymbol{p}[\boldsymbol{j}]\boldsymbol{u}^0_{\boldsymbol{j}+\boldsymbol{n}},\ H^{(\ell)}_{\boldsymbol{n}}=\sum_{\boldsymbol{j}\in\mathbb{Z}^2}\boldsymbol{q}^{(\ell)}[\boldsymbol{j}]\boldsymbol{u}^0_{\boldsymbol{j}+\boldsymbol{n}},\quad \boldsymbol{n}\in\mathbb{Z}^2,1\le\ell\le L;\\ &\boldsymbol{u}^1_{\boldsymbol{j}}=\sum_{\boldsymbol{n}\in\mathbb{Z}^2}\widetilde{\boldsymbol{p}}[\boldsymbol{j}-\boldsymbol{n}]L_{\boldsymbol{n}}+\sum_{\ell=1}^{L}\sum_{\boldsymbol{n}\in\mathbb{Z}^2}\widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}]\mathcal{S}_\ell(H^{(\ell)}_{\boldsymbol{n}}),\quad \boldsymbol{j}\in\mathbb{Z}^2,\end{aligned}\tag{2.22}$$

where $\mathcal{S}_\ell, 1 \le \ell \le L$ are the shrinkage functions. The first step in (2.22) is called the analysis process, and the second one is called the synthesis process after shrinkage. $L_{\boldsymbol{n}}$ and $H_{\boldsymbol{n}}^{(\ell)}$ are the lowpass and highpass outputs of $\boldsymbol{u}^0$. The shrinkage functions we consider in this paper are not limited to the well-known soft-thresholding that is frequently used in image restoration. As we shall see in later sections, different shrinkage operators will correspond to different PDEs.

Proposition 2.1. *Let $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}}, \widetilde{\boldsymbol{q}}^{(1)}, \cdots, \widetilde{\boldsymbol{q}}^{(L)}\}$ be a pair of bi-frame filter banks, and $\boldsymbol{u}^1_{\boldsymbol{j}}$ be the resulting signal given by* (2.22) *after 1-step frame shrinkage of $\boldsymbol{u}^0_{\boldsymbol{j}}$ with these filter banks. Then*

$$\boldsymbol{u}^1_{\boldsymbol{j}} = \boldsymbol{u}^0_{\boldsymbol{j}} + \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}]\big(\mathcal{S}_\ell(\xi) - \xi\big)\Big|_{\xi = H_{\boldsymbol{n}}^{(\ell)}}, \ \boldsymbol{j} \in \mathbb{Z}^2, \tag{2.23}$$

where $H_{\boldsymbol{n}}^{(\ell)}$ is defined by (2.22).

Proof. Since $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}}, \widetilde{\boldsymbol{q}}^{(1)}, \cdots, \widetilde{\boldsymbol{q}}^{(L)}\}$ satisfy the first equation of (2.4), $\boldsymbol{u}^0_{\boldsymbol{j}}$ can be recovered from the synthesis algorithm (2.22) with $\mathcal{S}_\ell(\xi) = \xi$, namely,

$$\boldsymbol{u}^0_{\boldsymbol{j}} = \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{p}}[\boldsymbol{j}-\boldsymbol{n}] L_{\boldsymbol{n}} + \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}] H_{\boldsymbol{n}}^{(\ell)}, \ j \in \mathbb{Z}^2.$$

Thus,

$$\begin{aligned}
\boldsymbol{u}^1_{\boldsymbol{j}} &= \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{p}}[\boldsymbol{j}-\boldsymbol{n}] L_{\boldsymbol{n}} + \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}] \mathcal{S}_\ell(H_{\boldsymbol{n}}^{(\ell)}) \\
&= \boldsymbol{u}^0_{\boldsymbol{j}} - \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}] H_{\boldsymbol{n}}^{(\ell)} + \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}] \mathcal{S}_\ell(H_{\boldsymbol{n}}^{(\ell)}) \\
&= \boldsymbol{u}^0_{\boldsymbol{j}} + \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}]\big(\mathcal{S}_\ell(\xi) - \xi\big)\Big|_{\xi = H_{\boldsymbol{n}}^{(\ell)}}, \ \boldsymbol{j} \in \mathbb{Z}^2.
\end{aligned}$$

□

The frame shrinkage process (2.22) can be applied iteratively:

$$\begin{aligned}
& H_{\boldsymbol{n}}^{(\ell),k-1} = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \boldsymbol{q}^{(\ell)}[\boldsymbol{j}] \boldsymbol{u}^{k-1}_{\boldsymbol{j}+\boldsymbol{n}}, \quad 1 \le \ell \le L; \\
& \boldsymbol{u}^k_{\boldsymbol{j}} = \boldsymbol{u}^{k-1}_{\boldsymbol{j}} + \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}]\Big(\mathcal{S}_\ell(H_{\boldsymbol{n}}^{(\ell),k-1}) - H_{\boldsymbol{n}}^{(\ell),k-1}\Big), \quad k = 1, 2, \cdots.
\end{aligned} \tag{2.24}$$

In this paper we will also consider the channel-mixed frame shrinkage (coupled frame shrinkage):

$$\boldsymbol{u}^1_{\boldsymbol{j}} = \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{p}}[\boldsymbol{j}-\boldsymbol{n}] L_{\boldsymbol{n}} + \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}] \mathcal{S}_\ell(H_{\boldsymbol{n}}^{(1)}, H_{\boldsymbol{n}}^{(2)}, \cdots, H_{\boldsymbol{n}}^{(L)}), \ \boldsymbol{j} \in \mathbb{Z}^2, \tag{2.25}$$

where $\mathcal{S}_\ell, 1 \le \ell \le L, \boldsymbol{n} \in \mathbb{Z}^2$ are functions of several variables. If $\boldsymbol{u}^1_{\boldsymbol{j}}, \boldsymbol{j} \in \mathbb{Z}^2$, are given by (2.25) after channel-mixed shrinkage, then one can obtain similarly to the proof of Proposition 2.1 that

$$\boldsymbol{u}^1_{\boldsymbol{j}} = \boldsymbol{u}^0_{\boldsymbol{j}} + \sum_{\ell=1}^{L} \sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}]\Big(\mathcal{S}_\ell(H_{\boldsymbol{n}}^{(1)}, H_{\boldsymbol{n}}^{(2)}, \cdots, H_{\boldsymbol{n}}^{(L)}) - H_{\boldsymbol{n}}^{(\ell)}\Big), \ \boldsymbol{j} \in \mathbb{Z}^2. \tag{2.26}$$

The channel-mixed frame shrinkage process can also be applied iteratively:

$$\begin{aligned} H_{\boldsymbol{n}}^{(\ell),k-1} &= \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \boldsymbol{q}^{(\ell)}[\boldsymbol{j}]\boldsymbol{u}_{\boldsymbol{j}+\boldsymbol{n}}^{k-1}, \quad 1 \le \ell \le L; \\ \boldsymbol{u}_{\boldsymbol{j}}^k &= \boldsymbol{u}_{\boldsymbol{j}}^{k-1} + \sum_{\ell=1}^{L}\sum_{\boldsymbol{n}\in\mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{n}]\Big(\mathcal{S}_\ell(H_{\boldsymbol{n}}^{(1),k-1},\cdots,H_{\boldsymbol{n}}^{(L),k-1}) - H_{\boldsymbol{n}}^{(\ell),k-1}\Big), \quad k=1,2,\cdots. \end{aligned} \tag{2.27}$$

Although the choice of shrinking functions $\mathcal{S}_\ell$ can be general, they need to be meaningfully chosen. In this paper, we will focus on a few specific shrinkage operators which are closed-form solutions of certain optimization problems.

Next, we rewrite (2.24) and (2.27) in operator form. We will also present the shrinkage operators that will be analyzed in this paper. Let $\boldsymbol{W}\boldsymbol{u} =: \boldsymbol{d}$ be the wavelet frame decomposition of $\boldsymbol{u}$ defined by (2.7), and $\widetilde{\boldsymbol{W}}^\top \boldsymbol{d}$ be the reconstruction operator using a duel wavelet frame. We have $\widetilde{\boldsymbol{W}}^\top\boldsymbol{W} = \boldsymbol{I}$. For simplicity, we assume the level of decomposition is 1, i.e. Lev = 1 in (2.7). For a given wavelet frame coefficients $\boldsymbol{d} = \{d_{\ell,\boldsymbol{n}} : \boldsymbol{n}\in\mathbb{Z}^2, 0\le\ell\le L\}$ and threshold $\boldsymbol{\alpha}(\boldsymbol{d}) = \{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d}) : \boldsymbol{n}\in\mathbb{Z}^2, 0\le\ell\le L\}$, define the *multiplicative shrinkage operator* $\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d})$ as

$$\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d}) = \{S_{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}(1-\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})) : \boldsymbol{n}\in\mathbb{Z}^2, 0\le\ell\le L\}. \tag{2.28}$$

Note that $S_{\alpha_{\ell,\boldsymbol{n}}}(d_{\ell,\boldsymbol{n}})$ in (2.28) denotes $\Big(\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d})\Big)_{\ell,\boldsymbol{n}}$, whose value may depend on more values than merely $d_{\ell,\boldsymbol{n}}$ (see e.g. (3.17)). Similarly, we denote the *anisotropic and isotropic soft-thresholding operator* as $\boldsymbol{\mathcal{T}}_{\boldsymbol{\alpha}}^1$ and $\boldsymbol{\mathcal{T}}_{\boldsymbol{\alpha}}^2$ respectively, where

$$\boldsymbol{\mathcal{T}}_{\boldsymbol{\alpha}}^1(\boldsymbol{d}) = \Big\{\mathcal{T}^1_{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = \frac{d_{\ell,\boldsymbol{n}}}{|d_{\ell,\boldsymbol{n}}|}\max\{|d_{\ell,\boldsymbol{n}}| - \alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d}), 0\} : \boldsymbol{n}\in\mathbb{Z}^2, 0\le\ell\le L\Big\}, \tag{2.29}$$

and

$$\boldsymbol{\mathcal{T}}_{\boldsymbol{\alpha}}^2(\boldsymbol{d}) = \Big\{\mathcal{T}^2_{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = \frac{d_{\ell,\boldsymbol{n}}}{R_{\ell,\boldsymbol{n}}}\max\Big\{R_{\ell,\boldsymbol{n}} - \alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d}), 0\Big\} : \boldsymbol{n}\in\mathbb{Z}^2, 0\le\ell\le L\Big\}, \tag{2.30}$$

where $R_{\ell,\boldsymbol{n}} = \Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|} |d_{\ell',\boldsymbol{n}}|^2\Big)^{\frac{1}{2}}$. Observe that $R_{\ell,\boldsymbol{n}} = R_{\ell',\boldsymbol{n}}$ if $|\boldsymbol{\beta}_{\ell'}| = |\boldsymbol{\beta}_\ell|$. For all shrinkage operators, we always fix the threshold $\alpha_{0,\boldsymbol{n}} = 0$, which means we never penalize the low frequency coefficients $d_{0,\boldsymbol{n}}$.

With the notation (2.7) and choosing the shrinking functions $\mathcal{S}_\ell$ as $\boldsymbol{S}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\mathcal{T}}_{\boldsymbol{\alpha}}^{\wp}$, we can rewrite both (2.24) and (2.27) as

$$\boldsymbol{u}^k = \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}), \quad k=1,2,\cdots, \tag{2.31}$$

which shall be referred to as *the iterative multiplicative wavelet frame shrinkage (algorithm)*; and

$$\boldsymbol{u}^k = \widetilde{\boldsymbol{W}}^\top \boldsymbol{\mathcal{T}}^{\wp}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}), \quad k=1,2,\cdots, \quad \wp=1,2, \tag{2.32}$$

which shall be referred to as *the iterative (anisotropic/isotropic) wavelet frame soft-thresholding (algorithm)*. Here, $\boldsymbol{\alpha}^{k-1} = \{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d}^{k-1}) : \boldsymbol{n}\in\mathbb{Z}^2, 0\le\ell\le L\}$ with $\boldsymbol{d}^{k-1} = \boldsymbol{W}\boldsymbol{u}^{k-1}$. When $\boldsymbol{W}$ is the transform associated with a tight wavelet frame system, we have $\widetilde{\boldsymbol{W}} = \boldsymbol{W}$ in (2.31) and (2.32). We note that the shrinkage operator (2.28) is in fact so general that it includes (2.29) and (2.30) as special cases. In other words, the iterative multiplicative shrinkage (2.31) includes (2.32) as a special case. We shall give more details in Section 5 where we present nonlinear diffusions that are in correspondence to (2.32) with various choices of thresholds. However, we will keep the two types of thresholding separated in notation and in our discussions, since they have different thresholding mechanism and have rather different optimality properties as will be shown in the following subsection.

2.5 Optimality of Iteration (2.31) and (2.32)

We shall show that the solutions of two consecutive time steps of (2.31) and (2.32) are linked by some optimization problem.

Given an $\boldsymbol{\alpha}$, we assume that

$$0 \leq \boldsymbol{\alpha} < 1, \quad \text{i.e. } 0 \leq \alpha_{\ell,\boldsymbol{n}} < 1, \text{ for all } \boldsymbol{n} \in \mathbb{Z}^2,\ 0 \leq \ell \leq L. \tag{2.33}$$

Consider the quadratic optimization problem

$$\min_{\boldsymbol{d}} \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{W}\boldsymbol{u}\|_2^2 + \frac{1}{2}\Big\|\sqrt{\frac{\boldsymbol{\alpha}}{1-\boldsymbol{\alpha}}}\cdot \boldsymbol{d}\Big\|_2^2,$$

where $\boldsymbol{d} = \{d_{\ell,\boldsymbol{n}} :\ \boldsymbol{n} \in \mathbb{Z}^2, 0 \leq \ell \leq L\}$ and $\sqrt{\frac{\boldsymbol{\alpha}}{1-\boldsymbol{\alpha}}}\cdot \boldsymbol{d} := \Big\{\sqrt{\frac{\alpha_{\ell,\boldsymbol{n}}}{1-\alpha_{\ell,\boldsymbol{n}}}} d_{\ell,\boldsymbol{n}} :\ \boldsymbol{n} \in \mathbb{Z}^2, 0 \leq \ell \leq L\Big\}$. We first observe that

$$\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{W}\boldsymbol{u}) = \arg\min_{\boldsymbol{d}} \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{W}\boldsymbol{u}\|_2^2 + \frac{1}{2}\Big\|\sqrt{\frac{\boldsymbol{\alpha}}{1-\boldsymbol{\alpha}}}\cdot \boldsymbol{d}\Big\|_2^2,$$

which is easy to derive by simple differentiation. Also, we have (see e.g. [8, 21, 33])

$$\boldsymbol{\mathcal{T}}^{\wp}_{\boldsymbol{\alpha}}(\boldsymbol{W}\boldsymbol{u}) = \arg\min_{\boldsymbol{d}} \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{W}\boldsymbol{u}\|_2^2 + \Big\|\boldsymbol{\alpha}\cdot \boldsymbol{d}\Big\|_{1,\wp}, \quad \wp = 1, 2,$$

where $\boldsymbol{\alpha}\cdot\boldsymbol{d} := \big\{\alpha_{\ell,\boldsymbol{n}} d_{\ell,\boldsymbol{n}} :\ \boldsymbol{n} \in \mathbb{Z}^2, 0 \leq \ell \leq L\big\}$,

$$\|\boldsymbol{d}\|_{1,1} := \sum_{\ell,\boldsymbol{n}} |d_{\ell,\boldsymbol{n}}| \quad \text{and} \quad \|\boldsymbol{d}\|_{1,2} = \sum_{\boldsymbol{n}}\sum_{l=1}^{m}\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=l} |d_{\ell',\boldsymbol{n}}|^2\Big)^{\frac{1}{2}}. \tag{2.34}$$

Here we regrouped $\boldsymbol{\beta}_\ell, 1 \leq \ell \leq L$, according to the order of vanishing moments $|\boldsymbol{\beta}_\ell|$ of highpass filter $q^{(\ell)}$, and we assume the largest number among $|\boldsymbol{\beta}_1|, \cdots, |\boldsymbol{\beta}_L|$ is m.

When we have a tight frame system, i.e. $\widetilde{\boldsymbol{W}} = \boldsymbol{W}$, by [9, Propositon 3], we have

$$\frac{1}{2}\|\boldsymbol{d} - \boldsymbol{W}\boldsymbol{u}\|_2^2 = \frac{1}{2}\|\boldsymbol{W}^\top\boldsymbol{d} - \boldsymbol{u}\|_2^2 + \frac{1}{2}\|(I - \boldsymbol{W}\boldsymbol{W}^\top)\boldsymbol{d}\|_2^2. \tag{2.35}$$

Therefore, the iteration (2.31) has the following optimality property

$$\boldsymbol{u}^k = \begin{cases} \widetilde{\boldsymbol{W}}^\top\Big[\arg\min_{\boldsymbol{d}} \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{W}\boldsymbol{u}^{k-1}\|_2^2 + \frac{1}{2}\Big\|\sqrt{\frac{\boldsymbol{\alpha}^{k-1}}{1-\boldsymbol{\alpha}^{k-1}}}\cdot\boldsymbol{d}\Big\|_2^2\Big] & \text{bi-frame} \\ \boldsymbol{W}^\top\Big[\arg\min_{\boldsymbol{d}} \frac{1}{2}\|\boldsymbol{W}^\top\boldsymbol{d} - \boldsymbol{u}^{k-1}\|_2^2 + \frac{1}{2}\|(I - \boldsymbol{W}\boldsymbol{W}^\top)\boldsymbol{d}\|_2^2 + \frac{1}{2}\Big\|\sqrt{\frac{\boldsymbol{\alpha}^{k-1}}{1-\boldsymbol{\alpha}^{k-1}}}\cdot\boldsymbol{d}\Big\|_2^2\Big] & \text{tight frame.} \end{cases} \tag{2.36}$$

Similarly, the iteration (2.32) has the following optimality property

$$\boldsymbol{u}^k = \begin{cases} \widetilde{\boldsymbol{W}}^\top\Big[\arg\min_{\boldsymbol{d}} \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{W}\boldsymbol{u}^{k-1}\|_2^2 + \big\|\boldsymbol{\alpha}^{k-1}\cdot\boldsymbol{d}\big\|_{1,\wp}\Big] & \text{bi-frame} \\ \boldsymbol{W}^\top\Big[\arg\min_{\boldsymbol{d}} \frac{1}{2}\|\boldsymbol{W}^\top\boldsymbol{d} - \boldsymbol{u}^{k-1}\|_2^2 + \frac{1}{2}\|(I - \boldsymbol{W}\boldsymbol{W}^\top)\boldsymbol{d}\|_2^2 + \big\|\boldsymbol{\alpha}^{k-1}\cdot\boldsymbol{d}\big\|_{1,\wp}\Big] & \text{tight frame.} \end{cases} \tag{2.37}$$

Judging from the formulas of the shrinkage operator, the parameter $\boldsymbol{\alpha}^{k-1}$ depends only on the wavelet frame coefficients $\boldsymbol{d}^{k-1}$. The first optimization problem in (2.37) is a synthesis based model [24, 36, 37, 39, 40] and the second optimization problem is a balanced model [7, 18].

2.6 Wavelet Frame Shrinkage and Nonlinear Evolution PDEs

In this subsection, we discuss how wavelet frame shrinkage is related to nonlinear evolution PDEs in general. We shall focus on motivation of such connection and leave the details to later sections. We will also discuss the difference and relation between our approach and some of the existing approaches, i.e. finite difference and wavelet Galerkin method.

In this paper, all evolution PDEs we shall consider take the following general form

$$u_t = \sum_{\ell=1}^{L} \frac{\partial^{\boldsymbol{\alpha}_\ell}}{\partial x^{\boldsymbol{\alpha}_\ell}}\Phi_\ell(\boldsymbol{D}u, u), \quad \text{with } \boldsymbol{D}u = (\frac{\partial^{\boldsymbol{\beta}_1}}{\partial x^{\boldsymbol{\beta}_1}}, \ldots, \frac{\partial^{\boldsymbol{\beta}_L}}{\partial x^{\boldsymbol{\beta}_L}}), \tag{2.38}$$

where we assume the PDE is defined on $\mathbb{R}^2$ for the moment, and $|\boldsymbol{\alpha}_\ell|, |\boldsymbol{\beta}_\ell| \geq 0$ for all $1 \leq \ell \leq L$. Note that (2.38) includes diffusion, hyperbolic and Hamilton-Jacobi equations as special cases. In this paper, we shall focus on nonlinear diffusions. However, we will also discuss a nonlinear hyperbolic equation used for image restoration called shock filter [51].

Given a pair of bi-frame filter bank $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}}, \widetilde{\boldsymbol{q}}^{(1)}, \ldots, \widetilde{\boldsymbol{q}}^{(L)}\}$, and their associated transforms $\boldsymbol{W}$ and $\widetilde{\boldsymbol{W}}$, we have $\widetilde{\boldsymbol{W}}^\top \boldsymbol{W} = \boldsymbol{I}$. We split the low and high frequency component of $\widetilde{\boldsymbol{W}}$ and $\boldsymbol{W}$ according to the notation in (2.7) as

$$\boldsymbol{W}_{\text{low}} = \boldsymbol{W}_0 \quad \text{and} \quad \boldsymbol{W}_{\text{high}} = \{\boldsymbol{W}_\ell : \ 1 \leq \ell \leq L\}.$$

We define $\widetilde{\boldsymbol{W}}_{\text{low}}$ and $\widetilde{\boldsymbol{W}}_{\text{high}}$ similarly. Assuming that $\boldsymbol{q}^{(\ell)}$ and $\widetilde{\boldsymbol{q}}^{(\ell)}$ having vanishing moment of order $\boldsymbol{\beta}_\ell$ and $\boldsymbol{\alpha}_\ell$ respectively, our key observation is based on the following relations which can be derived using Lemma 2.1 by choosing $\boldsymbol{\lambda}$ and $\widetilde{\boldsymbol{\lambda}}$ properly:

$$\boldsymbol{\lambda} \cdot \boldsymbol{W}_{\text{high}} \boldsymbol{u} \approx (\boldsymbol{D} u)\big| \quad \text{and} \quad \left(\widetilde{\boldsymbol{\lambda}} \cdot \widetilde{\boldsymbol{W}}_{\text{high}}\right)^\top \cdot \left(\boldsymbol{v}_\ell\right)_{\ell=1}^L \approx -\Big(\sum_{\ell=1}^L \frac{\partial^{\boldsymbol{\alpha}_\ell}}{\partial x^{\boldsymbol{\alpha}_\ell}} v_\ell\Big)\Big|, \tag{2.39}$$

where $\boldsymbol{u} = u|$ and $\boldsymbol{v}_\ell = (v_\ell)|$ for some smooth functions u and v_ℓ, $f|$ denotes the restriction of f on $\mathbb{Z}^2$, $\boldsymbol{\lambda} \cdot \boldsymbol{W}_{\text{high}} \boldsymbol{u} = \{\lambda_\ell \boldsymbol{W}_\ell \boldsymbol{u} : 1 \leq \ell \leq L\}$, and $\big(\widetilde{\boldsymbol{\lambda}} \cdot \widetilde{\boldsymbol{W}}_{\text{high}}\big)^\top \cdot (\boldsymbol{v}_\ell)_{\ell=1}^L = \sum_{\ell=1}^L \widetilde{\lambda}_\ell \widetilde{\boldsymbol{W}}_\ell^\top \boldsymbol{v}_\ell$. Note that the first approximation of (2.39) was observed earlier in [8], while both approximations of (2.39) in the 1-D setting was used in [47]. Using the observation (2.39), we can discretize (2.38) as

$$\widetilde{\boldsymbol{u}}^k = \widetilde{\boldsymbol{u}}^{k-1} - \tau \left(\widetilde{\boldsymbol{\lambda}} \cdot \widetilde{\boldsymbol{W}}_{\text{high}}\right)^\top \cdot \left(\Phi_\ell(\boldsymbol{\lambda} \cdot \boldsymbol{W}_{\text{high}} \widetilde{\boldsymbol{u}}^{k-1}, \widetilde{\boldsymbol{u}}^{k-1})\right)_{\ell=1}^L, \tag{2.40}$$

where $\widetilde{\boldsymbol{u}}_{\boldsymbol{j}}^k$ denotes an approximation to the value $u(h\boldsymbol{j}, \tau k)$ of $u(\boldsymbol{x}, t)$ at $(h\boldsymbol{j}, \tau k)$, where h and τ are the spatial step size and the time step size. Recall from Proposition 2.1 and the iterative shrinkage formula (2.26) that follows, we have the following general expression of iterative wavelet frame shrinkage algorithm

$$\boldsymbol{u}^k = \boldsymbol{u}^{k-1} - \widetilde{\boldsymbol{W}}_{\text{high}}^\top \cdot \left[\boldsymbol{W}_{\text{high}} \boldsymbol{u}^{k-1} - \left(\mathcal{S}_\ell(\boldsymbol{W}_{\text{high}} \boldsymbol{u}^{k-1})\right)_{\ell=1}^L\right]. \tag{2.41}$$

Comparing (2.40) with (2.41), we will show that, if the shrinkage operator $\mathcal{S}_\ell$ is properly chosen, the iterative wavelet frame shrinkage algorithm will match (2.40) in the sense that $\widetilde{\boldsymbol{u}}^k = \boldsymbol{u}^k$ for all $k \geq 1$ as long as $\widetilde{\boldsymbol{u}}^0 = \boldsymbol{u}^0$. This implies that (2.41) and its equivalent operator form (2.31) is a discrete approximation of the evolution PDE (2.38) if the shrinkage operator is properly chosen. On the other hand, for some type of given shrinkage, such as the soft-thresholding operator, we will work our way backwards and find the associated (new) evolution PDEs. The main ideas here will be followed and carried out in details in Section 3 and Section 4 to establish the connections between the nonlinear evolution PDEs (especially the nonlinear diffusions) and wavelet frame based approach for image restorations. Furthermore, in Section 5, we will discuss how can we generalize the relation between (2.38) and (2.41) and how new evolution PDEs and iterative wavelet frame shrinkage algorithms can be created.

2.6.1 Difference from finite difference approach

The main advantage of wavelet frame based discretization of PDEs is that it equips the nonlinear diffusion PDE approach with a space-frequency analysis and multi-scale analysis through the multiresolution analysis associated to wavelet frames. This is impossible to achieve by other finite difference methods. This provides a new angle to understand the nonlinear PDE approach. It further provides new types of nonlinear PDEs motivated from wavelet frame based approach. On the other hand, as we will see in this paper, the PDE based approach in turns, gives wavelet frame based approach a geometric explanation and motivates us to develop new wavelet frame based methods which are different from those available wavelet frame methods which are mainly based on the sparse approximation of underlying solutions in wavelet domain. We elaborate some of details here.

Judging from (2.39) and (2.40), the discretization by wavelet frame transform resembles that of the finite difference methods normally used for numerical PDEs. However, some properties that are unique to wavelet frame systems make the discretization have better structures and handy to use. Such properties include, for instance, the associated multiresolution analysis to each wavelet frame system and the perfect reconstruction property, i.e. $\widetilde{\boldsymbol{W}}^\top \boldsymbol{W} = \boldsymbol{I}$, of wavelet frames. In addition, the specific way of sampling associated to a wavelet frame systems makes the discretization by wavelet frame transform (2.39) merely the computation of sampled derivatives via a simple wavelet decomposition algorithm that takes the samples from function values as input. In other words, the discretization by (2.39) links the samples of derivatives directly with wavelet frame coefficients. This, in turn, makes it possible to detect singularities in solutions via wavelet analysis, which is very important in solving PDEs for image restoration and is different from the standard finite difference methods. All these properties of wavelet frames enable them to outperform standard finite differencing in image restoration, which is supported by our numerical simulations in Section 7 and some of early studies (see e.g. [13,28]). Here, we further elaborate these differences between wavelet frame transform and finite differencing.

1. Finite difference methods and our approach are different in how discrete data is sampled from the unknown function. For a standard finite difference method, discrete data $\boldsymbol{u}$ is sampled from its continuum counterpart u normally by $\boldsymbol{u} = R_h u$, where R_h is the restriction of u on a certain grid with meshsize h. However, the wavelet frame based approach samples u using the associated refinable function, that generates the underlying MRA of the wavelet frame system, as $\boldsymbol{u} = T_h u$ with $(T_h u)_{\boldsymbol{k}} := 2^n \langle u, \phi_{n,\boldsymbol{k}} \rangle$ (see (6.9)). The sampling used by wavelet frame based approach is more general and better than that used by finite difference methods in the following sense.
    (a) When u is continuous, the two samplings are equal to each other asymptotically (see Lemma 6.1 for details).
    (b) When $u \in L_2(\mathbb{R}^2)$, the sampling $T_h u$ is still well defined, while $R_h u$ is not. More important, with the samples of u, the sampled values of various differentiations of u can be obtained by applying a standard wavelet decomposition on $T_h u$. In this way, we are able to directly link the sampled values of derivatives to wavelet frame coefficients, which can be used to analyze various properties of underlying solution, e.g. singularities.
2. For standard finite difference approximation of the operator $\boldsymbol{D} = \left(\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial x^{\boldsymbol{\beta}_\ell}} \right)_{\ell=1}^{L}$ (differentiation) and $\sum_{\ell=1}^{L} \frac{\partial^{\boldsymbol{\alpha}_\ell}}{\partial x^{\boldsymbol{\alpha}_\ell}}$ (divergence), there is not a bi-frame structure for the finite differencing of differentiation and divergence. However, when we use wavelet frame transform to discretize differentiation and divergence, $\boldsymbol{\lambda} \cdot \boldsymbol{W}_{\text{high}}$ approximates the differentiation and $\widetilde{\boldsymbol{\lambda}} \cdot \widetilde{\boldsymbol{W}}_{\text{high}}^\top$ approximates the divergence. By augmenting the associated low pass filter, we always have $\widetilde{\boldsymbol{W}}^\top \boldsymbol{W} = \boldsymbol{I}$. This property is used extensively, and without it, we will not be able to rewrite the standard iterative wavelet frame shrinkage algorithm (2.22) as (2.23) (see the proof of Proposition 2.1). In other words, we will not have the general iterative wavelet frame shrinkage algorithm (2.41) that can be used to link with (2.40). Consequently, we will not have the optimality relation between two consecutive iterations as given in (2.36) and (2.37). Note that if we start with a certain finite difference scheme, we can complete the corresponding system to a certain bi-frame system. Therefore, implicitly, the finite difference method is also frame based, although it may not have the MRA structure and wavelet frames associate to it. However, the completion to a bi-frame system is not considered by finite difference methods.
3. Finally, the advantage of the discretization by wavelet frame transforms over the finite difference methods is the multiresolution structure of wavelet frames, or the multiple decomposition levels of wavelet frame transform. This automatically casts a multiscale analysis and space-frequency analysis to the PDE based approach. The multilevel decomposition allows us to detect singularities of the underly solution and its directives at the presence of noise. The ability of detecting singularities enables us to activate wavelet frame bands and the associated shrinkage algorithms adaptively according to the orders of singularities. These advantages of discretization by wavelet frame transforms over the finite difference methods will be illustrated in our numerical simulations in Section 7 (Tables 1 and 2), where multiple decomposition levels are used.

2.6.2 Difference from wavelet-Galerkin

The basic idea of wavelet Galerkin method [1, 43, 48, 53] is to represent the solution by a linear combination of a refinable function ϕ for an orthonormal wavelet system. One can form the weak form of the given PDE using refinable functions as test functions, and then one can obtain a linear system with a stiffness matrix, or an iterative scheme associated to the linear system, which has entries like $\langle \frac{\partial \phi^{\boldsymbol{\alpha}}}{\partial x^{\boldsymbol{\alpha}}}, \frac{\partial^{\boldsymbol{\beta}} \phi}{\partial x^{\boldsymbol{\beta}}} \rangle$ called connection coefficients. The value of u is sampled by a refinable function ϕ at resolution level n as $\langle u, \phi_{n,\boldsymbol{k}} \rangle$ and its differentiation of $\boldsymbol{D}u$ is sampled (up to a scaling) by a refinable function ϕ at resolution level n as $\langle u, \boldsymbol{D}^\top \phi_{n,\boldsymbol{k}} \rangle$ with $\boldsymbol{D}^\top$ the adjoint operator of $\boldsymbol{D}$. The key idea of wavelet Galerkin method is to make the stiffness matrix sparse in wavelet domain. This idea of making the stiffness matrix sparse is used in almost all other Galerkin methods. In order to make this idea work, the wavelet system used has to be close in some way to the eigenfunctions of the differential operator. However, it is hard to design such a wavelet system in general.

Our approach here never tries to make the stiffness matrix sparse. Instead, our methods are based on the properties of the underlying solutions. One example of such properties is that the underlying solutions for many differential equations considered in this paper are piecewise smooth functions, which can be sparsely approximated by a wide range of wavelet systems. Iterative algorithms are derived by carefully designing shrinkage operators based on prior knowledge of the underlying solutions. Such prior knowledge can be obtained either from the given data or by a prior analysis of the underlying solution. For example, when soft-thresholding is used, we assume that the underly solution has a good sparse approximation in wavelet frame domain.

Finally, to solve a given PDE using wavelet Galerkin method, one needs to compute the connection coefficients first, which is an additional complication in solving PDEs, especially given the fact that many refinable functions (except for B-splines) are defined by their refinement masks and do not have analytic forms. Comparing to the wavelet Galerkin method, our approach (2.41) (or (2.40)) using wavelet frame transforms is much easier to implement in practice. We can use the exact same algorithmic structure to solve a large class of (nonlinear) PDEs (e.g. the one in the form (2.38)). We do not need to alter the fast wavelet frame transforms, while we simply need to choose the right shrinkage operator and proper parameters $\boldsymbol{\lambda}$ and $\widetilde{\boldsymbol{\lambda}}$, which can be calculated analytically. In other words, the easiness in implementation and flexibility in solving general nonlinear PDEs are the major advantage of our approach over the wavelet Galerkin method.

3 Wavelet Frame Shrinkage for Nonlinear Evolution Equations

In this section, we focus on the relation between the iterative multiplicative wavelet frame shrinkage and nonlinear diffusion equations. A nonlinear hyperbolic equation known as the shock filters [51] will also be studied, which can be casted (formally) into the form of a nonlinear diffusion. In fact, the arguments we use to link wavelet frame shrinkage with nonlinear diffusions can be applied to general nonlinear evolution equations (2.38). Therefore, in the rest of this paper, we will most focus on nonlinear diffusions.

We will show that, with proper choice of the shrinkage functions, the wavelet frame shrinkage (2.24) (or equivalently (2.31)) is a discretization of the following nonlinear diffusion equation for $u = u(\boldsymbol{x}, t)$, $\boldsymbol{x} \in \mathbb{R}^2, t \geq 0$:

$$u_t = \sum_{\ell=1}^{L} (-1)^{1+|\boldsymbol{\alpha}_\ell|} \frac{\partial^{\boldsymbol{\alpha}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}_\ell}} \Big\{ g_\ell \Big((\frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}})^2 \Big) \frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \Big\}, \tag{3.1}$$

with $g_\ell : \mathbb{R} \mapsto \mathbb{R}^+$ smooth and f being the initial function: $u(\boldsymbol{x}, 0) = f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^2$. More precisely, let $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}}, \widetilde{\boldsymbol{q}}^{(1)}, \cdots, \widetilde{\boldsymbol{q}}^{(L)}\}$ be a pair of bi-frame filter banks. With $\boldsymbol{u}_{\boldsymbol{j}}^0 = f(h\boldsymbol{j})$, the sequence $\boldsymbol{u}_{\boldsymbol{j}}^k$ generated by (2.24) (or equivalently from (2.31)) approximates $u(h\boldsymbol{j}, k\tau)$, where h and τ are the spatial and temporal step sizes, provided that $\mathcal{S}_\ell$ in (2.24) satisfy

$$\mathcal{S}_\ell(\xi) = \xi \Big\{ 1 - \frac{\tau}{C^{(\ell)}_{\boldsymbol{\alpha}_\ell} C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}_\ell|}} g_\ell \Big(\frac{\xi^2}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}} \Big) \Big\}, \; \xi \in \mathbb{R}, 1 \leq \ell \leq L, \tag{3.2}$$

or equivalently the shrinkage operator $\boldsymbol{S}_{\boldsymbol{\alpha}}$ in (2.28) satisfy

$$S_{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}(1-\alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}})) = d_{\ell,\boldsymbol{n}}\Big(1 - \frac{\tau}{\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell} C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\alpha}_\ell|+|\boldsymbol{\beta}_\ell|}} g_\ell\big(\frac{d^2_{\ell,\boldsymbol{n}}}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}}\big)\Big), 1 \le \ell \le L, \tag{3.3}$$

where $\boldsymbol{\alpha}_\ell, \boldsymbol{\beta}_\ell$ in $\mathbb{Z}^2_+$ are the vanishing moment orders of $\widetilde{\boldsymbol{q}}^{(\ell)}, \boldsymbol{q}^{(\ell)}$, and $\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell}, C^{(\ell)}_{\boldsymbol{\beta}_\ell}$ are defined by (2.12).

In this section we will also consider the following channel-mixed nonlinear diffusion

$$u_t = \sum_{\ell=1}^{L} (-1)^{1+|\boldsymbol{\alpha}_\ell|} \frac{\partial^{\boldsymbol{\alpha}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}_\ell}} \Big\{ g_\ell\big(\frac{\partial^{\boldsymbol{\beta}_1} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_1}}, \frac{\partial^{\boldsymbol{\beta}_2} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_2}}, \cdots, \frac{\partial^{\boldsymbol{\beta}_L} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_L}}\big) \frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \Big\}, \tag{3.4}$$

with $g_\ell : \mathbb{R}^L \mapsto \mathbb{R}^+$ smooth, and show how to design some tight frame and bi-frame filter banks and choose appropriate shrinking functions/opertaor such that the iterative frame shrinkage (2.27)/(2.31) is a discretization of (3.4).

Finally, applying similar techniques, we will show how can we design tight frame and bi-frame filter banks and the associated shrinking functions/opertaors to discretize the Perona-Malik equation [52], Bertalmio-Sapiro-Caselles-Ballester's image inpainting diffusion [5] and a nonlinear hyperbolic equation known as the Osher-Rudin's shock filter [51].

3.1 Shrinkage for Nonlinear Diffusions

In this subsection, we discuss the correspondence between frame shrinkage and high-order nonlinear diffusion. Let $\boldsymbol{p}, \boldsymbol{q}^{(\ell)}, 1 \le \ell \le L$ be a tight frame filter bank. Suppose $\boldsymbol{q}^{(\ell)}, 1 \le \ell \le L$ have vanishing moments of orders $\boldsymbol{\beta}_\ell$ with $C^{(\ell)}_{\boldsymbol{\beta}_\ell} \neq 0$, where $C^{(\ell)}_{\boldsymbol{\beta}_\ell}$ are the constants defined by (2.12) with $\boldsymbol{q} = \boldsymbol{q}^{(\ell)}$ respectively. We will show that, with properly chosen shrinkage functions, the iterative shrinkage (2.24) or equivalently (2.31) is a discretization of the high order nonlinear diffusion equation

$$u_t = \sum_{\ell=1}^{L} (-1)^{1+|\boldsymbol{\beta}_\ell|} \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \Big\{ g_\ell\Big(\big(\frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}\big)^2\Big) \frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \Big\}, \tag{3.5}$$

for $u = u(\boldsymbol{x}, t)$, $\boldsymbol{x} \in \mathbb{R}^2, t \ge 0$ with $u(\boldsymbol{x}, 0) = f(\boldsymbol{x})$ and some smooth diffusivity function $g_\ell : \mathbb{R} \mapsto \mathbb{R}^+$.

Recall that h and τ denote the spatial step size and the time step size. Lemma 2.1 ensures that we can use FIR filter $\boldsymbol{q}^{(\ell)}$ to approximate partial derivatives $\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} u(\boldsymbol{x}, t)$ and $\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} G_\ell(\boldsymbol{x}, t)$, where

$$G_\ell(\boldsymbol{x}, t) := g_\ell\Big(\big(\frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}\big)^2\Big) \frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}.$$

Indeed,

$$\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} u(h\boldsymbol{j}, \tau k) \approx \frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}} \frac{1}{h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n} \in \mathbb{Z}^2} \boldsymbol{q}^{(\ell)}[n] u(h\boldsymbol{j} + h\boldsymbol{n}, \tau k), \tag{3.6}$$

$$\begin{aligned} \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} G_\ell(h\boldsymbol{j}, \tau k) &\approx \frac{(-1)^{|\boldsymbol{\beta}_\ell|}}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}} \frac{1}{h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m} \in \mathbb{Z}^2} \boldsymbol{q}^{(\ell)}[\boldsymbol{m}] G_\ell(h\boldsymbol{j} - h\boldsymbol{m}, k\tau) \qquad (3.7)\\ &= \frac{(-1)^{|\boldsymbol{\beta}_\ell|}}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}} \frac{1}{h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m} \in \mathbb{Z}^2} \boldsymbol{q}^{(\ell)}[\boldsymbol{j} - \boldsymbol{m}] G_\ell(h\boldsymbol{m}, k\tau), \end{aligned}$$

where (2.11) with $\varepsilon = h$ and $\varepsilon = -h$ has been used in (3.6) and (3.7) respectively.

Let $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j} \in \mathbb{Z}^2$. From (3.5), (3.6) and (3.7) with $k = 0$, $\widetilde{\boldsymbol{u}}^1_{\boldsymbol{j}}, \boldsymbol{j} \in \mathbb{Z}^2$ defined by

$$\frac{\widetilde{\boldsymbol{u}}^1_{\boldsymbol{j}} - \widetilde{\boldsymbol{u}}^0_{\boldsymbol{j}}}{\tau} = -\sum_{\ell=1}^{L} \frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m}} \boldsymbol{q}^{(\ell)}[\boldsymbol{j} - \boldsymbol{m}] g_\ell\Big(\big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n}} \boldsymbol{q}^{(\ell)}[n] \widetilde{\boldsymbol{u}}^0_{\boldsymbol{n}+\boldsymbol{m}}\big)^2\Big) \Big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n}} \boldsymbol{q}^{(\ell)}[n] \widetilde{\boldsymbol{u}}^0_{\boldsymbol{n}+\boldsymbol{m}}\Big),$$

give approximated values of the solution $u(\boldsymbol{x},t)$ at $(h\boldsymbol{j},\tau), \boldsymbol{j}\in\mathbb{Z}^2$. The above equation can be rewritten as

$$\widetilde{\boldsymbol{u}}^1_{\boldsymbol{j}} = \widetilde{\boldsymbol{u}}^0_{\boldsymbol{j}} - \tau\sum_{\ell=1}^{L}\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{m}\in\mathbb{Z}^2}\boldsymbol{q}^{(\ell)}[\boldsymbol{j}-\boldsymbol{m}]g_\ell\Big(\big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\mathcal{H}^{(\ell)}_{\boldsymbol{m}}\big)^2\Big)\Big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\mathcal{H}^{(\ell)}_{\boldsymbol{m}}\Big),\ \boldsymbol{j}\in\mathbb{Z}^2, \tag{3.8}$$

where $\mathcal{H}^{(\ell)}_{\boldsymbol{m}}$ is defined by (2.22): $\mathcal{H}^{(\ell)}_{\boldsymbol{m}} = \sum_{\boldsymbol{n}}\boldsymbol{q}^{(\ell)}[\boldsymbol{n}]\widetilde{\boldsymbol{u}}^0_{\boldsymbol{n}+\boldsymbol{m}} = \sum_{\boldsymbol{n}}\boldsymbol{q}^{(\ell)}[\boldsymbol{n}]f(h(\boldsymbol{n}+\boldsymbol{m}))$.

Repeating the above process, we have $\widetilde{\boldsymbol{u}}^1_{\boldsymbol{j}},\ \widetilde{\boldsymbol{u}}^2_{\boldsymbol{j}},\cdots$, which are approximated values of the solution $u(\boldsymbol{x},t)$ at $(h\boldsymbol{j},2\tau)$, $(h\boldsymbol{j},3\tau),\cdots$ respectively. More precisely, assume we have $\widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{j}}$, approximated values of $u(\boldsymbol{x},t)$ at $(h\boldsymbol{j},\tau(k-1)),\boldsymbol{j}\in\mathbb{Z}^2$. Then from (3.6), $\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_\ell}}u(h\boldsymbol{j},\tau(k-1)) \approx 1/(C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|})\sum_{\boldsymbol{n}\in\mathbb{Z}^2}\boldsymbol{q}^{(\ell)}[n]\widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{n}+\boldsymbol{j}}$. This, together with (3.7), implies that

$$\begin{aligned}&\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_\ell}}G_\ell(h\boldsymbol{j},\tau(k-1))\\ &\approx \frac{(-1)^{|\boldsymbol{\beta}_\ell|}}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}}\frac{1}{h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{m}}\boldsymbol{q}^{(\ell)}[\boldsymbol{j}-\boldsymbol{m}]g_\ell\Big(\big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{n}}\boldsymbol{q}^{(\ell)}[n]\widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{n}+\boldsymbol{m}}\big)^2\Big)\Big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{n}}\boldsymbol{q}^{(\ell)}[n]\widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{n}+\boldsymbol{m}}\Big).\end{aligned}$$

Thus, $\widetilde{\boldsymbol{u}}^k_{\boldsymbol{j}},\boldsymbol{j}\in\mathbb{Z}^2$ defined by

$$\frac{\widetilde{\boldsymbol{u}}^k_{\boldsymbol{j}}-\widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{j}}}{\tau} = -\sum_{\ell=1}^{L}\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{m}}\boldsymbol{q}^{(\ell)}[\boldsymbol{j}-\boldsymbol{m}]g_\ell\Big(\big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{n}}\boldsymbol{q}^{(\ell)}[n]\widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{n}+\boldsymbol{m}}\big)^2\Big)\Big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{n}}\boldsymbol{q}^{(\ell)}[n]\widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{n}+\boldsymbol{m}}\Big),$$

give approximated values of the solution $u(\boldsymbol{x},t)$ at $(h\boldsymbol{j},\tau k),\boldsymbol{j}\in\mathbb{Z}^2$. Hence the highpass filters $q^{(\ell)}, 1\le\ell\le L$ give a discretization of (3.5), which can be rewritten as: for $\boldsymbol{j}\in\mathbb{Z}^2$,

$$\widetilde{\boldsymbol{u}}^k_{\boldsymbol{j}} = \widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{j}} - \tau\sum_{\ell=1}^{L}\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{m}}\boldsymbol{q}^{(\ell)}[\boldsymbol{j}-\boldsymbol{m}]g_\ell\Big(\big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\mathcal{H}^{(\ell),k-1}_{\boldsymbol{m}}\big)^2\Big)\Big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\ h^{|\boldsymbol{\beta}_\ell|}}\mathcal{H}^{(\ell),k-1}_{\boldsymbol{m}}\Big),\ k=1,2,\cdots, \tag{3.9}$$

where $\mathcal{H}^{(\ell),k-1}, 1\le\ell\le L$ are the highpass outputs of $\widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{j}}$ defined by (2.22) with $\boldsymbol{u}^0$ replaced by $\widetilde{\boldsymbol{u}}^{k-1}$.

Let $\boldsymbol{u}^k$ be the resulting sequences of the wavelet frame shrinkage (2.24) with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$. Comparing (2.24) and (3.9), we have that $\boldsymbol{u}^k_{\boldsymbol{j}} = \widetilde{\boldsymbol{u}}^k_{\boldsymbol{j}}, \boldsymbol{j}\in\mathbb{Z}^2$ for all $k\ge 2$ as long as $\boldsymbol{u}^0=\widetilde{\boldsymbol{u}}^0$ and

$$S_\ell(\xi) = \xi\Big\{1-\frac{\tau}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2h^{2|\boldsymbol{\beta}_\ell|}}g_\ell\Big(\frac{\xi^2}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2h^{2|\boldsymbol{\beta}_\ell|}}\Big)\Big\},\ \xi\in\mathbb{R},\ 1\le\ell\le L. \tag{3.10}$$

If $\boldsymbol{u}^k$ is generated from (2.31), the equivalent operator form of (2.24), then condition (3.10) can be translated equivalently to

$$S_{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}(1-\alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}})) = d_{\ell,\boldsymbol{n}}\Big(1-\frac{\tau}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2h^{2|\boldsymbol{\beta}_\ell|}}g_\ell\big(\frac{d^2_{\ell,\boldsymbol{n}}}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2h^{2|\boldsymbol{\beta}_\ell|}}\big)\Big), 1\le\ell\le L. \tag{3.11}$$

In other words, the iterative wavelet frame shrinkage algorithm (2.24)/(2.31) generates the exact same sequence as (3.9) provided that $\boldsymbol{u}^0=\widetilde{\boldsymbol{u}}^0$ and (3.10)/(3.11) is satisfied, where (3.9) is in fact a discretization of the nonlinear diffusion (3.5). In the following theorem, we summarize this result on the relation between the wavelet frame shrinkage algorithm (2.24)/(2.31) and the high-order nonlinear diffusion (3.5).

Theorem 3.1. *Let $\boldsymbol{u}^k$ be the resulting sequence from the iterative wavelet frame shrinkage* (2.24)/(2.31) *with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$ and using a tight frame filter bank $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(L)}\}$ with $\boldsymbol{q}^{(\ell)}$ having vanishing moment $\boldsymbol{\beta}_\ell$. Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j},k\tau):\boldsymbol{j}\in\mathbb{Z}^2, k=1,2,\cdots\}$ with $u(\boldsymbol{x},t)$ the solution of* (3.5) *provided that the shrinkage functions satisfy* (2.24)/(3.11)*. Furthermore, if the $\boldsymbol{\alpha}$ given in* (3.11) *satisfies* (2.33)*, we have both the optimality properties in* (2.36) *hold for $\boldsymbol{u}^k$.*

Remark 3.1. *In Theorem 3.1 and many other results in this paper, the statement hat "$\boldsymbol{u}^k$ is a discrete approximation of* $\{u(h\boldsymbol{j}, k\tau) : \boldsymbol{j} \in \mathbb{Z}^2, k = 1, 2, \cdots\}$ *with* $u(\boldsymbol{x}, t)$ *the solution of a given PDE" means that the discretization by* (2.31) *is consistent with the given PDE. Since all the consistency proofs are very similar to each other, we present a detailed proof of consistency for a specific PDE later in Proposition 6.1, which can be easily modified to a proof for Theorem 3.1, as well as all other consistency results in this paper. To show that* $\boldsymbol{u}^k$ *indeed converges to the solution of the given PDE, we need it to be well-posed and the algorithm be stable in addition to consistency. A complete proof of convergence of the discretization given by* (2.31) *to a specific PDE is given in Section 6. Although the analysis is only applied to one particular PDE, one can easily see from the proofs in Section 6 that, as long as the given PDE is well-posed, we can always show convergence of* (2.31) *with properly chosen shrinkage functions.*

With Theorem 3.1, we have the following easy corollary.

Corollary 3.1. *Let* $\boldsymbol{u}^k$ *be the resulting signal from the iterative multiplicative wavelet frame shrinkage* (2.31) *with* $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j} \in \mathbb{Z}^2$ *and using a tight frame filter bank* $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(L)}\}$*, with* $L > 2$*,* $\boldsymbol{q}^{(\ell)}$ *having vanishing moment* $\boldsymbol{\beta}_\ell$*. Let* $\widetilde{L}$ *be an integer with* $2 \le \widetilde{L} < L$*. Then,* $\boldsymbol{u}^k$ *is a discrete approximation of* $\{u(h\boldsymbol{j}, k\tau) : \boldsymbol{j} \in \mathbb{Z}^2, k = 1, 2, \cdots\}$ *with* $u(\boldsymbol{x}, t)$ *the solution of*

$$u_t = \sum_{\ell=1}^{\widetilde{L}} (-1)^{1+|\boldsymbol{\beta}_\ell|} \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \Big\{ g_\ell\Big(\big(\frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}\big)^2 \Big) \frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \Big\}, \tag{3.12}$$

provided that the shrinkage operator of (2.31) *is chosen as*

$$S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}(1 - \alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}})) = \begin{cases} d_{\ell,\boldsymbol{n}} \Big\{ 1 - \frac{\tau}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}} g_\ell\Big(\frac{(d_{\ell,\boldsymbol{n}})^2}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}} \Big) \Big\}, & \text{for } 1 \le \ell \le \widetilde{L}, \\ d_{\ell,\boldsymbol{n}} \Big\{ 1 - \frac{\bar{C}^{(0)}_\ell \tau}{h^{s_\ell}} g_\ell\Big(\frac{\bar{C}^{(1)}_\ell (d_{\ell,\boldsymbol{n}})^2}{h^{2|\boldsymbol{\beta}_\ell|}} \Big) \Big\}, & \text{for } \widetilde{L} < \ell \le L, \end{cases} \tag{3.13}$$

where $\bar{C}^{(0)}_\ell, \bar{C}^{(1)}_\ell \ge 0$ *and* $s_\ell < 2|\boldsymbol{\beta}_\ell|$*. Furthermore, if the* $\boldsymbol{\alpha}$ *given in* (3.13) *satisfies* (2.33)*, we have both the optimality properties in* (2.36) *hold for* $\boldsymbol{u}^k$*.*

Remark 3.2. *Theorem 3.1 and Corollary 3.1 imply that we can either use the tight frame filter bank* $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(\widetilde{L})}\}$ *or* $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots, \boldsymbol{q}^{(L)}\}$ *with* $L > \widetilde{L}$*, to approximate the same PDE* (3.12)*. Numerically, the discretization by Corollary 3.1 can produce better image restoration results than the discretization by Theorem 3.1. This is in fact consistent with our earlier findings in [8] when we discretize variational models using B-spline tight wavelet frames.*

Similarly, when a pair of bi-frame filter banks $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}}, \widetilde{\boldsymbol{q}}^{(1)}, \cdots, \widetilde{\boldsymbol{q}}^{(L)}\}$ are used for frame shrinkage, the formulas used to discretize partial derivatives $\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} u(\boldsymbol{x}, t)$ and $\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} G_\ell(\boldsymbol{x}, t)$ in the diffusion equation (3.1), where $G_\ell(\boldsymbol{x}, t) = g_\ell\Big(\big(\frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}\big)^2 \Big) \frac{\partial^{\boldsymbol{\beta}_\ell} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}$, are

$$\begin{aligned} \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} u(h\boldsymbol{j}, \tau k) &\approx \frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}} \frac{1}{h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n} \in \mathbb{Z}^2} \boldsymbol{q}^{(\ell)}[n] u(h\boldsymbol{j} + h\boldsymbol{n}, \tau k), \\ \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} G_\ell(h\boldsymbol{j}, \tau k) &\approx \frac{(-1)^{|\boldsymbol{\beta}_\ell|}}{\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell}} \frac{1}{h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m} \in \mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j} - \boldsymbol{m}] G_\ell(h\boldsymbol{m}, \tau k), \end{aligned}$$

where $C^{(\ell)}_{\boldsymbol{\beta}_\ell}$ and $\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell}$ are the constants defined by (2.12) with $\boldsymbol{q} = \boldsymbol{q}^{(\ell)}$ and $\boldsymbol{q} = \widetilde{\boldsymbol{q}}^{(\ell)}$, respectively. Then $\widetilde{\boldsymbol{u}}^k_{\boldsymbol{j}}$ with $\widetilde{\boldsymbol{u}}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j} \in \mathbb{Z}^2$ defined by

$$\widetilde{\boldsymbol{u}}^k_{\boldsymbol{j}} = \widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{j}} - \tau \sum_{\ell=1}^{L} \frac{1}{\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m} \in \mathbb{Z}^2} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j} - \boldsymbol{m}] g_\ell\Big(\big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \mathcal{H}^{(\ell),k-1}_{\boldsymbol{m}} \big)^2 \Big) \Big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \mathcal{H}^{(\ell),k-1}_{\boldsymbol{m}} \Big), \tag{3.14}$$

for $k=1,2,\cdots$, provides a discretization of the diffusion equation (3.1), where $\mathcal{H}^{(\ell),k-1}, 1\le \ell\le L$ are the highpass outputs of $\widetilde{\boldsymbol{u}}^{k-1}$ defined by (2.22) with $\boldsymbol{u}^0$ replaced by $\widetilde{\boldsymbol{u}}^{k-1}$. Let $\boldsymbol{u}^k$ be the resulting sequences from wavelet frame shrinkage (2.24) with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$. Comparing (2.24) with (3.14), we have $\boldsymbol{u}^k = \widetilde{\boldsymbol{u}}^k$ for $k\ge 1$ if (3.2) holds. Therefore, in this case, the wavelet frame shrinkage algorithm (2.24) (or its equivalent form (2.31)) approximates the nonlinear diffusion (3.14).

Theorem 3.2. *Let $\boldsymbol{u}^k$ be the resulting sequences from wavelet frame shrinkage (2.24)/(2.31) with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$ and using a bi-frame filter bank $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}},\widetilde{\boldsymbol{q}}^{(1)},\cdots,\widetilde{\boldsymbol{q}}^{(L)}\}$ with $\boldsymbol{q}^{(\ell)}$ (resp. $\widetilde{\boldsymbol{q}}^{(\ell)}$) having vanishing moment $\boldsymbol{\beta}_\ell$ (resp. $\boldsymbol{\alpha}_\ell$). Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j},k\tau):\boldsymbol{j}\in\mathbb{Z}^2,k=1,2,\cdots\}$ with $u(\boldsymbol{x},t)$ the solution of (3.1) provided that (3.2)/(3.3) is satisfied. Furthermore, if the $\boldsymbol{\alpha}$ given in (3.3) satisfies (2.33), we have the first optimality property in (2.36) hold for $\boldsymbol{u}^k$.*

Similarly, one can easily derive the correspondence between the wavelet frame shrinkage and diffusion of (3.4). In this case, the discretization scheme for (3.4) with a bi-frame filter bank $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}},\widetilde{\boldsymbol{q}}^{(1)},\cdots,\widetilde{\boldsymbol{q}}^{(L)}\}$ is for $k=1,2,\cdots$,

$$\widetilde{\boldsymbol{u}}^k_{\boldsymbol{j}} = \widetilde{\boldsymbol{u}}^{k-1}_{\boldsymbol{j}} - \tau\sum_{\ell=1}^{L}\frac{1}{\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell}\,h^{|\boldsymbol{\beta}_\ell|}}\sum_{\boldsymbol{m}\in\mathbb{Z}^2}\widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{j}-\boldsymbol{m}]g_\ell\Big(\frac{\mathcal{H}^{(1),k-1}_{\boldsymbol{m}}}{C^{(1)}_{\boldsymbol{\beta}_1}\,h^{|\boldsymbol{\beta}_1|}},\cdots,\frac{\mathcal{H}^{(L),k-1}_{\boldsymbol{m}}}{C^{(L)}_{\boldsymbol{\beta}_L}\,h^{|\boldsymbol{\beta}_L|}}\Big)\Big(\frac{\mathcal{H}^{(\ell),k-1}_{\boldsymbol{m}}}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}\,h^{|\boldsymbol{\beta}_\ell|}}\Big), \tag{3.15}$$

where $\widetilde{\boldsymbol{u}}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$, and $\mathcal{H}^{(\ell),k-1}, 1\le\ell\le L$ are the highpass outputs of $\widetilde{\boldsymbol{u}}^{k-1}$ with $q^{(\ell)}$.

Theorem 3.3. *Let $\boldsymbol{u}^k$ be the resulting sequence from the wavelet frame shrinkage (2.27)/(2.31) with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$ and using a bi-frame filter bank $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}},\widetilde{\boldsymbol{q}}^{(1)},\cdots,\widetilde{\boldsymbol{q}}^{(L)}\}$ with $\boldsymbol{q}^{(\ell)}$ (resp. $\widetilde{\boldsymbol{q}}^{(\ell)}$) having vanishing moment $\boldsymbol{\beta}_\ell$ (resp. $\boldsymbol{\alpha}_\ell$). Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j},k\tau):\boldsymbol{j}\in\mathbb{Z}^2,k=1,2,\cdots\}$ with $u(\boldsymbol{x},t)$ the solution of (3.4) provided that the shrinkage functions S_ℓ of (2.27) satisfy*

$$S_\ell(\xi_1,\cdots,\xi_L) = \xi_\ell - \frac{\tau\xi_\ell}{\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell}C^{(\ell)}_{\boldsymbol{\beta}_\ell}h^{|\boldsymbol{\alpha}_\ell|+|\boldsymbol{\beta}_\ell|}}g_\ell\Big(\frac{\xi_1}{C^{(1)}_{\boldsymbol{\beta}_1}h^{|\boldsymbol{\beta}_1|}},\cdots,\frac{\xi_L}{C^{(L)}_{\boldsymbol{\beta}_L}h^{|\boldsymbol{\beta}_L|}}\Big),\ \xi_1,\cdots,\xi_L\in\mathbb{R}, 1\le\ell\le L, \tag{3.16}$$

or equivalently the shrinkage operator of (2.31) is chosen as

$$\begin{aligned}S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{1,\boldsymbol{n}},\cdots,d_{L,\boldsymbol{n}}) &= d_{\ell,\boldsymbol{n}}\big(1-\alpha_{\ell,\boldsymbol{n}}(d_{1,\boldsymbol{n}},\ldots,d_{L,\boldsymbol{n}})\big)\\ &= d_{\ell,\boldsymbol{n}}\Big(1-\frac{\tau}{\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell}C^{(\ell)}_{\boldsymbol{\beta}_\ell}h^{|\boldsymbol{\alpha}_\ell|+|\boldsymbol{\beta}_\ell|}}g_\ell\Big(\frac{d_{1,\boldsymbol{n}}}{C^{(1)}_{\boldsymbol{\beta}_1}h^{|\boldsymbol{\beta}_1|}},\cdots,\frac{d_{L,\boldsymbol{n}}}{C^{(L)}_{\boldsymbol{\beta}_L}h^{|\boldsymbol{\beta}_L|}}\Big)\Big),\end{aligned} \tag{3.17}$$

for $1\le\ell\le L$. Furthermore, if the $\boldsymbol{\alpha}$ given in (3.17) satisfies (2.33), we have the first optimality property in (2.36) hold for $\boldsymbol{u}^k$.

With Theorem 3.3, we have the following easy corollary.

Corollary 3.2. *Let $\boldsymbol{u}^k$ be the resulting signal from the iterative multiplicative wavelet frame shrinkage (2.31) with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$ and using a bi-frame filter bank $\{\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}},\widetilde{\boldsymbol{q}}^{(1)},\cdots,\widetilde{\boldsymbol{q}}^{(L)}\}$ with $L>\widetilde{L}\ge 2$, $\boldsymbol{q}^{(\ell)}$ (resp. $\widetilde{\boldsymbol{q}}^{(\ell)}$) having vanishing moment $\boldsymbol{\beta}_\ell$ (resp. $\boldsymbol{\alpha}_\ell$). Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j},k\tau):\boldsymbol{j}\in\mathbb{Z}^2,k=1,2,\cdots\}$ with $u(\boldsymbol{x},t)$ the solution of*

$$u_t = \sum_{\ell=1}^{\widetilde{L}}(-1)^{1+|\boldsymbol{\alpha}_\ell|}\frac{\partial^{\boldsymbol{\alpha}_\ell}}{\partial\boldsymbol{x}^{\boldsymbol{\alpha}_\ell}}\Big\{g_\ell\big(\frac{\partial^{\boldsymbol{\beta}_1}u}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_1}},\frac{\partial^{\boldsymbol{\beta}_2}u}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_2}}\big)\frac{\partial^{\boldsymbol{\beta}_\ell}u}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_\ell}}\Big\}, \tag{3.18}$$

provided that the shrinkage operator of (2.31) is chosen as

$$\begin{aligned}S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) &= d_{\ell,\boldsymbol{n}}\big(1-\alpha_{\ell,\boldsymbol{n}}(d_{1,\boldsymbol{n}},\ldots,d_{L,\boldsymbol{n}})\big)\\ &= \begin{cases} d_{\ell,\boldsymbol{n}}\left(1-\frac{\tau}{\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell}C^{(\ell)}_{\boldsymbol{\beta}_\ell}h^{|\boldsymbol{\alpha}_\ell|+|\boldsymbol{\beta}_\ell|}}g_\ell\left(\frac{d_{1,\boldsymbol{n}}}{C^{(1)}_{\boldsymbol{\beta}_1}h^{|\boldsymbol{\beta}_1|}},\frac{d_{2,\boldsymbol{n}}}{C^{(2)}_{\boldsymbol{\beta}_2}h^{|\boldsymbol{\beta}_2|}}\right)\right), & \text{for } 1\le\ell\le\widetilde{L},\\ d_{\ell,\boldsymbol{n}}\left(1-\frac{\bar{C}^{(0)}_\ell\tau}{h^{s_\ell}}g_\ell\left(\frac{\bar{C}^{(1)}_\ell d_{1,\boldsymbol{n}}}{h^{|\boldsymbol{\beta}_1|}},\frac{\bar{C}^{(2)}_\ell d_{2,\boldsymbol{n}}}{h^{|\boldsymbol{\beta}_2|}},\ldots,\frac{\bar{C}^{(L)}_\ell d_{L,\boldsymbol{n}}}{h^{|\boldsymbol{\beta}_L|}}\right)\right), & \text{for } \widetilde{L}<\ell\le L,\end{cases}\end{aligned} \tag{3.19}$$

where $\bar{C}_\ell^{(j)} \geq 0$, $j = 0, 1, \ldots, L$, and $s_\ell < |\boldsymbol{\alpha}_\ell| + |\boldsymbol{\beta}_\ell|$. Furthermore, if the $\boldsymbol{\alpha}$ given in (3.19) satisfies (2.33), we have the first optimality property in (2.36) hold for $\boldsymbol{u}^k$.

Remark 3.3. *Theorem 3.3 and Corollary 3.2 imply that we can either use a bi-frame filter bank $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(\widetilde{L})}\}$ and $\{\widetilde{\boldsymbol{p}}, \widetilde{\boldsymbol{q}}^{(1)}, \cdots, \widetilde{\boldsymbol{q}}^{(\widetilde{L})}\}$ or $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(L)}\}$ and $\{\widetilde{\boldsymbol{p}}, \widetilde{\boldsymbol{q}}^{(1)}, \ldots, \widetilde{\boldsymbol{q}}^{(L)}\}$ with $L > \widetilde{L}$, to approximate the same PDE (3.18). Numerically, the discretization by Corollary 3.2 can produce better image restoration results than the discretization by Theorem 3.3.*

3.2 PDE Models for Image Restoration

In this subsection we show how the iterative wavelet frame shrinkage can be used to discretize some PDE models commonly used in image restoration. We will consider the Perona-Malik equation [52] and a modified TV equation, Bertalmio-Sapiro-Caselles-Ballester's image inpainting diffusion [5] and Osher-Rudin's shock filter [51].

1. Perona-Malik's and TV models

The most commonly used diffusion equation is

$$u_t = \operatorname{div}\big(g(|\nabla u|^2)\nabla u\big), \tag{3.20}$$

namely,

$$u_t = \frac{\partial}{\partial x_1}\Big\{g\Big(\big(\frac{\partial u}{\partial x_1}\big)^2 + \big(\frac{\partial u}{\partial x_2}\big)^2\Big)\frac{\partial u}{\partial x_1}\Big\} + \frac{\partial}{\partial x_2}\Big\{g\Big(\big(\frac{\partial u}{\partial x_1}\big)^2 + \big(\frac{\partial u}{\partial x_2}\big)^2\Big)\frac{\partial u}{\partial x_2}\Big\},$$

where g is the diffusivity function. If $g(x^2) = \frac{c}{1+(x/\lambda)^2}$, where $c > 0, \lambda > 0$ are constants, (3.20) is Perona-Malik's model; while it is the (modified) TV model if $g(x^2) = \frac{c}{\sqrt{\varepsilon^2+x^2}}$, where $c > 0, \varepsilon > 0$ are constants.

Let $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}\}$ and $\{\widetilde{\boldsymbol{p}}, \widetilde{\boldsymbol{q}}^{(1)}, \widetilde{\boldsymbol{q}}^{(2)}\}$ be a pair of bi-frame filter banks defined by (2.19). Then $\boldsymbol{q}^{(1)}, \widetilde{\boldsymbol{q}}^{(1)}$ have vanishing moment of order $(1,0)$ and $\boldsymbol{q}^{(2)}, \widetilde{\boldsymbol{q}}^{(2)}$ have vanishing moment of order $(0,1)$ with $C_{(1,0)}^{(1)} = \widetilde{C}_{(1,0)}^{(1)} = C_{(0,1)}^{(2)} = \widetilde{C}_{(0,1)}^{(2)} = -\frac{1}{2}$. Thus, by Theorem 3.3 with $g_1(\xi_1, \xi_2) = g_2(\xi_1, \xi_2) = g(\xi_1^2 + \xi_2^2)$, if

$$S_1(\xi_1, \xi_2) = \xi_1 - \frac{4\tau\xi_1}{h^2}g\big(\frac{4(\xi_1^2 + \xi_2^2)}{h^2}\big),\ S_2(\xi_1, \xi_2) = \xi_2 - \frac{4\tau\xi_2}{h^2}g\big(\frac{4(\xi_1^2 + \xi_2^2)}{h^2}\big), \tag{3.21}$$

or equivalently

$$S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{1,\boldsymbol{n}}, d_{2,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}\left(1 - \frac{4\tau}{h^2}g\left(\frac{4(d_{1,\boldsymbol{n}})^2 + 4(d_{2,\boldsymbol{n}})^2}{h^2}\right)\right), \quad \ell = 1, 2, \tag{3.22}$$

then the iterative channel-mixed frame shrinkage (2.27)/(2.31) results in a discretization of Perona-Malik equation.

2. Image inpainting diffusion

The (slightly modified) diffusion equation for image inpainting is

$$u_t = \nabla(\triangle u) \cdot \nabla u^\perp + \varepsilon\triangle u, \quad (\boldsymbol{x}, t) \in D \times (0, T)$$

with initial condition $u(x, 0) = f^{ext}, \boldsymbol{x} \in \Omega$, an extension of f from D to $\Omega\backslash D$, where $\varepsilon > 0$ and $\nabla u^\perp = \begin{bmatrix} -\frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} \end{bmatrix}$.
We consider the more general equation

$$u_t = \nabla\big(g(\triangle u)\big) \cdot \nabla u^\perp + \varepsilon\triangle u, \tag{3.23}$$

where g is a smooth function on $\mathbb{R}$. Equation (3.23) can be written as

$$u_t = \frac{\partial}{\partial y}\Big(g(\triangle u)\frac{\partial u}{\partial x}\Big) + \frac{\partial}{\partial x}\Big(-g(\triangle u)\frac{\partial u}{\partial y}\Big) + \varepsilon\triangle u. \tag{3.24}$$

Thus equation (3.23), or equivalently (3.24), can be expressed in the form of (3.4) with $L = 4$,

$$\boldsymbol{\beta}_1 = (1,0),\ \boldsymbol{\beta}_2 = (0,1),\ \boldsymbol{\beta}_3 = (2,0),\ \boldsymbol{\beta}_4 = (0,2),$$
$$\boldsymbol{\alpha}_1 = (0,1),\ \boldsymbol{\alpha}_2 = (1,0),\ \boldsymbol{\alpha}_3 = (0,0),\ \boldsymbol{\alpha}_4 = (0,0),$$

and

$$g_1(\xi_1,\xi_2,\xi_3,\xi_4) = g(\xi_3+\xi_4),\ g_2(\xi_1,\xi_2,\xi_3,\xi_4) = -g(\xi_3+\xi_4),$$
$$g_3(\xi_1,\xi_2,\xi_3,\xi_4) = g_4(\xi_1,\xi_2,\xi_3,\xi_4) = -\varepsilon.$$

Therefore, in order for the wavelet frame shrinkage to be in corresponce to the diffusion equation (3.24), the analysis highpass filters $\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(4)}$ need to have vanishing moments of orders $(1,0)$, $(0,1)$, $(2,0)$, $(0,2)$ respectively. The synthesis highpass filters $\widetilde{\boldsymbol{q}}^{(1)}$ and $\widetilde{\boldsymbol{q}}^{(2)}$ should have vanishing moments of orders $(0,1)$ and $(1,0)$ respectively, and filters $\widetilde{\boldsymbol{q}}^{(3)}$ and $\widetilde{\boldsymbol{q}}^{(4)}$ are just the delta filter which have vanishing moment of $(0,0)$ order.

The bi-frame filters needed to discretize the equation (3.24) are given by (2.20). Indeed, the highpass filters $\boldsymbol{q}^{(1)},\boldsymbol{q}^{(2)},\boldsymbol{q}^{(3)},\boldsymbol{q}^{(4)}$ and $\widetilde{\boldsymbol{q}}^{(1)},\widetilde{\boldsymbol{q}}^{(2)},\widetilde{\boldsymbol{q}}^{(3)},\widetilde{\boldsymbol{q}}^{(4)}$ have vanishing moments of orders $(1,0)$, $(0,1)$, $(2,0)$, $(0,2)$ and $(0,1)$, $(1,0)$, $(0,0)$, $(0,0)$ respectively. One can easily obtain

$$C^{(1)}_{(1,0)} = -\frac{1}{2},\ C^{(2)}_{(0,1)} = \frac{1}{2},\ C^{(3)}_{(2,0)} = C^{(4)}_{(0,2)} = -\frac{1}{4}, \widetilde{C}^{(1)}_{(0,1)} = \widetilde{C}^{(2)}_{(1,0)} = -\frac{1}{2},\ \widetilde{C}^{(3)}_{(0,0)} = \widetilde{C}^{(4)}_{(0,0)} = 1.$$

Thus by Theorem 3.3, if the shrinkage functions S_ℓ of (2.27) and the diffusion function g satisfy

$$S_\ell(\xi_1,\xi_2,\xi_3,\xi_4) = \xi_\ell - \frac{4\tau\xi_\ell}{h^2}g\big(-\frac{4\xi_3+4\xi_4}{h^2}\big),\ \text{for } \ell = 1,2,$$
$$S_\ell(\xi_1,\xi_2,\xi_3,\xi_4) = \xi_\ell\big(1-\frac{4\tau\varepsilon}{h^2}\big),\ \text{for } \ell = 3,4;$$

or equivalently if the shrinkage operator $\mathcal{S}_{\boldsymbol{\alpha}}$ of (2.31) satisfies

$$S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = \begin{cases} d_{\ell,\boldsymbol{n}}\left(1-\frac{4\tau}{h^2}g\big(-\frac{4d_{3,\boldsymbol{n}}+4d_{4,\boldsymbol{n}}}{h^2}\big)\right), & \text{for } \ell = 1,2, \\ d_{\ell,\boldsymbol{n}}\left(1-\frac{4\tau\varepsilon}{h^2}\right), & \text{for } \ell = 3,4, \end{cases}$$

then, with the filters given by (2.20), the wavelet frame shrinkage (2.27)/(2.31) is a discretization of the diffusion (3.23).

3. Osher-Rudin's shock filter

Osher-Rudin's shock filter (nonlinear hyperbolic equation) is governed by

$$u_t + |\nabla u|F(\mathcal{L}(u)) = 0, \quad (\boldsymbol{x},t)\in\Omega\times(0,T), \tag{3.25}$$

with initial condition $u(x,0) = f(x), x\in\Omega$, where F is a Lipschitz continuous function satisfies

$$\begin{cases} F(0) = 0, \\ \operatorname{sign}(x)F(x) > 0,\ \text{for } x \neq 0. \end{cases}$$

The simplest example for F is $F(x) = x$. A desirable choice for $\mathcal{L}(u)$ is

$$\mathcal{L}(u) = \frac{u_x^2u_{xx} + 2u_xu_yu_{xy} + u_y^2u_{yy}}{|\nabla u|^2}, \tag{3.26}$$

which is the second derivative of u in the direction of $\mathbf{n} = \frac{\nabla u}{|\nabla u|}$. Here we consider the case $F(x) = x$ and show that the wavelet frame shrinkage is connected with the shock filtering of (3.25) with $\mathcal{L}(u)$ given by (3.26). In this case (3.25) is

$$u_t = -\frac{u_x^2u_{xx} + 2u_xu_yu_{xy} + u_y^2u_{yy}}{|\nabla u|},$$

which can be written as

$$u_t = \frac{\partial}{\partial x}\Big(g(|\nabla u|^2)\frac{\partial u}{\partial x}\Big) + \frac{\partial}{\partial y}\Big(g(|\nabla u|^2)\frac{\partial u}{\partial y}\Big) - g(|\nabla u|^2)\frac{\partial^2 u}{\partial x^2} - g(|\nabla u|^2)\frac{\partial^2 u}{\partial y^2}, \tag{3.27}$$

where $g(x) = -\sqrt{x},\ x \geq 0$. Thus equation (3.27) is an equation in the form of (3.4) with $L = 4$,

$$\boldsymbol{\beta}_1 = (1,0),\ \boldsymbol{\beta}_2 = (0,1),\ \boldsymbol{\beta}_3 = (2,0),\ \boldsymbol{\beta}_4 = (0,2),$$
$$\boldsymbol{\alpha}_1 = (1,0),\ \boldsymbol{\alpha}_2 = (0,1),\ \boldsymbol{\alpha}_3 = (0,0),\ \boldsymbol{\alpha}_4 = (0,0),$$

and $g_1 = g_2 = g_3 = g_4 = g(\xi_1^2+\xi_2^2)$. Therefore, in order that the frame shrinkage corresponds to the diffusion with equation (3.24), the analysis highpass filters $\boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(4)}$ have vanishing moments of orders $(1,0)$, $(0,1)$, $(2,0)$, $(0,2)$ respectively, while the synthesis highpass filters $\widetilde{\boldsymbol{q}}^{(1)}$ and $\widetilde{\boldsymbol{q}}^{(2)}$ should have vanishing moments of orders $(0,1)$ and $(1,0)$ respectively, and $\widetilde{\boldsymbol{q}}^{(3)}$ and $\widetilde{\boldsymbol{q}}^{(4)}$ are just the delta filter which have vanishing moment of $(0,0)$ order.

The filter bank that satisfies the above requirements is given by (2.21). Indeed, the highpass filters $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}, \boldsymbol{q}^{(4)}$ and $\widetilde{\boldsymbol{q}}^{(1)}, \widetilde{\boldsymbol{q}}^{(2)}, \widetilde{\boldsymbol{q}}^{(3)}, \widetilde{\boldsymbol{q}}^{(4)}$ have vanishing moments of orders $(1,0)$, $(0,1)$, $(2,0)$, $(0,2)$ and $(1,0)$, $(0,1)$, $(0,0)$, $(0,0)$, respectively. We have

$$C^{(1)}_{(1,0)} = C^{(2)}_{(0,1)} = -1,\ C^{(3)}_{(2,0)} = C^{(4)}_{(0,2)} = -\frac{1}{4}, \widetilde{C}^{(1)}_{(0,1)} = \widetilde{C}^{(2)}_{(1,0)} = -\frac{1}{4},\ \widetilde{C}^{(3)}_{(0,0)} = \widetilde{C}^{(4)}_{(0,0)} = 1.$$

Thus, by Theorem 3.3, if the shrinkage functions S_ℓ of (2.27) and the diffusion function g satisfy

$$S_\ell(\xi_1,\xi_2,\xi_3,\xi_4) = \xi_\ell - \frac{4\tau\xi_\ell}{h^2} g\Big(\frac{\xi_1^2+\xi_2^2}{h^2}\Big), \text{ for } \ell = 1,2,$$
$$S_\ell(\xi_1,\xi_2,\xi_3,\xi_4) = \xi_\ell + \frac{4\tau\xi_\ell}{h^2} g\Big(\frac{\xi_1^2+\xi_2^2}{h^2}\Big), \text{ for } \ell = 3,4;$$

or equivalently if the shrinkage operator $\mathcal{S}_{\boldsymbol{\alpha}}$ of (2.31) satisfies

$$S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = \begin{cases} d_{\ell,\boldsymbol{n}}\left(1 - \frac{4\tau}{h^2} g\left(\frac{(d_{1,\boldsymbol{n}})^2+(d_{2,\boldsymbol{n}})^2}{h^2}\right)\right), & \text{for } \ell = 1,2, \\ d_{\ell,\boldsymbol{n}}\left(1 + \frac{4\tau}{h^2} g\left(\frac{(d_{1,\boldsymbol{n}})^2+(d_{2,\boldsymbol{n}})^2}{h^2}\right)\right), & \text{for } \ell = 3,4, \end{cases}$$

then, with the filters given by (2.21), the sequence from the wavelet frame shrinkage (2.27)/(2.31) is a discretization of (3.27).

4 PDEs Derived from Wavelet Frame Shrinkage

In the previous section, we showed, in generic settings, how to construct wavelet frame system and choose appropriate shrinkage functions so that the wavelet frame shrinkage is a discrete approximation of a given PDE. In this section we present several specific (new) high-order diffusion equations that are derived from wavelet frame shrinkage (2.31) using some specific B-spline tight wavelet frame systems. Note that our techniques can be generalized easily to many other wavelet frame systems and other (possibly new) nonlinear evolution PDEs can be derived similarly.

4.1 Diffusions from B-spline Filter Banks

Let $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(8)}\}$ be the separable spline tight frame filter bank given in (2.15). The corresponding nonlinear diffusion equation is

$$\begin{aligned} u_t = {} & \frac{\partial}{\partial x_1}\Big\{g_1\Big(\big(\frac{\partial u}{\partial x_1}\big)^2\Big)\frac{\partial u}{\partial x_1}\Big\} + \frac{\partial}{\partial x_2}\Big\{g_2\Big(\big(\frac{\partial u}{\partial x_2}\big)^2\Big)\frac{\partial u}{\partial x_2}\Big\} - \frac{\partial^2}{\partial x_1^2}\Big\{g_3\Big(\big(\frac{\partial^2 u}{\partial x_1^2}\big)^2\Big)\frac{\partial^2 u}{\partial x_1^2}\Big\} \\ & -\frac{\partial^2}{\partial x_1\partial x_2}\Big\{g_4\Big(\big(\frac{\partial^2 u}{\partial x_1\partial x_2}\big)^2\Big)\frac{\partial^2 u}{\partial x_1\partial x_2}\Big\} - \frac{\partial^2}{\partial x_2^2}\Big\{g_5\Big(\big(\frac{\partial^2 u}{\partial x_2^2}\big)^2\Big)\frac{\partial^2 u}{\partial x_2^2}\Big\} + \frac{\partial^3}{\partial x_1^2\partial x_2}\Big\{g_6\Big(\big(\frac{\partial^3 u}{\partial x_1^2\partial x_2}\big)^2\Big)\frac{\partial^3 u}{\partial x_1^2\partial x_2}\Big\} \\ & +\frac{\partial^3}{\partial x_1\partial x_2^2}\Big\{g_7\Big(\big(\frac{\partial^3 u}{\partial x_1\partial x_2^2}\big)^2\Big)\frac{\partial^3 u}{\partial x_1\partial x_2^2}\Big\} - \frac{\partial^4}{\partial x_1^2\partial x_2^2}\Big\{g_8\Big(\big(\frac{\partial^4 u}{\partial x_1^2\partial x_2^2}\big)^2\Big)\frac{\partial^4 u}{\partial x_1^2\partial x_2^2}\Big\}, \end{aligned} \tag{4.1}$$

with the initial condition $u(\boldsymbol{x},0)=f(\boldsymbol{x}), \boldsymbol{x}\in\mathbb{R}^2$. More precisely, with

$$C^{(1)}_{\boldsymbol{\beta}_1}=C^{(2)}_{\boldsymbol{\beta}_2}=-\frac{\sqrt{2}}{2},\ C^{(3)}_{\boldsymbol{\beta}_3}=-\frac{1}{4},\ C^{(4)}_{\boldsymbol{\beta}_4}=\frac{1}{2}, C^{(5)}_{\boldsymbol{\beta}_5}=-\frac{1}{4},\ C^{(6)}_{\boldsymbol{\beta}_6}=C^{(7)}_{\boldsymbol{\beta}_7}=\frac{\sqrt{2}}{8},\ C^{(8)}_{\boldsymbol{\beta}_8}=\frac{1}{16},$$

we know, by Theorem 3.1, that the resulting signals from the wavelet frame shrinkage (2.24)/(2.31), with the filters (2.15), approximates the solution of diffusion equation (4.1) in discrete setting provided that the shrinking functions S_ℓ and the diffusivity g_ℓ have the relationship:

$$\begin{aligned}
&S_\ell(\xi)=\xi\Big\{1-\frac{2\tau}{h^2}g_\ell\big(\frac{2\xi^2}{h^2}\big)\Big\}\text{ for }\ell=1,2,\quad S_\ell(\xi)=\xi\Big\{1-\frac{16\tau}{h^4}g_\ell\big(\frac{16\xi^2}{h^4}\big)\Big\}\text{ for }\ell=3,5,\\
&S_\ell(\xi)=\xi\Big\{1-\frac{32\tau}{h^4}g_\ell\big(\frac{32\xi^2}{h^4}\big)\Big\}\text{ for }\ell=6,7,\\
&S_4(\xi)=\xi\Big\{1-\frac{4\tau}{h^4}g_\ell\big(\frac{4\xi^2}{h^4}\big)\Big\},\ S_8(\xi)=\xi\Big\{1-\frac{256\tau}{h^8}g_8\big(\frac{256\xi^2}{h^8}\big)\Big\},
\end{aligned}$$

or equivalently the shrinkage operator satisfies:

$$\begin{aligned}
&S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}})=d_{\ell,\boldsymbol{n}}\left\{1-\frac{2\tau}{h^2}g_\ell\left(\frac{2(d_{\ell,\boldsymbol{n}})^2}{h^2}\right)\right\},\ \text{for }\ell=1,2,\\
&S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}})=d_{\ell,\boldsymbol{n}}\left\{1-\frac{16\tau}{h^4}g_\ell\left(\frac{16(d_{\ell,\boldsymbol{n}})^2}{h^4}\right)\right\},\ \text{for }\ell=3,5,\\
&S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}})=d_{\ell,\boldsymbol{n}}\left\{1-\frac{32\tau}{h^6}g_\ell\left(\frac{32(d_{\ell,\boldsymbol{n}})^2}{h^6}\right)\right\},\ \text{for }\ell=6,7,\\
&S_{\boldsymbol{\alpha}_{4,\boldsymbol{n}}(\boldsymbol{d})}(d_{4,\boldsymbol{n}})=d_{4,\boldsymbol{n}}\left\{1-\frac{4\tau}{h^4}g_\ell\left(\frac{4(d_{\ell,\boldsymbol{n}})^2}{h^4}\right)\right\},\ S_{\boldsymbol{\alpha}_{8,\boldsymbol{n}}(\boldsymbol{d})}(d_{8,\boldsymbol{n}})=d_{8,\boldsymbol{n}}\left\{1-\frac{256\tau}{h^8}g_8\left(\frac{256(d_{8,\boldsymbol{n}})^2}{h^8}\right)\right\}.
\end{aligned}$$

Next we construct wavelet frame filter banks with fewer highpass filters, while the lowpass filter is still the same: $\boldsymbol{p}(\boldsymbol{\omega})=\boldsymbol{a}(\omega_1)\boldsymbol{a}(\omega_2)$. Denote

$$\xi=\big(\cos\frac{\omega_1}{2}\big)^2,\ \eta=\big(\cos\frac{\omega_2}{2}\big)^2. \tag{4.2}$$

Observe that

$$\begin{aligned}
\xi&=\Big|\frac{1+e^{-i\omega_1}}{2}\Big|^2,\quad 1-\xi=\big(\sin\frac{\omega_1}{2}\big)^2=\Big|\frac{1-e^{-i\omega_1}}{2}\Big|^2,\\
\eta&=\Big|\frac{1+e^{-i\omega_2}}{2}\Big|^2,\quad 1-\eta=\big(\sin\frac{\omega_2}{2}\big)^2=\Big|\frac{1-e^{-i\omega_2}}{2}\Big|^2,
\end{aligned}$$

and that

$$\begin{aligned}
4\xi(1-\xi)&=(\sin\omega_1)^2=\Big|\frac{e^{i\omega_1}-e^{-i\omega_1}}{2}\Big|^2,\quad (1-2\xi)^2=(\cos\omega_1)^2=\Big|\frac{e^{i\omega_1}+e^{-i\omega_1}}{2}\Big|^2,\\
4\eta(1-\eta)&=(\sin\omega_2)^2=\Big|\frac{e^{i\omega_2}-e^{-i\omega_2}}{2}\Big|^2,\quad (1-2\eta)^2=(\cos\omega_2)^2=\Big|\frac{e^{i\omega_2}+e^{-i\omega_2}}{2}\Big|^2.
\end{aligned}$$

From $|\boldsymbol{p}(\boldsymbol{\omega})|^2=\xi^2\eta^2$, and

$$\begin{aligned}
1&=\Big(\eta+(1-\eta)\Big)^2=\eta^2+2\eta(1-\eta)+(1-\eta)^2\\
&=\Big(\xi+(1-\xi)\Big)^2\eta^2+2\Big((1-2\xi)^2+4\xi(1-\xi)\Big)\eta(1-\eta)+(1-\eta)^2\\
&=\xi^2\eta^2+2\xi(1-\xi)\eta^2+(1-\xi)^2\eta^2+2(1-2\xi)^2\eta(1-\eta)+8\xi(1-\xi)\eta(1-\eta)+(1-\eta)^2,
\end{aligned}$$

we know that $\boldsymbol{q}^{(\ell)}, 1 \le \ell \le 5$ given by

$$
\begin{aligned}
&|\widehat{\boldsymbol{q}}^{(1)}(\boldsymbol{\omega})|^2 = 2\xi(1-\xi)\eta^2 = \frac{1}{2}\big(\sin\omega_1\big)^2\big(\cos\frac{\omega_2}{2}\big)^4, |\widehat{\boldsymbol{q}}^{(2)}(\boldsymbol{\omega})|^2 = 2(1-2\xi)^2\eta(1-\eta) = \frac{1}{2}\big(\cos\omega_1\big)^2\big(\sin\omega_2\big)^2,\\
&|\widehat{\boldsymbol{q}}^{(3)}(\boldsymbol{\omega})|^2 = (1-\xi)^2\eta^2 = \big(\sin\frac{\omega_1}{2}\big)^4\big(\cos\frac{\omega_2}{2}\big)^4, |\widehat{\boldsymbol{q}}^{(4)}(\boldsymbol{\omega})|^2 = 8\xi(1-\xi)\eta(1-\eta) = \frac{1}{2}\big(\sin\omega_1\big)^2\big(\sin\omega_2\big)^2,\\
&|\widehat{\boldsymbol{q}}^{(5)}(\boldsymbol{\omega})|^2 = (1-\eta)^2 = \big(\sin\frac{\omega_2}{2}\big)^4,
\end{aligned}
$$

which are the filters given by (2.16), together with the lowpass filter $\boldsymbol{p}$, form a tight frame filter bank. The highpass filters $\boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(5)}$ have the vanishing moments of orders

$$
\boldsymbol{\beta}_1 = (1,0),\ \boldsymbol{\beta}_2 = (0,1),\ \boldsymbol{\beta}_3 = (2,0),\ \boldsymbol{\beta}_4 = (1,1),\ \boldsymbol{\beta}_5 = (0,2),
$$

respectively, and one can obtain the constants $C^{(\ell)}_{\boldsymbol{\beta}_\ell}$ in (2.12) with $\boldsymbol{q} = \boldsymbol{q}^{(\ell)}$:

$$
C^{(1)}_{(1,0)} = C^{(2)}_{(0,1)} = -\frac{\sqrt{2}}{2},\ C^{(3)}_{(2,0)} = C^{(5)}_{(0,2)} = \frac{1}{4},\ C^{(4)}_{(1,1)} = \frac{\sqrt{2}}{2}. \tag{4.3}
$$

The nonlinear diffusion equation corresponding to this Ron-Shen type tight frame filter bank is

$$
\begin{aligned}
u_t = &\frac{\partial}{\partial x_1}\Big\{g_1\Big(\big(\frac{\partial u}{\partial x_1}\big)^2\Big)\frac{\partial u}{\partial x_1}\Big\} + \frac{\partial}{\partial x_2}\Big\{g_2\Big(\big(\frac{\partial u}{\partial x_2}\big)^2\Big)\frac{\partial u}{\partial x_2}\Big\} - \frac{\partial^2}{\partial x_1^2}\Big\{g_3\Big(\big(\frac{\partial^2 u}{\partial x_1^2}\big)^2\Big)\frac{\partial^2 u}{\partial x_1^2}\Big\}\\
&-\frac{\partial^2}{\partial x_1\partial x_2}\Big\{g_4\Big(\big(\frac{\partial^2 u}{\partial x_1\partial x_2}\big)^2\Big)\frac{\partial^2 u}{\partial x_1\partial x_2}\Big\} - \frac{\partial^2}{\partial x_2^2}\Big\{g_5\Big(\big(\frac{\partial^2 u}{\partial x_2^2}\big)^2\Big)\frac{\partial^2 u}{\partial x_2^2}\Big\},
\end{aligned} \tag{4.4}
$$

with $u(\boldsymbol{x}, 0) = f(\boldsymbol{x}),\ \boldsymbol{x} \in \mathbb{R}^2$. From Theorem 3.1, we have the conclusion in the following theorem.

Theorem 4.1. *Let $\boldsymbol{u}^k$ be the resulting signal from the wavelet frame shrinkage (2.24)/(2.31) with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j} \in \mathbb{Z}^2$ and using the spline tight frame filter bank $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(5)}\}$ given in (2.16). Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j}, k\tau) : \boldsymbol{j} \in \mathbb{Z}^2, k = 1, 2, \cdots\}$ with $u(\boldsymbol{x}, t)$ the solution of (4.4), provided that the shrinkage functions S_ℓ of (2.24) are chosen as*

$$
\begin{aligned}
&S_\ell(\xi) = \xi\Big\{1 - \frac{2\tau}{h^2}g_\ell\big(\frac{2\xi^2}{h^2}\big)\Big\} \text{ for } \ell = 1, 2, \quad S_\ell(\xi) = \xi\Big\{1 - \frac{16\tau}{h^4}g_\ell\big(\frac{16\xi^2}{h^4}\big)\Big\} \text{ for } \ell = 3, 5,\\
&S_4(\xi) = \xi\Big\{1 - \frac{2\tau}{h^4}g_\ell\big(\frac{2\xi^2}{h^4}\big)\Big\},\ \xi \in \mathbb{R};
\end{aligned}
$$

or equivalently the shrinkage operator $\boldsymbol{\mathcal{S}}_{\boldsymbol{\alpha}}$ of (2.31) is chosen as

$$
S_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}\big(1 - \alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}})\big) = \begin{cases} d_{\ell,\boldsymbol{n}}\left\{1 - \frac{2\tau}{h^2}g_\ell\left(\frac{2(d_{\ell,\boldsymbol{n}})^2}{h^2}\right)\right\}, & \text{for } \ell = 1, 2,\\ d_{\ell,\boldsymbol{n}}\left\{1 - \frac{16\tau}{h^4}g_\ell\left(\frac{16(d_{\ell,\boldsymbol{n}})^2}{h^4}\right)\right\}, & \text{for } \ell = 3, 5,\\ d_{\ell,\boldsymbol{n}}\left\{1 - \frac{2\tau}{h^4}g_\ell\left(\frac{2(d_{\ell,\boldsymbol{n}})^2}{h^4}\right)\right\}, & \text{for } \ell = 4.\end{cases}
$$

Furthermore, if the $\boldsymbol{\alpha}$ given above satisfies (2.33), both the optimality properties in (2.36) hold for $\boldsymbol{u}^k$.

4.2 Rotation-Invariant Diffusions

For some applications, the rotation invariant diffusion is preferred. In this subsection we will show that the iterative wavelet frame shrinkage (2.27)/(2.31) corresponds to rotation invariant diffusion if we choose the threshold $\boldsymbol{\alpha}$ for $\boldsymbol{S}_{\boldsymbol{\alpha}}$ properly. In particular, the rotation invariant diffusion equation corresponding to the spline tight frame filter banks (2.16) is

$$u_t = \frac{\partial}{\partial x_1}\Big\{g_1\Big((\frac{\partial u}{\partial x_1})^2 + (\frac{\partial u}{\partial x_2})^2\Big)\frac{\partial u}{\partial x_1}\Big\} + \frac{\partial}{\partial x_2}\Big\{g_1\Big((\frac{\partial u}{\partial x_1})^2 + (\frac{\partial u}{\partial x_2})^2\Big)\frac{\partial u}{\partial x_2}\Big\} \tag{4.5}$$
$$-\frac{\partial^2}{\partial x_1^2}\Big\{g_2\Big((\frac{\partial^2 u}{\partial x_1^2})^2 + (\frac{\partial^2 u}{\partial x_1 \partial x_2})^2 + (\frac{\partial^2 u}{\partial x_2^2})^2\Big)\frac{\partial^2 u}{\partial x_1^2}\Big\} - \frac{\partial^2}{\partial x_2^2}\Big\{g_2\Big((\frac{\partial^2 u}{\partial x_1^2})^2 + (\frac{\partial^2 u}{\partial x_1 \partial x_2})^2 + (\frac{\partial^2 u}{\partial x_2^2})^2\Big)\frac{\partial^2 u}{\partial x_2^2}\Big\}$$
$$-\frac{\partial^2}{\partial x_1 \partial x_2}\Big\{g_2\Big((\frac{\partial^2 u}{\partial x_1^2})^2 + (\frac{\partial^2 u}{\partial x_1 \partial x_2})^2 + (\frac{\partial^2 u}{\partial x_2^2})^2\Big)\frac{\partial^2 u}{\partial x_1 \partial x_2}\Big\},$$

with the initial condition $u(\boldsymbol{x}, 0) = f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^2$, where g_1 and g_2 are functions on $\mathbb{R}$.

Theorem 4.2. *Let $\boldsymbol{u}^k$ be the resulting signal from the wavelet frame shrinkage (2.27)/(2.31) with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j} \in \mathbb{Z}^2$ and using the spline type tight frame filter bank $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \cdots, \boldsymbol{q}^{(5)}\}$ given in (2.16). Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j}, k\tau) : \boldsymbol{j} \in \mathbb{Z}^2, k = 1, 2, \cdots\}$ with $u(\boldsymbol{x}, t)$ the solution of (4.5), provided that the shrinkage functions S_ℓ of (2.27) satisfy*

$$\begin{aligned} S_\ell(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5) &= \xi_\ell\Big\{1 - \frac{2\tau}{h^2}g_1\Big(\frac{2}{h^2}(\xi_1^2 + \xi_2^2)\Big)\Big\}, \ \text{for } \ell = 1, 2, \\ S_\ell(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5) &= \xi_\ell\Big\{1 - \frac{16\tau}{h^4}g_2\Big(\frac{2}{h^4}(8\xi_3^2 + \xi_4^2 + 8\xi_5^2)\Big)\Big\}, \ \text{for } \ell = 3, 5, \\ S_4(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5) &= \xi_4\Big\{1 - \frac{2\tau}{h^4}g_2\Big(\frac{2}{h^4}(8\xi_3^2 + \xi_4^2 + 8\xi_5^2)\Big)\Big\}, \end{aligned} \tag{4.6}$$

or equivalently the shrinkage operator $\mathcal{S}_{\boldsymbol{\alpha}}$ of (2.31) is chosen as

$$\mathcal{S}_{\boldsymbol{\alpha}_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}\big(1 - \alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}})\big) = \begin{cases} d_{\ell,\boldsymbol{n}}\left\{1 - \frac{2\tau}{h^2}g_\ell\left(\frac{2(d_{1,\boldsymbol{n}})^2 + 2(d_{2,\boldsymbol{n}})^2}{h^2}\right)\right\}, & \text{for } \ell = 1, 2, \\ d_{\ell,\boldsymbol{n}}\left\{1 - \frac{16\tau}{h^4}g_\ell\left(\frac{16(d_{3,\boldsymbol{n}})^2 + 2(d_{4,\boldsymbol{n}})^2 + 16(d_{5,\boldsymbol{n}})^2}{h^4}\right)\right\}, & \text{for } \ell = 3, 5, \\ d_{\ell,\boldsymbol{n}}\left\{1 - \frac{2\tau}{h^4}g_\ell\left(\frac{16(d_{3,\boldsymbol{n}})^2 + 2(d_{4,\boldsymbol{n}})^2 + 16(d_{5,\boldsymbol{n}})^2}{h^4}\right)\right\}, & \text{for } \ell = 4. \end{cases} \tag{4.7}$$

Furthermore, if the $\boldsymbol{\alpha}$ given in (4.7) satisfies (2.33), both the optimality properties in (2.36) hold for $\boldsymbol{u}^k$.

Next we construct tight filters which result in higher order rotation invariant diffusion. The scaling functions for the tight filter banks are the tensor products of the B-spline of an arbitrary order m with the two-scale symbol given by (2.17). Recall from (4.2) that ξ and η denote $\big(\cos\frac{\omega_1}{2}\big)^2$ and $\big(\cos\frac{\omega_2}{2}\big)^2$, respectively. Thus from the fact $\big|\widehat{\boldsymbol{p}}(\boldsymbol{\omega})\big|^2 = \xi^m\eta^m$ and the identity

$$\begin{aligned} 1 = \Big(\eta + (1-\eta)\Big)^m &= \sum_{r=0}^{m}\binom{m}{r}\eta^{m-r}(1-\eta)^r \\ &= \sum_{r=0}^{m}\binom{m}{r}\Big(\xi + (1-\xi)\Big)^{m-r}\eta^{m-r}(1-\eta)^r \\ &= \sum_{r=0}^{m}\binom{m}{r}\sum_{j=0}^{m-r}\binom{m-r}{j}\xi^{m-r-j}(1-\xi)^r\eta^{m-r}(1-\eta)^r \\ &= \sum_{r=0}^{m}\binom{m}{r}\sum_{s=r}^{m}\binom{m-r}{s-r}\xi^{m-s}\eta^{m-r}(1-\xi)^{s-r}(1-\eta)^r \quad (\text{by } s = r + j) \\ &= \sum_{s=0}^{m}\sum_{r=0}^{s}\binom{m}{s}\binom{s}{r}\xi^{m-s}\eta^{m-r}(1-\xi)^{s-r}(1-\eta)^r \\ &= \xi^m\eta^m + \sum_{s=1}^{m}\sum_{r=0}^{s}\binom{m}{s}\binom{s}{r}\xi^{m-s}\eta^{m-r}(1-\xi)^{s-r}(1-\eta)^r, \end{aligned}$$

we may choose tight frame highpass filters $\widehat{\boldsymbol{q}}^{(s,r)}(\boldsymbol{\omega}), 1 \le s \le m, 0 \le r \le s$ by

$$\big|\widehat{\boldsymbol{q}}^{(s,r)}(\boldsymbol{\omega})\big|^2 = \binom{m}{s}\binom{s}{r}\big(\cos\frac{\omega_1}{2}\big)^{2(m-s)}\big(\cos\frac{\omega_2}{2}\big)^{2(m-r)}\big(\sin\frac{\omega_1}{2}\big)^{2(s-r)}\big(\sin\frac{\omega_2}{2}\big)^{2r},$$

or by (2.18). Highpass filter $\boldsymbol{q}^{(s,r)}$ has vanishing moment of order $(s-r,s)$. Let $C_{s,r}$ denote the constants in (2.12) with $\boldsymbol{q} = \boldsymbol{q}^{(s,r)}$. Then it is not difficult to obtain

$$C_{s,r} = \frac{(-1)^s}{2^{2m-r}}\sqrt{\binom{m}{s}\Big/\binom{s}{r}}. \tag{4.8}$$

We consider the iterative channel-mixed shrinkage defined by

$$\boldsymbol{u}^k_{\boldsymbol{j}} = \sum_{\boldsymbol{n}\in\mathbb{Z}^2}\boldsymbol{p}[\boldsymbol{j}-\boldsymbol{n}]L^{k-1}_{\boldsymbol{n}} + \sum_{s=1}^{m}\sum_{r=0}^{s}\sum_{\boldsymbol{n}\in\mathbb{Z}^2}\boldsymbol{q}^{(s,r)}[\boldsymbol{j}-\boldsymbol{n}]S_{s,r}(H^{(s,0),k-1}_{\boldsymbol{n}}, H^{(s,2),k-1}_{\boldsymbol{n}}, \cdots, H^{(s,s),k-1}_{\boldsymbol{n}}), \tag{4.9}$$

where $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$, L^{k-1} and $H^{(s,r),k-1}$ are the lowpass and highpass outputs of $\boldsymbol{u}^{k-1}$ with lowpass filter p and highpass filters $\boldsymbol{q}^{(s,r)}$, and for an s with $1 \le s \le m$, each of $S_{s,r}, 0 \le r \le s$ is a function of $s+1$ variables. Note that (4.9) can be casted into the form of the generic iterative wavelet frame shrinkage (2.31). However, for notational convenience, we shall use the current form.

We can show as in the previous section that the tight frame filter bank given by (2.17) and (2.18) corresponds to the following general high order rotation invariant nonlinear diffusion,

$$u_t = \sum_{s=1}^{m}(-1)^{1+s}\sum_{|\boldsymbol{\alpha}|=s}\frac{\partial^{\boldsymbol{\alpha}}}{\partial\boldsymbol{x}^{\boldsymbol{\alpha}}}\Big\{g_s\Big(\sum_{|\boldsymbol{\beta}|=s}\big(\frac{\partial^{\boldsymbol{\beta}}u}{\partial\boldsymbol{x}^{\boldsymbol{\beta}}}\big)^2\Big)\frac{\partial^{\boldsymbol{\alpha}}u}{\partial\boldsymbol{x}^{\boldsymbol{\alpha}}}\Big\}, \tag{4.10}$$

where $m \ge 1$, and $g_s : \mathbb{R} \mapsto \mathbb{R}^+$ is smooth.

Theorem 4.3. *Let $\boldsymbol{u}^k$ be the resulting signal from the iterative wavelet frame shrinkage* (4.9) *with $\boldsymbol{u}^0_{\boldsymbol{j}} = f(h\boldsymbol{j}), \boldsymbol{j}\in\mathbb{Z}^2$ and using the tight frame filter bank $\{\boldsymbol{p}, \boldsymbol{q}^{(s,r)}, 1\le s\le m, 0\le r\le s\}$ given by* (2.17) *and* (2.18)*. Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j},\tau k) : \boldsymbol{j}\in\mathbb{Z}^2, k = 1,2,\cdots\}$ with $u(\boldsymbol{x},t)$ the solution of* (4.10)*, provided that*

$$S_{s,r}(\xi_0,\xi_1,\cdots,\xi_s) = \xi_r\Big\{1 - \frac{\tau}{\big(C_{s,k}\big)^2h^{2s}}g_s\Big(\frac{1}{h^{2s}}\sum_{j=0}^{s}\frac{1}{(C_{s,j})^2}\,\xi_j^2\Big)\Big\},\ r = 0,1,\cdots,s, \tag{4.11}$$

for $1 \le s \le m$.

5 Further Development

Up to this point, we have already known that given a certain nonlinear evolution equation, we can choose a wavelet frame system and the threshold $\boldsymbol{\alpha}$ in $\boldsymbol{S}_{\boldsymbol{\alpha}}$ properly such that (2.31) is an iterative finite difference scheme solving the given equation. On the other hand, new nonlinear diffusion equations can be derived using the iterative shrinkage (2.31) with certain wavelet frame filter banks and proper choices of the threshold $\boldsymbol{\alpha}$ for $\boldsymbol{S}_{\boldsymbol{\alpha}}$. In this section, we focus on the discussion of the iterative shrinkage (2.32) and also on other type of iterative shrinkage that is originated from (2.31) and (2.32). The main goal of this section is to:

1. show that the iterative wavelet frame soft-thresholding algorithms commonly used in image restoration can lead to new nonlinear diffusion equations;
2. show that by borrowing ideas from some of the algorithms used in image restoration, we can design new iterative multiplicative wavelet shrinkage algorithms which can also be understood as a discretization of a certain nonlinear diffusion equation;
3. discuss how we can borrow the idea of anisotropy of the PM nonlinear diffusion to design a new iterative wavelet frame soft-thresholding algorithm which is adaptive to local image features.

5.1 Nonlinear Diffusions from Soft-Thresholding

Throughout this subsection, we assume that $\boldsymbol{W}$ is the transform of a tight frame system, i.e. $\boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{I}$. We will find the corresponding diffusion equations to the iterative soft-thresholding algorithms in (2.32) with various choices of thresholds.

We start with the soft-thresholding operator with $\wp = 1$ given by (2.29). We will show that the diffusion equation corresponding to the iterative shrinkage algorithm (2.32) takes the form of (3.5) with some specific diffusivity functions g_ℓ. The connection between the diffusivity functions g_ℓ and the shrinkage operator is given by (3.10). In fact, the specific expression for g_ℓ can be solved from (3.10) by equating $\boldsymbol{S}_{\boldsymbol{\alpha}} = \boldsymbol{\mathcal{T}}^1_{\boldsymbol{\alpha}}$ and choosing an appropriate $\boldsymbol{\alpha}$. Indeed, for $\mathcal{T}^1_{\theta_\ell}(\xi) = \frac{\xi}{|\xi|}\max\{|\xi|-\theta_\ell, 0\}$, we have

$$\frac{\tau}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}} g_\ell\Big(\frac{\xi^2}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}}\Big) = 1 - \frac{\max\{|\xi|-\theta_\ell, 0\}}{|\xi|} = \min\Big\{1, \frac{\theta_\ell}{|\xi|}\Big\}.$$

Thus the diffusivity function g_ℓ in (3.5) satisfies

$$g_\ell(\xi^2) = \min\Big\{\frac{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}}{\tau}, \frac{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}\theta_\ell}{\tau|\xi|}\Big\}.$$

Now, we discuss how should we choose the threshold θ_ℓ properly such that when $h, \tau \to 0$, $g_\ell(\xi)$ is a function independent of h and τ. Note that it is reasonable to assume that there exists $C > 0$ such that $h^{2m}/\tau = C$ (see Proposition 6.2 in Section 6), where m is the largest number among $|\boldsymbol{\beta}_1|, \cdots, |\boldsymbol{\beta}_L|$. If we choose

$$\theta_\ell = \frac{\tau|\xi|}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}}\, \widetilde{g}_\ell(\xi^2), \quad \text{with some smooth function } \widetilde{g}_\ell : \mathbb{R} \mapsto \mathbb{R}^+,$$

then, we have the following formula for the diffusivity functions g_ℓ,

$$g_\ell(\xi^2) = \min\Big\{\frac{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 C}{h^{2(m-|\boldsymbol{\beta}_\ell|)}}, \widetilde{g}_\ell(\xi)\Big\} = \begin{cases} \min\Big\{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 C, \widetilde{g}_\ell(\xi^2)\Big\}, & \text{for } |\boldsymbol{\beta}_\ell| = m, \\ \widetilde{g}_\ell(\xi^2), & \text{for } 1 \le |\boldsymbol{\beta}_\ell| < m, \end{cases} \tag{5.1}$$

whenever h is small enough. Then, we have the following result.

Theorem 5.1. *Let $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(L)}\}$ be a given tight frame filter bank and let $\boldsymbol{q}^{(\ell)}$ have vanishing moment $\boldsymbol{\beta}_\ell$ with the associated constant $C^{(\ell)}_{\boldsymbol{\beta}_\ell}$ given by* (2.12) *and $m = \max_\ell\{|\boldsymbol{\beta}_\ell| : 1 \le \ell \le L\}$. Assume that the threshold $\boldsymbol{\alpha}(\boldsymbol{d})$ for wavelet frame coefficients $\boldsymbol{d}$ takes the form*

$$\boldsymbol{\alpha}(\boldsymbol{d}) = \Big\{\alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}}) = \frac{\tau|d_{\ell,\boldsymbol{n}}|}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}}\, \widetilde{g}_\ell((d_{\ell,\boldsymbol{n}})^2) : 1 \le \ell \le L, \boldsymbol{n} \in \mathbb{Z}^2\Big\},$$

with $\widetilde{g}_\ell : \mathbb{R} \mapsto \mathbb{R}^+$ being some smooth function; and we set $h^{2m}/\tau = C$ for some constant $C > 0$ with h and τ sufficiently small. Let $\boldsymbol{u}^k$ be generated from the iterative soft-thresholding algorithm (2.32) *with $\wp = 1$. Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j}, k\tau) : \boldsymbol{j} \in \mathbb{Z}^2, k = 1, 2, \cdots\}$ with $u(\boldsymbol{x}, t)$ being the solution of the diffusion equation* (3.5) *with the diffusivity functions given by* (5.1).

Note that in Theorem 5.1, the threshold $\boldsymbol{\alpha}$ depends on the wavelet frame coefficient $\boldsymbol{d}$, which means that the threshold in the iterative soft-thresholding algorithm (2.32) will be changing along with the iteration. However, the threshold used in the literature of wavelet or wavelet frame based image restoration is generally chosen independent of the iteration. The following corollary, which is a special case of Theorem 5.1, describes the type of differential equation that is approximated by the iterative soft-thresholding (2.32) when the threshold $\boldsymbol{\alpha}$ is independent of the iteration.

Corollary 5.1. *Under the same assumptions and notation as Theorem 5.1, if we choose the threshold $\boldsymbol{\alpha}$ as*

$$\boldsymbol{\alpha} = \Big\{\alpha_{\ell,\boldsymbol{n}} = \frac{\tau\lambda_\ell}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}}\ 1 \le \ell \le L, \boldsymbol{n} \in \mathbb{Z}^2\Big\},$$

for some constants $\lambda_\ell > 0$, then $\boldsymbol{u}^k$ generated from (2.32) *with $\wp = 1$ approximates the solution of the diffusion equation* (3.5) *with the diffusivity functions given by*

$$g_\ell(\xi^2) = \begin{cases} \min\Big\{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 C, \frac{\lambda_\ell}{|\xi|}\Big\}, & \text{for } |\boldsymbol{\beta}_\ell| = m, \\ \frac{\lambda_\ell}{|\xi|}, & \text{for } 1 \le |\boldsymbol{\beta}_\ell| < m. \end{cases}$$

In particular, when $L = 2$, $\boldsymbol{\beta}_1 = (1,0)$, $\boldsymbol{\beta}_2 = (0,1)$, and $C^{(1)}_{\boldsymbol{\beta}_1} = C^{(2)}_{\boldsymbol{\beta}_2} = C_{\boldsymbol{\beta}}$ (e.g. the Haar framelets), the corresponding diffusion equation to (2.32) *is*

$$u_t = \frac{\partial}{\partial x_1}\Big(\min\Big\{\bar{C}, \frac{\lambda_1}{|u_{x_1}|}\Big\}u_{x_1}\Big) + \frac{\partial}{\partial x_2}\Big(\min\Big\{\bar{C}, \frac{\lambda_2}{|u_{x_2}|}\Big\}u_{x_2}\Big), \tag{5.2}$$

with $\bar{C} = (C_{\boldsymbol{\beta}})^2 C$.

Remark 5.1.

1. *To have* (2.32) *approximate the 2nd order nonlinear diffusion* (5.2)*, we do not have to use Haar framelets. For example, we can use any B-spline tight frame filter bank, such as the piecewise linear framelets. We only need to properly adjust the threshold in a similar way as what was described in Corollary 3.1, i.e. introducing an additional power of h so that the terms corresponding to higher order derivatives vanish as $h \to 0$.*

2. *The diffusion* (5.2) *resembles the following (an anisotropic version of) mean curvature flow for $\lambda_1 = \lambda_2$,*

$$u_t = \lambda\left[\frac{\partial}{\partial x_1}\left(\frac{u_{x_1}}{|u_{x_1}|}\right) + \frac{\partial}{\partial x_2}\left(\frac{u_{x_2}}{|u_{x_2}|}\right)\right],$$

   except that the flow induced by iterative soft-thresholding, i.e. (5.2)*, is more regular in the sense that the diffusivity is bounded above.*

For $\wp = 2$, we consider the following rotation invariant diffusion

$$u_t = \sum_{\ell=1}^{L}(-1)^{1+|\boldsymbol{\beta}_\ell|}\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}\Big\{g_\ell\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\big(\frac{\partial^{\boldsymbol{\beta}_{\ell'}}u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_{\ell'}}}\big)^2\Big)\frac{\partial^{\boldsymbol{\beta}_\ell}u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}\Big\}. \tag{5.3}$$

Here, for a given $\boldsymbol{\beta}_\ell$, the summation $\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|} = \sum_{\{\ell' :\, |\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|\}}$. The relation between the shrinkage operator and the diffusivity functions g_ℓ is given by

$$S_\ell(\xi_\ell) = \xi_\ell\Big\{1 - \frac{\tau}{\big(C^{(\ell)}_{\boldsymbol{\beta}_\ell}\big)^2 h^{2|\boldsymbol{\beta}_\ell|}}g_\ell\Big(\frac{1}{h^{2|\boldsymbol{\beta}_\ell|}}\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\frac{\xi_{\ell'}^2}{\big(C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}}\big)^2}\Big)\Big\}, 1 \le \ell \le L. \tag{5.4}$$

Now, for

$$\mathcal{T}^2_{\theta_\ell}(\xi_\ell) = \frac{\xi_\ell}{\big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}|\xi_{\ell'}|^2\big)^{\frac{1}{2}}}\max\Big\{\big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}|\xi_{\ell'}|^2\big)^{\frac{1}{2}} - \theta_\ell, 0\Big\},$$

we have

$$\frac{\tau}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}}g_\ell\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\frac{\xi_{\ell'}^2}{(C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}})^2 h^{2|\boldsymbol{\beta}_{\ell'}|}}\Big) = 1 - \frac{\mathcal{T}^2_{\theta_\ell}(\xi_\ell)}{\xi_\ell} = \min\Big\{1, \frac{\theta_\ell}{\big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}|\xi_{\ell'}|^2\big)^{\frac{1}{2}}}\Big\},$$

or

$$g_\ell\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\xi_{\ell'}^2\Big) = \min\Big\{\frac{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{2|\boldsymbol{\beta}_\ell|}}{\tau}, \frac{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{|\boldsymbol{\beta}_\ell|}\theta_\ell}{\tau\big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}|C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}}\xi_{\ell'}|^2\big)^{\frac{1}{2}}}\Big\}.$$

Following the same idea as the case $\wp = 1$ and assuming $h^{2m}/\tau = C$ for some $C > 0$, we can choose

$$\theta_\ell = \frac{\tau\big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\big|C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}}\xi_{\ell'}\big|^2\big)^{\frac{1}{2}}}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{|\boldsymbol{\beta}_\ell|}}\widetilde{g}_\ell\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\xi^2_{\ell'}\Big),$$

so that

$$g_\ell\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\xi^2_{\ell'}\Big) = \begin{cases} \min\Big\{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 C, \widetilde{g}_\ell\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\xi^2_{\ell'}\Big)\Big\}, & \text{for } |\boldsymbol{\beta}_\ell| = m,\\ \widetilde{g}_\ell\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\xi^2_{\ell'}\Big), & \text{for } |\boldsymbol{\beta}_\ell| < m,\end{cases} \tag{5.5}$$

whenever h is small enough. Then, we have the following result.

Theorem 5.2. *Let $\{\boldsymbol{p}, \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(L)}\}$ be a given tight frame filter bank and let $\boldsymbol{q}^{(\ell)}$ have vanishing moment $\boldsymbol{\beta}_\ell$ with the associated constant $C^{(\ell)}_{\boldsymbol{\beta}_\ell}$ given by* (2.12) *and $m = \max_\ell\{|\boldsymbol{\beta}_\ell| : 1 \le \ell \le L\}$. Assume that the threshold $\boldsymbol{\alpha}(\boldsymbol{d})$ for wavelet frame coefficients $\boldsymbol{d}$ takes the form*

$$\boldsymbol{\alpha}(\boldsymbol{d}) = \Big\{\alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}}) = \frac{\tau\big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}\big|C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}}d_{\ell',\boldsymbol{n}}\big|^2\big)^{\frac{1}{2}}}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^{|\boldsymbol{\beta}_\ell|}}\widetilde{g}_\ell\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}(d_{\ell',\boldsymbol{n}})^2\Big) : 1 \le \ell \le L, \boldsymbol{n} \in \mathbb{Z}^2\Big\},$$

with $\widetilde{g}_\ell : \mathbb{R} \mapsto \mathbb{R}^+$ being some smooth function; and we set $h^{2m}/\tau = C$ for some constant $C > 0$ with h and τ sufficiently small. Let $\boldsymbol{u}^k$ be generated from the iterative soft-thresholding algorithm (2.32) *with $\wp = 2$ and the threshold given above. Then, $\boldsymbol{u}^k$ is a discrete approximation of $\{u(h\boldsymbol{j}, k\tau) : \boldsymbol{j} \in \mathbb{Z}^2, k = 1, 2, \cdots\}$ with $u(\boldsymbol{x}, t)$ being the solution of the diffusion equation* (5.3) *with the diffusivity functions given by* (5.5).

Similar as Corollary 5.1, we have the following corollary which is a special case of Theorem 5.2.

Corollary 5.2. *Under the same assumptions and notation as Theorem 5.2, if we choose the threshold $\boldsymbol{\alpha}$ as*

$$\boldsymbol{\alpha} = \Big\{\alpha_{\ell,\boldsymbol{n}} = \frac{\tau\lambda_\ell}{C^{(\ell)}_{\boldsymbol{\beta}_\ell}h^{|\boldsymbol{\beta}_\ell|}} \;: 1 \le \ell \le L, \boldsymbol{n} \in \mathbb{Z}^2\Big\},$$

for some constants $\lambda_\ell > 0$, then $\boldsymbol{u}^k$ generated from (2.32) *with $\wp = 2$ approximates the solution of the diffusion equation* (3.5) *with the diffusivity functions given by*

$$g_\ell(\xi^2) = \begin{cases} \min\Big\{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 C, \frac{\lambda_\ell}{\big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}}\xi^2_{\ell'}\big)^{1/2}}\Big\}, & \text{for } |\boldsymbol{\beta}_\ell| = m,\\ \frac{\lambda_\ell}{\big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}}\xi^2_{\ell'}\big)^{1/2}}, & \text{for } 1 \le |\boldsymbol{\beta}_\ell| < m.\end{cases}$$

In particular, when $L = 2$, $\boldsymbol{\beta}_1 = (1, 0)$, $\boldsymbol{\beta}_2 = (0, 1)$, and $C^{(1)}_{\boldsymbol{\beta}_1} = C^{(2)}_{\boldsymbol{\beta}_2} = C_{\boldsymbol{\beta}}$ (e.g. the Haar framelets), the corresponding diffusion equation to (2.32) *is*

$$u_t = \frac{\partial}{\partial x_1}\Big(\min\Big\{\bar{C}, \frac{\lambda_1/C_{\boldsymbol{\beta}}}{|\nabla u|}\Big\}u_{x_1}\Big) + \frac{\partial}{\partial x_2}\Big(\min\Big\{\bar{C}, \frac{\lambda_2/C_{\boldsymbol{\beta}}}{|\nabla u|}\Big\}u_{x_2}\Big), \tag{5.6}$$

with $\bar{C} = (C_{\boldsymbol{\beta}})^2 C$.

Remark 5.2. *The diffusion* (5.6) *resembles the following well-known mean curvature flow (see e.g. [2, 41, 42, 45, 59]) for $\lambda_1 = \lambda_2$,*

$$u_t = \lambda\left[\frac{\partial}{\partial x_1}\left(\frac{u_{x_1}}{|\nabla u|}\right) + \frac{\partial}{\partial x_2}\left(\frac{u_{x_2}}{|\nabla u|}\right)\right],$$

except that the flow induced by iterative soft-thresholding, i.e. (5.6)*, is more regular in the sense that the diffusivity is bounded above. In other words, the iterative isotropic soft-thresholding algorithm with threshold given in Corollary 5.2 solves a regularized mean curvature flow. The mean curvature flow has been used in image restoration as a regularizer that removes noise [50, 56, 61]. Since the equation is getting singular when $|\nabla u| \approx 0$, a regularized diffusivity was used to replace $|\nabla u|$, which is $|\nabla u|_\epsilon := \sqrt{|\nabla u|^2 + \epsilon^2}$. However, such regularization reduces the ability of the PDE model to preserve edges. It is known in the literature that soft-thresholding of wavelet frame coefficients of wavelet frame coefficients can well preserve edges. Therefore, Corollary 5.2 reveals that* (5.6) *is a better regularization of the mean curvature flow than using $|\nabla u|_\epsilon$ in place of $|\nabla u|$.*

5.2 Image-Restoration Embedded Diffusion

Other than the balanced and synthesis based model, the analysis based model (see [34,60]) is also frequently used in image restoration. This subsection shows that, inspired by algorithms solving the analysis based model, a new class of nonlinear diffusions, with the underlying image restoration model embedded in, can be derived.

Due to the diffusive nature of the diffusion equations, one drawback of the nonlinear diffusions we have seen so far is that, when $t \to \infty$, noise and all image features will vanish. For a good diffusion equation, image features diffuse slower than noise or other artifacts. Therefore in practice, an appropriate stopping time should be chosen for these diffusion equations. Also, since these diffusion equations do not have an idea of what image restoration problem it is solving, it is not guaranteed to produce relevant image restoration results. Therefore, we need to properly place the model of the underlying image restoration into the diffusion equations.

Image restoration is usually casted as the following linear inverse problem

$$Au = f + \eta, \tag{5.7}$$

where f is the observed image, η is additive noise and A is a linear operator corresponding to an image restoration problem (i.e. $A = I$ for denoising, $A = (a*)$ for deblurring, etc.). In discrete setting, we shall denote A, u and f as $\boldsymbol{A}$, $\boldsymbol{u}$ and $\boldsymbol{f}$ respectively. Since A is usually ill-posed, regularization based methods are usually adopted to find a reasonable solution. In the literature, various types of regularization have been used including the total variation model [56], and wavelet or wavelet frame based approach that includes synthesis based approach [24,36,37,39,40], analysis based approach [34,60] and balanced approach [7,18].

Here, we shall focus on the following analysis based model that was recently proposed by [38]:

$$\min_{\boldsymbol{u}} \; H_{\boldsymbol{\lambda}}(\boldsymbol{W}\boldsymbol{u}) + \frac{\mu}{2}\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{f}\|_2^2 + \frac{\mu^2}{2}\|\boldsymbol{A}^\top A\boldsymbol{u} - \boldsymbol{A}^\top \boldsymbol{f}\|_{\boldsymbol{D}}^2, \tag{5.8}$$

where $H_{\boldsymbol{\lambda}}$ is the Huber function (see [38] for detail) and $\|\boldsymbol{u}\|_{\boldsymbol{D}}^2 = \langle \boldsymbol{u}, \boldsymbol{D}\boldsymbol{u}\rangle$ with $\boldsymbol{D} = (I - \mu\boldsymbol{A}^\top\boldsymbol{A})^{-1}$. Two algorithms solving (5.8) were proposed by [38]: one is based on the proximal forward-backward splitting (PFBS) algorithm [7,18,21]; the other is based on the accelerated proximal gradient (APG) algorithm [58] (also known as FISTA [4]). Throughout the rest of this subsection, we assume again that $\boldsymbol{W}$ is a tight frame, i.e. $\boldsymbol{W}^\top\boldsymbol{W} = I$.

Start with the PFBS algorithm that solves (5.8)

$$\boldsymbol{u}^k = (I - \mu\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{W}^\top\boldsymbol{\mathcal{T}}_{\boldsymbol{\lambda}}(\boldsymbol{W}\boldsymbol{u}^{k-1}) + \mu\boldsymbol{A}^\top\boldsymbol{f}, \quad k = 1, 2, \cdots, \tag{5.9}$$

where $\boldsymbol{\mathcal{T}}_{\boldsymbol{\lambda}}$ is the soft-thresholding operator defined by either (2.29) or (2.30). Then, comparing the PFBS algorithm (5.9) with (2.31), it is natural to generalize (2.31) to the following algorithm

$$\boldsymbol{u}^k = (I - \mu\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{W}^\top\boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}) + \mu\boldsymbol{A}^\top\boldsymbol{f}, \quad k = 1, 2, \cdots. \tag{5.10}$$

It is not clear which energy function the algorithm (5.10) tries to minimize, since the threshold value $\boldsymbol{\alpha}^{k-1} = \boldsymbol{\alpha}(\boldsymbol{W}\boldsymbol{u}^{k-1})$ depends on $\boldsymbol{W}\boldsymbol{u}^{k-1}$ which is changing along with the iteration. However, if $\boldsymbol{\alpha}^k = \boldsymbol{\alpha}$ for all k, then it is not hard to show (similarly as in [38]) that (5.10) is a PFBS algorithm that solves the following optimization problem

$$\min_{\boldsymbol{u}} \; \frac{1}{2}\|\sqrt{\boldsymbol{\alpha}} \cdot \boldsymbol{W}\boldsymbol{u}\|_2^2 + \frac{\mu}{2}\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{f}\|_2^2 + \frac{\mu^2}{2}\|\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{u} - \boldsymbol{A}^\top\boldsymbol{f}\|_{\boldsymbol{D}}^2.$$

Therefore, when $\boldsymbol{\alpha}^k$ changes with k, algorithm (5.10) can be understood as an attempt to solve

$$\min_{\boldsymbol{u}} \; \frac{1}{2}\|\sqrt{\boldsymbol{\alpha}(\boldsymbol{W}\boldsymbol{u})} \cdot \boldsymbol{W}\boldsymbol{u}\|_2^2 + \frac{\mu}{2}\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{f}\|_2^2 + \frac{\mu^2}{2}\|\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{u} - \boldsymbol{A}^\top\boldsymbol{f}\|_{\boldsymbol{D}}^2. \tag{5.11}$$

Indeed, at each iteration k, if we let $\boldsymbol{\alpha}(\boldsymbol{W}\boldsymbol{u}) = \boldsymbol{\alpha}(\boldsymbol{W}\boldsymbol{u}^{k-1})$ in (5.11) and conduct one step of PFBS, we obtain the algorithm (5.10).

Now, we discuss the formulas of the nonlinear diffusions to which the algorithms (5.10) correspond. Once we have them, the corresponding diffusions to the algorithm (5.9) are automatically given by Theorem 5.1 and Theorem 5.2. We assume that the continuum and discrete version of the operator A satisfy the following consistency property

$$\boldsymbol{A}\boldsymbol{v} = Av + O(h) \quad \text{and} \quad \boldsymbol{A}^\top\boldsymbol{v} = A^\top v + O(h), \tag{5.12}$$

where $\boldsymbol{v_j} = v(\boldsymbol{j}h)$ for $\boldsymbol{j} \in \mathbb{Z}^2$ and v smooth enough. For the equations in (5.12) and below, their meaning is that the left-hand side (in sequence space) of the equation is a discretization of the smooth function on the right-hand side of the equation. We choose $\mu = \kappa\tau$ with some constant $\kappa > 0$. When the shrinkage given by (3.17) is used, we have

$$\mu \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{W}\boldsymbol{u}) = \kappa\tau A^\top A u + O(\tau h) + O(\tau^2).$$

Then, assuming (5.12), choosing the shrinkage given by (3.17) and following a similar derivation as in Section 3, we can obtain that the PDE approximated by algorithm (5.10) takes the form

$$u_t = \sum_{\ell=1}^{L} (-1)^{1+|\boldsymbol{\beta}_\ell|} \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \Big\{ g_\ell \Big(u, \frac{\partial^{\boldsymbol{\beta}_1} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_1}}, \cdots, \frac{\partial^{\boldsymbol{\beta}_L} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_L}} \Big) \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} u \Big\} - \kappa A^\top (Au - f). \tag{5.13}$$

In particular, when $L = 2$ with $\boldsymbol{\beta}_1 = (1,0), \boldsymbol{\beta}_2 = (0,1)$ and $g_1 = g_2 = g(\xi_1^2 + \xi_2^2)$, we have the following 2nd order nonlinear diffusion

$$u_t = \operatorname{div}\left(g\left(|\nabla u|^2 \right) \nabla u \right) - \kappa A^\top (Au - f),$$

which was considered by [63].

Now, recall the APG algorithm that solves (5.8) in [38]. Given some initial guess $\boldsymbol{v}^0$ and $\boldsymbol{v}^1$, and letting $t^0 = 1$, $t^{-1} = 0$, the algorithm reads

$$\begin{aligned} \boldsymbol{v}^k &= \boldsymbol{u}^{k-1} + \frac{t^{k-2} - 1}{t^{k-1}} (\boldsymbol{u}^{k-1} - \boldsymbol{u}^{k-2}) \\ \boldsymbol{u}^k &= (I - \mu \boldsymbol{A}^\top \boldsymbol{A}) \boldsymbol{W}^\top \boldsymbol{\mathcal{T}}_{\boldsymbol{\lambda}}(\boldsymbol{W}\boldsymbol{v}^k) + \mu \boldsymbol{A}^\top \boldsymbol{f}, \quad k = 1, 2, \cdots, \end{aligned} \tag{5.14}$$

where $t^k = \frac{1 + \sqrt{1 + 4(t^{k-1})^2}}{2}$. Following a similar argument as before, we have the following algorithm that generalizes (2.31)

$$\boldsymbol{u}^k = (I - \mu \boldsymbol{A}^\top \boldsymbol{A}) \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}} \Big((1 + \gamma^{k-1}) \boldsymbol{W} \boldsymbol{u}^{k-1} - \gamma^{k-1} \boldsymbol{W} \boldsymbol{u}^{k-2} \Big) + \mu \boldsymbol{A}^\top \boldsymbol{f}, \quad k = 1, 2, \cdots, \tag{5.15}$$

where $\gamma^{k-1} = \frac{t^{k-2} - 1}{t^{k-1}}$. The algorithm (5.15) (resp. (5.14)) is different from the PFBS algoirthm (5.10) (resp. (5.9)) in that it has both $\boldsymbol{u}^{k-1}$ and $\boldsymbol{u}^{k-2}$ involved in the iteration and the weighting given by t^k is changing along with the iteration. These differences makes APG require much less number of iterations to converge than the PFBS algorithm [4, 58]. The corresponding differential equation to the APG algorithm (5.15) will also be quite different from that of the PFBS algorithm (5.10). Once we find the corresponding nonlinear PDE to (5.15), the corresponding diffusions to the algorithm (5.14) are automatically given by Theorem 5.1 and Theorem 5.2.

We first have the following lemma about the asymptotical properties of t^k and γ^k, whose proof is provided in the end of this subsection.

Lemma 5.1. *Let $t^k, \gamma^k, k = 2, 3, \cdots$ be the sequences defined by $t^k = \frac{1 + \sqrt{1 + 4(t^{k-1})^2}}{2}$ with $t^1 = 1$, and $\gamma^k = (t^{k-1} - 1)/t^k$. Then*

$$\frac{t^k}{k} = \frac{1}{2} + O(\frac{\log k}{k}); \tag{5.16}$$

and

$$1 - \gamma^k = \frac{3}{k} + O(\frac{\log k}{k^2}). \tag{5.17}$$

For $\boldsymbol{u}_{\boldsymbol{j}}^k = u(\boldsymbol{j}h, \tau k)$ with u smooth enough, Taylor's expansion with respect to the time variable at $t = \tau(k-1)$ gives us

$$\begin{aligned} (1 + \gamma^{k-1}) \boldsymbol{u}_{\boldsymbol{j}}^{k-1} - \gamma^{k-1} \boldsymbol{u}_{\boldsymbol{j}}^{k-2} &= u(\boldsymbol{j}h, \tau(k-1)) + O(\tau), \\ \boldsymbol{u}^k - (1 + \gamma^{k-1}) \boldsymbol{u}_{\boldsymbol{j}}^{k-1} + \gamma^{k-1} \boldsymbol{u}_{\boldsymbol{j}}^{k-2} &= (1 - \gamma^{k-1}) \tau u_t(\boldsymbol{j}h, \tau(k-1)) + \tau^2 \tfrac{1 + \gamma^{k-1}}{2} u_{tt}(\boldsymbol{j}h, \tau(k-1)) + O(\tau^3). \end{aligned} \tag{5.18}$$

What makes algorithm (5.15) different from (5.10) is that u^{k-2} is also involved in the expression, which means we have $\boldsymbol{u}$ at three time steps present in the algorithm. This motivates us that, with a slightly different choice of the shrinkage operator, the corresponding PDE of algorithm (5.15) can be 2nd order in time. However, such correspondence turns out to be slightly different from all other correspondences we have established so far.

Suppose k is some large integer of size $O(\frac{1}{\tau})$, i.e. $k = C_1/\tau$ for some $C_1 > 0$. This is a reasonable assumption if we assume that the time interval $[0, T]$ is divided uniformly with size τ, and we focus on the behavior of (5.15) for large k. By Lemma 5.1, $\frac{1+\gamma^k}{2} \to 1$ as $k \to \infty$. Assume that (5.12) holds and we take $\mu = \tau^2\kappa$ with some $\kappa > 0$ and choose the following shrinkage function

$$S_{\boldsymbol{\alpha}_\ell,\boldsymbol{n}(\boldsymbol{d})}(d_{1,\boldsymbol{n}}, \cdots, d_{L,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}\Big(1 - \frac{\tau^2}{\widetilde{C}^{(\ell)}_{\boldsymbol{\alpha}_\ell} C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\alpha}_\ell|+|\boldsymbol{\beta}_\ell|}} g_\ell\Big(\frac{d_{1,\boldsymbol{n}}}{C^{(1)}_{\boldsymbol{\beta}_1} h^{|\boldsymbol{\beta}_1|}}, \cdots, \frac{d_{L,\boldsymbol{n}}}{C^{(L)}_{\boldsymbol{\beta}_L} h^{|\boldsymbol{\beta}_L|}}\Big)\Big), \tag{5.19}$$

for $1 \le \ell \le L$. We can see from (5.18) and Lemma 5.1 that, when τ and h are asymptotically small and $k = O(\frac{1}{\tau})$, the algorithm (5.15) *behaves asymptotically like* the following nonlinear evolution PDE that is *2nd order in time*

$$u_{tt} + Cu_t = \sum_{\ell=1}^{L} (-1)^{1+|\boldsymbol{\beta}_\ell|} \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \Big[g_\ell\Big(u, \frac{\partial^{\boldsymbol{\beta}_1} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_1}}, \cdots, \frac{\partial^{\boldsymbol{\beta}_L} u}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_L}}\Big) \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} u\Big] - \kappa A^\top(Au - f), \tag{5.20}$$

where $C = 3/(k\tau)$, the positive constant term in $(1 - \gamma^{k-1})/\tau = \frac{3}{k\tau} + O(\frac{\log k}{k^2\tau})$. For example, when $k = C_1/\tau$, then $C = 3/C_1$.

Remark 5.3.

1. *Both of the algorithms* (5.9) *and* (5.10) *discretize the PDE* (5.13). *For different types of shrinkage, we have different diffusivity functions* g_ℓ *(see e.g. the previous subsection for the form of* g_ℓ *that corresponds to the soft-thresholding).*
2. *What makes the nonlinear diffusion* (5.13) *different from the ones we have seen in earlier sections is that: (1) we have the underlying image restoration model embedded in the PDEs which lead to better image restoration results as supported by our numerical simulations; (2) it is now safe to seek the steady state solution, i.e.* $u(\boldsymbol{x}, \infty)$*, which makes the determination of stopping easier.*

We end this subsection by providing the proof of Lemma 5.1.

Proof of Lemma 5.1. For simplicity of presentation, denote $a_k = 2t^k, k = 1, 2, \cdots$. From

$$a_k = 1 + \sqrt{1 + a_{k-1}^2} \quad , \tag{5.21}$$

we have

$$1 + a_{k-1} < a_k < 2 + a_{k-1}, \ k \ge 2.$$

Thus,

$$(k-1) + a_1 < \cdots < 1 + 1 + a_{k-2} < a_k < 2 + 2 + a_{k-2} < \cdots < 2(k-1) + a_1,$$

and hence, we have

$$k < a_k < 2k, \ k = 2, 3, \cdots. \tag{5.22}$$

From (5.21) again, for $n \ge 1$,

$$\begin{aligned} a_{n+1} &= 1 + \sqrt{1 + a_n^2} = 1 + a_n\sqrt{1 + (1/a_n)^2} \\ &= 1 + a_n\Big(1 + \frac{1}{2}\frac{1}{a_n^2} - \frac{1}{8}\frac{1}{a_n^4} + O(\frac{1}{n^6})\Big) = 1 + a_n + \frac{1}{2}\frac{1}{a_n} - \frac{1}{8}\frac{1}{a_n^3} + O(\frac{1}{n^5}). \end{aligned}$$

Thus, we have

$$a_{n+1} - a_n = 1 + \frac{1}{2}\frac{1}{a_n} + O(\frac{1}{n^3}), \ n \ge 1, \tag{5.23}$$

which leads to that

$$a_k - a_1 = \sum_{n=1}^{k-1}(a_{n+1} - a_n) = k - 1 + \frac{1}{2}\sum_{n=1}^{k-1}\frac{1}{a_n} + O(1) = k - 1 + O(\log k),$$

where the last last equality follows from (5.22) and the fact $\sum_{n=1}^{k}\frac{1}{n} = O(\log k)$. Therefore, we have $a_k/k = 1 + O(\log k/k)$. This shows that (5.16) holds.

Applying (5.23) with $n = k-1$, we have

$$\begin{aligned}1 - \gamma^k = 1 - \frac{a_{k-1} - 2}{a_k} &= \frac{a_k - a_{k-1}}{a_k} + \frac{2}{a_k} = \frac{1 + 1/(2a_{k-1}) + O(1/k^3)}{a_k} + \frac{2}{a_k} \\ &= \frac{3}{a_k} + O(\frac{1}{k^2}) = \frac{3}{k + O(\log k)} + O(\frac{1}{k^2}) = \frac{3}{k} + O(\frac{\log k}{k^2}).\end{aligned}$$

This shows (5.17). □

5.3 Adaptive Thresholding for Wavelet Frame Shrinkage

Most of the iterative wavelet (frame) shrinkage algorithms for image restoration use one (or a few) fixed threshold for all coefficients. Since the original image to be restored can be sparsely approximated by wavelet frames, the threshold should be chosen such that the wavelet frame coefficients corresponding to features of the image, such as edges, are above the threshold. The rest of the coefficients are set to zero which does not hamper the restoration quality since the representation is sparse and thus most of the small nonzero coefficients correspond to noise instead of signal.

However, for a given nature image, it contains components with varied regularity. Therefore, a good threshold should be:

1. Large threshold should be chosen where the image is regular, while a small threshold should be chosen around singularities and the value of such threshold should depend on the type of the singularities.

2. The threshold near singularities should not only be moderately small, but should also be chosen such that we only introduce smoothing along the level sets of the image, while we are not introducing any smoothing or even sharpening in the directions normal to the level sets.

3. Since the restored image changes along with the iteration, the threshold should also be modified properly to incorporate with the updated information of local regularities of the image.

Therefore, to get a desirable image restoration through iterative wavelet frame shrinkage, we need a thresholding strategy satisfying the above three conditions.

It is not obvious how to design such thresholding strategy from only the perspective of wavelet frame transform. However, since we have the connection between wavelet frame shrinkage and nonlinear diffusions, we can borrow ideas from nonlinear diffusions, e.g. anisotropic diffusions, and use them to design a self-adaptive thresholding strategy for iterative wavelet shrinkage.

Take the Perona-Malik equation [52] as an example. This anisotropic diffusion equation reads

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

where g is a function satisfying

$$\begin{cases} g : [0,\infty) \mapsto (0,\infty) \text{ decreasing;} \\ g(0) = 1;\ g(x) \to 0 \text{ as } x \to \infty; \\ g(x) + 2xg'(x) > 0 \text{ for } x \le K; \quad g(x) + 2xg'(x) < 0 \text{ for } x > K. \end{cases} \tag{5.24}$$

One example of g is $g(x) = \frac{1}{1+x^p/K}$ for some constant $K > 0$ and $p > 1/2$. The diffusion coefficient of the Perona-Malik equation is $g(|\nabla u|^2)$ which controls the amount of diffusion at each location. By the specific assumption on g, we can see that at smooth regions ($|\nabla u|$ is small), $g(|\nabla u|^2)$ is large which means more diffusion is allowed; while near singularities ($|\nabla u|$ is large), $g(|\nabla u|^2)$ is small meaning less diffusion is allowed. If we rewrite the Perona-Malik equation as (see [52])

$$u_t = g(|\nabla u|^2)u_{TT} + \widetilde{g}(|\nabla u|^2)u_{NN},$$

with

$$\widetilde{g}(x) = g(x) + 2xg'(x), \quad N = \frac{\nabla u}{|\nabla u|} \quad \text{and} \quad T = N^{\perp},\ |T| = 1,$$

we can see that around singularities, the amount of diffusion across the singularities may be negative. A negative diffusivity means that we have a locally backward diffusion and hence the singularities such as edges are enhanced, which is desirable for image restoration.

An alternative way of interpreting such anisotropic diffusion, is that the diffusion at different locations of an image stops at different time or even moves in opposite directions: diffusion in regions with singularities stops earlier to prevent smearing or even moves backward to enhance sharp features, while diffusion in smooth regions stops late in order to remove sufficient amount of noise and other oscillatory artifacts.

Following the idea of Perona-Malik equation, we can choose the threshold $\boldsymbol{\alpha}$ for $\boldsymbol{\mathcal{T}}^{\wp}_{\boldsymbol{\alpha}}$ in the iterative wavelet frame soft-thresholding algorithm (2.32) to be adaptive to local features of the given image. Also, when we have a specific image restoration model (5.7) available, we can embed the model within the algorithm similarly as (5.10) and (5.15).

Now, we present the following iterative wavelet frame soft-thresholding algorithm with adaptive thresholds:

$$\boldsymbol{u}^k = (I - \mu \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{W}^\top \boldsymbol{\mathcal{T}}^{\wp}_{\boldsymbol{\theta}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}) + \mu \boldsymbol{A}^\top \boldsymbol{f}, \quad \text{with } \wp = 1 \text{ or } 2, \tag{5.25}$$

where $\boldsymbol{u}^0 = \boldsymbol{f}$ the initial data (e.g. the observed noisy image) and

$$\boldsymbol{\theta}^k := \boldsymbol{\theta}(\boldsymbol{G}_\sigma * \boldsymbol{W}\boldsymbol{u}^k) := \boldsymbol{\theta}(\boldsymbol{d}^k) := \Big\{\alpha_{0,\boldsymbol{n}} = 0; \alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d}^k) = C_\ell g_\ell\Big(h^{-2s_\ell} \sum_{|\boldsymbol{\beta}_{\ell'}| = |\boldsymbol{\beta}_\ell|} (d^k_{\ell',\boldsymbol{n}})^2\Big) : \boldsymbol{n} \in \mathbb{Z}^2, 1 \le \ell \le L\Big\} \tag{5.26}$$

with $C_\ell \ge 0$ being some fixed constant, $g_\ell(x)$ satisfying the first two conditions of (5.24), $\boldsymbol{d}^k = \boldsymbol{G}_\sigma * \boldsymbol{W}\boldsymbol{u}^k$ where $\boldsymbol{G}_\sigma$ denotes a discretized Gaussian with variance σ and s_ℓ satisfying

$$\begin{cases} s_\ell = |\beta_\ell|, & \text{for } \ell \in \big\{\ell : |\beta_\ell| = \min_{1\le \ell' \le L}\{|\beta_{\ell'}|\}\big\}, \\ s_\ell \le |\beta_\ell|, & \text{otherwise.} \end{cases}$$

The reason we only require g_ℓ satisfy the first two sets of conditions of (5.24), i.e. we do not require the diffusivity have a backward-diffusion mechanism, is because: (1) in contrast to the shrinkage operator $\boldsymbol{S}$, the soft-thresholding operator already has an edge-sharpening effect; (2) when the observed image $\boldsymbol{f}$ is blurred by the blurring operator $\boldsymbol{A}$, the presence of image restoration model in (5.25) (second term) also has a sharpening effect.

Similarly, we can have the following iterative wavelet frame shrinkage algorithm with adaptive thresholding that resembles (5.15) which is an accelerated version of (5.25):

$$\boldsymbol{u}^k = (I - \mu \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{W}^\top \boldsymbol{\mathcal{T}}^{\wp}_{\boldsymbol{\theta}^{k-1}}\left((1+\gamma^{k-1})\boldsymbol{W}\boldsymbol{u}^{k-1} - \gamma^{k-1}\boldsymbol{W}\boldsymbol{u}^{k-2}\right) + \mu \boldsymbol{A}^\top \boldsymbol{f}, \quad \wp = 1 \text{ or } 2, \tag{5.27}$$

where $\gamma^{k-1} = \frac{t^{k-2}-1}{t^{k-1}}$, $t^{-1} = 0$, $t^0 = 1$, and $\boldsymbol{\theta}^k$ is given either by (5.26) or

$$\boldsymbol{\theta}^k := \boldsymbol{\theta}\left(\boldsymbol{G}_\sigma * \left[(1+\gamma^{k-1})\boldsymbol{W}\boldsymbol{u}^k - \gamma^{k-1}\boldsymbol{W}\boldsymbol{u}^{k-1}\right]\right).$$

6 Convergence Analysis

This section is to provide convergence analysis of the iterative wavelet frame shrinkage (2.31) for a 2nd-order quasilinear parabolic equation; and the convergence of the iterative wavelet frame shrinkage algorithms (2.31) and (2.32) for generic thresholds.

Consider the following 2nd-order nonlinear diffusion equation

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial}{\partial x_1}\left[g_1\left(\left(\frac{\partial u}{\partial x_1}\right)^2\right)\frac{\partial u}{\partial x_1}\right] + \frac{\partial}{\partial x_2}\left[g_2\left(\left(\frac{\partial u}{\partial x_2}\right)^2\right)\frac{\partial u}{\partial x_2}\right] & \text{in } \Omega\times(0,t_e) \\ u = 0 & \text{on } \partial\Omega\times(0,t_e) \\ u(\boldsymbol{x},0) = u_0(\boldsymbol{x}) & \text{in } \Omega, \end{cases} \tag{6.1}$$

where $\Omega = (0,1)^2 \subset \mathbb{R}^2$, $t_e > 0$ and the initial data $u_0(\boldsymbol{x}) \in L_2(\Omega)$. We shall show that under suitable assumptions, we have that (2.31) converges to the solution of (6.1) with certain order of accuracy which depends on the choice of the corresponding wavelet frame system. Note that we can obtain similar convergence analysis for most of the nonlinear diffusions considered in this paper, provided that the given diffusion is well-posed. For simplicity and clarity, we shall focus on the diffusion (6.1).

We will also analyze the stability and convergence of the following generic iterative wavelet frame shrinkage

$$\boldsymbol{u}^k = \begin{cases} \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}) & \ell_2\text{-shrinkage} \\ \widetilde{\boldsymbol{W}}^\top \boldsymbol{\mathcal{T}}_{\boldsymbol{\theta}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}) & \ell_1\text{-shrinkage}, \end{cases} \tag{6.2}$$

with $\boldsymbol{\alpha}^k := \boldsymbol{\alpha}(\boldsymbol{d}^k)$, $\boldsymbol{\theta}^k := \boldsymbol{\theta}(\boldsymbol{d}^k)$, where $\boldsymbol{d}^k := \boldsymbol{W}\boldsymbol{u}^k$. In order to include most of the iterative shrinkage formulas we have discussed in the previous sections, we assume the thresholding function $\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d})$ and $\boldsymbol{\mathcal{T}}_{\boldsymbol{\theta}}(\boldsymbol{d})$ take the following forms

$$\begin{cases} \boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d}) = \left\{S_{\alpha_{\ell,\boldsymbol{n}}}(d_{\ell,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}(1-\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})) :\ 0\le\ell\le L, \boldsymbol{n}\in\mathbb{Z}^2\right\}, \\ \boldsymbol{\mathcal{T}}_{\boldsymbol{\theta}}(\boldsymbol{d}) = \left\{\mathcal{T}_{\theta_{\ell,\boldsymbol{n}}}(d_{\ell,\boldsymbol{n}}) = \frac{d_{\ell,\boldsymbol{n}}}{w}\max\left\{w-\theta_{\ell,\boldsymbol{n}}(\boldsymbol{d}),0\right\};\ w = w(d_{1,\boldsymbol{n}},\ldots,d_{L,\boldsymbol{n}}) :\ 0\le\ell\le L, \boldsymbol{n}\in\mathbb{Z}^2\right\}, \end{cases} \tag{6.3}$$

with

$$\begin{cases} \boldsymbol{\alpha}(\boldsymbol{d}) = \left\{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d}) = g_\ell(d_{1,\boldsymbol{n}},\cdots,d_{L,\boldsymbol{n}},h,\tau) :\ 0\le\ell\le L, \boldsymbol{n}\in\mathbb{Z}^2\right\}, \\ \boldsymbol{\theta}(\boldsymbol{d}) = \left\{\theta_{\ell,\boldsymbol{n}}(\boldsymbol{d}) = w\ g_\ell(d_{1,\boldsymbol{n}},\cdots,d_{L,\boldsymbol{n}},h,\tau) :\ 0\le\ell\le L, \boldsymbol{n}\in\mathbb{Z}^2\right\}. \end{cases} \tag{6.4}$$

We shall prove that, under suitable assumptions on the function g_ℓ, the iteration (6.2) has a subsequence converges weakly to a function in $L_2(\Omega)$ as $h,\tau\to 0$ and $k = O(\frac{1}{\tau})\to\infty$; and such convergence is stable.

6.1 Convergence of (2.31)

This subsection focus on the convergence of the iterative wavelet frame shrinkage (2.31), with a proper choice of thresholds, to a solution of the nonlinear diffusion equation (6.1) as $\tau, h \to 0$. To prove convergence of (2.31), we need to first make sure the corresponding PDE (6.1) is well-posed, for which we shall make the following assumptions.

Assumptions 6.1. *We assume that, for each* $\ell = 1,2$, $g_\ell(\xi) \in C^\infty(\mathbb{R})$ *are nonnegative functions satisfying* $0 < \left(\xi g_\ell(\xi^2)\right)' \le B, \xi\in\mathbb{R}$, *where* B *is a positive constant.*

Under Assumption 6.1, it is clear from the theory of nonlinear semigroups and monotone operators [6] that the quasilinear parabolic equation (6.1) is well-posed, i.e. there exists a unique solution for each given initial data, and it is continuously dependent on the initial data with respect to the L_2-norm (see e.g. [35] for a proof). With well-posedness of the differential equation, Lax's equivalent theorem implies that, to prove convergence of (2.31), we only need to show consistency and stability of this iterative method. For convenience, we shall also make the following assumptions on the underlying wavelet frame systems of (2.31). Note that the assumptions below are satisfied by all tensor-product B-spline tight frame systems constructed by [54].

Assumptions 6.2. *We shall consider the discretization given by a tensor-product tight frame system with FIR filters* $\{\boldsymbol{q}^{(\ell)} : 0\le\ell\le L\}$, *where* $\boldsymbol{q}^{(0)} = \boldsymbol{p}$ *and* $L\ge 2$. *Assume that each* $\boldsymbol{q}^{(\ell)}$ *with* $1\le\ell\le L$ *has vanishing moments of order* $\boldsymbol{\beta}_\ell$ *and has total vanishing moments of order* $K_\ell\backslash\{|\boldsymbol{\beta}_\ell|+1\}$ *with* $1\le|\boldsymbol{\beta}_\ell|<K_\ell$. *In particular, we require that* $\boldsymbol{\beta}_1 = (1,0), \boldsymbol{\beta}_2 = (0,1)$ *and* $K_1 = K_2$.

6.1.1 Consistency

To show consistency, we need to show that for any given $u \in C^\infty(\mathbb{R}^2)$, when the algorithm (2.31) is applied to $\{u(h\boldsymbol{j}, \tau k) : \boldsymbol{j} \in \mathbb{Z}^2, k = 1, 2, \ldots\}$ we will recover the PDE (6.1) plus the local truncation error which decreases to zero with a certain order of τ and h.

Lemma 2.1 implies that, for any $u \in C^\infty(\mathbb{R}^2)$ and $h > 0$ small enough, we have, for $1 \le \ell \le L$,

$$\frac{1}{h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \boldsymbol{q}^{(\ell)}[\boldsymbol{j}] u(h\boldsymbol{m} \pm h\boldsymbol{j}) = C^{(\ell)}_{\boldsymbol{\beta}_\ell} \frac{\partial^{\boldsymbol{\beta}_\ell} u(h\boldsymbol{m})}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} + O(h^{K_\ell - |\boldsymbol{\beta}_\ell|}). \tag{6.5}$$

For example, when the Haar wavelet frame system is used, we have $K_1 = K_2 = 2$; when the piecewise linear wavelet frame system is used, we have $K_1 = K_2 = 3$. By Theorem 3.2 and Corollary 3.1, we can take the following threshold $\boldsymbol{\alpha}$ for the shrinkage operator $\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d})$ that is in correspondence with the diffusivity functions of (6.1):

$$\begin{cases} \alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}}) = \frac{\tau}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^2} g_\ell\Big(\frac{(d_{\ell,\boldsymbol{n}})^2}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2 h^2}\Big) & \text{for } \ell = 1, 2; \\ \alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}}) = 0 & \text{for } 3 \le \ell \le L. \end{cases} \tag{6.6}$$

Then, we have the following consistency result stating that the iterative shrinkage method (2.31) using a tight wavelet frame satisfying Assumption 6.2, is consistent of order $O(\tau) + \sum_{\ell=1}^{2} O(h^{K_\ell - |\boldsymbol{\beta}_\ell|})$.

Proposition 6.1. *The numerical algorithm* (2.31) *with a tight wavelet frame system satisfying Assumption 6.2 is consistent of order* $O(\tau) + O(h^{K_\ell - |\boldsymbol{\beta}_\ell|})$ *to the diffusion equation* (6.1) *provided that the thresholds are chosen as* (6.6)*.*

Proof. We need to show that for any $u \in C^\infty(\mathbb{R}^2)$, we have

$$\begin{aligned} &\boldsymbol{u}^{k+1}_{\boldsymbol{j}} - \Big\{\boldsymbol{u}^k_{\boldsymbol{j}} - \tau \sum_{\ell=1}^{2} \frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m}} \boldsymbol{q}^{(\ell)}[\boldsymbol{j} - \boldsymbol{m}] g_\ell\Big(\big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n}} \boldsymbol{q}^{(\ell)}[n] \boldsymbol{u}^k_{\boldsymbol{n}+\boldsymbol{m}}\big)^2\Big)\Big(\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n}} \boldsymbol{q}^{(\ell)}[n] \boldsymbol{u}^k_{\boldsymbol{n}+\boldsymbol{m}}\Big)\Big\} \\ &= \tau\Big(\boldsymbol{u}^k_{t,\boldsymbol{j}} + \sum_{\ell=1}^{2} \Big(g_\ell\big((\boldsymbol{u}^k_{\boldsymbol{\beta}_\ell})^2\big) \boldsymbol{u}^k_{\boldsymbol{\beta}_\ell}\Big)_{\boldsymbol{\beta}_\ell, \boldsymbol{j}}\Big) + O(\tau^2) + \sum_{\ell=1}^{2} O(\tau h^{K_\ell - |\boldsymbol{\beta}_\ell|}), \end{aligned}$$

where $\boldsymbol{u}^k_{\boldsymbol{j}} = u(h\boldsymbol{j}, \tau k)$ and

$$\boldsymbol{u}^k_{t,\boldsymbol{j}} = u_t(\boldsymbol{j}h, \tau k), \quad \boldsymbol{u}^k_{\boldsymbol{\beta}_\ell} := \frac{\partial^{\boldsymbol{\beta}_\ell} u(\boldsymbol{x}, \tau k)}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \quad \text{and} \quad \boldsymbol{u}^k_{\boldsymbol{\beta}_\ell, \boldsymbol{m}} := \frac{\partial^{\boldsymbol{\beta}_\ell} u(\boldsymbol{m}h, \tau k)}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}}.$$

Denote

$$\widetilde{G}_\ell(\xi) = \xi g_\ell(\xi^2), \quad U^k_{\boldsymbol{j}} = \frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n}} \boldsymbol{q}^{(\ell)}[n] \boldsymbol{u}^k_{\boldsymbol{n}+\boldsymbol{j}}, \ \boldsymbol{j} \in \mathbb{Z}^2.$$

Then by (6.5),

$$U^k_{\boldsymbol{j}} = \boldsymbol{u}^k_{\boldsymbol{\beta}_\ell, \boldsymbol{j}} + O(h^{K_\ell - |\boldsymbol{\beta}_\ell|}).$$

Applying (6.5) to $\widetilde{G}_\ell$, we have

$$\begin{aligned} &\frac{1}{C^{(\ell)}_{\boldsymbol{\beta}_\ell} h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m}} \boldsymbol{q}^{(\ell)}[\boldsymbol{j} - \boldsymbol{m}] \widetilde{G}_\ell(U^k_{\boldsymbol{m}}) = \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \widetilde{G}_\ell(U^k_{\boldsymbol{j}}) + O(h^{K_\ell - |\boldsymbol{\beta}_\ell|}) \\ &= \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \widetilde{G}_\ell\Big(\boldsymbol{u}^k_{\boldsymbol{\beta}_\ell, \boldsymbol{j}} + O(h^{K_\ell - |\boldsymbol{\beta}_\ell|})\Big) + O(h^{K_\ell - |\boldsymbol{\beta}_\ell|}) = \frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial \boldsymbol{x}^{\boldsymbol{\beta}_\ell}} \widetilde{G}_\ell(\boldsymbol{u}^k_{\boldsymbol{\beta}_\ell, \boldsymbol{j}}) + O(h^{K_\ell - |\boldsymbol{\beta}_\ell|}) \\ &= \Big(g_\ell\big((\boldsymbol{u}^k_{\boldsymbol{\beta}_\ell})^2\big) \boldsymbol{u}^k_{\boldsymbol{\beta}_\ell}\Big)_{\boldsymbol{\beta}_\ell, \boldsymbol{j}} + O(h^{K_\ell - |\boldsymbol{\beta}_\ell|}), \end{aligned}$$

where the second to the last equality follows from the fact that $\widetilde{G}_\ell$ is differentiable. Note that $\boldsymbol{u}_{\boldsymbol{j}}^{k+1} - \boldsymbol{u}_{\boldsymbol{j}}^{k} = \tau u_t(\boldsymbol{j}h, \tau k) + O(\tau^2)$. Therefore, we have

$$\begin{aligned}
&\boldsymbol{u}_{\boldsymbol{j}}^{k+1} - \Big\{\boldsymbol{u}_{\boldsymbol{j}}^{k} - \tau \sum_{\ell=1}^{2} \frac{1}{C_{\boldsymbol{\beta}_\ell}^{(\ell)}\ h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m}} \boldsymbol{q}^{(\ell)}[\boldsymbol{j}-\boldsymbol{m}] g_\ell\Big(\big(\frac{1}{C_{\boldsymbol{\beta}_\ell}^{(\ell)}\ h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n}} \boldsymbol{q}^{(\ell)}[n]\boldsymbol{u}_{\boldsymbol{n}+\boldsymbol{m}}^{k}\big)^2\Big)\Big(\frac{1}{C_{\boldsymbol{\beta}_\ell}^{(\ell)}\ h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{n}} \boldsymbol{q}^{(\ell)}[n]\boldsymbol{u}_{\boldsymbol{n}+\boldsymbol{m}}^{k}\Big)\Big\} \\
&\quad = \boldsymbol{u}_{\boldsymbol{j}}^{k+1} - \boldsymbol{u}_{\boldsymbol{j}}^{k} + \tau \sum_{\ell=1}^{2} \frac{1}{C_{\boldsymbol{\beta}_\ell}^{(\ell)}\ h^{|\boldsymbol{\beta}_\ell|}} \sum_{\boldsymbol{m}} \boldsymbol{q}^{(\ell)}[\boldsymbol{j}-\boldsymbol{m}] \widetilde{G}_\ell(U_{\boldsymbol{m}}^{k}) \\
&\quad = \tau\left(\boldsymbol{u}_{t,\boldsymbol{j}}^{k} + \left(g_\ell\left((\boldsymbol{u}_{\boldsymbol{\beta}_\ell}^{k})^2\right)\boldsymbol{u}_{\boldsymbol{\beta}_\ell}^{k}\right)_{\boldsymbol{\beta}_\ell,\boldsymbol{j}}\right) + O(\tau^2) + \sum_{\ell=1}^{2} O(\tau h^{K_\ell - |\boldsymbol{\beta}_\ell|}).
\end{aligned}$$

This concludes the proof of the proposition. □

Remark 6.1.

1. *Note from Corollary 3.1 that, instead of setting the thresholds $\alpha_{\ell,\boldsymbol{n}}$ to zeros for $3 \le \ell \le L$, we can take the choices of $\boldsymbol{\alpha}$ given by* (3.13)*. The proof of consistency for these cases is very similar to that of Proposition 6.1.*
2. *The order of consistency we have just derived may not optimal. Optimal order of consistency can be obtained by a more careful local truncation error analysis. However, finding an (optimal) order of approximation of the solutions of PDEs is not what we want to focus in this paper. Although the analysis in this section can be modified to reveal the order of approximation, we will only prove convergence without addressing the order of convergence. The reason is that in image restoration, or image processing and analysis in general, all data are discrete to begin with, and most data does not have nor need a corresponding continuum version. Therefore, in this section, we show convergence, only to rigourously justify that wavelet frame shrinkage is indeed a discretization of the corresponding PDE.*

6.1.2 Stability

Same as standard finite difference discretization of nonlinear diffusions, the temporal step size τ and the spatial step size h should satisfy a certain stability condition. For the diffusivity functions in (6.1), the multiplicative shrinkage operator $\boldsymbol{S}_{\boldsymbol{\alpha}}$ defined in (2.28) takes the following form:

$$\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d}) = \{d_{\ell,\boldsymbol{n}}(1 - \alpha_{\ell,\boldsymbol{n}}(d_{\ell,\boldsymbol{n}})) : \ \boldsymbol{n} \in \mathbb{Z}^2, 0 \le \ell \le L\}. \tag{6.7}$$

with $\alpha_{\ell,\boldsymbol{n}}$ given in (6.6). Now, we address the stability condition for our iterative multiplicative shrinkage algorithm (2.31) as follows.

Assumptions 6.3. *Given the shrinkage operator $\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d})$ in* (6.7)*, we assume that $\alpha_{\ell,\boldsymbol{n}}(\xi)$ is differentiable for each ℓ and $\boldsymbol{n}$, and*

$$\left|\Big(\xi\big(1 - \alpha_{\ell,\boldsymbol{n}}(\xi)\big)\Big)'\right| = \big|1 - \alpha_{\ell,\boldsymbol{n}}(\xi) - \xi\alpha'_{\ell,\boldsymbol{n}}(\xi)\big| \le 1$$

for all $0 \le \ell \le L$ and $\boldsymbol{n} \in \mathbb{Z}^2$.

The following proposition shows that the stability requirements in Assumption 6.3 can be easily achieved in practice by assuming $\tau = O(h^2)$.

Proposition 6.2. *Suppose Assumption 6.1 hold for the diffusivity functions of* (6.1)*. Given the choice of thresholding $\boldsymbol{\alpha}$ in* (6.6)*, the stability requirements in Assumption 6.3 can be achieved by taking $\tau = Ch^2$, for some C independent of τ and h.*

Proof. First, observe that

$$\left|1-\alpha_{\ell,\boldsymbol{n}}(\xi)-\xi\alpha'_{\ell,\boldsymbol{n}}(\xi)\right|\leq 1 \quad\Leftrightarrow\quad 0\leq \alpha_{\ell,\boldsymbol{n}}(\xi)+\xi\alpha'_{\ell,\boldsymbol{n}}(\xi)\leq 2. \tag{6.8}$$

Let $\widetilde{C}:=\frac{\tau}{(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2h^2}$. Then, we have

$$\alpha_{\ell,\boldsymbol{n}}(\xi)=\widetilde{C}g_\ell\left(\widetilde{C}\xi^2/\tau\right)\quad\Rightarrow\quad \alpha'_{\ell,\boldsymbol{n}}(\xi)=\frac{2\widetilde{C}^2\xi}{\tau}g'_\ell\Big(\frac{\widetilde{C}\xi^2}{\tau}\Big).$$

Therefore,

$$\alpha_{\ell,\boldsymbol{n}}(\xi)+\xi\alpha'_{\ell,\boldsymbol{n}}(\xi)=\widetilde{C}g_\ell\Big(\frac{\widetilde{C}\xi^2}{\tau}\Big)+\frac{2\widetilde{C}^2\xi^2}{\tau}g'_\ell\Big(\frac{\widetilde{C}\xi^2}{\tau}\Big)=\widetilde{C}\Big[g_\ell\Big(\frac{\widetilde{C}\xi^2}{\tau}\Big)+2\frac{\widetilde{C}\xi^2}{\tau}g'_\ell\Big(\frac{\widetilde{C}\xi^2}{\tau}\Big)\Big].$$

Recall the assumption we imposed in Assumption 6.1:

$$0<(\xi g_\ell(\xi^2))'=g_\ell(\xi^2)+2\xi^2g_\ell(\xi^2)\leq B.$$

Thus, (6.8) is satisfied whence we have $\widetilde{C}\leq\frac{2}{B}$, which yields the condition

$$\tau=Ch^2\quad\text{with}\quad C\leq 2(C^{(\ell)}_{\boldsymbol{\beta}_\ell})^2/B.$$

This concludes the proof. ☐

Assume that the computation domain $\Omega=(0,1)^2$ is discretized by a dyadic grid

$$\{2^{-n}\boldsymbol{k}:0<k_1,k_2<2^n,n\geq 1\}.$$

Therefore, the meshsize for spatial discretization is $h=2^{-n}$. We denote the index set $\mathbb{O}^2_h\subset\mathbb{Z}^2$ as

$$\mathbb{O}^2_h:=\{\boldsymbol{k}\in\mathbb{Z}^2:0<k_1,k_2<2^n\}.$$

Given any function $u\in L_2(\Omega)$, and a compactly supported tensor-product B-spline function $\phi\in L_2(\mathbb{R}^2)$, we define the sampling operator $T_h:L_2(\Omega)\mapsto\ell_2(\mathbb{Z}^2)$ as

$$(T_hu)_{\boldsymbol{k}}:=2^n\langle u,\phi_{n,\boldsymbol{k}}\rangle,\quad\text{with }\boldsymbol{k}\in\mathbb{O}^2_h, \tag{6.9}$$

where $\phi_{n,\boldsymbol{k}}=2^n\phi(2^n\cdot-\boldsymbol{k})$. When a wavelet frame system associated to a refinable function ϕ is used in the shrinkage algorithms, the underlying sampling is given by the operator T_h. We denote the space $\ell_{2,h}(\mathbb{O}^2_h)$ as the collection of all vectors supported on the index set $\mathbb{O}^2_h$ equipped with the following norm

$$\|\boldsymbol{v}\|^2_{\ell_{2,h}(\mathbb{O}^2_h)}=\sum_{\boldsymbol{j}\in\mathbb{O}^2_h}|v_{\boldsymbol{j}}|^2h^2.$$

Note that we have $\|T_hu\|_{\ell_{2,h}(\mathbb{O}^2_h)}\leq C\|u\|_{L_2(\Omega)}$ with some C independent of h (see e.g. [8]).

Given a $t\in(0,t_e)$, let $K=\lfloor t/\tau\rfloor$. We define the discrete operator $M_{h,\tau}:\ell_{2,h}(\mathbb{O}^2_h)\mapsto\ell_{2,h}(\mathbb{O}^2_h)$ as

$$M_{h,\tau}\boldsymbol{v}:=\Big(\prod_{k=0}^{K}\boldsymbol{W}^\top\boldsymbol{S}_{\boldsymbol{\alpha}^k}\boldsymbol{W}\Big)\boldsymbol{v},\quad\text{for }\boldsymbol{v}\in\ell_{2,h}(\mathbb{O}^2_h).$$

Note that the shrinkage operator $\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d})$ defined in (2.28) is nonlinear in $\boldsymbol{d}$, since $\boldsymbol{\alpha}$ depends on $\boldsymbol{d}$ (see (6.7)). Thus the product above means a product of compositions of nonlinear operators.

Definition 6.1. *We say that the discrete algorithm* (2.31) *is stable, if for any $\epsilon>0$, there exist $\delta>0$ independent of τ,h, such that for any $u,v\in L_2(\Omega)$ with $\|u-v\|_{L_2(\Omega)}<\delta$, we have*

$$\|M_{h,\tau}T_hv-M_{h,\tau}T_hu\|_{\ell_{2,h}(\mathbb{O}^2_h)}<\epsilon,$$

i.e. the set of operators $\{M_{h,\tau}T_h\}_{h,\tau}$ is equi-continuous.

Now we show that $\{M_{h,\tau}T_h\}_{h,\tau}$ is indeed equi-continuous if Assumption 6.2 and 6.3 are satisfied.

Proposition 6.3. *Algorithm* (2.31)*, with threshold satisfying* (6.6)*, is stable if Assumption 6.2 and 6.3 are satisfied*

Proof. Note that $\|W\|_{\ell_{2,h}(\mathbb{O}_h^2)} = \|W^\top\|_{\ell_{2,h}(\mathbb{O}_h^2)} \leq 1$ which is in fact true for all tight wavelet frame systems, we have

$$\begin{aligned}\|M_{h,\tau}T_hv - M_{h,\tau}T_hu\|_{\ell_{2,h}(\mathbb{O}_h^2)} &= \Big\|\Big(\prod_{k=0}^{K} \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k}\boldsymbol{W}\Big)T_hv - \Big(\prod_{k=0}^{K} \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k}\boldsymbol{W}\Big)T_hu\Big\|_{\ell_{2,h}(\mathbb{O}_h^2)}\\ &\leq \Big\|\boldsymbol{S}_{\boldsymbol{\alpha}^K}\Big(\boldsymbol{W}\Big(\prod_{k=0}^{K-1} \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k}\boldsymbol{W}\Big)T_hv\Big) - \boldsymbol{S}_{\boldsymbol{\alpha}^K}\Big(\boldsymbol{W}\Big(\prod_{k=0}^{K-1} \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k}\boldsymbol{W}\Big)T_hu\Big)\Big\|_{\ell_{2,h}(\mathbb{O}_h^2)}\end{aligned}$$

Let

$$\boldsymbol{d}_v^K = \boldsymbol{W}\Big(\prod_{k=0}^{K-1} \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k}\boldsymbol{W}\Big)T_hv \quad \text{and} \quad \boldsymbol{d}_u^K = \boldsymbol{W}\Big(\prod_{k=0}^{K-1} \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k}\boldsymbol{W}\Big)T_hu.$$

Then, we have

$$\begin{aligned}\|M_{h,\tau}T_hv - M_{h,\tau}T_hu\|_{\ell_{2,h}(\mathbb{O}_h^2)} &\leq \big\|\boldsymbol{S}_{\boldsymbol{\alpha}^K}(\boldsymbol{d}_v^K) - \boldsymbol{S}_{\boldsymbol{\alpha}^K}(\boldsymbol{d}_u^K)\big\|_{\ell_{2,h}(\mathbb{O}_h^2)}\\ &= \big\|\boldsymbol{d}_v^K\cdot(1-\boldsymbol{\alpha}(\boldsymbol{d}_v^K)) - \boldsymbol{d}_u^K\cdot(1-\boldsymbol{\alpha}(\boldsymbol{d}_u^K))\big\|_{\ell_{2,h}(\mathbb{O}_h^2)}\\ \text{(Assumption 6.3)}\quad &\leq \big\|\boldsymbol{d}_v^K - \boldsymbol{d}_u^K\big\|_{\ell_{2,h}(\mathbb{O}_h^2)} \leq \big\|\boldsymbol{S}_{\boldsymbol{\alpha}^{K-1}}(\boldsymbol{d}_v^{K-1}) - \boldsymbol{S}_{\boldsymbol{\alpha}^{K-1}}(\boldsymbol{d}_u^{K-1})\big\|_{\ell_{2,h}(\mathbb{O}_h^2)}\\ &\cdots\\ &\leq \|T_hv - T_hu\|_{\ell_{2,h}(\mathbb{O}_h^2)} \leq C\|v-u\|_{L_2(\Omega)}.\end{aligned}$$

This shows that $\{M_{h,\tau}T_h\}_{h,\tau}$ is indeed equi-continuous on $L_2(\Omega)$. □

6.1.3 Convergence

We show that: well-posedness + consistency + stability ⇒ convergence. We first establish the following lemma regarding the consistency between the sampling T_h and the pointwise sampling for smooth functions. Given $u \in C(\Omega)$, we denote the pointwise sampling operator R_hu as

$$(R_hu)_{\boldsymbol{k}} = u(h\boldsymbol{k}), \quad \boldsymbol{k} \in \mathbb{O}_h^2. \tag{6.10}$$

Then we have the following lemma.

Lemma 6.1. *Let ϕ be a tensor-product B-spline refinable function, then we have, for every $u \in C(\Omega)$,*

$$\|T_hu - R_hu\|_{\ell_{2,h}(\mathbb{O}_h^2)} \to 0, \quad \textit{as } h \to 0. \tag{6.11}$$

Proof. We first note that $\|\cdot\|_{\ell_{2,h}(\mathbb{O}_h^2)} \leq \|\cdot\|_{\ell_1(\mathbb{O}_h^2)}$. Thus, we focus on showing that

$$\|T_hu - R_hu\|_{\ell_1(\mathbb{O}_h^2)} \to 0, \quad \text{as } h \to 0. \tag{6.12}$$

Let $\Box_{\boldsymbol{k}}$ be the rectangular domain $[\frac{k_1}{2^n}, \frac{k_1+1}{2^n}] \times [\frac{k_2}{2^n}, \frac{k_2+1}{2^n}]$ with $\boldsymbol{k} \in \mathbb{O}_h^2$. Then we have

$$\begin{aligned}\|T_hu - R_hu\|_{\ell_1(\mathbb{O}_h^2)} &= \sum_{\boldsymbol{k}\in\mathbb{O}_h^2} \Big|2^n\langle u, \phi_{n,\boldsymbol{k}}\rangle - u(2^{-n}\boldsymbol{k})\Big|2^{-2n}\\ &= \int_\Omega \sum_{\boldsymbol{k}\in\mathbb{O}_h^2} \Big|2^n\langle u, \phi_{n,\boldsymbol{k}}\rangle - u(2^{-n}\boldsymbol{k})\Big|\chi_{\Box_{\boldsymbol{k}}} d\boldsymbol{x}\\ &= \int_\Omega \Big|\sum_{\boldsymbol{k}\in\mathbb{O}_h^2} 2^n\langle u, \phi_{n,\boldsymbol{k}}\rangle\chi_{\Box_{\boldsymbol{k}}} - u(2^{-n}\boldsymbol{k})\chi_{\Box_{\boldsymbol{k}}}\Big| d\boldsymbol{x}.\end{aligned}$$

By [8, Lemma 4.1], for every $u \in L_1(\Omega) \supset C(\Omega)$ we have

$$\int_\Omega \Big| \sum_{\boldsymbol{k}\in\mathbb{O}_h^2} 2^n \langle u, \phi_{n,\boldsymbol{k}}\rangle \chi_{\square_{\boldsymbol{k}}}(\boldsymbol{x}) - u(\boldsymbol{x})\Big| d\boldsymbol{x} \to 0, \quad \text{as } n \to \infty.$$

On the other hand, the fact that for every $u \in C(\Omega)$

$$\int_\Omega \Big| \sum_{\boldsymbol{k}\in\mathbb{O}_h^2} u(2^{-n}\boldsymbol{k}) \chi_{\square_{\boldsymbol{k}}} - u(\boldsymbol{x})\Big| d\boldsymbol{x} \to 0, \quad \text{as } n \to \infty$$

is well-known in real analysis. We then have (6.12) by triangular inequality. □

Given any $v \in L_2(\Omega)$ and a $t \in (0, t_e)$, we define an operator $M : L_2(\Omega) \mapsto L_2(\Omega)$ as $u(\boldsymbol{x}, t) = M(v)$, where $u(\cdot, t) \in L_2(\Omega)$ is a solution of (6.1) with $u_0 = v$. Since the PDE (6.1) is well-posed w.r.t. the L_2-norm [35], then for any $\epsilon > 0$, there exists $\delta > 0$ such that for any $\|u - v\|_{L_2(\Omega)} < \delta$, we have $\|Mu - Mv\|_{L_2(\Omega)} < \epsilon$. In other words, M is well-defined and continuous on $L_2(\Omega)$.

Definition 6.2. *A method $M_{h,\tau}$ is said to converge to M, if for any $v \in L_2(\Omega)$, we have*

$$\|T_h M v - M_{h,\tau} T_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \to 0 \quad \text{as } \tau, h \to 0.$$

Theorem 6.1. *With Assumption 6.1-6.3 and assuming that $Mv \in C^\infty(\Omega)$ whenever $v \in C^\infty(\Omega)$, the iterative algorithm* (2.31) *with threshold* (6.6) *converges to the nonlinear diffusion* (6.1)*.*

Proof. Given any $u \in L_2(\Omega)$ and a given $\delta > 0$, we can always find $v \in C^\infty(\Omega)$ such that $\|u - v\|_{L_2(\Omega)} < \delta$. Then by triangular inequality, we have

$$\begin{aligned}\|T_h M u - M_{h,\tau} T_h u\|_{\ell_{2,h}(\mathbb{O}_h^2)} \le \|T_h M u - T_h M v\|_{\ell_{2,h}(\mathbb{O}_h^2)} + \|T_h M v - M_{h,\tau} T_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \\ + \|M_{h,\tau} T_h v - M_{h,\tau} T_h u\|_{\ell_{2,h}(\mathbb{O}_h^2)},\end{aligned}$$

where the three terms on the right-hand-side of the inequality correspond to well-posedness, consistency and stability respectively. Note that $\|T_h M u - T_h M v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \le C_T \|Mu - Mv\|_{L_2(\Omega)}$. Thus, the first term above can be made arbitrarily small for a properly chosen v due to well-posedness of the problem M. Same argument holds for the last term above as well, due to stability in Proposition 6.3. We now show that

$$\|T_h M v - M_{h,\tau} T_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \to 0, \quad \text{as } h, \tau \to 0.$$

Indeed, we have

$$\begin{aligned}\|T_h M v - M_{h,\tau} T_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \le \|T_h M v - R_h M v\|_{\ell_{2,h}(\mathbb{O}_h^2)} + \|R_h M v - M_{h,\tau} R_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \\ + \|M_{h,\tau} T_h v - M_{h,\tau} R_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)}.\end{aligned}$$

By (6.11) and that $Mv \in C^\infty(\Omega)$, the first term goes to zero. By our previous consistency analysis in Proposition 6.1, we have the second term go to zero as well. Finally, following a similar proof of stability in Proposition 6.3, we have

$$\|M_{h,\tau} T_h v - M_{h,\tau} R_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \le \|T_h v - R_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \to 0.$$

This concludes the proof of this theorem. □

6.2 Behavior of the Generic Wavelet Frame Shrinkage (6.2)

We will analyze the asymptotic behavior of the iterative wavelet frame shrinkage (6.2) under suitable assumptions, which is given as follows.

Assumptions 6.4. *We make the following assumptions on the wavelet frame system we use and the threshold function $g_\ell = g_\ell(\xi_1, \cdots, \xi_L, h, \tau)$:*

1. *Let $\boldsymbol{W}$ be the transform operator of a tensor-product wavelet frame system constructed from a univariate B-spline function, and assume that $\widetilde{\boldsymbol{W}}$ is the transform operator of the canonical duel system.*

2. *We assume that, for each $0 \leq \ell \leq L$, $g_\ell \in C^1(\mathbb{R}^L)$, $0 \leq g_\ell \leq 1$ and*

$$\Big|\frac{\partial}{\partial \xi_j}\Big(\xi_\ell\,\left(1 - g_\ell(\xi_1, \ldots, \xi_L, h, \tau)\right)\Big)\Big| \leq 1,\ \forall 1 \leq j \leq L,$$

for every $(\xi_1, \cdots, \xi_L) \in \mathbb{R}^L$, and $h, \tau > 0$ small enough.

Same as before, we assume $\Omega = (0,1)^2$ is discretized by the dyadic grids $\{\boldsymbol{k}2^{-n} : 0 < k_1, k_2 < 2^n, n \geq 1\}$. Then, $h = 2^{-n}$ and $\tau = O(2^{-np})$ for some $p > 0$. The value p comes naturally from the 2nd assumption of Assumption 6.4 (e.g. $p = 2$ in Section 6.1). We define the adjoint operator of T_h as $T_h^* : \ell_{2,h}(\mathbb{O}_h^2) \mapsto L_2(\Omega)$:

$$T_h^* \boldsymbol{v} = 2^{-n} \sum_{\boldsymbol{j} \in \mathbb{O}_h^2} v_{\boldsymbol{j}} \phi_{n,\boldsymbol{j}}.$$

We note that $\|T_h u\|_{\ell_{2,h}(\mathbb{O}_h^2)} \leq C_T \|u\|_{L_2(\Omega)}$ and $\|T_h^* \boldsymbol{u}\|_{L_2(\Omega)} \leq C_T \|\boldsymbol{u}\|_{\ell_{2,h}(\mathbb{O}_h^2)}$.

With Assumption 6.4, we have the following theorem.

Theorem 6.2. *Suppose all requirements in Assumption 6.4 are satisfied. For any given artificial time $t > 0$, let $K = \lfloor t/\tau \rfloor$ be the stopping iteration of (6.2).*

1. *For any given initial data $v \in L_2(\Omega)$, we denote the K step iterative procedure based on the first formula of (6.2) as a nonlinear operator $T_h^* M_h^{\boldsymbol{S}} T_h : L_2(\Omega) \mapsto L_2(\Omega)$:*

$$T_h^* M_h^{\boldsymbol{S}} T_h v := T_h^* \Big(\prod_{k=0}^{K} \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k} \boldsymbol{W} \Big) T_h v.$$

   Then, for every $v \in L_2(\Omega)$, the sequence $\{T_h^* M_h^{\boldsymbol{S}} T_h v\}_h$ has a weakly converging subsequence in $L_2(\Omega)$. Furthermore, the iterative algorithm (6.2) is stable in the follow sense: for any $\epsilon > 0$, there is a $\delta > 0$ independent of h, such that for any $\|u - v\|_{L_2(\Omega)} < \delta$, we have

$$\|T_h^* M_h^S T_h u - T_h^* M_h^S T_h v\|_{L_2(\Omega)} < \epsilon.$$

2. *Denote the K step iterative procedure based on the second formula of (6.2) as a nonlinear operator $T_h^* M_h^{\boldsymbol{\mathcal{T}}} T_h : L_2(\Omega) \mapsto L_2(\Omega)$:*

$$T_h^* M_h^{\boldsymbol{\mathcal{T}}} T_h v := T_h^* \Big(\prod_{k=0}^{K} \widetilde{\boldsymbol{W}}^\top \boldsymbol{\mathcal{T}}_{\boldsymbol{\theta}^k} \boldsymbol{W} \Big) T_h v.$$

   Same conclusions in 1. also hold for $T_h^* M_h^{\boldsymbol{\mathcal{T}}} T_h$.

Proof. We shall only prove part 1., since the proof of part 2. is similar. Note from the first set of assumptions of Assumption 6.4, we have $\|\boldsymbol{W}\|_{\ell_{2,h}(\mathbb{O}_h^2)} = C_W$ and $\|\widetilde{\boldsymbol{W}}\|_{\ell_{2,h}(\mathbb{O}_h^2)} = \frac{1}{C_W}$. Therefore, we have

$$\begin{aligned}
\|T_h^* M_h^{\boldsymbol{S}} T_h v\|_{L_2(\Omega)} &\leq \frac{C_T}{C_W} \Big\| \boldsymbol{S}_{\boldsymbol{\alpha}^k} \boldsymbol{W} \Big(\prod_{k=0}^{K-1} \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k} \boldsymbol{W} \Big) T_h v \Big\|_{\ell_{2,h}(\mathbb{O}_h^2)} \\
&\leq \frac{C_T}{C_W} \Big\| \boldsymbol{W} \Big(\prod_{k=0}^{K-1} \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k} \boldsymbol{W} \Big) T_h v \Big\|_{\ell_{2,h}(\mathbb{O}_h^2)} \leq C_T \Big\| \Big(\prod_{k=0}^{K-1} \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^k} \boldsymbol{W} \Big) T_h v \Big\|_{\ell_{2,h}(\mathbb{O}_h^2)} \\
&\leq \cdots \leq C_T \|T_h v\|_{\ell_{2,h}(\mathbb{O}_h^2)} \leq C_T^2 \|v\|_{L_2(\Omega)},
\end{aligned}$$

where the second inequality follows from the assumption $0 \leq g_\ell \leq 1$. This shows that $\{T_h^* M_h^{\boldsymbol{S}} T_h v\}_h$ is bounded, hence it has a weakly converging subsequence in $L_2(\Omega)$.

On the other hand, using the second inequality of 2. of Assumption (6.4) and following a similar derivation as in the proof of equi-continuity in Proposition 6.3, we can easily show that, for any $u, v \in L_2(\Omega)$ and some constant $C > 0$, we have

$$\|T_h^* M_h^{\boldsymbol{S}} T_h u - T_h^* M_h^{\boldsymbol{S}} T_h v\|_{L_2(\Omega)} \leq C\|u - v\|_{L_2(\Omega)},$$

which gives us the stability. □

7 Numerical Simulations and Comparisons

In this section, we conduct numerical experiments using some of the algorithms we discussed in earlier sections. Recall the general image restoration model

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{f} + \boldsymbol{\eta},$$

where $\boldsymbol{f}$ is the observed image and $\boldsymbol{\eta}$ is assumed to be Gaussian white noise. We take image deblurring as the specific image restoration problem, where $\boldsymbol{A}$ is the convolution operator with the kernel generated in MATLAB by "fspecial('gaussian',11,1.5)". To measure quality of the restored image, we use the PSNR value defined by

$$\text{PSNR} := -20 \log_{10} \frac{\|u - \widetilde{u}\|_2}{N},$$

where u and $\widetilde{u}$ are the original and restored images respectively, and N is total number of pixels in u.

We will compare performance of some of the algorithms discussed earlier in this paper for image deblurring. For convenience of the readers, we list these algorithms here. Note that all the parameters that are not specifically mentioned below are chosen manually for optimal image restoration quality. For Algorithm 1 and 2, we stop our iteration when we have the highest PSNR values of the restored images. For the rest of the algorithms, since we have image restoration model embedded in the algorithms, we can let the iteration run until convergence. In our experiments, we adopt the following stopping criterion for Algorithms 3-6:

$$\|\boldsymbol{u}^k - \boldsymbol{u}^{k-1}\|_2 / \|\boldsymbol{f}\|_2 < 10^{-5}.$$

All of the following algorithms are implemented on a Windows laptop with Intel Core i7 processor (1.73 GHz) and 8GB memory.

List of algorithms used for comparison.

1. Perona-Malik equation with standard discretization (**PM-SD**) [52]:

$$u_t = \operatorname{div}\big(g(|G_\sigma * \nabla u|^2)\nabla u\big). \qquad \textbf{(PM)}$$

We take $g(\xi) = \frac{1}{1+\xi/K}$.

2. Perona-Malik equation discretized by Haar (**PM-Haar**) and piecewise linear (**PM-Linear**) B-spline wavelet frame systems:

$$\boldsymbol{u}^k = \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}).$$

Here we choose the level of decomposition Lev = 1 for $\boldsymbol{W}$. The shrinkage operator $\boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{d}^{k-1}) = \{S_{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d}^{k-1})}(\boldsymbol{d}^{k-1}) : 1 \leq \ell \leq L\}$ is chosen as in (3.21) for $\ell = 1, 2$ with $g(\xi) = 1/(1 + \xi/K)$. We choose $S_{\alpha_{\ell,\boldsymbol{n}}}$ for $3 \leq \ell \leq L$ as in Corollary 3.2 with $s_\ell = |\boldsymbol{\alpha}_1| + |\boldsymbol{\beta}_1| = 2$ and

$$g_\ell\Big(\frac{\widetilde{C}_1^{(1)}\xi_1}{h^{|\boldsymbol{\beta}_1|}}, \frac{\widetilde{C}_2^{(2)}\xi_2}{h^{|\boldsymbol{\beta}_2|}}, \ldots, \frac{\widetilde{C}_2^{(L)}\xi_L}{h^{|\boldsymbol{\beta}_L|}}\Big) = g\Big(\sum_{|\boldsymbol{\beta}_{\ell'}| = |\boldsymbol{\beta}_\ell|} \frac{\xi_{\ell'}^2}{C_{\boldsymbol{\beta}_{\ell'}}^{(\ell')} h^{\boldsymbol{\beta}_{\ell'}}}\Big).$$

3. Iterative soft-thresholding algorithm by Haar (**IST-Haar**) and piecewise linear (**IST-Linear**) B-spline wavelet frame systems:

$$\boldsymbol{u}^k = (I - \mu\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{W}^\top \boldsymbol{\mathcal{T}}_{\boldsymbol{\lambda}}^2(\boldsymbol{W}\boldsymbol{u}^{k-1}) + \mu\boldsymbol{A}^\top\boldsymbol{f}.$$

Here we choose the level of decomposition Lev = 3 for $\boldsymbol{W}$ and $\boldsymbol{\lambda}$ is some fixed threshold.

4. Adaptive multiplicative-thresholding by Haar (**AMT-Haar**) and piecewise linear (**AMT-Linear**) B-spline wavelet frame systems:
$$\boldsymbol{u}^k = (I - \mu\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{W}^\top\boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}) + \mu\boldsymbol{A}^\top\boldsymbol{f}.$$
Here we choose the level of decomposition Lev = 3 for $\boldsymbol{W}$. We take
$$\boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}) = \{S_{\alpha_{l,\ell,\boldsymbol{n}}(\boldsymbol{W}_l\boldsymbol{u}^{k-1})}(\boldsymbol{W}_l\boldsymbol{u}^{k-1}):\ 0 \le l \le \text{Lev}-1, 1 \le \ell \le L\},$$
where $\boldsymbol{W}_l\boldsymbol{u} = \{\boldsymbol{W}_{l,\ell}\boldsymbol{u}:\ 1 \le \ell \le L\}$. The shrinkage operator $S_{\alpha_{l,\ell,\boldsymbol{n}}}$ is chosen the same as in (2) for $l = 0$ and $S_{\alpha_{l,\ell,\boldsymbol{n}}} = S_{\alpha_{0,\ell,\boldsymbol{n}}}$ for $l > 0$.

5. Adaptive soft-thresholding by Haar (**AST-Haar**) and piecewise linear (**AST-Linear**) B-spline wavelet frame systems:
$$\boldsymbol{u}^k = (I - \mu\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{W}^\top\boldsymbol{\mathcal{T}}^2_{\boldsymbol{\theta}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}) + \mu\boldsymbol{A}^\top\boldsymbol{f}.$$
Here, the threshold
$$\boldsymbol{\theta}^{k-1} = \left\{\theta_{l,\ell}(\boldsymbol{W}_{l,1}\boldsymbol{u}^{k-1}, \boldsymbol{W}_{l,2}\boldsymbol{u}^{k-1}, \ldots, \boldsymbol{W}_{l,L}\boldsymbol{u}^{k-1}):\ 0 \le l \le \text{Lev}-1, 1 \le \ell \le L\right\},$$
is chosen as
$$\theta_{l,\ell}(\xi_1, \xi_2, \ldots, \xi_L) = C_\ell g\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|} \frac{\xi_{\ell'}^2}{C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}} h^{\boldsymbol{\beta}_{\ell'}}}\Big),$$
for all $l = 0, 1, \ldots, \text{Lev}-1$. Here we choose $g = 1/(1 + x^{0.25}/K)$, Lev = 3 and $C_\ell > 0$ some fixed constants.

6. Image-restoration embedded diffusion with standard discretization (**IRED-SD**) [52]:
$$u_t = \operatorname{div}\big(g(|G_\sigma * \nabla u|^2)\nabla u\big) - \kappa A^\top(Au - f). \qquad (\textbf{IRED})$$

Table 1: Comparisons for image deconvolution. PSNR values of the restored images by Algorithm (1) and (2) (without the knowledge of image restoration model): PM-SD, PM-Haar and PM-Linear. The bolded number indicates the best PSNR value for each image example.

Image Name	PM-SD	PM-Haar	PM-Linear
Barbara	24.8097	24.9080	**24.9625**
Boat	23.4765	23.5915	**23.6089**
Peppers	23.6635	23.8096	**23.8203**

Table 2: Comparisons for image deconvolution. PSNR values of the restored images by algorithms (3)-(6): IST-Haar, IST-Linear, AMT-Haar, AMT-Linear, AST-Haar, AST-Linear and IRED-SD. The bolded number indicates the best PSNR value for each image example.

Image Name	IST-Haar	IST-Linear	AMT-Haar	AMT-Linear	AST-Haar	AST-Linear	IRED-SD
Barbara	25.5111	25.5204	**25.5987**	25.5651	25.4221	25.5398	25.5482
Boat	24.6767	24.7341	24.8784	24.7906	24.7158	**24.9417**	24.6955
Peppers	24.9834	25.0974	25.1419	**25.4677**	25.0221	25.3208	25.2149

Our results in Table 1 shows that the discretization provided by wavelet frame shrinkage is better than the standard discretization by [52] in terms of restoration quality. The results in Table 2 shows that with image restoration model being properly incorporated in the algorithms, better results can be obtained. More importantly, the new algorithms AMT and AST outperform the IST algorithm that is commonly used in wavelet frame based image restoration problems. In addition, the AMT algorithm, which is a discretization of the IRED, generates better results than the standard discretization of the IRED.

Figure 1: Test images: original (first row) and observed (second row). The PSNR values of the observed images are 23.4519 (Barbara), 22.4428 (Boat) and 22.5482 (Peppers). The image size of "Barbara" is 195×195, and both "Boat" and "Peppers" has size 256×256.

In Section 5.2, we proposed accelerated version of AMT and IRED. The accelerated algorithm and PDE produce results of similar quality as their corresponding non-accelerated versions, while the total number of iterations and computation time are greatly reduced. For the same image deblurring problem and with the same stopping criterion, we implemented the accelerated (APG) version of AMT and IRED following the generic formulas of (5.15) and (5.20). To be more precise, we implement the following accelerated AMT (**A-AMT**)

$$\boldsymbol{u}^k = (I - \mu \boldsymbol{A}^\top \boldsymbol{A}) \boldsymbol{W}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}} \Big((1+\gamma^{k-1}) \boldsymbol{W} \boldsymbol{u}^{k-1} - \gamma^{k-1} \boldsymbol{W} \boldsymbol{u}^{k-2} \Big) + \mu \boldsymbol{A}^\top \boldsymbol{f}, \quad k = 1, 2, \cdots,$$

where the shrinkage operator $\mathcal{S}_{\boldsymbol{\alpha}}$ is chosen exactly the same as that of Algorithm 4 above. The accelerated IRED takes the form

$$u_{tt} + Cu_t = \operatorname{div}\big(g(|G_\sigma * \nabla u|^2)\nabla u\big) - \kappa A^\top (Au - f). \qquad (\textbf{A-IRED}),$$

which is discretized by the standard finite differencing [52]. We shall call the A-IRED discretized by such finite differencing as **A-IRED-FD**. The parameters for A-AMT are chosen exactly the same as AMT, and we choose $C = 0.2$ for A-IRED. A comparison of efficiency is summarized in Table 3.

Table 3: Comparisons for image deconvolution. PSNR values of the restored images by algorithms (3)-(6): IST-Haar, IST-Linear, AMT-Haar, AMT-Linear, AST-Haar, AST-Linear and IRED-SD. The bolded number indicates the best PSNR value for each image example.

	AMT-Linear		A-AMT-Linear		IRED-FD		A-IRED-FD	
Image Name	Time (sec.)	PSNR	Time (sec)	PSNR	Time (sec)	PSNR	Time (sec)	PSNR
Barbara	23.38	25.5651	15.34	25.5657	7.11	25.5482	3.50	25.5555
Boat	40.44	24.7906	23.27	24.7322	9.62	24.6955	5.06	24.6916
Peppers	55.95	25.4677	31.46	25.4678	17.21	25.2149	8.73	25.2201

References

[1] K. Amaratunga, J. R. Williams, S. Qian, and J. Weiss, Wavelet Galerkin solutions for one-dimensional partial differential equations, International Journal for Numerical Methods in Engineering, 37 (1994) 2703–2716.

[2] S. Angenent, Parabolic equations for curves on surfaces: part II. Intersections, blow-up and generalized solutions, Annals of Mathematics, 133 (1991) 171–215.

[3] G. Aubert and K. P. Kornprobst, Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations, Vol. 147, Springer Science+ Business Media, 2006.

[4] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci., 2 (2009) 183–202.

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

[6] H. Brezis, Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Spaces de Hilbert, North-Holland Publishing Company (1973).

[7] J. Cai, R. Chan, L. Shen, and Z. Shen, Convergence analysis of tight framelet approach for missing data recovery, Adv. Comput. Math., 31 (2008) 1–27.

[8] J. Cai, B. Dong, S. Osher, and Z. Shen, Image restorations: total variation, wavelet frames and beyond, J. American Mathematical Society, 25 (2012) 1033–1089.

[9] J. Cai, S. Huang, H. Ji, Z. Shen, and G. Ye, Data-driven tight frame construction and iamge denoisng, preprint (2013).

[10] J. Cai and Z. Shen, Framelet based deconvolution, J. Comp. Math., 28 (2010) 289–308.

[11] J. Cai, R. Chan, and Z. Shen, A framelet-based image inpainting algorithm, Appl. Comput. Harmon. Anal., 24 (2008) 131–149.

[12] J. Cai, R. Chan, and Z. Shen, Simultaneous cartoon and texture inpainting, Inverse Problems and Imaging, 4 (2010) 379–395.

[13] J. Cai, S. Osher, and Z. Shen, Split Bregman methods and frame based image restoration, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 8 (2009) 337–369.

[14] J. Cai, R.H. Chan, L. Shen, and Z. Shen, Restoration of chopped and nodded images by framelets, SIAM J. Sci. Comput., 24 (2008) 1205–1227.

[15] F. Catte, P. Lions, J. Morel, and T. Coll, Image selective smoothing and edge detection by nonlinear diffusion, SIAM J. Numer. Anal., 29 (1992) 182–193.

[16] A. Chai and Z. Shen, Deconvolution: A wavelet frame approach, Numerische Mathematik, 106 (2007) 529–587.

[17] R. Chan, Z. Shen, and T. Xia, A framelet algorithm for enhancing video stills, Appl. Comput. Harmon. Anal., 23 (2007) 153–170.

[18] R. Chan, T. Chan, L. Shen, and Z. Shen, Wavelet algorithms for high-resolution image reconstruction, SIAM Journal Sci. Comput., 24 (2003) 1408–1432.

[19] R. Chan, T. Chan, L. Shen, and Z. Shen, Wavelet algorithms for high-resolution image reconstruction, SIAM Journal on Scientific Computing, 24 (2003) 1408–1432.

[20] R. Chan, S. Riemenschneider, L. Shen, and Z. Shen, Tight frame: an efficient way for high-resolution image reconstruction, Appl. Comput. Harmon. Anal., 17 (2004) 91–115.

[21] P. Combettes and V. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Modeling and Simulation, 4 (2006) 1168–1200.

[22] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Lecture Notes, Society for Industrial and Applied Mathematics (1992).

[23] I. Daubechies, B. Han, A. Ron, and Z. Shen, Framelets: MRA-based constructions of wavelet frames, Appl. Comput. Harmon. Anal., 14 (2003) 1–46.

[24] I. Daubechies, G. Teschke, and L. Vese, Iteratively solving linear inverse problems under general convex constraints, Inverse Problems and Imaging, 1 (2007) 29–46.

[25] S. Didas, Denoising and Enhancement of Digital Images–variational Methods, Integrodifferential Equations, and Wavelets, Ph.D. Dissertation, Saarland University (2008).

[26] S. Didas, J. Weickert, and B. Burgeth, Properties of higher order nonlinear diffusion filtering, J. Mathematical Imaging and Vision, 35 (2009) 208–226.

[27] B. Dong, H. Ji, J. Li, Z. Shen, and Y. Xu, Wavelet frame based blind image inpainting, Appl. Comput. Harmon. Anal., 32 (2011) 268–279.

[28] B. Dong, J. Li, and Z. Shen, X-ray CT image reconstruction via wavelet frame based regularization and Radon domain inpainting, J. Scientific Computing, 54 (2013) 333–349.

[29] B. Dong and Y. Zhang, An efficient algorithm for ℓ_0 minimization in wavelet frame based image restoration, J. Scientific Computing, 54 (2013) 350–368.

[30] B. Dong, A. Chien, and Z. Shen, Frame based segmentation for medical images, Commun. Math. Sci., 9 (2010) 551–559.

[31] B. Dong and Z. Shen, Frame based surface reconstruction from unorganized points, J. Computational Physics, 230 (2011) 8247–8255.

[32] 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).

[33] D. Donoho, De-noising by soft-thresholding, IEEE Trans. Inform. Th., 41 (1995) 613–627, 1995.

[34] M. Elad, J. Starck, P. Querre, and D. Donoho, Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA), Appl. Comput. Harmon. Anal., 19 (2005) 340–358.

[35] L. Evans, Partial Differential Equations, Graduate Studies in Mathematics, Vol 19, American Mathematical Society (1998).

[36] M. Fadili and J. Starck, Sparse representations and Bayesian image inpainting, Proc. SPARS, 5 (2005).

[37] M. Fadili, J. Starck, and F. Murtagh, Inpainting and zooming using sparse representations, The Computer Journal, 52 (2009) 64–79.

[38] M. Li, Z. Fan, H. Ji, and Z. Shen, Wavelet frame based algorithm for 3D reconstruction in electron microscopy, accepted by SIAM Journal on Scientific Computing (2013).

[39] M. Figueiredo and R. Nowak, An EM algorithm for wavelet-based image restoration, IEEE Trans. Image Proc., 12 (2003) 906–916.

[40] M. Figueiredo and R. Nowak, A bound optimization approach to wavelet-based image deconvolution, in: IEEE International Conference on Image Processing (ICIP 2005), Genoa, Italy, vol. 2, pp. 782–785.

[41] Y. Giga, S. Goto, H. Ishii, and M.-H. Sato, Comparison principle and convexity preserving properties for singular degenerate parabolic equations on unbounded domains, Indiana Univ. Math. J. 40 (1991) 443–470.

[42] Y. Giga and S. Goto, Geometric evolutions of phase-boundaries, in On the evolution of phase-boundaries, edited by M. Gurtin and Mc Fadden, IMA Vols. Math. Appl. Springer, Berlin, 43 (1992) 51–66.

[43] R. Glowinski, W. Lawton, M. Ravachol, and E. Tenenbaum. Wavelet solutions of linear and nonlinear elliptic, parabolic and hyperbolic problems in one space dimension. Computing methods in applied sciences and engineering (1990): 55–120.

[44] T. Goldstein and S. Osher, The split Bregman algorithm for L_1 regularized problems, SIAM J. Imaging Sci., 2 (2009) 323–343.

[45] D. W. Hoffman and J. W. Cahn, A vector thermodynamics for anisotropic surfaces, I: fundamentals and application to plane surface junctions, Surf. Sci., 31 (1972) 368–388.

[46] X. Jia, B. Dong, Y. Lou, and S. Jiang, GPU-based iterative cone-beam CT reconstruction using tight frame regularization, Phys. Med. Biol., 56 (2011) 3787–3807.

[47] Q. Jiang, Correspondence between frame shrinkage and high-order nonlinear diffusion, Appl. Numer. Math., 62 (2012) 51–66.

[48] A. Latto, H. L. Resnikoff, and E. Tenenbaum. The evaluation of connection coefficients of compactly supported wavelets. In Proceedings of the French-USA Workshop on Wavelets and Turbulence. NY: Springer-Verlag (1991, June) 76–89.

[49] M. Lysaker, A. Lundervold, and X. Tai, Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time, IEEE Trans. Image Proc., 12 (2003) 1579–1590.

[50] 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, 22(4) (2000) 387–405.

[51] S. Osher and L. Rudin, Feature-oriented image enhancement using shock filters, SIAM J. Numer. Anal., 27 (1990) 919–940.

[52] P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Machine Intelligence, 12 (1990) 629–639.

[53] S. Qian, and J. Weiss. Wavelets and the numerical solution of partial differential equations. Journal of Computational Physics, 106(1) (1993) 155–175.

[54] A. Ron and Z. Shen, Affine systems in $L_2(\mathbb{R}^d)$: the analysis of the analysis operators, J. Funct. Anal., 148 (1997) 408–447.

[55] A. Ron and Z. Shen, Affine systems in $L_2(\mathbb{R}^d)$ II: Dual systems, J. Fourier Anal. Appl., 3 (1997) 617–637.

[56] L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992) 259–268.

[57] Z. Shen, Wavelet frames and image restorations, Proceedings of the International Congress of Mathematicians, Hyderabad, India (2010).

[58] Z. Shen, K. Toh, and S. Yun, An accelerated proximal gradient algorithm for frame based image restorations via the balanced approach, SIAM J. Imaging Sci., 4 (2011) 573–596.

[59] H. Spohn, Interface motion in models with stochastic dynamics, J. Stat. Phys., 71 (1993) 1081–1132.

[60] J. Starck, M. Elad, and D. Donoho, Image decomposition via the combination of sparse representations and a variational approach, IEEE Trans. Image Proc., 14 (2005) 1570–1582.

[61] C. Vogel and M. Oman, Iterative methods for total variation denoising, SIAM Journal on Scientific Computing, 17(1) (1996) 227–238.

[62] G. Wei, Generalized Perona-Malik equation for image restoration, IEEE Signal Process. Lett., 6 (1999) 165–167.

[63] M. Welk, D. Theis, and J. Weickert, Variational deblurring of images with uncertain and spatially variant blurs, in: Pattern Recognition, Springer Berlin Heidelberg, 2005, pp.485–492.

[64] P. Mrázek, J. Weickert, and G. Steidl, Correspondences between wavelet shrinkage and nonlinear diffusion, in: L.D. Griffin and M. Lillholm (Eds.), Scale-Space Methods in Computer Vision, Lecture Notes in Computer Science, vol. 2695, Springer, Berlin, 2003, pp.101–116.

[65] P. Mrázek and J. Weickert, From two-dimensional nonlinear difussion to coupled Haar wavelet shrinkage, J. Visual Communication and Image Representation, 18 (2007) 162–175.

[66] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs, SIAM J. Numer. Anal., 42 (2004) 686–713.

[67] J. Weickert, Anisotropic Diffusion in Image Processing, B.G. Teubner, Stuttgart (1998).

[68] Y. You and M. Kaveh, Image enhancement using fourth order partial differential equations, in: 32nd Asilomar Conf. Signals, Systems, Computers, 2 (1998) 1677–1681.

[69] Y. You and M. Kaveh, Fourth-order partial differential equations for noise removal, IEEE Trans. Image Proc., 9 (2000) 1723–1730.

[70] Y. Zhang, B. Dong, and Z. Lu, ℓ_0 minimization of wavelet frame based image restoration, Math. Comp., 82 (2013) 995–1015.