APPROXIMATION OF FRAME BASED MISSING DATA RECOVERY

JIAN-FENG CAI*, ZUOWEI SHEN†, AND GUI-BO YE‡

Abstract. Recovering missing data from its partial samples is a fundamental problem in mathematics and it has wide range of applications in image and signal processing. While many such algorithms have been developed recently, there are very few papers available on their error estimations. This paper is to analyze the error of a frame based data recovery approach from random samples. In particular, we estimate the error between the underlying original data and the approximate solution that interpolates (or approximates with an error bound depending on the noise level) the given data that has the minimal ℓ_1 norm of the canonical frame coefficients among all the possible solutions.

1. Introduction. Recovering missing data from its partial samples is a fundamental problem in mathematics and it has wide range of applications in image and signal processing. The problem is to recover the underlying image or signal $\boldsymbol{p}$ from its partial observations given by

$$\boldsymbol{g}[k] = \begin{cases} \boldsymbol{p}[k] + \boldsymbol{\theta}[k], & k \in \Lambda, \\ \text{unknown}, & k \in \Omega \backslash \Lambda, \end{cases} \tag{1.1}$$

where $\boldsymbol{\theta}$ is the error contained in the observed data. Here the set Ω (see also (1.2)) is the domain where the underlying data is defined and Λ is a subset of Ω where we have the observed data. The observed data could be part of sound, images, time-varying measurement values and sensor data. The task is to recover the missing data on $\Omega \backslash \Lambda$. There are many methods to deal with this problem under many different settings, e.g., [3, 4, 10, 25, 38] for image inpainting, [11, 14, 15] for matrix completion, [29, 55] for regression in machine learning, [8, 9, 20, 23, 24] for framelet-based image deblurring, [41, 45] for surface reconstruction in computer graphics, and [16, 22, 26] for miscellaneous applications. We forgo to give a detailed survey on this fast developing area and the interested reader should consult the references mentioned above for the details. Instead, the focus of this paper is to establish the approximation properties of a frame based data recovery method.

The settings of (1.1) considered in this paper are as follows. Let

$$\Omega = \{k = (k_1, \dots, k_d) : k \in \mathbb{Z}^d, 0 \le k_i < N, i = 1, \dots, d\}, \tag{1.2}$$

where N is a given positive integer. Let

$$\Lambda \subset \Omega, \quad |\Lambda| = m, \quad \Lambda \text{ is uniformly randomly drawn from } \Omega. \tag{1.3}$$

Define $\rho := m/|\Omega|$ be the density of the known pixels. Then, in (1.1), the observed data $\boldsymbol{g}$ and the error $\boldsymbol{\theta}$ are given and fixed, although the error $\boldsymbol{\theta}$ may be viewed as a particular realization of some random variables, e.g., i.i.d. Gaussian noise. Hence, in this setting, the only random variables are Λ, which is uniformly randomly chosen from Ω.

One of the most important examples of our model is image recovery from random sampled pixels, which occurs when part of the pixel is randomly missing due to, e.g., the unliable communication channel [7, 26] or the corruption by a salt-and-pepper noise [12, 22]. One of such examples is shown in Figure 1.1. The task of image recovery is to restore the missing region from the incomplete pixels observed. Ideally, the restored image should possess shapes and patterns consistent with the given image in human vision. Therefore, we need to extract information such as edges and textures from the observed data to replace the corrupted part in such a way that it would look natural for human eyes. For this, it is often useful to restore images in a transform domain (e.g. tight frame transform) where the underlying image has a sparse approximation. This leads to a few frame based methods for image restorations as given in e.g. [10, 12, 38, 39].

In this paper, we give the error estimation for a frame based recovery method to solve (1.1)–(1.3). For this, we first introduce the concept of tight frame. See [35, 53] for an overview of tight frame. Let $\mathcal{H}$ be a

*Department of Mathematics, University of California, Los Angeles. Email: cai@math.ucla.edu.
†Department of Mathematics, National University of Singapore. Email: matzuows@nus.edu.sg.
‡Department of Computer Science, University of California, Irvine. Email: yeg@uci.edu.

(a) The 512×512 "peppers" image. (b) 50% pixels are randomly missing. (c) Recovered by (1.8). The algorithm employed is the split Bregman method in [13].

Fig. 1.1: Images

Hilbert space. A sequence $\{\boldsymbol{a}_n\}_{n\in\Gamma} \subset \mathcal{H}$ is a tight frame of $\mathcal{H}$ if for an arbitrary element $\boldsymbol{f} \in \mathcal{H}$

$$\|\boldsymbol{f}\|^2 = \sum_{n\in\Gamma} |\langle \boldsymbol{f}, \boldsymbol{a}_n\rangle|^2,$$

or, equivalently,

$$\boldsymbol{f} = \sum_{n\in\Gamma} \langle \boldsymbol{f}, \boldsymbol{a}_n\rangle \boldsymbol{a}_n. \tag{1.4}$$

For a given tight frame, the analysis operator $\mathcal{A}$ is defined as

$$\mathcal{A}\boldsymbol{f}[n] = \langle \boldsymbol{f}, \boldsymbol{a}_n\rangle, \quad \forall n \in \Gamma. \tag{1.5}$$

The sequence $\{\langle \boldsymbol{f}, \boldsymbol{a}_n\rangle\}_{n\in\Gamma}$ is called the canonical coefficients of the tight frame $\{\boldsymbol{a}_n\}_{n\in\Gamma}$. For recovery problem (1.1) with Ω defined by (1.2), we are working on the finite dimensional space $\mathcal{H} = \ell_2(\Omega)$. In this case, $\boldsymbol{a}_n$ is a sequence in $\ell_2(\Omega)$ and Γ is a finite set.

To measure the regularity of the underlying image or signal, one can employ the weighted ℓ_1 norm of the canonical frame coefficient. This is commonly used in image and signal processing literature. In this paper, we will use the weighted ℓ_1 norm of the canonical frame coefficient $\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)}$ for a given β in the form of the following

$$\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)} = \sum_{n\in\Gamma} 2^{\beta\Upsilon(n)} |\langle \boldsymbol{a}_n, \boldsymbol{f}\rangle|, \tag{1.6}$$

where Υ is a function mapping from Γ to $\mathbb{N}$ satisfying

$$\max\{\Upsilon(n) : n \in \Gamma\} \le \frac{1}{d}\log_2 |\Omega|. \tag{1.7}$$

Here the parameter β is to control the regularity of $\boldsymbol{f}$, and the function Υ is to make the weight more flexible so that it allows group weighting. It will be seen in Section 3.1 the usefulness and the explicit form of Υ in the case of framelet. As we know, signals and images are usually modeled by discontinuous functions, and the discontinuity possesses important information. Therefore, our assumption for β is always small in order to reflect the low regularity of the underlying signal. That is, we are only interested in signals of low regularity in this paper.

The focus of this paper is to study one of the analysis based approach using tight frame. We assume $\boldsymbol{p}$ satisfies $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)} < \infty$ and $\|\boldsymbol{p}\|_\infty \le M$, where M is a given constant. The first condition is the regularity of $\boldsymbol{p}$ and the second condition is the boundedness of each pixel of $\boldsymbol{p}$. In our model, the approximate solution $\boldsymbol{f}^\Lambda$ of the problem (1.1) is defined by:

$$\boldsymbol{f}^\Lambda = \arg\min\Big\{\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)} : \frac{1}{|\Lambda|}\sum_{k\in\Lambda}(\boldsymbol{f}[k]-\boldsymbol{g}[k])^2 \le \sigma^2,\quad \|\boldsymbol{f}\|_\infty \le M\Big\}. \tag{1.8}$$

It is clear that there exists at least one solution for the above minimization problem. Indeed, this follows from the facts that the constraint set $\{\mathbf{f} : \frac{1}{|\Lambda|}\sum_{k\in\Lambda}(\mathbf{f}[k]-\mathbf{g}[k])^2 \le \sigma^2, \|\mathbf{f}\|_\infty \le M\}$ is closed and bounded and the objective function $\|\mathcal{A}\mathbf{f}\|_{\ell_1(\beta,\Upsilon)}$ is continuous with respect to $\mathbf{f}$. Therefore, $\boldsymbol{f}^\Lambda$ is well defined and has a minimal weighted ℓ_1 norm of the canonical coefficient subject to reasonable constraints. Here, the constraint $\frac{1}{|\Lambda|}\sum_{k\in\Lambda}(\boldsymbol{f}[k]-\boldsymbol{g}[k])^2 \le \sigma^2$ is a data fitting term to (1.1) and σ^2 is the error bound. Therefore, $\boldsymbol{p}$ naturally satisfies this constraint. The constraint $\|\boldsymbol{f}\|_\infty \le M$ is to ensure that the recovered signal values are bounded by a preassigned number M. This constraint is usually inactive, i.e., solving (1.8) with or without this constraint gives the same solution in most numerical simulations as long as the original signal $\boldsymbol{p}$ also satisfies this constraint. When $m = |\Omega|$ and $\sigma = 0$, the unique solution of (1.8) is the original solution $\boldsymbol{p}$. Therefore, we are interested in the case when $m < |\Omega|$ and $\sigma \neq 0$. In figure 1.1, we give an example that shows (1.8) recovers the missing pixels of the image very well (the algorithm employed for solving (1.8) is the split Bregman method in [13]). The purpose of this paper is to show analytically that the errors of the recovered missing pixels are within the measurement error bound of the given pixels.

As for the energy $\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)}$ in (1.8), at top of the fact that it connects to the regularity of the underlying function where the data comes from, it can be interpreted as follows that links to the prior distribution of $\boldsymbol{f}$. In fact, we implicitly assume that the prior distribution of $\boldsymbol{f}$ satisfying

$$\text{Prob}\{\boldsymbol{f}\} = Const\cdot\exp\Big\{-\lambda\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)}\Big\}.$$

Hence, minimizing $\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)}$ is equivalent to maximizing the probability that the data occurs.

An efficient frame based algorithm is developed for some applications to solve (1.8) in [13]. The algorithm is implicitly based on the fact that $\boldsymbol{f}$ has a sparse approximation under the tight frame system used. A sparse approximation means majority of the canonical coefficients $\mathcal{A}\boldsymbol{f}$ are small and negligible. In this sense, (1.8) gives a sparse approximate solution of (1.1). However, there are big differences between the approach (1.8) here and compressed sensing (see e.g. [16–18, 32]) — one of the hottest research topics based on sparsity. Firstly, the requirement of sparsity here is much weaker than in compressed sensing. We do not require explicitly the sparsity of either $\boldsymbol{f}$ or its canonical frame coefficient. Instead, we assume the decay of the canonical frame coefficient in the sense that the weighted ℓ_1 norm (1.6) is bounded. Secondly, in basis pursuit of compressed sensing, the signal is synthesized by a sparse coefficient, hence it is a synthesis based approach. However, as mentioned before, the model (1.8) is an analysis based approach — the analyzed coefficient has a sparse approximation. There is a gap between the analysis and synthesis based approaches as pointed out in, e.g., [13, 37]. Last and most importantly, the matrix here does not satisfy the restricted isometry property (RIP) required in the theoretic analysis in compressed sensing. If we use a synthesis based approach instead of the analysis based approach (1.8), then the sensing matrix will be $\mathcal{P}_\Lambda\mathcal{A}^T$, where $\mathcal{P}_\Lambda$ is an operator satisfying $\mathcal{P}_\Lambda\boldsymbol{f}[k] = \boldsymbol{f}[k]$ for $k \in \Lambda$ and $\mathcal{P}_\Lambda\boldsymbol{f}[k] = 0$ for $k \in \Omega\backslash\Lambda$. Since usually each vector $\boldsymbol{a}_i$ (each row of $\mathcal{A}$) is locally supported, by a simple calculation, one finds that $\mathcal{P}_\Lambda\mathcal{A}^T$ has at least one column being the zero vector with a high probability. In turn, the sensing matrix does not satisfy the RIP with high probability. The matrix $\mathcal{P}_\Lambda$ does not satisfy the concentration inequality in [52]. Moreover, due to the compact support property of the frame elements $\boldsymbol{a}_i$, the incoherence conditions (see [33] for instance) between the column vectors of matrix $\mathcal{P}_\Lambda$ and the row vectors of $\mathcal{A}$ may not hold. This causes that there contains no enough information in the observed pixels for exact signal recovery. Therefore, the compressed sensing theory cannot be applied here, even the synthesis based approach is used.

This paper is to bound the error between the underlying unknown data $\boldsymbol{p}$ and the approximate solution $\boldsymbol{f}^\Lambda$ given by (1.8). It is clear that one can only expect that the recovered error is within the level of the measurement error up to a constant. It is trivially true when the density $\rho = 1$ (i.e., $\Lambda = \Omega$), since

$$\frac{1}{|\Omega|}\sum_{k\in\Omega}(\boldsymbol{f}^{\Lambda}[k]-\boldsymbol{p}[k])^2 \le \frac{2}{|\Omega|}\sum_{k\in\Omega}(\boldsymbol{f}^{\Lambda}[k]-\boldsymbol{g}[k])^2 + \frac{2}{|\Omega|}\sum_{k\in\Omega}(\boldsymbol{g}[k]-\boldsymbol{p}[k])^2 \le 4\sigma^2.$$

We are interested to know what will happen when the density $\rho < 1$. In fact, we will show that, under some mild assumptions, with probability $1-\delta$ for an arbitrary fixed $\delta \in (0,1)$, the error between $\boldsymbol{p}$ and $\boldsymbol{f}^{\Lambda}$ satisfies

$$\frac{1}{|\Omega|}\|\boldsymbol{f}^{\Lambda}-\boldsymbol{p}\|^2_{\ell_2(\Omega)} \le C\rho^{-\frac{1}{2}}\sqrt{\log_2|\Omega|}(|\Omega|)^{-b}\log\frac{1}{\delta} + \frac{16}{3}\sigma^2, \tag{1.9}$$

where b is a positive constant and will be given explicitly in Theorem 2.2, and C is a positive constant independent of $\rho, |\Omega|, \delta$ or σ when the tight framelets are used . Roughly, it says that as long as the data set is sufficiently large, one has a pretty good chance to recover the original data within the measurement error bound by solving (1.8).

The main difficulty here is that the underlying solution has a low regularity. The analysis here is based on the combination of the uniform law of large numbers, which is standard in classical empirical processes and statistical learning theory, and an estimation for its involved covering number. The covering number estimation given here is new involved, since the standard estimation for it is too large so that it is not good enough to derive the desired convergence rate. Our estimation for the covering number uses the special structure of the set and the max-flow min-cut theorem in graph theory. The error analysis here can be easily extended into more analysis based approaches, e.g. total variation method for imaging restorations.

The paper is organized as follows. In Section 2, we give our main results of approximation analysis for the frame based signal recovery method (1.8). Error estimations are given. Then, in Section 3, an application of our main results is illustrated. More precisely, we estimate the error of framelet based image recovery algorithms from random samples. Based on this, we further link the discrete approximation of solution to the function approximation of it in the content of multiresolution analysis and its associated tight frammlets given by [51]. Finally, the technical proofs of the critical lemmas and theorems are given in Section 4.

2. Error Analysis.

In this section, we give the error analysis of the model (1.8) for a given tight frame analysis operator $\mathcal{A}$. That is, we study the asymptotic property of $\|\boldsymbol{f}_{\Lambda}-\boldsymbol{p}\|_{\ell_2(\Omega)}$ with respect to $|\Omega|$. Here Λ is a data set with each element i.i.d drawn from uniform distribution of Ω and $|\Omega|$ denotes the cardinality of the set Ω. Such problem is well known in classical empirical processes [54] and statistical learning theory [55]. The most powerful tool used there is the uniform law of large numbers and our analysis is along this direction. To employ the uniform law of large numbers, the key issue is the capacity of the involved set. There are many tools to characterize the capacity of a set in the literature, e.g. VC-dimension [55], V_{γ}-dimension, P_{γ}-dimension [1], Rademacher complexities [2,46] and covering number [29]. As covering number is the most convenient and very powerful for metric space, we choose it to characterize the capacity of the involved set

$$\mathcal{M} = \left\{\boldsymbol{f}\in\ell_\infty(\Omega) : \|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)} \le \|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)}, \frac{1}{|\Lambda|}\sum_{k\in\Lambda}(\boldsymbol{f}[k]-\boldsymbol{g}[k])^2 \le \sigma^2,\ \|\boldsymbol{f}\|_\infty \le M\right\}. \tag{2.1}$$

Here, the constants M and σ are fixed. Notice that, with high probability, the underlying true solution $\boldsymbol{p}$ is in the set $\mathcal{M}$. Furthermore, according to the definition of $\boldsymbol{f}_{\Lambda}$ by (1.8), we have $\|\mathcal{A}\boldsymbol{f}_{\Lambda}\|_{\ell_1(\beta,\Upsilon)} \le \|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)}$ and obviously $\boldsymbol{f}_{\Lambda}\in\mathcal{M}$. Thus, the set $\mathcal{M}$ defined in (2.1) is the set we concerned.

To further illustrate our idea, we give the concept of the covering number, which is adapted to the settings of this paper.

Definition 2.1. *Let $\mathcal{M}\subset\mathbb{R}^{|\Omega|}$ and $\eta>0$ be given. The* **covering number** $\mathcal{N}(\mathcal{M},\eta)$ *is the minimal number of the ℓ_∞ balls with radius η in $\mathcal{M}$ that cover $\mathcal{M}$.*

The main difficulty of this paper is to give a tight estimate of the covering number $\mathcal{N}(\mathcal{M},\eta)$ of the set $\mathcal{M}$ defined in (2.1). At first glance, $\mathcal{M}$ is a subset of $\left\{\boldsymbol{f}\in\ell_\infty(\Omega) : \|\boldsymbol{f}\|_\infty \le M\right\}$, which is a ball in finite dimensional Banach space $\ell_\infty(\Omega)$. We have a simple bound for the covering number of this set, that is,

$$\mathcal{N}(\mathcal{M},\eta) \le \left(\frac{M}{\eta}\right)^{|\Omega|}, \tag{2.2}$$

see the details in [29]. However, this estimation is not tight enough to derive a convergence rate of the error $\|\boldsymbol{f}_\Lambda - \boldsymbol{p}\|_{\ell_2(\Omega)}$. We need to find a much tighter bound of $\mathcal{N}(\mathcal{M}, \eta)$ by further exploiting the conditions of the set $\mathcal{M}$. As mentioned before, $\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)}$ is a measure of regularity of $\boldsymbol{f}$, it is reasonable to get a much tighter bound by exploiting the condition $\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)} \le \|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)}$. However, things are becoming more complicated as this regularity condition is quite low from the functional point of view and any known results can not help us to achieve desired results. If we view the condition $\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)} \le \|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)}$ discretely and not connect it to its underlying function, it is too complicated to analyze because of the complicated structure of the frame operator $\mathcal{A}$. This motivates us to assume that the tight frame system $\mathcal{A}$ in (1.8) satisfy a mild regularity property — the discrete total variation has to be small. More specifically, we first give the definition of the discrete difference operator $\mathcal{D}$. For any $\boldsymbol{f} \in \ell^\infty(\Omega)$, we define

$$\mathcal{D}\boldsymbol{f} = \left\{\boldsymbol{f}[k_1,\dots,k_{i-1},k_i+1,k_{i+1}\dots,k_d] - \boldsymbol{f}[k_1,\dots,k_d]\right\}_{1\le i\le d,(k_1,\dots,k_d),(k_1,\dots,k_i+1\dots,k_d)\in\Omega}. \tag{2.3}$$

From the constraints $1 \le i \le d, (k_1,\dots,k_d), (k_1,\dots,k_i+1\dots,k_d) \in \Omega$, we know that $\mathcal{D}\boldsymbol{f}$ is a vector with total number of $d(|\Omega| - |\Omega|^{\frac{d-1}{d}})$ entries. The ℓ_1 norm of vector $\mathcal{D}\boldsymbol{f}$ is

$$\|\mathcal{D}\boldsymbol{f}\|_1 = \sum_{i=1}^{d} \sum_{\substack{(k_1,\dots,k_d)\in\Omega,\\(k_1,\dots,k_i+1\dots,k_d)\in\Omega}} |\boldsymbol{f}[k_1,\dots,k_{i-1},k_i+1,k_{i+1}\dots,k_d] - \boldsymbol{f}[k_1,\dots,k_d]|. \tag{2.4}$$

We call $\|\mathcal{D}\boldsymbol{f}\|_1$ a discrete total variation. In particular, when $d = 2$, it becomes

$$\|\mathcal{D}\boldsymbol{f}\|_1 = \sum_{\substack{(k_1,k_2)\in\Omega\\(k_1+1,k_2)\in\Omega}} |\boldsymbol{f}[k_1+1,k_2] - \boldsymbol{f}[k_1,k_2]| + \sum_{\substack{(k_1,k_2)\in\Omega\\(k_1,k_2+1)\in\Omega}} |\boldsymbol{f}[k_1,k_2+1] - \boldsymbol{f}[k_1,k_2]|.$$

For a given frame system $\{\boldsymbol{a}_n\}_{n\in\Gamma}$ of $\ell_2(\Omega)$, we say that it satisfies the bounded condition of the discrete total variation if there exists a positive constant C_d such that

$$\|\mathcal{D}\boldsymbol{a}_n\|_1 \le C_d 2^{\alpha\Upsilon(n)}, \quad n\in\Gamma, \quad \alpha \le d-1, \tag{2.5}$$

where Υ is defined by (1.7). This condition links to the regularity of tight frame systems and most tight frame systems satisfy (2.5) with certain α. This condition is also verifiable in many cases, and straightforward sometimes.

Under the condition (2.5), we can relax the set $\mathcal{M}$ to the set

$$\widetilde{\mathcal{M}} = \{\boldsymbol{f}\in\ell_\infty(\Omega) : \|\mathcal{D}\boldsymbol{f}\|_1 \le C_d|\Omega|^{\frac{\max\{\alpha-\beta,0\}}{d}}\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\gamma)}, \|\boldsymbol{f}\|_\infty \le M\} \tag{2.6}$$

by simple calculation and $\mathcal{M}\subset\widetilde{\mathcal{M}}$. Then we exploit the features of the set $\widetilde{\mathcal{M}}$ and use the famous max-flow min-cut in graph theory to derive the desired estimate of the covering numbers, see Section 4 for more details.

With all these notations, we can give the explicit form of our main result.

Theorem 2.2. *Let $\boldsymbol{f}^\Lambda$ be defined as (1.8), and $\mathcal{A}$ as (1.5). Assume that the frame $\{\boldsymbol{a}_n\}_{n\in\Gamma}$ satisfies (2.5) and $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)} \le C_{\boldsymbol{p}}|\Omega|^{\frac{1}{2}}$ with $\alpha - \frac{d}{2} \le \beta \le \alpha + \frac{d}{2}$. Then for an arbitrary $0<\delta<1$, the following inequality*

$$\frac{1}{|\Omega|}\|\boldsymbol{f}^\Lambda - \boldsymbol{p}\|^2_{\ell_2(\Omega)} \le \widetilde{c}\rho^{-\frac{1}{2}}|\Omega|^{-\min\{\frac{d+2(\beta-\alpha)}{4d},\frac{1}{4}\}}\sqrt{\log_2|\Omega|}\log\frac{1}{\delta} + \frac{16}{3}\sigma^2,$$

where $\widetilde{c} = \frac{256}{3}M^2 + 32M\sqrt{M(2M+2(d+1)C_dC_{\boldsymbol{p}})}$, holds with confidence $1-\delta$.

Note that the condition used in this theorem is quite general and only some low regularity condition for the frame $\mathcal{A}$ and original data $\boldsymbol{p}$ is required. The results is exciting as mentioned in the introduction. For fixed ρ, as long as the cardinality of Ω is large enough, $\boldsymbol{f}_\Lambda$ gives a good approximation of the original data $\boldsymbol{p}$. Furthermore, for fixed Ω, if we let ρ become larger, then we can get smaller error. This result is consistent with our common sense as we are given more data for fixed Ω with larger ρ.

In the following, we prove Theorem 2.2 — the main theorem of this paper. Following the same line as the technique used in statistical learning theory [55], instead of estimating the error $\frac{1}{|\Omega|}\|\boldsymbol{f}^{\Lambda}-\boldsymbol{p}\|^2_{\ell_2(\Omega)}$ directly, we first calculate the probability that the error $\frac{1}{|\Omega|}\|\boldsymbol{f}^{\Lambda}-\boldsymbol{p}\|^2_{\ell_2(\Omega)}$ is smaller than a fixed number by using the theorem of uniform law of large numbers. This leads to the following theorem, which estimates the probability of event $\frac{1}{|\Omega|}\|\boldsymbol{f}^{\Lambda}-\boldsymbol{p}\|^2_{\ell_2(\Omega)} \le \epsilon + \frac{16}{3}\sigma^2$ for an arbitrary given ϵ in terms of the covering numbers with its radius related to ϵ. We leave the proof in Section 4.

THEOREM 2.3. *Let $\mathcal{M}$ be defined by* (2.1) *and $\boldsymbol{f}^{\Lambda}$ by* (1.8). *Then for an arbitrary given $\epsilon > 0$, the inequality*

$$Prob\left\{\frac{1}{|\Omega|}\|\boldsymbol{f}^{\Lambda}-\boldsymbol{p}\|^2_{\ell_2(\Omega)} \le \epsilon + \frac{16}{3}\sigma^2\right\} \ge 1 - \mathcal{N}(\mathcal{M}, \frac{\epsilon}{12M}) \exp\left\{-\frac{3m\epsilon}{256M^2}\right\}$$

holds for an arbitrary m, where m is the number of samples.

Proof. See Section 4.1. ◻

In order to give the explicit convergence rate of $\|\boldsymbol{f}_{\Lambda}-\boldsymbol{p}\|_{\ell_2(\Omega)}$, we need an explicit estimate of the covering number $\mathcal{N}(\mathcal{M},\eta)$. The following theorem concerns an upper bound of $\mathcal{N}(\mathcal{M},\eta)$. As its proof is too complicated, we leave it in Section 4 for the reader more easy to understand the idea of this paper. The main difficulty we overcome is the low regularity of the sequence in the set $\mathcal{M}$ as β is not large enough here. We overcome it by using the powerful tool of discrete total variation and max-flow min-cut theorem. It should be noted that only discrete total variation is used to measure the regularity of the sequence in $\mathcal{M}$ for covering number estimation, so our analysis is still true for more general case such as similar TV based algorithms, see Section 4 for more details.

THEOREM 2.4. *Let $\mathcal{M}$ be defined as* (2.1) *and $\mathcal{A}$ as* (1.5). *Assume that the frame $\{\boldsymbol{a}_n\}_{n\in\Gamma}$ satisfies* (2.5) *and $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)} \le C_{\boldsymbol{p}}|\Omega|^{\frac{1}{2}}$ with $\beta \le \alpha + \frac{d}{2}$, then for any $\eta \ge |\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac{1}{2}\}-1}$,*

$$\log \mathcal{N}(\mathcal{M},\eta) \le \frac{C'_d|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac{1}{2}\}}\log_2|\Omega|}{\eta},$$

where $C'_d = 2M + 2(d+1)C_dC_{\boldsymbol{p}}$.

Proof. See Section 4.2. ◻

With all of these, we are now ready to prove Theorem 2.2. The technique used for the proof is somewhat similar to the one used in statistical learning theory [29]. The main difference is that we have some constraint for η in our bound for covering number given in Theorem 2.4, so we need to verify that this constraint will not influence the proof of Theorem 2.2.

Proof of Theorem 2.2. First, by Theorem 2.3, for an arbitrary given $\epsilon \ge 12M|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac{1}{2}\}-1}$, and $\eta = \frac{\epsilon}{12M}$ the inequality

$$\frac{1}{|\Omega|}\|\boldsymbol{f}^{\Lambda}-\boldsymbol{p}\|^2_{\ell_2(\Omega)} \le \epsilon + \frac{16}{3}\sigma^2$$

holds with the confidence at least

$$\begin{aligned}&1 - \mathcal{N}(\mathcal{M}, \frac{\epsilon}{12M}) \exp\left\{-\frac{3m\epsilon}{256M^2}\right\}\\ \ge\, &1 - \exp\left\{\frac{12MC'_d|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac{1}{2}\}}\log_2|\Omega|}{\epsilon}\right\}\exp\left\{-\frac{3m\epsilon}{256M^2}\right\}.\end{aligned}$$

The last inequality follows from Theorem 2.4. Next, choosing a special ϵ^* to be the unique positive solution of the following equation

$$\frac{12MC'_d|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac{1}{2}\}}\log_2|\Omega|}{\epsilon} - \frac{3m\epsilon}{256M^2} = \log\delta, \tag{2.7}$$

we have $\frac{1}{|\Omega|}\|\boldsymbol{f}^\Lambda-\boldsymbol{p}\|^2_{\ell_2(\Omega)} \le \epsilon^*+\frac{16}{3}\sigma^2$ with confidence $1-\delta$ if we can prove that $\epsilon^* \ge 12M|\Omega|^{\max\{\frac{d+2(\beta-\alpha)}{2d},\frac12\}-1}$. However, solving the equation (2.7) yields

$$\begin{aligned}\epsilon^* &= \frac{32M}{3m}\left(4M\log\frac{1}{\delta}+\sqrt{16M^2\log^2\frac{1}{\delta}+9mMC'_d|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac12\}}\log_2|\Omega|}\right) \\ &\le \frac{32M}{3m}\left(8M\log\frac{1}{\delta}+3\sqrt{mMC'_d|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac12\}}\log_2|\Omega|}\right) \\ &\le \widetilde{c}\rho^{-\frac12}\sqrt{\log_2|\Omega|}|\Omega|^{-\min\{\frac{d+2(\beta-\alpha)}{4d},\frac14\}}\log\frac{1}{\delta},\end{aligned} \tag{2.8}$$

where $\widetilde{c}=\frac{256}{3}M^2+32M\sqrt{MC'_d}$. Also, from (2.8), we know that

$$\epsilon^* \ge \frac{32M}{\sqrt{m}}\sqrt{MC'_d|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac12\}}\log_2|\Omega|} \ge 16M\sqrt{MC'_d}\rho^{-\frac12}\sqrt{\log_2|\Omega|}|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{4d},\frac14\}-\frac12},$$

which implies that $\epsilon^* \ge 12M|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac12\}-1}$. This concludes the proof. ∎

3. Image Recovery from Random Samples by Framelet. Before going to the proofs of the technical theorems in the previous section, we apply Theorem 2.2 to framelet based image recovery from random samples in this section. Various algorithms of framelet based image recovery algorithms have been developed in [9, 10, 13, 38, 39]. Especially, an efficient algorithm for framelet based image recovery by using splitting Bregman iteration is given in [13]. For this framelet based image recovery algorithm, we are able to link the approximation property of the algorithm to the regularity of the underlying function (in terms of the decay of its canonical coefficients of given tight frames) where the pixels come from. We start with discussions of the approximation of the framelet based recovery. It is then followed by the link of this analysis to the functional space. We restrict our discussions here for two variable functions, since the images can be viewed as a set of data sampled from two variable functions. For more general multi-variable functions, the discussions are the same.

3.1. Framelet. A wavelet (or affine) system $X(\Psi,\phi)$ derived from the multiresolution analysis generated by a refinable function ϕ is defined to be the collection of dilations and shifts of a finite set $\Psi=\{\psi^\ell : \ell=1,2,\dots,L\}\subset L_2(\mathbb{R}^2)$, i. e.,

$$X(\Psi,\phi)=\{\psi^\ell_{j,k}:=2^j\psi^\ell(2^jx-k) : j\in\mathbb{Z}, k\in\mathbb{Z}^2, \ell=1,2,\dots,L\}.$$

The elements in Ψ are called the *generators*. When $X(\Psi,\phi)$ is also a tight frame for $L_2(\mathbb{R}^2)$, then $\psi\in\Psi$ are called *(tight) framelets*, following the terminology used in [31]. Recall that $X(\Psi,\phi)$ is a tight frame for $L_2(\mathbb{R}^2)$ if, for any $f\in L_2(\mathbb{R}^2)$,

$$f=\sum_{\ell=1}^{L}\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}^2}\langle f,\psi^\ell_{j,k}\rangle\psi^\ell_{j,k}.$$

To construct compactly supported framelet systems, one starts with a compactly supported refinable function $\phi\in L_2(\mathbb{R}^2)$ with a refinement mask (low-pass filter) $\boldsymbol{h}_0$ such that ϕ satisfies the refinement equation:

$$\phi(x)=4\sum_k \boldsymbol{h}_0[k]\phi(2x-k). \tag{3.1}$$

Let V_0 be the closed shift invariant space generated by $\{\phi(\cdot-k) : k\in\mathbb{Z}^2\}$ and $V_j:=\{f(2^j\cdot) : f\in V_0, j\in\mathbb{Z}\}$. It is known that when ϕ is compactly supported, the sequence $\{V_j\}_{j\in\mathbb{Z}}$ forms a multiresolution analysis. Recall that $\{V_j\}_{j\in\mathbb{Z}}$ is said to generate a multiresolution analysis (MRA) if (a) $V_j\subset V_{j+1}$, (b) $\cup_j V_j$ is dense in $L_2(\mathbb{R}^2)$, (c) $\cap_j V_j=\{0\}$, see [44, 51] for more details.

In this paper, we assume that the refinable function ϕ satisfy the following conditions:

Assumption 1.

(a) $\phi : \mathbb{R}^2 \mapsto \mathbb{R}^2$ *is compactly supported with* $\int \phi = 1$.

(b) ϕ *is Hölder continuous with exponent 1, i.e. there exists a constant* $\widetilde{C}$ *such that for any* $x, y \in \mathbb{R}^2, |\phi(x) - \phi(y)| \leq \widetilde{C}\|x - y\|$, *where* $\| \cdot \|$ *is the Euclidean norm in* $\mathbb{R}^2$.

(c) $\{\phi(\cdot - k)\}_{k\in\mathbb{Z}^2}$ *is a Riesz Basis in the space* V_0.

There are many refinable functions satisfy the above assumptions, e.g. the tensor product pseudo splines (see e.g. [31, 34], or simply three directional box splines see e.g. [5]). The Riesz basis requirement is not so crucial. For example, it is not required in applying the unitary extension principle for the construction of tight framelets.

The compactly supported framelets $\boldsymbol{\Psi}$ are defined by

$$\psi^\ell(x) = 4 \sum_k \boldsymbol{h}_\ell[k]\phi(2x - k)$$

for some compactly supported sequence $\boldsymbol{h}_\ell$ in $\ell_\infty(\mathbb{Z}^2)$. When the filters $\{\boldsymbol{h}_i, i = 0, \ldots, L\}$ satisfy the following conditions

$$\sum_{\ell=0}^{L} |\widehat{\boldsymbol{h}_\ell}(\omega)|^2 = 1 \quad \text{and} \quad \sum_{\ell=0}^{L} \widehat{\boldsymbol{h}_\ell}(\omega)\overline{\widehat{\boldsymbol{h}_\ell}(\omega + \pi)} = 0, \quad \text{a.e.} \quad \omega \in [-\pi, \pi], \tag{3.2}$$

where $\widehat{\boldsymbol{h}_\ell}(\omega) := \sum_{k\in\mathbb{Z}^2} \boldsymbol{h}_\ell[k]e^{-ik\omega}$, then the wavelet system $X(\Psi, \phi)$ is a tight wavelet frame by the unitary extension principle (UEP) in [51]. The corresponding mask $\boldsymbol{h}_0$ is refinement mask which is a low pass filter and $\{\boldsymbol{h}_\ell : 1 \leq \ell \leq L\}$ are framelet masks which are high pass filters. Since the publication of UEP [51] and oblique extension principle (OEP) of [28] and [31], there are many construction of framelets using UEP and OEP, see [27, 47, 50] and references therein.

The advantage of framelet is that the discrete tight frame system for the computation is easy to derive by framelet decomposition and reconstruction algorithms of [31]. First, we construct

$$\widetilde{\boldsymbol{a}}_0 = 2^J \underbrace{\boldsymbol{h}_0 * \uparrow \ldots \boldsymbol{h}_0 * \uparrow}_{J} \boldsymbol{\delta} \tag{3.3}$$

and

$$\widetilde{\boldsymbol{b}}_j^\ell = 2^{J-j} \underbrace{\boldsymbol{h}_0 * \ldots \uparrow \boldsymbol{h}_0 * \uparrow}_{J-j-1} \boldsymbol{h}_\ell * \uparrow \boldsymbol{\delta}, \tag{3.4}$$

where $\boldsymbol{\delta}$ is a sequence with each component $\boldsymbol{\delta}[k] = 1$ when $k = (0, 0)$ and $\boldsymbol{\delta}[k] = 0$ otherwise, and $\boldsymbol{h}_0 * \uparrow$ is an upsampling operator, i.e., for a sequence $\boldsymbol{c} \in \ell_2(\mathbb{Z}^2)$,

$$\boldsymbol{h}_0 * \uparrow \mathbf{c} = \sum_k \boldsymbol{h}_0[n - 2k]\boldsymbol{c}[k].$$

Using these sequences, one can derive the standard framelet decomposition algorithm as suggested in [31].

Let $\boldsymbol{f} \in \ell_2(\Omega)$ be an image with

$$\Omega = \{k = (k_1, k_2) : 0 \leq k_1, k_2 < 2^J\}. \tag{3.5}$$

To make a suitable tight frame analysis, one needs to impose proper boundary conditions. Periodic boundary conditions are imposed here. We still use $\ell_2(\Omega)$ to denote the space of sequences defined on $\ell_2(\Omega)$ with periodic boundary conditions. Other boundary conditions can be discussed similarly, we forgo the discussion here and the interested reader should consult [8, 23] for more details. Let $\mathcal{P}$ be an operator that maps a vector in $\ell_2(\mathbb{Z}^2)$ into $\ell_2(\Omega)$

$$\mathcal{P}(\boldsymbol{v})[k] = \sum_{k_1', k_2' \in \mathbb{Z}} \boldsymbol{v}[k_1 + k_1' 2^J, k_2 + k_2' 2^J)], \quad \forall k = (k_1, k_2) \in \Omega.$$

Let

$$\boldsymbol{a}_0 = \mathcal{P}(\widetilde{\boldsymbol{a}}_0) \tag{3.6}$$

and

$$\boldsymbol{b}_j^{\ell,k} = \mathcal{P}(\widetilde{\boldsymbol{b}}_j^{\ell}[\cdot - k]), \tag{3.7}$$

where $\widetilde{\boldsymbol{a}}_0$ is defined by (3.3) and $\widetilde{\boldsymbol{b}}_j^{\ell}$ by (3.4). Then, the sequence $\{\boldsymbol{a}_0\} \cup \{\boldsymbol{b}_j^{k,\ell}\}_{0\le k_1,k_2<2^j,0\le j<J,1\le\ell\le L}$ is a tight frame system for the space $\ell_2(\Omega)$ with periodic boundary condition by the tight framelet theory (see e.g. [20]). With this tight frame system, the analysis operator $\mathcal{A}$ is defined as

$$\forall \boldsymbol{f}, \qquad \mathcal{A}\boldsymbol{f} = \Big\{\{\langle \boldsymbol{a}_0, \boldsymbol{f}\rangle\}, \quad \{\langle \boldsymbol{b}_j^{\ell,k}, \boldsymbol{f}\rangle\}_{0\le k_1,k_2<2^j,0\le j<J,1\le\ell\le L}\Big\}. \tag{3.8}$$

Denote the adjoint of $\mathcal{A}$ by $\mathcal{A}^*$. By the fact that filters $\{\mathbf{h}_i\}_{i=0}^{L}$ form a tight frame system, we have

$$\boldsymbol{f} = \mathcal{A}^*\mathcal{A}\boldsymbol{f} = \langle \boldsymbol{a}_0, \boldsymbol{f}\rangle \boldsymbol{a}_0 + \sum_{j=0}^{J-1}\sum_{k_1,k_2=0}^{2^j-1}\sum_{\ell=1}^{L}\langle \boldsymbol{b}_j^{\ell,k}, \boldsymbol{f}\rangle \boldsymbol{b}_j^{\ell,k}. \tag{3.9}$$

The operator $\mathcal{A}^*$ is also called synthesis operator. Once we have the analysis operator $\mathcal{A}$, we define the weighted norm $\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta)}$ for a given β by

$$\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta)} = |\langle \boldsymbol{a}_0, \boldsymbol{f}\rangle| + \sum_{0\le j<J} 2^{j\beta}\sum_{k,\ell}|\langle \boldsymbol{b}_j^{\ell,k}, \boldsymbol{f}\rangle|. \tag{3.10}$$

Note that the tight frame system $\{\boldsymbol{a}_0\} \cup \{\boldsymbol{b}_j^{k,\ell}\}_{0\le k_1,k_2<2^j,0\le j<J,1\le\ell\le L}$ is indexed by (j,k,ℓ), where $k = (k_1,k_2) \in \mathbb{Z}^2$. The same weight is used for the same subscript j in the above definition of $\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta)}$. More explicitly, using the notation in (1.6), we have chosen $\Gamma = \{0\} \cup \{(j,k,\ell) : 0 \le k_1,k_2 < 2^j, 0 \le j < J, 1 \le \ell \le L\}$ and the sequence Υ is defined as

$$\Upsilon(0) = 0, \text{ and } \Upsilon(j,k,\ell) = j.$$

Since $j < J := \frac{1}{2}\log_2|2^{2J}|$, the condition (1.7) naturally holds under this definition of Υ.

This weighted norm (3.10) links to regularity of the underlying function where the pixel $\boldsymbol{f}$ derived from, see [6, 40, 49] and Section 3.3 for more discussions.

3.2. Approximation by Framelet. Let $\boldsymbol{p}$ be a given sequence defined on $\ell_2(\Omega)$ satisfying $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta)} < \infty$ and $\|\boldsymbol{p}\|_\infty \le M$ for some preassigned constants β and M. Then, the approximation solution $\boldsymbol{f}^\Lambda$ defined by (1.8) becomes

$$\boldsymbol{f}^\Lambda = \arg\min\left\{\|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta)} : \frac{1}{|\Lambda|}\sum_{k\in\Lambda}(\boldsymbol{f}[k] - \boldsymbol{g}[k])^2 \le \sigma^2, \|\boldsymbol{f}\|_\infty \le M\right\}. \tag{3.11}$$

This section gives an error analysis for $\frac{1}{2^{2J}}\|\boldsymbol{f}^\Lambda - \boldsymbol{p}\|^2_{\ell_2(\Omega)}$ for the framelet based image recovery. To apply Theorem 2.2, we only need to verify (2.5). In fact, we have the following lemma which states that the condition (2.5) is satisfied with $\alpha = 0$ for $\mathcal{A}$ derived from $X(\Psi,\phi)$ satisfying Assumption 1.

LEMMA 3.1. *Assume that refinable function ϕ satisfies Assumption 1. Let $\mathcal{A}$ be defined as* (3.8) *by the compactly supported tight framelet system $X(\Psi,\phi)$ with compactly supported high-low filters derived by the unitary extension principle from the refinable function ϕ. Then*

$$\max\{\|\mathcal{D}\boldsymbol{a}_0\|_1, \sup_{j,\ell}\|\mathcal{D}\boldsymbol{b}_j^{\ell,k}\|_1\} \le C_d \tag{3.12}$$

for some constant $C_d \ge 1$, which is independent of J. Furthermore, for each $\boldsymbol{f} \in \mathcal{M}$, we have

$$\|\mathcal{D}\boldsymbol{f}\|_1 \le C_d\|\mathcal{A}\boldsymbol{f}\|_1. \tag{3.13}$$

Proof. See Section 4.3. □

Using Theorem 2.3 and Lemma 3.1, we can easily derive the following corollary.

COROLLARY 3.2. *Let $\mathcal{A}$ be defined as* (3.8) *by the compactly supported tight framelet system* $X(\Psi,\phi)$ *with compactly supported high-low filters derived by the unitary extension principle from refinable function* ϕ *that satisfies Assumption 1. Let* $\boldsymbol{f}^\Lambda$ *be defined in* (3.11). *Assume that* $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta)} \le C_{\boldsymbol{p}} 2^J$ *and* $-1<\beta<1$. *Then for any* $0<\delta<1$, *with confidence* $1-\delta$,

$$\frac{1}{2^{2J}}\|\boldsymbol{f}^\Lambda - \boldsymbol{p}\|^2_{\ell_2(\Omega)} \le \widetilde{c}\rho^{-\frac{1}{2}}\sqrt{J}2^{-J\min\{\frac{1+\beta}{2},\frac{1}{2}\}}\log\frac{1}{\delta} + \frac{16}{3}\sigma^2,$$

where $\widetilde{c} = \frac{256}{3}M^2 + 32M\sqrt{M(2M+6C_dC_{\boldsymbol{p}})}$.

Proof. By applying lemma 3.1, we know that the condition (2.5) satisfies with $\alpha=0$. Then the corollary can be deduced directly from Theorem 2.2 by letting $|\Omega| = 2^{2J}$ and $\alpha=0$. □

3.3. Connection to Function Approximation. This section is to link the estimate given before to the function approximation if we assume that the data is obtained by sampling a function which converts analog signal to digital signal. For example, for the image, the pixels are well modeled by local weighted averages of some underlying function p that closely fits the physics of CCD cameras. Furthermore, the pixel values of an image can be viewed as the inner product of p and some refinable function without much loss [21]. More specifically, let ϕ be a refinable function satisfying $\int\phi = 1$, and denote the scaled functions by $\phi_{J,k} := 2^J\phi(2^J\cdot -k)$ for $k\in\Omega$. Then, each pixel value is obtained by

$$\boldsymbol{p}[k] = 2^J\langle p, \phi_{J,k}\rangle, \quad k\in\Omega. \tag{3.14}$$

With $\boldsymbol{p}[k]$, implicitly, we use function

$$p_J = \sum_{k\in\Omega}\boldsymbol{p}[k]\phi(2^J\cdot -k) = \sum_{k\in\Omega}\langle p,\phi_{J,k}\rangle\phi_{J,k}$$

to approximate p. The approximation order of p_J to p has been studied extensively. Roughly, as long as ϕ meets the Strang and Fix condition of a certain order and the Fourier transform of ϕ is flat enough at the origin, then p_J will have a good approximation to p. For example, assume that ϕ satisfies the Strang and Fix condition with certain order, and $1-|\hat{\phi}(0)|^2$ has the same order of zeros, where $\hat{\phi}$ is the Fourier transform of ϕ, and p is sufficiently smooth, then p_J provides this order of approximation to p. Interested reader should consult [31] for details. However, this requires the underlying function has a high order of smoothness. In this case, minimizing the ℓ_2 norm of the canonical coefficients of the framelet system will work and the error analysis can be done similarly as that of [45]. In this paper, the underlying function we are interested in does not meet certain order of smoothness. Instead, we require here some decay condition of the wavelet system $X(\Psi,\phi)$ to analyze the approximation order of p_J to p. The decay condition here is so mild that the implicit assumption of the regularity of the underlying function is very weak.

Let $\boldsymbol{f}^\Lambda$ be the solution of (3.11). We take the function

$$f_J^\Lambda := \sum_{k\in\Omega}\boldsymbol{f}^\Lambda[k]\phi(2^J\cdot -k) \tag{3.15}$$

to approximate the underlying function p and find the error of $\|p-f_J^\Lambda\|_{L_2(I)}$. Note that

$$\|p-f_J^\Lambda\|_{L_2(I)} \le \|p_J - f_J^\Lambda\|_{L_2(I)} + \|p-p_J\|_{L_2(I)} \le \frac{C_\phi}{2^J}\|\boldsymbol{f}^\Lambda - \boldsymbol{p}\|_{\ell_2(\Omega)} + \|p-p_J\|_{L_2(I)}. \tag{3.16}$$

The second inequality follows from the fact that $\{\phi(\cdot - k)\}_{k\in\mathbb{R}^2}$ is a Bessel system in $L_2(\mathbb{R}^2)$, i.e.

$$\|p_J - f_J^\Lambda\|_{L_2(I)} \le \frac{C_\phi}{2^J}\|\boldsymbol{f}^\Lambda - \boldsymbol{p}\|_{\ell_2(\Omega)},$$

where C_ϕ is a constant independent of J.

Hence, to estimate $\|p_J - f_J^\Lambda\|_{L_2(I)}$, we need to apply Corollary 3.2 to derive the estimate of $\frac{1}{2^{2J}}\|\boldsymbol{f}^\Lambda - \boldsymbol{p}\|^2_{\ell_2(\Omega)}$. For this, we need a condition on p, so that $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta)} \le C_{\boldsymbol{p}} 2^J$ will be satisfied. Recall that $X(\Psi,\phi)$ is a tight framelet system and the intensity function p can be represented as

$$p = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}^2} \sum_{\ell=1}^{L} \langle p, \psi^\ell_{j,k}\rangle \psi^\ell_{j,k}. \tag{3.17}$$

The decay condition we assume here is that there is a $\beta \ge -1$ such that

$$C_p := \sum_k |\langle p, \phi_{0,k}\rangle| + \sum_{j\ge 0} 2^{\beta j} \sum_{k,\ell} |\langle p, \psi^\ell_{j,k}\rangle| < \infty. \tag{3.18}$$

This decay condition links to the regularity of the underlying function p when the framelet satisfy some mild conditions. We refrain to further discussion in this direction and interested reader should consult [6, 40] for the details. Under this mild decay condition of the canonical framelet coefficients, the approximation of underlying function can be stated below:

Corollary 3.3. *Let $\mathcal{A}$ be defined as* (3.8) *by the compactly supported tight framelet system $X(\Psi,\phi)$ with compactly supported high-low filters derived by the unitary extension principle from refinable function ϕ that satisfies Assumption 1. Assume that the underlying function p satisfies* (3.18) *with $-1<\beta<1$ and the samples $\{\boldsymbol{p}[k]\}_{k\in\Omega}$ are obtained by* (3.14). *Then, for an arbitrary $0<\delta<1$, the inequality*

$$\frac{1}{2^{2J}}\|\boldsymbol{f}^\Lambda - \boldsymbol{p}\|^2_{\ell_2(\Omega)} \le \widetilde{c}\rho^{-\frac{1}{2}}\sqrt{J}2^{-J\min\{\frac{1+\beta}{2},\frac{1}{2}\}}\log\frac{1}{\delta} + \frac{16}{3}\sigma^2 \tag{3.19}$$

holds with confidence $1-\delta$, where $\tilde{c}$ is a constant independent of J (i.e. independent of cardinality of Ω), ρ, δ or σ. Furthermore, let f_J^Λ be defined by (3.15). *If $\|p\|_\infty < \infty$, then*

$$\|p - f_J^\Lambda\|^2_{L_2(I)} \le C_1\rho^{-\frac{1}{2}}\sqrt{J}2^{-J\min\{\frac{1+\beta}{2},\frac{1}{2}\}}\log\frac{1}{\delta} + C_2\sigma^2 + C_3 2^{-(\beta+1)J} \tag{3.20}$$

with confidence $1-\delta$, where C_1, C_2, C_3 are three constants independent of J, ρ, δ or σ.

Proof. Inequality (3.19) can be derived directly from Corollary 3.2, as long as $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta)} \le C_{\boldsymbol{p}}2^J$ is verified under the assumption (3.18). In fact, the tight frames $\{\boldsymbol{a}_0\} \cup \{\boldsymbol{b}_j^{k,\ell}\}_{0\le k_1,k_2<2^j, 0\le j<J, 1\le \ell\le L}$ for the space $\ell_2(\Omega)$ are designed according to the standard framelet decomposition algorithm given in [31] with periodic boundary conditions. This observation leads to the fact that for any $p \in L_2(\mathbb{R}^2)$, $\langle \boldsymbol{b}_j^{k,\ell}, \boldsymbol{p}\rangle = 2^J\langle \psi^\ell_{j,k}, p\rangle$, $\langle \boldsymbol{a}_0, \boldsymbol{p}\rangle = 2^J\langle \phi_{0,k}, p\rangle$, where $\boldsymbol{p}$ is defined as (3.14). Hence, $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta)} \le C_{\boldsymbol{p}}2^J$ follows from (3.18) by setting $C_{\boldsymbol{p}} = C_p$. This leads to (3.19). Furthermore, as C_p and C_d are independent of J (i.e. the cardinality of Ω), $\widetilde{c} = \frac{256}{3}M^2 + 32M\sqrt{M(2M+6C_dC_p)}$ is independent of J, ρ, δ or σ.

For (3.20), we need to estimate $\|p - p_J\|_{L_2(I)}$ by (3.16). A standard tight framelet decomposition leads to (see e.g. [31])

$$p_J = \sum_k \langle p, \phi_{0,k}\rangle \phi_{0,k} + \sum_{0\le j<J} \sum_{k,\ell} \langle p, \psi^\ell_{j,k}\rangle \psi^\ell_{j,k}.$$

This, together with (3.17) and the Bessel property of the tight frame system $X(\Psi,\phi)$, we have

$$\|p - p_J\|_2^2 \le \sum_{j\ge J}\sum_{k,\ell} |\langle p, \psi^\ell_{j,k}\rangle|^2.$$

Note that

$$|\langle p, \psi^\ell_{j,k}\rangle| \le \|p\|_\infty \|\psi^\ell_{j,k}\|_{L_1(I)} = \|p\|_\infty 2^{-j}\|\psi^\ell\|_{L_1(I)}.$$

This further leads to

$$\begin{aligned}
\|p - p_J\|_2^2 &\leq \|p\|_\infty \max_\ell \|\psi^\ell\|_{L_1(I)} \sum_{j\geq J} \sum_{k,\ell} 2^{-j} |\langle p, \psi_{j,k}^\ell \rangle| \\
&\leq \|p\|_\infty \max_\ell \|\psi^\ell\|_{L_1(I)} \sum_{j\geq J} \sum_{k,\ell} 2^{-j} \frac{2^{(\beta+1)j}}{2^{(\beta+1)J}} |\langle p, \psi_{j,k}^\ell \rangle| \\
&\leq C_p \|p\|_\infty \max_\ell \|\psi^\ell\|_{L_1(I)} 2^{-(\beta+1)J}. \qquad (3.21)
\end{aligned}$$

Thus, inequality (3.20) follows by setting $C_1 = 2(C_\phi)^2\widetilde{c}$, $C_2 = 16(C_\phi)^2/3$ and $C_3 = 2C_p\|p\|_\infty \max_\ell \|\psi^\ell\|_{L_1(I)}$. ∎

This corollary says that as long as the data set is sufficiently dense, one has a pretty good chance to derive a good approximation of the underlying true solution $\boldsymbol{p}$ by solving (1.8). Furthermore, the approximation of the function constructed from the recovered data gives a good approximation of the underlying function where the original data comes from with high probability.

4. Proof of Critical Lemmas and Theorems. This section is devoted to the technical details we left in the pervious sections. In particular, this section gives the proofs of Theorem 2.3, Theorem 2.4 and Lemma 3.1.

4.1. Proof of Theorem 2.3. It requires several lemmas and propositions to prove Theorem 2.3. The idea follows the same line as in statistical learning theory [29]. However, the setting is somewhat different and we still give the proof for the completeness. We start with the following ratio probability inequality concerning only one random variable. It can be deduced from Bernstein inequality directly, see [29, 55] for more details.

LEMMA 4.1. *Suppose a random variable ξ on Z satisfies $E\xi = \mu \geq 0$, and $|\xi - \mu| \leq B$ almost everywhere. Assume that $E\xi^2 \leq cE\xi$. Then for any $\epsilon > 0$ and $0 < \gamma \leq 1$, we have*

$$Prob_{z\in Z^m} \left\{ \frac{\mu - \frac{1}{m}\sum_{i=1}^m \xi(z_i)}{\sqrt{\mu+\epsilon}} > \gamma\sqrt{\epsilon} \right\} \leq \exp\left\{ -\frac{\gamma^2 m\epsilon}{2c + \frac{2}{3}B} \right\}.$$

Next, let $\xi = (\boldsymbol{f}[\zeta] - \boldsymbol{p}[\zeta])^2$ where $\boldsymbol{f} \in \mathcal{M}$ and ζ is a random variable i. i. d drawn from the uniform distribution on Ω. Then ξ is a random variable satisfying $0 \leq \xi \leq M^2$. Define

$$\mathcal{E}(\boldsymbol{f}) := E\xi = E(\boldsymbol{f}[\zeta] - \boldsymbol{p}[\zeta])^2 = \frac{1}{|\Omega|}\sum_{k\in\Omega} (\boldsymbol{f}[k] - \boldsymbol{p}[k])^2 \qquad (4.1)$$

and

$$\mathcal{E}_\Lambda(\boldsymbol{f}) = \frac{1}{|\Lambda|}\sum_{k\in\Lambda} (\boldsymbol{f}[k] - \boldsymbol{p}[k])^2. \qquad (4.2)$$

We have the following lemma:

LEMMA 4.2. *Let $\mathcal{M}$ be defined by (2.1) and $\boldsymbol{f}_1, \boldsymbol{f}_2 \in \mathcal{M}$. Then,*

$$|\mathcal{E}(\boldsymbol{f}_1) - \mathcal{E}(\boldsymbol{f}_2)| \leq 2M\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_\infty, \qquad |\mathcal{E}_\Lambda(\boldsymbol{f}_1) - \mathcal{E}_\Lambda(\boldsymbol{f}_2)| \leq 2M\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_\infty.$$

Proof. Note that

$$\sum_{k\in\Omega} (\boldsymbol{f}_1[k] - \boldsymbol{p}[k])^2 - \sum_{k\in\Omega} (\boldsymbol{f}_2[k] - \boldsymbol{p}[k])^2 \leq \sum_{k\in\Omega} |\boldsymbol{f}_1[k] - \boldsymbol{f}_2[k]||\boldsymbol{f}_1[k] - \boldsymbol{p}[k] + \boldsymbol{f}_2[k] - \boldsymbol{p}[k]|.$$

Hence,

$$\begin{aligned}
|\mathcal{E}(\boldsymbol{f}_1) - \mathcal{E}(\boldsymbol{f}_2)| &= \left| \frac{1}{|\Omega|}\sum_{k\in\Omega} (\boldsymbol{f}_1[k] - \boldsymbol{p}[k])^2 - \frac{1}{|\Omega|}\sum_{k\in\Omega} (\boldsymbol{f}_2[k] - \boldsymbol{p}[k])^2 \right| \\
&\leq \frac{2M}{|\Omega|}\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_1 \leq 2M\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_\infty.
\end{aligned}$$

The second inequality can be proved similarly by replacing Ω with Λ. □

Now we give a ratio probability inequality involving the space $\mathcal{M}$. For this, we recall that $\mathcal{N}(\mathcal{M},\eta)$ is the covering number of $\mathcal{M}$ with respect to the metric $\ell_\infty(\Omega)$.

PROPOSITION 4.3. *Let $\mathcal{M}$, $\mathcal{E}(\boldsymbol{f})$ and $\mathcal{E}_\Lambda(\boldsymbol{f})$ be defined by* (2.1), (4.1) *and* (4.2) *respectively. Then for every $\epsilon>0$ and $0<\gamma<1$, we have*

$$Prob\left\{\sup_{\boldsymbol{f}\in\mathcal{M}}\frac{\mathcal{E}(\boldsymbol{f})-\mathcal{E}_\Lambda(\boldsymbol{f})}{\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}>4\gamma\sqrt{\epsilon}\right\}\le\mathcal{N}(\mathcal{M},\frac{\gamma\epsilon}{2M})\exp\left\{-\frac{3\gamma^2m\epsilon}{8M^2}\right\}.$$

Proof. Let $\{\boldsymbol{f}_j\}_{j=1}^K$, where $K=\mathcal{N}(\mathcal{M},\frac{\gamma\epsilon}{2M})$, be a sequence such that $\mathcal{M}$ is covered by ℓ_∞ balls in $\mathcal{M}$ centered at $\boldsymbol{f}_j$ with radius $\frac{\gamma\epsilon}{2M}$. For each j, consider the random variable $\xi=(\boldsymbol{f}_j[\zeta]-\boldsymbol{p}[\zeta])^2$, where ζ is an i.i.d random variable drawn from the uniform distribution on Ω. Note that $\boldsymbol{f}_j,\boldsymbol{p}\in\mathcal{M}$ implies $\|\boldsymbol{f}_j\|_\infty\le M$ and $\|\boldsymbol{p}\|_\infty\le M$. Thus we have $|\xi-E\xi|\le M^2$. Furthermore,

$$E\xi^2=E(\boldsymbol{f}_j[\zeta]-\boldsymbol{p}[\zeta])^4\le M^2E(\boldsymbol{f}_j[\zeta]-\boldsymbol{p}[\zeta])^2=M^2E\xi.$$

Applying Lemma 4.1 to ξ with $B=c=M^2$, we have

$$\text{Prob}\left\{\frac{\mathcal{E}(\boldsymbol{f}_j)-\mathcal{E}_\Lambda(\boldsymbol{f}_j)}{\sqrt{\mathcal{E}(\boldsymbol{f}_j)+\epsilon}}>\gamma\sqrt{\epsilon}\right\}\le\exp\left\{-\frac{3\gamma^2m\epsilon}{8M^2}\right\}. \tag{4.3}$$

For an arbitrary $\boldsymbol{f}\in\mathcal{M}$, there is some $j\in\{1,\dots,K\}$ such that $\|\boldsymbol{f}-\boldsymbol{f}_j\|_\infty\le\frac{\gamma\epsilon}{2M}$. This, together with Lemma 4.2, yields

$$|\mathcal{E}_\Lambda(\boldsymbol{f})-\mathcal{E}_\Lambda(\boldsymbol{f}_j)|\le2M\|\boldsymbol{f}-\boldsymbol{f}_j\|_\infty\le\gamma\epsilon,\quad|\mathcal{E}(\boldsymbol{f})-\mathcal{E}(\boldsymbol{f}_j)|\le2M\|\boldsymbol{f}-\boldsymbol{f}_j\|_\infty\le\gamma\epsilon.$$

Therefore,

$$\frac{|\mathcal{E}_\Lambda(\boldsymbol{f})-\mathcal{E}_\Lambda(\boldsymbol{f}_j)|}{\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}\le\gamma\sqrt{\epsilon}\quad\text{and}\quad\frac{|\mathcal{E}(\boldsymbol{f})-\mathcal{E}(\boldsymbol{f}_j)|}{\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}\le\gamma\sqrt{\epsilon}.$$

The latter implies that

$$\begin{aligned}\mathcal{E}(\boldsymbol{f}_j)+\epsilon&=\mathcal{E}(\boldsymbol{f}_j)-\mathcal{E}(\boldsymbol{f})+\mathcal{E}(\boldsymbol{f})+\epsilon\le\gamma\sqrt{\epsilon}\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}+\mathcal{E}(\boldsymbol{f})+\epsilon\\&\le\sqrt{\epsilon}\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}+\mathcal{E}(\boldsymbol{f})+\epsilon\le2(\mathcal{E}(\boldsymbol{f})+\epsilon).\end{aligned}$$

This leads to $\sqrt{\mathcal{E}(\boldsymbol{f}_j)+\epsilon}\le2\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}$ for any $\boldsymbol{f}\in\{\boldsymbol{f}:\|\boldsymbol{f}-\boldsymbol{f}_j\|_\infty\le\frac{r\epsilon}{2M}\}$.

Next, assume that $\frac{\mathcal{E}(\boldsymbol{f})-\mathcal{E}_\Lambda(\boldsymbol{f})}{\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}>4\gamma\sqrt{\epsilon}$, then

$$\begin{aligned}\frac{\mathcal{E}(\boldsymbol{f}_j)-\mathcal{E}_\Lambda(\boldsymbol{f}_j)}{2\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}&\ge\frac{\mathcal{E}(\boldsymbol{f})-\mathcal{E}_\Lambda(\boldsymbol{f})}{2\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}-\frac{\mathcal{E}(\boldsymbol{f})-\mathcal{E}(\boldsymbol{f}_j)}{2\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}-\frac{\mathcal{E}_\Lambda(\boldsymbol{f}_j)-\mathcal{E}_\Lambda(\boldsymbol{f})}{2\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}\\&>2\gamma\sqrt{\epsilon}-\frac{\gamma\sqrt{\epsilon}}{2}-\frac{\gamma\sqrt{\epsilon}}{2}=\gamma\sqrt{\epsilon}.\end{aligned}$$

This together with the fact $\sqrt{\mathcal{E}(\boldsymbol{f}_j)+\epsilon}\le2\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}$ implies that for any $\boldsymbol{f}\in\{\boldsymbol{f}:\|\boldsymbol{f}-\boldsymbol{f}_j\|_\infty\le\frac{\gamma\epsilon}{2M}\}$, if the condition $\frac{\mathcal{E}(\boldsymbol{f})-\mathcal{E}_\Lambda(\boldsymbol{f})}{\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}>4\gamma\sqrt{\epsilon}$ holds, then the following inequality

$$\frac{\mathcal{E}(\boldsymbol{f}_j)-\mathcal{E}_\Lambda(\boldsymbol{f}_j)}{\sqrt{\mathcal{E}(\boldsymbol{f}_j)+\epsilon}}\ge\frac{\mathcal{E}(\boldsymbol{f}_j)-\mathcal{E}_\Lambda(\boldsymbol{f}_j)}{2\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}>\gamma\sqrt{\epsilon}$$

holds. Hence, for each fixed j, $1\le j\le K$,

$$\text{Prob}\left\{\sup_{\boldsymbol{f}\in\{\boldsymbol{f}:\|\boldsymbol{f}-\boldsymbol{f}_j\|_\infty\le\frac{\gamma\epsilon}{2M}\}}\frac{\mathcal{E}(\boldsymbol{f})-\mathcal{E}_\Lambda(\boldsymbol{f})}{\sqrt{\mathcal{E}(\boldsymbol{f})+\epsilon}}>4\gamma\sqrt{\epsilon}\right\}\le\text{Prob}\left\{\frac{\mathcal{E}(\boldsymbol{f}_j)-\mathcal{E}_\Lambda(\boldsymbol{f}_j)}{\sqrt{\mathcal{E}(\boldsymbol{f}_j)+\epsilon}}>\gamma\sqrt{\epsilon}\right\}.$$

Since $\mathcal{M} \subseteq \cup_j \{\boldsymbol{f} : \|\boldsymbol{f} - \boldsymbol{f}_j\|_\infty \leq \frac{\gamma\epsilon}{2M}\}$, we have

$$\text{Prob}\left\{\sup_{\boldsymbol{f}\in\mathcal{M}} \frac{\mathcal{E}(\boldsymbol{f}) - \mathcal{E}_\Lambda(\boldsymbol{f})}{\sqrt{\mathcal{E}(\boldsymbol{f}) + \epsilon}} > 4\gamma\sqrt{\epsilon}\right\} \leq \sum_{j=1}^{K} \text{Prob}\left\{\sup_{\boldsymbol{f}\in\{\boldsymbol{f}:\|\boldsymbol{f}-\boldsymbol{f}_j\|_\infty \leq \frac{\gamma\epsilon}{2M}\}} \frac{\mathcal{E}(\boldsymbol{f}) - \mathcal{E}_\Lambda(\boldsymbol{f})}{\sqrt{\mathcal{E}(\boldsymbol{f}) + \epsilon}} > 4\gamma\sqrt{\epsilon}\right\}.$$

Therefore,

$$\text{Prob}\left\{\sup_{\boldsymbol{f}\in\mathcal{M}} \frac{\mathcal{E}(\boldsymbol{f}) - \mathcal{E}_\Lambda(\boldsymbol{f})}{\sqrt{\mathcal{E}(\boldsymbol{f}) + \epsilon}} > 4\gamma\sqrt{\epsilon}\right\} \leq \sum_{j=1}^{K} \text{Prob}\left\{\frac{\mathcal{E}(\boldsymbol{f}_j) - \mathcal{E}_\Lambda(\boldsymbol{f}_j)}{\sqrt{\mathcal{E}(\boldsymbol{f}_j) + \epsilon}} > \gamma\sqrt{\epsilon}\right\}.$$

The right hand side can be further bounded by $\mathcal{N}(\mathcal{M}, \frac{\gamma\epsilon}{2M}) \exp\{-\frac{3\gamma^2 m\epsilon}{8M^2}\}$ by using the fact $K = \mathcal{N}(\mathcal{M}, \frac{r\epsilon}{2M})$ and (4.3). □

Finally, we give the proof of Theorem 2.3.

Proof of Theorem 2.3: By Proposition 4.3, for every $\epsilon > 0$ and $0 < \gamma \leq 1$, the inequality

$$\sup_{\boldsymbol{f}\in\mathcal{M}} \frac{\mathcal{E}(\boldsymbol{f}) - \mathcal{E}_\Lambda(\boldsymbol{f})}{\sqrt{\mathcal{E}(\boldsymbol{f}) + \epsilon}} \leq 4\gamma\sqrt{\epsilon}$$

holds with probability at least

$$1 - \mathcal{N}(\mathcal{M}, \frac{\gamma\epsilon}{2M}) \exp\left\{-\frac{3\gamma^2 m\epsilon}{8M^2}\right\}.$$

Therefore, for all $\boldsymbol{f} \in \mathcal{M}$, the inequality

$$\mathcal{E}(\boldsymbol{f}) - \mathcal{E}_\Lambda(\boldsymbol{f}) \leq 4\gamma\sqrt{\epsilon}\sqrt{\mathcal{E}(\boldsymbol{f}) + \epsilon} \tag{4.4}$$

holds with the same probability. Since the original data $\boldsymbol{p}$ satisfies the constraint

$$\frac{1}{|\Lambda|}\sum_{k\in\Lambda}(\boldsymbol{f}[k] - \boldsymbol{g}[k])^2 \leq \sigma^2 \text{ and } \|\boldsymbol{f}\|_\infty \leq M,$$

and $\boldsymbol{f}^\Lambda$ is the solution of (1.8), we have

$$\|\mathcal{A}\boldsymbol{f}^\Lambda\|_{\ell_1(\beta,\Upsilon)} \leq \|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)}.$$

Therefore, $\boldsymbol{f}^\Lambda \in \mathcal{M}$. Taking $\gamma = \sqrt{2}/8$ and $\boldsymbol{f} = \boldsymbol{f}^\Lambda$ in (4.4), we know that

$$\mathcal{E}(\boldsymbol{f}^\Lambda) - \mathcal{E}_\Lambda(\boldsymbol{f}^\Lambda) \leq \frac{1}{2}\sqrt{2\epsilon(\mathcal{E}(\boldsymbol{f}^\Lambda) + \epsilon)} \tag{4.5}$$

holds with probability at least $1 - \mathcal{N}(\mathcal{M}, \frac{\epsilon}{8\sqrt{2}M}) \exp\{-\frac{3m\epsilon}{256M^2}\}$. Furthermore, since

$$\begin{aligned}
\mathcal{E}_\Lambda(\boldsymbol{f}^\Lambda) &= \frac{1}{|\Lambda|}\sum_{k\in\Lambda}(\boldsymbol{f}^\Lambda[k] - \boldsymbol{p}[k])^2 \\
&\leq \frac{2}{|\Lambda|}\sum_{k\in\Lambda}(\boldsymbol{f}^\Lambda[k] - \boldsymbol{g}[k])^2 + \frac{2}{|\Lambda|}\sum_{k\in\Lambda}(\boldsymbol{p}[k] - \boldsymbol{g}[k])^2 \\
&\leq 4\sigma^2. \qquad (4.6)
\end{aligned}$$

Combining (4.5) with (4.6) yields

$$\mathcal{E}(\boldsymbol{f}^\Lambda) \leq \frac{1}{2}\sqrt{2\epsilon(\mathcal{E}(\boldsymbol{f}^\Lambda) + \epsilon)} + 4\sigma^2.$$

This together with the fact $\mathcal{N}(\mathcal{M}, \frac{\epsilon}{8\sqrt{2}M}) \leq \mathcal{N}(\mathcal{M}, \frac{\epsilon}{12M})$ implies $\mathcal{E}(\boldsymbol{f}^\Lambda) \leq \frac{16}{3}\sigma^2 + \epsilon$ with probability at least $1 - \mathcal{N}(\mathcal{M}, \frac{\epsilon}{12M}) \exp\{-\frac{3m\epsilon}{256M^2}\}$. Thus we get the conclusion of Theorem 2.3. □

4.2. Proof of Theorem 2.4. Theorem 2.4 is to estimate the covering number. As pointed out in Section 2, it is not easy to analyze $\mathcal{M}$ directly because of the complexity of the frame operator $\mathcal{A}$. Note that $\mathcal{D}$ is a linear operator. If we assume $\mathcal{A}$ satisfy the condition (2.5), then together with the definition of $\mathcal{M}$ and (1.4), we have

$$\|\mathcal{D}\boldsymbol{f}\|_1 \le \sum_{n\in\Gamma} |\langle \boldsymbol{f}, \boldsymbol{a}_n\rangle| \|\mathcal{D}\boldsymbol{a}_n\|_1 \le C_d \sum_{n\in\Gamma} |\langle \boldsymbol{f}, \boldsymbol{a}_n\rangle| 2^{\alpha\Upsilon(n)} \tag{4.7}$$

which can be further bounded by

$$C_d 2^{\max\{\alpha-\beta,0\}\max\{\Upsilon(n):,n\in\Gamma\}} \sum_{n\in\Gamma} 2^{\beta\Upsilon(n)} |\langle \boldsymbol{f}, \boldsymbol{a}_n\rangle|.$$

Recall that Υ is a function mapping from Γ to $\mathbb{N}$ satisfying (1.7). Thus, for any $\boldsymbol{f} \in \mathcal{M}$, we have

$$\|\mathcal{D}\boldsymbol{f}\|_1 \le C_d |\Omega|^{\frac{\max\{\alpha-\beta,0\}}{d}} \|\mathcal{A}\boldsymbol{f}\|_{\ell_1(\beta,\Upsilon)} \le C_d |\Omega|^{\frac{\max\{\alpha-\beta,0\}}{d}} \|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)}.$$

Therefore, $\mathcal{M} \subset \widetilde{\mathcal{M}}$ and $\mathcal{N}(\mathcal{M},\eta) \le \mathcal{N}(\widetilde{\mathcal{M}},\eta)$, where $\widetilde{\mathcal{M}}$ is defined by (2.6). Now we only need to bound the covering number $\mathcal{N}(\widetilde{\mathcal{M}},\eta)$.

By the definition of covering number $\mathcal{N}(\widetilde{\mathcal{M}},\eta)$, it is easy to see that when there is a finite set $F \subseteq \widetilde{\mathcal{M}}$ such that $\widetilde{\mathcal{M}} \subseteq \cup_{\boldsymbol{q}\in F}\{\boldsymbol{f} : \|\boldsymbol{f}-\boldsymbol{q}\|_\infty \le \eta\}$, then $\mathcal{N}(\widetilde{\mathcal{M}},\eta) \le |F|$, where $|F|$ is the number of elements in set F. What we need now is to construct a good set F by exploiting the specific structure of $\widetilde{\mathcal{M}}$, so that $|F|$ has a good upper bound that gives a good estimate of the covering number $\mathcal{N}(\widetilde{\mathcal{M}},\eta)$, and further the one of $\mathcal{N}(\mathcal{M},\eta)$. To do so, we need the following lemma. First, we introduce the set R. Let $r = \lceil \frac{2M}{\eta} \rceil$, i.e. r is the smallest integer greater than $\frac{2M}{\eta}$. Define

$$R = \{-r\eta/2, -(r-1)\eta/2, \ldots, r\eta/2\}. \tag{4.8}$$

LEMMA 4.4. *Let $\widetilde{\mathcal{M}}$ be defined as* (2.6) *and R as* (4.8). *Then for each $\boldsymbol{f} \in \widetilde{\mathcal{M}}$, there exists a vector $Q(\boldsymbol{f})$ taking values in R and satisfying $\|\boldsymbol{f} - Q(\boldsymbol{f})\|_\infty \le \frac{\eta}{2}$ and $\|\mathcal{D}Q(\boldsymbol{f})\|_1 \le \|\mathcal{D}\boldsymbol{f}\|_1$, where $\mathcal{D}$ is the discrete total variation (DTV) operator defined by* (2.3).

This lemma is vital in constructing the set F, as shown in the following proof of Theorem 2.4. The proof of Lemma 4.4 is delayed to the end of this subsection.

Proof of Theorem 2.4: As shown in the above discussion that $\mathcal{N}(\mathcal{M},\eta) \le \mathcal{N}(\widetilde{\mathcal{M}},\eta)$, we only need to bound the covering number $\mathcal{N}(\widetilde{\mathcal{M}},\eta)$. The major part of the proof is to construct a set $F \subseteq \widetilde{\mathcal{M}}$ such that $\widetilde{\mathcal{M}} \subseteq \cup_{\boldsymbol{q}\in F}\{\boldsymbol{f} : \|\boldsymbol{f}-\boldsymbol{q}\|_\infty \le \eta\}$ and a good upper bound of the total number of elements in F provides a desired upper bound of the covering number $\mathcal{N}(\widetilde{\mathcal{M}},\eta)$.

Lemma 4.4 says that for each $\boldsymbol{f} \in \widetilde{\mathcal{M}}$, there exists a vector $Q(\boldsymbol{f})$ whose range is R and satisfying $\|\boldsymbol{f} - Q(\boldsymbol{f})\|_\infty \le \frac{\eta}{2}$ and $\|\mathcal{D}Q(\boldsymbol{f})\|_1 \le \|\mathcal{D}\boldsymbol{f}\|_1$. Let

$$\widetilde{F} = \{\boldsymbol{q} \in \ell_\infty(\Omega) : \boldsymbol{q} = Q(\boldsymbol{f}), \text{ for some } \boldsymbol{f} \in \widetilde{\mathcal{M}}\}.$$

This is a subset of the set of sequences defined on Ω and whose range is R. For each fixed element $\boldsymbol{q}$ in $\widetilde{F}$, there may have more than one element in $\widetilde{\mathcal{M}}$ satisfying $\boldsymbol{q} = Q(\boldsymbol{f})$.

For each fixed $\boldsymbol{q} \in \widetilde{F}$, choose a vector $\boldsymbol{f}_{\boldsymbol{q}} \in \widetilde{\mathcal{M}}$ such that $\|\boldsymbol{f}_{\boldsymbol{q}} - \boldsymbol{q}\|_\infty \le \frac{\eta}{2}$ and define $F = \{\boldsymbol{f}_{\boldsymbol{q}} : \boldsymbol{q} \in \widetilde{F}\}$. Then, F is a subset of $\widetilde{\mathcal{M}}$. For an arbitrary given $\boldsymbol{f} \in \widetilde{\mathcal{M}}$, there exists a function $\boldsymbol{q} \in \widetilde{F}$ such that $\|\boldsymbol{f} - \boldsymbol{q}\|_\infty \le \frac{\eta}{2}$ which implies

$$\|\boldsymbol{f} - \boldsymbol{f}_{\boldsymbol{q}}\|_\infty \le \|\boldsymbol{f} - \boldsymbol{q}\|_\infty + \|\boldsymbol{q} - \boldsymbol{f}_{\boldsymbol{q}}\|_\infty \le \eta$$

by the definition of $\boldsymbol{f}_{\boldsymbol{q}}$. Therefore,

$$\widetilde{\mathcal{M}} \subset \cup_{\boldsymbol{f}_{\boldsymbol{q}}\in F}\{\boldsymbol{f} : \|\boldsymbol{f} - \boldsymbol{f}_{\boldsymbol{q}}\|_\infty \le \eta\} \text{ and } \mathcal{N}(\widetilde{\mathcal{M}},\eta) \le |F| \le |\widetilde{F}|.$$

Thus, the upper bound of $|\widetilde{F}|$ will give an upper bound of $|F|$, hence the covering number $\mathcal{N}(\widetilde{\mathcal{M}},\eta)$ is bounded by any upper bound of $|\widetilde{F}|$. Note that for an arbitrary vector $\boldsymbol{q}\in\widetilde{F}$, it is uniquely determined by the sequence $\mathcal{D}\boldsymbol{q}$ and $\boldsymbol{q}[1,\dots,1]$. Therefore, in order to bound the number of elements in set $\widetilde{F}$, we only need to bound the number of the choices of the sequence $\mathcal{D}\boldsymbol{q}$ and $\boldsymbol{q}[1,\dots,1]$. As the range of $\boldsymbol{q}$ is the set R defined as (4.8), we have $2r+1$ choices of $\boldsymbol{q}[1,\dots,1]$. What left is to count the number of choices of the vector $\mathcal{D}\boldsymbol{q}$. Define

$$\mathcal{D}\widetilde{F}=\{\mathcal{D}\boldsymbol{q}:\boldsymbol{q}\in\widetilde{F}\}.$$

Then, we need to estimate an upper bound of $|\mathcal{D}\widetilde{F}|$, i.e., an upper bound of the total number of elements in $\mathcal{D}\widetilde{F}$. To do that, we first find a uniform bound of $\|\mathcal{D}\boldsymbol{q}\|_1$ for $\boldsymbol{q}\in\widetilde{F}$.

By the definition of $\widetilde{F}$ and lemma 4.4, for a fixed $\boldsymbol{q}\in\widetilde{F}$, there exists a function $\boldsymbol{f}\in\widetilde{\mathcal{M}}$ such that $\|\mathcal{D}\boldsymbol{q}\|_1\le\|\mathcal{D}\boldsymbol{f}\|_1$. Since $\|\mathcal{A}\boldsymbol{p}\|_{\ell_1(\beta,\Upsilon)}\le C_{\boldsymbol{p}}|\Omega|^{\frac{1}{2}}$ by assumption, $\|\mathcal{D}\boldsymbol{q}\|_1$ can be further bounded by $C_dC_{\boldsymbol{p}}|\Omega|^{\frac{\max\{2(\alpha-\beta)+d,d\}}{2d}}$. With this bound, we now can estimate the number of choices of the sequence $\mathcal{D}\boldsymbol{q}$, $\boldsymbol{q}\in\widetilde{F}$. Let

$$K=\left\lceil\frac{2C_dC_{\boldsymbol{p}}|\Omega|^{\frac{\max\{2(\alpha-\beta)+d,d\}}{2d}}}{\eta}\right\rceil.$$

Hence for an arbitrary element $\boldsymbol{q}\in\widetilde{F}$, $\|\mathcal{D}\boldsymbol{q}\|_1\le K\eta$. Furthermore, since the range of $\boldsymbol{q}$ is R, the range of the vector $\mathcal{D}\boldsymbol{q}$ is the set $\{-K\eta,-(K-1)\eta,\dots,K\eta\}$. Recall that there are $d(|\Omega|-|\Omega|^{\frac{d-1}{d}})$ entries in vector $\mathcal{D}\boldsymbol{q}$. Therefore, the bound of $|\mathcal{D}\widetilde{F}|$ can be estimated as the total number of the following events: Consider $d(|\Omega|-|\Omega|^{\frac{d-1}{d}})$ ordered balls, we choose K balls from them with replacement. After this, assign each ball with possible value either $0,\eta$ or $-\eta$. This implies that the $|\mathcal{D}\widetilde{F}|$ can be bounded by

$$\begin{aligned}&3^KK!\binom{d(|\Omega|-|\Omega|^{\frac{d-1}{d}})+K-1}{K}\\&=3^K(d(|\Omega|-|\Omega|^{\frac{d-1}{d}})+K-1)\cdot(d(|\Omega|-|\Omega|^{\frac{d-1}{d}})+K-2)\dots(d(|\Omega|-|\Omega|^{\frac{d-1}{d}})-1).\end{aligned}\tag{4.9}$$

As $\eta\ge|\Omega|^{\max\{\frac{d+2(\alpha-\beta)}{2d},\frac{1}{2}\}-1}$ by the assumption, we have $K-1\le 2C_dC_{\boldsymbol{p}}|\Omega|$. Hence, (4.9) can be further bounded by $(3(d+2C_dC_{\boldsymbol{p}})|\Omega|)^K$. Recall that the number of the choices of $\boldsymbol{q}[1,\dots,1]$ is bounded by $2r+1$, we have

$$|\widetilde{F}|\le(2r+1)(3(d+2C_dC_{\boldsymbol{p}})|\Omega|)^K.$$

Therefore,

$$\log\mathcal{N}(\mathcal{M},\eta)\le\log\mathcal{N}(\widetilde{\mathcal{M}},\eta)\log|\widetilde{F}|\le(4M+3(d+2C_dC_{\boldsymbol{p}}))\frac{|\Omega|^{\max\{\frac{2(\alpha-\beta)+d}{2d},\frac{1}{2}\}}\log_2|\Omega|}{\eta}.$$

This leads to the desired inequality (2.7) by letting $C_d'=4M+3(d+2C_dC_{\boldsymbol{p}})$. □

Next, we prove Lemma 4.4.

Proof of Lemma 4.4: We prove this lemma in a constructive way. Recall that $r=\lceil\frac{2M}{\eta}\rceil$ and $R=\{-r\eta/2,-(r-1)\eta/2,\dots,r\eta/2\}$. First, we note that for $\boldsymbol{f}\in\widetilde{\mathcal{M}}$, if its range is in R, then we simply choose $Q(\boldsymbol{f})=\boldsymbol{f}$.

For general $\boldsymbol{f}\in\widetilde{\mathcal{M}}$, one can participate the domain Ω into $2r+1$ parts. Indeed, let $U_i=\{k\in\Omega:(i-1)\eta/2\le\boldsymbol{f}[k]<i\eta/2\}$, $i=-r+1,\dots,r+1$. Then $U_i\cap U_j=\emptyset$ and $\cup_iU_i=\Omega$, since $|\boldsymbol{f}[k]|\le M$ for $k=(k_1,\dots,k_d)\in\Omega$ whenever $\boldsymbol{f}\in\widetilde{\mathcal{M}}$. We observe that in order to make the range of $Q(\boldsymbol{f})$ be in R and $\|Q(\boldsymbol{f})-\boldsymbol{f}\|_\infty\le\frac{\eta}{2}$, one needs to move the value of $\boldsymbol{f}[k]$, $k\in U_i$ to either $(i-1)\eta/2$ or $i\eta/2$. The choice is finally determined by $\|\mathcal{D}Q(\boldsymbol{f})\|_1\le\|\mathcal{D}\boldsymbol{f}\|_1$, that is quite involved. We want to define $Q(\boldsymbol{f})$ on each $\{U_i\}_{-r+1\le i\le r+1}$, however, it involves the behavior of $\boldsymbol{f}$ on the whole Ω.

In order to overcome this difficulty, we define set

$$B_{i,j,n} = \{(k,k') : k \in U_i, k' \in U_j, k_n = k'_n + 1, k_{n'} = k'_{n'}, n' \neq n\}$$

for each $-r+1 \leq i,j \leq r+1, 1 \leq n \leq d$. Here the vector (k,k') has $2d$ entries and the difference between vector k and k' is the nth entry. Let $B_i = \cup_{j\neq i} \cup_{1\leq n\leq d} B_{ijn}$. Then,

$$B_i := \{(k,k') : k \in U_i, k' \in \Omega\backslash U_i,\ k,\ k' \text{ only differ by one entry and } \sum_{n=1}^{d} k'_n - k_n = 1\}$$

For each $(k,k') \in B_{i,j,n}$, if $f[k'] \neq \frac{i\eta}{2}$, we add $|i-j|$ new points between $k = (k_1,\ldots,k_d)$ and $k' = (k'_1,\ldots,k'_d)$:

$$x_\ell^{(k,k')} = (k_1,\ldots,k_{n-1},k_n + \frac{\ell}{|i-j|+1}, k_{n+1},\ldots,k_d),\ \forall 1 \leq \ell \leq |i-j|.$$

Let

$$X^{(k,k')} = \begin{cases} \{x_1^{(k,k')},\ldots,x_{|i-j|}^{(k,k')}\}, & \text{if } f[k'] \neq \frac{i\eta}{2}, \\ \emptyset, & \text{if } f[k'] = \frac{i\eta}{2}. \end{cases}$$

and extend $\boldsymbol{f}$ to $X^{(k,k')}$ as follows:

$$\boldsymbol{f}[x_\ell^{(k,k')}] = (i + \ell\text{sign}(j-i))\eta/2,\quad \forall 1 \leq \ell \leq |i-j|. \tag{4.10}$$

This definition indicates that the sequence $\{\boldsymbol{f}[k], \boldsymbol{f}[x_1^{(k,k')}], \boldsymbol{f}[x_2^{(k,k')}],\ldots,\boldsymbol{f}[x_{|i-j|}^{(k,k')}], \boldsymbol{f}[k']\}$ is monotonically increasing or decreasing. Therefore, for any $(k,k') \in B_{i,j,n}$, we have

$$|\boldsymbol{f}[k] - \boldsymbol{f}[k']| = |\boldsymbol{f}[k] - \boldsymbol{f}[x_1^{(k,k')}]| + |\boldsymbol{f}[x_1^{(k,k')}] - \boldsymbol{f}[x_2^{(k,k')}]| + \cdots + |\boldsymbol{f}[x_{|i-j|}^{(k,k')}] - \boldsymbol{f}[k']|, \tag{4.11}$$

which will be used later.

Let $\Omega' := \cup_{(k,k')\in\cup_i B_i} X^{(k,k')} \cup \Omega$. While Ω is a subset of the lattice $\mathbb{Z}^d$, Ω' is more complicated. It is a subset of a more dense lattice and it is nonuniform. Nevertheless, by (4.10) we have extended $\boldsymbol{f}$ defined on Ω to a sequence defined on Ω'. To avoid confusion, in what follows, we use $\boldsymbol{f}|_\Omega$ to represent the original sequence $\boldsymbol{f}$ defined on Ω, when we write $\boldsymbol{f}$ indicating the domain is Ω'. Let

$$\widetilde{B} = \{(k,k') \in \Omega \times \Omega :\ k,\ k' \text{ only differ by one entry and } \sum_{\ell=1}^{d} k'_\ell - k_\ell = 1\}.$$

Then $\|\mathcal{D}\boldsymbol{f}|_\Omega\|_1 = \sum_{(k,k')\in\widetilde{B}} |\boldsymbol{f}_\Omega[k'] - \boldsymbol{f}_\Omega[k]|$ by (2.4). One can extend the definition of discrete total variation of (2.4) to the more complicated set Ω', but we will not do it, since we do not need it. However, we need to use the following number

$$\begin{aligned} Z = & \sum_{(k,k')\in\widetilde{B}\backslash(\cup_i B_i)} |\boldsymbol{f}[k'] - \boldsymbol{f}[k]| \\ & + \sum_{(k,k')\in\cup_i B_i} |\boldsymbol{f}[k] - \boldsymbol{f}[x_1^{(k,k')}]| + |\boldsymbol{f}[x_1^{(k,k')}] - \boldsymbol{f}[x_2^{(k,k')}]| + \cdots + |\boldsymbol{f}[x_{|i-j|}^{(k,k')}] - \boldsymbol{f}[k']|. \end{aligned} \tag{4.12}$$

Then $Z = \|\mathcal{D}\boldsymbol{f}|_\Omega\|_1$ by (4.11).

Next, we regroup the terms in Z so that Z can be written as a sum of $2r$ terms with each term only involves the points in

$$\widetilde{U}_i = \{x \in \Omega' : (i-1)\eta/2 \leq \boldsymbol{f}[x] \leq i\eta/2\}.$$

For this, let $x = (x_1, \dots, x_d) \in \Omega'$ and $n \in \{1, \dots, d\}$ be given. Define

$$O_x^n = \{x' \in \Omega' : x'_n > x_n, x_{n'} = x'_{n'}, n' \neq n\}, \quad \text{and} \quad N_x^n = \{y \in \Omega' : y = \arg\min_{x' \in O_x^n} \|x' - x\|_1\}.$$

Let

$$N_x := \cup_{n=1}^d N_x^n \text{ and } U'_i = \{x \in \Omega' : (i-1)\eta/2 \leq \boldsymbol{f}[x] < i\eta/2\}.$$

Then $\|\mathcal{D}\boldsymbol{f}|_\Omega\|_1$ can be written as follows:

$$\|\mathcal{D}\boldsymbol{f}|_\Omega\|_1 = Z = \sum_{x \in \Omega'} \sum_{x' \in N_x} |\boldsymbol{f}[x'] - \boldsymbol{f}[x]| = \sum_{i=-r+1}^{r} \sum_{x \in U'_i} \sum_{x' \in N_x} |\boldsymbol{f}[x'] - \boldsymbol{f}[x]|. \tag{4.13}$$

Note that the right hand side is a summation of $2r$ parts and each part only involves the points in $\widetilde{U}_i$. This property is important and it makes us feasible to deal with each part separately. In fact, the main purpose of the extension of Ω to Ω' is to insert sufficient points into Ω so that (4.13) holds.

Next, for each $i \in \{-r+1, -r+2, \dots, r\}$, we construct a sequence f_i^* defined on $\widetilde{U}_i$ satisfying the following three conditions: the range of f_i^* is $\{(i-1)\eta/2, i\eta/2\}$; $\boldsymbol{f}_i^*$ coincide with $\boldsymbol{f}$ on the set $\{x \in \widetilde{U}_i : \boldsymbol{f}[x] = i\eta/2 \text{ or } \boldsymbol{f}[x] = (i-1)\eta/2\}$; and

$$\sum_{x \in U'_i} \sum_{x' \in N_x} |\boldsymbol{f}_i^*[x'] - \boldsymbol{f}_i^*[x]| \leq \sum_{x \in U'_i} \sum_{x' \in N_x} |\boldsymbol{f}[x'] - \boldsymbol{f}[x]|. \tag{4.14}$$

Then the desired result follows by letting

$$Q(\boldsymbol{f}|_\Omega)[k] = \boldsymbol{f}_i^*[k], k \in \Omega, i = -r+1, -r+2, \dots, r. \tag{4.15}$$

For each fixed $i \in \{-r+1, -r+2, \cdots, r\}$, we construct a simple graph $G = (V, E)$. Let the set of vertices $V = \widetilde{U}_i$ and we link $x \in E$ and $y \in E$ by $e(x, y)$, an edge connecting x and y, if either $x \in N_y \cap \widetilde{U}_i$ or $y \in N_x \cap \widetilde{U}_i$ holds, i.e. the set of edges is $E = \{e(x, y) : x \in N_y \cap \widetilde{U}_i \text{ or } y \in N_x \cap \widetilde{U}_i\}$. Obviously, graph $G = (V, E)$ is connected.

Consider the vertices $P_i := \{x \in \widetilde{U}_i : \boldsymbol{f}[x] = (i-1)\eta/2\} \subset E$ and $Q_i := \{x \in \widetilde{U}_i : \boldsymbol{f}[x] = i\eta/2\} \subset E$. Let K_1 be the minimum size of an edge set in E whose removal disconnects P_i and Q_i and K_2 be the maximum size of a family of pairwise edge disjoint paths from P_i to Q_i. By max-flow min-cut theorem [36], $K_1 = K_2$.

Let W be a subset in E whose removal disconnects P_i and Q_i and satisfying $|W| = K_1$. We construct a new graph $G' = (V, E')$ with $E' = E \backslash W$. Then the vertices of graph G' can naturally be divided into two parts according to their connection with P_i and Q_i in graph G'. More specifically, choosing S_i to be the largest set satisfying that $P_i \subseteq S_i$ and there exists a path in G' connecting x and P_i for each vertex $x \in S_i \backslash P_i$, and T_i be the largest set satisfying that $Q_i \subseteq T_i$ and there exists a path connecting y and Q_i for each vertex $y \in T_i \backslash Q_i$, we have $T_i \cup S_i = \widetilde{U}_i$.

Define

$$\boldsymbol{f}_i^*[x] = \begin{cases} (i-1)\eta/2, & \text{if } x \in S_i, \\ i\eta, & \text{if } x \in T_i. \end{cases} \tag{4.16}$$

It is clear that

$$\sum_{x \in U'_i} \sum_{x' \in N_x} |\boldsymbol{f}_i^*[x'] - \boldsymbol{f}_i^*[x]| = K_1\eta/2. \tag{4.17}$$

Consider a pairwise edge disjoint path in G connecting P_i to Q_i with ordered vertices $\{x^1, \dots x^K\}$ for some $K > 0$ that satisfy $\boldsymbol{f}[x^1] = (i-1)\eta/2, \boldsymbol{f}[x^K] = i\eta/2$ and $\{e(x^1, x^2), e(x^2, x^3), \dots, e(x^{K-1}, x^K)\} \subseteq E$. Applying the triangle inequality, we have

$$\sum_{n=1}^{K-1} |\boldsymbol{f}[x^{n+1}] - \boldsymbol{f}[x^n]| \geq \sum_{n=1}^{K-1} (\boldsymbol{f}[x^{n+1}] - \boldsymbol{f}[x^n]) = \boldsymbol{f}[x^K] - \boldsymbol{f}[x^1] = \eta/2.$$

This together with the definition of the graph G and K_2 yields

$$\sum_{x\in U_i'}\sum_{x'\in N_x}|\boldsymbol{f}[x']-\boldsymbol{f}[x]|\geq K_2\eta/2.$$

By applying the fact $K_1=K_2$ and (4.17), inequality (4.14) follows with $\boldsymbol{f}_i^*$ defined by (4.16) and the desired function $Q(\boldsymbol{f}|_\Omega)$ can then be constructed by the equation (4.15). ∎

Note that in our proof, the regularity of the sequence in $\widetilde{\mathcal{M}}$ is measured by discrete total variation and our estimation for the covering number $\mathcal{N}(\mathcal{M},\eta)$ is exactly done by estimating $\mathcal{N}(\widetilde{\mathcal{M}},\eta)$. So our analysis used here is also applicable for other general setting whose involved set is contained in a set of the form $\{\boldsymbol{f}\in\ell_\infty(\Omega):\|\mathcal{D}\boldsymbol{f}\|_1\leq C'_{\boldsymbol{p}}\}$ for some constant $C'_{\boldsymbol{p}}>0$. Obviously, some TV based algorithm is included.

4.3. Proof of Lemma 3.1. To apply Theorem 2.4 to the framelet case, we need to know α, explicitly, in condition (2.5). In this section, we show Lemma 3.1 which states that the tight frame system defined in (3.6) and (3.7) derived from filters of the tight framelets $X(\Psi,\phi)$ satisfies (2.5) with $\alpha=0$. For this, we need some discussions of the convergent rate of a stationary subdivision algorithm.

Let $\phi:\mathbb{R}^2\mapsto\mathbb{R}^2$ be a compactly supported refinable function with mask $\{\boldsymbol{h}_0[k],k\in\mathbb{Z}^2\}$ and satisfies refinement equation (3.1). For $\boldsymbol{d}\in\ell_\infty(\mathbb{Z}^2)$, we define subdivision operator $\mathcal{S}$ as

$$(\mathcal{S}\boldsymbol{d})[k]=4\sum_{i\in\mathbb{Z}^2}\boldsymbol{h}_0[k-2i]\boldsymbol{d}[i]. \tag{4.18}$$

We say that the subdivision algorithm converges if

$$\|f_{\boldsymbol{d}}(\frac{\cdot}{2^n})-\mathcal{S}^n\boldsymbol{d}|_\infty\to 0$$

as $n\to\infty$, where

$$f_{\boldsymbol{d}}(x)=\sum_{k\in\mathbb{Z}^2}\boldsymbol{d}[k]\phi(x-k), x\in\mathbb{R}^2.$$

The convergence of the subdivision algorithm is a very well studied subject. In fact, there is a complete characterization of the convergence in terms of refinement mask. The interested reader should consult [42,48] for details in this direction. It is even known earlier that when the refinable function ϕ and its shifts form a Riesz system, the corresponding subdivision algorithm converges (see e.g. [19]). Here, we need a convergence rate which is given below. The proof is standard and can be modified from that of Proposition 2.3 in [19].

PROPOSITION 4.5. *Assume the refinable function ϕ satisfies Assumption 1. Then the subdivision algorithm defined by* (4.18) *satisfies*

$$\|f_{\boldsymbol{d}}(\frac{\cdot}{2^n})-\mathcal{S}^n\boldsymbol{d}|_\infty\leq C_{\boldsymbol{d}}2^{-n}, \tag{4.19}$$

where $C_{\boldsymbol{d}}$ is a constant independent of n.

Proof. By a standard argument of the subdivision, one has that

$$|(\mathcal{S}^n\boldsymbol{d})[k]-f_{\boldsymbol{d}}(\frac{k}{2^n})|\leq\|\boldsymbol{d}\|_\infty\sum_{j\in\mathbb{Z}^2}|(\mathcal{S}^n\boldsymbol{\delta})[k-2^nj]-\phi(\frac{k}{2^n}-j)|, \tag{4.20}$$

where $\boldsymbol{\delta}$ is a vector with $\boldsymbol{\delta}[k]=1$ for the case $k=(0,0)$ and $\boldsymbol{\delta}[k]=0$ otherwise.

Therefore, once one proves that there exists a constant $\widetilde{C}_1\geq 0$ such that

$$\|\phi(\frac{k}{2^n})-\mathcal{S}^n\boldsymbol{\delta}[k]\|_\infty\leq\widetilde{C}_12^{-n}, \tag{4.21}$$

then (4.21) together with (4.20) yields that for all $k\in\mathbb{Z}^2$,

$$|(\mathcal{S}^n\boldsymbol{d})[k]-f_{\boldsymbol{d}}(\frac{k}{2^n})|\leq\|\boldsymbol{d}\|_\infty\widetilde{C}_1N2^{-n},$$

where N is the number of non-zero terms in the sum on the right hand side of (4.20). Since ϕ is of compact support and $supp\mathcal{S}^n\boldsymbol{\delta}$ is contained in some ball of radius $2^n r$ with r being a positive constant independent of n, N is a constant independent of n. This leads to the desired result, i.e.

$$\|f_{\boldsymbol{d}}(\frac{\cdot}{2^n}) - \mathcal{S}^n\boldsymbol{d}\|_\infty \le C_{\boldsymbol{d}}2^{-n}$$

by choosing $C_{\boldsymbol{d}} = \|\boldsymbol{d}\|_\infty\widetilde{C}_1 N$.

Finally, we prove (4.21). By using the fact $\phi(x) = 4\sum_{k\in\mathbb{Z}^2}\boldsymbol{h}_0[k]\phi(2x-k)$, one gets

$$\phi(x) = \sum_{k\in\mathbb{Z}^2}\boldsymbol{\delta}[k]\phi(x-k) = \sum_{k\in\mathbb{Z}^2}\mathcal{S}^n\boldsymbol{\delta}[k]\phi(2^n x-k). \tag{4.22}$$

Since ϕ is Hölder continuity with exponent 1 and compactly supported, there exists a constant $\widetilde{C}_2$ independent of n such that

$$|\sum_{k\in\mathbb{Z}^2}(\phi(\frac{k}{2^n}) - \phi(x))\phi(2^n x-k)| \le \widetilde{C}_2 2^{-n},\ x\in\mathbb{R}^s. \tag{4.23}$$

Note that by the assumption of ϕ, we have $\sum_{k\in\mathbb{Z}^2}\phi(\cdot-k) = 1$ (see [30, 49]). This together with (4.22) and (4.23) yields

$$|\sum_{k\in\mathbb{Z}^2}(\phi(\frac{k}{2^n}) - \mathcal{S}^n\boldsymbol{\delta}[k])\phi(2^n x-k)| \le \widetilde{C}_3 2^{-n}. \tag{4.24}$$

Since ϕ is compactly supported and $\{\phi(\cdot-k)\}_k$ is a Riesz basis in $L_2(\mathbb{R}^2)$, according to Theorem 3.5 in [43], the stability condition

$$\|\boldsymbol{c}\|_\infty \le C_\infty\|\sum_{k\in\mathbb{Z}^2}\boldsymbol{c}[k]\phi(2^n x-k)\|_\infty \tag{4.25}$$

holds for some constant C_∞ and all $\boldsymbol{c}\in\ell_\infty(\mathbb{Z}^2)$. Therefore, we obtain (4.21) from (4.24) and (4.25) by letting $\widetilde{C}_1 = \widetilde{C}_3 C_\infty$. The desired result then follows. ◻

Finally, we use the above proposition 4.5, to obtain the bounds of $\|\mathcal{D}\boldsymbol{a}_0\|_1$ and $\|\mathcal{D}\boldsymbol{b}_j^\ell\|_1$.

Proof of Lemma 3.1: The definition $\widetilde{\boldsymbol{a}}_0$ of (3.3) gives $2^J\widetilde{\boldsymbol{a}}_0 = S^J\boldsymbol{\delta}$. Denote $\boldsymbol{\Phi} = (\phi(\frac{k}{2^J}))_{k\in\Omega}$. Then (4.19) leads to, by taking $\boldsymbol{d} = \boldsymbol{\delta}$,

$$\|\boldsymbol{\Phi} - 2^J\widetilde{\boldsymbol{a}}_0\|_\infty \le C_{\boldsymbol{\delta}}2^{-J}. \tag{4.26}$$

Let $\widetilde{B} = \{(k,k') = (k_1,\dots,k_d,k_1',\dots,k_d')\in\Omega\times\Omega : k,k'$ only differ by one entry and $\sum_{n=1}^d k_n' - k_n = 1\}$. Then the number of elements in $\widetilde{B}$ is bounded by $2(2^{2J}-2^J)$ and

$$\|\mathcal{D}\widetilde{\boldsymbol{a}}_0\|_1 = \sum_{(k,k')\in\widetilde{B}}\left|\widetilde{\boldsymbol{a}}_0\left(\frac{k'}{2^J}\right) - \widetilde{\boldsymbol{a}}_0\left(\frac{k}{2^J}\right)\right|. \tag{4.27}$$

For any $(k,k')\in\widetilde{B}$, using (4.26), we have

$$-C_{\boldsymbol{\delta}}2^{-J} \le \phi\left(\frac{k}{2^J}\right) - 2^J\boldsymbol{a}_0\left(\frac{k}{2^J}\right) \le C_{\boldsymbol{\delta}}2^{-J},\quad -C_{\boldsymbol{\delta}}2^{-J} \le \phi\left(\frac{k'}{2^J}\right) - 2^J\boldsymbol{a}_0\left(\frac{k'}{2^J}\right) \le C_{\boldsymbol{\delta}}2^{-J}.$$

Therefore,

$$\left|\boldsymbol{a}_0\left(\frac{k'}{2^J}\right) - \boldsymbol{a}_0\left(\frac{k}{2^J}\right)\right| \le 2^{-J}\left|\phi\left(\frac{k'}{2^J}\right) - \phi\left(\frac{k}{2^J}\right)\right| + 2C_{\boldsymbol{\delta}}2^{-2J}.$$

The right hand side of the inequality can further be bounded by $(\widetilde{C}+2C_{\boldsymbol{\delta}})2^{-2J}$ by using the assumption that ϕ is Hölder continuous with exponent 1, i.e. there exists a constant $\widetilde{C}$ such that for any $x, y \in \mathbb{R}^2, |\phi(x)-\phi(y)| \leq \widetilde{C}\|x-y\|$. Together with (4.27) and the fact that $|\widetilde{B}| \leq 2(2^{2J}-2^J)$, we have

$$\|\mathcal{D}\widetilde{\boldsymbol{a}}_0\|_1 \leq \sum_{(k,k')\in\widetilde{B}} (\widetilde{C}+2C_{\boldsymbol{\delta}})2^{-2J} \leq 2\widetilde{C}+4C_{\boldsymbol{\delta}}.$$

Note that equation (3.4) for calculating $\widetilde{\boldsymbol{b}}_j^\ell$ can be seen as a subdivision with the mask $\{\boldsymbol{h}_0[k], k \in \mathbb{Z}^2\}$ and $\boldsymbol{d} = \boldsymbol{h}_\ell * \uparrow \boldsymbol{\delta}$. As done for $\widetilde{\boldsymbol{a}}_0$, we can similarly prove that there exists a constant $\widetilde{C}_3$ independent of j and ℓ such that $\|\mathcal{D}\widetilde{\boldsymbol{b}}_j^\ell\|_1 \leq \widetilde{C}_3$. Furthermore, using equations (3.6) and (3.7), there exists a constant $\widetilde{C}_4$ depending on the restriction operator $\mathcal{P}$ such that

$$\|\mathcal{D}\boldsymbol{a}_0\|_1 \leq \widetilde{C}_4\|\mathcal{D}\widetilde{\boldsymbol{a}}_0\|_1, \text{ and } \|\mathcal{D}\boldsymbol{b}_j^{\ell,k}\|_1 \leq \widetilde{C}_4\|\mathcal{D}\widetilde{\boldsymbol{b}}_j^\ell\|_1.$$

Thus, we get the first desired inequality (3.12) by letting $C_d = \max\{2\widetilde{C}_4(\widetilde{C}+2C_{\boldsymbol{\delta}}), \widetilde{C}_3\widetilde{C}_4, 1\}$.

Now we prove the second inequality (3.13). Equation (3.9) together with the linearity of the total variation operator $\mathcal{D}$ yields

$$\|\mathcal{D}\boldsymbol{f}\|_1 \leq |\langle \boldsymbol{a}_0, \boldsymbol{f}\rangle| \|\mathcal{D}\boldsymbol{a}_0\|_1 + \sum_{j=0}^{J-1}\sum_{k_1,k_2=0}^{2^j-1}\sum_{\ell=1}^{L} \langle \boldsymbol{b}_j^{\ell,k}, \boldsymbol{f}\rangle \|\mathcal{D}\boldsymbol{b}_j^{\ell,k}\|_1.$$

This together with equations (3.12) yields

$$\|\mathcal{D}\boldsymbol{f}\|_1 \leq C_d(|\langle \boldsymbol{a}_0, \boldsymbol{f}\rangle| + \sum_{j,k,\ell} |\langle \boldsymbol{b}_j^{\ell,k}, \boldsymbol{f}\rangle|) = C_d\|\mathcal{A}\boldsymbol{f}\|_1.$$

We get the desired estimate. ∎

REFERENCES

[1] N. Alon, S. Ben-David, S. N. Cesa-Bianchi, and D. Hanssler. *Scale-sensitive dimensions, uniform convergence and learnability*, Journal of the ACM, 1997(44), 615–631.

[2] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe, *Convexity, classification, and risk bounds*, Journal of the American Statistical Association, 2006(101), pp. 138–156.

[3] 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. New York, NY, USA, 2000, pp. 417–424.

[4] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, *Simultaneous structure and texture image inpainting*, IEEE Trans. Image Process., 12(2003), pp. 882–889.

[5] C, de Boor, K. Höllig, and S. Riemenschneider, *Box Splines*, Applied Mathematical Sciences 98, Springer-Verlag, 1993.

[6] L. Borup, R. Gribonval, and M. Nielsen, *Bi-framelet systems with few vanishing moments characterize Besov spaces*, Appl. Comput. Harmon. Anal., 17 (2004), pp. 3–28.

[7] A. Bovik, *Handbook of image and video processing (communications, networking and multimedia)*, Academic Press, Inc. Orlando, FL, USA, 2005.

[8] J.-F. Cai, R. Chan, L. Shen, and Z. Shen, *Restoration of chopped and nodded images by framelets*, SIAM J. Sci. Comput., 30 (2008), pp. 1205–1227.

[9] J.-F. Cai, R. H. Chan, L. Shen, and Z. Shen, *Simultaneously Inpainting in Image and Transformed Domains*, Numer. Math., 112 (2009), pp. 509–533.

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

[11] J.-F. Cai, E. J. Candès, and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim., 20 (2010), pp. 1956–1982.

[12] J.-F. Cai, R. H. Chan, L. Shen and Z. Shen, *Convergence analysis of tight framelet approach for missing data recovery*, Adv. Comput. Math., 31 (2009), 87-113.

[13] J.-F. Cai, S. Osher, and Z. Shen, *Split bregman methods and frame based image restoration*, Multiscale Model. Simul., 8 (2009), pp. 337–369.

[14] E.J. Candes and Y. Plan, *Matrix completion with noise*, Proceedings of the IEEE, (2009).

[15] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.

[16] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[17] E. J. Candès and T. Tao. *Decoding by linear programming.* IEEE Trans. Inform. Theory, 51(12):4203–4215, 2005.

[18] E. J. Candès and T. Tao. *Near-optimal signal recovery from random projections: universal encoding strategies?* IEEE Trans. Inform. Theory, 52(12):5406–5425, 2006.

[19] A. S. Cavaretta, W. Dahmen, and C. A. Micchelli, *Stationary subdivision*, Mem. Amer. Math. Soc., 93 (1991), pp. vi+186.

[20] A. Chai, and Z. Shen, *Deconvolution: A wavelet frame approach*, Numer. Math., 106 (2007), pp. 529–587.

[21] A. Chambolle, R. A. DeVore, N.-y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process., 7 (1998), pp. 319–335.

[22] R. Chan, C. Ho, and M. Nikolova, *Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization*, IEEE Trans. Image Process., 14 (2005), pp. 1479–1485.

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

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

[25] T. F. Chan, S. H. Kang, and J. Shen, *Euler's elastica and curvature-based inpainting*, SIAM J. Appl. Math., 63 (2002), pp. 564–592 (electronic).

[26] T. F. Chan, J. Shen, and H.-M. Zhou, *Total variation wavelet inpainting*, J. Math. Imaging Vision, 25 (2006), pp. 107–125.

[27] C.K. Chui and W. He, *Compactly supported tight frames associated with refinable functions*, Appl. Comput. Harmon. Anal., 8 (2000), pp. 293–319.

[28] C.K. Chui, W. He, and J. Stöckler, *Compactly supported tight and sibling frames with maximum vanishing moments*, Appl. Comput. Harmon. Anal., 13 (2002), 224–262.

[29] F. Cucker and S. Smale, *On the mathematical foundations of learning*, Bull. Amer. Math. Soc. (N.S.), 39 (2002), pp. 1–49 (electronic).

[30] I. Daubechies, *Ten lectures on wavelets*, vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

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

[32] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[33] D.L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory, 47 (2001), 2845–2862.

[34] B. Dong, and Z. Shen, *Pseudo-splines, wavelets and framelets*, Appl. Comput. Harmon. Anal., 22 (1) (2007), 78-104.

[35] 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.

[36] R. Diestel, *Graph theory*, Graduate Texts in Mathematics, vol. 173, third edition, Springer-Verlag, Heidelberg, 2005.

[37] M. Elad, P. Milanfar, and R. Rubinstein, *Analysis versus synthesis in signal priors*, Inverse Problems, 23 (2007), pp. 947–968.

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

[39] M. Fadili, J. Starck, and F. Murtagh, *Inpainting and zooming using sparse representations*, The Comput. J., 52 (2007), pp. 64–79.

[40] B. Han, Z. Shen, *Dual wavelet frames and Riesz bases in Sobolev spaces*, Constructive Approximation, 29(3) (2009), pp. 369-406.

[41] H. Ji, Z. Shen, and Y. Xu, *Split Bregman iteration for scene reconstruction*, preprint, 2009.

[42] R. Q. Jia *Subdivision schemes in L^p spaces*, Adv. Comput. Math. 3 (1995), no. 4, 309–341.

[43] R. Q. Jia and C. A. Micchelli, *Using the refinement equations for the construction of pre-wavelets. II. Powers of two*, in Curves and surfaces (Chamonix-Mont-Blanc, 1990), Academic Press, Boston, MA, 1991, pp. 209–246.

[44] R. Q. Jia and Z. W. Shen, *Multiresolution and Wavelets*, Proc. Edinburgh Math. Soc., 37 (1994), 271–300.

[45] M. J. Johnson, Z. Shen, and Y. Xu, *Scattered data reconstruction by regularization in B-spline and associated wavelet spaces*, J. Approx. Theory, 159 (2009), 197–223.

[46] V. Koltchinskii, *Rademacher penalties and structural risk minimization*, IEEE Transactions on Information Theory, 2001(47), 1902–1914.

[47] M. J. Lai and A. Petukhov, *Method of virtual components for constructing redundant filter banks and wavelet frames*, Appl. Comput. Harmon. Anal., 22 (2007), 304–318.

[48] W. Lawton, S. L. Lee, Zuowei Shen, *Convergence of multidimensional cascade algorithm*, Numerische Mathematik, 78 (1998) 3, 427-438.

[49] S. Mallat, *A wavelet tour of signal processing*, Academic Press Inc., San Diego, CA, 1998.

[50] A. Petukhov, *Explicit construction of framelets*, Appl. Comput. Harmon. Anal., 11 (2001), 313–327.

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

[52] H. Rauhut, K. Schnass, and P. Vandergheynst, *Compressed Sensing and Redundant Dictionaries*, IEEE Trans. Inform.

Theory, 54 (2008), pp. 2210–2219.
[53] Z. Shen, *Wavelet Frames and Image Restorations*, Proceedings of the International Congress of Mathematicians, Vol IV, Hyderabad, India, (2010), Hindustan Book Agency, (Rajendra Bhatia eds), 2834-2863.
[54] A. W. van der Vaart and J. A. Wellner, *Weak convergence and empirical processes*, Springer-Verlag, 1996.
[55] V. N. Vapnik, *Statistical learning theory*, Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons Inc., New York, 1998. A Wiley-Interscience Publication.