APPROXIMATION FROM NOISY DATA*

BIN DONG†, ZUOWEI SHEN‡, AND JIANBIN YANG §

Abstract. In most applications, functions are often given by sampled data. Approximation of functions from observed data is often needed. This has been widely studied in the literature when data is exact, and the underlying function is smooth. However, the observed data is often contaminated with noise and the underlying function may be non-smooth (e.g. contains singularities). To properly handle noisy data, any effective approximation scheme must contain a noise removal component. To well approximate non-smooth functions, one needs to have a sparse approximation in, for example, the wavelet domain. Sparsity based noise removal schemes have been proven effective empirically. This paper presents theoretical analysis of such noise removal schemes through the lens of function approximation. For a given sample size, approximation from uniform grid data and scattered data are investigated. The error of the approximation scheme, the bias of the denoising model, and the noise level of data are analyzed, respectively. In addition, when the amount of data is large enough, a new approximation scheme is proposed to grant sufficient reduction on the noise level and ensure asymptotic convergence.

Key words. approximation analysis, wavelet frame, random sampling, noisy data

AMS subject classifications. 42C40, 65D05, 65D10, 65D15

1. Introduction.

For many scientific and engineering problems, such as signal and image processing [25], computer graphics [20] and machine learning [17], data come in large quantities and are often corrupted by noise. Approximation of functions from the data is often needed. When the data is noise-free, and the function is smooth, this has been extensively studied in the literature. However, in many important applications, data is noisy, and the underlying function is non-smooth.

In a typical sampling model, we are given a data set $\Xi = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset \Omega$ and associated function values

$$\boldsymbol{y}_i = (\mathcal{S}_n f)(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i, \qquad \text{for } i = 1, 2, \ldots, n,$$

where Ω is a bounded domain of $\mathbb{R}^d$, $\mathcal{S}_n$ is a sampling operator with $(\mathcal{S}_n f)(\boldsymbol{x}_i)$ being the sampling value of f at $\boldsymbol{x}_i$, and $\boldsymbol{\epsilon}_i$ denotes a sampling noise. By applying some denoising scheme, for example, the wavelet frame based denoising method [5, 6, 20, 29], we obtain a denoised result $\boldsymbol{y}^* = \{\boldsymbol{y}_i^*\}_{i=1}^n$. Then, through some approximation scheme, an approximation function f_n^* can be obtained from $\{(\boldsymbol{x}_i, \boldsymbol{y}_i^*)\}_{i=1}^n$. To evaluate the result of this procedure, we discuss the following two questions:

1. How to understand the denoising and approximation schemes?
2. How to quantify the approximation error $\|f_n^* - f\|$ in terms of a given sample size n, and when n is sufficiently large?

This paper attempts to develop a rigorous analysis of the approximation problem of two types of sampling procedures, uniform grid sampling, and random sampling. Here, we summarize the main results and leave full details of the analysis to the subsequent sections.

1.1. Approximation on uniform grids.

Let $f \in L_2(\Omega)$ with $\Omega = (0,1)^d$ being a unit cube, and $(\mathcal{S}_n f)(2^{-n}\alpha)$ be the sampling of f in uniform grids of Ω with step size 2^{-n}, $n \in \mathbb{N}$. Suppose we are given a sequence of function values,

$$\boldsymbol{y}[\alpha] = (\mathcal{S}_n f)(2^{-n}\alpha) + \boldsymbol{\epsilon}[\alpha], \quad \alpha \in \mathbb{I}_n,$$

where $\boldsymbol{\epsilon}[\alpha]$ denotes a random noise with $\mathcal{E}(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) \le \sigma^2$, and the index set

$$\mathbb{I}_n = \{\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d) \in \mathbb{Z}^d,\ 0 \le \alpha_1, \alpha_2, \ldots, \alpha_d \le 2^n - 1\}.$$

Then, after applying the following denoising scheme

$$\min_{\boldsymbol{u}} \ E(\boldsymbol{u}) = \|\boldsymbol{u} - \boldsymbol{y}\|_{\ell_2}^2 + \|\mathrm{diag}(\lambda)\mathcal{W}\boldsymbol{u}\|_{\ell_1}, \tag{1.1}$$

*Submitted to the editors on December 30, 2020.

Funding: This work was partially supported by the National Science Foundation grant 11771120, 11671022, China, and the Singapore MOE AcRF Research Grant MOE2011-T2-1-116, R-146-000-165-112.

†Beijing International Center for Mathematical Research, Peking University, China (dongbin@math.pku.edu.cn).

‡Department of Mathematics, National University of Singapore, Singapore (matzuows@nus.edu.sg).

§Department of Mathematics, Hohai University, China (jbyang@hhu.edu.cn, corresponding author).

we obtain a denoised result $\boldsymbol{y}^*$ from $\boldsymbol{y}$. Here, $\mathcal{W}$ is the discrete framelet transform and $\mathrm{diag}(\lambda)$ is a diagonal matrix which scales different wavelet channels. In the model (1.1), the first term tries to fit the data. The second term penalizes the roughness of the solution on the one hand and preserves discontinuity features of signals on the other hand [6, 13, 20, 29]. Since in real-world applications many signals have a sparse approximation in the wavelet domain, this model finds various applications, such as in signal and image processing (see e.g., [5, 15]).

Let $\mathcal{A}_n : \ell_2(\mathbb{I}_n) \to L_2(\Omega)$ be an approximation scheme, and $f_n^* = \mathcal{A}_n \boldsymbol{y}^*$ be the function approximated from $\boldsymbol{y}^*$. Then, the approximation error

$$\begin{aligned} \|f_n^* - f\| &= \|\mathcal{A}_n \boldsymbol{y}^* - f\| \\ &\leq \|\mathcal{A}_n(\mathcal{S}_n f) - f\| + \|\mathcal{A}_n\| \|\boldsymbol{y}^* - (\mathcal{S}_n f)\|. \end{aligned} \tag{1.2}$$

The first term of (1.2) depends on the approximation result from noise-free data, while the second term depends on the denoising result and approximation operator. Assume that the wavelets in model (1.1) have enough vanishing moments and the wavelet coefficients of f satisfy certain decay conditions, for example for $|i| \geq 1$,

$$\sum_{n \in \mathbb{N}} \sum_{\alpha \in \mathbb{I}_n} 2^{(2s-d)n} |\mathcal{W}_i \mathbf{f}[\alpha]|^2 \leq \mathrm{C},$$

then we can find an approximation scheme $\mathcal{A}_n$ such that for any $\mu > 0$, with probability at least $1 - \frac{1}{\mu}$, the approximation error

$$\|f_n^* - f\| \leq \mathrm{C}\big((2^{-n})^s + \lambda^{1/2}(2^{-n})^{\frac{s}{2}} + \mu^{1/2}\sigma\big),$$

which depends on the approximation ability of $\mathcal{A}_n$, the bias of the denoising model and the noise level. We shall give a detailed analysis of this approach in section 2, and show how to choose an approximation scheme and a way to reduce the noise level in order to achieve convergence.

1.2. Approximation on randomly sampled data. In recent years, with the rapid development of machine learning (especially deep learning [17]), function approximation on randomly sampled data is more often seen in practice. The second part of this paper is to investigate the approximation from scattered data $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n$. Here, $\{\boldsymbol{x}_i\}_{i=1}^n$ is uniformly randomly drawn from $\Omega \subset \mathbb{R}^d$, $\boldsymbol{y}_i = (\mathcal{S}_n f)(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i$ is the sampling value, and $\boldsymbol{\epsilon}_i$ is the random noise. Similar as the first part, we choose a denoising scheme and obtain an approximant f_n^* through some approximation scheme. We focus on estimating the approximation error $\|f_n^* - f\|$ in terms of n and the noise level.

Interpolation and quasi-interpolation schemes based on the noise-free data have been extensively studied in the literature (see [12, 16, 18, 21] and the references therein). For scattered data, many approximation schemes through properly defined functions spaces have been proposed and studied in the literature, including the reproducing kernel Hilbert/Banach spaces [33], spline subspaces [24, 31], radial basis functions [32], bandlimited functions [37], finite element spaces [8, 9] and the shift-invariant spaces [11, 21, 34].

The approximation schemes for functions on scattered data are usually nonlinear, and most approaches determine f_n^* by solving an optimization problem on a prescribed function space [23, 31]

$$\min_{g \in V} \frac{1}{n} \sum_{i=1}^n (g(\boldsymbol{x}_i) - \boldsymbol{y}_i)^2 + \Gamma(g),$$

where the first term tries to fit $f_n^*(\boldsymbol{x}_i)$ to $\boldsymbol{y}_i$, and the second term describes a prior knowledge (or regularization) on the approximant f_n^*.

We choose the principal shift invariant system and its dilations, i.e.,

$$S^h(\varphi, \Omega) = \{\sum_{\alpha \in I} \boldsymbol{u}[\alpha] \varphi(\frac{\cdot}{h} - \alpha) : \boldsymbol{u}[\alpha] \in \mathbb{R}\}.$$

Besides its structural simplicity, $S^h(\varphi, \Omega)$ has the advantage that for special choices of φ, such as B-spline, it provides sparse system and good approximation orders to smooth functions [11, 24].

Consider the following optimization model

$$\min_{\boldsymbol{u}} \sum_{i=1}^{n} w_i(\sum_{\alpha\in I} \boldsymbol{u}[\alpha]\varphi(\frac{\boldsymbol{x}_i}{h}-\alpha)-\boldsymbol{y}_i)^2+\rho\|\mathrm{diag}(\lambda)\mathcal{W}\boldsymbol{u}\|_{\ell_1}\ , \tag{1.3}$$

where $\{\boldsymbol{u}[\alpha]\}_{\alpha\in I}$ are the coefficients which we want to solve, h is the scaling parameter, and w_i is the weight to balance the penalties of different $\boldsymbol{u}[\alpha]$ according to the density of sampling points in the support of $\varphi(\frac{\cdot}{h}-\alpha)$. This leads to an approximation function $f_n^*\in S^h(\varphi,\Omega)$:

$$f_n^*=\sum_{\alpha\in I}\boldsymbol{u}^*[\alpha]\varphi(\frac{\cdot}{h}-\alpha)$$

with $\boldsymbol{u}^*$ being the minimizer of (1.3).

For properly chosen parameters, it can be shown that for any $\mu>0$, the following inequality

$$\begin{aligned}&\|f_n^*-f\|_{L_2(\Omega)}\\ &\le \mathrm{C}\Big(n^{-\frac{(1-\gamma_1)}{d}(k-\frac{d}{2})}|f|_{W_1^k(\Omega)}+\sqrt{\rho}|f|^{\frac12}_{W_1^k(\Omega)}+n^{-\frac{3(1-\gamma_1)(2k-d)}{2d}}\rho^{-1}|f|^2_{W_1^k(\Omega)}\\ &+\mu^{1/2}\sigma+n^{\frac{-(1-\gamma_1)(2k-d)}{2d}}\rho^{-1}\mu\sigma^2\Big)\end{aligned}$$

holds with probability at least

$$(1-\frac{1}{\mu})\Big(1-n^{1-\gamma_1}\exp(-\frac{(n^{\gamma_1}-2)}{2})\Big),$$

where $0<\gamma_1<1$. We shall discuss this in full detail in section 3, and analyze $\|f_n^*-f\|_{L_2(\Omega)}$ in terms of the approximation ability of $S^h(\varphi,\Omega)$, the bias induced by the regularization, and the noise level. Furthermore, we consider the case when n is sufficiently large, how to choose the approximation scheme such that the noise level can be reduced and the convergence is guaranteed.

Note that the regularized least squares models are frequently used to fit noisy data and avoid overfitting. Most of the previous methods mainly impose regularity conditions directly on the functions to be approximated. The regularization is often chosen as the Sobolev semi-norm which is discretized by numerical integration methods [10, 23, 24, 31, 32]. In contrast, the regularization of model (1.3) is imposed on the discrete coefficients of wavelet transform which is able to preserve discontinuities of the functions to be approximated. Moreover, noting that basis functions with large supports can fit the scattered data while they fail to represent the details of the functions. On the other hand, basis functions with small supports can detect the details of the functions to be approximated, whereas the approximant may fluctuate in areas with fewer sampled data. Therefore, multiresolution wavelet frames are often preferred to represent these functions, and the redundancy of the system offers more resilience to noise [36].

Similar denoising scheme was considered in [35] in which the data density and accumulation level of sampling set was given and an asymptotic approximation analysis of (1.3) was discussed in the case $w_i\equiv 1$. The model (1.3) was used to approximate range data which is known to contain discontinuities [20], and recently was applied to fit coarse-grained force functions in structural biology [36].

1.3. Organization of the paper.

The remaining part of this paper is organized as follows. In section 2 we first present the necessary notation and review some basic properties of wavelet frames. Then, we consider the approximation from data on uniform grids and analyze the convergence of the solution. In section 3 we investigate the approximation from randomly sampled data, and characterize the approximation error in terms of the sample size and the noise level.

2. Approximation on Uniform Grids.

2.1. Notation and preliminaries.

Let $\mathbb{N}$ denote the set of nonnegative integers and $B(r)=\{|x|<r, x\in\mathbb{R}^d\}$. Let $\#S$ denote the cardinality of a finite set S and $|E|$ denote the Lebesgue measure of a measurable set $E\subset\mathbb{R}^d$.

For a compactly supported function $\varphi \in L_2(\mathbb{R}^d)$, the shift invariant space $S(\varphi)$ generated by φ is defined as

$$S(\varphi) := closure\{\varphi *' \boldsymbol{a} : \boldsymbol{a} \in \ell_0(\mathbb{Z}^d)\},$$

where

$$\varphi *' \boldsymbol{a} := \sum_{\alpha \in \mathbb{Z}^d} \boldsymbol{a}[\alpha]\varphi(\cdot - \alpha)$$

and $\ell_0(\mathbb{Z}^d)$ denotes the set of all finitely supported sequences in $\mathbb{Z}^d$. The shifts of φ are called stable if there exist two positive constants C_1 and C_2 such that for all sequences $\boldsymbol{a} \in \ell_2(\mathbb{Z}^d)$,

$$\mathrm{C}_1\|\boldsymbol{a}\|_{\ell_2} \le \|\sum_{\alpha \in \mathbb{Z}^d} \boldsymbol{a}[\alpha]\varphi(\cdot - \alpha)\|_{L_2} \le \mathrm{C}_2\|\boldsymbol{a}\|_{\ell_2}. \tag{2.1}$$

The Sobolev space $W_1^k(\mathbb{R}^d)$ is the set of all distributions f such that $D^\mu f \in L_1(\mathbb{R}^d)$ for all $|\mu| \le k$, and the Sobolev semi-norm is defined as $|f|_{W_1^k} = \sum_{|\mu|=k} \|D^\mu f\|_{L_1}$. A function f is said to satisfy the Strang-Fix conditions [30] of order k if

$$\hat{f}(0) \neq 0 \quad \text{and} \quad D^\mu \hat{f}(2\pi\alpha) = 0, \quad \forall \alpha \in \mathbb{Z}^d \backslash \{0\},\ |\mu| < k.$$

A wavelet system $X(\Psi)$ is defined to be a collection of dilations and shifts of a finite set of functions $\Psi = \{\psi_1, \dots, \psi_m\} \subset L_2(\mathbb{R})$, where

$$X(\Psi) := \{\psi_{\ell;j,k} = 2^{j/2}\psi_\ell(2^j \cdot - k),\ \ell = 1, \dots, m; j, k \in \mathbb{Z}\}.$$

When the set of functions $X(\Psi)$ forms a tight frame of $L_2(\mathbb{R})$, i.e.,

$$\|f\|^2_{L_2(\mathbb{R})} = \sum_{\ell=1}^{m} \sum_{j,k \in \mathbb{Z}} |\langle f, \psi_{\ell;j,k}\rangle|^2,$$

it is called a wavelet tight frame. To construct a wavelet system, one usually starts with a refinable function ϕ satisfying

$$\phi(x) = 2\sum_{\alpha \in \mathbb{Z}} \mathbf{h}_0[\alpha]\phi(2x - \alpha),$$

where $\mathbf{h}_0 \in \ell_0(\mathbb{Z})$ is called a refinement mask. Then the construction of a wavelet frame system is to find the set of framelets $\Psi = \{\psi_1, \dots, \psi_m\}$ defined by

$$\psi_\ell(x) = 2\sum_{\alpha \in \mathbb{Z}} \mathbf{h}_\ell[\alpha]\phi(2x - \alpha), \qquad \ell = 1, \dots, m. \tag{2.2}$$

Let B_m be the B-spline of order m, which in the frequency domain is given by

$$\widehat{B_m}(\xi) = e^{-ij_m\frac{\xi}{2}} \frac{\sin^m(\frac{\xi}{2})}{(\frac{\xi}{2})^m}, \tag{2.3}$$

where

$$j_m = \begin{cases} 1, & m \text{ is odd}, \\ 0, & m \text{ is even}. \end{cases} \tag{2.4}$$

It is easy to check that B_m is refinable with refinement mask

$$\widehat{\mathbf{h}}_0(\xi) = e^{-ij_m\frac{\xi}{2}} \cos^m(\frac{\xi}{2}). \tag{2.5}$$

By B_m and $\mathbf{h}_0$, a family of wavelet tight frame can be derived by the Unitary Extension Principle (UEP) [27]. Let m framelet masks be given by

$$\widehat{\mathbf{h}}_\ell(\xi) := -i^\ell e^{-ij_m\frac{\xi}{2}} \sqrt{\binom{m}{\ell}} \sin^\ell(\frac{\xi}{2}) \cos^{m-\ell}(\frac{\xi}{2}), \quad \ell = 1, 2, \dots, m, \tag{2.6}$$

then $X(\Psi)$ forms a tight frame with $\mathbf{h}_\ell$ in (2.2) defined above.

By the $m+1$ filters $\{\mathbf{h}_\ell\}_{\ell=0,1,\ldots,m}$, we can define the discrete wavelet frame transform on $\ell_1(\mathbb{Z}^d)$ by tensor product. For index $i=(i_1,i_2,\ldots,i_d)$ with $0\le i_1,i_2,\ldots,i_d\le m$, the wavelet filters $(\mathbf{h}_i[k])_{k\in\mathbb{Z}^d}$ are defined as

$$\mathbf{h}_i[k] := \mathbf{h}_{i_1}[k_1]\mathbf{h}_{i_2}[k_2]\ldots\mathbf{h}_{i_d}[k_d],$$

where i_r denotes the i_r-th vanishing moment of $\mathbf{h}_{i_r}$ corresponding to the r-th variable and $k=(k_1,k_2,\ldots,k_d)\in\mathbb{Z}^d$. For the $(\ell+1)$-th level of undecimated wavelet frame transform, the filters are given by $\mathbf{h}_{\ell,i} := \tilde{\mathbf{h}}_{\ell,i} * \tilde{\mathbf{h}}_{\ell-1,i} * \ldots * \tilde{\mathbf{h}}_{0,i}$, where

$$\tilde{\mathbf{h}}_{\ell,i}[k] = \begin{cases} \mathbf{h}_i[2^{-\ell}k], & k\in 2^\ell\mathbb{Z}^d,\\ 0, & k\notin 2^\ell\mathbb{Z}^d.\end{cases}$$

Let $\boldsymbol{u}\in\ell_1(\mathbb{Z}^d)$, the 1-th level of wavelet frame decomposition is defined as

$$\mathcal{W}_i\boldsymbol{u} = \mathbf{h}_i[-\cdot] * \boldsymbol{u} \qquad \text{for} \quad i=(i_1,i_2,\ldots,i_d).$$

In general, we denote $\mathcal{W}_{\ell,i}\boldsymbol{u} = \mathbf{h}_{\ell,i}[-\cdot] * \boldsymbol{u}$ and the wavelet frame decomposition with L levels as

$$\mathcal{W}\boldsymbol{u} = \{\mathcal{W}_{\ell,i}\boldsymbol{u} : 0\le\ell\le L-1, 0\le i_1,i_2,\ldots,i_d\le m\}.$$

For simplicity, in this paper we choose $L=1$, and the analysis can be extended to general cases with $L>1$. See [11,14,27] for more details on wavelet frames.

2.2. Error analysis.

2.2.1. Approximation error analysis. Let $f\in L_2(\Omega)$ with $\Omega=(0,1)^d$. Suppose that the discrete observation of f at Euclidean point $2^{-n}\alpha$ is given by

$$\boldsymbol{y}[\alpha] = (\mathcal{S}_n f)(2^{-n}\alpha) + \boldsymbol{\epsilon}[\alpha], \quad \alpha\in\mathbb{I}_n, \tag{2.7}$$

where 2^{-n} represents the step size of the uniform grid for some $n\in\mathbb{N}$, $\boldsymbol{\epsilon}[\alpha]$ is a random noise, and $\mathcal{S}_n : L_2(\Omega)\to\ell_2(\mathbb{I}_n)$ is a sampling operator defined as

$$(\mathcal{S}_n f)(2^{-n}\alpha) = 2^{dn}\langle f, \phi(2^n\cdot-\alpha)\rangle, \quad \alpha\in\mathbb{Z}^d. \tag{2.8}$$

Here, $\phi\in L_2(\mathbb{R}^d)$ is a refinable function with compact support satisfying $\int_{\mathbb{R}^d}\phi(x)dx=1$, and the index set

$$\mathbb{I}_n := \{\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_d)\in\mathbb{Z}^d,\ 0\le\alpha_1,\alpha_2,\ldots,\alpha_d\le 2^n-1\}.$$

If the sampling process is assumed noiseless, i.e., $\boldsymbol{\epsilon}\equiv 0$, f can be well approximated by the discrete observations. Let $\mathbf{f}[\alpha]=(\mathcal{S}_n f)(2^{-n}\alpha)$. Then we can define an approximation scheme $\mathcal{A}_n : \ell_2(\mathbb{I}_n)\to L_2(\Omega)$ as follows

$$\mathcal{A}_n\mathbf{f} := \sum_{\alpha\in\mathbb{I}_n}\mathbf{f}[\alpha]\varphi(2^n\cdot-\alpha), \tag{2.9}$$

where φ is a compactly supported function in $L_2(\mathbb{R}^d)$.

In the following, we show that if the wavelet coefficients of $\mathbf{f}$ satisfy some mild decay condition, $\mathcal{A}_n\mathbf{f}$ converges to f with some approximation order when φ is properly chosen. Let $X(\Psi)$ be a wavelet system generated by UEP with masks $\{\mathbf{h}_\ell\}$, $\ell=1,\ldots,m$. For $\tau>0$, we say that the filter $\mathbf{h}_\ell$ has τ vanishing moments if the following condition holds

$$\widehat{\mathbf{h}_\ell}(\xi) = O(\|\xi\|^\tau), \quad \text{as } \xi\to 0. \tag{2.10}$$

Let $\tilde{\mathbb{I}}_n = \{\alpha\in\mathbb{I}_n : \text{support of } \psi_{i;n,\alpha}\subset\overline{\Omega}, \forall |i|\ge 1\}$. For any $\epsilon>0$, let

$$\Omega_\epsilon = \{x\in\Omega : dist(x,\partial\Omega)>\epsilon\}.$$

PROPOSITION 2.1. *Let* $\mathbf{f}[\alpha] = (\mathcal{S}_n f)(2^{-n}\alpha)$, $\alpha \in \mathbb{I}_n$, *be the discrete sampling of* f, *and* $\mathcal{A}_n$ *be the approximation scheme given by* (2.9). *Let* $X(\Psi)$ *be a wavelet system satisfying the conditions of vanishing moments of order* τ. *Suppose that there exists* $s > 0$ *such that for* $|i| \geq 1$,

$$\sum_{n\in\mathbb{N}} \sum_{\alpha\in\tilde{\mathbb{I}}_n} 2^{(2s-d)n} |\mathcal{W}_i \mathbf{f}[\alpha]|^2 \leq \mathrm{C}, \tag{2.11}$$

where C *is a positive constant. Then for any* $\epsilon > 0$ *and* $0 < \zeta \leq \min\{s, \tau\}$, *we have*

$$\|\mathcal{A}_n \mathbf{f} - f\|_{L_2(\Omega_\epsilon)} \leq \mathrm{C}(2^{-n})^{\min\{\zeta, r\}}$$

provided that φ *is chosen satisfying the following conditions:*

$$\sum_{\alpha\in\mathbb{Z}^d\setminus\{0\}} |\hat{\varphi}(\xi + 2\pi\alpha)|^2 = O(\|\xi\|^{2r}), \tag{2.12}$$

and

$$1 - \overline{\hat{\varphi}}(\xi)\hat{\phi}(\xi) = O(\|\xi\|^r) \tag{2.13}$$

for some $r > 0$.

Proof. Since $\widehat{\psi}_i = \widehat{\mathbf{h}}_i(\frac{\cdot}{2})\widehat{\phi}(\frac{\cdot}{2})$, we obtain that

$$\begin{aligned}
&2^{2sn}\big(2^{-dn} \sum_{\alpha\in\tilde{\mathbb{I}}_n} |\mathcal{W}_i \mathbf{f}[\alpha]|^2\big) \\
&= \sum_{\alpha\in\tilde{\mathbb{I}}_n} 2^{2sn} 2^{-dn} |\mathcal{W}_i (\mathcal{S}_n f)[\alpha]|^2 \\
&= 2^{2sn}\big(\sum_{\alpha\in\tilde{\mathbb{I}}_n} |2^{-dn/2} \mathcal{W}_i \langle f(2^{-n}\cdot), \phi(\cdot - \alpha)\rangle|^2\big) \\
&= \sum_{\alpha\in\tilde{\mathbb{I}}_n} 2^{2sn} |\langle f, \psi_{i,n-1,\alpha}\rangle|^2.
\end{aligned}$$

By (2.11), we have

$$\sum_{n\in\mathbb{N}} \sum_{\alpha\in\tilde{\mathbb{I}}_n} 2^{2sn} |\langle f, \psi_{i,n,\alpha}\rangle|^2 = 2^{-2s} \sum_{n\geq 1} \sum_{\alpha\in\tilde{\mathbb{I}}_n} 2^{2sn} |\langle f, \psi_{i,n-1,\alpha}\rangle|^2 \leq \mathrm{C}.$$

Moreover, by the conditions of vanishing moments of ψ_i, we have $f|_{\Omega_\epsilon} \in H^\zeta(\Omega_\epsilon)$ (see e.g. [18, 25]). Therefore, by virtue of (2.12), (2.13) and [11, 21], we have

$$\|\mathcal{A}_n \mathbf{f} - f\|_{L_2(\Omega_\epsilon)} \leq \mathrm{C}(2^{-n})^{\min\{\zeta, r\}}.$$

This concludes the proof. ∎

If ϕ and φ satisfy the properties of (2.12) and (2.13) for some $r > 0$, we have $\lim_{n\to\infty} \|\mathcal{A}_n \mathbf{f} - f\|_{L_2(\Omega)} = 0$ for any $f \in L_2(\Omega)$ ([5, Lemma 4.1]). It is well known that the smoothness of functions can be characterized by the decay of their wavelet coefficients [4,10,18,25,35], and in many applications such as signal and image processing, small ζ is preferred to reflect the low regularity of these functions [26,28]. In the case $\zeta \geq 1$, wavelets with high order vanishing moments should be applied to characterize the conditions (2.10) and (2.11). Moreover, if we know $f \in H^\zeta(\mathbb{R}^d)$ in advance, and it is sampled on the uniform grids in $\mathbb{R}^d$, then $\mathcal{A}_n(\mathcal{S}_n)$ is the quasi-projection operator (see [11, 12, 21]) defined as

$$\mathcal{A}_n(\mathcal{S}_n f) = 2^{dn} \sum_{\alpha\in\mathbb{Z}^d} \langle f, \phi(2^n \cdot - \alpha)\rangle \varphi(2^n \cdot - \alpha).$$

In this case, we can obtain the same approximation result as in the previous proposition.

Note that φ is a compact supported function with $\hat{\varphi}(0) \neq 0$. Then, the condition (2.12) of φ is equivalent to the Strang-Fix conditions of order r. If ϕ is chosen as B-spline of order m, we have $|\hat{\phi}(\xi)| = \frac{\sin^m(\frac{\xi}{2})}{(\frac{\xi}{2})^m}$, and $\phi \in H^s(\mathbb{R})$ for any $s < m - \frac{1}{2}$. In addition, if φ is also chosen as B-spline of order m, we have $\sum_{\alpha \in \mathbb{Z}^d \setminus \{0\}} |\hat{\varphi}(\xi + 2\pi\alpha)|^2 = O(\|\xi\|^{2m})$, and $1 - \overline{\hat{\varphi}}(\xi)\hat{\phi}(\xi) = O(\|\xi\|^2)$ for $m \geq 2$.

The sampling scheme (2.8) has been discussed in [5, 7, 25, 35]. In particular, if f is a smooth function and ϕ is chosen as the Dirac delta function, $(\mathcal{S}_n f)(\alpha) = f(2^{-n}\alpha)$. In this case, $\hat{\phi}(\xi) \equiv 1$ and φ can be explicitly constructed such that (2.12) and (2.13) hold for any $r > 0$, and the approximation order of $\mathcal{A}_n(\mathcal{S}_n f)$ to f depends on the smoothness of f and r [11]. The consistency of various sampling methods is analyzed in [13, Lemma 6.1].

2.2.2. Statistical error analysis. Due to physical sampling processes and system errors, even with high-precision devices, the acquired data inevitably contains noise. Thus any effective approximate scheme should contain a noise removal component.

Let $\boldsymbol{y}$ be the discrete sampling of f, and $\mathcal{D}_n$ be the denoising scheme given by the following model

$$\min_{\boldsymbol{u}} \; E(\boldsymbol{u}) = \|\boldsymbol{u} - \boldsymbol{y}\|^2_{\ell_2(\mathbb{I}_n)} + \|\mathrm{diag}(\lambda)\mathcal{W}\boldsymbol{u}\|_{\ell_1(\tilde{\mathbb{I}}_n)}, \tag{2.14}$$

where $\mathcal{W}$ is the discrete framelet transform, and the parameter $\mathrm{diag}(\lambda)$ is a diagonal matrix based on the vector $\lambda = [\lambda_0, \lambda_1, \ldots]$ which scales different wavelet channels. We then obtain a denoised result from $\boldsymbol{y}$, i.e., $\mathcal{D}_n\boldsymbol{y} = \boldsymbol{y}^*$, where $\boldsymbol{y}^*$ is the optimal solution of (2.14).

PROPOSITION 2.2. *Let $\boldsymbol{y} = (\mathcal{S}_n f) + \boldsymbol{\epsilon}$ be the noisy observations of f given by (2.7). Assume that the random noise $\boldsymbol{\epsilon}$ are independent with $\mathcal{E}(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) \leq \sigma^2$. Let $\boldsymbol{y}^* = \mathcal{D}_n\boldsymbol{y}$ be the denoised result obtained by (2.14). Taking expectation w.r.t. the random variable $\boldsymbol{\epsilon}$, we obtain*

$$\mathcal{E}\big(2^{-\frac{dn}{2}}\|\boldsymbol{y}^* - (\mathcal{S}_n f)\|_{\ell_2(\mathbb{I}_n)}\big) \leq 2\Big(2^{-\frac{dn}{2}}\|\mathrm{diag}(\lambda)\mathcal{W}(\mathcal{S}_n f)\|^{1/2}_{\ell_1(\tilde{\mathbb{I}}_n)} + \sigma\Big). \tag{2.15}$$

Proof. Since $\boldsymbol{y}^*$ is the minimizer of (2.14), the following applies

$$\begin{aligned}
\|\boldsymbol{y}^* - (\mathcal{S}_n f)\|^2_{\ell_2(\mathbb{I}_n)} &\leq 2(\|\boldsymbol{y}^* - \boldsymbol{y}\|^2_{\ell_2(\mathbb{I}_n)} + \|\boldsymbol{\epsilon}\|^2_{\ell_2(\mathbb{I}_n)}) \\
&\leq 2(\|(\mathcal{S}_n f) - \boldsymbol{y}\|^2_{\ell_2(\mathbb{I}_n)} + \|\mathrm{diag}(\lambda)\mathcal{W}(\mathcal{S}_n f)\|_{\ell_1(\tilde{\mathbb{I}}_n)} + \|\boldsymbol{\epsilon}\|^2_{\ell_2(\mathbb{I}_n)}) \\
&\leq 2\|\mathrm{diag}(\lambda)\mathcal{W}(\mathcal{S}_n f)\|_{\ell_1(\tilde{\mathbb{I}}_n)} + 4\|\boldsymbol{\epsilon}\|^2_{\ell_2(\mathbb{I}_n)}.
\end{aligned}$$

By Jensen's inequality [3] and the independence of $\boldsymbol{\epsilon}[\alpha]$, $\alpha \in \mathbb{I}_n$, we have

$$\mathcal{E}(2^{-\frac{dn}{2}}\|\boldsymbol{\epsilon}\|_{\ell_2(\mathbb{I}_n)}) \leq \big(\mathcal{E}(2^{-dn}\sum_{\alpha \in \mathbb{I}_n}|\boldsymbol{\epsilon}[\alpha]|^2)\big)^{1/2} \leq \sigma.$$

Thus, we conclude that (2.15) holds. ∎

2.3. Approximation from data on uniform grids. Let $\boldsymbol{y}$ be the noisy observations of f. By the denoising scheme $\mathcal{D}_n$ (2.14) and the approximation scheme $\mathcal{A}_n$ (2.9), we can obtain an approximation function

$$f_n^* = \mathcal{A}_n\boldsymbol{y}^* = \mathcal{A}_n(\mathcal{D}_n\boldsymbol{y}).$$

Since φ is a compactly supported function in $L_2(\mathbb{R}^d)$, by [22, Theorem 2.1], for all sequences $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \ell_2(\mathbb{I}_n)$, there exists a positive constant C independent of n such that

$$\|\mathcal{A}_n\boldsymbol{y}_1 - \mathcal{A}_n\boldsymbol{y}_2\|_{L_2(\Omega)} \leq \mathrm{C}2^{-\frac{dn}{2}}\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_{\ell_2(\mathbb{I}_n)}. \tag{2.16}$$

Thus, we have

$$\begin{aligned}
\|f_n^* - f\|_{L_2(\Omega)} &= \|\mathcal{A}_n\boldsymbol{y}^* - f\|_{L_2(\Omega)} \\
&\leq \|\mathcal{A}_n(\mathcal{S}_n f) - f\|_{L_2(\Omega)} + \|\mathcal{A}_n\boldsymbol{y}^* - \mathcal{A}_n(\mathcal{S}_n f)\|_{L_2(\Omega)} \\
&\leq \|\mathcal{A}_n(\mathcal{S}_n f) - f\|_{L_2(\Omega)} + \mathrm{C}2^{-\frac{dn}{2}}\|\boldsymbol{y}^* - (\mathcal{S}_n f)\|_{\ell_2(\mathbb{I}_n)}.
\end{aligned}$$

It follows that the approximation error depends on the properties of approximation scheme in section 2.2.1 and the denoising result in section 2.2.2.

THEOREM 2.3. *Let $\boldsymbol{y} = (\mathcal{S}_n f) + \boldsymbol{\epsilon}$ be the noisy observations of f given by (2.7). Suppose that the random noise $\boldsymbol{\epsilon}$ are independent with $\mathcal{E}(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) \leq \sigma^2$. Let the wavelet system $X(\psi)$ satisfy the vanishing moments conditions of order τ. In addition, we assume that there exists $s > 0$ such that for $|i| \geq 1$,*

$$\sum_{n\in\mathbb{N}}\sum_{\alpha\in\tilde{\mathbb{I}}_n} 2^{(2s-d)n}|\mathcal{W}_i(\mathcal{S}_n f)[\alpha]|^2 \leq \mathrm{C}.$$

Let $\boldsymbol{y}^* = \mathcal{D}_n\boldsymbol{y}$ be the denoised result obtained by (2.14) with $\lambda_0 = 0$ and $0 < \lambda_i \leq 2^{(s-s_0)n}$ for some $0 < s_0 \leq 2s$ and all $|i| \geq 1$. Let $f_n^* = \mathcal{A}_n\boldsymbol{y}^*$ be the approximation function. Then for any $\epsilon > 0$ and $0 < \zeta \leq \min\{s,\tau\}$,

(i) $\mathcal{E}(\|f_n^* - f\|_{L_2(\Omega_\epsilon)}) \leq \mathrm{C}\big((2^{-n})^{\min\{\zeta,r\}} + (2^{-n})^{\frac{s_0}{2}} + \sigma\big)$,

and

(ii) for any $\mu > 0$, with probability at least $1 - \frac{1}{\mu}$,

$$\|f_n^* - f\|_{L_2(\Omega_\epsilon)} \leq \mathrm{C}\big((2^{-n})^{\min\{\zeta,r\}} + (2^{-n})^{\frac{s_0}{2}} + \mu^{1/2}\sigma\big),$$

provided that ϕ and φ satisfy the conditions of order r in (2.12) and (2.13). Here, C is a positive constant independent of n.

Proof. By Proposition 2.1 and (2.16), we have

$$\begin{aligned}\|f_n^* - f\|_{L_2(\Omega_\epsilon)} &\leq \|\mathcal{A}_n(\mathcal{S}_n f) - f\|_{L_2(\Omega_\epsilon)} + \|\mathcal{A}_n\mathbf{y}^* - \mathcal{A}_n(\mathcal{S}_n f)\|_{L_2(\Omega_\epsilon)}\\ &\leq \mathrm{C}_1(2^{-n})^{\min\{\zeta,r\}} + \mathrm{C}_2 2^{-\frac{dn}{2}}\|\boldsymbol{y}^* - (\mathcal{S}_n f)\|_{\ell_2(\mathbb{I}_n)}.\end{aligned}$$

Moreover, by Proposition 2.2, we have

$$\|f_n^* - f\|_{L_2(\Omega_\epsilon)} \leq \mathrm{C}_3\big((2^{-n})^{\min\{\zeta,r\}} + 2^{-\frac{dn}{2}}\|\mathrm{diag}(\lambda)\mathcal{W}(\mathcal{S}_n f)\|^{1/2}_{\ell_1(\tilde{\mathbb{I}}_n)} + 2^{-\frac{dn}{2}}\|\boldsymbol{\epsilon}\|_{\ell_2(\mathbb{I}_n)}\big).$$

By the Cauchy-Schwarz inequality, for $0 < s_0 \leq 2s$ and $|i| \geq 1$,

$$\begin{aligned}&2^{-dn}\big(\lambda_i \sum_{\alpha\in\tilde{\mathbb{I}}_n}|\mathcal{W}_i(\mathcal{S}_n f)[\alpha]|\big)\\ &\leq 2^{(s-s_0)n}\big(2^{-dn}\sum_{\alpha\in\tilde{\mathbb{I}}_n}|\mathcal{W}_i(\mathcal{S}_n f)[\alpha]|\big)\\ &\leq 2^{(s-s_0-\frac{d}{2})n}\big(\sum_{\alpha\in\tilde{\mathbb{I}}_n}|\mathcal{W}_i(\mathcal{S}_n f)[\alpha]|^2\big)^{\frac{1}{2}}\\ &\leq \mathrm{C}_4 2^{-s_0 n}.\end{aligned}$$

Therefore, we conculde

$$\|f_n^* - f\|_{L_2(\Omega_\epsilon)} \leq \mathrm{C}_5\big((2^{-n})^{\min\{\zeta,r\}} + (2^{-n})^{\frac{s_0}{2}} + 2^{-\frac{dn}{2}}\|\boldsymbol{\epsilon}\|_{\ell_2(\mathbb{I}_n)}\big).$$

By Jensen's inequality [3] and the independence of $\boldsymbol{\epsilon}[\alpha]$, $\alpha \in \mathbb{I}_n$, we obtain

$$\mathcal{E}(2^{-\frac{dn}{2}}\|\boldsymbol{\epsilon}\|_{\ell_2(\mathbb{I}_n)}) \leq \big(\mathcal{E}(2^{-dn}\sum_{\alpha\in\mathbb{I}_n}|\boldsymbol{\epsilon}[\alpha]|^2)\big)^{1/2} \leq \sigma.$$

In addition, by Markov's inequality, for any $\mu > 0$,

$$\mathcal{P}\big((2^{-dn}\sum_{\alpha\in\mathbb{I}_n}|\boldsymbol{\epsilon}[\alpha]|^2) > \mu\big) \leq \frac{\sigma^2}{\mu}.$$

This completes the proof of the theorem. ∎

Theorem 2.3 shows that the approximation error is completely determined by the approximation scheme, the regularization of the denoising model (bias) and the noise level (variance). In particular, if the wavelets have enough vanishing moments and the approximation scheme $\mathcal{A}_n$ is chosen to satisfy the conditions in (2.12) and (2.13) with $r \geq s$, then Theorem 2.3 implies that for any $\mu > 0$, with probability at least $1 - \frac{1}{\mu}$,

$$\|f_n^* - f\| \leq \mathrm{C}\big((2^{-n})^s + \lambda^{1/2}(2^{-n})^{\frac{s}{2}} + \mu^{1/2}\sigma\big).$$

Under the assumption that $2^{-\frac{dn}{2}}\|\boldsymbol{\epsilon}\|_{\ell_2(\mathbb{I}_n)} \to 0$, a similar result was established in [35, Proposition 3.1]. However, if the sampling noise $\boldsymbol{\epsilon}$ are independently and identically distributed (i.i.d.) with $\mathcal{E}(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) = \sigma^2$, then

$$\mathcal{E}(\boldsymbol{\epsilon}^2) = \sigma^2 \quad \text{and} \quad \mathcal{E}(2^{-dn}\|\boldsymbol{\epsilon}\|^2_{\ell_2(\mathbb{I}_n)}) = \sigma^2.$$

Thus, in general $2^{-\frac{dn}{2}}\|\boldsymbol{\epsilon}\|_{\ell_2(\mathbb{I}_n)}$ can not be neglected.

In the following subsection, we consider the case when n is sufficiently large, how to choose an approximation scheme to reduce noise level to ensure convergence. The main idea is that when the data is dense enough, the observed values in high resolution grids can be filtered to generate "new" sampling values. Compared to the original observations, these values are on a coarser grid. Nevertheless, the noise level is decreased. Then, we may approximate f from these weighted values with reduced noise. In this way, although the approximation speed becomes slower, the noise level is decreased, and finally $\|f_n - f\|_{L_2(\Omega)}$ converges to 0 instead of a positive constant.

2.4. Reducing the noise level. Let $\boldsymbol{y}[\alpha] = (\mathcal{S}_n f)(2^{-n}\alpha) + \boldsymbol{\epsilon}[\alpha]$, $\alpha \in \mathbb{I}_n$, be the noisy observations of f on a fine grid $2^{-n}\mathbb{I}_n$ given by (2.7). Here, $(2^{-n}\alpha)$ denotes the Euclidean coordinate of the sampling point, and $[\alpha]$ denotes the index of the sequence.

We can choose $n_1 \in \mathbb{N}$ (e.g. $n_1 = \lfloor \ln(n) \rfloor$) such that

$$\lim_{n\to\infty} n_1 = +\infty \quad \text{and} \quad \lim_{n\to\infty} n - n_1 = +\infty.$$

For any given n and $n_1 \in \mathbb{N}$, let $\boldsymbol{a_n}$ be a low pass filter satisfying

$$\widehat{\boldsymbol{a}_n}[0] = 1,\ \|\boldsymbol{a}_n\|^2_{\ell_2} \leq \frac{d}{2^{n-n_1}},\ \text{and}\ \operatorname{supp}\boldsymbol{a}_n \subset \mathrm{C}[-2^{n-n_1}, 2^{n-n_1}]^d, \tag{2.17}$$

where C is a positive constant independent of n and n_1. Then, we define

$$\tilde{\phi}(x) = 2^{d(n-n_1)} \sum_{\alpha\in\mathbb{Z}^d} \boldsymbol{a}_n[\alpha]\phi(2^{n-n_1}x - \alpha), \quad x \in \mathbb{R}^d. \tag{2.18}$$

By the properties of $\boldsymbol{a}_n$, we have $\int_{\mathbb{R}^d}\tilde{\phi}(x)dx = \int_{\mathbb{R}^d}\phi(x)dx$ and $\operatorname{supp}\tilde{\phi} \sim \operatorname{supp}\phi$. In particular, if ϕ is constructed by tensor product from a univariate B-spline function (2.3), we can choose $\tilde{\phi} = \phi$ and $\boldsymbol{a}_n$ as the refinement mask of (2.18).

For sequences $\boldsymbol{y}$ and $\boldsymbol{a}_n$, the discrete convolution on Euclidean points $2^{-n}\mathbb{I}_n$ is defined as

$$(\boldsymbol{a}_n[-\cdot] \circledast |_{\mathbb{I}_n}\boldsymbol{y})[\alpha] := \sum_{\beta\in\mathbb{Z}^d} \boldsymbol{a}_n[\beta]\boldsymbol{y}[\alpha+\beta], \quad \alpha \in \mathbb{I}_n.$$

Let

$$\mathbb{I}_{n_1} = \{\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d) \in \mathbb{Z}^d,\ 0 \leq \alpha_1, \alpha_2, \ldots, \alpha_d \leq 2^{n_1} - 1\}.$$

In the uniform grids $2^{-n_1}\mathbb{I}_{n_1} \subset 2^{-n}\mathbb{I}_n$, we define $\tilde{\boldsymbol{y}} \in \ell_2(\mathbb{I}_{n_1})$ as

$$\tilde{\boldsymbol{y}}[\alpha] = \big(\boldsymbol{a}_n[-\cdot] \circledast |_{\mathbb{I}_n}\boldsymbol{y}\big)[2^{n-n_1}\alpha], \quad \alpha \in \mathbb{I}_{n_1}. \tag{2.19}$$

Here for $\alpha \in \mathbb{I}_{n_1}$, $\tilde{\boldsymbol{y}}[\alpha]$ can be seen as a "new" sampling value of f at Euclidean coordinate $(2^{-n_1}\alpha) \in 2^{-n_1}\mathbb{I}_{n_1}$.

Then based on the above values $\tilde{\boldsymbol{y}}$ on Euclidean coordinates $2^{-n_1}\mathbb{I}_{n_1}$, we can find an approximation function with convergence after applying the denoising and approximation scheme as in section 2.3. That is, we obtain a denoised result $\boldsymbol{y}^*_{n_1}$ by solving

$$\min_{\boldsymbol{u}} \|\boldsymbol{u}-\tilde{\boldsymbol{y}}\|^2_{\ell_2(\mathbb{I}_{n_1})} + \|\mathrm{diag}(\lambda)\mathcal{W}\boldsymbol{u}\|_{\ell_1(\tilde{\mathbb{I}}_{n_1})}, \tag{2.20}$$

and get an approximation function

$$g^*_{n_1} = \sum_{\alpha\in\mathbb{I}_{n_1}} \boldsymbol{y}^*_{n_1}[\alpha]\varphi(2^{n_1}\cdot -\alpha). \tag{2.21}$$

THEOREM 2.4. *Let ϕ be the tensor product from a univariate B-spline function, and $\boldsymbol{y}=\mathcal{S}_n f+\boldsymbol{\epsilon}$ be the noisy observations of f on Euclidean coordinates $2^{-n}\mathbb{I}_n$. Suppose that $\mathcal{S}_n f$ and $\boldsymbol{\epsilon}$ satisfy the conditions in Theorem 2.3.*

Assume that $\boldsymbol{a}_n$ is a low pass filter satisfying the properties in (2.17) and $\tilde{\boldsymbol{y}}=\boldsymbol{a}_n[-\cdot]\circledast|_{\mathbb{I}_n}\boldsymbol{y}$ is given by (2.19). Let $\boldsymbol{y}^*_{n_1}=\mathcal{D}_{n_1}\tilde{\boldsymbol{y}}$ be the denoised result obtained by (2.20) with $\lambda_0=0$ and $0<\lambda_i\le 2^{(s-s_0)n_1}$ for some $0<s_0\le 2s$ and all $|i|\ge 1$. Let $f^*_{n_1}=\mathcal{A}_{n_1}\boldsymbol{y}^*_{n_1}$ be the approximation function given by (2.21). Then, if ϕ and φ satisfy the conditions in Proposition 2.1, we have

$$\mathcal{E}\big(\|f^*_{n_1}-f\|_{L_2(\Omega_\epsilon)}\big) \le \mathrm{C}\big((2^{-n_1})^{\min\{\zeta,r\}} + 2^{-\frac{s_0}{2}n_1} + \sqrt{d}\sigma 2^{-\frac{(n-n_1)}{2}}\big).$$

In particular, for any $\mu>0$, with probability at least $1-\frac{1}{\mu}$,

$$\|f^*_{n_1}-f\|_{L_2(\Omega_\epsilon)} \le \mathrm{C}\big((2^{-n_1})^{\min\{\zeta,r\}} + 2^{-\frac{s_0}{2}n_1} + \sqrt{d\mu}\sigma 2^{-\frac{(n-n_1)}{2}}\big).$$

Proof. By the definition of $\tilde{\phi}$ in (2.18), we have

$$\begin{aligned}\tilde{\phi}(2^{n_1}x-\alpha) &= 2^{d(n-n_1)}\sum_{\beta\in\mathbb{Z}^d}\boldsymbol{a}_n[\beta]\phi\big(2^{n-n_1}(2^{n_1}x-\alpha)-\beta\big)\\ &= 2^{d(n-n_1)}\sum_{\beta\in\mathbb{Z}^d}\boldsymbol{a}_n[\beta]\phi\big(2^{n}x-2^{(n-n_1)}\alpha-\beta\big).\end{aligned}$$

Thus, we obtain the following sequence on $2^{-n_1}\mathbb{I}_{n_1}$ with $\alpha\in\mathbb{I}_{n_1}$,

$$\begin{aligned}(\boldsymbol{a}_n[-\cdot]\circledast|_{\mathbb{I}_n}\mathcal{S}_n f)[\alpha] &= \sum_{\beta\in\mathbb{Z}^d}\boldsymbol{a}_n[\beta](\mathcal{S}_n f)[2^{n-n_1}\alpha+\beta]\\ &= 2^{dn}\sum_{\beta\in\mathbb{Z}^d}\boldsymbol{a}_n[\beta]\langle f,\phi(2^n\cdot -2^{(n-n_1)}\alpha-\beta)\rangle\\ &= 2^{dn_1}\langle f,\tilde{\phi}(2^{n_1}\cdot-\alpha)\rangle\\ &= (\mathcal{S}_{n_1}f)[\alpha].\end{aligned}$$

It follows that

$$\tilde{\boldsymbol{y}} = \boldsymbol{a}_n[-\cdot]\circledast|_{\mathbb{I}_n}\boldsymbol{y} = \boldsymbol{a}_n[-\cdot]\circledast|_{\mathbb{I}_n}(\mathcal{S}_n f+\boldsymbol{\epsilon}) = \mathcal{S}_{n_1}f + \boldsymbol{a}_n[-\cdot]\circledast|_{\mathbb{I}_n}\boldsymbol{\epsilon}.$$

Since the random noise $\boldsymbol{\epsilon}[\alpha]$ are independent with $\mathcal{E}(\boldsymbol{\epsilon})=0$ and $Var(\boldsymbol{\epsilon})\le\sigma^2$, by the properties of $\boldsymbol{a}_n$ in (2.17), we have

$$\mathcal{E}(2^{-dn_1}\|\boldsymbol{a}_n[-\cdot]\circledast|_{\mathbb{I}_n}\boldsymbol{\epsilon}\|^2_{\ell_2(\mathbb{I}_{n_1})}) \le \frac{d\sigma^2}{2^{(n-n_1)}}.$$

Moreover, by Markov's inequality, for any $\mu>0$,

$$\mathcal{P}(2^{-dn_1}\|\boldsymbol{a}_n[-\cdot]\circledast|_{\mathbb{I}_n}\boldsymbol{\epsilon}\|^2_{\ell_2(\mathbb{I}_{n_1})} > \mu) \le \frac{d\sigma^2}{2^{(n-n_1)}\mu}.$$

Then, we apply Theorem 2.3 to the data $\tilde{\boldsymbol{y}}$ on $2^{-n_1}\mathbb{I}_{n_1}$ and conclude that

$$\mathcal{E}\big(\|f^*_{n_1}-f\|_{L_2(\Omega_\epsilon)}\big) \le \mathrm{C}\big(\|\mathcal{A}_{n_1}(\mathcal{S}_{n_1}f)-f\|_{L_2(\Omega_\epsilon)} + 2^{-\frac{s_0}{2}n_1} + \sqrt{d}\sigma 2^{-\frac{(n-n_1)}{2}}\big).$$

This together with Proposition 2.1 completes the proof. ∎

Theorem 2.4 shows that if the data density is high enough and the sampling noise are independent with zero mean and bounded variance, we can find a way to approximate the function with convergence. Approximation from the coarse grid space may slow down the convergence rate, however the noise level can be reduced. In fact, if we choose $\mathrm{diag}(\lambda) = \mathbf{0}$, the denoising and approximation scheme in Theorem 2.4 is exactly the quasi-interpolation scheme [12, 22].

3. Approximation on Randomly Sampled Data. In this section we discuss the approximation of functions from random sampled data. Let $\Omega \subset \mathbb{R}^d$ be a bounded open set. Suppose that we are given a noisy sample $f|_\Xi$ at the data set $\Xi \subset \Omega$, i.e.,

$$\boldsymbol{y}_i = f(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \dots, n,$$

where $\Xi = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n\}$ is a set of points drawn from Ω.

We are interested in how to find an approximation function f_n^* from the noisy data $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n$ and estimate the approximation error $\|f_n^* - f\|_{L_2(\Omega)}$ in terms of the number of samples. Furthermore, we consider when n is large enough, how to choose the denoising and approximation scheme such that the noise level can be reduced and convergence is guaranteed.

We choose the shift invariant subspace $S^h(\varphi, \Omega)$ as the approximation space, which is spanned by the integer translates of $\varphi(\cdot/h)$, i.e.,

$$S^h(\varphi, \Omega) = \{\sum_{\alpha \in I} \boldsymbol{u}[\alpha]\varphi(\frac{\cdot}{h} - \alpha) : \boldsymbol{u}[\alpha] \in \mathbb{R}\},$$

with

$$I = \{\alpha \in \mathbb{Z}^d : \mathrm{supp}\, \varphi(\frac{\cdot}{h} - \alpha) \cap \Omega \neq \emptyset\}.$$

Moreover, we assume that φ is a compactly supported function with stable shifts and satisfies the Strang-Fix conditions [30] of order k, which ensures that every smooth functions can be approximated by $S^h(\varphi)$ with high order. An explicit example is the tensor product of B-splines,

$$\varphi(x) = B_m(x_1)B_m(x_2)\cdots B_m(x_d), \quad \text{with} \quad x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d.$$

The assumption on φ ensures that there exists $\boldsymbol{b} \in \ell_0(\mathbb{Z}^d)$ such that for all $q \in \pi_{k-1}^d$,

$$q = \sum_{\alpha \in \mathbb{Z}^d} (q *' \boldsymbol{b})[\alpha]\varphi(\cdot - \alpha),$$

where π_{k-1}^d denotes the set of all polynomials in d variables with degree $\leq k - 1$ (see e.g. [11]). Let

$$\tilde{f} = \sum_{\alpha \in \mathbb{Z}^d} (f(h\cdot) *' \boldsymbol{b})[\alpha]\varphi(\frac{\cdot}{h} - \alpha). \tag{3.1}$$

We can find an approximation function from $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n$,

$$f_n^* = \sum_{\alpha \in I} \boldsymbol{u}^*[\alpha]\varphi(\frac{\cdot}{h} - \alpha) \in S^h(\varphi, \Omega) \tag{3.2}$$

with $\boldsymbol{u}^*$ being the minimizer of the following model

$$\min_{\boldsymbol{u}} \; E_w(\boldsymbol{u}) = \sum_{i=1}^n w_i(\sum_{\alpha \in I} \boldsymbol{u}[\alpha]\varphi(\frac{\boldsymbol{x}_i}{h} - \alpha) - \boldsymbol{y}_i)^2 + \rho\|\mathrm{diag}(\lambda)\mathcal{W}\boldsymbol{u}\|_{\ell_1} \, , \tag{3.3}$$

where $\mathcal{W}$ is the discrete framelet transform and ρ is the regularization parameter. The weight w_i is to balance the penalties of different $\boldsymbol{u}[\alpha]$ according to the density of sampling points in the support of $\varphi(\frac{\cdot}{h} - \alpha)$, and $\mathrm{diag}(\lambda)$ is a diagonal matrix based on the vector λ which scales the different wavelet channels. The ℓ_1-norm of the wavelet frame transform induces the model a preference to a solution whose wavelet coefficients is sparse, and to preserve important features of the function.

In the following subsections, we first consider the approximation ability of the shift invariant subspace on the sampling points, then discuss the error of the denoising model (3.3), i.e., $\|f_n^* - f\|_{\ell_{2,w}(\Xi)}$. In the end, we investigate the error $\|f_n^* - f\|_{L_2(\Omega)}$ when $\{\boldsymbol{x}_i\}_{i=1}^n$ is randomly drawn from Ω for a given sample size n and when n goes to infinity.

3.1. Error analysis.

3.1.1. Approximation error analysis. We assume that $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain satisfying the cone property, that is, there exist positive constants d_Ω, r_Ω such that for all $\xi \in \Omega$, there exists $\eta \in \Omega$ such that $|\xi - \eta| = d_\Omega$ and

$$\xi + t(\eta - \xi + r_\Omega B(1)) \subset \Omega, \qquad \forall\, t \in [0,1].$$

For $0 < h < 1$, let

$$L_h = \{hj, \quad j \in \mathbb{Z}^d\}$$

be a set of lattice nodes, and

$$\mathrm{Q}_O^h = \bigcup_{\ell_\alpha, r_\alpha \in L_h, r_\alpha - \ell_\alpha = h} [\ell_1, r_1) \times [\ell_2, r_2) \cdots \times [\ell_d, r_d) \supset \Omega \tag{3.4}$$

be the minimal set of cubes with nodes in L_h that cover Ω.

According to Q_O^h and Ξ, we define the weight of points $\{\boldsymbol{x}_i\}$ in cube $V_\beta \in \mathrm{Q}_O^h$ by

$$w_i = h^d (\#(\Xi \cap V_\beta))^{-1}, \tag{3.5}$$

if $\Xi \cap V_\beta \neq \emptyset$; and $w_i = 0$, if $\Xi \cap V_\beta = \emptyset$. The weighted ℓ_2-norm of a continuous function g on Ξ is defined by

$$\|g\|_{\ell_{2,w}(\Xi)} = (\sum_{x_i \in \Xi} w_i |g(\boldsymbol{x}_i)|^2)^{1/2}. \tag{3.6}$$

Using the above weights, the model (3.3) can be rewritten as

$$\min_{\boldsymbol{u}} \|\overline{f}(\boldsymbol{x}_i) - \boldsymbol{y}_i\|^2_{\ell_{2,w}(\Xi)} + \rho \|\mathrm{diag}(\lambda) \mathcal{W} \boldsymbol{u}\|_{\ell_1},$$

where $\overline{f} = \sum_{\alpha \in I} \boldsymbol{u}[\alpha] \varphi(\frac{\cdot}{h} - \alpha) \in S^h(\varphi, \Omega)$.

It is easy to check that for functions $g, \tilde{g} \in C(\Omega)$, we have

$$\|\tilde{g} - g\|_{\ell_{2,w}(\Xi)} \leq \|\tilde{g}\|_{\ell_{2,w}(\Xi)} + \|g\|_{\ell_{2,w}(\Xi)},$$

and for any constant c,

$$\|cg\|_{\ell_{2,w}(\Xi)} = |c| \|g\|_{\ell_{2,w}(\Xi)}.$$

In particular, for any vector $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n)^T \in \mathbb{R}^n$,

$$\|g - \boldsymbol{z}\|_{\ell_{2,w}(\Xi)} \leq \|g\|_{\ell_{2,w}(\Xi)} + \|\boldsymbol{z}\|_{\ell_{2,w}(\Xi)},$$

where

$$\|\boldsymbol{z}\|_{\ell_{2,w}} = (\sum_{i=1}^n w_i |\boldsymbol{z}_i|^2)^{1/2}.$$

PROPOSITION 3.1. *Let $f \in W_1^k(\mathbb{R}^d)$ with $k \geq d$ and $\tilde{f}$ be given by (3.1). Assume that $\varphi \in L_2(\mathbb{R}^d)$ has compact support and satisfies the Strang-Fix conditions of order k. Then*

$$\|\tilde{f} - f\|_{\ell_{2,w}(\Xi)} \leq \mathrm{C} h^{(k - \frac{d}{2})} |f|_{W_1^k(\Omega)},$$

where C *is a constant independent of h.*

Proof. For every $V_\beta \in \mathrm{Q}_O^h$, let x_β be the point in $\Xi \cap V_\beta$ such that

$$\boldsymbol{x}_\beta = \arg\max\{|\tilde{f}(\boldsymbol{x}_j) - f(\boldsymbol{x}_j)| : \boldsymbol{x}_j \in \Xi \cap V_\beta\}.$$

We pick one $\boldsymbol{x}_\beta$ for every $V_\beta \in \mathrm{Q}_O^h$ and set $\breve{\Xi} = \bigcup_{V_\beta \in \mathrm{Q}_O^h} \{\boldsymbol{x}_\beta\}$. It follows that

$$\breve{\Xi} \subseteq \Xi \quad \text{and} \quad \|\tilde{f} - f\|_{\ell_{2,w}(\Xi)} \leq h^{d/2} \|\tilde{f} - f\|_{\ell_2(\breve{\Xi})}.$$

Then, following the line of the proof of [35, Proposition 2.2], we have

$$\|\tilde{f} - f\|^2_{\ell_{2,w}(\Xi)} \leq h^d \|\tilde{f} - f\|^2_{\ell_2(\breve{\Xi})} \leq \sum_{V_\beta \in \mathrm{Q}_O^h} h^d \|\tilde{f} - f\|^2_{L_\infty(V_\beta)} \leq \mathrm{C} h^{(2k-d)} |f|^2_{W_1^k(\Omega)}. \qquad \square$$

3.1.2. Statistical error analysis. Let $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n$ be a noisy observation of f, where $\boldsymbol{y}_i = f(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i$, and $\boldsymbol{\epsilon}_i$ is a random noise. By applying the denoising scheme (3.3) to the noisy data, we obtain an approximation function f_n^* given by (3.2). In the following proposition we discuss the error of f_n^* on Ξ.

PROPOSITION 3.2. *Let $f \in W_1^k(\mathbb{R}^d)$ with $k \geq d$, and $\boldsymbol{y}_i = f(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i$. Suppose that $\{\boldsymbol{\epsilon}_i\}$ are independent random noise with*

$$\mathcal{E}(\boldsymbol{\epsilon}_i) = 0 \quad \textit{and} \quad Var(\boldsymbol{\epsilon}_i) \leq \sigma^2.$$

Let $E_w(\boldsymbol{u})$ be the denoising scheme given by (3.3) with $\operatorname{diag}(\lambda) \sim \operatorname{diag}(h^{d-k})$, and $\mathcal{W}\boldsymbol{u}$ being given by those $\mathcal{W}_i\boldsymbol{u}$ for which $|i| \geq k$. Let $\boldsymbol{u}^*$ be the minimizer of $E_w(\boldsymbol{u})$, and

$$f_n^* = \sum_{\alpha \in I} \boldsymbol{u}^*[\alpha]\varphi(\frac{\cdot}{h} - \alpha).$$

Then taking expectation w.r.t. the random variable $\boldsymbol{\epsilon}$, we have

$$\mathcal{E}\big(\|f_n^* - f\|^2_{\ell_{2,w}(\Xi)} + \rho|f_n^*|_{W_1^k(\Omega)}\big) \leq \mathrm{C}\big(h^{(2k-d)}|f|^2_{W_1^k(\Omega)} + \sigma^2 h^d \#\{\mathrm{Q}_O^h\} + \rho|f|_{W_1^k(\Omega)}\big),$$

where C *is a positive constant independent of h, and $\#\{\mathrm{Q}_O^h\}$ denotes the number of cubes in Q_O^h.*

Proof. Since $\boldsymbol{u}^*$ is the minimizer of $E_w(\boldsymbol{u})$, we have

$$\begin{aligned}
&\|f_n^* - \boldsymbol{y}\|^2_{\ell_{2,w}(\Xi)} + \rho\|\operatorname{diag}(\lambda)\mathcal{W}\boldsymbol{u}^*\|_{\ell_1}\\
&= E_w(\boldsymbol{u}^*) \leq E_w(f(h\cdot) *' \boldsymbol{b})\\
&= \|\tilde{f} - \boldsymbol{y}\|^2_{\ell_{2,w}(\Xi)} + \rho\|\operatorname{diag}(\lambda)\mathcal{W}(f(h\cdot) *' \boldsymbol{b})\|_{\ell_1}\\
&\leq 2(\|\tilde{f} - f\|^2_{\ell_{2,w}(\Xi)} + \|f - \boldsymbol{y}\|^2_{\ell_{2,w}(\Xi)}) + \mathrm{C}_1\rho\|\operatorname{diag}(\lambda)\mathcal{W}(f(h\cdot))\|_{\ell_1}\\
&= 2(\|\tilde{f} - f\|^2_{\ell_{2,w}(\Xi)} + \|\boldsymbol{\epsilon}\|^2_{\ell_{2,w}(\Xi)}) + \mathrm{C}_1\rho\|\operatorname{diag}(\lambda)\mathcal{W}(f(h\cdot))\|_{\ell_1}.
\end{aligned}$$

In addition, noting that

$$\|f_n^* - f\|_{\ell_{2,w}(\Xi)} \leq \|f_n^* - \boldsymbol{y}\|_{\ell_{2,w}(\Xi)} + \|\boldsymbol{\epsilon}\|_{\ell_{2,w}(\Xi)},$$

we obtain

$$\begin{aligned}
&\|f_n^* - f\|^2_{\ell_{2,w}(\Xi)} + \rho\|\operatorname{diag}(\lambda)\mathcal{W}\boldsymbol{u}^*\|_{\ell_1}\\
&\leq 6(\|\tilde{f} - f\|^2_{\ell_{2,w}(\Xi)} + \|\boldsymbol{\epsilon}\|^2_{\ell_{2,w}(\Xi)}) + 2\mathrm{C}_1\rho\|\operatorname{diag}(\lambda)\mathcal{W}(f(h\cdot))\|_{\ell_1}.
\end{aligned}$$

By [35, Proposition 2.1] and the choice of $\operatorname{diag}(\lambda)$, we conclude that

$$\begin{aligned}
&\|f_n^* - f\|^2_{\ell_{2,w}(\Xi)} + \rho|f_n^*|_{W_1^k(\Omega)}\\
&\leq \mathrm{C}_2\big(\|\tilde{f} - f\|^2_{\ell_{2,w}(\Xi)} + \|\boldsymbol{\epsilon}\|^2_{\ell_{2,w}(\Xi)} + \rho|f|_{W_1^k(\Omega)}\big).
\end{aligned}$$

This together with Proposition 3.1 implies that

$$\|f_n^* - f\|^2_{\ell_{2,w}(\Xi)} + \rho|f_n^*|_{W_1^k(\Omega)} \leq \mathrm{C}_3\big(h^{(2k-d)}|f|^2_{W_1^k(\Omega)} + \|\boldsymbol{\epsilon}\|^2_{\ell_{2,w}(\Xi)} + \rho|f|_{W_1^k(\Omega)}\big).$$

Moreover, by the properties of $\{\boldsymbol{\epsilon}_i\}$ and the definition of the weighted ℓ_2-norm in (3.6), we have

$$\mathcal{E}(\|\boldsymbol{\epsilon}\|^2_{\ell_{2,w}(\Xi)}) \leq \sigma^2 h^d \#\{\mathrm{Q}_O^h\}.$$

This completes the proof of the proposition. ∎

3.2. Approximation from random sampled data. Let $f \in W_1^k(\mathbb{R}^d)$ with $k \geq d$, and $\boldsymbol{y}_i = f(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i$, $i = 1, 2, \ldots, n$, be the noisy observation of f. Assume that the sampling set $\Xi = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ is uniformly randomly drawn from a bounded set $\Omega \subset \mathbb{R}^d$, i.e., for any measurable subset $\Omega_0 \subset \Omega$,

$$\mathcal{P}(\{\boldsymbol{x}_i \in \Omega_0\}) = \frac{|\Omega_0|}{|\Omega|}, \quad i = 1, 2, \ldots, n.$$

For the sampling set Ξ, the density level of Ξ in Ω is defined as

$$\delta(\Xi, \Omega) = \sup_{\boldsymbol{x} \in \Omega} \inf_{\boldsymbol{\xi} \in \Xi} |\boldsymbol{x} - \boldsymbol{\xi}|.$$

Without loss of generality, in the following we assume that $|\Omega| = 1$. We first give a probability estimate of the density level of the sampling data, then determine the scale parameter of the approximation space and give a detailed analysis of the approximation error from randomly sampled data.

For every cube $\mathcal{C} \subset \Omega$ with volume $|\mathcal{C}| = \frac{1}{n}$, by the properties of the binomial distribution, we have

$$\mathcal{P}(\{\boldsymbol{x}_i \notin \mathcal{C}, \forall \boldsymbol{x}_i \in \Xi\}) = (1 - \frac{1}{n})^n \sim \frac{1}{e}$$

and the expectation

$$\mathcal{E}(\#\{\boldsymbol{x}_i \in \mathcal{C}\}) = 1.$$

Noting that there exist cubes of the form $[\ell_1, r_1) \times [\ell_2, r_2) \cdots \times [\ell_d, r_d)$ which are of the same size and cover Ω, and in order to guarantee that there exist sampling points in most cubes with high probability, the volume of the cube should be larger than $\frac{1}{n}$. Let $0 < \gamma_1 < 1$ be a small positive number. We define a set of lattice nodes as follows:

$$L_{\gamma_1} = \{n^{\frac{-(1-\gamma_1)}{d}} j, \quad j \in \mathbb{Z}^d\}.$$

According to L_{γ_1}, there are two sets of cubes

$$Q_I^{\gamma_1} = \bigcup_{\ell_i, r_i \in L_{\gamma_1}, r_i - \ell_i = n^{\frac{-(1-\gamma_1)}{d}}} [\ell_1, r_1) \times [\ell_2, r_2) \cdots \times [\ell_d, r_d) \subset \Omega \tag{3.7}$$

and

$$Q_O^{\gamma_1} = \bigcup_{\ell_i, r_i \in L_{\gamma_1}, r_i - \ell_i = n^{\frac{-(1-\gamma_1)}{d}}} [\ell_1, r_1) \times [\ell_2, r_2) \cdots \times [\ell_d, r_d) \supset \Omega, \tag{3.8}$$

which are the most cubes inside Ω and the minimal cubes that cover Ω. It is easy to check that $Q_O^{\gamma_1} = \mathrm{Q}_O^h$ with $h = n^{\frac{-(1-\gamma_1)}{d}}$ in (3.4), and the volume of each cube is equal to $n^{\gamma_1 - 1}$.

Let $\#Q_I^{\gamma_1}$ and $\#Q_O^{\gamma_1}$ be the number of cubes in $Q_I^{\gamma_1}$ and $Q_O^{\gamma_1}$ respectively. Then

$$\#Q_I^{\gamma_1} \leq n^{1-\gamma_1},$$

and we assume that there exists a positive constant C independent of n such that

$$\#Q_O^{\gamma_1} \leq \mathrm{C}(\#Q_I^{\gamma_1}).$$

LEMMA 3.3. *Let $\Omega \subset \mathbb{R}^d$ be a bounded set satisfying the cone property and $\Xi = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ be uniformly drawn from Ω. Let $Q_I^{\gamma_1}$ be the set of cubes defined by (3.7), and $\Xi_I = \Xi \cap Q_I^{\gamma_1}$ be the sampling points in $Q_I^{\gamma_1}$. For every cube $V_\alpha \in Q_I^{\gamma_1}$, let $\#(\Xi \cap V_\alpha)$ denote the number of points in $\Xi \cap V_\alpha$.*
(i) For an arbitrary $0 \leq \gamma_2 < \gamma_1 < 1$,

$$\mathcal{P}\{\textit{every}\, V_\alpha \in Q_I^{\gamma_1}, \#(\Xi \cap V_\alpha) > n^{\gamma_2}\} \geq 1 - n^{1-\gamma_1} \exp\big(-\frac{(n^{\gamma_1} - 2n^{\gamma_2})}{2}\big).$$

(ii) With probability greater than

$$1 - n^{1-\gamma_1} \exp\big(-\frac{(n^{\gamma_1} - 2)}{2}\big),$$

the density level of Ξ_I in Ω

$$\delta(\Xi_I, \Omega) \leq \mathrm{C} n^{\frac{-(1-\gamma_1)}{d}},$$

and thus

$$\delta(\Xi, \Omega) \leq \mathrm{C} n^{\frac{-(1-\gamma_1)}{d}},$$

where C *is a positive constant dependent only on Ω.*

Proof. (i) For any given $V_{\alpha_0} \in Q_I^{\gamma_1}$, let

$$p = \mathcal{P}\{\boldsymbol{x}_i \in V_{\alpha_0}\} = |V_{\alpha_0}| = n^{-(1-\gamma_1)}$$

be the probability of "success" in the sequence of n independent experiments, and the cumulative distribution function be expressed as

$$F(n^{\gamma_2}; n, p) = \mathcal{P}\{\#(\Xi \cap V_{\alpha_0}) \leq n^{\gamma_2}\}.$$

By Chernoff's inequality of the binomial distribution [19], we have

$$F(n^{\gamma_2}; n, p) \leq \exp\big(-\frac{1}{2p}\frac{(np - n^{\gamma_2})^2}{n}\big).$$

Since the number of cubes V_α in $Q_I^{\gamma_1}$ is less than $\frac{|\Omega|}{|V_\alpha|} = n^{1-\gamma_1}$, we obtain that

$$\begin{aligned}
&\mathcal{P}\{\exists\, V_\alpha \in Q_I^{\gamma_1}, \text{ s.t. } \#(\Xi \cap V_\alpha) \leq n^{\gamma_2}\} \\
&\leq n^{1-\gamma_1} F(n^{\gamma_2}; n, p) \\
&\leq n^{1-\gamma_1} \exp\big(-\frac{n^{1-\gamma_1}}{2}\frac{(np - n^{\gamma_2})^2}{n}\big) \\
&= n^{1-\gamma_1} \exp\big(-\frac{(n^{\gamma_1} - n^{\gamma_2})^2}{2n^{\gamma_1}}\big) \\
&\leq n^{1-\gamma_1} \exp\big(-\frac{(n^{\gamma_1} - 2n^{\gamma_2})}{2}\big).
\end{aligned}$$

Thus,

$$\mathcal{P}\{\forall\, V_\alpha \in Q_I^{\gamma_1}, \#(\Xi \cap V_\alpha) > n^{\gamma_2}\} \geq 1 - n^{1-\gamma_1} \exp\big(-\frac{(n^{\gamma_1} - 2n^{\gamma_2})}{2}\big).$$

(ii) By the cone property of Ω, for every $\boldsymbol{x} \in \Omega$, there exists a ball $B(o, r) \subset \Omega$ with center o and radius $r = 2n^{\frac{-(1-\gamma_1)}{d}}$ satisfying

$$|\boldsymbol{x} - o| \leq \frac{d_\Omega}{r_\Omega} r.$$

Besides, there exists a cube $V_{\alpha_1} \in Q_I^{\gamma_1}$ such that $V_{\alpha_1} \subset B(o, r)$. By (i), with probability at least $1 - n^{1-\gamma_1} \exp\big(-\frac{(n^{\gamma_1}-2)}{2}\big)$, there exists a point $\boldsymbol{x}_{i_1} \in \Xi \cap V_{\alpha_1}$ inside $B(o, r)$ and

$$|\boldsymbol{x} - \boldsymbol{x}_{i_1}| \leq (1 + \frac{d_\Omega}{r_\Omega}) r = 2(1 + \frac{d_\Omega}{r_\Omega}) n^{\frac{-(1-\gamma_1)}{d}}.$$

Thus,

$$\delta(\Xi_I, \Omega) \leq 2(1 + \frac{d_\Omega}{r_\Omega}) n^{\frac{-(1-\gamma_1)}{d}}.$$

In addition, since $\Xi_I \subset \Xi$, it is obvious that $\delta(\Xi, \Omega) \leq \delta(\Xi_I, \Omega)$. ∎

THEOREM 3.4. *Let $f \in W_1^k(\mathbb{R}^d)$ with $k \geq d$, and $\Omega \subset \mathbb{R}^d$ be a bounded set satisfying the cone property. Let $\{\boldsymbol{x}_i\}_{i=1}^n$ be uniformly drawn from Ω, and $\boldsymbol{y}_i = f(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i$. Suppose that $\{\boldsymbol{\epsilon}_i\}$ are independent random noise with*

$$\mathcal{E}(\boldsymbol{\epsilon}_i) = 0 \quad \textit{and} \quad Var(\boldsymbol{\epsilon}_i) \leq \sigma^2.$$

Let $E_w(\boldsymbol{u})$ be the denoising model given by (3.3) with $h = n^{\frac{-(1-\gamma_1)}{d}}$ for some $0 < \gamma_1 < 1$, $\mathrm{diag}(\lambda) \sim \mathrm{diag}(h^{d-k})$, and $\mathcal{W}\boldsymbol{u}$ being given by those $\mathcal{W}_i\boldsymbol{u}$ for which $|i| \geq k$. Moreover, suppose that the weight w_i of (3.3) is given by (3.5), and ρ denotes the regularization parameter. Let $\boldsymbol{u}^*$ be the minimizer of $E_w(\boldsymbol{u})$ and

$$f_n^* = \sum_{\alpha\in I} \boldsymbol{u}^*[\alpha]\varphi(\frac{\cdot}{h} - \alpha).$$

Then for any $\mu > 0$, the following inequality

$$\begin{aligned}
(3.9) \qquad & \|f_n^* - f\|_{L_2(\Omega)} \\
&\leq \mathrm{C}\Big(n^{-\frac{(1-\gamma_1)}{d}(k-\frac{d}{2})}|f|_{W_1^k(\Omega)} + \sqrt{\rho}|f|^{\frac{1}{2}}_{W_1^k(\Omega)} + n^{-\frac{3(1-\gamma_1)(2k-d)}{2d}}\rho^{-1}|f|^2_{W_1^k(\Omega)} \\
&\quad + \mu^{1/2}\sigma + n^{\frac{-(1-\gamma_1)(2k-d)}{2d}}\rho^{-1}\mu\sigma^2\Big)
\end{aligned}$$

holds with probability at least

$$(1-\frac{1}{\mu})\Big(1 - n^{1-\gamma_1}\exp(-\frac{(n^{\gamma_1}-2)}{2})\Big),$$

where C is a positive constant independent of n.

Proof. By Lemma 3.3, with probability at least

$$1 - n^{1-\gamma_1}\exp(-\frac{(n^{\gamma_1}-2)}{2}),$$

the density level of $\Xi = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ in Ω,

$$\delta(\Xi, \Omega) \leq \mathrm{C}_1 n^{\frac{-(1-\gamma_1)}{d}}. \tag{3.10}$$

For any continuous function g, let

$$\widetilde{\Xi}_g = \bigcup_{V_\alpha \in Q_O^{\gamma_1}} \{\boldsymbol{x}_\alpha : \boldsymbol{x}_\alpha = \arg\min\{|g(\boldsymbol{x}_j)| : \boldsymbol{x}_j \in \Xi \cap V_\alpha\},$$

where $Q_O^{\gamma_1}$ is the set of cubes defined by (3.8). Then we have

$$\|g\|_{\ell_2(\widetilde{\Xi}_g)} = (\sum_{\boldsymbol{x}_\alpha\in\widetilde{\Xi}_g} |g(\boldsymbol{x}_\alpha)|^2)^{1/2} \leq h^{-\frac{d}{2}}\|g\|_{\ell_{2,w}(\Xi)}.$$

Moreover, we can check that (3.10) implies that the density level of $\widetilde{\Xi}_g$ in Ω,

$$\delta(\widetilde{\Xi}_g, \Omega) \sim \delta(\Xi, \Omega).$$

By Duchon's inequality [2, Theorem 4.1], for any $g \in W_1^k(\Omega)$,

$$\begin{aligned}
\|g\|_{L_2(\Omega)} &\leq \mathrm{C}_2\big(\delta(\widetilde{\Xi}_g, \Omega)^{k-\frac{d}{2}}|g|_{W_1^k(\Omega)} + \delta(\widetilde{\Xi}_g, \Omega)^{\frac{d}{2}}\|g\|_{\ell_2(\widetilde{\Xi}_g)}\big) \\
&\leq \mathrm{C}_2\big(h^{k-\frac{d}{2}}|g|_{W_1^k(\Omega)} + \|g\|_{\ell_{2,w}(\Xi)}\big).
\end{aligned}$$

It follows that

$$\|f_n^* - f\|_{L_2(\Omega)} \leq \mathrm{C}_2\big(h^{k-\frac{d}{2}}(|f_n^*|_{W_1^k(\Omega)} + |f|_{W_1^k(\Omega)}) + \|f_n^* - f\|_{\ell_{2,w}(\Xi)}\big).$$

In addition, by Proposition 3.2 and Markov's inequality, for any $\mu > 0$, with probability at least $1 - \frac{1}{\mu}$, we have

$$\|f_n^* - f\|_{\ell_{2,w}(\Xi)}^2 + \rho |f_n^*|_{W_1^k(\Omega)} \leq \mathrm{C}_3 \big(h^{(2k-d)} |f|_{W_1^k(\Omega)}^2 + \mu\sigma^2 + \rho |f|_{W_1^k(\Omega)}\big).$$

Therefore, we conclude that the inequality (3.9) holds with probability at least

$$(1 - \frac{1}{\mu})\Big(1 - n^{1-\gamma_1} \exp(-\frac{(n^{\gamma_1} - 2)}{2})\Big). \qquad \square$$

Theorem 3.4 shows that the approximation error $\|f_n^* - f\|_{L_2(\Omega)}$ consists of three parts, in which $h^{(k-\frac{d}{2})}|f|_{W_1^k(\Omega)}$ is the error determined by the approximation ability of $S^h(\varphi, \Omega)$, $\rho^{1/2}|f|_{W_1^k(\Omega)}^{\frac{1}{2}} + \rho^{-1} h^{3(k-\frac{d}{2})} |f|_{W_1^k(\Omega)}^2$ is the regularization error of the model, and $\mu^{1/2}\sigma + h^{k-\frac{d}{2}}\rho^{-1}\mu\sigma^2$ is the noise error. Here, the scale $h = n^{\frac{-(1-\gamma_1)}{d}}$ is determined by the density level of sampling data. We can choose the regularization parameter such that $n^{\frac{-(1-\gamma_1)(2k-d)}{2d}} \leq \rho \to 0$ as $n \to \infty$. Then, when the data density is high enough, the regularization error can be negligible and $\|f_n^* - f\|_{L_2(\Omega)}$ is bounded by the noise level.

3.3. Reducing the noise level. For a given sample size and noise level, Theorem 3.4 provides a denoising scheme to approximate functions from the random sampled data based on the analysis of data density, and gives an approximation analysis of the solution. In the following, we consider the case when there are multiple sensors and the number of sampling points is large enough, how to tackle the problem of sampling noise. The idea is similar to section 2.4. We first filter the neighbouring points in every local area at a high resolution level, then approximate f from these filtered values at a relatively coarse level. The advantage of this process is that the noise level will be sufficiently reduced and meanwhile the convergence can be guaranteed.

Let $0 < \gamma_1 < 1$ be a small positive number and $Q_I^{\gamma_1}$ be given by (3.7). For every $V_\beta \in Q_I^{\gamma_1}$, let $\Xi \cap V_\beta = \{\boldsymbol{x}_{\beta_1}, \boldsymbol{x}_{\beta_2}, \ldots, \boldsymbol{x}_{\beta_{\beta_s}}\}$ be the β_s sampling points in the cube V_β. We define a sampling point $(\boldsymbol{x}_\beta, \boldsymbol{y}_\beta)$ of V_β as follows

$$\boldsymbol{x}_\beta = \frac{1}{\beta_s} \sum_{j=1}^{\beta_s} \boldsymbol{x}_{\beta_j} \tag{3.11}$$

and

$$\boldsymbol{y}_\beta = \frac{1}{\beta_s} \sum_{j=1}^{\beta_s} \boldsymbol{y}_{\beta_j}.$$

Let

$$\overline{\overline{\Xi}} = \bigcup_{V_\beta \in Q_I^{\gamma_1}} \{x_\beta : \text{ for every } V_\beta \in Q_I^{\gamma_1}, \boldsymbol{x}_\beta \in V_\beta \text{ is defined by (3.11)}\}.$$

Then based on the filtered data $\{(\boldsymbol{x}_\beta, \boldsymbol{y}_\beta) : \boldsymbol{x}_\beta \in \overline{\overline{\Xi}}\}$, we apply the following denoising scheme

$$\min_{\boldsymbol{u}} \tilde{E}(\boldsymbol{u}) = h^d \sum_{\boldsymbol{x}_\beta \in \overline{\overline{\Xi}}} (\sum_{\alpha \in I} \boldsymbol{u}[\alpha] \varphi(\frac{\boldsymbol{x}_\beta}{h} - \alpha) - \boldsymbol{y}_\beta)^2 + \rho \|\mathrm{diag}(\lambda) \mathcal{W} \boldsymbol{u}\|_{\ell_1} \tag{3.12}$$

with $h = n^{\frac{-(1-\gamma_1)}{d}}$.

Theorem 3.5. *Let $f \in W_1^k(\mathbb{R}^d)$ with $k > d$, and $\Omega \subset \mathbb{R}^d$ be a bounded set satisfying the cone property. Let $\{\boldsymbol{x}_i\}_{i=1}^n$ be uniformly drawn from Ω, and $\boldsymbol{y}_i = f(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i$. Suppose that $\{\boldsymbol{\epsilon}_i\}$ are independent random noise with*

$$\mathcal{E}(\boldsymbol{\epsilon}_i) = 0 \quad \textit{and} \quad Var(\boldsymbol{\epsilon}_i) \leq \sigma^2.$$

Let $0 < \gamma_1 < 1$ and $\{(\boldsymbol{x}_\beta, \boldsymbol{y}_\beta) : \boldsymbol{x}_\beta \in \overline{\overline{\Xi}}\}$ be the data obtained by (3.11) from $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n$. Let $\boldsymbol{u}^*$ be the minimizer of $\tilde{E}(\boldsymbol{u})$ in (3.12) with $\mathrm{diag}(\lambda) \sim \mathrm{diag}(h^{d-k})$ and $\mathcal{W}\boldsymbol{u}$ being given by those $\mathcal{W}_i \boldsymbol{u}$ for which $|i| \geq k$. Let

$$f_n^* = \sum_{\alpha \in I} \boldsymbol{u}^*[\alpha] \varphi(\frac{\cdot}{h} - \alpha).$$

(i) For any $\mu > 0$ and $0 < \gamma_2 < \gamma_1$, the following inequality

$$\begin{aligned}
&\|f_n^* - f\|_{L_2(\Omega)} \\
&\le \mathrm{C}\Big(n^{-\frac{(1-\gamma_1)}{d}(k-\frac{d}{2})}|f|_{W_1^k(\Omega)} + \sqrt{\rho}|f|^{\frac{1}{2}}_{W_1^k(\Omega)} + n^{-\frac{3(1-\gamma_1)(2k-d)}{2d}}\rho^{-1}|f|^2_{W_1^k(\Omega)} \\
&+ \mu^{1/2}\big(n^{-\frac{(1-\gamma_1)}{d}}\|f\|_{W_1^k(\Omega)} + n^{\frac{-\gamma_2}{2}}\,\sigma\big) \\
&+ n^{\frac{-(1-\gamma_1)(2k-d)}{2d}}\rho^{-1}\mu\big(n^{-\frac{2(1-\gamma_1)}{d}}\|f\|^2_{W_1^k(\Omega)} + n^{-\gamma_2}\,\sigma^2\big)\Big)
\end{aligned}$$

holds with probability at least

$$(1 - \frac{1}{\mu})\Big(1 - n^{1-\gamma_1}\exp\big(-\frac{(n^{\gamma_1} - 2n^{\gamma_2})}{2}\big)\Big).$$

(ii) If the regularization parameter ρ is chosen such that

$$n^{-\frac{(1-\gamma_1)}{d}(k-\frac{d}{2})} \le \rho \to 0$$

as $n \to \infty$, and $\gamma_2 = \frac{\gamma_1}{2}$, then when n is large enough, the following inequality

$$\|f_n^* - f\|_{L_2(\Omega)} \le \mathrm{C}\Big(\sqrt{\rho}|f|^{\frac{1}{2}}_{W_1^k(\Omega)} + n^{-\frac{(1-\gamma_1)}{2d}}\|f\|_{W_1^k(\Omega)} + n^{-\frac{(1-\gamma_1)}{d}}\|f\|^2_{W_1^k(\Omega)} + n^{-\frac{\gamma_1}{8}}\sigma\Big)$$

holds with probability at least

$$\Big(1 - \max\{n^{-\frac{(1-\gamma_1)}{d}}, n^{-\frac{\gamma_1}{4}}\}\Big)\Big(1 - n^{1-\gamma_1}\exp\big(-n^{\frac{\gamma_1}{2}}\big)\Big).$$

In particular, for all $\omega > 0$, we have

$$\lim_{n\to\infty}\mathcal{P}\big(\|f_n^* - f\|_{L_2(\Omega)} > \omega\big) = 0.$$

Proof. (i) For every $\boldsymbol{x}_\beta \in \overline{\overline{\Xi}}$, we have

$$\begin{aligned}
|f(\boldsymbol{x}_\beta) - \boldsymbol{y}_\beta| &= |f(\boldsymbol{x}_\beta) - \frac{1}{\beta_s}\sum_{j=1}^{\beta_s}(f(\boldsymbol{x}_{\beta_j}) + \boldsymbol{\epsilon}_{\beta_j})| \\
&\le |f(\boldsymbol{x}_\beta) - \frac{1}{\beta_s}\sum_{j=1}^{\beta_s} f(\boldsymbol{x}_{\beta_j})| + |\frac{1}{\beta_s}\sum_{j=1}^{\beta_s}\boldsymbol{\epsilon}_{\beta_j}| \\
&\le |\frac{1}{\beta_s}\sum_{j=1}^{\beta_s}(f(\boldsymbol{x}_\beta) - f(\boldsymbol{x}_{\beta_j}))| + |\frac{1}{\beta_s}\sum_{j=1}^{\beta_s}\boldsymbol{\epsilon}_{\beta_j}| \\
&\le \max_{j=1,2,\ldots,\beta_s}|f(\boldsymbol{x}_\beta) - f(\boldsymbol{x}_{\beta_j})| + |\frac{1}{\beta_s}\sum_{j=1}^{\beta_s}\boldsymbol{\epsilon}_{\beta_j}|.
\end{aligned}$$

Since $\boldsymbol{x}_\beta$ and $\{\boldsymbol{x}_{\beta_j}\}_{j=1}^{\beta_s}$ are in the same cube V_β, by the Sobolev embedding theorem [1], we obtain

$$|f(\boldsymbol{x}_\beta) - f(\boldsymbol{x}_{\beta_j})| \le \mathrm{C}_1\|f\|_{W_1^k(V_\beta)} n^{\frac{-(1-\gamma_1)}{d}}.$$

Moreover, by Lemma 3.3, with probability at least

$$1 - n^{1-\gamma_1}\exp\big(-\frac{(n^{\gamma_1} - 2n^{\gamma_2})}{2}\big),$$

there are more than n^{γ_2} points in every cube $V_\beta \in Q_I^{\gamma_1}$, i.e., $\beta_s \geq n^{\gamma_2}$. Thus, taking expectation over $\boldsymbol{\epsilon}_j$, we obtain

$$\mathcal{E}\big(|\frac{1}{\beta_s}\sum_{j=1}^{\beta_s}\boldsymbol{\epsilon}_j|^2\big) \leq \frac{\sigma^2}{n^{\gamma_2}}.$$

It follows that with probability at least

$$1 - n^{1-\gamma_1}\exp\big(-\frac{(n^{\gamma_1}-2n^{\gamma_2})}{2}\big),$$

we have

$$\begin{aligned}\mathcal{E}\big((\#\{\Xi\})^{-1}\|f-y_\beta\|^2_{\ell_2(\Xi)}\big) &\leq 2\big((\mathrm{C}_1)^2 n^{\frac{-2(1-\gamma_1)}{d}}\|f\|^2_{W_1^k(\Omega)} + \frac{\sigma^2}{n^{\gamma_2}}\big)\\ &\leq \mathrm{C}_2\big(n^{-\frac{2(1-\gamma_1)}{d}}\|f\|^2_{W_1^k(\Omega)} + n^{-\gamma_2}\sigma^2\big).\end{aligned}$$

Then we apply Theorem 3.4 to the data Ξ, and conclude that for any $\mu > 0$, the result of (i) holds.

(ii) If we choose $\gamma_2 = \frac{\gamma_1}{2}$, $\mu = \min\{n^{\frac{(1-\gamma_1)}{d}}, n^{\frac{\gamma_1}{4}}\}$ and ρ such that

$$n^{-\frac{(1-\gamma_1)}{d}(k-\frac{d}{2})} \leq \rho \to 0$$

as $n \to \infty$, then when n is large enough, the approximation error can be simplified by ignoring high order infinitesimal. By the result of (i), we can check that the following inequality

$$\|f_n^* - f\|_{L_2(\Omega)} \leq \mathrm{C}_4\Big(\sqrt{\rho}|f|^{\frac{1}{2}}_{W_1^k(\Omega)} + n^{-\frac{(1-\gamma_1)}{2d}}\|f\|_{W_1^k(\Omega)} + n^{-\frac{(1-\gamma_1)}{d}}\|f\|^2_{W_1^k(\Omega)} + n^{-\frac{\gamma_1}{8}}\sigma\Big)$$

holds with probability at least

$$(1-\frac{1}{\mu})\Big(1 - n^{1-\gamma_1}\exp\big(-n^{\frac{\gamma_1}{2}}\big)\Big).$$

The desired result then follows. ∎

Remark 3.6. *In Theorem 3.4 and 3.5 we discussed the approximation of analog signals from random sampled data. Based on the estimate of data density, we proposed an ℓ_1-regularized weighted least squares model which makes additional use of the wavelet frame transform in order to preserve the discontinuity features. The weight in the model is to balance the penalties of different coefficients. When the density of sampling points is high enough, the filtering process of the original data can reduce the noise level and the convergence can be obtained at a relatively coarse level. In the special case when points are in the uniform grids, and the approximation function is chosen to be an interpolatory function, that is $\varphi(0) = 1$ and $\varphi(m) = 0$ for all $m \in \mathbb{Z}^d\backslash\{0\}$, the denoising model (3.3) is the same as (2.14) with $h = 2^{-n}$.*

We assumed that $f \in W_1^k(\mathbb{R}^d)$ and the approximation result was restrict to a bounded Lipschitz domain, so no boundary conditions were considered. If we assume that f is defined on Ω, but no information about f outside Ω is known, the boundary problem will become more subtle. The parameters λ in Theorem 3.4 were chosen of the same order to make all of the wavelet channels $\mathcal{W}_i\boldsymbol{u}^*$ decay for $|i| \geq 1$ and $\|\mathrm{diag}(\lambda)\mathcal{W}\boldsymbol{u}^*\|_{\ell_1} \sim |f_n^*|_{W_1^k}$. The regularization parameter ρ should not be large in order to fit the observations, nor too small in order to control the smoothness and avoid overfitting.

REFERENCES

[1] R. A. Adams and J. J. Fournier, *Sobolev spaces*, Academic Press, 2003.
[2] R. Arcangéli, M. C. L. de Silanes, and J. J. Torrens, *An extension of a bound for functions in sobolev spaces, with applications to (m,s)-spline interpolation and smoothing*, Numerische Mathematik, 107 (2007), pp. 181–211.
[3] A. Ben-Tal and E. Hochman, *More bounds on the expectation of a convex function of a random variable*, Journal of Applied Probability, 9 (1972), pp. 803–812.

[4] L. Borup, R. Gribonval, and M. Nielsen, *Bi-framelet systems with few vanishing moments characterize besov spaces*, Applied and Computational Harmonic Analysis, 17 (2004), pp. 3–28.
[5] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, *Image restoration: total variation, wavelet frames, and beyond*, Journal of the American Mathematical Society, 25 (2012), pp. 1033–1089.
[6] J.-F. Cai, S. Osher, and Z. Shen, *Split bregman methods and frame based image restoration*, Multiscale modeling & simulation, 8 (2010), pp. 337–369.
[7] A. Chambolle, R. A. De Vore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Transactions on Image Processing, 7 (1998), pp. 319–335.
[8] Z. Chen, R. Tuo, and W. Zhang, *Stochastic convergence of a nonconforming finite element method for the thin plate spline smoother for observational data*, Siam Journal on Numerical Analysis, 56 (2018), pp. 635–659.
[9] Z. Chen, R. Tuo, and W. Zhang, *A balanced oversampling finite element method for elliptic problems with observational boundary data*, Journal of Computational Mathematics, 38 (2020), pp. 355–374.
[10] W. Dahmen, *Multiscale and wavelet methods for operator equations*, in Multiscale problems and methods in numerical simulations, Springer, 2003, pp. 31–96.
[11] I. Daubechies, B. Han, A. Ron, and Z. Shen, *Framelets: MRA-based constructions of wavelet frames*, Applied and computational harmonic analysis, 14 (2003), pp. 1–46.
[12] C. de Boor, R. A. Devore, and A. Ron, *Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$*, Transactions of the American Mathematical Society, 341 (1994), pp. 787–806.
[13] B. Dong, Q. Jiang, and Z. Shen, *Image restoration: Wavelet frame shrinkage, nonlinear evolution pdes, and beyond*, Multiscale Modeling & Simulation, 15 (2017), pp. 606–660.
[14] 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).
[15] B. Dong, Z. Shen, and P. Xie, *Image restoration: A general wavelet frame based model and its asymptotic analysis*, SIAM Journal on Mathematical Analysis, 49 (2017), pp. 421–445.
[16] W. Gao, X. Sun, Z. Wu, and X. Zhou, *Multivariate monte carlo approximation based on scattered data*, SIAM Journal on Scientific Computing, 42 (2020), pp. A2262–A2280.
[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT press Cambridge, 2016.
[18] B. Han and Z. Shen, *Dual wavelet frames and riesz bases in sobolev spaces*, Constructive Approximation, 29 (2009), pp. 369–406.
[19] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Publications of the American Statistical Association, 58 (1963), pp. 13–30.
[20] H. Ji, Z. Shen, and Y. Xu, *Wavelet frame based scene reconstruction from range data*, Journal of Computational Physics, 229 (2010), pp. 2093–2108.
[21] R.-Q. Jia, *Approximation with scaled shift-invariant spaces by means of quasi-projection operators*, Journal of Approximation Theory, 131 (2004), pp. 30–46.
[22] 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, Elsevier, 1991, pp. 209–246.
[23] M. J. Johnson, *Scattered date interpolation from principal shift-invariant spaces*, Journal of Approximation Theory, 113 (2001), pp. 172–188.
[24] M. J. Johnson, Z. Shen, and Y. Xu, *Scattered data reconstruction by regularization in b-spline and associated wavelet spaces*, Journal of Approximation Theory, 159 (2009), pp. 197–223.
[25] S. Mallat, *A wavelet tour of signal processing: the sparse way*, Academic press, 2008.
[26] D. B. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on pure and applied mathematics, 42 (1989), pp. 577–685.
[27] A. Ron and Z. Shen, *Affine systems in $L_2(\mathbb{R}^d)$: the analysis of the analysis operator*, Journal of Functional Analysis, 148 (1997), pp. 408–447.
[28] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: nonlinear phenomena, 60 (1992), pp. 259–268.
[29] Z. Shen, *Wavelet frames and image restorations*, in Proceedings of the International Congress of Mathematicians 2010, World Scientific, 2010, pp. 2834–2863.
[30] G. Strang and G. Fix, *A fourier analysis of the finite element variational method*, in Constructive aspects of functional analysis, Springer, 1973, pp. 793–840.
[31] G. Wahba, *Spline models for observational data*, Society for Industrial and Applied Mathematics, 1990.
[32] H. Wendland, *Scattered data approximation*, Cambridge University Press, 2005.
[33] Y. Xu and Q. Ye, *Generalized mercer kernels and reproducing kernel banach spaces*, Memoirs of the American Mathematical Society, 258 (2019).
[34] J. Yang, *Random sampling and reconstruction in multiply generated shift-invariant spaces*, Analysis and Applications, 17 (2019), pp. 323–347.
[35] J. Yang, D. Stahl, and Z. Shen, *An analysis of wavelet frame based scattered data reconstruction*, Applied and Computational Harmonic Analysis, 42 (2017), pp. 480–507.
[36] J. Yang, G. Zhu, D. Tong, L. Lu, and Z. Shen, *B-spline tight frame based force matching method*, Journal of Computational Physics, 362 (2018), pp. 208–219.
[37] Q. Zhang, L. Wang, and W. Sun, *Signal denoising with average sampling*, Digital Signal Processing, 22 (2012), pp. 226–232.