FINDING NUMERICAL DERIVATIVES FOR UNSTRUCTURED AND NOISY DATA BY MULTISCALE KERNELS

LEEVAN LING*

Abstract. The recently developed multiscale kernel of R. Opfer is applied to approximate numerical derivatives. The proposed method is truly mesh-free and can handle unstructured data with noise in any dimension. The method of Tikhonov and the method of L-curve are employed for regularization; no information about the noise level is required. An error analysis is provided in a general setting for all dimensions. Numerical comparisons are given in two dimensions which show competitive results with recently published thin plate spline methods.

Key words. Numerical differentiation, multiscale kernel, multivariate interpolation, unstructured data, inverse problems, Tikhonov regularization, L-curve.

AMS subject classifications. 65D05, 65D25, 65J20, 65J22

1. Introduction. Evaluating derivatives of a function using only information from discrete function values is a typical ill-posed problem. Small measurement errors, including rounding errors, will be greatly amplified during the numerical differentiation process. The problem of *numerical differentiation* arises in many branches of science and engineering. Some practical examples are the identification of discontinuities in image reconstruction [10, 13], resolution enhancement of spectra [17], solving Abel integral equations [7, 12], determination of peaks in chemical spectroscopy [24], determination of discontinuous points of the exact solutions [33], solving integral equations [8], determination of source parameter and diffusion coefficient in parabolic differential equations [6, 14], simulation of constrained mechanical systems of particles [19], singular convolution [25], and many other inverse problems in Mathematical Physics. The previous literature on numerical differentiation featured plenty of nicely calculated practical solutions, but most research papers on this topic are limited to one dimension or highly structured grids [4, 14, 20, 26, 27, 30, 33, etc.]. Numerical methods for higher dimensions are very limited. In particular, many existing methods are based on finite difference schemes [2], wavelet methods [5], and thin plate splines approximation [34]. The goal of this paper is to supply a new, efficient and practical alternative for scientists and engineers who need to compute numerical differentiation from real-life, large-scale and noisy multivariate data.

Given some set of real-life data in any dimension, multivariate functions are reconstructed from unstructured data by some specially designed multiscale kernels

$$\Phi(x,\cdot)=\sum_{j=0}^{u}\sum_{k\in\mathbb{Z}^d}\lambda_\sigma^j\varphi(2^jx-k)\varphi(2^j\cdot-k).$$

Since multiscale kernels are proven to be positive definite, for every set of data points we can solve an interpolation problem and write the interpolant in the form of the *kernel representation*:

$$s=\sum_{i=1}^{n}\beta_i\Phi(x_i,\cdot). \tag{1.1}$$

*Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. (lling@hkbu.edu.hk).

The multiscale property, found in wavelet analysis, is considered a major breakthrough in the development of kernel-based mesh-free methods. We can go one step further and express (1.1) in its *frame representation*:

$$s = \sum_{j=1}^{u} \sum_{k \in \mathbb{Z}^d} \lambda_\sigma^j c_k^j \varphi(2^j \cdot - k), \tag{1.2}$$

where $c_k^j = c_k^j(\{x_i\}, \beta_i)$ are called the frame coefficients. The interpolant obtained will have a frame representation on structured grids instead of the unstructured data. The solution process involves solving a sparse matrix system if the multiscale kernel is compactly supported. Once we determine the multivariate function that interpolates the noisy data, this newly developed method has potential applications in many branches of science and engineering. The well-developed wavelet techniques (e.g. denoising, compression, shape detection, and etc.) can be applied thereafter. In this paper, we focus on a classical ill-posed numerical differentiation problem. The derivative of (1.2) can be obtained by replacing φ by $D^\gamma \varphi$. An overview of multiscale kernels will be given in Section 2.

In Section 3, the instability of numerical differentiation is regularized by the Tikhonov regularization method that seeks a stable approximate interpolant. Error estimates in Section 3.1 show that the errors of numerical derivatives blow up when the noise level is high or when the minimum separation distance of the data points is small. This agrees with the ill-posed nature of numerical differentiation. On the other hand, both errors in interpolation and in the derivatives can be minimized with an optimal regularization parameter. In Section 4, the L-curve method is employed to numerically located the optimal regularization parameter. Finally, two bivariate examples are given in Section 5 to conclude the paper.

2. Finding Numerical Derivatives. Consider a symmetric function of the form $\Phi : \Omega \times \Omega \to \mathbb{R}$ for some $\Omega \subset \mathbb{R}^d$ and let $\mathcal{N}_\Phi$ be the *reproducing kernel* of a *native* Hilbert space [29] of Φ. It is proven in the same article that the native space $\mathcal{N}_\Phi$ for a given symmetric positive definite kernel Φ is unique if it exists, and it coincides with the closure of the space of finite linear combination of functions $\Phi(x, \cdot)$, $x \in \Omega$ under the inner product defined via

$$(\Phi(x, \cdot),\, \Phi(y, \cdot))_{\mathcal{N}_\Phi} = \Phi(x, y) \quad \text{for all} \quad x, y \in \Omega.$$

That is, for every fixed point $x \in \Omega$ and function $\Phi(x, \cdot)$ belongs to $\mathcal{N}_\Phi$, every $f \in \mathcal{N}_\Phi$ can be recovered by an inner product of the form $f(x) = \langle f, \Phi(x, \cdot) \rangle$, $x \in \Omega$. For a detailed treatise of reproducing kernel Hilbert spaces see Aronszajn [3] or Meschkowski [21].

To begin, we reconstruct multivariate functions from unstructured data by a multiscale technique. The basic concepts of this technique were first investigated by Opfer [23]. The implementation of MSK is out of the scope of this paper and the developments of MSK are only sketched here. We refer the reader to the original dissertation of Opfer for the details.

A function $\varphi : \mathbb{R}^d \to \mathbb{R}$ is called *refinable* if there is a sequence $\{\mathfrak{h}_k\}_{k \in \mathbb{Z}^d}$ of real numbers such that

$$\varphi = \sum_{k \in \mathbb{Z}^d} \mathfrak{h}_k \varphi(2 \cdot - k). \tag{2.1}$$

For every level-$j \in \mathbb{Z}$ we define the *shift invariant space*

$$\mathcal{V}_j := \Big\{ \sum_{k\in\mathbb{Z}^d} c_k \varphi(2^j \cdot -k) : c_k \in \mathbb{R}, \sum_{k\in\mathbb{Z}^d} (c_k)^2 < \infty \Big\}. \tag{2.2}$$

By standard wavelet arguments it follows from (2.1) that the spaces $\{\mathcal{V}_j\}_{j\in\mathbb{Z}}$ form a nested sequence, i.e. $\mathcal{V}_0 \subset \mathcal{V}_1 \subset \cdots \subset \mathcal{V}_u$. The main idea here involves several levels of $\mathcal{V}_j$ in *one* reconstruction scheme.

Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a function in $L^2(\mathbb{R}^d)$ with decay $\varphi(x) = \mathcal{O}\big((1+\|x\|)^{-(d+1)/2}\big)$. Let $u \geq 0$ be a fixed integer, $\sigma > d/2$ be a positive real number. Then the kernel $\Phi_\sigma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ given by

$$\Phi_\sigma(x,y) := \sum_{j=0}^{u} \lambda_\sigma^j \underbrace{\Big(\sum_{k\in\mathbb{Z}^d} \varphi(2^j x - k)\varphi(2^j y - k) \Big)}_{\Phi_{\sigma,j}} \text{ where } \lambda_\sigma := 2^{d-2\sigma}, \tag{2.3}$$

is called a *multiscale kernel* (MSK).

THEOREM 2.1. *[23, Theorem 5.4] Every MSK in the form of* (2.3) *is positive semidefinite. Let $B_\rho(c)$ be a ball of radius ρ with center $c \in \mathbb{R}^d$ such that* $\operatorname{supp}(\varphi) \subset B_\rho(c)$. *If the point set $X \subset \mathbb{R}^d$ satisfies*

$$h_{X,\min} := \min_{i\neq j} \|x_i - x_j\|_2 > \rho\, 2^{-u+1}, \tag{2.4}$$

then the matrix $A_X := \big(\Phi_\sigma(x_i,x_k)\big)_{1\leq i,k\leq n}$ is positive definite.

In this paper, we are mainly interested in compactly supported refinable functions φ that clearly satisfy the decay condition required in the Theorem 2.1. The resulting MSK are therefore positive definite.

We can find to any given data Y an interpolant of the form (1.1) by solving a sparse symmetric linear collocation system for $\beta \in \mathbb{R}^n$,

$$y_j = \sum_{i=1}^{n} \beta_i \Phi_\sigma(x_i, x_j), \quad 1 \leq j \leq n. \tag{2.5}$$

Theorem 2.1 implies that (2.5) has a unique solution if the integer $u = u(h_{X,\min})$ is large enough with respect to the density of the data points X. The MSK scheme is based on the following idea: The kernel representation can be decomposed into a frame representation due to the specially designed structure of Φ_σ. Firstly, $s \in \mathcal{N}_\Phi$ is decomposed into a sequence of functions $s_j \in \mathcal{V}_j$,

$$s = \sum_{i=1}^{n} \beta_i \Phi_\sigma(x_i,\cdot) = \sum_{i=1}^{n} \beta_i \sum_{j=0}^{u} \lambda_\sigma^j \Phi_{\sigma,j}(x_i,\cdot) = \sum_{j=0}^{u} \lambda_\sigma^j \underbrace{\sum_{i=1}^{n} \beta_i \Phi_{\sigma,j}(x_i,\cdot)}_{s_j} = \sum_{j=0}^{u} \lambda_\sigma^j s_j, \tag{2.6}$$

such that each $s_j \in \mathcal{V}_j$ can be further decomposed into

$$s_j = \sum_{i=1}^{n} \beta_i \Phi_{\sigma,j}(x_i,\cdot) = \sum_{k\in\mathbb{Z}^d} \underbrace{\Big(\sum_{i=1}^{n} \beta_i \varphi(2^j x_i - k) \Big)}_{c_k^j} \varphi(2^j \cdot - k) = \sum_{k\in\mathbb{Z}^d} c_k^j \varphi(2^j \cdot - k). \tag{2.7}$$

Combining (2.6) and (2.7) gives us the frame representation in the form of (1.2). Functions in lower levels capture the smooth structure of f while the higher levels contain the fine structure of f, including noise. Furthermore, the refinability of the function φ allows the *frame coefficients* c_k^j for $0 \le j \le u-1$ to be computed via

$$c_k^j = \lambda_\sigma^{-j} \sum_{\mu \in \mathbb{Z}^d} \mathfrak{h}_{\mu-2k} c_\mu^{j+1}, \quad k \in \mathbb{Z}^d.$$

Computation of frame coefficients c_k^j requires a nearest neighbor search, e.g. kd-tree [35, Chapter 14], to locate all $x \in X$ inside the support of $\varphi(2^u \cdot -k)$. Note that the number of nonzero c_k^j is finite due to the fact that $|X|$ is finite and φ is compactly supported. The native space $\mathcal{N}_\Phi$ and each $\mathcal{V}_j$ in (2.2) can be equipped with a norm, respectively,

$$\|s\|_{\mathcal{N}_\Phi}^2 = \sum_{j=0}^{u} \lambda_\sigma^{-j} \|s_j\|_{\mathcal{V}_j}^2 \quad \text{and} \quad \|s_j\|_{\mathcal{V}_j}^2 = \sum_{k \in \mathbb{Z}^d} (c_k^j)^2.$$

Let $h_{X,\Omega}$ denote the fill distance of the data points $X \subset \Omega$ given by

$$h_{X,\Omega} := \sup_{y \in \overline{\Omega}} \inf_{x_i \in X_h} \|y - x_i\|_2.$$

If φ satisfies certain smoothness and decay properties, then $\mathcal{N}_\Phi \simeq W^{\sigma,2}$ are norm equivalent and the interpolant obtained by MSK satisfies the standard native space error bound:

THEOREM 2.2. *[23, Theorem 5.21] Let the multiscale kernel Φ_σ be constructed with a scaling function φ of an r-regular multiscale analysis of $L^2(\Omega^d)$ with $r > d/2$. Fix an σ with $d/2 < \sigma < r$. Further we assume that $X := \{x_1, \ldots, x_n\} \subset \Omega$ is a set of points with fill distance $h_{X,\Omega}$ where $\Omega \subset \mathbb{R}^d$ is a compact set with Lipschitz boundary which satisfies an interior cone condition. Let $f \in H^\sigma(\mathbb{R}^d)$ and s be the interpolant. Let $1 \le q \le \infty$ and $\gamma = (\gamma_1, \ldots, \gamma_d)$ be a multi-index such that $|\gamma| < \lfloor \sigma \rfloor - d/2$. Then, there is a constant $C > 0$ independent of f and $h_{X,\Omega}$ such that*

$$\|s - f\|_{W^{|\gamma|,q}(\Omega)} \le C_1 h_{X,\Omega}^{\sigma - |\gamma| - d\left(1/2 - 1/q\right)_+} \|f\|_{\mathcal{N}_\Phi},$$

where $(x)_+ = x$ if $x \ge 0$ and $(x)_+ = 0$ if $x < 0$.

2.1. Noise Data. Let us assume we have points $X := \{x_1, \ldots, x_n\} \subset \Omega \subset \mathbb{R}^d$ and noisy data

$$Y_\eta := \{\widetilde{y}_1, \ldots, \widetilde{y}_n\} \subset \mathbb{R},$$

where

$$\widetilde{y}_i = y_i + \delta_i = f(x_i) + \eta(x_i),$$

and δ_i are random noise. The noise function η here is not necessarily classically differentiable or even continuous. Assume that we obtain an interpolant in the frame representation

$$s_{\delta,X} = \sum_{j=0}^{u} \lambda_\sigma^j s_{\delta,X,j} = \sum_{j=0}^{u} \sum_{k \in \mathbb{Z}^d} \lambda_\sigma^j c_k^j \varphi(2^j \cdot -k), \tag{2.8}$$

for some noisy data (X, Y_η) by MSK with the following conditions satisfied.

ASSUMPTION 2.3. *The MSK in* (2.3) *is constructed by*

1. $\sigma \geq 2$ *and* $\sigma > \frac{d}{2}$,
2. *a* r*-regular* φ *smooth enough such that* $r > \left(2 + \frac{d}{2}\right)$, *i.e.,* $\varphi \in C^r(\Omega)$ *with compact support up to order* r, *and*
3. *for any given data points* X, $u = \left\lceil 1 + \log_2 \frac{\rho}{h_{X,\min}} \right\rceil$ *where* $h_{X,\min}$ *is given in* (2.4) *and* ρ *as in Theorem 2.1.*

The reasons for the above assumptions will soon become clear when we look at the error estimates in Section 3.1. Throughout the paper, let γ with $|\gamma| = \gamma_1 + \ldots + \gamma_d = 1$ be a multi-index. Our interest is to approximate or reconstruct the derivatives of f from the noisy data Y_η via

$$(X,\, Y_\eta) \longrightarrow D^\gamma f.$$

From the frame representation (2.8), the numerical derivatives are given by

$$D^\gamma s_{\delta,X} = \sum_{j=0}^{u} \lambda_\sigma^j \, D^\gamma \, s_{\delta,X,j} = \sum_{j=0}^{u} \sum_{k \in \mathbb{Z}^d} \lambda_\sigma^j c_k^j \, D^\gamma \varphi(2^j \cdot - k). \tag{2.9}$$

This numerical procedure is highly unstable. Since the input data Y_η contains noise, the resulting approximated derivatives $D^\gamma s_{\delta,X}$ will contain large error and therefore are not trustworth. We select a subset of frame coefficients $\{r_k^j\} \subset \{c_k^j\}$ to *regularize* the numerical derivatives.

Any regularized interpolant g to $s_{\delta,X}$ is in the form of

$$g = \sum_{j=0}^{u} \sum_{k \in \mathbb{Z}^d} \lambda_\sigma^j r_k^j \, \varphi(2^j \cdot - k) \text{ where } r_k^j \in \{0,\, c_k^j\}. \tag{2.10}$$

For some threshold $t_\sigma(j) > 0$ for $0 \leq j \leq u$ and a fixed regularization parameter α, the regularized interpolant is defined to be

$$s_\alpha = \sum_{j=0}^{u} \sum_{k \in \mathbb{Z}^d} \lambda_\sigma^j \, r_k^j \, \varphi(2^j \cdot - k) \;\text{ such that }\; r_k^j = \begin{cases} c_k^j & \text{if } |c_k^j| > t_\sigma(j)\, \alpha, \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

For practical problems, the optimal regularization parameter α^* is not attainable unless η is known *a priori* . In the next section, we specify our choice of threshold $t_\sigma(j)$ using the Tikhonov regularization method. After giving a concrete formula of the threshold $t_\sigma(j)$, we make sure the errors in interpolation and in the gradient of the regularized interpolant in (2.11) is both bounded and well behaved for some suitable α.

3. Regularization. The classical Tikhonov regularization method [31] is a common tool for finding solution from an unstable system. Using some *a priori* choice strategy for regularization parameters, Hofmann and Yamamoto [18] prove convergence rates for the Tikhonov regularization method. Despite the differences with the classical problem, we seek a *regularized interpolant* s_α to $s_{\delta,X}$ (considered to be fixed here) by the Tikhonov regularization method. For any

$$\tilde{g} = \sum_{j=0}^{u} \sum_{k \in \mathbb{Z}^d} \lambda_\sigma^j \tilde{r}_k^j \, \varphi(2^j \cdot - k) \in \mathcal{V}_u,$$

we define the *error measure* by

$$E(\tilde{g}) = E(\tilde{g}; s_{\delta,X}) := \|s_{\delta,X}\|^2_{\mathcal{N}_\Phi} - \|g\|^2_{\mathcal{N}_\Phi} = \sum_{j=0}^{u} \sum_{k\in\mathbb{Z}^d} \lambda_\sigma^{-j}\big((c_k^j)^2 - (\tilde{r}_k^j)^2\big), \tag{3.1}$$

and the *roughness measure* by

$$R(\tilde{g}) := \sum_{j=0}^{u} \sum_{k\in\mathbb{Z}^d} \lambda_\sigma^j |\tilde{r}_k^j|\, |\varphi(2^j\cdot - k)|_{W^{2,2}(\Omega)}, \tag{3.2}$$

such that $|\tilde{g}|^2_{W^{2,2}(\Omega)} \le R(\tilde{g})$ for any $\tilde{g} \in \mathcal{V}_u$. The error measure depends on the interpolant $s_{\delta,X}$ but both are independent of α.

Given any regularization parameter $\alpha \ge 0$ (consider to be fixed here), the *regularized interpolant* s_α is defined to be the minimizer of $E(\cdot) + \alpha\, R(\cdot)$ over all functions in the form of (2.10), i.e.,

$$E(s_\alpha) + \alpha\, R(s_\alpha) = \inf\Big\{E(g) + \alpha\, R(g) \;\text{ for all }\; g \text{ as in (2.10)}\Big\}. \tag{3.3}$$

Although the number of nonzero functions in the form of (2.10) is finite, we have the following theorem to simplify our selection process.

THEOREM 3.1. *For any given $\alpha \ge 0$ the optimizer to* (3.3) *is given by* (2.11) *with*

$$t_\sigma(j) := \big(2^{d-2\sigma+4}\, |\varphi|^2_{W_{2,2}}\big)^j < \infty \;\; \textit{for all} \;\; 0 \le j \le u < \infty.$$

Proof. First by changing variables, we obtain

$$|\varphi(2^j\cdot - k)|^2_{W_{2,2}(\Omega)} = \Big\| \sum_{|\gamma|=2} D^\gamma \varphi(2^j\cdot - k)\Big\|^2_{L^2(\Omega)} = 2^{j(4-d)}\, |\varphi|^2_{W_{2,2}(\Omega)}. \tag{3.4}$$

For any g in the form of (2.10), we have

$$\begin{aligned} E(g) + \alpha\, R(g) &= \sum_{j=0}^{u} \sum_{k\in\mathbb{Z}^d} \Big(\big(\lambda_\sigma^{-j}(c_k^j)^2 - (r_k^j)^2\big) + \alpha\, \lambda_\sigma^j\, |r_k^j|\, 2^{j(4-d)}\, |\varphi|^2_{W_{2,2}(\Omega)}\Big) \\ &= \underbrace{\left(\sum_{j=0}^{u} \sum_{k\in\mathbb{Z}^d} \lambda_\sigma^{-j}(c_k^j)^2\right)}_{=\|s_{\delta,X}\|^2_{\Phi_\sigma}} - \left(\sum_{j=0}^{u} \sum_{k\in\mathbb{Z}^d} \lambda_\sigma^{-j}(r_k^j)^2 - \alpha\, \lambda_\sigma^j\, |r_k^j|\, 2^{j(4-d)}\, |\varphi|^2_{W_{2,2}(\Omega)}\right). \end{aligned}$$

Since $\|s_{\delta,X}\|^2_{\Phi_\sigma}$ is a fixed quantity, the minimizer of (3.3) corresponds to the following condition on r_k^j:

$$\lambda_\sigma^{-j}(r_k^j)^2 - \alpha\, \lambda_\sigma^j\, |r_k^j|\, 2^{j(4-d)}\, |\varphi|^2_{W_{2,2}(\Omega)} > 0.$$

After simplification, we obtain $(r_k^j)^2 > t_\sigma(j)|r_k^j|\alpha$. □

Once α is determined, Theorem 3.1 allows us to select $\{r_k^j\}$ from $\{c_k^j\}$ and construct the regularized interpolant and its derivatives.

3.1. Error Estimate. In general, interpolation does not make sense on $L^2(\Omega)$ and there are many possibilities of projecting $L^2(\Omega)$ to $\mathcal{N}_\Phi$. Moreover, there are many new results on interpolation in cases where f is not in the native space [22, 28]. For our problem, we will define the necessary projection by interpolation.

Let $\Omega \subset \mathbb{R}^d$ be a domain satisfying the conditions in Theorem 2.2. Suppose that the multiscale kernel Φ_σ also satisfies Assumption 2.3 and $f \in \mathcal{N}_\Phi = H^\sigma(\Omega)$. For any fixed center X and noise function $\eta \in L^2(\Omega) \cap C(\Omega)$, the *noise level* is defined as

$$\delta := \sup_{x\in\Omega} |\eta(x)|.$$

It is easy to verify that $\|\eta\|_{L^2(\Omega)} \le V^{1/2}(\Omega)\,\delta$, where $V(\Omega)$ is the volume of $\Omega \subset \mathbb{R}^d$. The noisy input data for interpolation at the points $X \subset \Omega$ is given by $Y_\eta := (f+\eta)\big|_X$ under the assumption that f and η are both well defined at all points $x \in \Omega$.

We define a finite dimensional subspace $V_X \subset \mathcal{N}_\Phi$ to be the span of $\Phi_\sigma(z,\,\cdot)$ and $V_X^{(\gamma)}$ to be the span of $D^\gamma \Phi_\sigma(z,\,\cdot)$ where differentiation acts upon the second variable of Φ_σ for all $z \in X$. Furthermore, we define a *projection map*

$$P_X : L^2(\Omega) \cap C(\Omega) \to \mathbb{R}^{|X|} \;\text{ such that }\; P_X f = \{f(x) :\; x \in X\}$$

that extracts discrete values from a function in $L^2(\Omega)\cap C(\Omega)$ at X so that interpolation is possible and makes sense, and an *interpolation map*

$$I_X : \mathbb{R}^{|X|} \to V_X \;\text{ such that }\; I_X P_X f = I_X f \;\text{ for all }\; f \in \mathcal{N}_\Phi,$$

which maps discrete function values at X to a function in V_X by interpolation using MSK. Last, we define a *truncation map*,

$$T_\alpha : \{1\}^{\mathbb{N}\times\mathbb{Z}^d} \to \{0,\,1\}^{\mathbb{N}\times\mathbb{Z}^d} \;\text{ for all }\; \alpha \ge 0$$

that smoothes out functions by truncating some of their frame coefficients. Furthermore, when no confusion arises, we treat T_α as a map from V_X and $V_X^{(\gamma)}$ onto themselves in the sense that,

$$T_\alpha\left(\sum_{j=0}^{u}\sum_{k\in\mathbb{Z}^d} \lambda_\sigma^j\, c_k^j\, \phi(2^j\cdot - k)\right) := \sum_{j=0}^{u}\sum_{k\in\mathbb{Z}^d} \lambda_\sigma^j\, T_\alpha(c_k^j)\, \phi(2^j\cdot - k),\;\; \phi = \{\varphi,\, D^\gamma\varphi\}.$$

The truncation map T_α, as in (2.11), is a nonlinear map whose actual form depends on the parameter α and the data $(X,\, Y_\eta)$. It can also be interpreted as a countable set $\{\tau_k^j\} \subset \{0,\,1\}^{\mathbb{N}\times\mathbb{Z}^d}$ such that $T_\alpha(c_k^j) = \tau_k^j(\alpha) c_k^j = r_k^j(\alpha)$ where

$$\tau_k^j = \tau_k^j(\alpha) = \begin{cases} 1 & \text{if } r_k^j = c_k^j, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Since the number of nonzero $c_k^j \in \{0,\,1\}^{\mathbb{N}\times\mathbb{Z}^d}$ is finite, there are infinitely many $c_k^j = 0$ and the corresponding $\tau_k^j = 1$ because $r_k^j = 0 = c_k^j$ for all $\alpha \ge 0$ by (3.5). Thus, there are infinitely many $\tau_k^j = 1$ (frame coefficients being kept) and only a finite number of $\tau_k^j = 0$ (frame coefficients being truncated) for the selected frame coefficients.

With the newly introduced notation, the unknown full interpolant can be expressed by $s := I_X P_X f$. Furthermore, we can write the regularized interpolant in Theorem 3.1 as

$$s_{\delta,X} := I_X P_X (f - \eta) \quad\text{and}\quad s_\alpha = T_\alpha s_{\delta,X}.$$

Moreover, Equation (2.11) can be restated as

$$s_\alpha = T_\alpha s_{\delta,X} = \sum_{j=0}^{u} \sum_{k\in\mathbb{Z}^d} \lambda_\sigma^j \, \tau_k^j c_k^j \, \varphi(2^j \cdot - k).$$

Without any extra assumptions on the noise function η, the threshold $t_\sigma(j)$ and the data points X, the truncation map has the following properties.

PROPOSITION 3.2. *Let $|\gamma| = 1$ and $\mathrm{nz}_j(\cdot)$ be a function with respect to j that returns the number of zero elements in the level-j of a set in $\{0,1\}^{\mathbb{N}\times\mathbb{Z}^d}$. Denote the $L^2(\Omega)$-induced norm for maps on V_X by $\|\cdot\|_{L^2(\Omega)}$ and define*

$$u_\alpha := \sup\left\{ j \,\middle|\, \tau_k^j \neq 0 \text{ for some } k\in\mathbb{Z}^d, 0\le j\le u \right\}, \tag{3.6}$$

to be the maximum nonzero frame level after truncation. Then the truncation map T_α satisfies:

1. *$\|T_\alpha\|_{L^2(\Omega)} = \|T_0 - T_\alpha\|_{L^2(\Omega)} = 1$ for $\alpha > 0$.*
2. *$\|D^\gamma T_\alpha\|_{L^2(\Omega)} = \|T_\alpha D^\gamma\|_{L^2(\Omega)} = 2^{u_\alpha} \|D^\gamma\varphi\|_{L^2(\Omega)} \|\varphi\|_{L^2(\Omega)}^{-1}$.*
3. *For any given data (X, Y_η), the number of frame coefficients being truncated by T_α, denoted by $\mathrm{nz}_j(1-\tau_k^j(\alpha)) < \infty$, is a bounded nondecreasing simple function in α and $\mathrm{nz}_j(1-\tau_k^j(0)) = 0$.*

Proof. The perfect candidate to evaluate the above norms is the scaled function in the frame. For each nested space $\mathcal{V}_j$ $(0\le j\le u)$, such function is given by

$$g_{j,k} = \left(2^{jd/2}\,\|\varphi\|_{L^2(\Omega)}^{-1}\right)\varphi(2^j\cdot - k) \in \mathcal{V}_j,\ 0\le j\le u,$$

such that $\|g_{j,k}\|_{L^2(\Omega)} = 1$ and $\|D^\gamma g_{j,k}\|_{L^2(\Omega)} = 2^j \|D^\gamma\varphi\|_{L^2(\Omega)} \|\varphi\|_{L^2(\Omega)}^{-1}$.

For Proposition 3.2-1 follows directly from the fact that $T_\alpha \neq 0$ for all $\alpha \ge 0$; there exists some (j_1,k_1) and (j_2,k_2) such that $\tau_{k_1}^{j_1} = 1$ and $\tau_{k_2}^{j_2} = 0$ for $0\le j_i\le u$ and $k_i\in\mathbb{Z}^d$ corresponding to a frame coefficient that is kept and truncated by T_α, respectively. Hence, we have

$$\|T_\alpha I_X P_X g_{j_1,k_1}\|_{L^2(\Omega)} = 1, \text{ and } \|(T_0 - T_\alpha) I_X P_X g_{j_2,k_2}\|_{L^2(\Omega)} = 1.$$

To prove Proposition 3.2-2, we first note that the differential operator acts on each φ independently as in (2.9); thus, c_k^j and τ_k^j are independent of the truncation process. *Differentiation after truncation* is the same as *truncation after differentiation*, namely we have $D^\gamma T_\alpha s_j = T_\alpha D^\gamma s_j$ for all $s_j\in\mathcal{V}_j$. For numerical efficiency, the operation $D^\gamma T_\alpha$ is preferred for efficiency.

Since $\|D^\gamma\varphi\|_{L^2(\Omega)} \|\varphi\|_{L^2(\Omega)}^{-1}$ is a fixed quantity once φ is fixed, without regularization the noise in the level-j will be greatly amplified as expected,

$$\|D^\gamma I_X P_X g_{j,k}\|_{L^2(\Omega)} = \|D^\gamma g_{j,k}\|_{L^2(\Omega)} \le 2^j \|D^\gamma\varphi\|_{L^2(\Omega)} \|\varphi\|_{L^2(\Omega)}^{-1}. \tag{3.7}$$

Let u_α be the highest nonzero frame level appearing in the regularized interpolant as in (3.6). Applying the regularization map T_α will "cut off" all levels higher than u_α exclusively and we arrive at the conclusion.

Lastly, Proposition 3.2-3 follows from the fact that the number of nonzero c_k^j is finite and no regularization is applied when $\alpha = 0$. □

We now turn our focus to the error estimate for $\|f - s_\alpha\|$. First of all,

$$\begin{aligned}\|f - s_\alpha\|_{L^2(\Omega)} &\le \|f - I_X P_X f\|_{L^2(\Omega)} + \|I_X P_X f - s_{\delta,X}\|_{L^2(\Omega)} + \|s_{\delta,X} - T_\alpha s_{\delta,X}\|_{L^2(\Omega)} \\ &= \underbrace{\|f - I_X P_X f\|_{L^2(\Omega)}}_{\text{interp. error}} + \underbrace{\|I_X P_X \eta\|_{L^2(\Omega)}}_{\text{noise}} + \underbrace{\|(T_0 - T_\alpha) s_{\delta,X}\|_{L^2(\Omega)}}_{\text{reg. error}}.\end{aligned}$$

The last inequality uses the fact that

$$\|I_X P_X f - s_{\delta,X}\| = \|I_X P_X f - I_X P_X (f - \eta)\| = \|I_X P_X \eta\|.$$

By Theorem 2.2 with $q = 2$ and $|\gamma| = 0$, the first term (interpolation error) can be bounded by

$$\|I_X P_X f - f\|_{L^2(\Omega)} \le C_1 h_{X,\Omega}^{\sigma} \|f\|_{\mathcal{N}_\Phi},$$

and the second term (noise) is bounded by our assumption on η,

$$\|I_X \eta\|_{L^2(\Omega)} \le V^{1/2}(\Omega)\, \delta.$$

It is straight forward to verify that

$$\|s_j\|_{L^2(\Omega)}^2 \le 2^{-jd} \|\varphi\|_{L^2(\Omega)}^2 \|s_j\|_{\mathcal{V}_j}^2 \quad \text{for all} \quad s_j \in \mathcal{V}_j. \tag{3.8}$$

For the third term (regularization error), by Theorem 3.1 and (3.8) we have

$$\begin{aligned}\|(T_0 - T_\alpha) s_{\delta,X}\|_{L^2(\Omega)}^2 &\le \sum_{j=0}^{u} \|(T_0 - T_\alpha) s_{\delta,X,j}\|_{L^2(\Omega)}^2 \qquad (3.9) \\ &\le \|\varphi\|_{L^2(\Omega)}^2 \sum_{j=0}^{u} \sum_{k \in \mathbb{Z}^d} 2^{-jd} \big((1 - \tau_k^j) c_k^j\big)^2 \\ &\le \|\varphi\|_{L^2(\Omega)}^2 \sum_{j=0}^{u} 2^{-jd} \mathrm{nz}_j (1 - \tau_k^j)\, t_\sigma(j)^2 \alpha^2 \\ &\le \sum_{j=0}^{u} 2^{-2(\sigma-2)j} \mathrm{nz}_j (1 - \tau_k^j) |\varphi|_{W_{2,2}}^{2j} \|\varphi\|_{L^2(\Omega)}^{2(j+1)} \alpha^2 \\ &:= \big(C_2(\alpha)\alpha\big)^2\end{aligned}$$

An immediate fact from Proposition 3.2-3 is that $C_2(\alpha)$ is a bounded positive nondecreasing simple function with $C_2(0) = 0$.

For the error in the gradient, we have

$$\begin{aligned}\|\nabla f - \nabla s_\alpha\|_{L^2(\Omega)} &\le \|\nabla f - \nabla I_X P_X f\|_{L^2(\Omega)} + \|\nabla I_X P_X f - \nabla T_\alpha I_X f\|_{L^2(\Omega)} + \|\nabla T_\alpha I_X P_X f - \nabla s_\alpha\|_{L^2(\Omega)} \\ &\le \underbrace{\|\nabla f - \nabla I_X P_X f\|_{L^2(\Omega)}}_{\text{interp. error}} + \underbrace{\|\nabla (T_0 - T_\alpha) I_X P_X f\|_{L^2(\Omega)}}_{\text{reg. error}} + \underbrace{\|\nabla T_\alpha I_X P_X P_X \eta\|_{L^2(\Omega)}}_{\text{noise}}\end{aligned}$$

Using Theorem 2.2 with $q = 2$ and $|\gamma| = 1$, the interpolation error in gradient is again bounded by

$$\|\nabla I_X P_X f - \nabla f\|_{L^2(\Omega)} \le C_1 h_{X,\Omega}^{\sigma-1} \|f\|_{\mathcal{N}_\Phi}.$$

Next, we need a stronger assumption than $\sigma \geq 2$ such that $\mathcal{N}_\Phi \subseteq W^{2,2}(\Omega)$ to make use of an inequality in [1, Theorem 4.14]: For any $0 < \epsilon_0$ there exists a constant $C_3 = C_3(\epsilon_0, \Omega, d) > 0$ such that for $g \in W^{2,2}(\Omega)$ and for all $0 < \epsilon < \epsilon_0$,

$$\|\nabla g\|_{L^2(\Omega)} \leq C_3\big(\epsilon\, |g|_{W^{2,2}(\Omega)} + \epsilon^{-1}\|g\|_{L^2(\Omega)}\big). \tag{3.10}$$

By assumption, the unknown function f is "smoother" than the random noise η. Hence, for all $\alpha \geq 0$ the following statement holds

$$\|\nabla(T_0 - T_\alpha) I_X P_X f\|_{L^2(\Omega)} \leq \|\nabla(T_0 - T_\alpha) s_{\delta,X}\|_{L^2(\Omega)}.$$

Similar to (3.9), by (3.4) we have

$$\begin{aligned}
|(T_0 - T_\alpha)s_{\delta,X}|_{W^{2,2}} &\leq \sum_{j=0}^{u} |(T_0 - T_\alpha)s_{\delta,X,j}|_{W^{2,2}} \\
&\leq |\varphi|^2_{W^{2,2}} \sum_{j=0}^{u} \sum_{k\in\mathbb{Z}^d} 2^{j(2-d/2)}\big((1-\tau_k^j)c_k^j\big)^2 \\
&\leq |\varphi|^2_{W^{2,2}} \sum_{j=0}^{u} 2^{j(2-d/2)} \mathrm{nz}_j (1-\tau_k^j)\, t_\sigma(j)\alpha \\
&\leq \sum_{j=0}^{u} 2^{(6-2\sigma+d/2)j} \mathrm{nz}_j (1-\tau_k^j) |\varphi|^{2(j+1)}_{W^{2,2}} \alpha \\
&:= C_4(\alpha)\alpha.
\end{aligned} \tag{3.11}$$

We choose $\epsilon = 1 < \epsilon_0$ for some fixed ϵ_0. Putting (3.9) and (3.11) into (3.10) yields

$$\|\nabla(T_0 - T_\alpha) I_X f\|_{L^2(\Omega)} \leq C_5(\alpha)\alpha.$$

Namely, $C_5(\alpha) = C_3(C_2(\alpha) + C_4(\alpha))$ that is a bounded positive nondecreasing simple function with $C_5(0) = 0$.

All the terms considered so far are stable. Last, but most importantly, we consider the error in gradient due to the presence of noise. By Proposition 3.2-2, if there exist some (j,k) such that $c_k^j \neq 0$ and $\tau_k^j = 1$, we have

$$\|\nabla T_\alpha I_X P_X \eta\|_{L^2(\Omega)} \leq 2^{d/2}\, 2^{u_\alpha}\, \|\nabla\varphi\|_{L^2(\Omega)}\, \|\varphi\|^{-1}_{L^2(\Omega)}\, V^{1/2}(\Omega)\, \delta := C_6(\alpha)\, \delta. \tag{3.12}$$

Otherwise $s_{\delta,X} = 0$, we clearly have $\|\nabla T_\alpha I_X P_X \eta\|_{L^2(\Omega)} = 0$ and $C_6(\alpha) = 0$.

The function $C_6(\alpha)$ in (3.12) is a bounded positive nonincreasing simple function. Since $2^u \geq \frac{2\rho}{h_{X,\min}}$ is the requirement of a positive definite kernel, the gradient error in (3.7) will blow up when one takes finer and finer data points if the noise level $\delta > 0$ is fixed and no regularization is applied.

We summarize all results by the following theorem.

THEOREM 3.3. *For any given data (X, Y_η), let s_α be the regularized interpolant obtained by a MSK satisfying Assumption 2.3 and regularized by Theorem 3.1. There exist a constant C_1, two bounded positive nondecreasing simple functions $C_2^{\nearrow}(\alpha) \geq C_5^{\nearrow}(\alpha)$ such that $C_2^{\nearrow}(0) = 0 = C_5^{\nearrow}(0)$, and a bounded nonnegative noincreasing simple function $C_6^{\searrow}(\alpha)$ with $C_6^{\searrow}(0) > 0$ such that the errors in regularized interpolant are bounded by*

$$\|f - s_\alpha\|_{L^2(\Omega)} \leq C_1 h^\sigma_{X,\Omega}\, \|f\|_{\mathcal{N}_\Phi} + V^{1/2}(\Omega)\, \delta + C_2^{\nearrow}(\alpha)\alpha, \tag{3.13}$$

and

$$\|\nabla f - \nabla s_\alpha\|_{L^2(\Omega)} \le C_1 h_{X,\Omega}^{\sigma-1} \|f\|_{\mathcal{N}_\Phi} + C_5^{\nearrow}(\alpha)\alpha + C_6^{\searrow}(\alpha)\,\delta, \tag{3.14}$$

for all $\alpha \ge 0$. *Furthermore, if the noise level* $\delta \ge K(f,\sigma)$, *there exists a* nonzero optimizer α^* *that minimizes the sum of the upper bounds in* (3.13) *and* (3.14).

Proof. For any given data $(X,\, Y_\eta)$, the minimizer α^* in the theorem is also a minimizer to the function

$$\big(C_2^{\nearrow}(\alpha) + C_5^{\nearrow}(\alpha)\big)\alpha + C_6^{\searrow}(\alpha)\,\delta. \tag{3.15}$$

By the properties of $C_2^{\nearrow}(\alpha)$ and $C_5^{\nearrow}(\alpha)$, we know that the term $\big(C_2^{\nearrow}(\alpha)+C_5^{\nearrow}(\alpha)\big)\alpha$ is a monotone increasing piecewise linear function. Its jump discontinuities are governed by the term $\mathrm{nz}_j(1-\tau_k^j(\alpha))$.

The terms $C_6^{\searrow}(\alpha)\delta$ is a nonnegative nonincreasing simple function having jump discontinuities at $0 =: \alpha_{u+1} < \alpha_u \le \ldots \le \alpha_0 < \infty$ where α_j is the infimum over α such that j-th level is completely truncated, i.e. for all $0 \le j \le u$

$$\alpha_j := \inf\{\, \alpha \,\big|\, r_k^j(\alpha) = \tau_k^j c_k^j = 0 \ \text{ for all } \ k \in \mathbb{Z}^d\}.$$

Define $\Delta_k G(\alpha) = G(\alpha_{u-k}) - G(\alpha_{u-k+1})$ for all $0 \le k < u$. If, for sufficiently large δ, the accumulated drop due to term $C_6^{\searrow}(\alpha)\delta$ is larger than the accumulated grow due to the term $\big(C_2^{\nearrow}(\alpha) + C_5^{\nearrow}(\alpha)\big)\alpha$, i.e.,

$$\delta > K(f,\sigma) := \min_{0\le j<u} \left\{ \sum_{k=1}^{j} \Delta_k \frac{(C_2^{\nearrow}(\alpha) + C_5^{\nearrow}(\alpha))\alpha}{C_6^{\searrow}(\alpha)} \right\}, \tag{3.16}$$

then an optimizer $\alpha^* > 0$ exists. □

To end this section, note that the constant term $K(f,\sigma)$ in (3.16) decreases as σ increases. If the unknown function f is sufficient smooth with respect to the noise level δ, our MSK scheme is able to regularize the interpolant. Consider $\delta < K(f,\sigma)$. These cases correspond to small noise levels that are negligible to our regularization technique. As shown in Section 5 when $\delta = 0$, while $\alpha^* = 0$ is the theoretical optimizer to (3.15), we would numerically obtain an approximation α_{LC} to α^* such that $0 < \alpha_{LC} < \epsilon_{mach}$ (machine epsilon). In these cases, we set the approximation $\alpha_{LC} = \varepsilon_{mach}$ to filter out extremely small frame coefficients for efficiency.

4. L-curve Method. The theoretical existence of α^* does not help us pinpointing its whereabouts. Choosing an optimal α^*, or an approximation α_{LC}, is a separate topic that will be considered in this section.

The L-curve (LC) method was investigated by Hansen and O'Leary [16] to regularize ill-posed systems under different values of the regularization parameter α. The knowledge of the noise level δ is *not necessary*. Vogel [32] shows that the L-curve regularization parameter selection method may fail to converge for a certain class of problems. In our numerical experiments, however, we find that the L-curve method provides a stable algorithm to find the regularization parameter α.

Our version of the L-curve method is derived from simplifying both measures in (3.1) and (3.2) for the ease of computation. First, we order the frame coefficients c_k^j by defining an ordered set,

$$\{\,(\xi_\ell,\, \eta_\ell)\,\}_{\ell=1}^{\mathrm{nz}(c_k^j)} = \left\{ \left(\left\| c_k^j \varphi(2^j - k) \right\|_{L^2(\Omega)}^2,\, R\big(c_k^j \varphi(2^j - k)\big) \right) \,:\, c_k^j \ne 0 \right\}_{0\le j\le u,\, k\in\mathbb{Z}^d}$$

Level-j	0	1	2	3	4	5	6	7
$\lvert c_k^j\rvert > 0$	64	144	400	1296	4624	17424	26896	26896
$\lvert c_k^j\rvert > \epsilon_{mach}$	56	121	361	1225	4489	17161	24025	24025
$\lvert r_k^j\rvert > 0$ by α_{LC}	49	121	350	0	0	0	0	0
$\lvert r_k^j\rvert > 0$ by ϵ_{mach}	49	121	361	1204	0	0	0	0

TABLE 4.1
MSK(3,3) frame coefficients among all levels on a 41×41 *uniform grids for Section 5.1.*

such that η_ℓ/ξ_ℓ forms a monotone nondecreasing sequence where $\mathrm{nz}(\cdot)$ returns the number of nonzero elements in the set and $R(\cdot)$ is the roughness measure in (3.2). Then we compute a finite set of points in $\mathbb{R}^2$ by

$$L = \left\{ \left(\|s_{\delta,X}\|^2_{\Phi_\sigma} - \sum_{\ell=0}^{p} \xi_\ell,\ \sum_{\ell=0}^{p} \eta_\ell \right) \subset \mathbb{R}^2,\ p = 0, 1, \ldots, \mathrm{nz}(c_k^j) \right\},$$

which is known as the *L-curve*.

A suitable regularization parameter α_{LC} is the one near the *corner* on a log-log scale of the L-curve [15]. In numerical computation, finite difference schemes are applied to (the log-values of) these discrete points in order to approximate the curvature of the L-curve. The point with maximum curvature will be labeled as the corner of the L-curve. For numerical efficiency, we impose an extra condition that

$$\alpha_{LC} \geq \epsilon_{mach}.$$

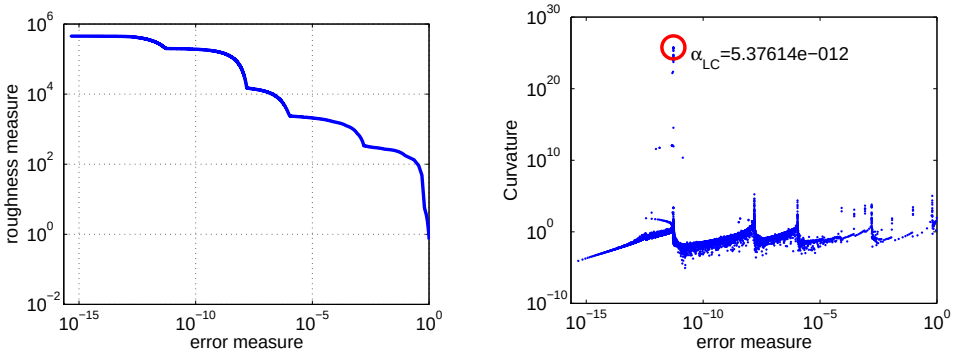
We show some results with the L-curve method in Figure 4.1. The L-curve is shown in Figure 4.1(a) with a corner at $\alpha_{LC} = 5.3761\text{e-}12$. This value is chosen from the curvature of the L-curve, see Figure 4.1(b).

The number of nonzero frame coefficients in the regularized interpolant s_α is 1735 and 520 for $\alpha = \epsilon_{mach}$ and $\alpha = \alpha_{LC}$, respectively. Figure 4.2(a) for ϵ_{mach} and Figure 4.2(b) for α_{LC} show all $|c_k^j|$ and label the selected r_k^j in boldface dots. All c_k^j are ordered by levels, from level-0 on the left to level-u on the right. In both cases, only the c_k^j in the lower few levels with large absolute values are chosen.

At first glance, the computation of *all* nonzero c_k^j may look tremendous. In fact, we are showing all 77744 nonzero frame coefficients in Figure 4.2 but some are extremely small, e.g. 2.4e-42. If we are only interested in frame coefficients whose sizes are larger than machine epsilon, we are looking at 71463 coefficients. The distribution of the frame coefficients among all levels are in Table 4.1. After regularization, the maximum levels appears in $\{r_k^j\}$ are $u_\alpha = 2$ for $\alpha = \alpha_{LC}$ and $u_\alpha = 3$ for $\alpha = \epsilon_{mach}$, readers may already see how this can be computed efficiently.

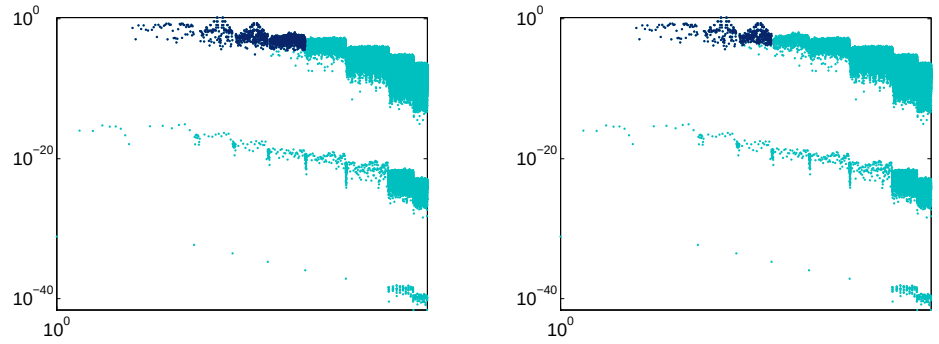
Our L-curve only makes use of the local property of each function $c_k^j \varphi(2^j \cdot - k)$. Pre-truncation does not affect the final outcome. One could pick an intermediate value $0 < \upsilon < u$ and compute frame coefficients up to level-υ only. A safeguard of this approach is that the maximum level appearing in the regularized interpolant should be strictly less than υ. If this is not the case, one can compute the frame coefficients for level-$(\upsilon+1)$ and reapply the L-curve method.

5. Numerical Comparison and Demonstration. We demonstrate some bivariate examples in this section. All codes are written in MATLAB. Random noise is



(a) *L-Curve in log-log scale.* **(b) *Corner of L-Curve.***

FIG. 4.1. *L-curve method applied to MSK(3,3) in Section 5.1 with* $\delta = 1.018 \times 10^{-3}$ *on a* 41×41 *uniform grids.*



(a) *1735 frame coefficients for* $\alpha = \epsilon_{mach}$**.** **(b) *520 frame coefficients for*** $\alpha = \alpha_{LC}$**.**

FIG. 4.2. *Selected frame coefficients* $\{r_k^j\} \subset \{c_k^j\}$ *corresponds to Figure 4.1.*

generated by the built-in routine RAND with STATE reset to 0. Generated random numbers are scaled to $[-1, 1]$ and multiplied by the noise level δ. For problem in $\mathbb{R}^2$, tested values for σ are 2 or 3, see Assumption 2.3. The multiscale kernel Φ_σ in (2.3) is constructed with the univariate B-spline of order m defined on the knot sequence $[0, 1, \ldots, m]$, denoted by b_m, see [9],

$$\varphi(x,\, y) = b_m(x)\, b_m(y) \;\text{ such that }\; x, y \in \mathbb{R},\; m = \{3, 4\},$$

that fulfills all assumptions in the previous discussion. Values of σ and m are specified by the notation $MSK(m,\, \sigma)$ throughout the section.

5.1. Comparison with TPS-based Method. The recent work of Wei et al. [34] uses the thin plate spline (TPS) to compute numerical derivatives. The presented TPS-based method requires triangular partitions of data points; the authors claim that the method can become truly mesh-free with additional assumptions. Two regularization parameters are studied in the same paper: $\alpha_1 = \delta^2$ obtained by *a priori* rule and $\alpha_2(\delta)$ obtained by Morozov's discrepancy principle. We denote them by *TPS-AP* and *TPS-DP*, respectively, hereafter. TPS-DP is reported to be the more

	$\delta = 1.018 \times 10^{-3}$			$\delta = 1.020 \times 10^{-2}$		
Method	$\varepsilon(s_\alpha)$	$\varepsilon(\nabla s_\alpha)$	α_{LC}	$\varepsilon(s_\alpha)$	$\varepsilon(\nabla s_\alpha)$	α_{LC}
TPS-AP	0.0028	0.0195	–	0.0699	0.3736	–
TPS-DP	0.0019	0.0157	–	0.0100	0.0659	–
MSK(3,2)	0.0011	0.0072	1.5543e-11	0.0042	0.0310	4.4899e-11
MSK(3,3)	0.0010	0.0075	9.1833e-12	0.0040	0.0260	5.1559e-13
MSK(4,2)	0.0014	0.0071	1.0479e-10	0.0042	0.0300	8.4749e-11
MSK(4,3)	0.0009	0.0048	7.4298e-11	0.0039	0.0242	7.8693e-13

TABLE 5.1
Comparison to TPS-based methods on a 21×21 *uniform grids with different noise levels.*

effective and stable method between the two.

The clear advantages of MSK with L-curve are that it is already in a truly mesh-free setting for any dimension and it does not require any *a priori* knowledge about the noise level δ. Moreover, resultant linear systems of MSK in (2.5) are sparse. To make the comparison as fair as possible, we compare the accuracies of all methods on uniformly distributed grids among many given examples in their papers. Please be reminded that there are still some differences between the problem settings here and in [34].

Let $\Omega = [-2, 2]^2$. The noise levels are chosen to be the reported $\delta = 1.018\text{e-}3$ and $\delta = 1.020\text{e-}2$. The unknown function to be approximated is given by

$$f(x, y) = \sin(\pi\, x) \sin(\pi\, y) \exp(-x^2 - y^2), \quad (x, y) \in \mathbb{R}^2,$$

with $\|f\|_{L^2(\Omega)} \approx 0.387$ and $\|\nabla f\|_{L^2(\Omega)} \approx 4.235$. Since the number of evaluation points is not reported in [34], we use the same *root mean square* (RMS) errors on a 100×100 uniformly distributed grids $x'_i \in \Omega$ to measure accuracy for interpolation,

$$\varepsilon(s_\alpha) = \frac{1}{100} \Big(\sum_{i=1}^{100^2} (s_\alpha(x'_i) - f(x'_i))^2 \Big)^{1/2},$$

and for gradient approximation,

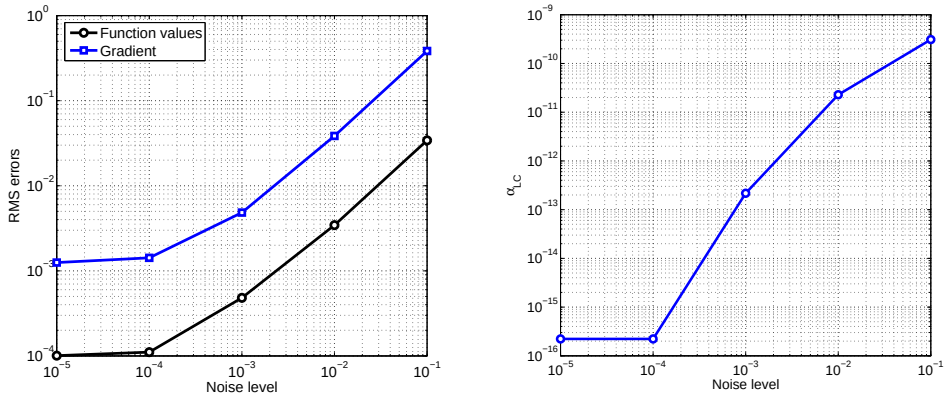
$$\varepsilon(\nabla s_\alpha) = \frac{1}{100} \Big(\sum_{i=1}^{100^2} \|\nabla s_\alpha(x'_i) - \nabla f(x'_i)\|_{\ell^2}^2 \Big)^{1/2}.$$

Table 5.1 shows the RMS errors for both tested noise levels on a 21×21 uniform grids. The differences in error should not be overinterpreted as they are influenced by the regularization parameter α_{LC} and the noise function η. It is more important to note that all choices of m and σ result in the same order of accuracy. Under this point density, MSK shows competitive results and seems to outperform TPS.

For 1609 unstructured data points, see Figure 5.2(a), with minimum separation distance $h_{X,\min} = 5.092\text{e-}2$ and fill distance $h_{X,\Omega} = 1.317\text{e-}1$. We apply MSK(3,2) to various noise levels. Results are listed in Table 5.2 and graphically demonstrated in Figure 5.1. All regularization parameters are chosen by the L-curve method except the first row of Table 5.2: $\alpha = 0$ indicates the result of the full interpolant without regularization. Our algorithm runs in the same way as if the data points were structured. The number of selected frame coefficients is listed under the column of $\mathrm{nz}(r_k^j)$ in the table.

δ	α_{LC}	$\mathrm{nz}(r_k^J)$	$\varepsilon(s_\alpha)$	$\varepsilon(\nabla s_\alpha)$
0	$\alpha = 0$	100921	*8.5518e-5*	1.5045e-3
0	2.2204e-16	6081	1.0032e-4	*1.2479e-3*
1e-5	2.2204e-16	6076	1.0066e-4	1.2511e-3
1e-4	2.2204e-16	6158	1.1065e-4	1.4226e-3
1e-3	2.1649e-13	1800	4.8194e-4	4.8393e-3
1e-2	2.2794e-11	1678	3.4443e-3	3.8510e-2
1e-1	3.0885e-10	1633	3.4145e-2	3.8377e-1

TABLE 5.2
MSK(3,2) RMS errors and α_{LC} on an 1609 unstructured data points with different noise levels.



(a) *RMS errors.* **(b) α_{LC}.**

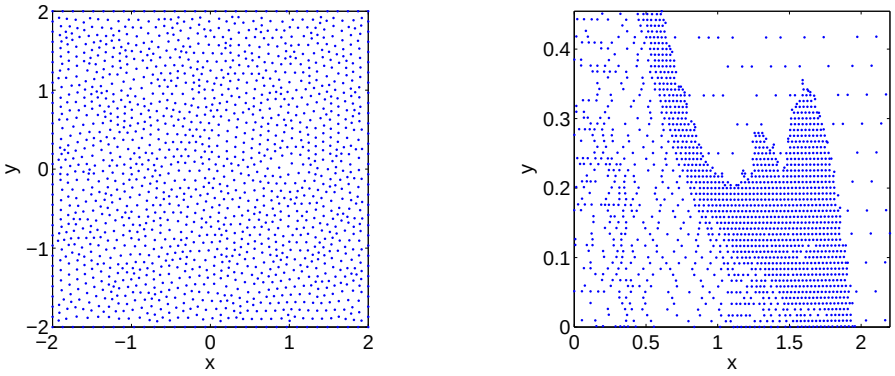
FIG. 5.1. *RMS and α_{LC} errors as functions of the noise level δ.*

Comparing the two noise-free results in Table 5.2, the interpolation error when $\alpha = 0$ is the smallest since the regularization error no longer exists. On the other hand, due to the presence of rounding errors, the regularized interpolant gives better approximation to the gradient than the unregularized full interpolant. In fact, this is true up to $\delta =$ 1e-4. When $\delta \geq$ 1e-3, we have $\alpha_{LC} > \epsilon_{mach}$ and our regularization technique is functioning in these examples, see Theorem 3.3. Overall, the error profile is extremely similar to the TPS-DP, see [34, Figure 5]. The monotonic trend shown in α_{LC} suggests that the proposed L-curve method is capable of balancing the increasing noise with an increasing regularization parameter.

Our MSK scheme performs equally well when the noise function η is smooth[1]. For completion, MSK(3,2) results in $\varepsilon(s_\alpha) = 0.0025$ and $\varepsilon(\nabla s_\alpha) = 0.0046$ on a 41×41 uniformly distributed grid. Whereas, TPS-DP results in $\varepsilon(s_\alpha) = 0.0035$ and $\varepsilon(\nabla s_\alpha) = 0.0159$.

5.2. Derivative of a Landscape Data. We demonstrate another example with a set of landscape data [11], see Figure 5.2(b). The data set, containing 1669 data points, is processed by MSK(3,2) and MSK(3,3) in order to estimate its derivatives. Unlike the previous example, data points are unevenly distributed and there is no exact solution for this example. Hence, the full interpolant $s_{\delta,X}$ will be used for

[1] $\eta(x, y) = 0.005 \sin\left(\frac{1}{2}\pi x\right) \sin\left(\frac{1}{2}\pi y\right)$, see [34, Table 1].



(a) *1609 unstructured points.* **(b) *Landscape data points.***

FIG. 5.2. *Data points distribution for examples in Section 5.1 and Section 5.2.*

comparison. We only demonstrate the x-derivatives; results for the y-derivatives are similar and are omitted here.

The full interpolant $s_{\delta,X}$ and its x-derivative are shown in Figure 5.3. As we see in Section 3.1, the presence of noise does not introduce instability to the interpolation problem. On the other hand, we observe serious oscillations in the derivatives of the full interpolant, see Figure 5.3(b).

The MSK(3,2) regularized interpolants with $\alpha_{LC} = 5.0626\text{e-}14$ (566 nonzero frame coefficients) are shown in Figure 5.4. The regularized interpolant in Figure 5.4 is very similar to Figure 5.3 but with less local structures. The derivative of the regularized interpolant in Figure 5.4(b) clearly reveal the local features of the landscape.

The MSK(m,σ) method assumes the unknown function f lies in $\mathcal{N}_\Phi$ and LC regularizes the interpolant accordingly. If σ is too large, the multiscale kernel Φ_σ is very smooth and the MSK scheme will over-regularize the interpolant. Fortunately, nothing will become unbounded. To see this, if we can write the unknown function $f \notin \mathcal{N}_\Phi$ as $f = f_1 + f_2$ where $f_1 \in W^{\sigma,2}$ and $f_2 \in L^2(\Omega) \cap C(\Omega)$ then our results in Section 3.1 apply consequently. As an example, Figure 5.5 shows the regularized interpolant of MSK(3,3). The regularization parameter is $\alpha_{LC} = 3.8654\text{e-}12$ resulting in 122 frame coefficients. The resulting regularized interpolant in Figure 5.5 is much smoother than that of MSK(3,2) in Figure 5.4. In fact, it seems too smooth for the landscape data.

For rough data from a function $f \notin \mathcal{N}_\Phi$, we shall treat α_{LC} as an upper estimated parameter. To capture more local features, we could use a regularization parameter $0 < \alpha < \alpha_{LC}$ and obtain results similar to the one from MSK(3,2). The resulting interpolant will contain more local features with any $0 < \alpha < \alpha_{LC}$, while the oscillation in its derivatives are still relatively well behaved. However, we have no robust routine for choosing an optimal regularization parameter in this case.

For unevenly distributed data points, the tolerance to roughness should be proportional to the local density of data points, e.g. a threshold of the form $t_\sigma(j,k)$. Regions with high data points density are expected to have more local features and higher roughness should therefore be allowed. This allows smooth kernels to capture more local features of the given data set in certain regions. An example of such a density measure is the number of data points in the support of each function $\varphi(2^j \cdot -k)$;

the information is already available after computing the frame coefficients. We leave this as an open question for future study.

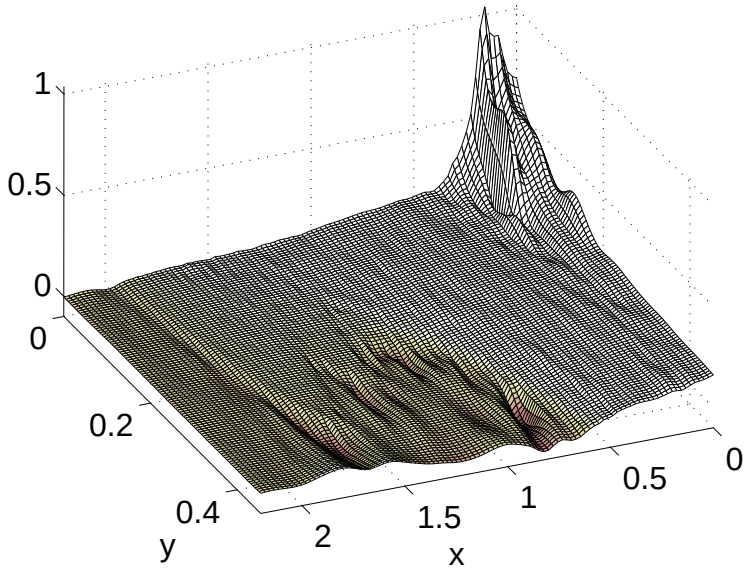
6. Conclusion. We solve a classical ill-posed numerical differentiation problem by a state-of-the-art matrix-free multiscale kernel based multivariate interpolation method. The theoretical stability for this ill-posed problem is investigated. The Tikhonov regularization and the L-curve method are employed to obtain a regularized interpolant. The advantages of the proposed method are (1) the ability to handle problems in higher dimensions, (2) the flexibility to handle real-life, noisy and multiple-valued data, and (3) the efficiency due to the resultant sparse matrix systems. Numerical examples are given for a bivariate test problem that shows results competitive with the thin plate spline based method and a landscape data set that shows the stability of our scheme even when the unknown function may not be smooth enough for our assumptions.

Acknowledgement. The authors would like to thank M. Yamamoto, R. Schaback, R. Opfer, M. R. Trummer, S. Ruuth and T. Takeuchi for their helpful comments. Moreover, we thank the reviewers for improving the academic quality and readability of this manuscript. This research was partially supported by a Postdoctoral Fellowship from the Japan Society for the Promotion of Science.

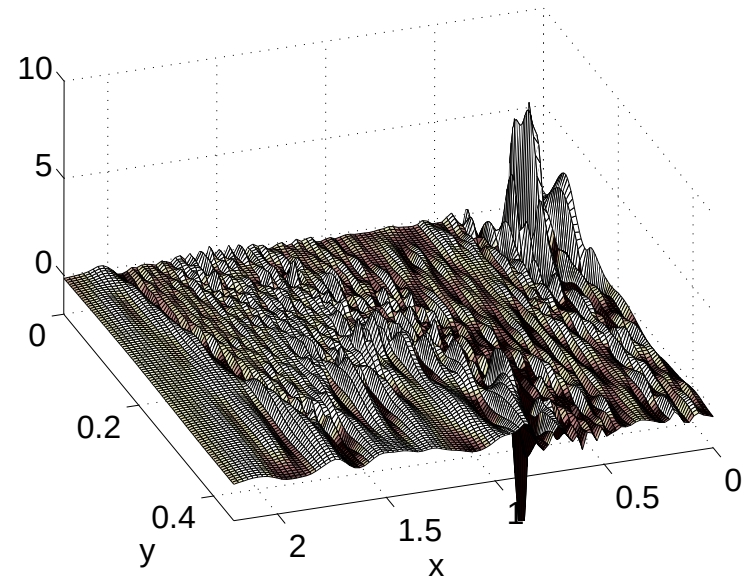
REFERENCES

[1] R. A. Adams, *Sobolev spaces*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975. Pure and Applied Mathematics, Vol. 65.
[2] R. S. Anderssen and M. Hegland, *For numerical differentiation, dimensionality can be a blessing!*, Math. Comp., 68 (1999), pp. 1121–1141.
[3] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68 (1950), pp. 337–404.
[4] R. Baltensperger and M. R. Trummer, *Spectral differencing with a twist*, SIAM J. Sci. Comput., 24 (2003), pp. 1465–1487 (electronic).
[5] M. Bozzini and M. Rossini, *Numerical differentiation of 2D functions from noisy data*, Comput. Math. Appl., 45 (2003), pp. 309–327.
[6] J. R. Cannon, Y. P. Lin, and S. Xu, *Numerical procedures for the determination of an unknown coefficient in semi-linear parabolic differential equations*, Inverse Problems, 10 (1994), pp. 227–243.
[7] J. Cheng, Y. C. Hon, and Y. B. Wang, *A numerical method for the discontinuous solutions of Abel integral equations*, in Inverse problems and spectral theory, vol. 348 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2004, pp. 233–243.
[8] J. Cullum, *Numerical differentiation and regularization*, SIAM J. Numer. Anal., 8 (1971), pp. 254–265.
[9] C. de Boor, *A practical guide to splines*, vol. 27 of Applied Mathematical Sciences, Springer-Verlag, New York, revised ed., 2001.
[10] S. R. Deans, *The Radon transform and some of its applications*, A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1983.
[11] R. Franke, mbay.mat. Available at `http://www.math.nps.navy.mil/~rfranke/`.
[12] R. Gorenflo and M. Yamamoto, *Operator-theoretic treatment of linear Abel integral equations of first kind*, Japan J. Indust. Appl. Math., 16 (1999), pp. 137–161.
[13] C. W. Groetsch and O. Scherzer, *Iterative stabilization and edge detection*, in Inverse problems, image analysis, and medical imaging (New Orleans, LA, 2001), vol. 313 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2002, pp. 129–141.
[14] M. Hanke and O. Scherzer, *Inverse problems light: numerical differentiation*, Amer. Math. Monthly, 108 (2001), pp. 512–521.
[15] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.
[16] P. C. Hansen and D. P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.

[17] M. Hegland and R. S. Anderssen, *Resolution enhancement of spectra using differentiation*, Inverse Problems, 21 (2005), pp. 915–934.
[18] B. Hofmann and M. Yamamoto, *Convergence rates for Tikhonov regularization based on range inclusions*, Inverse Problems, 21 (2005), pp. 805–820.
[19] C. Itiki and J. J. Neto, *Complete automation of the generalized inverse method for constrained mechanical systems of particles*, Appl. Math. Comput., 152 (2004), pp. 561–580.
[20] L. Ling, *Multivariate quasi-interpolation schemes for dimension-splitting multiquadric*, Appl. Math. Comput., 161 (2005), pp. 195–209.
[21] H. Meschkowski, *Hilbertsche Räume mit Kernfunktion*, Die Grundlehren der mathematischen Wissenschaften, Bd. 113, Springer-Verlag, Berlin, 1962.
[22] F. J. Narcowich, J. D. Ward, and H. Wendland, *Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting*, Math. Comp., 74 (2005), pp. 743–763 (electronic).
[23] R. Opfer, *Multiscale Kernels*, PhD thesis, Georg-August-Universität zu Göttingen, Göttingen, 2004.
[24] M. Piana, R. Barrett, J. C. Brown, and S. W. McIntosh, *A non-uniqueness problem in solar hard x-ray spectroscopy*, Inverse Problems, 15 (1999), pp. 1469–1486.
[25] D. A. Popov and D. V. Sushko, *Computation of singular convolutions*, in Applied problems of Radon transform, vol. 162 of Amer. Math. Soc. Transl. Ser. 2, Amer. Math. Soc., Providence, RI, 1994, pp. 43–127.
[26] A. G. Ramm and A. B. Smirnova, *On stable numerical differentiation*, Math. Comp., 70 (2001), pp. 1131–1153 (electronic).
[27] T. J. Rivlin, *Optimally stable Lagrangian numerical differentiation*, SIAM J. Numer. Anal., 12 (1975), pp. 712–725.
[28] R. Schaback, *Approximation by radial basis functions with finitely many centers*, Constr. Approx., 12 (1996), pp. 331–340.
[29] ———, *Native Hilbert spaces for radial basis functions I*, in New Developments in Approximation Theory, B. M.D., M. D. H., F. M., Müller, and M.W., eds., vol. 132 of International Series of Numerical Mathematics, Birkhäuser Verlag, 1999, pp. 255–282.
[30] L. Tang and J. D. Baeder, *Uniformly accurate finite difference schemes for p-refinement*, SIAM J. Sci. Comput., 20 (1999), pp. 1115–1131 (electronic).
[31] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*, V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.
[32] C. R. Vogel, *Non-convergence of the L-curve regularization parameter selection method*, Inverse Problems, 12 (1996), pp. 535–547.
[33] Y. B. Wang, X. Z. Jia, and J. Cheng, *A numerical differentiation method and its application to reconstruction of discontinuity*, Inverse Problems, 18 (2002), pp. 1461–1476.
[34] T. Wei, Y. C. Hon, and Y. B. Wang, *Reconstruction of numerical derivatives from scattered noisy data*, Inverse Problems, 21 (2005), pp. 657–672.
[35] H. Wendland, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics (No. 17), Cambridge University Press, Cambridge, 2005.

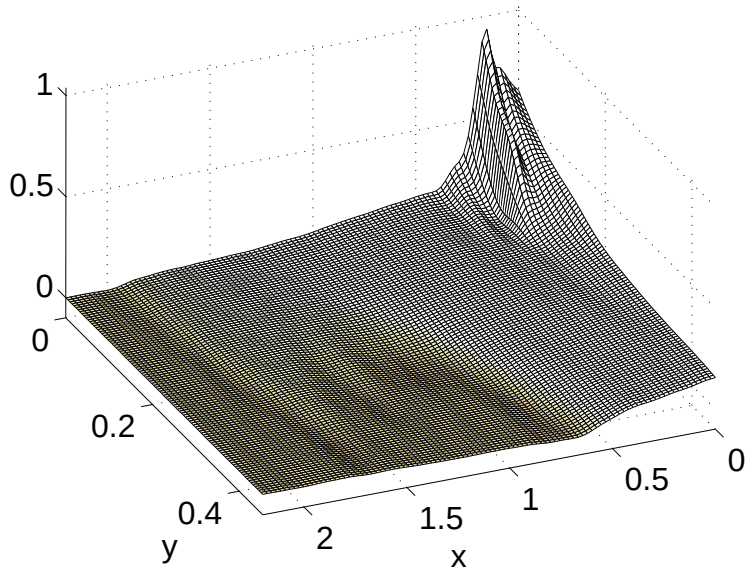


(a) *Full interpolant* $s_{\delta,X}$.

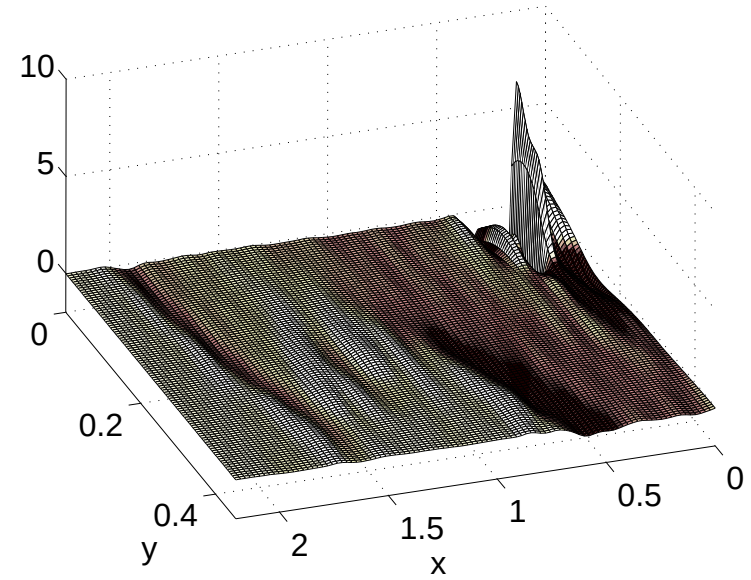


(b) x-derivatives.

FIG. 5.3. *Full interpolant for the landscape data and its x-derivatives.*

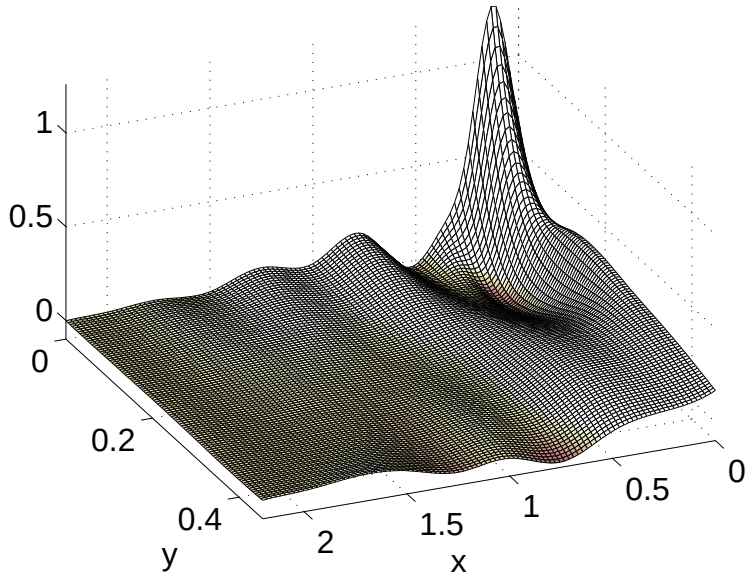


(a) *Regularized interpolant s_α with 566 frame coefficients.*

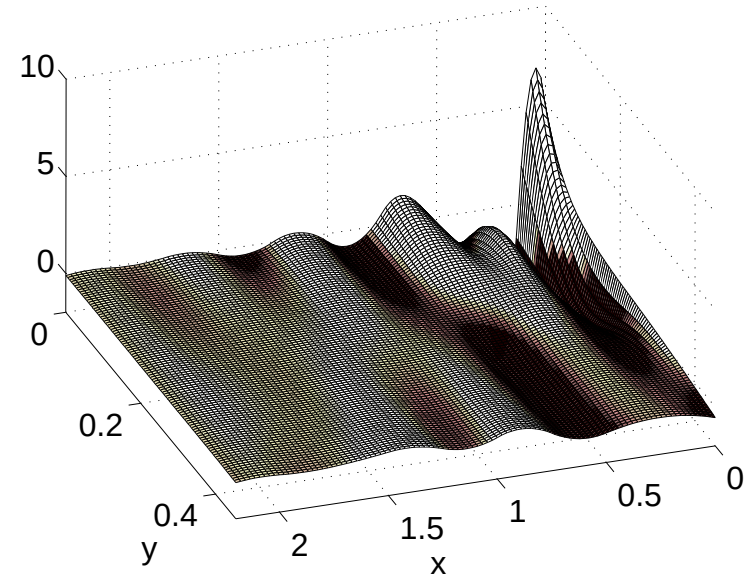


(b) *x-derivatives.*

FIG. 5.4. *MSK(3,2) regularized interpolant for the landscape data and its x-derivatives.*



(a) *Regularized interpolant s_α with 122 frame coefficients.*



(b) *x-derivatives.*

FIG. 5.5. *MSK(3,3) regularized interpolant for the landscape data and its x-derivatives.*