Multiscale Support Vector Regression Method On Spheres with Data Compression

Min Zhong
Department of Mathematics
Southeast University
Nanjing, 210096, People’s Republic of China

Q. T. Le Gia
School of Mathematics and Statistics
University of New South Wales
Sydney, NSW 2052, Australia

Wei Wang† (Corresponding Author)
College of Mathematics, Physics and Information Engineering
Jiaxing University
Jiaxing, 314001, People’s Republic of China

MIN.ZHONG@SEU.EDU.CN
QLEGIA@UNSW.EDU.AU
WEIWANG_MATH@126.COM

Abstract

In this manuscript, we investigate the multiscale support vector regression (SVR) method with data compression for approximation of functions on the unit sphere. The data are obtained at scattered sites on the sphere and may contain noise. The Vapnik ε-intensive loss function, which has been well-developed in learning theory, is introduced to obtain a local regularized approximation at each step, a data compression method is applied to discard small coefficients during the computation. We will discuss the convergence of the algorithm, and prove additional errors can be controlled so that the discarding strategy does not lead to significant errors. Numerical simulations support the theoretical predictions.

Keywords: multiscale, support vector regression, unit sphere

1. Introduction

Large data sets measured on the sphere arise in many applications such as computer graphics, geophysics, astrophysics and quantum chemistry, (see McEwen et al., 2011) and references therein. Spherical basis functions (SBFs) (see Narcowich et al., 2002; Le Gia et al., 2010) centered at scattered sites can be utilized as efficient tools for constructing the approximations when the data are scattered.

When SBFs with compact supports are used, a general observation shows that a small support radius leads to a well-conditioned sparse linear system with low computational costs but the approximation result is not satisfactory, while a large support radius yields good result but at the expense of an ill-conditioned collocation matrix (see Schaback, 1997). Therefore, a natural idea is to vary the support radius and consequently a multilevel or
multiscale approximation scheme is introduced (see Floater et al., 1996; Schaback, 1995, 1997; Wendland, 2005). The multiscale approximation scheme studied here follows an idea by Floater et al. (1996), but the original root goes back to Schaback (1995). The basic idea is that one begins with SBFs of large scale, and obtains a first approximation at widely separate points, then forms the residual and find a correction by approximating the residual with SBFs of a finer scale and more points. When the iteration terminates by a pre-specified number, we obtain an approximation which is a sum of SBFs with finer scales. Unlike the work of Floater et al. (1996); Narcowich et. al. (1999), there is no assumption that point sets at different levels are nested. In 2010, Wendland (2010) developed a multiscale algorithm on bounded domains based on least-square methods. For the first time, rigorous proofs demonstrated that the multiscale approach obtained an excellent trade-off between superior convergence rate and low computational costs. Later on, utilizing compact spherical SBFs, the multiscale method on the sphere was proposed and analyzed by Le Gia et al. (2010). Such idea then has been successfully extended by Le Gia et al. 2012b; Xu et al. 2015 where the regularity of object function on sphere was not known, and has been used in solving elliptic PDEs on a sphere (see Le Gia et al., 2012a), in solving second-order elliptic PDEs by Galerkin method on bounded domains (see Cherni et al., 2012; Chernih et al., 2013), with boundary zero boundary conditions (see Townsend et al., 2013). The multiscale approach in solving linear ill-posed problems was also investigated by Zhong et al. (2015, 2012).

Another problem in approximation on spheres is that the observations always contain noises, thus regularization methods should be taken into consideration. In Wendland (2010); Xu et al. (2015), the authors considered a noisy case, the regularized-least squares method and support vector regression (SVR) method were used respectively, but such idea has not been applied to the sphere. In this manuscript, we considered the SVR method applying to the approximation of noisy, scattered data on the unit sphere. The SVR is a variant of the support vector machine, which has been developed in learning theory by Boser et al. (1992) and has been further investigated by Cortes et al. (1995); Vapnik (2000). The SVR method usually is accomplished with a cut-off parameter $\varepsilon$ and the so-called Vapnik $\varepsilon$–intensive loss function defined by

$$|x|_\varepsilon = \begin{cases} 
0, & \text{for } |x| \leq \varepsilon; \\
|x| - \varepsilon, & \text{for } |x| > \varepsilon.
\end{cases}$$

By using the $\varepsilon$–intensive loss function in the data-fitting term we intend to overcome overfitting of noisy data which destroys the general characteristic of the target functions. An overview of SVR methods, the classification of the Vapnik $\varepsilon$–intensive loss function and parameter choice rule can be found in (e.g. Christianini et al., 2000; Hastie et al., 2009; B. Schölkopf et. al., 2002; Smola et al., 2004; Rieger et al., 2009) and references therein.

In practical situations, the data on the unit sphere are scattered, the structure and the information are impossible to be detected, thus an efficient and accurate compression of data on the sphere becomes increasingly necessary for the storage of the data. Data compression on sphere has been considered by McEwen et al. (2011); Schröder and Sweldens (1995), where a wavelet representation of function were utilized. In the work of Le Gia and Wendland (2014), two data compression techniques with thresholding strategies called ”discarding at the end” and ”discarding dynamically” were proposed respectively. They employed a mesh free approximation method, which is perfectly suitable to deal with the real scattered data.
on spheres. To the authors’ best knowledge, it was the first time such technique was used with data compression.

In this manuscript, we consider the multiscale SVR method with data compression, the amount of data we obtained at scattered sites on the sphere is large and the data may contain noise. In each step, we use the SVR method to obtain a local regularized approximation to the residual of previous level, the "discarding dynamically" was applied to discard small coefficients during the computation. We will discuss the convergence of the algorithm, and prove additional errors can be controlled so that discarding strategy does not lead to significant errors.

The manuscript is organized as follows. Section 2 contains an overview of the preliminary knowledge and presents the multiscale SVR approximation algorithm with data compression. Convergence analysis for the approximation algorithm is carried out in Sections 3 and 4 for smooth and rough functions respectively. Finally, some numerical simulations in Section 5 show the efficiency of our proposed multiscale algorithm.

2. Preliminaries

2.1 Polynomial spaces on the unit sphere

The unit sphere \( S^n \) is defined by \( S^n = \{ x \in \mathbb{R}^{n+1} : |x| = 1 \} \), where \(|x|\) denotes the Euclidean norm of \( x \). Spherical harmonics are the restrictions to \( S^n \) of the (real) homogeneous polynomials \( Y(x) \) in \( \mathbb{R}^{n+1} \) which satisfy \( \Delta Y(x) = 0 \), where \( \Delta \) is the Laplacian operator in \( \mathbb{R}^{n+1} \). For \( l \geq 0 \), let \( \mathbb{H}_l \) be the space of all spherical harmonics of degree \( l \) in \( S^n \), it has orthonormal basis

\[
\{ Y_{lk} : k = 1, 2, \ldots, N(n, l) \},
\]

where

\[
N(n, 0) = 1, \quad \text{and} \quad N(n, l) = \frac{(2l + n - 1)\Gamma(l + n - 1)}{\Gamma(l + 1)\Gamma(n)} \quad \text{for} \ l \geq 1.
\]

The space of spherical harmonics of degree not exceed \( L \) will be denoted by \( \mathbb{P}_L = \bigoplus_{l=0}^L \mathbb{H}_l \), it has dimension \( N(n+1, L) \). The set of spherical harmonics

\[
\{ Y_{lk} : k = 1, 2, \ldots, N(n, l), \ l = 1, 2, \ldots \}
\]

is a complete orthonormal basis of \( L_2(S^n) \). It follows that an arbitrary \( f \in L_2(S^n) \) can be represented in \( L_2 \) sense by its Fourier series (Gronwall, 1914) with respect to spherical harmonics

\[
f = \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} \hat{f}_{lk} Y_{lk},
\]

with its Fourier coefficients given by

\[
\hat{f}_{lk} = (f, Y_{lk})_{L_2(S^n)} = \int_{S^n} f(x) Y_{lk}(x) d\omega(x),
\]

where \( d\omega(x) \) is the measure of the \( S^n \). More detailed discussions related to spherical harmonics can be found in Müller (1966).
2.2 Positive definite bizonal kernels on $\mathbb{S}^n$ and RBF kernels

Bizonal kernels defined on $\mathbb{S}^n \times \mathbb{S}^n$ are functions that can be represented as $\phi(x \cdot y)$ for $x, y \in \mathbb{S}^n$, where $\phi(t)$ is continuous on $[-1,1]$. We shall be concerned exclusively with bizonal kernels of the type

$$\phi(x \cdot y) = \sum_{l=0}^{\infty} a_l P_l(n+1;x \cdot y), \quad a_l > 0, \quad \sum_{l=0}^{\infty} a_l < \infty,$$

where $\{P_l(n+1;t)\}_{l=0}^{\infty}$ is the sequence of $(n+1)$-dimensional Legendre polynomials normalized to $P_l(n+1;1) = 1$. For a positive integer $M$ and a sequence of distinct points $\{x_1, \ldots, x_M\}$ on $\mathbb{S}^n$, the functions $\Phi(\cdot, x_j) = \phi(\cdot \cdot x_j)$ for $1 \leq j \leq M$ are called spherical (radial) basis functions (SBFs). Thanks to the Schoenberg’s work Schoenberg (1942) and later work Xu et al. (1992), we know that such kernel is positive definite on $\mathbb{S}^n$, i.e., the matrix $\{\Phi(x_i, x_j)\}_{i,j=1}^{M}$ is positive definite for every $M$ and distinct points set.

Using the additional theorem Mller (1966), we can write

$$\Phi(x, y) = \phi(x \cdot y) = \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} \hat{\phi}(l) Y_{lk}(x) \bar{Y}_{lk}(y), \quad \text{with} \quad \hat{\phi}(l) = \omega_n \frac{\omega_n}{N(n,l)} a_l,$$

where $\omega_n$ is the surface area of $\mathbb{S}^n$, and $\hat{\phi}(l)$ is the Fourier coefficients of $\Phi$. We will assume $\hat{\phi}(l) \sim (1 + l)^{-2\sigma}$, where $\sigma$ is the smoothness index of the target function we aim to reconstruct. More precisely, we assume there exists two constants $c_1, c_2 > 0$ such that

$$c_1 (1 + l)^{-2\sigma} \leq \hat{\phi}(l) \leq c_2 (1 + l)^{-2\sigma}, \quad l \geq 0. \quad (1)$$

The reproducing kernel Hilbert space (RKHS) induced by $\Phi$ is defined to be

$$\mathcal{N}_\Phi = \left\{ f \in \mathcal{D}'(\mathbb{S}^n) : \|f\|_\Phi^2 = \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} |\hat{f}_{lk}|^2 \hat{\phi}(l) < \infty \right\},$$

where $\mathcal{D}'(\mathbb{S}^n)$ is the set of all tempered distributions defined on $\mathbb{S}^n$. We emphasize that $\mathcal{N}_\Phi$ is the completion of span $\{\Phi(\cdot, x), x \in \mathbb{S}^n\}$ with respect to the norm $\| \cdot \|_\Phi$.

The Sobolev space on the unit sphere $H^\sigma(\mathbb{S}^n)$ is defined by

$$H^\sigma(\mathbb{S}^n) = \left\{ f \in \mathcal{D}'(\mathbb{S}^n) : \|f\|_{H^\sigma(\mathbb{S}^n)}^2 = \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} (1 + l)^{2\sigma} |\hat{f}_{lk}|^2 < \infty \right\}.$$ 

By assumption (1), the norms in the space $\mathcal{N}_\Phi$ and the Sobolev space $H^\sigma(\mathbb{S}^n)$ are equivalent, i.e., for $f \in H^\sigma(\mathbb{S}^n)$, we have

$$c_1^{1/2} \|f\|_\Phi \leq \|f\|_{H^\sigma(\mathbb{S}^n)} \leq c_2^{1/2} \|f\|_\Phi.$$

In addition, by Sobolev embedding theorem, we know functions in $H^\sigma(\mathbb{S}^n)$ are continuous if $\sigma > n/2$. 

4
We illustrate that the assumption (1) is reasonable. Let $\Phi$ be defined from a compactly supported Wendland’s radial basis functions $W : \mathbb{R}^{n+1} \to \mathbb{R}$, (see Wendland, 1995; Wu, 1995), with the associated RKHS $H^{\sigma + \frac{1}{2}}(\mathbb{R}^{n+1})$. By restricting $W$ to $\mathbb{S}^n \subset \mathbb{R}^{n+1}$, we obtain a positive, bizonal kernel on unit sphere

$$
\Phi(x, y) = W(|x - y|) = \phi(x \cdot y), \quad x, y \in \mathbb{S}^n.
$$

It is known from Narcowich et al. (2002) that the Fourier symbols of $\Phi$ satisfies (1), i.e., $\Phi$ is the reproducing kernel of RKHS $H^\sigma(\mathbb{S}^n)$.

For a given $\eta > 0$, we define the scaled version of kernel $\Phi$ on $\mathbb{S}^n$ by

$$
\Phi_\eta(x, y) = \eta^{-n}W\left(\frac{|x - y|}{\eta}\right), \quad x, y \in \mathbb{S}^n.
$$

(2)

We can expand the scaled kernel $\Phi_\eta$ into a series of spherical harmonics

$$
\Phi_\eta(x, y) = \sum_{l=0}^{\infty} \sum_{k=1}^{N(n, l)} \phi_\eta(l)Y_{lk}(x)\overline{Y_{lk}(y)},
$$

in which the Fourier symbols $\phi_\eta(l)$ satisfy

$$
c_1(1 + \eta l)^{-2\sigma} \leq \phi_\eta(l) \leq c_2(1 + \eta l)^{-2\sigma}, \quad l \geq 0,
$$

(3)

with constants $c_1$ and $c_2$ from (1) possibly relaxed (see Le Gia et al., 2010). We define the RKHS $\mathcal{N}_{\Phi_\eta}$ induced by scaled kernel $\Phi_\eta$, with the norm

$$
\|f\|_{\Phi_\eta}^2 = \sum_{l=0}^{\infty} \sum_{k=1}^{N(n, l)} \frac{|f_{lk}|^2}{\phi_\eta(l)}.
$$

Then, for all $f \in H^\sigma(\mathbb{S}^n)$, the application of (3) yields,

$$
c_1^{1/2}\|f\|_{\Phi_\eta} \leq \|f\|_{H^\sigma(\mathbb{S}^n)} \leq c_2^{1/2}\eta^{-\sigma}\|f\|_{\Phi_\eta}.
$$

(4)

The following lemma is of a similar fashion.

**Lemma 1** (see Le Gia and Wendland, 2014) Assume that $\Phi(x, x) = W(0) = 1$ is a reproducing kernel of $H^\sigma(\mathbb{S}^n)$ with $\sigma > n/2$ and let $\Phi_\eta$ be the scaled kernel (2) with $\eta > 0$. Then the following estimates hold:

(a) For every $x \in \mathbb{S}^n$, we have $\|\Phi_\eta(\cdot, x)\|_{L^2(\mathbb{S}^n)} \leq c_2^{1/2}\eta^{-n/2}$.

(b) For every $x \in \mathbb{S}^n$ and every $\eta' \in (0, \eta)$ we have

$$
\|\Phi_\eta(\cdot, x)\|_{\Phi_{\eta'}} \leq (c_2/c_1)^{1/2}\eta^{-n/2}.
$$

(c) For every $x \in \mathbb{S}^n$, we have $\|\Phi_\eta(\cdot, x)\|_{\Phi_\eta} \leq \eta^{-n/2}$. 

5
2.3 Function approximation by the SVR method

The distance between any two points \(x, y\) on \(S^n\) is measured by the geodesic distance \(\theta(x, y) \in [0, \pi]\):

\[
\theta(x, y) = \arccos(x \cdot y), \quad \text{or} \quad \cos \theta = x \cdot y.
\]

The geodesic distance \(\theta \in [0, \pi]\) and Euclidean distance \(||x - y|| \in [0, 2]\) are related by

\[
\frac{2\theta}{\pi} \leq ||x - y|| = \sqrt{2 - 2x \cdot y} = 2 \sin \frac{\theta}{2} \leq \theta, \quad x, y \in S^n.
\]

Let \(X = \{x_1, \ldots, x_M\}\) be a set of \(M\) distinct scattered points on \(S^n \subset \mathbb{R}^{n+1}\). The quality of the geometric distribution of \(X\) is often characterized by the following two quantities and their ratio. The mesh norm

\[
h_X := \sup_{y \in S^n} \min_{x_j \in X} \theta(y, x_j),
\]

and the separation distance

\[
q_X = \min_{x_j, x_k \in X, j \neq k} \theta(x_j, x_k).
\]

The mesh norm \(h_X\) is the covering radius for covering the sphere with spherical caps of the smallest possible equal radius centered at points in \(X\), and \(q_X\) is the packing radius thus \(h_X \geq q_X\). If we pack \(M\) spherical caps of equal radius \(q_X\) on the sphere then the spherical caps are disjoint and hence the total volume of \(M\) caps is less than volume of \(S^n\), that is \(Mq^n_X \leq C\), or equivalently \(q_X \leq CM^{-1/n}\). The mesh ratio \(\rho_X = h_X/q_X \geq 1\) provides a measure of how uniformly points in \(X\) are distributed on \(S^n\). In the case \(n = 1\), \(\rho_X = 1\) means the observation points are equally spaced, in other cases, we have \(\rho_X > 1\). If \(\rho_X\) is bounded above by a constant \(\rho\), we say \(X\) is quasi-uniform, then

\[
h_X \leq \rho q_X \leq cM^{-\frac{1}{n}}, \quad \text{or equivalently} \quad M \leq c h_X^{-n}.
\]

The goal to learn a function is to approximate its value from some known discrete training observations data (which may be noisy) on one hand, and predict it on the unknown data-set on the other hand (see Poggio et al., 2003; B. Schölkopf et. al., 2002; De Vito et al., 2005). Concerning the regression problem on \(S^n\) with the Vapnik \(\varepsilon\)-intensive loss function, we concentrate on the following cost functional

\[
\min_{p \in H^\varepsilon(S^n)} \left\{ \frac{1}{M} \sum_{j=1}^{M} |p(x_j) - f^\delta(x_j)|_\varepsilon + \gamma \|p\|_\Phi^2 \right\}.
\]  

For a fixed regularization parameter \(\gamma > 0\) and a cut-off parameter \(\varepsilon > 0\), the minimizer of the above cost functional (5) exists, which can be represented as the linear combination of SBFs \(\Phi(\cdot, x_j)\) (see Rieger et al., 2009). To numerically calculate the approximate solution,
we introduce slack variables $a_j, b_j \geq 0$ for $j = 1, 2, \ldots, M$. The minimization problem can be parameterized and yield a quadratic optimal problem (see Schölkopf et al., 2000),

$$\min_{a,b,c \in \mathbb{R}^M, a,b \geq 0} \left( \frac{1}{M} \sum_{j=1}^{M} (a_j + b_j) + \gamma c^T \Phi_X c \right)$$

s.t. $(\Phi_X)_{j} c - f_j^\delta \leq \varepsilon + a_j, j = 1, \ldots, M,$

$$-[(\Phi_X)_{j} c - f_j^\delta] \leq \varepsilon + b_j, j = 1, \ldots, M,$$

where $\Phi_X \in \mathbb{R}^{M \times M}$ is the symmetric collocation matrix with entries

$$[\Phi_X]_{ij} = [\Phi(x_i, x_j)]_{1 \leq i,j \leq M}.$$  

As in Xu et al. (2015), when we denote $1_M = (1, \ldots, 1)^T \in \mathbb{R}^M$ and $z = (c, a, b)^T \in \mathbb{R}^{3M}$, the quadratic problem (6) can be written in the matrix form

$$\min_{z \in \mathbb{R}^{3M}} \left\{ \frac{1}{2} z^T H z + d^T z \right\}$$

s.t. $Cz \leq g$, $z_{M+1}, \ldots, z_{3M} \geq 0$

with

$$H = 2 \left( \begin{array}{cc} \gamma \Phi_X & O_{M,2M} \\ O_{2M,M} & \gamma O_{2M,2M} \end{array} \right) \in \mathbb{R}^{3M \times 3M}, \quad d = \frac{1}{M} \left( \begin{array}{c} O_M \\ 1_{2M} \end{array} \right) \in \mathbb{R}^{3M},$$

$$C = \left( \begin{array}{cc} \Phi_X & -I_{M,M} \\ -\Phi_X & O_{M,M} \end{array} \right) \in \mathbb{R}^{2M \times 3M}, \quad g = \left( \begin{array}{c} f_\delta^\delta |X_1 + \varepsilon 1_M \\ -f_\delta^\delta |X_k + \varepsilon 1_M \end{array} \right) \in \mathbb{R}^{2M},$$

where $O_{M,M}$ and $I_{M,M}$ denote the zero and the identity matrices, respectively.

3. Multiscale SVR approximation method in $\mathbb{S}^n$

3.1 Fundamental multiscale SVR method

The main idea in the multiscale method is that multiple support radii of the kernel function are used by Wendland (2010). More precisely, instead of a single sampling data set, we are dealing with a sequence of point sets $X_1, X_2, \ldots, X_k \subset \mathbb{S}^n$, where $X_j = \{x_{j1}, \ldots, x_{jM_j}\}$, with mesh norms $h_1, h_2, \ldots, h_k$ respectively. We will assume $h_{j+1} \approx \mu h_j$ for $1 \leq j \leq k$. Noisy observations are realized on the sampling data sets and denoted by $f_\delta^\delta |X_1, \ldots, f_\delta^\delta |X_k$.

We choose a positive definite bizonal SBF $\Phi(x, y)$ on $\mathbb{S}^n$, which is the restriction of a compactly supported RBF with associated RKHS $H^{n+1/2}(\mathbb{R}^{n+1})$ ($\sigma > n/2$) on the unit sphere. Let $\eta_1, \eta_2, \ldots$ be a decreasing sequence of positive real numbers, each $\eta_j$ will be called a scaling parameter defined by $\eta_j = \nu h_j$ for some $\nu > 0$. For every $j = 1, 2, \ldots$, we define the scaled SBFs as in (2). Taking the scaling parameter proportional to the mesh norm is desirable for both accuracy and efficiency, although the size of the kernel matrix $\Phi_{X_j}$ is increase with respect to $j$, the decrease of the support radius maintains its sparsity.
The multiscale algorithm starts with a widely spread set of sampling points $X_1$ and a basis kernel with scale $\eta_1$ generating a global reconstruction $f_1^\delta = s_1^\delta$ by the SVR method. The residual $e_1^\delta = f_1^\delta - s_1^\delta$ exists in the next level. A finer data-set $X_2$ and a smaller support radius is chosen, with the application of the SVR method we obtain a correction $s_2^\delta$ with even finer details, and a new approximation $f_2^\delta = f_1^\delta + s_2^\delta$, so the new residual is $e_2^\delta = f_2^\delta - f_1^\delta$ and so on. The whole process is iterated until a pre-specified level number is reached.

When we set the starting value $f_0^\delta = 0$ and $e_0^\delta = f_0^\delta$, we know that the algorithm makes $f_j^\delta + e_j^\delta$ independent of $j$, i.e.,

$$f_0^\delta + e_0^\delta = f_1^\delta + e_1^\delta = \ldots = f_j^\delta + e_j^\delta,$$

and $f_j^\delta$ as the approximation to $f$ at $j$-th level, which is the linear combination of SBFs at all scales up to $j$. The local correction $s_j^\delta$ can be regarded as an additional detail to $f_{j-1}^\delta$ to produce $f_j^\delta$. We can also put the algorithm into a multiresolution framework. For $j \geq 1$, let

$$W_j = \text{span} \{ \Phi_{\eta_j}(\cdot, x), \ x \in X_j \}, \quad \text{and} \quad V_j = \text{span} \{ \Phi_{\eta_i}(\cdot, x), \ x \in X_i, \ i \leq j \}.$$

In the language of wavelets, $W_j$ is wavelet space and $V_j$ is scale space. If $V_0 = \{0\}$, we have $V_j = V_{j-1} \oplus W_j$, $j \geq 1$, where the sum is direct. In addition, $\bigcup_{j=1}^{\infty} V_j$ with respect to the $\| \cdot \|_{L_2}$ norm is $L_2(\mathbb{S}^n)$ (see Le Gia et al., 2010).

### 3.2 Multiscale algorithm with data compression

In our algorithm, after $j$ steps, we finish with

$$f_j^\delta = \sum_{i=1}^{j} s_i^\delta = \sum_{i=1}^{j} \sum_{m=1}^{M_i} a_{im} \Phi_i(\cdot, x_{im}),$$

where $\Phi_i$ is actually $\Phi_{\eta_i}$ for simplicity. To reduce the storage, we can discard insignificant coefficients without introducing large errors. In Le Gia and Wendland (2014) the authors considered the data compression on the unit sphere, two strategies considering how to discard small coefficients were introduced: discarding at the end and discarding dynamically. In this manuscript, we applied the second type.

Choosing the thresholds $\tau_1, \ldots, \tau_k$ beforehand, depending only on levels, we proceed as follows. Let $\tilde{e}_0^\delta = f_0^\delta$ and $\tilde{f}_0^\delta = 0$ be the starting values, we first construct the local approximation $s_1^\delta = f_1^\delta$ by the SVR method, and obtain

$$s_1^\delta = \sum_{m=1}^{M_1} a_{1m} \Phi_1(\cdot, x_{1m}).$$

After discarding $\{a_{1m}\}$ with $|a_{1m}| \leq \tau_1$, $m = 1, \ldots, M_1$, we obtain

$$\tilde{e}_2^\delta = \sum_{|a_{1m}| > \tau_1} a_{1m} \Phi_1(\cdot, x_{1m}),$$

8
and define the residual $\tilde{e}_2^\delta = \tilde{e}_1^\delta - s_1^\delta$. In the next step, we compute $s_2^\delta$, the local correction, which again is in the form

$$s_2^\delta = \sum_{m=1}^{M_2} a_{2m} \Phi_1(\cdot, x_{2m}).$$

After discarding $|a_{2m}| \leq \tau_2$ for $m = 1, \ldots, M_2$, we obtain $\tilde{s}_2^\delta$ and then the residual $\tilde{e}_2^\delta = \tilde{e}_1^\delta - \tilde{s}_2^\delta$. Again the invariance of $\tilde{f}_j^\delta + \tilde{e}_j^\delta$ with respect to $j$ can be obtained, this allows us to identify $\tilde{e}_j^\delta = f^\delta - \tilde{f}_j^\delta$ as the error at the $j$–th level. The algorithm can be proposed as follows.

**Algorithm 1** Multiscale SVR method on $\mathbb{S}^n$ with data compression

- **Input:** Number of levels $k$, right-hand side $f^\delta|_{X_1}, \ldots, f^\delta|_{X_k}$
- **Output:** Approximate solution $\tilde{f}_k = \tilde{s}_1^\delta + \ldots + \tilde{s}_k^\delta$

1. Set $f_0^\delta = 0, \tilde{f}_0^\delta = f^\delta$
2. for $j = 1, 2, \ldots, k$ do
3. Determine the local correction $s_j^\delta \in W_j$ to $\tilde{e}_{j-1}^\delta$ by
   $$s_j^\delta = \arg \min_{s \in H^\delta(\mathbb{S}^n)} \left( \frac{1}{M_j} \sum_{m=1}^{M_j} |e_{j-1}^\delta(x_{jm}) - s(x_{jm})|_{\epsilon_j} + \gamma_j \| s \|^2_{\Phi_j} \right),$$
4. Drop all coefficients $|a_{jm}| \leq \tau_j$ in $s_j^\delta$ and define $\tilde{s}_j^\delta = \sum |a_{jm}| \geq \tau_j a_{jm} \Phi_j(\cdot, x_{jm})$.
5. Set $\tilde{f}_j^\delta = \tilde{f}_{j-1}^\delta + \tilde{s}_j^\delta$.
6. Set $\tilde{e}_j^\delta = \tilde{e}_{j-1}^\delta - \tilde{s}_j^\delta$.
7. end for

4. Multiscale analysis for smooth functions on $\mathbb{S}^n$

In this section, we will investigate the performance of the multiscale SVR approximation algorithm on $\mathbb{S}^n$ with data compression towards noisy observations of a smooth target function $f^* \in H^\delta(\mathbb{S}^n)$. Before the formal theoretical analysis, we introduce the following sampling inequality in unit sphere.

**Lemma 2** (see Le Gia et al., 2006) Suppose $X \subset \mathbb{S}^n$ is finite and has a sufficiently small mesh norm $h := h_X$, then for any $u \in H^\delta(\mathbb{S}^n)$ with $\sigma > n/2$, we have

$$\| u \|_{L_2(\mathbb{S}^n)} \leq C \left( h^\delta \| u \|_{H^\delta(\mathbb{S}^n)} + \| u \|_{L_\infty(X)} \right).$$

For readers’ convenience, we also collect the main assumptions we will use subsequently.

**Assumption 3**

- **A1** $\{X_1, X_2, \ldots, X_k\}$ forms a sequence of observation data sets, each $X_j$ has $M_j$ points, the mesh norms satisfy $c_\mu h_j \leq h_{j+1} \leq \mu h_j$ for $j = 1, 2, \ldots, k$ with fixed $c_\mu, \mu \in (0, 1)$ and $h_1$ sufficiently small. Assume all the data sets are quasi-uniform, i.e., for each $j$, $h_j/\rho_j \leq \rho$, or equivalently, $M_j \leq c\rho_j^{-n}$.

- **A2** At each step $j$, the observations $f^\delta|_{X_j}$ contain noise, satisfying $\| f^\delta - f^*\|_{L_\infty(X_j)} \leq \delta_j$. 


At each step $j$, the scaled SBFs $\Phi_j$ defined by (2), their Fourier symbols satisfy (3).

The scaling parameter is $\eta_j = \nu h_j$ with a constant $\nu > 0$. Moreover, assume $1 \geq T/\mu \leq \nu \leq 1/h_1$ with a constant $T > 0$.

At each step $j$, the algorithm discards insignificant coefficients $|a_{jm}| \leq \tau_j$ for $m = 1, 2, \ldots, M_j$ dynamically.

In the following theoretical analysis, we will use $c_1, c_2, \ldots$ for special constants, but $c$ and $C$ as general constants possibly relaxed.

Recalling the Algorithm, at the $j$-th level, the local correction $s_j^\delta \in H^\sigma(\mathbb{S}^n)$ satisfies

$$s_j^\delta = \arg \min_{s \in H^\sigma(\mathbb{S}^n)} \left( \frac{1}{M_j} \sum_{m=1}^{M_j} |\tilde{e}_{j-1}^\delta(x_{jm}) - s(x_{jm})|_{\varepsilon_j} + \gamma_j \|s\|^2_{\Phi_j} \right), \tag{7}$$

For each fixed $j$, defining a new function $e_j := f^* - \tilde{s}_1^\delta - \cdots - \tilde{s}_j^\delta$ belongs to $H^\sigma(\mathbb{S}^n)$. Putting $e_{j-1}$ into (7), since $s_j^\delta$ is the minimizer of (7), we have

$$\frac{1}{M_j} \sum_{m=1}^{M_j} |\tilde{e}_{j-1}^\delta(x_{jm}) - s_j^\delta(x_{jm})|_{\varepsilon_j} + \gamma_j \|s_j^\delta\|^2_{\Phi_j} \leq \frac{1}{M_j} \sum_{m=1}^{M_j} |\tilde{e}_{j-1}^\delta(x_{jm}) - e_{j-1}(x_{jm})|_{\varepsilon_j} + \gamma_j \|e_{j-1}\|^2_{\Phi_j}
= \frac{1}{M_j} \sum_{m=1}^{M_j} |f^\delta(x_{jm}) - f^*(x_{jm})|_{\varepsilon_j} + \gamma_j \|e_{j-1}\|^2_{\Phi_j}
\leq \frac{1}{M_j} \cdot M_j \cdot (\delta_j - \varepsilon_j)^+ + \gamma_j \|e_{j-1}\|^2_{\Phi_j},$$

where $(\cdot)^+$ denotes the positive part. This provides us a motivation to choose the cut-off parameter $\varepsilon_j$, which is stated as the following lemma:

**Lemma 4** Let the Assumption 3 hold. With the choice $\varepsilon_j = \delta_j$, we have the following estimate

$$\frac{1}{M_j} \sum_{m=1}^{M_j} |\tilde{e}_{j-1}^\delta(x_{jm}) - s_j^\delta(x_{jm})|_{\varepsilon_j} + \gamma_j \|s_j^\delta\|^2_{\Phi_j} \leq \gamma_j \|e_{j-1}\|^2_{\Phi_j}. \tag{8}$$

The following lemma follows directly from the basic estimate (8).

**Lemma 5** Let the Assumption 3 hold. At each level $j$, choosing the cut-off parameter $\varepsilon_j = \delta_j$, then the following estimates hold

$$\|s_j^\delta\|_{\Phi_j} \leq \|e_{j-1}\|_{\Phi_j},$$
$$\|e_{j-1} - s_j^\delta\|_{H^\sigma(\mathbb{S}^n)} \leq 2c_1^{1/2} \eta_j^{-\sigma} \|e_{j-1}\|_{\Phi_j},$$
$$\|e_{j-1} - s_j^\delta\|_{L^\infty(X_j)} \leq 2\delta_j + M_j \gamma_j \|e_{j-1}\|^2_{\Phi_j}.$$
Proof Due to the positivity of the two terms in the left hand side of the basic estimate (8), the first estimate is derived consequently. The second estimate follows from (4) and the first estimate,

$$
\|e_{j-1} - s_j^\delta\|_{H^\sigma(S^n)} \leq c_2^{1/2} \eta_j^{-\sigma} \|e_{j-1} - s_j^\delta\|_{\Phi_j}
\leq c_2^{1/2} \eta_j^{-\sigma} \left( \|e_{j-1}\|_{\Phi_j} + \|s_j^\delta\|_{\Phi_j} \right)
\leq 2c_2^{1/2} \eta_j^{-\sigma} \|e_{j-1}\|_{\Phi_j}.
$$

For the last estimate, referring to (8) and the choice for \(\varepsilon_j\), it follows that,

$$
\|e_{j-1} - s_j^\delta\|_{l^\infty(X_j)} \leq \|e_{j-1} - \bar{e}_{j-1}\|_{l^\infty(X_j)} + \|\bar{e}_{j-1} - s_j^\delta\|_{l^\infty(X_j)}
\leq \delta_j + \max_{1 \leq m \leq M_j} \left| \bar{e}_{j-1}(x_{jm}) - s_j^\delta(x_{jm}) \right|
\leq \delta_j + \max_{1 \leq m \leq M_j} \left| \bar{e}_{j-1}(x_{jm}) - s_j^\delta(x_{jm}) \right| + \varepsilon_j
\leq 2\delta_j + \sum_{m=1}^{M_j} \left| \bar{e}_{j-1}(x_{jm}) - s_j^\delta(x_{jm}) \right| \varepsilon_j
\leq 2\delta_j + M_j \gamma_j \|e_{j-1}\|_{\Phi_j}^2.
$$

\[\boxplus\]

**Theorem 6** Let the Assumption 3 hold. In addition, assume that the target function \(f^* \in H^\sigma(S^n)\) with \(\|f^*\|_{H^\sigma(S^n)} < c_1^{1/2}\). At each level \(j\), choose the cut-off parameter \(\varepsilon_j\), the regularization parameter \(\gamma_j\) satisfy

$$
\varepsilon_j = \delta_j, \quad \gamma_j \leq \left( \frac{1}{n} \right) \sigma_h^n\delta_j^n,
$$

Then, there exist constants \(C\) and \(\bar{M}\) independent of \(j\) and \(f^*\), such that

$$
\|e_j\|_{\Phi_{j+1}} \leq C\mu^\sigma \|e_{j-1}\|_{\Phi_j} + C\delta_j + \bar{M} \tau_j (c_2/c_1)^{1/2} \eta_j^{-n/2}, \text{ for } j = 1, 2, \ldots.
$$

(10)

Moreover, for a given constant \(\tau\), assume \(\alpha + C\delta + \tau < 1\) with \(\alpha := C\mu^\sigma\), then, if compression parameter \(\tau_j\) satisfies

$$
\tau_j \leq \frac{\tau}{\bar{M}} \left\{ \left( \frac{c_1}{c_2} \right)^{1/2} + \left( \frac{1}{c_2} \right)^{1/2} \right\} \eta_j^{n/2}
$$

we have

$$
\|e_j\|_{\Phi_{j+1}} \leq \alpha \|e_{j-1}\|_{\Phi_j} + C\delta_j + \tau < 1, \text{ for } j = 1, 2, \ldots.
$$

(12)

Proof Firstly, we have \(\|e_0\|_{\Phi_1} = \|f^*\|_{\Phi_1} \leq c_1^{-1/2} \|f^*\|_{H^\sigma(S^n)} < 1\), assume the induction hypothesis \(\|e_{j-1}\|_{\Phi_j} < 1\), we are aiming to prove the recursive formula (12).
Since
\[ \|e_j\|_{\Phi_{j+1}} = \|e_{j-1} - \tilde{s}_j\|_{\Phi_{j+1}} \leq \|e_{j-1} - s_j\|_{\Phi_{j+1}} + \|s_j - \tilde{s}_j\|_{\Phi_{j+1}}, \]  
we will try to estimate \(\|e_{j-1} - s_j\|_{\Phi_{j+1}}\) first, which can be split into
\[
\|e_{j-1} - s_j\|_{\Phi_{j+1}} = \left( \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} \frac{|(e_{j-1})_{lk} - (s_j)_{lk}|^2}{\phi_{j+1}(l)} \right)^{1/2}
\leq \left( \frac{1}{c_1} \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} |(e_{j-1})_{lk} - (s_j)_{lk}|^2 (1 + \eta_{j+1}l)^{2\sigma} \right)^{1/2}
\leq \sqrt{S_1} + \sqrt{S_2},
\]
where
\[
S_1 := \frac{1}{c_1} \sum_{l \leq 1/\eta_{j+1}}^{N(n,l)} \sum_{k=1}^{N(n,l)} |(e_{j-1})_{lk} - (s_j)_{lk}|^2 (1 + \eta_{j+1}l)^{2\sigma},
\]
\[
S_2 := \frac{1}{c_1} \sum_{l > 1/\eta_{j+1}}^{N(n,l)} \sum_{k=1}^{N(n,l)} |(e_{j-1})_{lk} - (s_j)_{lk}|^2 (1 + \eta_{j+1}l)^{2\sigma}.
\]
Since \(\eta_{j+1}l \leq 1\), concerning the choice for parameters \(\varepsilon_k, \gamma_k\), Assumption 3, sampling inequality in Lemma 2 and Lemma 5, then an estimate for \(\sqrt{S_1}\) can be derived as
\[
\sqrt{S_1} \leq c_1^{-1/2} 2^\sigma \left( \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} |(e_{j-1})_{lk} - (s_j)_{lk}|^2 \right)^{1/2}
\leq C \|e_{j-1} - s_j\|_{L_2(\mathbb{G})}^\sigma
\leq C \left( h_j^\sigma \|e_{j-1} - s_j\|_{H^\sigma(\mathbb{G})} + \|e_{j-1} - s_j\|_{L^\infty(X_j)} \right)
\leq C \left( 2c_2 1_{\nu}^{1/2} \left( \frac{h_j}{\eta_j} \right)^\sigma \|e_{j-1}\|_{\Phi_j} + 2\delta_j + M_j \gamma_j \|e_{j-1}\|_{\Phi_j} \right)
\leq C \left( 2c_2 1_{\nu}^{1/2} \left( \frac{1}{\nu} \right)^\sigma \|e_{j-1}\|_{\Phi_j} + 2\delta_j + c \left( \frac{1}{\nu} \right)^\sigma \|e_{j-1}\|_{\Phi_j} \right)
\leq C \left( 2c_2 1_{\nu}^{1/2} \left( \frac{\mu}{T} \right)^\sigma \|e_{j-1}\|_{\Phi_j} + 2\delta_j + c \left( \frac{\mu}{T} \right)^\sigma \|e_{j-1}\|_{\Phi_j} \right)
\leq C \left( \mu^\sigma \|e_{j-1}\|_{\Phi_j} + \delta_j \right).
\]
For the sum \(\sqrt{S_2}\), since \(\eta_{j+1}l > 1\), we have
\[
(1 + \eta_{j+1}l)^{2\sigma} < (2\eta_{j+1}l)^{2\sigma} \leq (2\eta_{j+1})^{2\sigma}(1 + l)^{2\sigma}.
\]
Using (4) and Lemma 5 again, the term $\sqrt{S_2}$ can be estimated as

$$
\sqrt{S_2} < c_1^{-1/2} (2n_{j+1})^\sigma \left( \sum_{l=0}^{\infty} \sum_{k=1}^{N(l)} \left( e_{j-1} - (s^\delta_j)_{lk} \right)^2 (1 + l)^{2\sigma} \right)^{1/2}
$$

$$
= c_1^{-1/2} 2^{\sigma} \eta_{j+1} \|e_{j-1} - s^\delta_j\|_{H^\sigma(S^n)}
$$

$$
\leq c_1^{-1/2} 2^{\sigma+1} c_2^{1/2} \left( \frac{\eta_{j+1}}{\eta_j} \right)^\sigma \|e_{j-1}\|_{\Phi_j}
$$

Combining the two terms, we have

$$
\|e_{j-1} - s^\delta_j\|_{\Phi_{j+1}} \leq C \mu^\sigma \|e_{j-1}\|_{\Phi_j} + C \delta_j := \alpha \|e_{j-1}\|_{\Phi_j} + C \delta_j.
$$

(14)

Then, recalling (13), the second term $\|s^\delta_j - \tilde{s}^\delta_j\|_{\Phi_{j+1}}$ can be estimated as

$$
\|s^\delta_j - \tilde{s}^\delta_j\|_{\Phi_{j+1}} = \| \sum_{m=1}^{M_j} a_{jm} \Phi_j(\cdot, x_{jm}) - \sum_{|a_{jm}| \neq \tau_j} a_{jm} \Phi_j(\cdot, x_{jm}) \|_{\Phi_{j+1}}
$$

$$
= \| \sum_{|a_{jm}| \leq \tau_j} a_{jm} \Phi_j(\cdot, x_{jm}) \|_{\Phi_{j+1}}
$$

$$
\leq M'_j \tau_j \| \Phi_j(\cdot, x_{jm}) \|_{\Phi_{j+1}}.
$$

Here, $M'_j$ is defined by

$$
M'_j := \sup_{x \in S^n} \# \{ m \in \{1, 2, \ldots, M_j\} : |a_{jm}| \leq \tau_j; \theta(x, x_{jm}) < R(\eta_j) \},
$$

where $\#X$ denotes the cardinality of a set $X$ and $R(\eta_j)$ is the radius (in geodesic distance) of the support of the SBF $\Phi_j(\cdot, x_{jm})$. An obvious upper bound for $M'_j$ is $M'_j \leq M_j$. However, due to the local supports of the SBF's, there exists a better upper bound (see Le Gia and Wendland, 2014) Lemma 4.1, especially when $n = 2$, under the assumption $\eta_j = \nu h_j$ and $h_j \leq \rho q_j$, $M'_j$ can be bounded by a constant, i.e.,

$$
M'_j \leq \pi^2 \left( \frac{\eta_j}{d_j} \right)^2 \leq (\pi \nu)^2.
$$

A similar result has already been derived in Le Gia et al. (2010) even for arbitrary space dimension but with weaker constants. Hence, $M'_j$ can be bounded by a constant $\tilde{M}$ if $\eta_j$ is proportional to $q_j$. The following estimate can be derived by Lemma 1 (b) and the choice strategy for $\tau_j$:

$$
\|s^\delta_j - \tilde{s}^\delta_j\|_{\Phi_{j+1}} \leq \tilde{M} \tau_j (c_2/c_1)^{1/2} \eta_j^{-n/2} < \tau.
$$

(15)

The combination of (14) and (15) gives the desired conclusion.  

$\blacksquare$
Theorem 7 Under the condition of Theorem 6, there exists a constant $C$ independent of level index and exact solution $f^*$, such that the $L_2$—error between exact solution $f^*$ and regularized solution $\tilde{f}_k^\delta$ can be bounded by

$$\|f^* - \tilde{f}_k^\delta\|_{L_2(S^n)} \leq \alpha^k \|f^*\|_{\Phi_1} + C \sum_{i=1}^{k} \alpha^{k-i} \delta_i + \frac{1 - \alpha^k}{1 - \alpha} \tau.$$ 

Proof We first split $\|f^* - \tilde{f}_k^\delta\|_{L_2(S^n)}$ into

$$\|f^* - \tilde{f}_k^\delta\|_{L_2(S^n)} = \|e_{k-1} - s_k^\delta\|_{L_2(S^n)} \leq \|e_{k-1} - s_k^\delta\|_{L_2(S^n)} + \|s_k^\delta - \tilde{s}_k^\delta\|_{L_2(S^n)}.$$ 

Referring to Lemma 2 and Lemma 5, the similar technique in Theorem 6 gives the estimate of the first term

$$\|e_{k-1} - s_k^\delta\|_{L_2(S^n)} \leq C \left( h_k^\delta \|e_{k-1} - s_k^\delta\|_{H^\alpha(S^n)} + \|e_{k-1} - s_k^\delta\|_{L^\infty(\mathcal{X}_k)} \right) \leq \alpha \|e_{k-1}\|_{\Phi_k} + C \delta_k.$$ 

Then, referring to Lemma 1 and the strategy of choosing parameters (9), the second term can be bounded by

$$\|s_k^\delta - \tilde{s}_k^\delta\|_{L_2(S^n)} = \sum_{|a_{km}| \leq \tau_k} a_{km} \phi_k(\cdot, x_{km}) \|\phi_m\|_{L_2(S^n)} \leq \tilde{M} \tau_k c_2 \frac{1}{\eta_k} \frac{\eta_k}{\tau} < \tau,$$

where the last inequality is derived from the choice for $\tau_k$ in (11). Therefore, with induction in Theorem 6, we have

$$\|f^* - \tilde{f}_k^\delta\|_{L_2(S^n)} \leq \alpha \|e_{k-1}\|_{\Phi_k} + C \delta_k + \tau \leq \alpha^2 \|e_{k-1}\|_{\Phi_{k-1}} + C(\alpha \delta_{k-1} + \delta_k) + (\alpha + 1) \tau \leq \cdots \leq \alpha^k \|f^*\|_{\Phi_1} + C \sum_{i=1}^{k} \alpha^{k-i} \delta_i + \sum_{i=1}^{k} \alpha^{k-i} \tau.$$ 

Remark 8 In many applications, the data are usually observed at once, which allows us to obtain a full observation data $f^\delta|\mathcal{X}$ with $\|f^\delta - f^*\|_{L^\infty(\mathcal{X})} \leq \delta$. The sampling data set $X$ into a sequence of data sets $X_1, X_2, \ldots$, satisfying $\bigcup_{j=1}^{k} X_j = X$. In this case, we have $\delta_j = \delta$ for $j = 1, 2, \ldots k$ and the conclusion in Theorem 7 turns into

$$\|f^* - \tilde{f}_k^\delta\|_{L_2(S^n)} \leq \alpha^k \|f^*\|_{\Phi_1} + \frac{(1 - \alpha^k)}{1 - \alpha} (C \delta + \tau).$$ 

Moreover, if $\alpha < 1$ and $k$ goes to infinity, the error between multiscale SVR approximation on $S^n$ with data compression and target function $f^*$ can be bounded by $(C \delta + \tau)/(1 - \alpha)$.
5. Multiscale analysis for rough functions on $S^n$

The theoretical analysis in the previous section depends on the \textit{a priori} information of the target function $f^* \in H^\sigma(S^n)$, the regularity index $\sigma$ should be known a priori, which allows us to choose an appropriate RBF $\Phi$ generating $H^{\sigma+1/2}(\mathbb{R}^{n+1})$, then to obtain a reproducing kernel of $H^\sigma(S^n)$ by restricting $\Phi$ on the unit sphere, scaled SBFs will be generated at the same time.

However, in practical situations, the regularity information of $f^*$ is usually unknown beforehand. In this section, we analyze the error estimate for multiscale SVR approximation with data compression under the following conditions. The algorithm is performed using the kernel $\Phi$, which generates $H^\sigma(S^n)$ with $\sigma > n/2$. The target function $f^*$, however, belongs only to $\beta$ for some $n/2 < \beta \leq \sigma$.

We note that the assumption $\beta > n/2$ is necessary to guarantee the boundedness of point-wise evaluation, i.e., $f^*$ is continuous. Let $\Psi$ be the compactly supported RBF, which is the reproducing kernel of $H^{\beta+1/2}(\mathbb{R}^{n+1})$, restricting $\Psi$ to the unit sphere, it becomes the reproducing kernel of $H^\beta(S^n)$. Let the scaled kernels in the form

$$
\Psi_j(x, y) = \eta_j^{-n} \Psi\left(\frac{|x - y|}{\eta_j}\right), \quad j = 1, 2, \ldots, k,
$$

as in (2). It is worth to emphasize that, since $\beta$ is unknown, both kernel $\Psi$ and scaled kernels $\Psi_j$ for $j = 1, 2, \ldots, k$ are only be utilized in theoretical analysis, but not in practical calculation.

We need to summarize some results before the formal convergence.

**Lemma 9** For $0 < \eta_j \leq 1$ and $f \in H^\beta(S^n)$, we have

$$
c_3^{1/2} \|f\|_{\Psi_j} \leq \|f\|_{H^\beta(S^n)} \leq c_4^{1/2} \eta_j^{-\beta} \|f\|_{\Psi_j}, \quad j = 1, 2, \ldots, k. \quad (16)
$$

This result follows from (4). The proof is essentially that of Le Gia et al. (2010) with $\sigma$ replaced by $\beta$. Furthermore, the Lemma 1 also valid by replacing $\sigma$ to $\beta$.

**Lemma 10** (see Le Gia et al., 2012b) Let $f \in H^\beta(S^n)$, $\beta > n/2$ and let $X$ be a finite subset of $S^n$ with separation distance $q_X$, let $\eta \in (0, 1]$ be given. There exists a constant $\kappa$, which depends only on $n$ and $\beta$, such that if $L \geq \kappa \max\{\eta/q_X, 1/\eta\}$, then there is a spherical polynomial $\tilde{p} \in \mathcal{P}_L$ such that $\tilde{p}|_X = f|_X$ and

$$
\|f - \tilde{p}\|_{\Psi_n} \leq 5 \|f\|_{\Psi_n}.
$$

Recalling the Algorithm 1, at the $j$-th level, the local correction $s_j^\delta \in H^\sigma(S^n)$ satisfies (7). However, in this section, $e_{j-1} := f^* - \tilde{s}_1^\delta - \ldots - \tilde{s}_{j-1}^\delta$ only belong to $H^\beta(S^n)$, thus can not be regarded as candidate. Thanks to Lemma 10, choosing $L = \gamma \kappa \max\{\eta_j/q_j, 1/\eta_j\}$, there is a spherical polynomial $\tilde{p}_{j-1} \in \mathcal{P}_L$ such that

$$
\tilde{p}_{j-1}|x_j = e_{j-1}|x_j, \quad \text{and} \quad \|\tilde{p}_{j-1} - e_{j-1}\|_{\Psi_j} \leq 5 \|e_{j-1}\|_{\Psi_j}. \quad (17)
$$
Taking $\bar{p}_{j-1}$ as candidate, and put it into (7), it follows that,

$$\frac{1}{M_j} \sum_{m=1}^{M_j} \| \tilde{e}_{j-1}(x_{jm}) - s_j^\delta(x_{jm}) \|_{\Phi_j} + \gamma_j s_j^\delta \|_{\Phi_j} \leq \frac{1}{M_j} \sum_{m=1}^{M_j} \| \tilde{e}_{j-1}(x_{jm}) - \bar{p}_{j-1}(x_{jm}) \|_{\epsilon_j} + \gamma_j \| \bar{p}_{j-1} \|_{\Phi_j},$$

$$= \frac{1}{M_j} \sum_{m=1}^{M_j} \| \tilde{e}_{j-1}(x_{jm}) - e_{j-1}(x_{jm}) \|_{\epsilon_j} + \gamma_j \| \bar{p}_{j-1} \|_{\Phi_j},$$

$$\leq \frac{1}{M_j} \cdot M_j \cdot (\delta_j - \epsilon_j)^+ + \gamma_j \| \bar{p}_{j-1} \|_{\Phi_j}^2.$$

In addition, using Assumption 3 we have

$$\| \bar{p}_{j-1} \|_{\Phi_j}^2 = \sum_{l=0}^{L} \sum_{k=1}^{N(n,l)} \left| \frac{(\bar{p}_{j-1})_{lk}}{\phi_j(l)} \right|^2 \leq \frac{1}{c_1} \sum_{l=0}^{L} \sum_{k=1}^{N(n,l)} \left| \frac{(\bar{p}_{j-1})_{lk}}{\phi_j(l)} \right|^2 (1 + \eta_j l)^{2\sigma}$$

$$\leq \frac{1}{c_1} (1 + \eta_j L)^{2(\sigma - \beta)} \sum_{l=0}^{L} \sum_{k=1}^{N(n,l)} \left| \frac{(\bar{p}_{j-1})_{lk}}{\phi_j(l)} \right|^2 (1 + \eta_j l)^{2\beta}$$

$$\leq \frac{c_4}{c_1} (1 + \eta_j L)^{2(\sigma - \beta)} \sum_{l=0}^{L} \sum_{k=1}^{N(n,l)} \left| \frac{(\bar{p}_{j-1})_{lk}}{\phi_j(l)} \right|^2 \| \bar{p}_{j-1} \|_{\Phi_j}^2$$

$$= \frac{c_4}{c_1} (1 + \eta_j L)^{2(\sigma - \beta)} \| e_{j-1} \|_{\Phi_j}^2$$

$$\leq \frac{36c_4}{c_1} (1 + \eta_j L)^{2(\sigma - \beta)} \| e_{j-1} \|_{\Phi_j}^2$$

where in the last step we use $\gamma a^\gamma \leq 2a$ for $a \geq 1$. Therefore, as in (8), if we choose cut-off parameter $\epsilon_k = \delta_k$, the following basic estimate can be derived:

$$\frac{1}{M_j} \sum_{m=1}^{M_j} \| \tilde{e}_{j-1}(x_{jm}) - s_j^\delta(x_{jm}) \|_{\epsilon_j} + \gamma_j s_j^\delta \|_{\Phi_j} \leq c^2 \gamma_j \| e_{j-1} \|_{\Phi_j}^2,$$

and consequently, the following lemma is valid.

**Lemma 11** Let the Assumption 3 hold. At each level $j$, choosing the cut-off parameter $\epsilon_j = \delta_j$, then the following estimates hold:

$$\| s_j^\delta \|_{\Psi_j} \leq \left( \frac{c^2 c_2}{c_3} \right)^{1/2} \| e_{j-1} \|_{\Psi_j},$$

$$\| e_{j-1} - s_j^\delta \|_{H^\beta(\mathbb{R}^n)} \leq c_4^{1/2} \left( 1 + \left( \frac{c^2 c_2}{c_3} \right)^{1/2} \right) \eta_j^{-\beta} \| e_{j-1} \|_{\Psi_j},$$

$$\| e_{j-1} - s_j^\delta \|_{L^\infty(\Omega_j)} \leq 2 \delta_j + c^2 M_j \gamma_j \| e_{j-1} \|_{\Psi_j}^2.$$

---

ZHONG, LE GIA AND WANG.
Proof It is directly from basic estimate (18) that
\[ \|s_j^{\delta}\|_{\Psi_j} \leq \overline{c}\|e_{j-1}\|_{\Psi_j}, \]

combining with
\[ \|s_j^{\delta}\|_{\Psi_j}^2 \leq \frac{1}{c_3} \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} |(s_j^{\delta})_{lk}|^2 (1 + \eta_j l)^{2\beta} \leq \frac{1}{c_3} \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} |(s_j^{\delta})_{lk}|^2 (1 + \eta_j l)^{2\sigma} \leq \frac{c_2}{c_3}\|s_j^{\delta}\|_{\Psi_j}^2, \]

the first estimate is consequently obtained. For the second one, the application of Lemma 9 yields,
\[ \|e_{j-1} - s_j^{\delta}\|_{H^\beta(S^n)} \leq c_4^{1/2} \eta_j^{-\beta} \left(\|e_{j-1}\|_{\Psi_j} + \|s_j^{\delta}\|_{\Psi_j}\right) \leq c_4^{1/2} \left(1 + \left(\frac{c_2^2 c_2}{c_3}\right)^{1/2}\right) \eta_j^{-\beta}\|e_{j-1}\|_{\Psi_j}. \]

For the last one, similar to the third estimate in Lemma 5
\[ \|e_{j-1} - s_j^{\delta}\|_{l^\infty(X_j)} \leq \|e_{j-1} - e_{j-1}^{\ast}\|_{l^\infty(X_j)} + \|e_{j-1}^{\ast} - s_j^{\delta}\|_{l^\infty(X_j)} \leq \delta_j + \sum_{m=1}^{M_j} \|e_{j-1}^{\ast}(x_{jm}) - s_j^{\delta}(x_{jm})\|_{\psi_j} + \varepsilon_j \leq 2\delta_j + c_2^2 M_j \gamma_j\|e_{j-1}\|_{\Psi_j}. \]

\[ \mathbf{Theorem~12} \text{ Let the Assumption 3 hold, in addition, assume that the target function } f^* \in H^\beta(S^n) \text{ with } \|f^*\|_{H^\beta(S^n)} < c_3^{1/2}. \text{ At each level } j, \text{ under the parameter choice strategies} \]
\[ \varepsilon_j = \delta_j, \quad \gamma_j \leq \left(\frac{1}{\nu}\right)^\sigma h_j^n, \quad \tau_j < \frac{\tau}{M} \min \left\{\left(\frac{c_1 c_3}{c_2}\right)^{1/2}, \left(\frac{1}{c_2}\right)^{1/2}\right\} \eta_j^{n/2}, \quad (19) \]

where \( \tau \) is a given constant. Then, there exists constant \( C \) independent of level \( j \) and \( f^* \), such that by denoting \( q = C\mu^{-\beta} \), and assume \( q + C\delta + \tau < 1 \), we have
\[ \|e_j\|_{\Psi_{j+1}} \leq q\|e_{j-1}\|_{\Psi_j} + C\delta_j + \tau\eta_j^{n/2} < 1, \quad \text{for } j = 1, 2, \ldots. \quad (20) \]

Proof Firstly, \( \|e_0\|_{\Psi_1} \leq c_3^{-1/2} \|f^*\|_{H^\beta(S^n)} < 1 \), assume \( \|e_{j-1}\|_{\Psi_j} < 1 \), then we are aiming to prove the recursive formula. Similarly to the proof in Theorem 6, we split \( \|e_j\|_{\Psi_{j+1}} \) into
\[ \|e_j\|_{\Psi_{j+1}} = \|e_{j-1} - \Gamma_j\|_{\Psi_{j+1}} \leq \|e_{j-1} - s_j^{\delta}\|_{\Psi_{j+1}} + \|s_j^{\delta} - \Gamma_j\|_{\Psi_{j+1}}, \]

and the first term \( \|e_{j-1} - s_j^{\delta}\|_{\Psi_{j+1}} \) can be decomposed into
\[ \|e_{j-1} - s_j^{\delta}\|_{\Psi_{j+1}} \leq \frac{1}{c_3} \sum_{l=0}^{\infty} \sum_{k=1}^{N(n,l)} \left|\left(e_{j-1}^{\ast} - s_j^{\delta}\right)_{lk}\right|^2 (1 + \eta_j l)^{2\beta} \right)^{1/2} \leq \sqrt{S_1} + \sqrt{S_2}, \]
Under the condition of Theorem 12 and Lemma 12, solution $e_j$ of level $j$.

Therefore, the recursive (20) is obtained by combining these estimates.

For the first sum $\sqrt{S_1}$, under the choice for $\varepsilon_k, \gamma_k$, Assumption 3, Lemma 2 and Lemma 11, it follows that,

$$\sqrt{S_1} \leq c_3^{-1/2} 2^\beta \| e_{j-1} - s_j^\beta \|_{L_2(\mathbb{S}^n)}$$

$$\leq C \left( h_j^\beta \| e_{j-1} - s_j^\beta \|_{H^\beta(\mathbb{S}^n)} + \| e_{j-1} - s_j^\beta \|_{L^\infty(x_j)} \right)$$

$$\leq C \left( c_4^{1/2} \left( 1 + \left( \frac{c_2^2 c_2}{c_3} \right)^{1/2} \right) \left( \frac{1}{\nu} \right)^\beta \| e_{j-1} \|_{\Psi_j} + 2 \delta_j + \bar{c} \left( \frac{1}{\nu} \right)^\beta \| e_{j-1} \|_{\Psi_j} \right)$$

$$\leq C \left( \mu \| e_{j-1} \|_{\Psi_j} + \delta_j \right),$$

and the second sum $\sqrt{S_2}$ can be estimated as

$$\sqrt{S_2} < c_3^{-1/2} 2^\beta \eta_j^\beta \| e_{j-1} - s_j^\beta \|_{H^\beta(\mathbb{S}^n)} \leq C \left( \frac{\eta_j+1}{\eta_j} \right)^\beta \| e_{j-1} \|_{\Psi_j} \leq C \mu \| e_{j-1} \|_{\Psi_j}.$$

Putting the estimate together, we obtain

$$\| e_{j-1} - s_j^\beta \|_{\Psi_j} \leq C \mu \| e_{j-1} \|_{\Psi_j} + C \delta_j := q \| e_{j-1} \|_{\Psi_j} + C \delta_j . \tag{21}$$

The estimate for the second term $\| s_j^\delta - s_j^\beta \|_{\Psi_{j+1}}$ will be derived as

$$\| s_j^\delta - s_j^\beta \|_{\Psi_{j+1}} \leq \left( \frac{c_2}{c_3} \right)^{1/2} \| s_j^\delta - s_j^\beta \|_{\phi_{j+1}}$$

$$= \left( \frac{c_2}{c_3} \right)^{1/2} \| \sum_{|a_{jm}| \leq \tau_j} a_{jm} \phi_j(\cdot, x_{jm}) \|_{\phi_{j+1}}$$

$$\leq \frac{c_2}{(c_1 c_3)^{1/2}} \bar{M} \tau_j \eta_j^{-n/2} < \tau .$$

Therefore, the recursive (20) is obtained by combining these estimates.

**Theorem 13** Under the condition of Theorem 12, there exists a constant $C$ independent of level $j$ and $f^*$, such that the $L_2$—error between rough exact solution $f^*$ and regularized solution $\tilde{f}_k^*$ can be bounded by

$$\| f^* - \tilde{f}_k^* \|_{L_2(\mathbb{S}^n)} \leq q^k \| f^* \|_{\Psi_i} + C \sum_{i=1}^k q^{k-i} \delta_i + \frac{1 - q^k}{1 - q} \tau .$$
**Proof** As in proof in Theorem 7, we split \( \| f^* - \tilde{f}_k^\delta \|_{L^2(S^n)} \) into
\[
\| f^* - \tilde{f}_k^\delta \|_{L^2(S^n)} = \| e_{k-1} - s_k^\delta \|_{L^2(S^n)} \leq \| e_{k-1} - s_k^\delta \|_{L^2(S^n)} + \| s_k^\delta - \tilde{s}_k^\delta \|_{L^2(S^n)}.
\]
Utilizing Lemma 2 and Lemma 11 again, we know
\[
\| e_{k-1} - s_k^\delta \|_{L^2(S^n)} \leq C \left( h_k^\delta \| e_{k-1} - s_k^\delta \|_{H^3(S^n)} + \| e_{k-1} - s_k^\delta \|_{H^\infty(X_k)} \right)
\leq q \| e_{k-1} \| \psi_k + C \delta_k,
\]
and the second term
\[
\| s_k^\delta - \tilde{s}_k^\delta \|_{L^2(S^n)} = \| \sum_{|a_{km}| \leq \tau_k} a_{km} \Phi_k(\cdot, x_{km}) \|_{L^2(S^n)} \leq \tilde{M} \tau_k c_2^{1/2} \eta_k^{-n/2} < \tau.
\]
where the last inequality is derived by the choice for \( \tau_k \) in (19). Therefore, with recursive formula (20), we have
\[
\| f^* - \tilde{f}_k^\delta \|_{L^2(S^n)} \leq q \| e_{k-1} \| \psi_k + C \delta_j + \tau
\leq \cdots \leq q^k \| f^* \| \psi_1 + C \sum_{i=1}^k q^{k-i} \delta_i + \sum_{i=1}^k q^{k-i} \tau.
\]

**Remark 14** Under the conditions in Theorem 13, moreover, assume \( \delta_j = \delta \) for \( j = 1, 2, \ldots, k \) and \( q < 1 \), if \( k \) goes to infinity, the error between multiscale SVR approximation on \( S^n \) with data compression and rough target function \( f^* \) can be bounded by \((C \delta + \tau)/(1 - q)\).

6. Numerical Experiments

6.1 Exact observations

In our first example, we will test our compression multiscale SVR algorithm on the topographical data of the earth, which is available from the National Geophysical Data Center, NOAA US Department of Commerce. The data is available in the Matlab file called `topo.mat`, which is a matrix of size (180, 360). The exact observation values (i.e., \( \delta = 0 \)) represent the average heights in meters above and below the mean sea level.

We generate the sets \( X_1, \ldots, X_5 \), the number of points \( M_j \), mesh norms \( h_j \) and separation radii \( q_j \) of these sets are listed in Table 1.

The RBF used in the experiment is the Wendland function
\[
\Phi(x, y) = (1 - |x - y|)^4_+ (4|x - y| + 1),
\]
which generates \( H^3(\mathbb{R}^3) \), when we restrict it into the unit sphere \( S^2 \), it becomes an SBF generating the RKHS \( H^{2.5}(S^2) \), the scaled kernels are defined as in (2). We remark that the target function does not belong to the RKHS.
Table 1: (Exact observation) Quantity information at each level for exact observation.

<table>
<thead>
<tr>
<th>level</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>64</td>
<td>250</td>
<td>1100</td>
<td>4000</td>
<td>7500</td>
</tr>
<tr>
<td>$h$</td>
<td>0.3337</td>
<td>0.1643</td>
<td>0.0830</td>
<td>0.0480</td>
<td>0.0368</td>
</tr>
<tr>
<td>$q$</td>
<td>0.2120</td>
<td>0.0933</td>
<td>0.0413</td>
<td>0.0179</td>
<td>0.0093</td>
</tr>
</tbody>
</table>

According to our algorithm of discarding dynamically, at level $j$, we set the cut-off parameter $\varepsilon_j = 0$, the regularized parameter $\gamma_j = O(h_j^2)$ and the threshold to $\tau_j = K\eta_j$, where $K$ is a level-independent given constant. More precisely, at level $j$ all the coefficients $a_{jm}$ with $|a_{jm}| \leq K\eta_j$ are set to 0. It is obvious that for $K = 0$, no extra coefficients are discarded and the number of discarded coefficients increases with $K$. In our experiments, we set $K = 0, 50, 100$ respectively. In each case, we denote $D_j$ be the number of discarded coefficients at the $j$-th level. The reconstructions results are given in Table 2, Figure 1 and Figure 2. The numerical errors are computed via

$$
\|\tilde{e}_j\|_{L^2} := \left(\frac{1}{4\pi} \frac{2\pi^2}{|\mathcal{G}|} \sum_{x \in \mathcal{G}} |f^*(\theta, \varphi) - \tilde{f}_j^S(\theta, \varphi)|^2 \sin \theta \right)^{1/2},
$$

where $\theta$ and $\phi$ are spherical coordinate components and $\mathcal{G}$ is the grid containing the centers of rectangles of size 1 degree times 1 degree and $|\mathcal{G}| = 64800$.

Table 2: Results for discarding dynamically.

<table>
<thead>
<tr>
<th>$K$</th>
<th>$D$</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$|\tilde{e}|_{L^2}$</td>
<td>3359</td>
<td>2929</td>
<td>1382</td>
<td>493</td>
<td>349</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0</td>
<td>15</td>
<td>112</td>
<td>1496</td>
<td>6775</td>
</tr>
<tr>
<td></td>
<td>$|\tilde{e}|_{L^2}$</td>
<td>3359</td>
<td>2929</td>
<td>1382</td>
<td>515</td>
<td>419</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>64</td>
<td>32</td>
<td>200</td>
<td>2449</td>
<td>7340</td>
</tr>
<tr>
<td></td>
<td>$|\tilde{e}|_{L^2}$</td>
<td>3449</td>
<td>3001</td>
<td>1436</td>
<td>676</td>
<td>640</td>
</tr>
</tbody>
</table>
Figure 1: (Exact observation) Multiscale SVR approximation without data compression \( \tilde{f}_3 - \tilde{f}_5 \), and final error \( \tilde{f}_5 \).
Figure 2: (Exact observation) Compression Multiscale SVR approximation $\tilde{f}_5$, and final errors $\tilde{e}_5$, with $K = 50$ and $K = 100$, respectively.
6.2 Noisy observations

In our second example, we test the Multiscale SVR Algorithm with data compression under noisy observations. The target function is the sum of the Franke function modified by Renka (1988)

\[
f_1(x, y, z) = 0.75 \exp\left(-\frac{(9x - 2)^2}{4} - \frac{(9y - 2)^2}{4} - \frac{(9z - 2)^2}{4}\right) \\
+ 0.75 \exp\left(-\frac{(9x + 1)^2}{49} - \frac{(9y + 1)^2}{10} - \frac{(9z + 1)^2}{10}\right) \\
+ 0.5 \exp\left(-\frac{(9x - 7)^2}{4} - \frac{(9y - 3)^2}{4} - \frac{(9z - 5)^2}{4}\right) \\
- 0.2 \exp\left(-\frac{(9x - 4)^2 - (9y - 7)^2 - (9z - 5)^2}{4}\right).
\]

and the function \( f_0 : \mathbb{S}^2 \to \mathbb{C} \) for \( x = (x, y, z) \in \mathbb{S}^2 \) by

\[
f_0(x) = \begin{cases} 
0.5, & a \leq z \leq 1; \\
0, & -1 \leq z < a,
\end{cases}
\]

with a fixed \( 0 < a < 1 \), in our example, we choose \( a = 0.5 \). The function \( f_0 \) belongs to \( H^s(\mathbb{S}^2) \) for all \( s < 1/2 \) (see Le Gia and McLean, 2012). The simulated noise values are given as the components of a random vector, each component is uniformly distributed on \([-0.2, 0.2]\], with noise level \( \delta = 0.2 \).

We generate the sets \( X_1, \ldots, X_4 \), the number of points \( M_j \), mesh norms \( h_j \) and separation radii \( q_j \) of these sets are listed in Table 3.

Table 3: (Noisy observation) Quantity information at each level for noisy observation.

<table>
<thead>
<tr>
<th>level</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M )</td>
<td>32</td>
<td>250</td>
<td>400</td>
<td>2000</td>
</tr>
<tr>
<td>( h )</td>
<td>0.4371</td>
<td>0.2288</td>
<td>0.1327</td>
<td>0.0613</td>
</tr>
<tr>
<td>( q )</td>
<td>0.3019</td>
<td>0.1280</td>
<td>0.0632</td>
<td>0.0262</td>
</tr>
</tbody>
</table>

The RBF used in the experiment is the Wendland function

\[
\Phi(x, y) = (1 - |x - y|)^2_+,
\]

which generates \( H^2(\mathbb{R}^3) \), when we restrict it into the unit sphere \( \mathbb{S}^2 \), it becomes a SBF generating the RKHS \( H^{1.5}(\mathbb{S}^2) \), the scaled version as in (2). We remark that the target function does not belong to the RKHS.

According to our algorithm of discarding dynamically, at level \( j \), we set the cut-off parameter \( \varepsilon_j = 0.2 \), the regularized parameter \( \gamma_j = O(h_j^2) \), the threshold to \( \tau_j = 1 \times 10^{-5} \). The reconstructions results are given in Table 4 and Figure 3. This time we compute the relative errors defined by \( \|e_j^\delta\|_{L_2}/\|f^*\|_{L_2(\mathbb{S}^2)} \). Referring to the results, it is clear that most of the coefficients are discarded at higher levels. Moreover, the error at each level does not vary significantly from the error without discarding though, for \( \tau_j = 1 \times 10^{-5} \), a total number of 2000 of 2682 coefficients are discarded, which corresponds to approximately 75%.

Acknowledgments

23
Figure 3: (Noisy observation) Compression Multiscale SVR approximation $\tilde{f}_4^\delta$ and exact solution $f^*$, errors $e_1^\delta - \tilde{e}_4^\delta$, with $\tau_j = 1 \times 10^{-5}$.

Table 4: Results for discarding dynamically.

<table>
<thead>
<tr>
<th>$K$</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$D$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$|e_j^\delta|_{L_2}$</td>
<td>0.2404</td>
<td>0.1528</td>
<td>0.0911</td>
</tr>
<tr>
<td>$1 \times 10^{-5}$</td>
<td>$D$</td>
<td>18</td>
<td>97</td>
<td>328</td>
</tr>
<tr>
<td></td>
<td>$|e_j^\delta|_{L_2}$</td>
<td>0.2404</td>
<td>0.1528</td>
<td>0.0911</td>
</tr>
</tbody>
</table>
The author M. Zhong was supported by the National Natural Science Foundation of China (No. 11501102) and Natural Science Foundation of Jiangsu Province (No. BK20150594). The author W. Wang was supported by the National Natural Science Foundation of China (No. 11401257) and Natural Science Foundation of Zhejiang Province (No. LQ14A010013).
References


