Research Article

Continuous Regularized Least Squares Polynomial Approximation on the Sphere

Yang Zhou and Yanan Kong

1School of Mathematics and Statistics, Shandong Normal University, Jinan, China
2Foreign Languages Department, Shandong College of Tourism and Hospitality, Jinan, China

Correspondence should be addressed to Yanan Kong; yanankong@126.com

Received 16 April 2020; Revised 9 July 2020; Accepted 21 July 2020; Published 20 August 2020

Abstract

In this paper, we consider the problem of polynomial reconstruction of smooth functions on the sphere from their noisy values at discrete nodes on the two-sphere. The method considered in this paper is a weighted least squares form with a continuous regularized least squares scheme (CRLS) as function approximation on the sphere. The sets of all functions spanned by spherical polynomials with degree up to a fixed integer \( L \) and order \( k \), which means

\[
\mathbb{P}_L = \{ \text{spherical polynomials with degree} \leq L \}
\]

where \( Y_{\ell k} \) is a fixed \( L^2 \)-orthonormal real spherical harmonic of degree \( \ell \) and order \( k \), which means

\[
\int_{S^2} Y_{\ell k}(x)Y_{\ell' k'}(x)\,d\omega(x) = \delta_{\ell,\ell'}\delta_{k,k'}, \quad \ell, \ell' = 0, \ldots, \ell + 1, \quad k, k' = 1, \ldots, 2\ell + 1,
\]

where \( d\omega(x) \) denotes the surface measure on \( S^2 \) and \( \delta_{\ell,k} \) is the Kronecker delta. It is well known that the dimension of \( \mathbb{P}_L \) is

\[
d_L := \dim(\mathbb{P}_L) = \sum_{\ell=0}^{L} (2\ell + 1) = (L + 1)^2.
\]

The Sobolev space \( \mathbb{H}^s = \mathbb{H}^s(S^2) \) is defined for \( s \geq 0 \) as the set of all functions \( f \in L^2(S^2) \) whose Laplace–Fourier coefficients

1. Introduction

In recent decades, the problem of finding a polynomial approximation \( g : S^2 \to \mathbb{R} \) with degree of \( g \) up to an integer \( L \) of a continuous function \( f \in C(S^2) \) on the sphere \( S^2 = \{ x \in \mathbb{R}^3 \colon ||x|| = 1 \} \) has received much attention [1–6]. Given noisy data \( f^\delta(x_i) \) at points \( x_i \in S^2, i = 1, \ldots, N \), regularized least squares schemes are usually considered, whereas the regularization varies for different tasks. In this paper, we consider approximating the function \( f \) with a continuous regularized least squares scheme (CRLS) as

\[
\min_{g \in \mathbb{P}_L} \sum_{i=1}^{N} w_i [g(x_i) - f^\delta(x_i)]^2 + \lambda ||g||^2_{H^s},
\]

where \( w_i, i = 1, \ldots, N \) are positive scalars as weights of the data fitting term. For normalization, we can always assume that \( \sum_{i=1}^{N} w_i = N \). Moreover, \( \lambda \) is the regularized parameter to balance the data fitting term and the regularized term and \( \mathbb{P}_L \) denotes the polynomial space.
\[
\tilde{f}_{\ell k} = \langle f, Y_{\ell k} \rangle_{L_2(S^2)} = \int_{S^2} f(x) Y_{\ell k}(x) \, d\omega(x)
\]  
(5)

satisfy
\[
\sum_{\ell=0}^{\infty} \sum_{k=1}^{2\ell+1} (1 + \lambda_\ell) |\tilde{f}_{\ell k}| < \infty,
\]  
(6)

where \( \lambda_\ell = \ell (\ell + 1) \), and the norm of \( H^q(S^2) \) can be defined as
\[
\|f\|_{H^q} \approx \left[ \sum_{\ell=0}^{\infty} \sum_{k=1}^{2\ell+1} \phi_{\ell}^2 \tilde{f}_{\ell k}^2 \right]^{1/2},
\]  
(7)

where the positive parameters \( \phi_{\ell} \) are decreasing in \( \ell \) and satisfy
\[
\phi_{\ell}^2 (1 + \lambda_\ell)^{-2} = (\ell + 1)^{-2}.  
\]  
(8)

Here, we say that \( a \approx b \) if there exist positive constants \( c_1 \) and \( c_2 \), independent of \( n \), such that \( c_1 a_n \leq b_n \leq c_2 a_n \) for all \( n \), and \( \delta > 0 \). Obviously, by letting \( \delta = 0 \), we can obtain \( H^q(S^2) = L_2(S^2) \). CRLS is to obtain the approximation of \( f \) with fitting the data values \( f^0(x_i), w_i = 1, \ldots, N \), and at the same time, we keep the approximation smoother. The motivation is often from the geophysical quantity problems, such as satellite gravity gradiometry problem ([7], pp. 120, 262), in which a smoother polynomial representation of the gravitational potential than the Earth’s surface is to be found. On the other hand, CRLS can be seen as a generalization or approximation of some smooth-oriented regularized least squares models (see the appendix for discussions on some polynomial approximation models relative to this paper).

Polynomial approximation problems arise in a variety of engineering areas, for which different models are used. To recover smooth functions on \( S^2 \), a discrete regularized least squares method (DRLS) is proposed in [1, 5] as
\[
D_f = \arg \min_{g \in P_L} \sum_{j=1}^{N} w_j \left[ g(x_j) - f^0(x_j) \right]^2
\]
\[
+ \lambda \sum_{j=1}^{N} w_j \left( L_1 g(x_j) \right)^2,
\]  
(9)

where \( R_1 \colon P_L \rightarrow P_L \) is a linear “penalization” operator, which can be chosen in different ways. The points used are positive-weight cubature formula that is exact for all spherical polynomials of degree up to \( 2L \), which will guarantee the uniqueness of the solution and furthermore derive its explicit form. A reconstruction error bound is obtained for the case that the nodes and weights establish a cubature formula that is exact for all spherical polynomials of degree up to \( 2L \). To our best knowledge, the number of points needed to construct a \( 2L \)-degree polynomial-exact cubature on \( S^2 \) is around \((2L + 1)^2/3 \) [8], which is bigger than the dimension of the polynomial space \( P_L \) in most practical cases. However, in real application, it is usually costly to acquire as much data points as expected, which motivates us to study the case that the number of data points is not large. In this paper, we do not assume that \( N \geq (L + 1)^2 \), that is, the number of data values is not necessary to reach or exceed the dimension of the polynomial space \( P_L \). The regularization term could still guarantee the uniqueness of the minimizer of CRLS when \( N < (L + 1)^2 \) and avoid the overfitting.

The rest of this paper is organized as follows. In Section 2, we discuss the solution and approximation error of CRLS and the condition number of the linear system it derives. Based on the results, we further give a discussion on spherical \( t_\epsilon \)-designs applied to this model. Numerical experiments are proposed in Section 3 in which the model is tested for its efficiency and accuracy.

## 2. Solution, Approximation Error, and Condition Number

### 2.1. Solution of CRLS.

For simplicity, denote \( \beta_{\ell} = 1/\phi_{\ell}^2 \) for \( \ell = 0, \ldots, L \). A matrix form of problem (1) could be written as
\[
\min_{c \in \mathbb{R}^{(L+1)^2}} \| W^{1/2} Y c - W^{1/2} f^0 \|^2 + \lambda \| B^{1/2} c \|^2,
\]  
(10)

where \( Y \in \mathbb{R}^{N \times (L+1)^2} \) is defined as \( Y_{\ell \epsilon + k} = Y_{\ell k}(x_i) \), \( W = \text{diag}(w_1, \ldots, w_N) \in \mathbb{R}^{N \times N} \), \( f^0 = (f^0(x_1), \ldots, f^0(x_N))^T \in \mathbb{R}^N \), and
\[
B = \text{diag}(\beta_0, \beta_1, \beta_1, \ldots, \beta_{\ell}, \ldots, \beta_{L}, \ldots, \beta_L)
\]  
\[
\in \mathbb{R}^{(L+1)^2 \times (L+1)^2}.
\]  
(11)

Problem (1) is convex, and thus by deriving its first-order optimality condition, we can obtain that solving problem (1) leads to find a solution of the linear system
\[
T c = Y^T W f^0,
\]  
(12)

with
\[
T = Y^T W Y + \lambda B.
\]  
(13)

**Remark 1.** For \( \lambda > 0 \), equation (12) will always have a unique solution by the fact that \( Y^T W Y = (W^{1/2} Y)^T W^{1/2} Y \) is positive semidefinite and \( B \) is positive definite.

By the uniqueness of the solution of problem (1), we denote the minimizer of it by
\[
f^0 = \arg \min_{g \in P_L} \sum_{i=1}^{N} w_i \left[ g(x_i) - f^0(x_i) \right]^2 + \lambda \| g \|^2_{H^q}.
\]  
(14)

Furthermore, when \( \lambda = 0 \), problem (1) may have multiple solutions. In this case, \( f^0 \) represents the uniquely determined minimizer in \( P_L \) of (1) with minimal norm.

### 2.2. Approximation Error.

For the case that \( N \geq (L + 1)^2 \) and \( X \) is a spherical \( 2L \)-design or a cubature rule with \( 2L \)-degree polynomial exactness, the error bounds are discussed in
[1, 3]. Model (1) can be seen as a variant of models considered in [1, 3], in which the difference is that we choose different regularization terms. Furthermore, to propose the error bounds, we do not need the assumption about the polynomial exactness the data point needs to have. The only assumption we need is on the mesh norm of the data point set $X \subseteq \mathbb{S}^d$, which can be defined as follows.

**Definition 1** (mesh norm). The mesh norm $h_X$ of a point set $X \subseteq \mathbb{S}^d$ is

$$h_X := \max_{y \in \mathbb{S}^d} \min_{x \in X} \text{dist}(y, x).$$

The mesh norm could be regarded as the largest geodesic distance of all points on $\mathbb{S}^d$ to their nearest $x \in X$ or as the radius of the largest spherical cap that does not contain any $x \in X$. The results in this section are based on a slight generalization of Theorem 5.1 in [9] which presents an $L_2$ sampling inequality on the sphere $\mathbb{S}^d$.

**Theorem 1.** Let $d \geq 2$ and let $s > d/2$. There exist positive constants $\varepsilon_s$ and $\bar{c}_s$ (depending on $d$ and $s$) such that for any finite point set $X = \{x_1, \ldots, x_N\}$ of $N$ distinct points on $\mathbb{S}^d$ with $h_X \leq \varepsilon_s$ and for any function $g \in \mathbb{H}^s$,

$$\|g\|_{L_2} \leq \bar{c}_s \left[ h_X^2 \|g\|_{L_2} + h_X^{d/2} \left( \sum_{i=1}^N \|g(x_i)^2\| \right)^{1/2} \right].$$

(15)

For the setting of weighted least squares, we slightly revise the above theorem as the following.

**Theorem 2.** Let $d \geq 2$, $s > d/2$, and $w \in [1 - \varepsilon, (1 - \varepsilon)^{-1}]^N$ with $0 \leq \varepsilon < 1$. There exist positive constants $\varepsilon_s$ and $\bar{c}_s$ (depending on $d$ and $s$) such that for any finite point set $X = \{x_1, \ldots, x_N\}$ of $N$ distinct points on $\mathbb{S}^d$ with $h_X \leq \varepsilon_s$ and for any function $g \in \mathbb{H}^s$,

$$\|g\|_{L_2} \leq \bar{c}_s \left[ h_X^2 \|g\|_{L_2} + h_X^{d/2} \left( \sum_{i=1}^N w_i \|g(x_i)^2\| \right)^{1/2} \right].$$

(16)

Proof. The gap between Theorem 5.1 in [9] and the above one is the involvement of the weight vector $w = (w_1, \ldots, w_N)$. By the fact that $w_i \geq 1 - \varepsilon$ for $\forall i = 1, \ldots, N$ and inequality (16), we can obtain the conclusion.

**Theorem 3.** Let $s > 1$. For a point set $X = \{x_1, \ldots, x_N\} \subseteq \mathbb{S}^d$ of $N$ distinct points, a weight vector $w = (w_1, \ldots, w_N)^T \in [1 - \varepsilon, (1 - \varepsilon)^{-1}]^N$ with $0 \leq \varepsilon < 1$ and a function $f \in \mathbb{H}^s$ with values $f = (f(x_1), \ldots, f(x_N))^T$ on $X$, and given a vector of deterministic noise $\delta = (\delta_1, \ldots, \delta_N)^T$ and given $\lambda \geq 0$, there exist positive constants $\varepsilon_s$ and $\bar{c}_s$ (dependent on $s$) such that if the point set $X$ satisfies $h_X \leq \varepsilon_s$, the following holds:

$$\left\| f^\delta - f \right\|_{L_2} \leq c_s \left[ h_X^2 \left( \frac{1}{2 \sqrt{\lambda}} \|\delta\|_2 + \|f\|_{L_2} \right) + \frac{h_X^{d/2}}{1 - \varepsilon} \left( \|\delta\|_2 + \sqrt{\lambda} \|f\|_{L_2} \right) \right].$$

(18)

Proof. For simplicity, we define the norm $\|w\|_w$ as $y^T W y$ for a vector $y \in \mathbb{R}^N$. Note that the norm is well defined since $W$ is positive definite. Define the quadratic functional $\mu(g)$ as

$$\mu(g) := \mu_{\lambda \delta}(g) = \sum_{i=1}^N w_i \left[ g(x_i) - f^\delta(x_i) \right]^2 + \lambda \|g\|_{L_2}^2, \quad g \in P_L.$$  

(19)

Let $f^\delta, f \in \mathbb{R}^N$ be the vector of function value of $f^\delta, f$ at the point set $X$, respectively. By sampling inequality (17) in Theorem 2, we can obtain that there exists a positive constant $\bar{c}_s$ such that

$$\left\| f^\delta - f \right\|_{L_2} \leq \bar{c}_s \left[ h_X^2 \left\| f^\delta - f \right\|_{L_2} + \frac{h_X^{d/2}}{1 - \varepsilon} \||f^\delta - f||_{L_2} \right].$$

(20)

Then, by the fact that $f^\delta$ is the minimizer of $\mu(g)$, the following holds:

$$\left\| f^\delta - f \right\|_{L_2}^2 = \left\| f^\delta - f + \delta \right\|_{L_2}^2$$

$$\leq 2 \left\| f^\delta - f \right\|_{L_2}^2 + 2 \|\delta\|_w^2$$

$$= 2 \mu(f^\delta) - 2 \lambda \|f^\delta\|_{L_2}^2 + 2 \|\delta\|_w^2$$

$$\leq 2 \mu(f) + 2 \|\delta\|_w^2$$

$$= 2 \lambda \|f\|_{L_2}^2 + 4 \|\delta\|_w^2.$$  

(21)

Moreover, we have

$$\left\| f^\delta \right\|_{L_2}^2 = \frac{1}{\lambda} \left[ \mu(f^\delta) - \left| f^\delta - f \right|_{L_2}^2 \right]$$

$$\leq \frac{1}{\lambda} \left[ \mu(f) - \left| f^\delta - f \right|_{L_2}^2 \right]$$

$$= \frac{1}{\lambda} \left[ \|\delta\|_w^2 + \lambda \|f\|_{L_2}^2 - \left| f^\delta - f \right|_{L_2}^2 \right]$$

$$\leq \frac{1}{\lambda} \|\delta\|_w^2 + \lambda \|f\|_{L_2}^2.$$  

(22)

Thus, the term $\left\| f^\delta - f \right\|_{L_2}$ can be bounded as

$$\left\| f^\delta - f \right\|_{L_2} \leq \frac{1}{\lambda} \left[ \|\delta\|_w^2 + \lambda \|f\|_{L_2}^2 \right]$$

$$= \left( \frac{1}{\lambda} \|\delta\|_w^2 + \lambda \|f\|_{L_2}^2 \right)^{1/2} + \|f\|_{L_2}$$

$$\leq \frac{1}{\lambda} \|\delta\|_w^2 + 2 \lambda \|f\|_{L_2}.$$  

(23)

Substitute inequalities (21) and (23) into (20) with $\bar{c}_s = 2c_s$, and we have
where the last inequality is derived by the fact that \( \|\delta\|_{\infty} \leq (1/1 - \epsilon)\|\delta\|_{2} \).

From the theorem, it is easy to see that the data point set \( X = \{x_1, x_2, \ldots, x_N\} \) and the weight \( w = (w_1, w_2, \ldots, w_N)^T \in \mathbb{R}^N \) need to be chosen properly to improve the approximation quality of the model. The data set \( X \) should have its mesh norm as small as possible and \( w \) should be near to its mean value. Furthermore, one can assume that \( X \) and \( w \) establish a cubature rule with polynomial exactness so that it can give a fast way to find the solution of the problem. By this view, spherical \( t_{\epsilon} \)-designs which are proposed and studied in \([6, 10–14]\) are a suitable choice in this paper.

**Definition 2** (spherical \( t_{\epsilon} \)-design). A spherical \( t_{\epsilon} \)-design with \( 0 < \epsilon < 1 \) on \( S^d \) is a set of points \( X: = \{x_1, \ldots, x_N\} \subset S^d \), such that the quadrature rule

\[
\frac{1}{N} \sum_{i=1}^{N} w_i \rho(x_i) = \frac{1}{|S^d|} \int_{S^d} \rho(x) d\sigma_{d}(x) \quad (25)
\]

is exact for all spherical polynomials \( \rho \) of degree at most \( t \), with the weight vector \( w = (w_1, \ldots, w_N)^T \) satisfying

\[
1 - \epsilon \leq w_i \leq (1 - \epsilon)^{-1}, \quad i = 1, \ldots, N. \quad (26)
\]

Naturally, we can choose a spherical \( t_{\epsilon} \)-design as the data set and its corresponding weight \( w \) as the weight in problem (1). Moreover, solving problem (1) can become trivial when \( t \) is large enough.

**Theorem 4.** Let \( X \) be a spherical \( t_{\epsilon} \)-design with \( t \geq 2L \) and \( w \) be the vector of weights satisfying (25) and (26) with respect to \( X \). Then,

\[
(\mathbf{Y}^T \mathbf{W})_{\ell_{2} \rightarrow \ell_{2}} \leq \frac{N}{4\pi} \mathbf{I}_{(L+1)^2},
\]

and (1) has the unique solution \( g = \sum_{\ell=0}^{L} \sum_{k=1}^{2^\ell+1} a_{\ell k} Y_{\ell, k}(x) \) with

\[
a_{\ell k} = \frac{4\pi(x_k)}{N + 4\pi\lambda_{\ell}^2} \sum_{i=1}^{N} w_i Y_{\ell, k}(x_i) f^\delta(x_i). \quad (28)
\]

**Proof.** Note that when \( X \) is a spherical \( t_{\epsilon} \)-design,
\[
\min_{a \in \mathbb{R}^2} \|Aa - b\|_2^2,
\]
with
\[
A = \begin{bmatrix} W^{1/2}Y \sqrt{1} \end{bmatrix} \in \mathbb{R}^{(N+d_1)xN},
\]
\[
b = \begin{bmatrix} W^{1/2}f \ 0 \end{bmatrix} \in \mathbb{R}^{N+d_1}.
\]

The perturbation bound could then be achieved by applying the standard least squares perturbation bound ([16], Theorem 1.4.6) as
\[
\|a - a^0\|_2 \leq \frac{\|A^T(b - b^0)\|_2}{\|A^Tb\|_2} \kappa(A^TA) = \frac{\|\gamma^TW(f - f^0)\|_2}{\|YWf\|_2} \kappa(T).
\]

When \(X\) is a spherical \(t_x\)-design with corresponding weight \(w = (w_1, \ldots, w_N)^T\) with \(t \geq 2L\), by (27), we have that \(Y^TWY\) is a scalar multiple of the identity matrix, and from (13), we find
\[
\kappa(T) = \frac{N + 4\pi\lambda^2}{N + 4\pi\lambda^2}. \quad (37)
\]

Therefore, estimate (31) of the condition number \(\kappa(T)\) is sharp. It can be directly obtained by the above theorem that \(\kappa(T) = \tilde{\omega},\) which induces that \(s\) should be chosen properly as a balance of smoothing and solvability of the problem. However, the choice of \(s\), which may depend on the nature of target function \(f \in C(S^2)\) and the scale of noise, will be an interesting question to be solved in the future. \(\square\)

3. Numerical Experiments

In this section, we report the numerical results for the CRLS model proposed in this paper. To illustrate the effectiveness of the model, we first compare the approximation errors of our model and the DRRLS model (9) as proposed in [1, 5]. The computation is implemented in Matlab 2016a and performed on a HP notebook computer equipped with Intel Core i7-6500U 2.5G Hz CPU, 16 GB RAM running Windows 10. The solver for linear systems generated by the two models is selected as the left matrix divide command embedded in Matlab, which indeed performs fast in the experiment.

For the choice of the regularized parameter \(\lambda\) and the penalized parameters \(\beta, \kappa, \iota = 0, \ldots, L\) for the DRRLS model, one can refer to [5], in which \(\beta, \kappa\) are chosen based on the theoretical error bounds and real applications and \(\lambda\) is chosen by the balancing principle. The parameter choice of CRLS is also of great importance for the approximation quality of our model, and this will be a crucial problem we will tackle in the future work. The task for the numerical experiment here is to compare the numerical performance between CRLS and DRRLS. What should also be noticed in the numerical experiments is that for an approximation \(\mathcal{H}f\) of \(f\), both the uniform and \(L_2\) approximation errors as \(\|\mathcal{H}f - f\|_{C(S^2)}\) and \(\|\mathcal{H}f - f\|_{L_2}\) are difficult to obtain exactly. In such case, the errors will be approximately computed by
\[
\|f - \mathcal{H}f\|_{L_2} = \left[ \frac{4\pi}{N} \sum_{i=1}^{N} (f(x_i) - \mathcal{H}f(x_i))^2 \right]^{1/2},
\]
(38)

where \(X = \{x_1, \ldots, x_N\} \subset \mathbb{S}^2\) is a large-scaled and well-distributed point set and we use it to estimate the errors. In this paper, \(X\) is selected as a maximal determinant (MD) (see [6, 17]) point set with \(10^4\) points.

Example 1. In the first example, we will test the two models, CRLS and DRRLS, and record their approximation errors for comparison. The data point set \(X\) is also selected as the MD node of size 900. The penalized parameters are selected as \(\beta_t = (1 + \ell((+1))^t\) and \(L = 29\) so that both models will achieve unique minimizers. The target function supposed to approximate is the Franke function [18]:
\[
f_1(x) = \sum_{i=1}^{4} a_i \exp\left(-\sum_{i=1}^{4} b_i(x-i)^2\right) + \sum_{i=1}^{4} c_i(x-i)^d,
\]
\[
(x, y, z) \in \mathbb{S}^2,
\]
\[
\mathcal{H}f - f \|_{L_2} \equiv \|f - \mathcal{H}f\|_{L_2},
\]
which is not a spherical polynomial but continuously differentiable on the whole sphere. The noises in the data \(f_\delta\) obey a uniform distribution in \([-0.5, 0.5]\) i.i.d. Images of \(f_1\) and its noisy version are shown in Figures 1(a) and 1(b).

Figure 2 shows uniform and \(L_2\) approximation errors of the both models for \(s = 0.15\), \(j = 1, \ldots, 20\) and \(\lambda = 10^{-3.5+x.134}, k = 1, \ldots, 20.\) It should be noted that both models will only be effective to recover the target function by choosing the parameters \(s, \lambda\) properly. Therefore, thresholds as a success of the approximation are introduced into the figure, that is, \(L_2\) error \(\leq 1.75\) and uniform error \(\leq 1\). As can be seen from the figure, CRLS shows its effectiveness and robustness in this example, which seems to be achieving success with more choices of the parameters than
Also, the deep blue area of the first row is larger than that in the second row, which indicates that CRLS has more chance than DRLS to achieve small errors.

**Example 2.** In this example, we will focus on the solving speed of the linear systems induced by the models. The condition number and CPU time about the linear systems derived by CRLS and DRLS will be reported. Two types of data point set with different scales will be chosen in this example. One is a well-distributed type which is used in Example 1, and the other type is ill-distributed, called Gauss–Legendre point set (GL) [19] with Gauss points in cos(polar angle) and equally spaced points in azimuthal angle.

In Example 1, the two models cannot achieve each one’s smallest errors in the same setting $\lambda$ and $s$, even when other settings are the same. To deal with this point, for a
fixed data set, we search for the optimal parameter setting of \( \lambda \) and \( s \) for each model to achieve least uniform errors separately with the scheme shown in Example 1 and report their errors, condition numbers, and CPU time in Table 1. From Table 1, it is obvious to see that for both MD and GL cases, CRLS and DRLS can achieve similar uniform errors at their best situations. However, in most cases, CRLS is more well conditioned than DRLS and costs less CPU time.

It is an interesting question to explain that CRLS performs better than DRLS in the simulation. First, we can compare the theoretical results on the condition numbers of the linear systems induced by CRLS and DRLS and then get the reason that CRLS has smaller condition numbers and solving times. However, the theoretical underlying reason that CRLS achieves smaller errors is unknown to us until now. We can try to explain it in a heuristic way: we know that the difference between CRLS and DRLS is the choice of the regularization term. In CRLS, we use the \( \| \cdot \|_{\ell_p} \) norm of the polynomial approximation \( g \) while in DRLS, we use the \( \ell_2 \) norm of the value of \( g \) with some operator transform. Thus, it can be easy to obtain that the regularization term in DRLS is determined on the choice of the data point set \( X_N \) but in CRLS, it is not. This may possibly be the reason for CRLS obtaining smaller errors.

### 4. Conclusion and Future Work

In this paper, a continuous regularized least squares polynomial approximation model on the two-sphere is discussed. Condition number and error bound of the model are proposed in the situation that the data are noisy and their cardinality is not necessarily larger than the dimension of the polynomial space. To guarantee the error bound, we need two assumptions: the mesh norm of the data point set should be small enough and the weights should be close to their mean. Numerical results show that the novel model is effective to approximate smooth functions on the sphere.

We conclude this paper by pointing out a few problems for future directions. The first is that it is of great importance to study the strategy of choosing the parameters \( s \) and \( \lambda \), which is crucial to the approximation quality of the model. Secondly, it is an interesting topic to explain that CRLS could achieve smaller errors than DRLS in the numerical experiments, since the difference between them is only the choices of the regularization term.

### Appendix

**Two Relative Polynomial Approximation Models**

We will introduce two related regularized polynomial approximation models in this part. The first one can be seen as a special case of the CRLS model discussed in this paper.

**Example A.1** (Euclidean norm regularization). Let \( s = 0 \) and \( \varphi^{(s)}_\ell = 1 \). Let

\[
g = \sum_{\ell=0}^{L} \sum_{k=1}^{2\ell+1} c_{\ell,k} Y_{\ell,k},
\]

and \( c = (c_{0,1}, c_{1,1}, c_{1,2}, c_{1,3}, \ldots, c_{L,1}, \ldots, c_{L,2L+1})^T \). Then by (7), problem (1) can be rewritten as

\[
\min_{g \in \mathcal{P}_l} \sum_{i=1}^{N} w_i |g(x_i) - f^s(x_i)|^2 + \lambda \|c\|_2^2,
\]

in which the regularized term is the Euclidean norm of \( c \).

The second example is that we replace the regularized term by the \( \| \cdot \|_{L_1} \) norm of the Laplace–Beltrami transform of \( g \), which is slightly different with another special case of the CRLS model.

**Example A.2** (Laplace–Beltrami operator regularization). Let \( \varphi^{(s)}_\ell = |\ell(\ell+1)|^{-s} \) for \( \ell > 0 \). By omitting the term \( \lambda (1/|\ell(\ell+1)|)^{s} \), problem (1) can be rewritten as

\[
\min_{g \in \mathcal{P}_l} \sum_{i=1}^{N} w_i \left[ g(x_i) - f^s(x_i) \right]^2 + \lambda \|(-\Delta^*)^s g\|_{L_1},
\]

in which \( \Delta^* \) denotes the Laplace–Beltrami operator ([20], pp. 38–39), which is the angular part of the Laplace operator in three dimensions.
\[ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (A.4) \]

This can be directly derived by the characterization of Laplace–Beltrami operator that the spherical harmonics are its eigenfunctions, that is,

\[ \Delta^* Y_{\ell k}(x) = -\ell(\ell + 1)Y_{\ell k}(x). \quad (A.5) \]

The proof will be completed if \( T \) is nonsingular. Let \( \delta_1(T) \geq \delta_2(T) \geq \ldots \geq \delta_{d_2}(T) \) be eigenvalues of \( T \). Note that by (13), matrix \( T \) is semipositive definite, and thus we only need to prove that \( \delta_{d_2}(T) > 0 \). Note that

\[ \delta_{d_2}(T) \geq \delta_{d_2}(Y^TWY) + \lambda \delta_{d_2}(B) = \delta_{d_2}(Y^TWY) + \lambda \beta_0 \geq 0. \quad (A.7) \]

By this condition, we can prove the conclusion by showing that

\[ \text{Null}(Y^TWY) \cap \text{Null}(B) \cap \{0\} = \emptyset, \quad (A.8) \]

since under (A.8), for any \( \xi \neq 0 \), the two equalities

\[ \xi^T Y^TWY \xi = 0 \quad \text{and} \quad \xi^T B \xi = 0 \]

cannot hold at the same time, and thus \( T \) is nonsingular.

In the following part, we will prove (A.8) by contradiction. Assume there exists \( \xi = (\xi_1, \ldots, \xi_{d_2})^T \in \text{Null}(Y^TWY) \cap \text{Null}(B) \cap \{0\} \), i.e., \( \xi \neq 0 \), which satisfies

\[ B \xi = (0, \beta_1 \xi_2, \ldots, \beta_{d_2} \xi_{d_2})^T = 0, \quad (A.9) \]

\[ Y^TWY \xi = 0. \quad (A.10) \]

Equality (A.9) means that \( \xi_1 \neq 0 \) and \( \xi_2 = \cdots = \xi_{d_2} = 0 \). By this fact, we can rewrite (A.10) as

\[ Y^TWY \xi = \frac{\xi_1}{\sqrt{A_1}} \left( \sum_{i=1}^N w_i Y_{\ell_1}(x_i), \ldots, \sum_{i=1}^N w_i Y_{\ell_k}(x_i), \ldots, \sum_{i=1}^N w_i Y_{L_{L+1}}(x_i) \right)^T. \quad (A.11) \]

Note that the first component does not equal to 0, which contradicts equality (A.10). The proof is completed.

Data Availability

The Matlab codes for the numerical experiments in Section 3 of this paper are available from Dr. Yang Zhou (zhyg1212@163.com) upon request, who is the first author of the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

In fact, when \( \beta_0 = 0 \) and \( \beta_{\ell} > 0 \) for \( \ell = 1, \ldots, L \), such as in problem (A.3), the solution will also be unique, as is shown in the following lemma.

**Proposition A.1.** The solution of problem (A.3) is unique when \( \lambda > 0 \).

**Proof.** Problem (A.3) could be reformulated as form (10) with

\[ B = \text{diag} \begin{pmatrix} 0, 2^\ell, 2^\ell, \ldots, [\ell(\ell + 1)]^\ell \ldots, [\ell(\ell + 1)]^\ell \ldots, [L(L + 1)]^\ell \ldots, [L(L + 1)]^\ell \end{pmatrix} \quad (A.6) \]

Acknowledgments

This research was supported by the National Natural Science Foundation of China (no. 11626147) and Shandong Provincial Natural Science Foundation, China (no. ZR2019PA004).

References

[1] C. An, X. Chen, I. H. Sloan, and R. S. Womersley, “Regularized least squares approximations on the sphere using spherical designs,” *SIAM Journal on Numerical Analysis*, vol. 50, no. 3, pp. 1513–1534, 2012.

[2] V. Cammarota and D. Marinucci, “The stochastic properties of \( \ell \)-regularized spherical Gaussian fields,” *Applied and Computational Harmonic Analysis*, vol. 38, no. 2, pp. 262–283, 2015.

[3] S. Ji, D. Xu, L. Guo, M. Li, and D. Zhang, “The seeding algorithm for spherical \( k \)-means clustering with penalties,” *Journal of Combinatorial Optimization Accepted*, pp. 1–8, 2020.

[4] Q. T. Le Gia, I. H. Sloan, R. S. Womersley, and Y. G. Wang, “Isotropic sparse regularization for spherical harmonic representations of random fields on the sphere,” *Applied and Computational Harmonic Analysis*, vol. 49, no. 1, pp. 257–278, 2020.

[5] S. V. Pereverzyev, I. H. Sloan, and P. Tkachenko, “Parameter choice strategies for least-squares approximation of noisy smooth functions on the sphere,” *SIAM Journal on Numerical Analysis*, vol. 53, no. 2, pp. 820–835, 2015.

[6] Y. Zhou and X. Chen, “Spherical \( t \)-designs for approximations on the sphere,” *Mathematics of Computation*, vol. 87, no. 314, pp. 2831–2855, 2018.

[7] N. K. Pavlis, S. A. Holmes, S. C. Kenyon, and J. K. Factor, “An earth gravitational model to degree 2160: egm2008,” *EGU General Assembly*, vol. 10, pp. 13–18, 2008.

[8] V. I. Lebedev, “Quadratures on a sphere,” *USSR Computational Mathematics and Mathematical Physics*, vol. 16, no. 2, pp. 10–24, 1976.

[9] K. Hesse, I. H. Sloan, and R. S. Womersley, “Radial basis function approximation of noisy scattered data on the sphere,” *Numerische Mathematik*, vol. 137, no. 3, pp. 579–605, 2017.

[10] X. Chen, S. Du, and Y. Zhou, “A smoothing trust region filter algorithm for nonsmooth least squares problems,” *Science China Mathematics*, vol. 59, no. 5, pp. 999–1014, 2016.
[11] Y. Zhou, "Quadrature rules with neighborhood of spherical designs on the two-sphere," *Applied Mathematics and Computation*, vol. 367, pp. 1–8, 2020.

[12] Q. Liu and C. Wang, "An augmented Lagrangian algorithm for solving semi-infinite programming," *Journal of Applied Mathematics*, vol. 2012, Article ID 145083, 11 pages, 2012.

[13] L. Han, D. Xu, M. Li, and D. Zhang, "Approximation algorithms for the robust facility leasing problem," *Optimization Letters*, vol. 12, no. 3, pp. 625–637, 2018.

[14] J. Zhang, M. Li, Y. Wang, C. Wu, and D. Xu, "Approximation algorithm for squared metric two-stage stochastic facility location problem," *Journal of Combinatorial Optimization*, vol. 38, no. 2, pp. 618–634, 2019.

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 2012.

[16] A. Bjorck, *Numerical Methods for Least Squares Problems*, Siam, Philadelphia, PA, USA, 1996.

[17] I. H. Sloan and R. S. Womersley, "Extremal systems of points and numerical integration on the sphere," *Advances in Computational Mathematics*, vol. 21, no. 1-2, pp. 107–125, 2004.

[18] R. J. Renka, "Multivariate interpolation of large sets of scattered data," *ACM Transactions on Mathematical Software (TOAMS)*, vol. 14, no. 2, pp. 139–148, 1988.

[19] V. I. Krylov and A. H. Stroud, *Approximate Calculation of Integrals*, Courier Corporation, Chelmsford, MA, USA, 2006.

[20] C. Müller, *Spherical Harmonics*, Springer, Berlin, Germany, 2006.