INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 12, Number 2, Pages 328–342

DISCRETE LEAST SQUARES HYBRID APPROXIMATION WITH REGULARIZATION ON THE TWO-SPHERE

YANG ZHOU

Abstract. In this paper we consider the discrete constrained least squares problem coming from numerical approximation by hybrid scheme on the sphere, which applies both radial basis functions and spherical polynomials. We propose a novel $l_2 - l_1$ regularized least square model for this problem and show that it is a generalized model of the classical "saddle point" model. We apply the alternating direction algorithm to solve the $l_2 - l_1$ model and propose a convenient stopping criterion for the algorithm. Numerical results show that our model is more efficient and accurate than other models.

Key words. Regularized least squares, hybrid approximation, alternating direction method.

1. Introduction

Numerical approximation on the sphere is nowadays a widely studied problem arising in plenty of science landscapes such as geophysics, astrophysics, and surface reconstruction. Amongst varieties of different approaches, the hybrid approximation scheme [7, 11, 16, 19] seems an attractive method, which employs both the radial basis functions (RBF) and spherical harmonic polynomials. Often the underlying motivation has been the need to approximate geophysical quantities. It is well understood that the radial basis functions could approximate rapidly varying data over short distance effectively, whereas the spherical harmonic polynomials are more suitable for slowly varying data on a global scale.

In this paper we will discuss approximating a continuous function $f \in C(\mathbb{S}^2)$ using both radial basis functions and spherical harmonic polynomials, where $\mathbb{S}^2$ represents the unit sphere in three dimensional space as

$$\mathbb{S}^2 = \{\mathbf{x} = (x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\},$$

and $C(\mathbb{S}^2)$ denotes the space of all continuous functions defined on $\mathbb{S}^2$. We assume that the values of f are given at a distinct data point set

$$X_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N,\ N \in \mathbb{N}\} \subset \mathbb{S}^2.$$

To construct the radial basis functions, we choose all points in X_N as the center points and employs a (strictly) positive definite kernel ϕ [15, 19, 20] which satisfies

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \phi(\mathbf{x}_i, \mathbf{x}_j)\alpha_j \geq 0, \tag{1}$$

for any point set of $X_N \subset \mathbb{S}^2$ and for all $N \in \mathbb{N}$, with equality for distinct points $\mathbf{x}_j$ only if $\alpha_1 = \alpha_2 = \ldots = \alpha_N = 0$. Then the RBFs are defined as $\phi(\cdot, \mathbf{x}_j)$ with $j = 1, \ldots, N$. Additionally, we assume that kernel ϕ is zonal, which means

$$\phi(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i \cdot \mathbf{x}_j),$$

Received by the editors January 28, 2014, and in revised form, August 19, 2014.
2000 *Mathematics Subject Classification.* 65K05, 90C30.
This research was supported by the National Science Foundation of China (11301222).

for arbitrary $i, j = 1, \ldots, N$, where $\mathbf{x}_i \cdot \mathbf{x}_j$ denotes the Euclidean inner product in $\mathbb{R}^3$. Then we can define a space of RBFs as

$$\mathcal{X}_{X_N,\phi} = \mathcal{X}_N = \text{span}\{\phi(\cdot, \mathbf{x}_j) : \ \mathbf{x}_j \in X_N\}.$$

Further more, denote by

$$\mathcal{F}_\phi = \text{span}\{\phi(\cdot, \mathbf{x}_j) : \ \mathbf{x}_j \in \mathbb{S}^2, j \in 1, \ldots, N, \ N \in \mathbb{N}\},$$

which is a reproducing kernel pre-Hilbert space [19] under the inner product

$$\left\langle \sum_{i=1}^{N} \alpha_i \phi(\cdot, \mathbf{x}_i), \sum_{j=1}^{N} \alpha'_j \phi(\cdot, \mathbf{x}_j) \right\rangle_\phi = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha'_j \phi(\mathbf{x}_i, \mathbf{x}_j), \tag{2}$$

and the norm

$$\left\| \sum_{i=1}^{N} \alpha_i \phi(\cdot, \mathbf{x}_i) \right\|_\phi = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \phi(\mathbf{x}_i, \mathbf{x}_j), \tag{3}$$

with $\alpha_j \in \mathbb{R}$, $j = 1, \cdots, N$. Let $\mathcal{N}_\phi$ be the completion of $\mathcal{F}_\phi$ and then we can obtain that $\mathcal{N}_\phi$ is a reproducing kernel Hilbert space (RKHS). It is well known and could be easily verified that the reproducing kernel [11] ϕ of $\mathcal{N}_\phi$ satisfies

$$\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{y}, \mathbf{x}), \quad \forall \ \mathbf{x}, \ \mathbf{y} \in \mathbb{S}^2,$$

$$\phi(\cdot, \mathbf{y}) \in \mathcal{N}_\phi, \qquad \mathbf{y} \in \mathbb{S}^2,$$

and

$$\langle f, \phi(\cdot, \mathbf{y}) \rangle_\phi = f(\mathbf{y}), \quad \mathbf{y} \in \mathbb{S}^2, \quad f \in \mathcal{N}_\phi.$$

Another space we will apply is defined as

$$\begin{aligned} \mathbb{P}_L &= \{\textit{spherical polynomials of degree } \leq L\} \\ &= \text{span}\{Y_{\ell,k} : \ell = 0, \ldots, L, k = 1, \ldots, 2\ell + 1\}, \end{aligned}$$

with its dimension denoted by

$$d_L = \sum_{\ell=0}^{L} (2\ell + 1) = (L + 1)^2.$$

Here $Y_{\ell,k}$ is a fixed $\mathbb{L}_2$-orthonormal real spherical harmonic polynomial [1] of degree ℓ and order k defined on $\mathbb{S}^2$, which we can express by the denotation of $\mathbb{L}_2(\mathbb{S}^2)$ inner product on $\mathbb{S}^2$ as

$$\langle Y_{\ell,k}, Y_{\ell',k'} \rangle_{\mathbb{L}_2} = \int_{\mathbb{S}^2} Y_{\ell,k} Y_{\ell',k'} d\omega(\mathbf{x}) = \delta_{\ell,\ell'} \delta_{k,k'},$$

with

$$\ell, \ell' = 0, \ldots, L, \ k = 1, \ldots, 2\ell + 1, k' = 1, \ldots, 2\ell' + 1,$$

where $d\omega(\mathbf{x})$ denotes the normalized surface measure, and $\delta_{\ell,\ell'}$ is the Kronecker delta. According to the addition theorem, a zonal radial basis function has a expansion of the form

$$\phi(\cdot, \mathbf{x}) = \sum_{\ell=0}^{\infty} \hat{\phi}_\ell P_\ell(\cdot, \mathbf{x}) = \sum_{\ell=0}^{\infty} \frac{\hat{\phi}_\ell}{2\ell + 1} \sum_{k=1}^{2\ell+1} Y_{\ell,k}(\cdot) Y_{\ell,k}(\mathbf{x}), \tag{4}$$

where $\hat{\phi}_\ell > 0, \ \ell = 0, \ldots, \infty$ when ϕ is a strictly positive kernel. Here P_ℓ is the Legendre polynomial of degree ℓ in 3-dimension normalized to $P_\ell(1) = 1$.

Now let $X_N \subset \mathbb{S}^2$ be fixed, then the approximation of $f \in C(\mathbb{S}^2)$ with both RBFs and spherical harmonics can be defined as

$$\Lambda_{X,L} f = \sum_{j=1}^{N} \alpha_j \phi(\cdot, \mathbf{x}_j) + \sum_{\ell=0}^{L} \sum_{k=1}^{2\ell+1} \beta_{\ell,k} Y_{\ell,k}, \tag{5}$$

with the orthogonal condition

$$\sum_{j=1}^{N} \alpha_j v(\mathbf{x}_j) = 0, \quad \forall v \in \mathbb{P}_L, \tag{6}$$

where $\alpha_j,\ j = 1, \ldots, N,\ \beta_{\ell,k},\ \ell = 0, \ldots, L,\ k = 1, \ldots, 2\ell + 1$, are the coefficients of the RBFs and spherical harmonics, and L denotes the maximal degree of the spherical harmonics applied in the approximation. The values of f are given at the N-point set $X_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbb{S}^2$, and we insist that X_N is a fundamental system of degree L, see [2, 4, 16]. Thus the approximation $\Lambda_{X,L} f$ is to find an element $u \in \mathcal{X}_N$ and $v \in \mathbb{P}_L$ as

$$u = \sum_{j=1}^{N} \alpha_j \phi(\cdot, \mathbf{x}_j),$$

$$v = \sum_{\ell=0}^{L} \sum_{k=1}^{2l+1} \beta_{\ell,k} Y_{\ell,k},$$

satisfying condition (5) and (6). Usually, we can directly force $\Lambda_{X,L} f(\mathbf{x}_i) = f(\mathbf{x}_i), i = 1, \ldots, N$. Then let $\mathbf{f} := \mathbf{f}(X_N)$ be the column vector with

$$\mathbf{f} := [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N)]^T \in \mathbb{R}^N,$$

$A := A(X_N) \in \mathbb{R}^{N \times N}$ and $Q := Q(X_N) \in \mathbb{R}^{N \times (L+1)^2}$ with their entries as

$$A_{i,j} := \phi(\mathbf{x}_i, \mathbf{x}_j), \qquad i, j = 1, \ldots, N, \tag{7}$$

and

$$Q_{i,\ell k} := Y_{\ell,k}(\mathbf{x}_i), \quad i = 1, \ldots, N,\ \ell = 0, \ldots, L,\ k = 1, \ldots, 2\ell + 1, \tag{8}$$

and substitute (7), (8) to (5), (6), we could obtain a saddle point linear system as

$$\begin{bmatrix} A & Q \\ Q^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}. \tag{9}$$

Since the kernel we choose for constructing the RBFs is strictly positive definite, the matrix A here is positive definite. Moreover, the fact that X_N is a fundamental system implies that Q is of full column rank. Under these two conditions, equation (9) is well-posed and has a unique solution.

Condition (6) is added to guarantee the polynomial accuracy and the efficiency of the approximation (5). Under the assumption that Q is of full column rank, the approximation form (5) with condition (6) has an algebraic accuracy with degree L. That means, when the function f is a spherical polynomial with its order no greater than L, the condition will guarantee that $\Lambda_{X,L} f \equiv f$ holds no matter how the point set X_N is chosen.

To keep the efficiency of the approximation with both RBFs and spherical harmonics, the obtained linear combination of RBFs u is expected to be ϕ–orthogonal or $\mathbb{L}_2$–orthogonal to the spherical harmonic space $\mathbb{P}_L$. In this sense we can have that $\langle u, v \rangle_\phi = 0$ or $\langle u, v \rangle_{L_2} = 0$ for arbitrary element $v \in \mathbb{P}_L$. Since $Y_{\ell,k},\ \ell =$

$1,\ldots,L,\ k=1,\ldots,2\ell+1$ is a basis of $\mathbb{P}_L$, the orthogonal condition could be presented as

$$\langle u, Y_{\ell,k}\rangle_\phi = 0 \quad \text{for } \ell = 0,\ldots,L,\ k=1,\ldots,2\ell+1, \tag{10}$$

or

$$\langle u, Y_{\ell,k}\rangle_{\mathbb{L}_2} = 0 \quad \text{for } \ell = 0,\ldots,L,\ k=1,\ldots,2\ell+1. \tag{11}$$

From the definition of the two different inner products and $\hat{\phi}_\ell > 0,\ \ell = 0,\ldots,\infty$ in (4) we can obtain that

$$\begin{aligned}
\langle u, Y_{\ell,k}\rangle_\phi &= \left\langle \sum_{j=1}^{N} \alpha_j \phi(\cdot,\mathbf{x}_j), Y_{\ell,k} \right\rangle_\phi \\
&= \sum_{j=1}^{N} \alpha_j \langle \phi(\cdot,\mathbf{x}_j), Y_{\ell,k}\rangle_\phi \\
&= \sum_{j=1}^{N} \alpha_j Y_{\ell,k}(\mathbf{x}_j) = 0, \quad \text{for } \ell = 1,...,L,\ k=1,...2\ell+1,
\end{aligned}$$

and

$$\begin{aligned}
\langle u, Y_{\ell,k}\rangle_{\mathbb{L}_2} &= \langle \sum_{j=1}^{N} \alpha_j \phi(\cdot,\mathbf{x}_j), Y_{\ell,k}\rangle_{\mathbb{L}_2} \\
&= \sum_{j=1}^{N} \alpha_j \int_{\mathbb{S}^2} \phi(\mathbf{x},\mathbf{x}_j) Y_{\ell,k}(\mathbf{x}) \mathrm{d}\omega(\mathbf{x}) \\
&= \sum_{j=1}^{N} \alpha_j \int_{\mathbb{S}^2} \sum_{\ell'=0}^{\infty} \frac{\hat{\phi}_\ell}{2\ell+1} \sum_{k'=1}^{2\ell+1} Y_{\ell',k'}(\mathbf{x}) Y_{\ell',k'} Y_{\ell,k}(\mathbf{x}) \mathrm{d}\omega(\mathbf{x}) \\
&= \sum_{j=1}^{N} \alpha_j \frac{\hat{\phi}_\ell}{2\ell+1} Y_{\ell,k}(\mathbf{x}_j) = 0, \quad \text{for } \ell = 1,...,L,\ k=1,...2\ell+1,
\end{aligned}$$

which are both equivalent to the equation

$$Q^T\alpha = 0. \tag{12}$$

Le Gia, Watson, Sloan in [11] declared that problem (9) can also be explained as a continuous constrained least squares problem as

$$\begin{aligned}
&\min_{u\in\mathcal{X}_N} \frac{1}{2}\|u-f\|_\phi^2 \\
&\text{s.t.} \quad \langle u,\ Y_{\ell,k}\rangle_\phi = 0,\ \ell = 0,\ldots,L,\ k=1,\ldots,2\ell+1,
\end{aligned} \tag{13}$$

where the $\|\cdot\|_\phi$ is defined as in (3). With notations (7)(8) the problem can be reformulated as

$$\begin{aligned}
&\min_{\alpha} \alpha^T A\alpha - \alpha^T \mathbf{f} \\
&\text{s.t.} \quad Q^T\alpha = 0.
\end{aligned} \tag{14}$$

By introducing the vector β as the Lagrangian multiplier and deriving the KKT condition of this problem we can also obtain system (9).

The paper is organized as follows. In Section 2 we develop the "saddle point" model (9) into a $l_2 - l_1$ regularized least squares form as (24) to improve the approximation. Then we apply the alternating direction method (ADM) to solve

the problem presented in this section. We obtain that in each iteration of ADM, the problem can be decomposed to solving a linear system and a separable convex programming. Especially, a simply-formed and effective stopping criterion is given for this algorithm in this section. In Section 3 several numerical tests are established to show the efficient performance of the model conclusion is given in Section 4.

2. Problem formulation

In this section we pay our attention to study the least square approximation problem which also employs both the RBFs and the spherical harmonics. It is well known that (9) is an effective model for approximation problem in many cases. Now we consider a generalized least squares formula for hybrid approximation. Firstly, instead of using the data point set X_N as the center point set of RBFs, we use a different point set $X_{N_*} = \{\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_{N_*}^*\}$ as the center point set to construct the RBFs. As we declared in the last section, in the original saddle point model (9) the data point set is directly chosen as the center point set when we creating RBFs. As far as we know, no theoretical result has been established to insist this choice as far as we know. Naturally, we have some new notations about this new center point set as matrix $A \in \mathbb{R}^{N\times N_\star}$ and $Q_* \in \mathbb{R}^{N\times(L+1)^2}$ defined by

$$A_{i,j} := \phi(\mathbf{x}_i, \mathbf{x}_j^*), \qquad i = 1, \ldots, N,\ j = 1, \ldots, N_*, \tag{15}$$

and

$$(Q_*)_{i,\ell^2+k} := Y_{\ell,k}(\mathbf{x}_i^*), \quad i = 1, \ldots, N_*,\ k = 1, \ldots, 2\ell+1,\ \ell = 0, \ldots, L. \tag{16}$$

Then the orthogonal condition (6) is equivalent to

$$Q_*^T\alpha = 0. \tag{17}$$

In this case, we still always assume that X_{N_*} to be a fundamental system, which implies that Q_* is of full column rank. Now we consider the new linear system

$$\begin{bmatrix} A & Q \\ Q_* & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}. \tag{18}$$

If we here assume that $N > N_*$ and A is of full column rank, this system will be over determined and has no solution. In this case, instead of equation (18), we consider its least squares form as

$$\min_{\alpha,\beta} \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_2^2 \quad \text{s.t.} \quad Q_*^T\alpha = 0, \tag{19}$$

or

$$\min_{\alpha,\beta} \frac{1}{2}\alpha Q_* Q_*^T\alpha \quad \text{s.t.} \quad A\alpha + Q\beta - \mathbf{f} = 0. \tag{20}$$

The solution of problem (19) satisfies the condition (6) strictly whereas the solution of problem (20) interpolates the data $\mathbf{f}$ exactly. To meet a balance of this two aspects, we consider its penalized form as a more general case. we plus the two objective functions in the two models with a penalized parameter multiplied on one of them, and then can obtain its $l_2 - l_2$ penalized form as

$$\min_{\alpha,\beta} \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_2^2 + \lambda\|Q_*^T\alpha\|_2^2, \tag{21}$$

where λ is the regularized or penalized parameter to balance the approximation and the orthogonal condition. This problem is a smooth convex unconstrained

programming. By deriving its first order necessary condition, we can obtain a linear system which is easy to solve.

Moreover, we usually want the orthogonal condition to hold for as many ℓ and k, $\ell = 0, \ldots, L,\ k = 1, \ldots, 2\ell + 1$, as possible. That is to say, we want the vector $Q_*^T\alpha$ as sparse as possible. Based on this point of view, the $l_2 - l_0$ is more advisable to consider:

$$\min_{\alpha,\beta} \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_2^2 + \lambda\|Q_*^T\alpha\|_0. \tag{22}$$

This problem is non-lipschitz nonconvex and can have multiple local minimizers. However, we should note that problem (22) is non-convex and NP hard [5, 9, 12]. Hence we should consider an approximation of this model instead. A good approximation is that replacing the l_0 norm by l_p norm, with the form as

$$\min_{\alpha,\beta} \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_2^2 + \lambda\|Q_*^T\alpha\|_p, \tag{23}$$

where $0 < p < 1$, which is named low order penalty problem. However, the problem is nonsmooth and nonconvex. For the convenience of computing, a convex approximation form is expected to make this problem easier to solve. As the closest convex form of model (22), we take the l_1 regularization replacing the l_0 one. Thus problem (19) becomes non-smooth but convex programming as

$$\min_{\alpha,\beta} \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_2^2 + \lambda\|Q_*^T\alpha\|_1. \tag{24}$$

Remark 1. *The solution of the saddle point system (9), if exists, is also an optimal solution of optimization problem (19), (21), (22),(23) and (24) when choosing* $X_{N_*} = \{\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_{N_*}^*\} = X_N$.

The existence of solution for system (9) guarantees that the optimal values of all objective functions in (24), (19), (23) and (22) equals to 0.

Remark 2. *Systems (19), (23), (22) and (24) all guarantee the exactness for polynomials of degree* $\leq L$. *That is, for* $\forall f \in \mathbb{P}_L$, (α^*, β^*) *is the optimal solution for all the four problems, in which* $\alpha^* = 0$ *and* β^* *satisfies*

$$f = \sum_{l=0}^{L} \sum_{k=1}^{2\ell+1} \beta_{\ell,k}^* Y_{\ell,k},$$

where $\beta^* = (\beta_{\ell,k}^*),\ \ell = 0, \ldots, L,\ k = 1, \ldots, 2\ell + 1$. *In this situation,* β^* *is unique and all the objective functions equal to 0.*

The $\|\cdot\|_1$ regularizer in (24) is to guarantee condition (12) which forces u, the linear combination of RBFs, separated from the spherical harmonic polynomial space $\mathbb{P}_L$ and λ is the regularization parameter to balance the two parts. This leads to (24) as a non-smooth and convex optimization problem. The problem also requires that both X_N and X_{N_*} are fundamental systems [2] which guarantee that both Q and Q_* are of full column rank.

Now we consider using the alternating direction method (ADM) to solve the problem (24). The basic idea of ADM goes back to the work of Gabay and Mercier [8]. The motivation of this method is to solve a separable programming by separating it into two or more easier subproblems. Since the method requires that the objective function is separable, first we introduce an auxiliary variable vector

$\mathbf{y} = Q_*^T\alpha \in \mathbb{R}^{(L+1)^2}$ and (30) can be reformulated into a constrained optimization problem as

$$\min_{\alpha,\beta} \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_2^2 + \lambda\|\mathbf{y}\|_1 \tag{25}$$
$$\text{s.t.} \quad Q_*^T\alpha - \mathbf{y} = 0,$$

which is a structured convex constrained optimization problem and can be solved by the ADM method. Let $\mathbf{z}$ be the Lagrangian multiplier and $0 < \rho \leq 1$ be the augmented Lagrangian parameter for the linear constraint $Q_*^T\mathbf{x} - \mathbf{y} = 0$, then the augmented lagrangian function of (25) is

$$L(\mathbf{x},\mathbf{y},\mathbf{z}) = \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_2^2 + \lambda\|\mathbf{y}\|_1 - \mathbf{z}^T(Q_*^T\alpha - \mathbf{y}) + \frac{\rho}{2}\|Q_*^T\alpha - \mathbf{y}\|_2^2. \tag{26}$$

Then a framework of the alternating direction method for problem (24) could be given as Algorithm 1.

Algorithm 1: ADM for solving $l_2 - l_1$ regularized model (24)

Step 0: Initialization. Make an initial guess $\mathbf{v}^0 = (\mathbf{y}^0, \mathbf{z}^0)$.

Step 1: Find a new x. For given $(\mathbf{y}^k, \mathbf{z}^k)$, solve the convex quadratic programming

$$\mathbf{x}^{k+1} = \begin{pmatrix} \alpha^{k+1} \\ \beta^{k+1} \end{pmatrix} = \arg\min \left\{ \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_2^2 - (\mathbf{z}^k)^T(Q_*^T\alpha - \mathbf{y}^k) + \frac{\rho}{2}\|Q_*^T\alpha - \mathbf{y}^k\|_2^2 \right\}. \tag{27}$$

Step 2: Find a new y. Use $\mathbf{z}^k$ and the obtained $\mathbf{x}^{k+1}$ to solve the convex separable quadratic programming

$$\mathbf{y}^{k+1} = \arg\min \left\{ \lambda\|\mathbf{y}\|_1 - (\mathbf{z}^k)^T(Q_*^T\alpha^{k+1} - \mathbf{y}) + \frac{\rho}{2}\|Q_*^T\alpha^{k+1} - \mathbf{y}\|_2^2 \right\}. \tag{28}$$

Step 3: Update the lagrangian operator. Update $\mathbf{z}^k$ as $\mathbf{z}^{k+1} = \mathbf{z}^k - \rho(Q_*^T\alpha^{k+1} - \mathbf{y}^{k+1})$ and go back to step 1 until convergence.

Since our problem (24) is convex, the convergence of Algorithm 1 is covered by the analysis given in [6] where each ADM subproblem is required to be solved more and more accurately as the algorithm proceeds. In step 2, by deriving its first-order optimality condition, solving (27) could lead to seeking the solution of the following linear system:

$$\begin{bmatrix} A^TA + \rho Q_*Q_*^T & A^TQ \\ Q^TA & Q^TQ \end{bmatrix} \begin{bmatrix} \alpha^{k+1} \\ \beta^{k+1} \end{bmatrix} = \begin{bmatrix} A^T\mathbf{f} + \rho Q_*\mathbf{y}^k + Q_*\mathbf{z}^k \\ Q^T\mathbf{f} \end{bmatrix}. \tag{29}$$

Now we denote by

$$M = \begin{bmatrix} A & Q \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad B = \begin{bmatrix} Q_*^T & 0 \end{bmatrix}$$

and

$$J = \begin{bmatrix} M \\ B \end{bmatrix} = \begin{bmatrix} A & Q \\ Q_*^T & 0 \end{bmatrix}$$

where $M \in \mathbb{R}^{N\times(N_*+(L+1)^2)}$, $B \in \mathbb{R}^{(L+1)^2\times(N_*+(L+1)^2)}$ and $J \in \mathbb{R}^{(N+(L+1)^2)\times(N_*+(L+1)^2)}$.

Lemma 3. *Linear system (29) has a unique solution if*

$$\operatorname{rank}(J) = N_* + (L+1)^2.$$

Especially, when we assume that $X_{N_*} = X_N$ and they are fundamental systems with order L, we have that the solution of (24) is unique and also an optimal solution of problem (9), (19), (22), (23) and (21).

Proof. For (29) we could obtain that

$$\begin{aligned}\begin{bmatrix} A^TA+\rho Q_*Q_*^T & A^TQ \\ Q^TA & Q^TQ \end{bmatrix} &= \begin{bmatrix} A^T \\ Q^T \end{bmatrix}\begin{bmatrix} A & Q \end{bmatrix} + \rho \begin{bmatrix} Q_* \\ 0 \end{bmatrix}\begin{bmatrix} Q_*^T & 0 \end{bmatrix} \\ &= M^TM + \rho B^TB.\end{aligned}$$

The above matrix is invertible if and only if the zero vector is the unique solution of the system

$$J\mathbf{x} = 0.$$

Then we can get that

$$N_* + (L+1)^2 \le \operatorname{rank}(J) \le N_* + (L+1)^2.$$

Especially, when we assume that $X_{N_*} = X_N$ and they are fundamental systems, we can obtain that $\operatorname{rank}(J) = N + (L+1)^2 = N_* + (L+1)^2$. Therefore, the solution of problem (24) is unique. Moreover, in this case we can have $A\alpha + Q\beta - \mathbf{f} = 0$ and $Q^T\alpha = 0$ hold at the same time because the equation (9) is well posed. Thus, the unique solution of (24) is also solution of problem (9), (19), (22) and (23).

The proof of the lemma is completed. □

Now we consider the subproblem (28) in step 3. The format of problem (24) could be simplified as

$$\begin{aligned}\min_{\mathbf{x}} \frac{1}{2}\|M\mathbf{x} - \mathbf{f}\|_2^2 + \lambda\|B\mathbf{x}\|_1 \\ \text{s.t.} \quad B\mathbf{x} - \mathbf{y} = 0.\end{aligned} \tag{30}$$

We denote by $(B\mathbf{x}^{k+1})_i$ as the ith column of the vector $B\mathbf{x}^{k+1}$. Then problem (28) can be reformulated as

(31)

$$\begin{aligned}\mathbf{y}^{k+1} &= \arg\min \left\{ \lambda \sum_{i=1}^{(L+1)^2} |\mathbf{y}_i| - \sum_{i=1}^{(L+1)^2} (\mathbf{z}^k)_i((B\mathbf{x}^{k+1})_i - \mathbf{y}_i) \right. \\ &\quad \left. + \frac{\rho}{2} \sum_{i=1}^{(L+1)^2} ((B\mathbf{x}^{k+1})_i - \mathbf{y}_i)^2 \right\} \\ &= \arg\min \left\{ \sum_{i=1}^{(L+1)^2} (\lambda|\mathbf{y}_i| - (\mathbf{z}^k)_i((B\mathbf{x}^{k+1})_i - \mathbf{y}_i) + \frac{\rho}{2}((B\mathbf{x}^{k+1})_i - \mathbf{y}_i)^2) \right\}.\end{aligned}$$

We see that (31) is a separable optimization problem. Then the problem can be separated to $(L+1)^2$ one-dimension subproblems as

$$\mathbf{y}_i^{k+1} = \arg\min \left\{ \lambda|\mathbf{y}_i| + (\mathbf{z}^k)_i\mathbf{y}_i + \frac{\rho}{2}((B\mathbf{x}^{k+1})_i - \mathbf{y}_i)^2 \right\}, \tag{32}$$

and by the first order optimal condition, we have that

$$0 \in \lambda\partial(|\mathbf{y}_i|) - \mathbf{z}_i^k - \rho(B\mathbf{x}^{k+1})_i + \rho\mathbf{y}_i,$$

where $\partial(|\mathbf{y}_i|)$ denotes the subdifferential of the nondifferentiable convex function $|\mathbf{y}_i|$. We could note that this step is equivalent to a scalar shrinkage process and the solution of (31) is developed based on the relative conclusion in [17] as following.

Remark 4. *The solution of (32) could be given by the formulation*

$$\begin{aligned}\mathbf{y}_i^{k+1} =& \frac{1}{\rho}\left(\max\left\{0, \rho(B\mathbf{x}^{k+1})_i + (\mathbf{z}^k)_i - \lambda\right\}\right. \\ & \left. - \max\left\{0, -\rho(B\mathbf{x}^{k+1})_i - (\mathbf{z}^k)_i - \lambda\right\}\right).\end{aligned} \tag{33}$$

Now we consider the stopping criterion of Algorithm 1. By deriving the first-order optimality condition (25), we have

$$\begin{cases} M^T M\mathbf{x} - M^T\mathbf{f} - B^T\mathbf{z} = 0, \\ 0 \in \lambda\partial(\|\mathbf{y}\|_1) + \mathbf{z}, \\ B\mathbf{x} - \mathbf{y} = 0. \end{cases} \tag{34}$$

Hence problem (25) has the following variational inequality characterization: find $\omega \in \Omega := \mathbb{R}^{N_*+3(L+1)^2}$ such that

$$\omega \in \Omega, \quad \lambda(|\mathbf{y}'| - |\mathbf{y}|) + \langle \omega' - \omega, F(\omega)\rangle \geq 0, \quad \forall \omega' \in \Omega, \tag{35}$$

where

$$\omega = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} \quad \textit{and} \quad F(\omega) = \begin{pmatrix} M^T M\mathbf{x} - M^T\mathbf{f} - B^T\mathbf{z} \\ \mathbf{z} \\ B\mathbf{x} - \mathbf{y} \end{pmatrix}. \tag{36}$$

Let $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}) \in \Omega$ be generated by Algorithm 1. We denote by $\Omega^* = \{(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)\}$ the solution set of problem (25). Note that system (29) is equivalent to find an $\mathbf{x}^{k+1}$ satisfying

$$\begin{aligned}\left\langle \mathbf{x}' - \mathbf{x}^{k+1}, M^T M\mathbf{x}^{k+1} - M^T\mathbf{f} - B^T\mathbf{z}^k + \rho B^T B\mathbf{x}^{k+1} - \rho B^T\mathbf{y}^k \right\rangle \geq 0, \\ \forall \mathbf{x}' \in \mathbb{R}^{N_*+(L+1)^2}.\end{aligned}$$

Simultaneously, we have

$$\begin{aligned}&\lambda(|\mathbf{y}'| - |\mathbf{y}^{k+1}|) - \left\langle \begin{pmatrix} \mathbf{x}' - \mathbf{x}^{k+1} \\ \mathbf{y}' - \mathbf{y}^{k+1} \\ \mathbf{z}' - \mathbf{z}^{k+1} \end{pmatrix}, \right. \\ &\left. \begin{pmatrix} M^T M\mathbf{x}^{k+1} - M^T\mathbf{f} - B^T\mathbf{z}^{k+1} \\ \mathbf{z}^{k+1} \\ \rho(B\mathbf{x}^{k+1} - \mathbf{y}^{k+1}) \end{pmatrix} + \begin{pmatrix} \rho B^T(\mathbf{y}^{k+1} - \mathbf{y}^k) \\ 0 \\ \mathbf{z}^{k+1} - \mathbf{z}^k \end{pmatrix} \right\rangle \geq 0,\end{aligned} \tag{37}$$

for any $(\mathbf{x}', \mathbf{y}', \mathbf{z}') \in \Omega$. Therefore, $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1})$ is a solution of (25) if and only if $\mathbf{y}^k = \mathbf{y}^{k+1}$ and $\mathbf{z}^{k+1} = \mathbf{z}^k$. Then we could establish a stopping criterion for Algorithm 1 according to this conclusion:

$$\max\{\|\mathbf{y}^k - \mathbf{y}^{k+1}\|_2, \|\mathbf{z}^k - \mathbf{z}^{k+1}\|_2\} \leq \epsilon, \tag{38}$$

where $\epsilon > 0$.

3. Numerical results

In this section we establish some numerical experiments to test the efficiency of the approximation model proposed in the above section. The models and algorithms are implemented in Matlab 2011b equipped on a Lenovo Thinkcenter PC equipped with Intel Core i7-3770 3.4G Hz CPU, 8 GB RAM running Windows 7.

The choice of kernels for constructing positive definite RBFs is multiple. In this paper we choose the "Wendland" function as the kernel of RBFs, which is a kind of piece wise function with compact support and is proved to be positive definite on $\mathbb{S}^2$, see [19]. The function is defined as

$$\phi(\mathbf{x},\mathbf{y}) = \psi(|\mathbf{x}-\mathbf{y}|) = \psi(\sqrt{2-2\mathbf{x}\cdot\mathbf{y}}), \quad \mathbf{x},\mathbf{y}\in\mathbb{S}^2,$$

where $\psi(r)$ could be one of the three choices as

$$\begin{aligned}
\psi_0(r) &= (1-r)_+^2 \in C^0(\mathbb{R}),\\
\psi_2(r) &= (1-r)_+^4(4r+1) \in C^2(\mathbb{R}),\\
\psi_4(r) &= (1-r)_+^6(35r^2+18r+3) \in C^4(\mathbb{R}).
\end{aligned}$$

Note that $\psi_i, i = 0, 2, 4$ are all continuous but ψ_0 is not differentiable at $r = 0$. In our experiment, numerical tests has proved that approximation residuals do not vary so much by choosing different RBFs, so here we will only apply the kernel ψ_2 which is continuous differentiable on $\mathbb{R}$ to construct the RBFs. Sometime scaling of the compact support is employed to improve the approximation, but scaling also results in large condition number of A, which will lead to long solving time. That may influence the solving time for each different method differently. Hence for fairness we will not use scaling in this experiment if nothing announced.

In this paper we do not pay much attention to the choice of regularized parameter λ in (23), though it may influence the result of approximation. However, how to choose λ is still a significant and open problem which is worth studying in the future. As a part of remedy of that, we would try several different selections of λ, as $\lambda = 10^{-3}, 10^{-2}, \cdots, 10^3$, and record the best results among them.

We choose two different types of point sets as the data set and center point set for RBFs. One type is a set with its points (approximately) uniformly distributed on the whole sphere. The minimal energy [14], extremal spherical design [4] and well conditioned spherical t-design [2] systems are all uniformly distributed point systems. Here we will use equal area partitioning method (EAP) which is proposed by Saff [14] to generate the points.

The other type is a non-uniformly distributed set with points distributed densely in a small cap region and relatively sparsely in the rest region of the sphere. Obviously, this kind of point sets will lead to ill conditioned matrix and more solving time. We still can generate this type of point set X_N using EAP method. The basic scheme is that we first generate some points densely and uniformly in the small cap region using EAP and then generate points sparsely in the rest region. In each time of test, we apply the same type of the data set and center point set. That means if a non-uniformly distributed point set is selected as data point set in one test, then the center point set will also be selected as non-uniformly with the same cap region and vice versa. We should also note that the theory of keeping matrix M of full column rank is still not studied. But in practical tests it is not difficult to choose the data set and center set to make this condition hold.

3.1. Approximate Franke function. In the first experiment, to test the model (24) and Algorithm 1 mentioned in our paper, the Franke function with the form

$$\begin{aligned}
f(\mathbf{x}) = & f(x,y,z) = 0.75\exp(-(9x-2)^2/4 - (9y-2)^2/4 - (9z-2)^2/4)\\
& + 0.75\exp(-(9x+1)^2/49 - (9y+1)/10 - (9z+1)/10)\\
& + 0.5\exp(-(9x-7)^2/4 - (9y-3)^2/4 - (9z-5)^2/4)\\
& - 0.2\exp(-(9x-4)^2 - (9y-7)^2 - (9z-5)^2), \ (x,y,z)\in\mathbb{S}^2
\end{aligned} \tag{39}$$

is chosen as the target function to approximate, which is modified by Renka [13]. Note that the Franke function is continuously differentiable on $\mathbb{S}^2$. We choose the both two types of point set mentioned above: scattered data systems (SD), which represent the non-uniformly distributed point sets, and equal area partitioning systems (EAP), which represent the uniformly distributed point sets.

Especially, we should note that here the infinity norms of the residuals are approximate values. We all know that it is a complicated and time-consuming process to calculate the precise value of infinity norm of a function on $\mathbb{S}^2$. The scheme to approximate the values is that we choose a uniformly distributed test point set X_t with large cardinalities and then use the maximal residual evaluated at X_t to approximate the infinity norm. Define $\Lambda_{p,q}$ as the approximation obtained by problem

$$\min_{\alpha,\beta} \frac{1}{2}\|A\alpha + Q\beta - \mathbf{f}\|_p^p + \lambda\|Q_*^T\alpha\|_q^q, \tag{40}$$

where $p, q = 1, 2$. Then the residuals are obtained as

$$R_{p,q} = \|\Lambda_{p,q}f - f\|_\infty \approx \max_{\mathbf{x}\in X_t} |\Lambda_{p,q}f(\mathbf{x}) - f(\mathbf{x})|.$$

Simultaneously, we also present the residual norms by applying the saddle point model (9) to get approximate the target function defined as

$$R_{X,L} = \|\Lambda_{X,L}f - f\|_\infty \approx \max_{\mathbf{x}\in X_t} |\Lambda_{X,L}f(\mathbf{x}) - f(\mathbf{x})|,$$

in the table, of which both the center point set and data set are X_{N_*}. Here we use an equal area partitioning point set with 10^6 points as the test set X_t.

Besides the model (24) we investigate in this paper, the $l_2 - l_2$, $l_1 - l_1$ and the saddle point model (9) are also tested in this experiment. Since the $l_2 - l_2$ model can be reformulated as a linear system, we will solve it by the minimal residual method (MINRES), as what is applied to equation (9). For $l_1 - l_1$ model, the problem is non-smooth but convex. There are many popular methods and packages to solve a convex problem, such as the SPG [3], which is written in Fortran 77 and is proved to be efficient for many continuous differentiable optimization problems. In this experiment we would apply a package called CVX which is written by Matlab for comparison. The SDPT3 solver [18] is chosen in this package, which employs an infeasible primal-dual predictor-corrector path-following method and could deal with varieties of convex problem. To test the efficiency of Algorithm 1, we also solve model (24) using SDPT3 for comparison.

We collect the infinity norms of the approximation residuals and CPU times in seconds for solving process for each model as presented in Table 1. From the table it is natural to see that all residuals $R_{2,1}$, $R_{2,2}$, $R_{1,1}$ and $R_{X,L}$ generally decrease when N_* turns larger, which means that the approximation is more accurate. For the same scale it is obviously to see that $R_{2,1}$ is always the smallest among the residuals. Also we could see that ADM is an efficient solver for our model. For same scale and same type of point set, ADM needs the least time to solve its model. For similar scale but different types of point sets, $l_2 - l_1$ model using ADM costs similar time for both the SD and the EAP case, whereas other models need much more time to solve for SD case than the EAP case.

Moreover, for solving model (24) using different algorithms we can conclude from the second and third column of the table that ADM spends less time than SDPT3 in any case. And also, model (24) is solved faster than $l_1 - l_1$ model when same algorithm is applied, seen from the third and fourth column.

TABLE 1. Residuals ($R_{.,.}$) and CPU time ($T_{.,.}$) with different models for hybrid approximation with $L = 10$.

Type(N,N_*)	$R_{2,1}(T_{2,1})$	$R_{2,1}(T_{2,1})$	$R_{2,2}(T_{2,2})$	$R_{1,1}(T_{1,1})$	$R_{X,L}(T_{X,L})$
Methods	ADM	SDPT3	MINRES	SDPT3	MINRES
EAP(2000,400)	0.0433(0.02)	0.0433(4.50)	0.0471(0.05)	0.687(55)	0.0446(0.01)
EAP(4000,800)	0.0165(0.23)	0.0165(18.9)	0.0360(0.14)	0.0305(198)	0.0221(0.08)
EAP(6000,1200)	0.0068(0.25)	0.0068(110.7)	0.0372(0.39)	0.0076(490)	0.0095(0.25)
EAP(8000,1600)	0.0040(0.56)	0.0040(276.2)	0.0377(0.81)	0.0076(927)	0.0071(0.63)
EAP(10000,2000)	0.0018(0.69)	0.0018(738.6)	0.0368(0.87)	0.0031(2438)	0.0031(0.69)
SD(1993,399)	0.0535(0.03)	0.0535(0.03)	0.0563(0.06)	0.0845(46)	0.0514(0.04)
SD(3979,801)	0.0261(0.13)	0.0261(0.13)	0.0450(0.22)	0.0336(216)	0.0336(0.10)
SD(5974,1194)	0.0109(0.27)	0.0109(0.27)	0.0437(0.79)	0.0162(698)	0.0150(0.52)
SD(7964,1592)	0.0080(0.57)	0.0080(0.57)	0.0438(1.5)	0.0151(1132)	0.0126(1.39)
SD(9954,1993)	0.0040(0.95)	0.0040(0.95)	0.0438(2.5)	0.0076(1828)	0.0071(2.47)

3.2. Approximate Franke function with noise. In this subsection we will approximate the Franke function with some noise. We denote the noisy function as

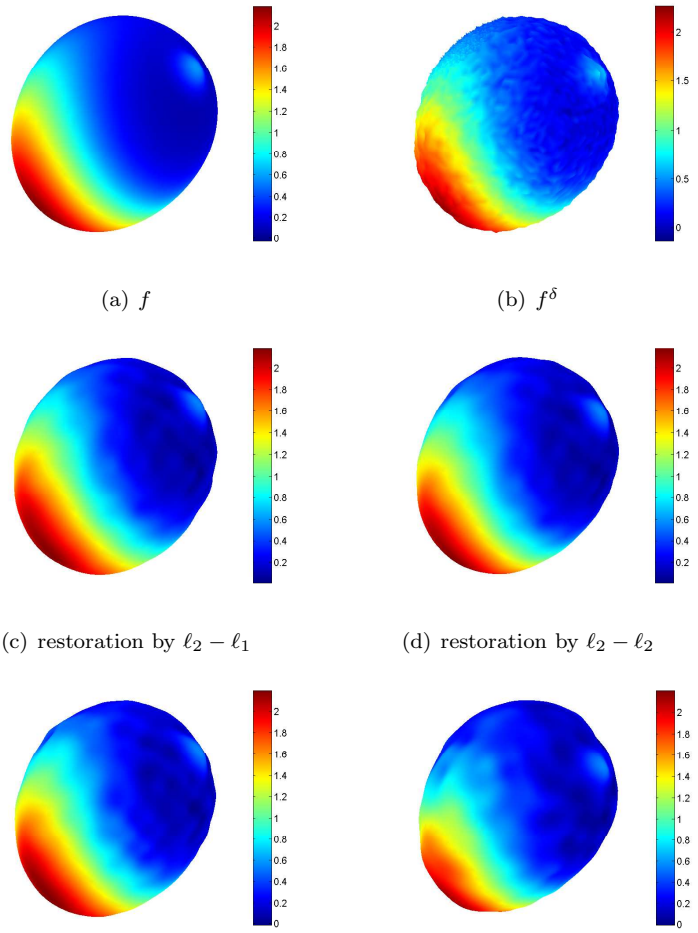
$$f^{\delta}(\mathbf{x}) = f(\mathbf{x}) + \delta(\mathbf{x}),$$

where for each $\mathbf{x}$, $\delta(\mathbf{x})$ is a sample of a normal random variable with mean 0 and standard deviation $\sigma = 0.05$. In this experiment we aim to restore the function f through approximating f^{δ} using the proposed model (23). The saddle point model (9) will also be compared with (23) for restoration of the original function. The approximate infinity norms of residuals are still recorded to measure the approximation quality. Since we have found that the equal area partition point systems behave well in the first experiment, in this experiment we only choose the EAP system for test. The infinity norms of the residual are still recorded to measure the quality of the approximation.

TABLE 2. Residuals of approximation for f^{δ}.

N,N_*,L	$R_{2,1}$	$R_{2,2}$	$R_{1,1}$	$R_{X,L}$
121,36,5	0.3312	0.3187	0.3915	0.3968
441,121,5	0.1129	0.1129	0.1708	0.1646
961,256,5	0.1035	0.1035	0.1134	0.1395
961,256,10	0.1039	0.1039	0.1152	0.1449
1681,441,5	0.0959	0.0959	0.1179	0.1585
1681,441,10	0.0861	0.0861	0.1255	0.1591

Numerical Results for this experiment are shown in Table 2. From the table we could find that for noisy case, the regularized models, including $l_2 - l_1$, $l_2 - l_2$ and $l_1 - l_1$, provide more accurate restoration for the original target function. In Figure



(a) f (b) f^δ

(c) restoration by $\ell_2 - \ell_1$ (d) restoration by $\ell_2 - \ell_2$

(e) restoration by $\ell_1 - \ell_1$ (f) restoration by saddle point system

FIGURE 1. Shapes of f, f^δ and its restoration.

1 we present the shapes of Franke function f, the noisy function f^δ and restorations using model appearing in the above Table 2. Especially, the $l_2 - l_1$, $l_2 - l_2$ models still own similar residuals and perform better than the $l_1 - l_1$ model. However, it is hard to say that for noisy case, which type of regularized model performs best. It may depend on the property of the noise, which probably will be a long standing problem.

3.3. Approximate Franke function plus cap function. In this subsection we seek to approximate the Franke function plus a cap function as

$$f_{cap}(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) = f(\mathbf{x}) + \begin{cases} \rho \cos\left(\frac{\pi \arccos(\mathbf{x}_c \cdot \mathbf{x})}{2r}\right), & \mathbf{x} \in C(\mathbf{x}_c, r), \\ 0, & \text{otherwise.} \end{cases}$$

In this experiment we choose $\mathbf{x}_c = (-0.5, 0.5, \sqrt{0.5})^T$,$r = 0.5$ and $\rho = \frac{1}{r}$. It is easy to see that f_{cap} is a continuous but non-differentiable function on the sphere. Thus, when we use continuous differentiable functions to approximate the f_{cap},

TABLE 3. Residuals of approximation for f_{cap} with $L = 10$.

N,N_*	$R_{2,1}$	$R_{2,2}$	$R_{1,1}$	$R_{X,L}$
1993,399	0.1007	0.1040	0.1456	0.1467
2987,598	0.0708	0.0836	0.1096	0.1232
3979,801	0.0673	0.0814	0.0793	0.1101
4980,994	0.0565	0.0745	0.0826	0.0925
5974,1194	0.0490	0.0749	0.0625	0.0745

errors near the cap boundary often are much larger than other parts. To get a close estimate of the uniform residual for this case, we choose the test set X_t to be a type with whose points are distributed densely around the cap boundary. Similarly, to obtain better approximation, we will choose points denser in the cap region than the rest region. Also we should note that in each region the points could be uniformly distributed, which may lead to good approximation. In this sense, we apply the SD point systems as the center point set and data point set. In this experiment we set the scaling parameter $\sigma = 0.5$, which means that in the RBFs kernel $\psi_i(r)$ the variable r is replaced by $\frac{r}{\sigma}$ to reduce the compact support area of each function.

Table 3 shows the numerical results of this experiment. From the table we can see that for all point sets we apply to the approximation, the $l_2 - l_1$ model (24) keeps the best approximation among the four models.

4. Conclusion

In this paper we generalize the original hybrid approximation "saddle point" model to a regularized $l_2 - l_1$ least squares problem. We apply the ADM method to solve this problem and give a convenient and effective stopping criterion for this problem. Numerical results show that the proposed model and algorithm are efficient for this kind of problem.

Acknowledgement

The author would like to give heartfelt thanks to Professor Xiaojun Chen whose suggestions were of inestimable value for the new model we proposed in this paper. The author also would thank Wei Bian for her comments and suggestions to improve the quality of the manuscript.

References

[1] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1970.
[2] C. An, X. Chen, I. H. Sloan and R. S. Womersley, Well conditioned spherical designs for integration and interpolation on the two-sphere, SIAM J. Numer. Anal., 48 (2010) 2135–2157.
[3] E. G. Birgin, J. M. Martínez and M. Raydan, Algorithm 813: SPG–software for convex-constrained optimization, ACM Trans. Math. Softw., 27 (2001) 340–349.
[4] X. Chen, R. S. Womersley and J. J. Ye, Minimizing the condition number of a Gram matrix, SIAM J. Optim., 21 (2011) 127–148.
[5] X. Chen, F. Xu and Y. Ye, Lower bound theory of nonzero entries in solutions of $\ell_2 - \ell_p$ minimization, SIAM J. Sci. Comput., 32 (2010) 2832–2852.
[6] J. Eckstein and D. Bertsekas, On the DouglasRachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Prog., 55 (1992) 293–318.

[7] W. Freeden, T. Gervens and M. Schreiner, Constructive Approximation on the Sphere (With Applications to Geomathematics), Oxford Science (Clarendon), Oxford, UK, 1998.
[8] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, Comput. Math. Appl., 2 (1976) 17–40.
[9] M. R. Garey and D. S. Johnson , Computers and Intractability – A Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.
[10] M. C. Grantand S. Boyd, The CVX Users' Guide, 2013.
[11] Q. T. Le Gia, I. H. Sloan and A. J. Wathen, Stability and preconditioning for a hybrid approximation on the sphere, Numer. Math., 118 (2011) 695–711.
[12] B. K. Natarajan, Sparse spproximate solutions to linear systems, SIAM J. Comput., 24 (1995) 227–234.
[13] R. J. Renka, Multivariate interpolation of large sets of scattered data, ACM Trans. Math. Softw., 14 (1998) 671–680.
[14] E. B. Saff and A. B. Kuijlaars, Distributing many points on a sphere, Math. Intell., 19 (1997) 5–11.
[15] I. J. Schoenberg, Positive definite functions on spheres, Duke Math. J., 9 (1942) 96–108.
[16] I. H. Sloan and A. Sommariva, Approximation on the sphere using radial basis functions plus polynomials, Adv. Comput. Math., 29 (2008) 147–177.
[17] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. Royal. Statist. Soc B., 58 (1996) 267–288.
[18] K. C. Toh, R. H. Tütüncü and M. J. Todd, On the Implementation of SDPT3 (version 3.1) – a Matlab Software Package for Semidefinite-Quadratic-Linear-Programming, IEEE Conferenc on Computer-Aided Control System Design, 2004.
[19] H. Wendland, Scattered Data Approximation, Cambridge University Press, 2005.
[20] Y. Xu and E. W. Cheney, Strictly positive definite functions on spheres, Proceedings of the AMS, 116 (1992) 977–981.

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong
E-mail: zhyg1212@163.com