

# SPHERICAL DESIGNS AND NON-CONVEX MINIMIZATION FOR RECOVERY OF SPARSE SIGNALS ON THE SPHERE\*

XIAOJUN CHEN<sup>†</sup> AND ROBERT S. WOMERSLEY<sup>‡</sup>

**Abstract.** This paper considers the use of spherical designs and non-convex minimization for recovery of sparse signals on the unit sphere  $\mathbb{S}^2$ . The available information consists of low order, potentially noisy, Fourier coefficients for  $\mathbb{S}^2$ . As Fourier coefficients are integrals of the product of a function and spherical harmonics, a good cubature rule is essential for the recovery. A spherical  $t$ -design is a set of points on  $\mathbb{S}^2$ , which are nodes of an equal weight cubature rule integrating exactly all spherical polynomials of degree  $\leq t$ . We will show that a spherical  $t$ -design provides a sharp error bound for the approximation signals. Moreover, the resulting coefficient matrix has orthonormal rows. In general the  $\ell_1$  minimization model for recovery of sparse signals on  $\mathbb{S}^2$  using spherical harmonics has infinitely many minimizers, which means that most existing sufficient conditions for sparse recovery do not hold. To induce the sparsity, we replace the  $\ell_1$ -norm by the  $\ell_q$ -norm ( $0 < q < 1$ ) in the basis pursuit denoise model. Recovery properties and optimality conditions are discussed. Moreover, we show that the penalty method with a starting point obtained from the re-weighted  $\ell_1$  method is promising to solve the  $\ell_q$  basis pursuit denoise model. Numerical performance on nodes using spherical  $t$ -designs and  $t_\epsilon$ -designs (extremal fundamental systems) are compared with tensor product nodes. We also compare the basis pursuit denoise problem with  $q = 1$  and  $0 < q < 1$ .

**Key words.** sparse recovery, quasi-norm, spherical design, non-convex minimization, spherical cubature, re-weighted  $\ell_1$

**AMS subject classifications.** 90C26, 90C90, 65D32

## 1. Introduction. Recovery of sparse signals on the unit sphere

$$\mathbb{S}^2 := \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\|_2 = 1\}$$

has many applications in astrophysics, gravitational sensing, geophysics, climate modelling, global navigation, as numerous signals live on the earth's surface and satellite signals affect our everyday life. Those signals are often noisy and corrupted by outliers with different scales. Research on sparse recovery of signals on the sphere is very active [2, 3, 16, 21, 33]. Moreover, three dimensional recognition of human faces with sparse spherical representation has numerous applications in authentication and surveillance, which has advantages over 2D face recognition since it is more accurate and natural to describe and analyse human face data on the sphere [7, 34, 35].

Let  $\mathbb{P}_t(\mathbb{S}^2)$  denote the space of all spherical polynomials of degree at most  $t$  on the unit sphere  $\mathbb{S}^2$ . The dimension of the space of homogeneous harmonic polynomials of exact degree  $\ell$  on  $\mathbb{S}^2$  is  $2\ell + 1$ , thus  $\dim \mathbb{P}_t(\mathbb{S}^2) = \sum_{\ell=0}^t (2\ell + 1) = (t + 1)^2$ . In this paper, we consider spherical polynomials of degree up to  $L$  and  $\dim \mathbb{P}_L(\mathbb{S}^2) = (L + 1)^2$ . Let  $Y_{\ell,k}$ ,  $\ell = 0, \dots, L$ ,  $k = 1, \dots, 2\ell + 1$  be an orthonormal basis for  $\mathbb{P}_L(\mathbb{S}^2)$ .

Suppose that the signal  $F : \mathbb{S}^2 \rightarrow \mathbb{R}$  is observed via the Fourier coefficients

$$(1) \quad c_{\ell,k} = \int_{\mathbb{S}^2} F(\mathbf{x}) Y_{\ell,k}(\mathbf{x}) d\mu(\mathbf{x}) + \eta_{\ell,k}, \quad \ell = 0, \dots, L, \quad k = 1, \dots, 2\ell + 1,$$

\*13 September, 2017, revised 16 January, 2018

**Funding:** This work was supported in part by Hong Kong Research Council Grant PolyU153016/16p.

<sup>†</sup>Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong (max-jchen@polyu.edu.hk, <http://www.polyu.edu.hk/ama/staff/xjchen/ChenXJ.htm>).

<sup>‡</sup>School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052 Australia (r.womersley@unsw.edu.au, <http://www.maths.unsw.edu.au/~rsw/>).

with noise  $\eta_{\ell,k}$ .

We are interested in discrete approximation of (1) using a cubature rule for  $\mathbb{S}^2$  with weights  $w_j > 0, j = 1, \dots, n$  and a set of  $n$  nodes

$$(2) \quad \mathcal{X}_n := \{\mathbf{x}_j \in \mathbb{S}^2, j = 1, \dots, n\},$$

which is exact for all spherical polynomials of degree at most  $2L$ . The nodes should also respect the intrinsic rotational invariance of the sphere and have good separation. For signal processing,  $\mathcal{X}_n$  is a grid where the signals are located.

If  $F \in \mathbb{P}_L(\mathbb{S}^2)$  then a cubature rule which is exact for all spherical polynomials of degree at most  $2L$  gives

$$(3) \quad \int_{\mathbb{S}^2} F(\mathbf{x}) Y_{\ell,k}(\mathbf{x}) d\mu(\mathbf{x}) = \sum_{j=1}^n w_j F(\mathbf{x}_j) Y_{\ell,k}(\mathbf{x}_j), \quad \ell = 0, \dots, L, \quad k = 1, \dots, 2\ell + 1.$$

Applying such a cubature rule to (1), we have the discrete approximation

$$(4) \quad c_{\ell,k} = \sum_{j=1}^n w_j F(\mathbf{x}_j) Y_{\ell,k}(\mathbf{x}_j) + \eta_{\ell,k}, \quad \ell = 0, \dots, L, \quad k = 1, \dots, 2\ell + 1.$$

In matrix notation, (4) is

$$(5) \quad c = YWf + \eta,$$

where  $c \in \mathbb{R}^m$  is the noisy Fourier coefficient vector,  $Y \in \mathbb{R}^{m \times n}$  is the spherical harmonic basis matrix,  $W = \text{diag}(w) \in \mathbb{R}^{n \times n}$  is the diagonal matrix of cubature weights,  $\eta \in \mathbb{R}^m$  is the noise vector, and  $f = (F(\mathbf{x}_1), \dots, F(\mathbf{x}_n))^T \in \mathbb{R}^n$  is the signal vector which we want to recover. Here  $m = (L+1)^2$  is the dimension of  $\mathbb{P}_L(\mathbb{S}^2)$  and  $n$  is the number of nodes of the cubature rule.

The nodes of the cubature rule for  $\mathbb{P}_{2L}(\mathbb{S}^2)$  determine the coefficient matrix  $Y$ , while the weights determine  $W$ . Choosing a good cubature rule is essential for recovery. In Section 2, we discuss cubature rules with positive weights which are exact for all spherical polynomials of degree at most  $2L$ . They include spherical designs [20, 45], spherical  $t_\epsilon$ -designs [50] from extremal fundamental systems [39], tensor product, equally spaced in spherical coordinates  $\theta, \phi$ , HealPix points [27], tensor product, Gauss-Legendre in  $\cos(\theta)$  and equally spaced in  $\phi$  [40]. Recovery of sparse signals on the sphere has been investigated [2, 3] for tensor product grids, under the assumption that the support set of the underlying signal is well-separated. Tensor product grids on the sphere intrinsically do not respect the rotational invariance of the sphere, making it harder to maintain good separation. In Section 2, we show that the spherical  $t$ -design with equal weights and good separation [45, 46] provides good grids.

To use a normalized matrix in the optimization model, we let

$$A = YW^{1/2} \quad \text{and} \quad v = W^{1/2}f.$$

Then  $YWf = Av$ . The  $m$  by  $n$  matrix  $A$  is the spherical harmonic basis matrix with the columns scaled by the square root of the cubature weights. Since the nodes  $\mathcal{X}_n$  and weights  $W$  form a positive weight cubature rule on the sphere that is exact for all spherical polynomials of degree at most  $t \geq 2L$ , the matrix  $A$  satisfies

$$(6) \quad AA^T = YWY^T = I \in \mathbb{R}^{m \times m}$$

so its rows are orthonormal. See Section 2 for more details.

The underlying vector  $v \in \mathbb{R}^n$  has  $v_j$  associated with the node  $\mathbf{x}_j$  for  $j = 1, \dots, n$ . In addition the signal is sparse and the weights are positive, so that its support set

$$J = \text{supp}(v) := \{j \in \{1, \dots, n\} : v_j \neq 0\} = \text{supp}(f) := \{j \in \{1, \dots, n\} : f_j \neq 0\}$$

has few elements  $\|v\|_0 = \|f\|_0 = |J| \ll n$ .

Consider the following constrained optimization problem, known as basis pursuit when  $\sigma = 0$ , so  $Av = c$ , and basis pursuit de-noising when  $\sigma > 0$ ,

$$(7) \quad \begin{aligned} \min \quad & \|v\|_1 \\ \text{s.t.} \quad & \|Av - c\|_2 \leq \sigma. \end{aligned}$$

This has become a classical problem building on the work of Donoho and Huo [23] and Candés and Tao [9], using sufficient conditions on  $A$  to guarantee that minimizing the  $\ell_1$ -norm, a convex problem, gives a solution to the  $\ell_0$ -norm minimization problem. See, for example, [6, 25]. Such sufficient conditions guarantee that the  $\ell_0$ -norm solution is also the unique minimizer for a class of sparsity measures [32], including

$$\|v\|_q^q := \sum_{i=1}^n |v_i|^q \quad \text{for } 0 < q \leq 1.$$

The sphere is invariant under rotations and problem (7) may have multiple solutions, which means that these sufficient conditions do not hold or only hold for unrealistically sparse solutions. In this paper, we use the optimization problem

$$(8) \quad \begin{aligned} \min \quad & \|v\|_q^q \\ \text{s.t.} \quad & \|Av - c\|_2 \leq \sigma, \end{aligned}$$

with  $0 < q < 1$  to recover sparse signals.

*Example 1.1.* This example, for degree  $L = 0$ , has infinitely many solutions of problem (7) that have no zero components, while problem (8) has only two solutions that are the sparsest feasible solutions. Let

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad c = \frac{1}{\sqrt{2}}, \quad 0 \leq \sigma < \frac{1}{\sqrt{2}}.$$

For  $q = 1$  the solution set of (7) is

$$(9) \quad S_1 = \left\{ \begin{bmatrix} v_1 \\ 1 - \sqrt{2}\sigma - v_1 \end{bmatrix}, v_1 \in [0, 1 - \sqrt{2}\sigma] \right\}.$$

For  $0 < q < 1$  the solution set of (8) is

$$(10) \quad S_q = \left\{ \begin{bmatrix} 0 \\ 1 - \sqrt{2}\sigma \end{bmatrix}, \begin{bmatrix} 1 - \sqrt{2}\sigma \\ 0 \end{bmatrix} \right\}.$$

For a regularization form:  $\min \|Av - c\|_2^2 + \lambda \|v\|_q^q$ , see Example 2.1 in [12].

Using  $\|v\|_q^q$  has advantages when seeking sparse solutions, see for example [14, 24, 19, 36, 44], but at the expense of making the minimization problem non-convex. We show that to solve (8), the penalty approach in Chen, Lu and Pong [13] with a

starting point obtained from the re-weighted  $\ell_1$  minimization [10, 15, 18, 51, 49] is promising.

Recovery of sparse signals on the sphere includes three key roles: cubature rules, optimization models and optimization methods, in particular when the sufficient conditions for sparse recovery by  $\ell_1$  minimization are not satisfied. The main contribution of this paper is to show in theory and computation that using a cubature rule with equal weights (spherical  $t$ -design) to build the optimization model (8) and then using the penalty method with a starting point from re-weighted  $\ell_1$  to solve (8) can find better sparse solutions than other cubature rules and the  $\ell_1$  optimization model (7).

The paper is organized as follows. In Section 2, we present properties of cubature rules in building the optimization model (8), that is, matrices  $Y$ ,  $A$ , vector  $c$  and noise control  $\sigma$ . In Section 3, we prove that the quality of the recovery depends on the condition number of the weight matrix  $W$ , so equal weights give the smallest error bound for recovery. In Section 4, we give a new first order optimality condition for problem (8) and describe the penalty method with a starting point from the re-weighted  $\ell_1$  method and its convergence. In Section 5, we present numerical results to compare cubature rules using spherical  $t$ -designs and  $t_\epsilon$ -designs (extremal fundamental systems) with tensor product rules. We also compare solutions of problem (7) and problem (8).

## 2. Cubature rules with positive weights for spherical signal recovery.

A cubature rule with positive weights for  $\mathbb{S}^2$  is a set of nodes  $\mathbf{x}_j \in \mathbb{S}^2$  and weights  $w_j > 0, j = 1, \dots, n$  such that

$$\mathcal{Q}_n(g) := \sum_{j=1}^n w_j g(\mathbf{x}_j) \approx I(g) := \int_{\mathbb{S}^2} g(\mathbf{x}) d\mu(\mathbf{x}),$$

where  $g : \mathbb{S}^2 \rightarrow \mathbb{R}$  and  $d\mu(\mathbf{x})$  denotes the surface (Lebesgue) measure on  $\mathbb{S}^2$ . The cubature rule has *degree of precision*  $t$  if it is exact for all spherical polynomials of degree at most  $t$ , (and is not exact for some polynomial of degree  $t + 1$ ), that is

$$\mathcal{Q}_n(p) = I(p), \quad \text{for all } p \in \mathbb{P}_t(\mathbb{S}^2), \quad \mathcal{Q}_n(\bar{p}) \neq I(\bar{p}), \quad \text{for some } \bar{p} \in \mathbb{P}_{t+1}(\mathbb{S}^2).$$

A cubature rule which is exact for constant polynomials, so

$$\sum_{j=1}^n w_j = |\mathbb{S}^2| = 4\pi,$$

has degree of precision at least 0. Thus an equal weight (or average weight if degree of precision is at least 0) rule has

$$(11) \quad w_j = w_{av} := \frac{|\mathbb{S}^2|}{n} = \frac{4\pi}{n}, \quad j = 1, \dots, n.$$

Given a set  $\mathcal{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{S}^2$  of points, the spherical harmonic basis matrix for degree  $t \geq 0$  is  $Y \in \mathbb{R}^{(t+1)^2 \times n}$  defined by

$$(12) \quad Y_{ij} = Y_{\ell,k}(\mathbf{x}_j), \quad i = \ell^2 + k, \quad k = 1, \dots, 2\ell + 1, \quad \ell = 0, \dots, t, \quad j = 1, \dots, n.$$

When given scattered data, so the nodes are fixed, one would like to find positive weights  $w_j > 0, j = 1, \dots, n$  so that polynomials of degree at most  $t$  are integrated

exactly, giving

$$(13) \quad Yw = \bar{e}_1, \quad \text{where} \quad \bar{e}_1 = \sqrt{|\mathbb{S}^2|} (1, 0, 0, \dots, 0)^T \in \mathbb{R}^{(t+1)^2}.$$

These conditions come from the orthonormality of the spherical harmonics  $Y_{\ell,k}$ , as

$$\int_{\mathbb{S}^2} Y_{\ell,k}(\mathbf{x}) d\mu(\mathbf{x}) = 0, \quad k = 1, \dots, 2\ell + 1, \quad \ell \geq 1.$$

Ideally the weights are chosen to be as close to equal as possible, while still having degree of precision  $t$ . This can be achieved by solving the convex optimization problem

$$(14) \quad \begin{aligned} \min \quad & \left\| \frac{w}{w_{av}} - e \right\|_{\infty} \\ \text{s.t.} \quad & Yw = \bar{e}_1, \end{aligned}$$

where  $e \in \mathbb{R}^n$  is the vector with all elements equal to 1 and  $\bar{e}_1$  is defined in (13). If the resulting weights are not positive, then the requested degree of precision must be reduced, so there are fewer constraints. How difficult it is to solve (14) depends on the point set  $\mathcal{X}_n$  and hence the conditioning of the matrix  $Y$ . We find that problem (14) has positive solutions, all be it with varying degrees of precision for the same number of points, for the following sets of points on the sphere.

- SF spherical designs
- MD maximum determinant
- TP tensor product point with equally spaced points in polar angle and azimuthal angle
- GL Gauss-Legendre points with Gauss points in  $\cos(\text{polar angle})$  and equally spaced points in azimuthal angle
- HealPix points
- minimum energy points

We give a brief introduction for the first four sets of points.

SF A spherical  $t$ -design, introduced in [20], is a set of  $n$  points  $\mathcal{X}_n = \{\mathbf{x}_j, j = 1, \dots, n\}$  on  $\mathbb{S}^2$  such that equal weight cubature has degree of precision  $t$ , that is

$$(15) \quad w_{av} \sum_{j=1}^n p(\mathbf{x}_j) = \int_{\mathbb{S}^2} p(\mathbf{x}) d\mu(\mathbf{x}), \quad \text{for all } p \in \mathbb{P}_t(\mathbb{S}^2).$$

A lower bound [20]  $n \geq n^*(t)$  on the number of points is

$$(16) \quad n^*(t) := \begin{cases} \frac{(t+1)(t+3)}{4} & \text{if } t \text{ odd,} \\ \frac{(t+2)^2}{4} & \text{if } t \text{ even,} \end{cases}$$

but it is known that these lower bounds are not achievable except in a few special cases. Although the existence of spherical  $t$ -designs for  $n$  sufficiently large has been known for some time, it was only recently established [4] that spherical  $t$ -designs on  $\mathbb{S}^d$  exist for all  $n \geq c_d t^d$ , the optimal order, and moreover [5] that well-separated spherical  $t$ -designs also exist. Interval methods [11] have established the existence of spherical  $t$ -designs on  $\mathbb{S}^2$  with  $n = (t+1)^2$  points. If  $\mathcal{X}_n$  is a spherical  $t$ -design, so is the set  $\{\mathcal{X}_n, H\mathcal{X}_n\}$  of  $2n$  points for any orthogonal matrix  $H \in \mathbb{R}^{3 \times 3}$ . Thus spherical  $t$ -designs with more than twice the minimum number of points may have poor separation. Computed

spherical  $t$ -designs with  $n = t^2/2 + O(t)$  and good geometric properties (including separation) are available for  $t \leq 325$  [45, 46]. Computed spherical designs with fewer sampling points are given in [29, 30].

MD Maximum determinant points, also called extremal fundamental systems [39], are sets of  $n \leq (t+1)^2$  points  $\mathcal{X}_n = \{\mathbf{x}_j, j = 1, \dots, n\}$  on  $\mathbb{S}^2$  maximizing the determinant of the Gram matrix

$$\det Y^T Y,$$

where the spherical harmonic basis matrix  $Y \in \mathbb{R}^{(t+1)^2 \times n}$  is defined in (12). The extremal fundamental systems of Sloan and Womersley [39] which have  $n = (t+1)^2$  belong to the class of  $t_e$ -designs studied in Zhou and Chen [50], which provide cubature rules that are exact for all spherical polynomials of degree at most  $t$ , but where flexibility is allowed in the choice of weights,

$$w_{av}(1 - \epsilon) \leq w_i \leq w_{av}(1 - \epsilon)^{-1}, \quad i = 1, \dots, n, \quad \epsilon \in [0, 1].$$

A spherical  $t$ -design corresponds to  $\epsilon = 0$ , so the weights are all equal.

TP and GL use the standard spherical parametrization of  $\mathbf{x} \in \mathbb{S}^2$ , which is, for polar angle  $\theta \in [0, \pi]$  and azimuthal angle  $\phi \in [0, 2\pi)$ ,

$$(17) \quad \mathbf{x} = (\sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi), \cos(\theta))^T.$$

If TP uses equally spaced points  $\theta_i = \pi i/n_\theta, i = 0, \dots, n_\theta - 1$  and  $\phi_j = 2\pi j/n_\phi, j = 0, \dots, n_\phi - 1$ , then there are  $n_\phi$  points at the north pole. Shifting the polar angles away from the poles, so that

$$\theta_i = \frac{\pi(2i+1)}{2n_\theta}, \quad i = 0, \dots, n_\theta - 1, \quad \phi_j = \frac{j2\pi}{n_\phi}, \quad j = 0, \dots, n_\phi - 1,$$

gives  $n = n_\theta n_\phi$  distinct points.

GL uses the Gauss-Legendre nodes  $z_j \in (-1, 1)$  with  $\cos(\theta_j) = z_j$  for  $j = 1, \dots, n_z$ , and still produces sets with points too close together near the poles. Such tensor product nodes fail to respect the rotational invariance of the sphere as illustrated in the first plots in Figures 5 and 6.

There are other spherical quadrature rules and computed spherical designs [28, 29].

The natural distance on the sphere is the *geodesic distance*

$$(18) \quad \text{dist}(\mathbf{x}, \mathbf{y}) = \arccos(\mathbf{x} \cdot \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathbb{S}^2,$$

where  $\mathbf{x} \cdot \mathbf{y}$  is the standard Euclidean inner product and  $|\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x}$ . The Euclidean distance

$$|\mathbf{x} - \mathbf{y}|^2 = 2(1 - \mathbf{x} \cdot \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathbb{S}^2$$

is related to the geodesic distance by

$$|\mathbf{x} - \mathbf{y}| = 2 \sin(\text{dist}(\mathbf{x}, \mathbf{y})/2),$$

illustrating the fact that the geodesic and Euclidean distances are very similar when the distances are small.

A key factor in the quality of nodes  $\mathcal{X}_n = \{\mathbf{x}_j, j = 1, \dots, n\}$  is its separation  $\delta(\mathcal{X}_n)$  defined by

$$\delta(\mathcal{X}_n) := \min_{i \neq j} \text{dist}(\mathbf{x}_i, \mathbf{x}_j),$$

where the (geodesic) distance between two points on the sphere is defined by (18). A sequence of point sets  $\{\mathcal{X}_n\}$  is well-separated if there exists a constant  $c_2$ , independent of  $n$ , such that all the point sets satisfy

$$(19) \quad \delta(\mathcal{X}_n) \geq c_2 n^{-\frac{1}{2}}.$$

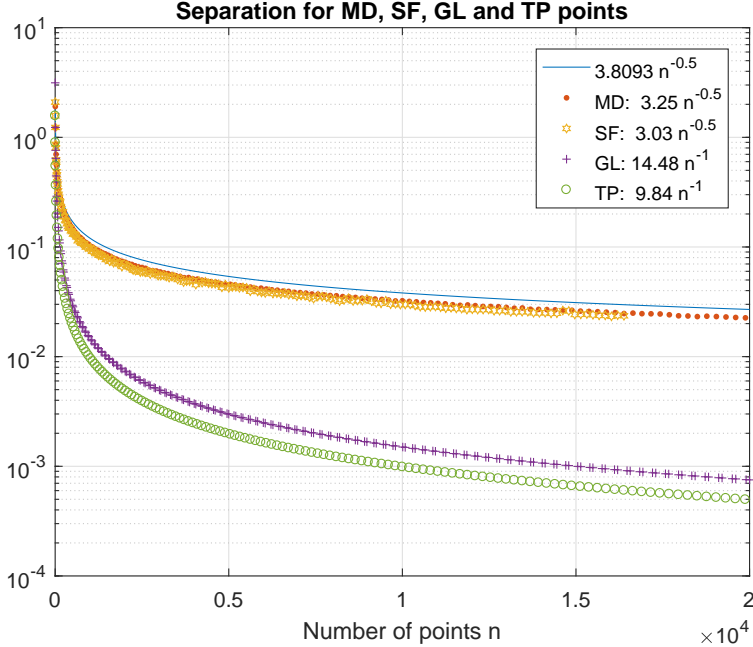


FIG. 1. Separation distance of two classes of point sets on  $\mathbb{S}^2$

The problem of choosing  $n$  points on  $\mathbb{S}^2$  to maximize their separation, see [38, 17] for example, is also known as the best packing problem as the separation is twice the packing radius. The best possible separation for a set of  $n$  points on  $\mathbb{S}^2$  is

$$\delta_n := \max_{\mathcal{X}_n \subset \mathbb{S}^2} \delta(\mathcal{X}_n) \approx c_{\text{sep}} n^{-1/2}.$$

Using the Euclidean distance,  $c_{\text{sep}} = (8\pi/\sqrt{3})^{1/2}$ . Figure 1 shows the minimal separation  $\delta_n$  as a function of the number of points  $n$ , for two classes of point sets on  $\mathbb{S}^2$ . The first class, which are well-separated as  $\delta_n$  has the optimal order  $n^{-1/2}$ , include the MD and SF points. The second class are tensor product point sets including GL and TP. The minimal separation for the GL and TP points decays as  $n^{-1}$ , reflecting the poor separation near the poles for tensor-product point sets. A separation condition

$$\delta(\mathcal{X}_n) \geq \frac{\nu}{L},$$

where  $\nu > 0$  is independent of  $n$  and  $L$  as in [2], implies that  $L = O(n^{\frac{1}{2}})$  for well-separated point sets, but for tensor product nodes  $L = O(n)$ . Note that good separation of the grid points  $\mathcal{X}_n$  is not sufficient to guarantee that they form a fundamental system, as on  $\mathbb{S}^2$  two points may be interchanged, hence changing the sign of the determinant of  $Y$ , whilst keeping separation.

If the cubature rule with nodes  $\mathcal{X}_n$  and weights  $w_j, j = 1, \dots, n$  has degree of precision  $t \geq 2L$ , the orthonormality of the spherical harmonics implies (6) for a spherical harmonic basis matrix  $Y$  of degree  $L$ . In particular, if the points  $\mathcal{X}_n$  form a spherical  $t$ -design, so  $W = w_{av}I_n$ , with  $t \geq 2L$ , then

$$(20) \quad YY^T = (w_{av})^{-1} I_m,$$

for  $m = (L+1)^2$ . The requirement that  $t \geq 2L$  and the lower bounds (16) imply that  $n > n^*(2L) \approx L^2$ .

For  $n \geq m = (L+1)^2$ , the singular values of  $Y$  are  $\sigma_i = (w_{av})^{-\frac{1}{2}}$  for  $i = 1, \dots, (L+1)^2$ , so  $Y$  has the ideal condition number 1. The Gram matrix  $G = Y^T Y \in \mathbb{R}^{n \times n}$ , is symmetric and positive semi-definite with singular values  $\sigma_i(G) = (w_{av})^{-1}$  for  $i = 1, \dots, m$  and  $\sigma_i = 0$  for  $i = m+1, \dots, n$ . Thus maximizing the determinant or minimizing the condition number of  $G$  only makes sense when the points do not form a  $2L$ -design, for example  $n = (L+1)^2$  as in [1].

**3. Conditions for recovery.** Problem (7) is a convex optimization problem, and widely used in signal processing [25]. However, from our numerical experiments, problem (8) with the folded concave objective function is more efficient for recovering sparse signals on the sphere than problem (7). We also find that several sufficient conditions which guarantee that the solution of a convex  $\ell_1$ -norm minimization problem will give the solution to the  $\ell_0$ -norm problem fail. For example, sufficient conditions<sup>1</sup> on  $\text{spark}(A)$ , mutual incoherence  $M(A)$ , restricted isometry property(RIP) and null space of  $A$  ensure that

$$\text{argmin}\{\|v\|_1 : Av = c\} = \text{argmin}\{\|v\|_0 : Av = c\}.$$

In this section, we consider sufficient conditions on local minimizers of the non-convex non-Lipschitz  $\ell_q$  minimization problem (8) for recovery of sparse signals on the sphere.

Denote the feasible set of problem (8) by

$$\Omega = \{v : \|Av - c\|_2 \leq \sigma\}.$$

ASSUMPTION 3.1.  $\Omega$  has an interior point  $\tilde{v}$  such that  $\|A\tilde{v} - c\|_2 < \sigma$ .

Let  $S_q$  be the set of local minimizers of problem (8).

For a given set  $J \subset \{1, 2, \dots, m\}$ , consider the convex optimization problem

$$(21) \quad \begin{aligned} \min \quad & \|v\|_2^2 \\ \text{s.t.} \quad & \|Av - c\|_2 \leq \sigma, \quad v_{J^c} = 0, \end{aligned}$$

where  $v_{J^c}$  is the vector of the elements  $v_i, i \notin J$ . Let  $v^\circ$  be the unique solution of (21).

---

1

- Spark: Donoho, Elad [22],  $\text{spark}(A) = \min\{\|v\|_0 : Av = 0, v \neq 0\}$ .
- Mutual incoherence: Donoho, Huo [23],  $M(A) = \max_{i \neq j} |A_i^T A_j| / (\|A_i\|_2 \|A_j\|_2)$ .
- RIP: Candes, Romberg, Tao [8, 25] Smallest  $\delta_s$  s.t.

$$(1 - \delta_s)\|v\|_2^2 \leq \|Av\|_2^2 \leq (1 + \delta_s)\|v\|_2^2, \quad \forall v : \|v\|_0 \leq s.$$

- Null space: Zhang [48],  $(\|v\|_0)^{\frac{1}{2}} < \frac{1}{2} \min\{\|v\|_1 / \|v\|_2 : Av = 0, v \neq 0\}$ .



If  $J$  is the support set of a true signal, we call  $v^\circ$  an oracle solution. In what follows, we estimate the error between an oracle solution and a local minimizer of problem (8), that is an upper bound for  $\|v^\circ - \hat{v}\|_q^q$  for  $\hat{v} \in S_q$ .

Let

$$(22) \quad Q = A^T A \quad \text{and} \quad P = I - Q,$$

which are projection matrices for the range space of  $A^T$  and the null space of  $A$  respectively, so  $QQ = Q$ ,  $PP = P$  and  $PQ = 0$ . From (6), the matrix  $A$  has full row rank.

Let

$$\Omega_J(\hat{v}) = \{v \in \Omega : \|\hat{v}\|_q^q \leq \|v\|_q^q, \quad v_{J^c} = 0\}.$$

For any  $\hat{v} \in S_q$ , from the continuity of  $\|v\|_q^q$  and Assumption 3.1,  $\Omega_J(\hat{v})$  is nonempty.

LEMMA 3.1. *Given  $\hat{v} \in S_q$  and  $v^* \in \Omega_J(\hat{v})$ , let  $r = (I - Q)(\hat{v} - v^*)$ . If*

$$(23) \quad \|r_J\|_q \leq \gamma \|r_{J^c}\|_q,$$

for some  $\gamma \in [0, 1)$ , then

$$(24) \quad \|\hat{v} - v^*\|_q^q \leq \frac{2(2\beta\sigma)^q}{1 - \gamma^q},$$

where  $\beta = n^{\frac{2-q}{2q}}$ .

*Proof.* As  $QQ = Q = A^T A$ , the feasibility of  $\hat{v}$  and  $v^*$  implies

$$\|Q(\hat{v} - v^*)\|_2^2 = \|A(\hat{v} - v^*)\|_2^2 \leq (\|A\hat{v} - c\|_2 + \|Av^* - c\|_2)^2 \leq 4\sigma^2.$$

Using this bound and  $v_{J^c}^* = 0$  with the inequality

$$\| \|v\|_q^q - \|u\|_q^q \| \leq \|v - u\|_q^q,$$

from Lemma 2.4 in [13], we have

$$\begin{aligned} \|\hat{v}\|_q^q &= \|v^* + (I - Q)(\hat{v} - v^*) + Q(\hat{v} - v^*)\|_q^q \\ &\geq \|v^* + (I - Q)(\hat{v} - v^*)\|_q^q - \|Q(\hat{v} - v^*)\|_q^q \\ &= \|(v^* + r)_J\|_q^q + \|r_{J^c}\|_q^q - \|Q(\hat{v} - v^*)\|_q^q \\ &\geq \|v^*\|_q^q - \|r_J\|_q^q + \|r_{J^c}\|_q^q - \|Q(\hat{v} - v^*)\|_q^q \\ &= \|v^*\|_q^q + (1 - \gamma^q)\|r_{J^c}\|_q^q - \|Q(\hat{v} - v^*)\|_q^q \\ &\geq \|\hat{v}\|_q^q + (1 - \gamma^q)\|r_{J^c}\|_q^q - \|Q(\hat{v} - v^*)\|_q^q, \end{aligned}$$

where the last inequality uses  $\|\hat{v}\|_q^q \leq \|v^*\|_q^q$ . This gives

$$(25) \quad (1 - \gamma^q)\|r_{J^c}\|_q^q \leq \|Q(\hat{v} - v^*)\|_q^q.$$

Using Hölder's inequality<sup>2</sup> and the definition of  $\beta$ , we have

$$\|Q(\hat{v} - v^*)\|_q \leq \beta \|Q(\hat{v} - v^*)\|_2 \leq 2\beta\sigma.$$

---

2

$$\|v\|_q^q = \sum_{i=1}^n |v_i|^q \cdot 1 \leq \left( \sum_{i=1}^n (|v_i|^q)^{\frac{2}{q}} \right)^{\frac{q}{2}} \left( \sum_{i=1}^n 1^{\frac{2}{2-q}} \right)^{\frac{2-q}{2}} = n^{\frac{2-q}{2}} \|v\|_2^q.$$

Hence, we can derive the error bound (24) as follows.

$$\begin{aligned}
\|(I - Q)(\hat{v} - v^*)\|_q^q &= \|r_J\|_q^q + \|r_{J^c}\|_q^q \\
&\leq (1 + \gamma^q) \|r_{J^c}\|_q^q \\
&\leq \frac{1 + \gamma^q}{1 - \gamma^q} \|Q(\hat{v} - v^*)\|_q^q \\
&\leq \frac{1 + \gamma^q}{1 - \gamma^q} (2\beta\sigma)^q,
\end{aligned}$$

and

$$\|\hat{v} - v^*\|_q^q \leq \|(I - Q)(\hat{v} - v^*)\|_q^q + \|Q(\hat{v} - v^*)\|_q^q \leq \frac{2(2\beta\sigma)^q}{1 - \gamma^q}.$$

□

We now show how the weights for the cubature rule affect the recovery error. Let  $A = YW^{\frac{1}{2}}$  and  $\max_i w_i = \kappa \frac{4\pi}{n}$ , with  $\kappa \geq 1$ . Consider the following problem

$$\begin{aligned}
(26) \quad &\min \|u\|_q^q \\
&\text{s.t. } \|Yu - c\|_2 \leq \sigma.
\end{aligned}$$

Let

$$\nu = \frac{\max_{1 \leq i \leq n} w_i}{\min_{1 \leq i \leq n} w_i}.$$

**THEOREM 3.2.** *Given  $\hat{v} \in S_q$  and  $v^* \in \Omega_J(\hat{v})$  satisfying (23), the vectors of  $\hat{u} = W^{\frac{1}{2}}\hat{v}$  and  $u^* = W^{\frac{1}{2}}v^*$  are feasible points of (26), and*

$$(27) \quad \|\hat{u}\|_q \leq \sqrt{\nu} \min\{\|u\|_q : \|Yu - c\|_2 \leq \sigma\}.$$

Moreover, we have

$$(28) \quad \|\hat{u} - u^*\|_q^q \leq \frac{2(4\beta_1 \sqrt{\kappa\pi}\sigma)^q}{1 - \gamma^q},$$

where  $\beta_1 = n^{\frac{1-q}{q}}$ .

*Proof.* From  $A = YW^{\frac{1}{2}}$ ,  $\hat{u}$  and  $u^*$  are feasible solutions of (26). Moreover, for any feasible solution  $\tilde{u}$  of (26), there is a feasible solution  $\tilde{v}$  of (8) such that  $\tilde{u} = W^{\frac{1}{2}}\tilde{v}$ . Hence, we can obtain (27) from the following

$$\begin{aligned}
\|\hat{u}\|_q^q &= \|W^{\frac{1}{2}}\hat{v}\|_q^q = \sum_{i=1}^n w_i^{\frac{q}{2}} |\hat{v}_i|^q \leq \max w_i^{\frac{q}{2}} \sum_{i=1}^n |\hat{v}_i|^q \\
&\leq \max w_i^{\frac{q}{2}} \sum_{i=1}^n |\tilde{v}_i|^q \\
&\leq \frac{\max w_i^{\frac{q}{2}}}{\min w_i^{\frac{q}{2}}} \sum_{i=1}^n w_i^{\frac{q}{2}} |\tilde{v}_i|^q \\
&= \nu^{\frac{q}{2}} \|W^{\frac{1}{2}}\tilde{v}\|_q^q \\
&= \nu^{\frac{q}{2}} \|\tilde{u}\|_q^q.
\end{aligned}$$

Using  $\beta = \beta_1 \sqrt{n}$  and

$$\|\hat{v} - v^*\|_q^q \leq \frac{2(2\beta_1 \sigma \sqrt{n})^q}{1 - \gamma^q},$$

together with  $\max_i w_i = \kappa \frac{4\pi}{n}$ , we find

$$\begin{aligned} \|\hat{u} - u^*\|_q^q &= \|W^{\frac{1}{2}}(\hat{v} - v^*)\|_q^q \leq \max_i w_i^{\frac{q}{2}} \|\hat{v} - v^*\|_q^q \\ &\leq \left( \sqrt{\kappa \frac{4\pi}{n}} \right)^q \frac{2(2\beta\sigma)^q}{1 - \gamma^q} = \frac{2(4\beta_1 \sqrt{\kappa\pi}\sigma)^q}{1 - \gamma^q}. \end{aligned}$$

□

REMARK 3.1. If we choose a spherical  $t$ -design with  $t \geq 2L$ , we have  $\kappa = 1$  and  $\nu = 1$ . Hence,  $\hat{u} = \frac{4\pi}{n}\hat{v}$  is a local minimizer of problem (26),  $u^* = \frac{4\pi}{n}v^*$  and

$$(29) \quad \|\hat{u} - u^*\|_q \leq \frac{2(4\beta_1 \sqrt{\pi}\sigma)^q}{1 - \gamma^q}.$$

Moreover, for  $q = 1$ , we have  $\beta_1 = 1$  and  $\hat{u} = \frac{4\pi}{n}\hat{v}$  is a minimizer of problem (26). The error bound in (29) reduces to

$$(30) \quad \|\hat{u} - u^*\|_1 \leq \frac{8\sqrt{\pi}\sigma}{1 - \gamma}.$$

REMARK 3.2. The set  $S_q$  of local minimizers of problem (8) with  $\sigma = 0$  has only a finite number of elements. See Corollary 2.1 in [14] and Theorem 3 in [26]. This implies that the probability of getting a  $\hat{v} \in S_q$  satisfying  $\|\hat{v}\|_q \leq \|v^o\|_q$  and condition (23) is positive for  $q \in (0, 1)$ . However, the probability of getting a  $\hat{v} \in S_q$  satisfying condition (23) may be 0 for  $q = 1$ . Consider Example 1.1, for which

$$I - Q = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

For  $J = \{1\}$  the oracle solution is  $v^o = (1 - \sqrt{2}\sigma, 0)$ . For  $q \in (0, 1)$ ,  $S_q$  is given in (10) which has only two minimizers, while for  $q = 1$ ,  $S_1$  is given in (9) which has infinitely many minimizers. Moreover,  $v^o \in S_q$ , for  $0 < q < 1$ . However, for  $q = 1$ ,

$$r = (I - Q)(\hat{v} - v^o) = \begin{bmatrix} \hat{v}_1 + \sqrt{2}\sigma - 1 \\ 1 - \hat{v}_1 - \sqrt{2}\sigma \end{bmatrix}.$$

Hence, the probability of getting a  $\hat{v} \in S_q$  satisfying condition (23) is  $\frac{1}{2}$  for  $q \in (0, 1)$ . However, the probability of getting a  $\hat{v} \in S_1$  satisfying condition (23) is 0 for  $q = 1$ . The conclusion also holds for Example 1.1 with  $J = \{2\}$ . Figure 2 shows the solution set of Example 1.1 with  $q = 1$  and  $q \in (0, 1)$ .

**4. Optimality conditions and algorithms.** Problem (8) minimizes a non-Lipschitz function over a convex set. Due to the separability of the objective function  $\varphi(v) := \|v\|_q^q$  from [37, Definition 8.3] and [13, Lemma 2.5], the (limiting) subdiffe-

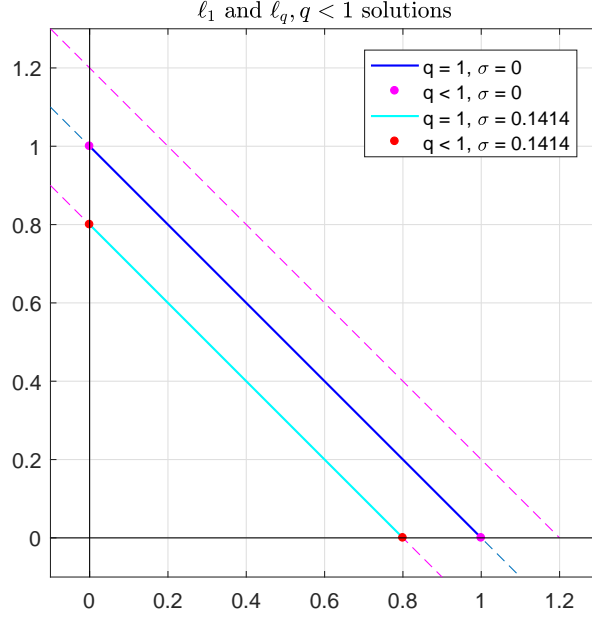


FIG. 2. Example 1.1 with multiple solutions for  $q = 1$  and  $q \in (0, 1)$ :  $S_1 = \{(v_1, 1 - \sqrt{2}\sigma - v_1)^T, v_1 \in [0, 1 - \sqrt{2}\sigma]\}$  containing a segment between  $(0, 1 - \sqrt{2}\sigma)^T$  and  $(1 - \sqrt{2}\sigma, 0)^T$ , and  $S_q = \{(0, 1 - \sqrt{2}\sigma)^T, (1 - \sqrt{2}\sigma, 0)^T\}$  containing only two points.

rential and horizon subdifferential of  $\varphi(v)$  can be given respectively as

$$\begin{aligned} \partial\varphi(v) &:= \left\{ u : \exists v^k \xrightarrow{\varphi} v, u^k \rightarrow u \text{ with } \liminf_{z \rightarrow v^k} \frac{\varphi(z) - \varphi(v^k) - \langle u^k, z - v^k \rangle}{\|z - v^k\|} \geq 0 \forall k \right\} \\ &= \partial|v_1|^q \times \dots \times \partial|v_n|^q, \\ \partial^\infty\varphi(v) &:= \left\{ u : \exists v^k \xrightarrow{\varphi} v, \lambda_k u^k \rightarrow u, \lambda_k \downarrow 0 \text{ with} \right. \\ &\quad \left. \liminf_{z \rightarrow v^k} \frac{\varphi(z) - \varphi(v^k) - \langle u^k, z - v^k \rangle}{\|z - v^k\|} \geq 0 \forall k \right\} \\ &= \partial^\infty|v_1|^q \times \dots \times \partial^\infty|v_n|^q, \end{aligned}$$

where  $\lambda_k \downarrow 0$  means  $\lambda_k > 0$  and  $\lambda_k \rightarrow 0$ , and  $v^k \xrightarrow{\varphi} v$  means both  $v^k \rightarrow v$  and  $\varphi(v^k) \rightarrow \varphi(v)$ . By the definition of the (limiting) subdifferential and horizon subdifferential, we have

$$\begin{aligned} \partial|v_i|^q &= \{q|v_i|^{q-1}\text{sign}(v_i)\}, \quad \partial^\infty|v_i|^q = \{0\}, \quad \text{if } v_i \neq 0, \\ \partial|v_i|^q &= \partial^\infty|v_i|^q = (-\infty, \infty), \quad \text{if } v_i = 0. \end{aligned}$$

**DEFINITION 4.1 (KKT point of problem (8)).** We say that  $v$  is a KKT point of (8) if  $v \in \Omega$  and

$$(31) \quad 0 \in \mathcal{N}_\Omega(v) + \partial\|v\|_q^q,$$

where  $\mathcal{N}_\Omega(v)$  is the normal cone at  $v \in \Omega$  defined as

$$\mathcal{N}_\Omega(v) := \{y : y^T(u - v) \leq 0 \quad \forall u \in \Omega\}.$$

From [37, Theorem 8.15], any locally optimal solution  $\bar{v}$  of (8) is a KKT point assuming the following constraint qualification holds:

$$(32) \quad -\partial^\infty(\|\bar{v}\|_q^q) \cap \mathcal{N}_\Omega(\bar{v}) = \{0\}.$$

If  $\text{int}(\Omega) \neq \emptyset$ , then

$$\mathcal{N}_\Omega(v) = \begin{cases} \{\tau A^T(Av - c) : \tau \geq 0\} & \text{if } \|Av - c\| = \sigma, \\ \{0\} & \text{if } \|Av - c\| < \sigma. \end{cases}$$

In this case, condition (31) can be written as

$$(33) \quad 0 = \tau(A^T(Av - c))_i + q|v_i|^{q-1}, \quad \text{if } v_i \neq 0, \quad \text{for some } \tau \geq 0$$

and the constraint qualification (32) can be written as

$$(34) \quad (A^T(Av - c))_i \neq 0, \quad \text{if } v_i \neq 0.$$

Now we propose a new first order optimality condition for problem (8) without a constraint qualification via the approximate problem

$$(35) \quad \begin{aligned} \min \quad & \| |v| + \mu e \|_q^q := \sum_{i=1}^n (|v_i| + \mu)^q \\ \text{s.t.} \quad & \|Av - c\|_2 \leq \sigma, \end{aligned}$$

where  $\mu > 0$  and  $e \in \mathbb{R}^n$  is the vector with all its entries being one.

For  $v \in \mathbb{R}^n$  and  $\mu > 0$ , the subdifferential of  $\| |v| + \mu e \|_q^q$  is the set-valued function

$$\mathcal{W}(v, \mu) = \left\{ q \begin{pmatrix} (|v_1| + \mu)^{q-1} \alpha_1 \\ \vdots \\ (|v_n| + \mu)^{q-1} \alpha_n \end{pmatrix} : \alpha_i \in \begin{cases} \{1\} & \text{if } v_i > 0 \\ \{-1\} & \text{if } v_i < 0 \\ [-1, 1] & \text{if } v_i = 0 \end{cases} \right\}.$$

LEMMA 4.2. *If  $v$  is a local minimizer of (35), then*

$$(36) \quad 0 \in (I - Q)\mathcal{W}(v, \mu) := \{y : y = (I - Q)u, u \in \mathcal{W}(v, \mu)\}.$$

*Proof.* Since  $Q = A^T A$  and  $Q^2 = Q$ ,  $I - Q$  is a symmetric positive semi-definite matrix and  $(I - Q)(I - Q) = (I - Q)$ . This implies

$$(37) \quad ((I - Q)u, u) = ((I - Q)u, (I - Q)u) \geq 0, \quad \text{for } u \in \mathbb{R}^n.$$

For a fixed positive number  $\xi > 0$ ,  $(|t| + \xi)^q$  is a folded concave function. Thus we have

$$(|t^+| + \xi)^q \leq (|t| + \xi)^q + q(|t| + \xi)^{q-1} \text{sign}(t)(t^+ - t), \quad \text{for } tt^+ \geq 0, t \neq 0.$$

Moreover, there is  $\alpha_t \in [-1, 1]$  such that

$$(|t^+| + \xi)^q \leq \xi^q + q\xi^{q-1} \alpha_t t^+, \quad \text{for } t = 0.$$

Let

$$v^+ = v - \frac{\eta}{\|(I - Q)\omega^*(v, \mu)\|_2} (I - Q)\omega^*(v, \mu),$$

where

$$0 < \eta \leq \min_{v_i \neq 0} |v_i|$$

and  $\omega^*(v, \mu)$  is a solution of the following problem

$$(38) \quad \begin{aligned} \|(I - Q)\omega^*(v, \mu)\|_2 &= \min && \|(I - Q)\omega\|_2 \\ \text{s.t.} &&& \omega \in \mathcal{W}(v, \mu). \end{aligned}$$

Then we have  $v_i v_i^+ \geq 0$ . Hence, from (37) we obtain

$$\begin{aligned} \||v^+| + \mu e\|_q^q &\leq \||v| + \mu e\|_q^q + (\omega^*(v, \mu), v^+ - v) \\ &= \||v| + \mu e\|_q^q - \frac{\eta}{\|(I - Q)\omega^*(v, \mu)\|_2} ((I - Q)\omega^*(v, \mu), (I - Q)\omega^*(v, \mu)) \\ &\leq \||v| + \mu e\|_q^q - \eta \|(I - Q)\omega^*(v, \mu)\|_2. \end{aligned}$$

Moreover, from  $\|Av - c\|_2 \leq \sigma$  and  $AA^T = I$ , we have

$$\begin{aligned} A(v^+ - v) &= -A \frac{\eta}{\|(I - Q)\omega^*(v, \mu)\|_2} (I - Q)\omega^*(v, \mu) \\ &= \frac{\eta}{\|(I - Q)\omega^*(v, \mu)\|_2} A(I - A^T A)\omega^*(v, \mu) = 0, \end{aligned}$$

which implies

$$\|Av^+ - c\|_2 = \|Av + A(v^+ - v) - c\|_2 = \|Av - c\|_2 \leq \sigma.$$

Hence, if  $0 \notin (I - Q)\mathcal{W}(v, \mu)$ , then  $\|(I - Q)\omega^*(v, \mu)\|_2 \neq 0$ , and for all sufficiently small  $\eta$ ,  $v^+$  is in a neighborhood of  $v$ ,  $\||v^+| + \mu e\|_q^q < \||v| + \mu e\|_q^q$  and  $\|Av^+ - c\|_2 \leq \sigma$ . This means that  $v$  is not a local minimizer of (35).  $\square$

Now we present the relation between problems (8) and (35) regarding global minimizers, and a new first order optimality condition for problem (8) using (35).

**THEOREM 4.3.** *(i) Let  $v^*$  and  $v_\mu$  be global minimizers of problems (8) and (35), respectively. Then  $v^*$  and  $v_\mu$  are  $n\mu^q$ -global minimizers of (35) and (8), respectively, that is,*

$$\||v^*| + \mu e\|_q^q \leq \min_{v \in \Omega} \||v| + \mu e\|_q^q + n\mu^q$$

and

$$\|v_\mu\|_q^q \leq \min_{v \in \Omega} \|v\|_q^q + n\mu^q.$$

*(ii) Let  $\hat{v}$  be a local minimizer of problem (8). Then there is  $\bar{\mu} > 0$ , such that for any  $\mu \in (0, \bar{\mu})$ ,*

$$(39) \quad 0 \in (I - Q)\mathcal{W}(\hat{v}, \mu).$$

*(iii) Let  $\hat{v}$  be a local minimizer of problem (8) with  $J = \{i : \hat{v}_i \neq 0\}$ . Suppose that there is  $v_J^0 \in \mathbb{R}^{|J|}$  such that  $\|A_J v_J^0 - c\| < \sigma$ . Then there is  $\lambda^* > 0$  such that whenever  $\lambda \geq \lambda^*$ ,  $\hat{v}$  is a local minimizer of the penalty problem*

$$(40) \quad \min F_\lambda(v) := \|v\|_q^q + \lambda(\|Av - c\|_2^2 - \sigma^2)_+$$

and for  $i \in J$ ,

$$(41) \quad |\hat{v}_i| \geq C_L := \left( \frac{q}{2\sqrt{\lambda}\sqrt{F_\lambda(v^\diamond) + \lambda\sigma^2}} \right)^{\frac{1}{1-q}},$$

where  $v^\diamond$  is chosen such that  $F_\lambda(\hat{v}) \leq F_\lambda(v^\diamond)$ .

*Proof.* (i) Note that problems (8) and (35) have the same feasible set. Let  $v^*$  and  $v_\mu$  be global minimizers of problems (8) and (35), respectively. Then from

$$\| |v^*| + \mu e \|_q^q \leq \|v^*\|_q^q + \|\mu e\|_q^q \leq \|v_\mu\|_q^q + n\mu^q \leq \| |v_\mu| + \mu e \|_q^q + n\mu^q = \min_{v \in \Omega} \| |v| + \mu e \|_q^q + n\mu^q$$

and

$$\|v_\mu\|_q^q \leq \| |v_\mu| + \mu e \|_q^q \leq \| |v^*| + \mu e \|_q^q \leq \|v^*\|_q^q + \|\mu e\|_q^q = \min_{v \in \Omega} \|v\|_q^q + n\mu^q$$

we find that  $v^*$  is a  $n\mu^q$ -global minimizer of (35) and  $v_\mu$  is a  $n\mu^q$ -global minimizer of (8).

(ii) Since  $\hat{v}$  is a local minimizer of problem (8), which is an isolated local minimizer by the folded concavity of  $\|\cdot\|_q^q$ , there is neighborhood  $N_{\hat{v}} = \{v : \|v - \hat{v}\|_2 \leq \eta\}$  of  $\hat{v}$  for some  $\eta \in (0, \min_{\hat{v}_i \neq 0} |\hat{v}_i|)$  such that for any  $v \in N_{\hat{v}} \cap \Omega$ ,  $v \neq \hat{v}$ , we have  $\|\hat{v}\|_q^q < \|v\|_q^q$ . Specially, we have  $\min\{\|v\|_q^q : \|v - \hat{v}\|_2 = \eta, v \in \Omega\} - \|\hat{v}\|_q^q > \epsilon$  for some  $\epsilon > 0$ .

By the same argument as in the proof of Lemma 4.2, if  $0 \notin (I - Q)\mathcal{W}(\hat{v}, \mu)$ , then  $\|(I - Q)\omega^*(\hat{v}, \mu)\|_2 \neq 0$ . Let

$$v^+ = \hat{v} - \frac{\eta}{\|(I - Q)\omega^*(\hat{v}, \mu)\|_2} (I - Q)\omega^*(\hat{v}, \mu).$$

Then  $v^+ \in N_{\hat{v}} \cap \Omega$ ,  $v_i^+ \hat{v}_i \geq 0$  and  $\eta = \|v^+ - \hat{v}\|_2$ , which implies  $\|v^+\|_q^q - \|\hat{v}\|_q^q > \epsilon$ . Hence using the same argument as in the proof of Lemma 4.2 again, we have

$$\|v^+\|_q^q \leq \| |v^+| + \mu e \|_q^q \leq \| |\hat{v}| + \mu e \|_q^q - \eta \|(I - Q)\omega^*(\hat{v}, \mu)\|_2 < \|\hat{v}\|_q^q + n\mu^q.$$

This cannot hold for  $\mu \leq (\epsilon/n)^{\frac{1}{q}}$ . Thus we derive (ii).

(iii) This part is from Theorem 3.2 in [13].  $\square$

Berg and Friedlander developed an efficient package SPGL1 [41, 42, 43] (Version 1.9) for solving the  $\ell_1$  basis pursuit de-noising problem (7). Chen, Lu and Pong developed a penalty method [13] for solving the  $\ell_q$  problem (8) using (40). Motivated by Theorem 4.3, we propose the following algorithm.

ALGORITHM 4.1. Choose  $k^* > 1$  and  $\mu > 0$ . Let  $\alpha_i^0 = 1, i = 1, \dots, n$  and  $k = 0$ .

1. Using SPGL1, solve

$$(42) \quad \begin{aligned} v^k \in \operatorname{Argmin} \quad & \sum_{i=1}^n \alpha_i^k |v_i| \\ \text{s.t.} \quad & \|Av - c\|_2 \leq \sigma. \end{aligned}$$

2. If  $k + 1 = k^*$ , go to Step 3. Otherwise let  $\alpha_i^{k+1} = q(|v_i^k| + \mu)^{q-1}$ ,  $i = 1, \dots, n$  and go to Step 1.

3. Using  $v^{k^*}$  as an initial point, solve problem (8) by the penalty method with (40) in [13].

Algorithm 4.1 is the penalty method in [13] for solving problem (8) with a starting point generated by the re-weighted  $\ell_1$  method. By Theorem 4.2 in [13], the sequence generated by Algorithm 4.1 is bounded and any accumulation point of the sequence is a KKT point of (8) which satisfies (31).

In Section 5, we show that the performance of Algorithm 4.1 for solving problem (8) is better than SPGL1 for problem (7) and the penalty method for problem (8) with other initial points.

**5. Experimental results.** In this section, we present numerical results for recovery of sparse signals on the sphere. We use four cubature rules on  $\mathbb{S}^2$  to generate the matrix  $A$  and the vector  $c$  in problem (8) and three methods to solve problems (8) generated by the four cubature rules.

The four cubature rules, as detailed in Section 2, are

SF Spherical designs [20, 45, 46]

MD Spherical  $t_\epsilon$ -designs [50] from extremal fundamental systems [39]

GL Tensor product, Gauss-Legendre points in  $z \in (-1, 1)$  and equally spaced points in  $\phi$  [40]

TP Tensor product, equally spaced in spherical coordinates  $\theta, \phi$

The three methods:

SPGL1 Solve the  $\ell_1$  basis pursuit de-noising problem (7) [41, 42, 43]

RWL1-8 Steps 1-2 in Algorithm 4.1 with  $k^* = 8$

penalty  $(\ell_q, \text{ones})$ ,  $(\ell_q, \text{SPGL1})$  and  $(\ell_q, \text{RWL1-8})$  solve the  $\ell_q$  problem (8) using penalty method (40) with ones, a solution from SPGL1 and a solution from RWL1-8 as starting points, respectively

All codes are written in MATLAB, and the experiments were performed in MATLAB version R2016b.

**5.1. Four cubature rules.** The properties of the node sets (grids) of the four cubature rules for  $n \approx 1024$  are summarized in Table 1.

Code	$n$	$t$	$\delta(\mathcal{X}_n)$	$\delta(\mathcal{X}_n)n^{\frac{1}{2}}$	$\text{cond}(Y)$	$\min w_i/w_{av}$	$\max w_i/w_{av}$
SF	1014	44	9.8e-2	3.12	1.9	1.000	1.000
MD	1024	31	1.0e-1	3.24	5.9	0.705	1.350
GL	1058	45	1.4e-2	0.453	4.0e+16	0.154	1.537
TP	1024	30	9.6e-3	0.308	3.7e+17	0.067	1.569

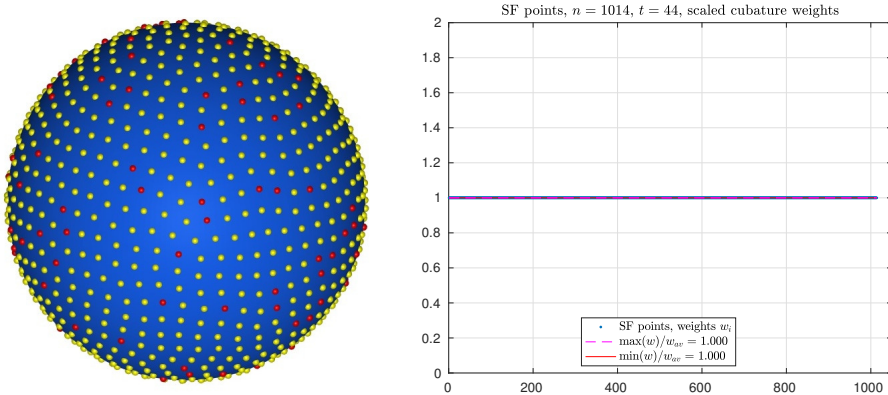
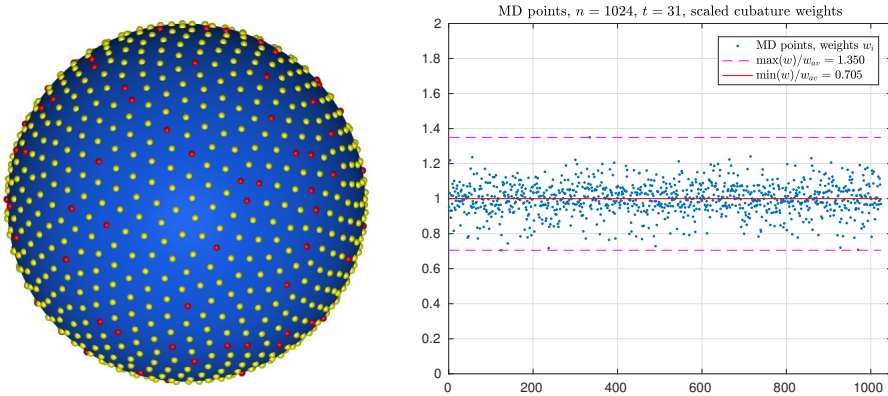
TABLE 1  
Characteristics of node sets

In Figures 3-6, we show the distribution of these nodes on the sphere and the corresponding cubature weights.

Examples of spherical  $t_\epsilon$ -designs include the extremal fundamental systems of Sloan and Womersley [39], with  $\epsilon = 0.5$  up to degree  $t = 165$ , see the first plot in Figure 7. The extremal systems of points have the additional advantage of having a provable lower bound on the separation, as illustrated in the second plot in Figure 7. If the solution to the optimization problem (14) is feasible and satisfies  $\|\frac{w}{w_{av}} - e\|_\infty \leq \bar{\epsilon}$ , for  $\bar{\epsilon} \in [0, 1)$ , then the points and weights form a  $t_{\bar{\epsilon}}$ -design (as  $(1 + \epsilon) \leq (1 - \epsilon)^{-1}$  for  $\epsilon \in [0, 1)$ ).

**5.2. Three algorithms.** The  $\ell_1$  basis pursuit de-noising problem (7) is a convex problem, and could be solved by a variety of other methods, including CVX [31] a



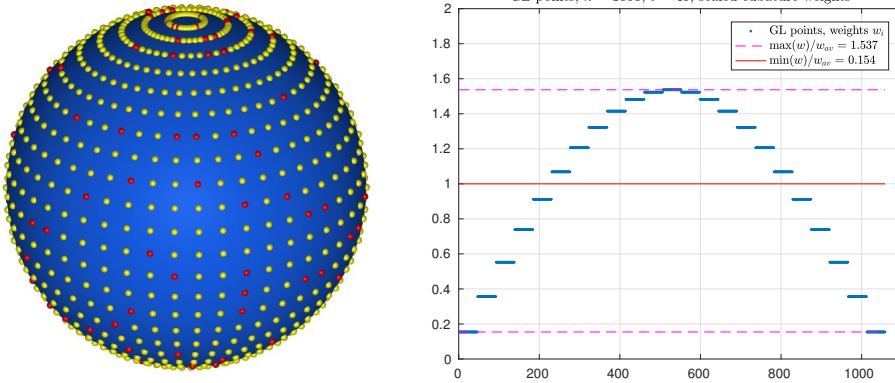
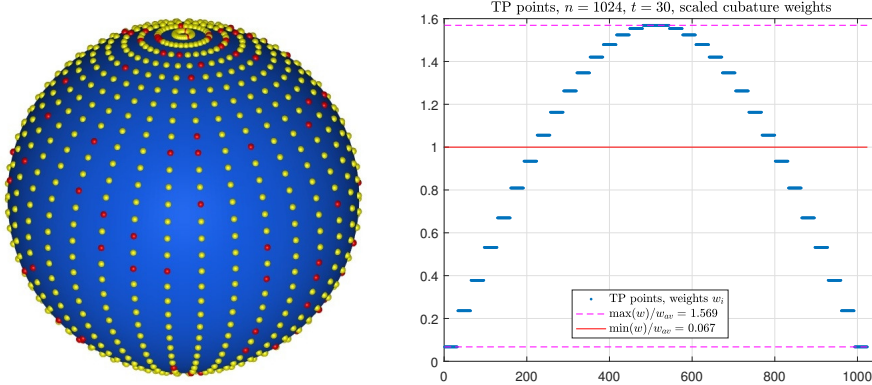
FIG. 3. *SF points and cubature weights*FIG. 4. *MD points and cubature weights*

package for disciplined convex programming based on interior point solvers. However for problem (7) the specialized package SPGL1 was much faster than the general purpose package. Also as CVX uses an interior point solver, it will not produce sparse solutions, so a tolerance must be used in counting the number of elements in the support of  $v$ .

For  $0 < q < 1$ , problem (40) is solved using the penalty method proposed in [13], whose subproblems are solved via a nonmonotone proximal gradient (NPG) method [47] with a suitable update scheme for the penalty parameters.

The penalty method in particular is sensitive to the choices of initial points. We use a variety of starting points, including the vector of all ones, the  $\ell_1$  solution produced by SPGL1, and the solution produced by the RWL1 process with  $\mu = 10^{-6}$ . RWL1-8 means that Steps 1-2 of Algorithm 4.1 were performed  $k^* = 8$  times. This is indicated immediately after the method name in Tables 2, 3, 4, 5. Other parameters are same as those in [13, 42].

We chose an index set  $J$  of size  $|J|$  at random. We generate a vector  $u \in \mathbb{R}^{|J|}$  with i.i.d. standard Gaussian entries, and define a vector  $u^* \in \mathbb{R}^n$  by setting  $u_j^* = u$  and  $u_{j^c}^* = 0$ . The measurement  $c$  is then set to be  $AWu^* + \delta\xi$ , so  $v^* = W^{\frac{1}{2}}u^*$  where  $W$  is the diagonal matrix of weights. The noise uses a  $\delta > 0$ , with each entry of  $\xi$  following the standard Gaussian distribution. This corresponds to an additive noise

FIG. 5. *GL points and cubature weights*FIG. 6. *TP points and cubature weights*

model for the Fourier coefficients. Finally, we set  $\sigma = \delta \|\xi\|_2$  so that the resulting feasible set will contain the sparse vector  $v^*$ .

Tables 2, 3, 4, 5 summarize the performance of different methods for recovery of a sparse signal  $v^*$ . The quantity  $\varphi(v) := \|Av - c\|_2^2 - \sigma^2$  satisfies  $\varphi(v) \leq 0$  for a feasible solution  $v$ . While  $\|\hat{v}\|_0$  is the number of non-zeros in the calculated solution,  $\|v^* \& \hat{v}\|_0$  is the number of non-zero values that  $v^*$  and  $\hat{v}$  have in common, so the number of “false positives” where  $\hat{v}_i$  is non-zero, but  $v_i^*$  is zero, is  $\|\hat{v}\|_0 - \|v^* \& \hat{v}\|_0$ . These tables show that  $\ell_q$ , RWL1-8 (Algorithm 4.1 with  $k^* = 8$ ) performs best, which solves the  $\ell_q$  problem (8) using the penalty method (40) with a solution from re-weighted  $\ell_1$  minimization algorithm (Steps 1-2 in Algorithm 4.1).

Figure 8 illustrates the performance of the different solution methods over 100 trials with noise  $\eta_{\ell,k} = \delta N(0, 1)$ . The behaviour on individual trials is consistent with the sample mean data in Table 2. From Figure 8 and Table 2, we can see that  $\ell_q$ , RWL1-8 performs best regarding the values of objective function at the calculated solutions.

In Figures 9, 10, 11 and 12 the first plot illustrates the function obtained from

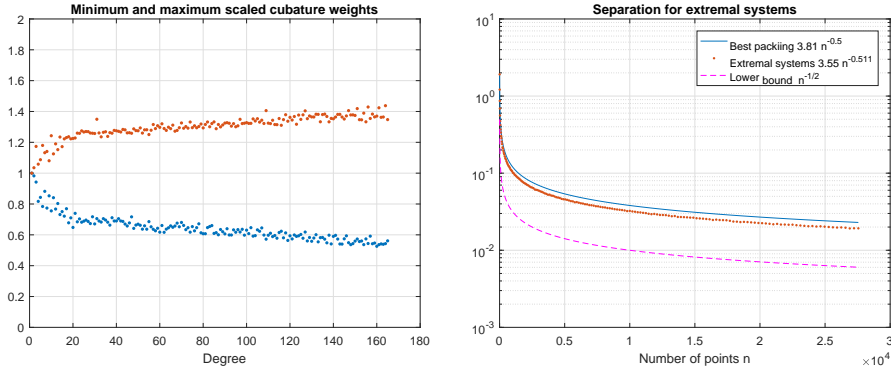


FIG. 7. Scaled cubature weights and separation for maximum determinant systems

Method	$\ \hat{v}\ _q^q$	$\varphi(\hat{v})$	$\ v^* - \hat{v}\ _2$	$\ \hat{v}\ _0$	$\ v^* \& \hat{v}\ _0$	false
SPGL1	99.688	-1.1e-08	3.21	217.4	101.7	115.7
$\ell_q$ , ones	94.715	2.1e-17	7.65	150.7	42.8	107.9
$\ell_q$ , SPGL1	92.315	-1.0e-17	6.91	143.9	50.3	93.6
RWL1-8	85.262	-1.2e-11	3.16	125.3	89.5	35.8
$\ell_q$ , RWL1-8	83.999	5.9e-19	3.19	119.9	87.7	<b>32.2</b>

TABLE 2

SF nodes:  $m = 256$ ,  $n = 1014$ ,  $q = 0.5$ ,  $|J^*| = 120$ ,  $\|v^*\|_q^q = 84.639$ , averages of 100 trials with  $\delta = 0.01$ ,  $\sigma = 0.1604$ .

the noisy Fourier coefficient  $c \in \mathbb{R}^m$  up to degree  $L$  by

$$f(\mathbf{x}_j) = \sum_{\ell=0}^L \sum_{k=1}^{2\ell+1} c_{\ell,k} Y_{\ell,k}(\mathbf{x}_j), \quad j = 1, \dots, n$$

or  $f = Yc$  using the degree  $L$  spherical harmonic basis matrix. The second plot in these figures is the true signal  $F$ . The third plot is the approximation

$$\hat{f} = W^{-\frac{1}{2}} \hat{v}$$

where  $\hat{v}$  is obtained using Algorithm 4.1 with  $k^* = 8$ , which is the penalty method starting from the solution after using 8 iterations of the iteratively re-weighted  $\ell_1$  algorithm. The last plot is the errors  $\hat{f} - F$  at the nodes of the cubature rule. The recovery of the underlying function values  $\hat{f}$  depends critically on the condition number of the diagonal scaling matrix  $W^{-\frac{1}{2}}$  coming from the cubature weights. As illustrated in Table 1,  $\text{cond}(W^{-\frac{1}{2}}) = 1$  for nodes based on spherical designs,  $\text{cond}(W^{-\frac{1}{2}}) \leq (1-\epsilon)^{-1}$  for nodes based on spherical  $t_\epsilon$ -designs. For nodes based on tensor product point sets  $\text{cond}(W^{-\frac{1}{2}})$  may be large, especially the TP nodes based on equally spaced points in  $\theta$  and  $\phi$ . This accounts for the significantly larger errors in Figure 12.

**6. Conclusion remarks.** Numerical integration and optimization methods are two important factors in signal processing. In this paper, we give comprehensive investigation on four spherical cubature rules and three optimization methods for recovery of sparse signals on the unit sphere  $\mathbb{S}^2$ . In theory and numerical computation, we find that using an equal weight cubature rule (spherical  $t$ -design) to build the

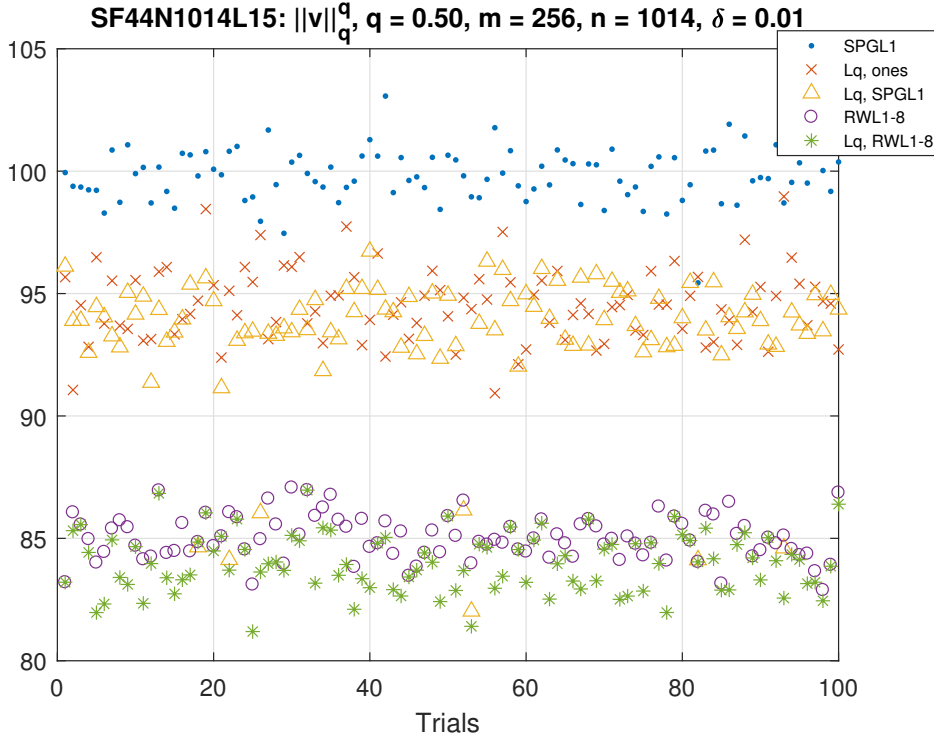


FIG. 8. Values of  $\|v\|_q^q$  at the calculated solutions using SF nodes:  $m = 256$ ,  $n = 1014$ ,  $q = 0.5$ ,  $|J^*| = 120$ , for 100 trials with  $\delta = 0.01$ . The values of  $\|v\|_q^q$  at the randomly generated sparse solution  $v^*$  is  $\|v^*\|_q^q = 84.639$ .

Method	$\ \hat{v}\ _q^q$	$\varphi(\hat{v})$	$\ v^* - \hat{v}\ _2$	$\ \hat{v}\ _0$	$\ v^* \& \hat{v}\ _0$	false
SPGL1	104.964	-7.1e-10	4.65	230.5	88.6	141.9
$\ell_q$ , ones	94.905	1.8e-17	7.59	151.6	45.6	106
$\ell_q$ , SPGL1	95.362	-8.7e-18	7.45	152.7	43.4	109.3
RWL1-8	90.355	-1.3e-11	4.71	140.6	77.8	62.8
$\ell_q$ , RWL1-8	88.779	1.2e-17	4.77	133.1	75.4	<b>57.7</b>

TABLE 3

MD nodes:  $m = 256$ ,  $n = 1024$ ,  $q = 0.5$ ,  $|J^*| = 120$ ,  $\|v^*\|_q^q = 86.101$ , averages of 100 trials with  $\delta = 0.01$ ,  $\sigma = 0.1604$ .

optimization model (8) and then using Algorithm 4.1 to solve (8) can find better sparse solutions than other cubature rules and other optimization methods for the  $\ell_1$  optimization model (7) and the  $\ell_q$  optimization model (8).

**Acknowledgments.** The authors would like to thank the referees and associate editor for their helpful comments and constructive suggestions. This work was carried out while the second author was visiting the Department of Applied Mathematics, The Hong Kong Polytechnic University. Their hospitality is gratefully acknowledged.

Method	$\ \hat{v}\ _q^q$	$\varphi(\hat{v})$	$\ v^* - \hat{v}\ _2$	$\ \hat{v}\ _0$	$\ v^* \& \hat{v}\ _0$	false
SPGL1	92.310	-1.2e-07	3.15	213.8	87.3	126.5
$\ell_q$ , ones	85.952	5.0e-18	6.76	134.3	48.9	85.4
$\ell_q$ , SPGL1	79.618	1.2e-17	5.16	114.2	62.8	51.4
RWL1-8	74.911	-1.4e-11	3.86	99.4	75.3	24.1
$\ell_q$ , RWL1-8	74.757	4.6e-18	3.87	98.2	75.1	<b>23.1</b>

TABLE 4

GL nodes:  $m = 256$ ,  $n = 1058$ ,  $q = 0.5$ ,  $|J^*| = 120$ ,  $\|v^*\|_q^q = 82.308$ , averages of 100 trials with  $\delta = 0.01$ ,  $\sigma = 0.1604$ .

Method	$\ \hat{v}\ _q^q$	$\varphi(\hat{v})$	$\ v^* - \hat{v}\ _2$	$\ \hat{v}\ _0$	$\ v^* \& \hat{v}\ _0$	false
SPGL1	92.275	-1.7e-07	3.57	206.1	86.6	119.5
$\ell_q$ , ones	84.746	1.5e-18	7.04	129.5	48.1	81.4
$\ell_q$ , SPGL1	82.863	2.4e-19	6.72	124.4	51.9	72.5
RWL1-8	76.139	-1.2e-11	4.16	101.7	73.8	27.9
$\ell_q$ , RWL1-8	75.741	6.9e-20	4.20	100.0	73.3	<b>26.7</b>

TABLE 5

TP nodes:  $m = 256$ ,  $n = 1024$ ,  $q = 0.5$ ,  $|J^*| = 120$ ,  $\|v^*\|_q^q = 82.340$ , averages of 100 trials with  $\delta = 0.01$ ,  $\sigma = 0.1604$ .

## REFERENCES

- [1] 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), pp. 2135–2157.
- [2] T. BENDORY, S. DEKEL, AND A. FEUER, *Exact recovery of Dirac ensembles from the projection onto spaces of spherical harmonics*, Constr. Approx., 42 (2015), pp. 183–207.
- [3] T. BENDORY, S. DEKEL, AND A. FEUER, *Super-resolution on the sphere using convex optimization*, IEEE Trans. Signal Process., 63 (2015), pp. 2253–2262.
- [4] A. BONDARENKO, D. RADCHENKO, AND M. VIAZOVSKA, *Optimal asymptotic bounds for spherical designs*, Ann. Math., 178 (2013), pp. 443–452.
- [5] A. BONDARENKO, D. RADCHENKO, AND M. VIAZOVSKA, *Well-separated spherical designs*, Constr. Approx., 41 (2015), pp. 93–112.
- [6] A. M. BRUCKSTEIN, D. L. DONOHO, AND M. ELAD, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Rev., 51 (2009), pp. 34–81.
- [7] T. BÜLOW, *Spherical diffusion for 3D surface smoothing*, IEEE Trans. Pattern Anal. Machine Intell., 26 (2004), pp. 1650–1654.
- [8] E. J. CANDÈS, J. K. ROMBERG, AND T. TAO, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.
- [9] E. J. CANDÈS AND T. TAO, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.
- [10] E. J. CANDÈS, M. B. WAKIN, AND S. P. BOYD, *Enhancing sparsity by reweighted  $\ell_1$  minimization*, J. Fourier Anal. Appl., 14 (2008), pp. 877–905.
- [11] X. CHEN, A. FROMMER, AND B. LANG, *Computational existence proofs for spherical  $t$ -designs*, Numer. Math., 117 (2011), pp. 289–305.
- [12] X. CHEN, D. GE, Z. WANG, AND Y. YE, *Complexity of unconstrained  $L_2$ - $L_p$  minimization*, Math. Program., 143 (2014), pp. 371–383.
- [13] X. CHEN, Z. LU, AND T. K. PONG, *Penalty methods for a class of non-Lipschitz optimization problems*, SIAM J. Optim., 26 (2016), pp. 1465–1492.
- [14] 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), pp. 2832–2852.
- [15] X. CHEN AND W. ZHOU, *Convergence of the reweighted  $\ell_1$  minimization algorithm for  $\ell_2$ - $\ell_p$  minimization*, Comput. Optim. Appl., 59 (2014), pp. 47–61.
- [16] Y. CHI, L. SCHARF, A. PEZESHKI, AND A. CALDERBANK, *Sensitivity to basis mismatch in compressed sensing*, IEEE Trans. Signal Process., 59 (2011), pp. 2182–2195.
- [17] J. H. CONWAY AND N. J. A. SLOANE, *Sphere Packings, Lattices and Groups*, Springer-Verlag,

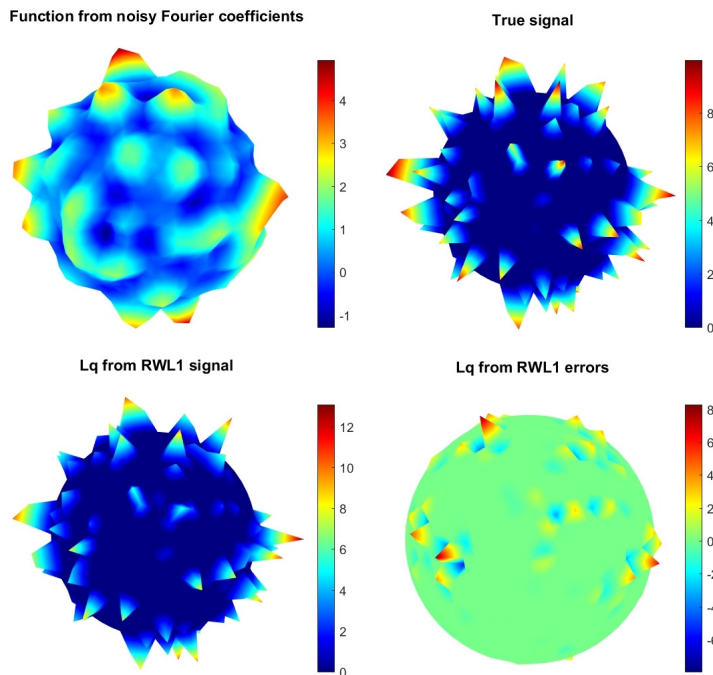


FIG. 9. *SF nodes: function from noisy Fourier coefficients, true signal, signal minimizing  $\|v\|_q^q$  from solution of Algorithm 4.1 with  $k^* = 8$ , and error*

- New York, third ed., 1999.
- [18] I. DAUBECHIES, R. DeVORE, M. FORNASIER, AND C. S. GÜNTÜRK, *Iteratively reweighted least squares minimization for sparse recovery*, *Comm. Pure Appl. Math.*, 63 (2010), pp. 1–38.
- [19] M. E. DAVIES AND R. GRIBONVAL, *Restricted isometry constants where  $\ell^p$  sparse recovery can fail for  $0 < p \leq 1$* , *IEEE Trans. Inform. Theory*, 55 (2009), pp. 2203–2214.
- [20] P. DELSARTE, J. M. GOETHALS, AND J. J. SEIDEL, *Spherical codes and designs*, *Geometriae Dedicata*, 6 (1977), pp. 363–388.
- [21] I. DOKMANIĆ AND Y. LU, *Sampling sparse signals on the sphere: algorithms and applications*, *IEEE Trans. Signal Process.*, 64 (2016), pp. 189–202.
- [22] D. L. DONOHO AND M. ELAD, *Optimally sparse representation in general (nonorthogonal) dictionaries via  $l^1$  minimization*, *Proc. Natl. Acad. Sci. USA*, 100 (2003), pp. 2197–2202.
- [23] D. L. DONOHO AND X. HUO, *Uncertainty principles and ideal atomic decomposition*, *IEEE Trans. Inform. Theory*, 47 (2001), pp. 2845–2862.
- [24] S. FOUCART AND M.-J. LAI, *Sparsest solutions of underdetermined linear systems via  $l_q$ -minimization for  $0 < q \leq 1$* , *Appl. Comput. Harmon. Anal.*, 26 (2009), pp. 395–407.
- [25] S. FOUCART AND H. RAUHUT, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser/Springer, New York, 2013.
- [26] D. GE, X. JIANG, AND Y. YE, *A note on the complexity of  $L_p$  minimization*, *Math. Program.*, 129 (2011), pp. 285–299.
- [27] K. M. GÓRSKI, E. HIVON, A. J. BANDAY, B. D. WANDELT, F. K. HANSEN, M. REINECKE, AND M. BARTELMANN, *HEALPix: A framework for high resolution discretisation and fast analysis of data distributed on the sphere*, *The Astrophys. J.*, 622 (2005), pp. 759–771.
- [28] M. GRÄF, *Efficient Algorithms for the Computation of Optimal Quadrature Points on Riemannian Manifolds*, Dissertation, Universitätsverlag, 2013.
- [29] M. GRÄF, *Quadrature rules on manifolds*. <http://www.tu-chemnitz.de/~potts/workgroup/graef/quadrature/>, 2013.
- [30] M. GRÄF AND D. POTTS, *On the computation of spherical designs by a new optimization approach based on fast spherical fourier transforms*, *Numer. Math.*, 119 (2011), pp. 699–724.
- [31] M. GRANT AND S. BOYD, *CVX: Matlab software for disciplined convex programming, version*

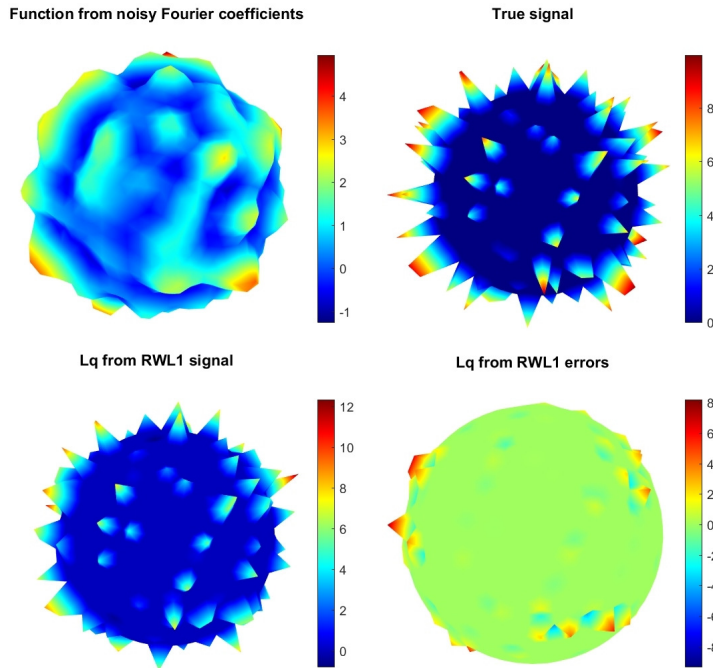


FIG. 10. MD nodes: function from noisy Fourier coefficients, true signal, signal minimizing  $\|v\|_q^q$  from solution of Algorithm 4.1 with  $k^* = 8$ , and error

2.1. <http://cvxr.com/cvx>, Mar. 2014.

- [32] R. GRIBONVAL AND M. NIELSEN, *Highly sparse representations from dictionaries are unique and independent of the sparseness measure*, Appl. Comput. Harmon. Anal., 22 (2007), pp. 335–355.
- [33] S. KUNIS, H. MÖLLER, AND U. VON DER OHE, *Prony’s method on the sphere*. arXiv:160302020, 2016.
- [34] P. LIU, Y. WANG, D. HUANG, Z. ZHANG, AND L. CHEN, *Learning the spherical harmonic features for 3-D face recognition*, IEEE Trans. Image Processing, 22 (2013), pp. 914–925.
- [35] R. LONCH, E. KOKIOPOULOU, I. TOSIĆ, AND P. FROSSARD, *3D face recognition with sparse spherical representations*, Pattern Recogn., 43 (2010), pp. 824–834.
- [36] M. NIKOLOVA, *Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares*, Multiscale Model. Simul., 44 (2005), pp. 960–991.
- [37] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
- [38] E. B. SAFF AND A. B. J. KUIJLAARS, *Distributing many points on a sphere*, Math. Intelligencer, 19 (1997), pp. 5–11.
- [39] I. H. SLOAN AND R. S. WOMERSLEY, *Extremal systems of points and numerical integration on the sphere*, Adv. Comput. Math., 21 (2004), pp. 107–125.
- [40] A. H. STROUD, *Approximate calculation of multiple integrals*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971. Prentice-Hall Series in Automatic Computation.
- [41] E. VAN DEN BERG AND M. P. FRIEDLANDER, *SPGL1: A solver for large-scale sparse reconstruction*, June 2007. <http://www.cs.ubc.ca/labs/scl/spgl1>.
- [42] E. VAN DEN BERG AND M. P. FRIEDLANDER, *Probing the Pareto frontier for basis pursuit solutions*, SIAM J. Sci. Comput., 31 (2008), pp. 890–912.
- [43] E. VAN DEN BERG AND M. P. FRIEDLANDER, *Sparse optimization with least-squares constraints*, SIAM J. Optim., 21 (2011), pp. 1201–1229.
- [44] M. WANG, W. XU, AND A. TANG, *On the performance of sparse recovery via  $\ell_p$ -minimization ( $0 \leq p \leq 1$ )*, IEEE Trans. Inform. Theory, 57 (2011), pp. 7255–7278.
- [45] R. S. WOMERSLEY, *Efficient spherical designs with good geometric properties*. <http://web.maths.unsw.edu.au/~rsw/Sphere/EffSphDes/>, 2015.
- [46] R. S. WOMERSLEY, *Efficient spherical designs with good geometric properties*, in Festschrift for

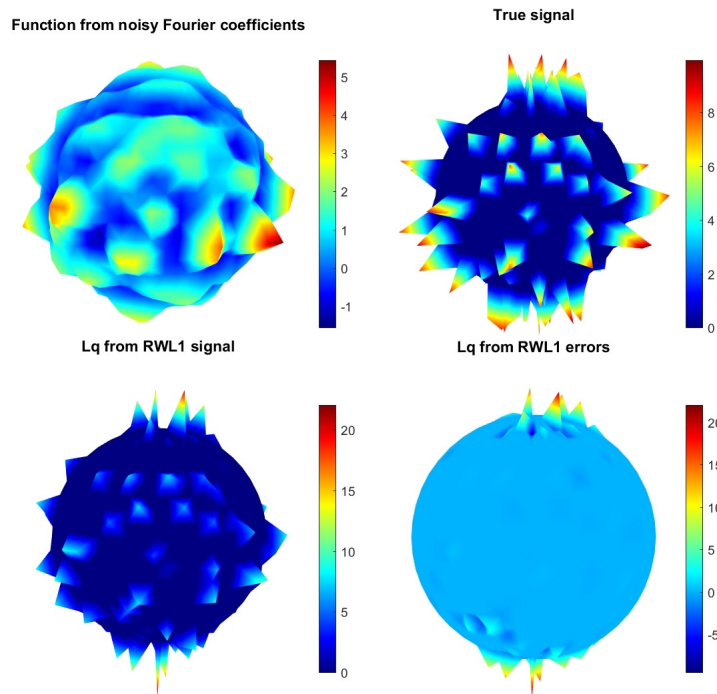


FIG. 11. *GL nodes: function from noisy Fourier coefficients, true signal, signal minimizing  $\|v\|_q^q$  from solution of Algorithm 4.1 with  $k^* = 8$ , and error*

the 80th Birthday of Ian Sloan, J. Dick, F. Y. Kuo, and H. Wozniakowski, eds., Springer, 2018.

- [47] S. J. WRIGHT, R. D. NOWAK, AND M. A. T. FIGUEIREDO, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493.
- [48] Y. ZHANG, *Theory of compressive sensing via  $\ell_1$  minimization: a non-RIP analysis and extensions*, J. Oper. Res. Soc. China, 1 (2013), pp. 79–105.
- [49] Y.-B. ZHAO AND Z.-Q. LUO, *Constructing new weighted  $\ell_1$ -algorithms for the sparsest points of polyhedral sets*, Math. Oper. Res., 42 (2017), pp. 57–76.
- [50] Y. ZHOU AND X. CHEN, *Spherical  $t_\epsilon$ -designs for applications on the sphere*, Math. Comp., (to appear).
- [51] J. ZHU, X. LI, F. ARROYO, AND E. ARROYO, *Error analysis of reweighted  $\ell_1$  greedy algorithm for noisy reconstruction*, J. Comput. Appl. Math., 286 (2015), pp. 93–101.



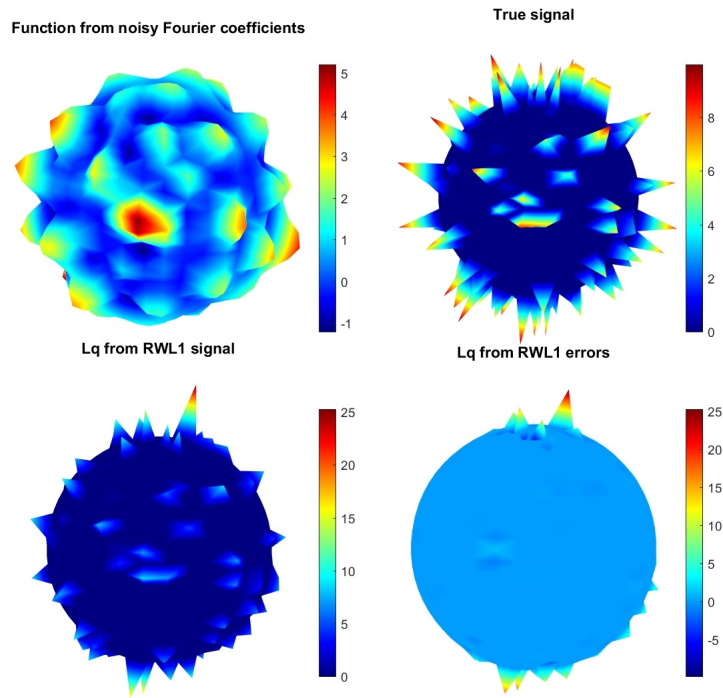


FIG. 12. TP nodes: function from noisy Fourier coefficients, true signal, signal minimizing  $\|v\|_q^q$  from solution of Algorithm 4.1 with  $k^* = 8$ , and error