SIAM J. IMAGING SCIENCES
Vol. 11, No. 2, pp. 1390–1415

Spherical Designs and Nonconvex Minimization for Recovery of Sparse Signals on the Sphere*

Xiaojun Chen† and Robert S. Womersley‡

Abstract. This paper considers the use of spherical designs and nonconvex 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 ℓ_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 ℓ_1-norm by the ℓ_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 reweighted ℓ_1 method is promising to solve the ℓ_q basis pursuit denoise model. Numerical performance on nodes using spherical t-designs and t_ϵ-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, nonconvex minimization, spherical cubature, reweighted ℓ_1

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

DOI. 10.1137/17M1147378

1. Introduction.

Recovery of sparse signals on the unit sphere

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

has many applications in astrophysics, gravitational sensing, geophysics, climate modeling, and 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 two dimensional face recognition since it is more accurate and natural to describe and analyze human face data on the sphere [7, 34, 38].

*Received by the editors September 13, 2017; accepted for publication (in revised form) April 3, 2018; published electronically May 24, 2018.
http://www.siam.org/journals/siims/11-2/M114737.html
Funding: This work was supported in part by Hong Kong Research Council grant PolyU153016/16p.
†Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (maxjchen@polyu.edu.hk, http://www.polyu.edu.hk/ama/staff/xjchen/ChenXJ.htm).
‡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/).

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 ℓ 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 $F : \mathbb{S}^2 \to \mathbb{R}$ is observed via the Fourier coefficients

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

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

$$\mathcal{X}_n := \{\boldsymbol{x}_j \in \mathbb{S}^2, j = 1,\dots,n\}, \tag{2}$$

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

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

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

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

In matrix notation, (4) is

$$c = YWf + \eta, \tag{5}$$

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, $W = \operatorname{diag}(w) \in \mathbb{R}^{n\times n}$ is the diagonal matrix of cubature weights, $\eta \in \mathbb{R}^m$ is the noise vector, and $f = (F(\boldsymbol{x}_1),\dots,F(\boldsymbol{x}_n))^T \in \mathbb{R}^n$ 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_ϵ-designs [50] from extremal fundamental systems [39], a tensor product that is equally spaced in spherical coordinates θ, ϕ, HealPix points [27], and a tensor product that is Gauss–Legendre in $\cos(\theta)$ and equally spaced in ϕ [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

$$AA^T = YWY^T = I \in \mathbb{R}^{m\times m}, \tag{6}$$

and 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 $\boldsymbol{x}_j$ for $j = 1, \ldots, n$. In addition, the signal is sparse and the weights are positive, so that its support set

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

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 denoising when $\sigma > 0$:

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

This has become a classical problem building on the work of Donoho and Huo [23] and Candès and Tao [9], using sufficient conditions on A to guarantee that minimizing the ℓ_1-norm, a convex problem, gives a solution to the ℓ_0-norm minimization problem. See, for example, [6, 25]. Such sufficient conditions guarantee that the ℓ_0-norm solution is also the unique minimizer for a class of sparsity measures [32], including

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

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

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

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

$$A = \tfrac{1}{\sqrt{2}}[1 \quad 1], \quad c = \tfrac{1}{\sqrt{2}}, \quad 0 \le \sigma < \tfrac{1}{\sqrt{2}}.$$

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

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

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

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

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, 35, 44]) but at the expense of making the minimization problem nonconvex. We show that to solve (8), the penalty approach in Chen, Lu, and Pong [13] with a starting point obtained from the reweighted ℓ_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 ℓ_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 starting point from reweighted ℓ_1 to solve (8) can result in better sparse solutions than using other cubature rules and the ℓ_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 σ. 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 reweighted ℓ_1 method and its convergence. In section 5, we present numerical results to compare cubature rules using spherical t-designs and t_ϵ-designs (extremal fundamental systems) with tensor product rules. We also compare solutions of problem (7) with solutions of 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 $\boldsymbol{x}_j \in \mathbb{S}^2$ and weights $w_j > 0, j = 1, \ldots, n$, such that

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

where $g : \mathbb{S}^2 \to \mathbb{R}$ and $\mathrm{d}\mu(\boldsymbol{x})$ 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 the degree of precision is at least 0) rule has

$$w_j = w_{av} := \frac{|S^2|}{n} = \frac{4\pi}{n}, \quad j = 1, \ldots, n. \tag{11}$$

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

$$Y_{ij} = Y_{\ell,k}(\boldsymbol{x}_j), \qquad i = \ell^2 + k,\ k = 1, \ldots, 2\ell+1,\ \ell = 0, \ldots, t, \qquad j = 1, \ldots, n. \tag{12}$$

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

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

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

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

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

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

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, albeit 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(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 $\mathcal{X}_n = \{\boldsymbol{x}_j, j = 1, \ldots, n\}$ on $\mathbb{S}^2$ such that equal weight cubature has degree of precision t, that is,

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

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

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

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 $\mathcal{X}_n = \{\boldsymbol{x}_j, j = 1, \ldots, n\}$ 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_ϵ-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,

$$w_{av}(1-\epsilon) \leq w_i \leq w_{av}(1-\epsilon)^{-1}, \quad j = 1, \ldots, n, \qquad \epsilon \in [0, 1).$$

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

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

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

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

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

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 $j = 1, \ldots, n_z$ 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*

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

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

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

is related to the geodesic distance by

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

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 $\mathcal{X}_n = \{\boldsymbol{x}_j, j = 1, \ldots, n\}$ is its separation $\delta(\mathcal{X}_n)$ defined by

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

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

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

The problem of choosing n points on $\mathbb{S}^2$ to maximize their separation (see [37, 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 δ_n as a function of the number of points n for two classes of point sets on $\mathbb{S}^2$. The first class, whose point sets are well separated as δ_n has the optimal order $n^{-1/2}$, includes the MD and

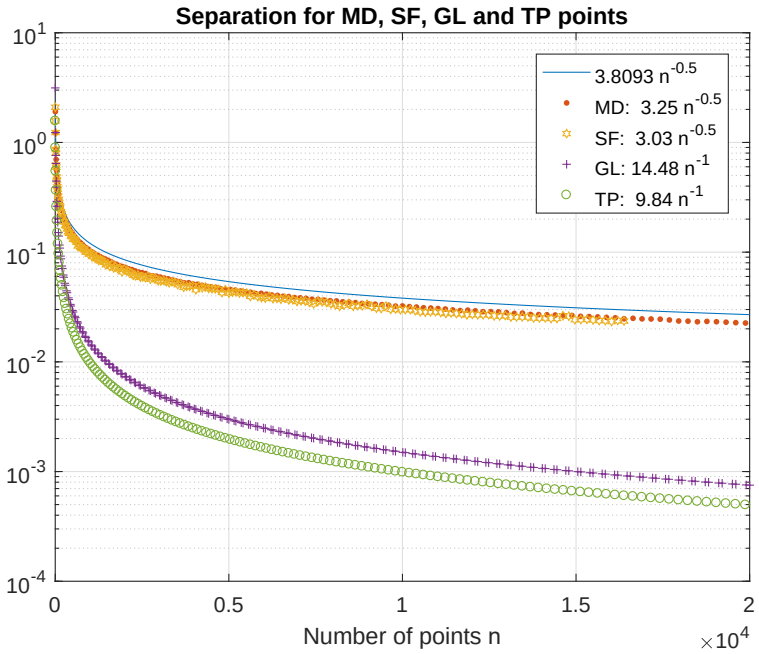

Figure 1. *Separation distance of two classes of point sets on* $\mathbb{S}^2$.

SF points. The second class is comprised of 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, while keeping separation.

If the cubature rule with nodes $\mathcal{X}_n$ and weights $w_j, j = 1, \ldots, n$, 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

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

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 $i = 1, \ldots, (L+1)^2$, 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 semidefinite with singular values $\sigma_i(G) = (w_{av})^{-1}$ for $i = 1, \ldots, m$ and $\sigma_i = 0$

for $i = m+1, \ldots, n$. 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 is 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 ℓ_1-norm minimization problem will give the solution to the ℓ_0-norm problem fail. For example, sufficient conditions[1] on spark(A), mutual incoherence $M(A)$, the restricted isometry property (RIP), and the null space of A ensure that

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

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

Denote the feasible set of problem (8) by

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

Assumption 3.1. *Ω 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 $J \subset \{1, 2, \ldots, m\}$, consider the convex optimization problem

$$\begin{array}{ll} \min & \|v\|_2^2 \\ \text{s.t.} & \|Av - c\|_2 \le \sigma, \quad v_{J^c} = 0, \end{array} \tag{21}$$

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

If J is the support set of a true signal, we call v^{o} 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 $\|v^{\mathrm{o}} - \hat{v}\|_q^q$ for $\hat{v} \in S_q$.

Let

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

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

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

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

[1]
- Spark: Donoho, Elad [22], spark$(A) = \min\{\|v\|_0 : Av = 0, v \ne 0\}$.
- Mutual incoherence: Donoho, Huo [23], $M(A) = \max_{i \ne j} |A_i^T A_j|/(\|A_i\|_2 \|A_j\|_2)$.
- RIP: Candes, Romberg, Tao [8, 25], Smallest δ_s s.t.
$$(1 - \delta_s)\|v\|_2^2 \le \|Av\|_2^2 \le (1 + \delta_s)\|v\|_2^2 \quad \forall v : \|v\|_0 \le s.$$
- Null space: Zhang [48], $(\|v\|_0)^{\frac{1}{2}} < \frac{1}{2} \min\{\|v\|_1/\|v\|_2 : Av = 0, v \ne 0\}$.

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

$$\|r_J\|_q \le \gamma \|r_{J^c}\|_q \tag{23}$$

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

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

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

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

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

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

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

from Lemma 2.4 in [13] we have

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

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

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

Using Hölder's inequality[2] and the definition of β, we have

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

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

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

and

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

This completes the proof. ∎

[2] $\|v\|_q^q = \sum_{i=1}^n |v_i|^q \cdot 1 \le (\sum_{i=1}^n (|v_i|^q)^{\frac{2}{q}})^{\frac{q}{2}} (\sum_{i=1}^n 1^{\frac{2}{2-q}})^{\frac{2-q}{2}} = n^{\frac{2-q}{2}} \|v\|_2^q$.

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:

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

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*

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

Moreover, we have

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

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

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

This completes the proof. ∎

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

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

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

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

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 $\|\hat{v}\|_q \leq \|v^{\rm o}\|_q$ 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 $v^{\rm o} = (1-\sqrt{2}\sigma, 0)$. For $q \in (0,1)$, S_q as given in (10) has only two minimizers, while for $q = 1$, S_1 as given in (9) has infinitely many minimizers. Moreover, $v^{\rm o} \in S_q$ for $0 < q < 1$. However, for $q = 1$,

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

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 [36, Definition 8.3] and [13, Lemma 2.5], the (limiting) subdifferential and horizon subdifferential of $\varphi(v)$ can be given, respectively, as

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

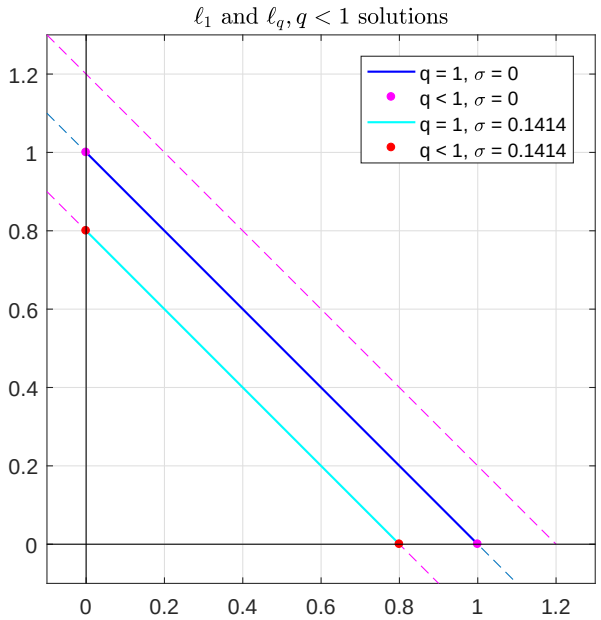

Figure 2. *Example 1.1 with multiple solutions for $q = 1$ and $q \in (0,1)$: $S_1 = \{(v_1, 1-\sqrt{2}\sigma-v_1)^T, v_1 \in [0, 1-\sqrt{2}\sigma]\}$ containing a segment between $(0, 1-\sqrt{2}\sigma)^T$ and $(1-\sqrt{2}\sigma, 0)^T$, and $S_q = \{(0, 1-\sqrt{2}\sigma)^T, (1-\sqrt{2}\sigma, 0)^T\}$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

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

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

$$\partial |v_i|^q = \partial^\infty |v_i|^q = (-\infty, \infty) \qquad \text{if } v_i = 0.$$

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

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

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

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

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

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

and the constraint qualification (32) can be written as

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

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

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

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

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

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

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

Proof. Since $Q = A^TA$ and $QQ = Q$, $I - Q$ is a symmetric positive semidefinite matrix and $(I-Q)(I-Q) = (I-Q)$. This implies

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

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

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

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

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

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:

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

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

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

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 η, 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). ∎

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_μ be global minimizers of problems (8) and (35), respectively. Then v^* and v_μ 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 $\mu \in (0,\bar{\mu})$,*

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

(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*

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

and for $i \in J$,

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

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_μ be global minimizers of problems (8) and (35), respectively. Then from

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

and

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

we find that v^* is an $n\mu^q$-global minimizer of (35) and v_μ is an $n\mu^q$-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 a neighborhood $N_{\hat v} = \{v : \ \|v - \hat v\|_2 \le \eta\}$ of $\hat v$ for some $\eta \in (0, \min_{\hat v_i \ne 0} |\hat v_i|)$ such that for any $v \in N_{\hat v} \cap \Omega$, $v \ne \hat v$, we have $\|\hat v\|_q^q < \|v\|_q^q$. In particular, 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 $\|(I - Q)\omega^*(\hat v, \mu)\|_2 \ne 0$. Let

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

Then $v^+ \in N_{\hat v} \cap \Omega$, $v_i^+ \hat v_i \ge 0$, 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

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

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

(iii) This part is from Theorem 3.2 in [13]. ∎

Van den Berg and Friedlander developed an efficient package SPGL1 [41, 42, 43] (version 1.9) for solving the ℓ_1 basis pursuit denoising problem (7). Chen, Lu, and Pong developed a penalty method [13] for solving the ℓ_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*

$$v^k \in \operatorname{Argmin} \sum_{i=1}^n \alpha_i^k |v_i| \quad \text{s.t.} \quad \|Av - c\|_2 \le \sigma. \tag{42}$$

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 reweighted ℓ_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 as follows:

`SF` Spherical designs [20, 45, 46].

`MD` Spherical t_ϵ-designs [50] from extremal fundamental systems [39].

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

`TP` Tensor product, equally spaced in spherical coordinates θ, ϕ.

The three methods are as follows:

`SPGL1` Solve the ℓ_1 basis pursuit denoising problem (7) [41, 42, 43].

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

`penalty` (ℓ_q, ones), (ℓ_q, SPGL1), and (ℓ_q, RWL1-8) solve the ℓ_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.

Table 1
Characteristics of node sets.

Code	n	t	$\delta(\mathcal{X}_n)$	$\delta(\mathcal{X}_n)n^{\frac{1}{2}}$	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

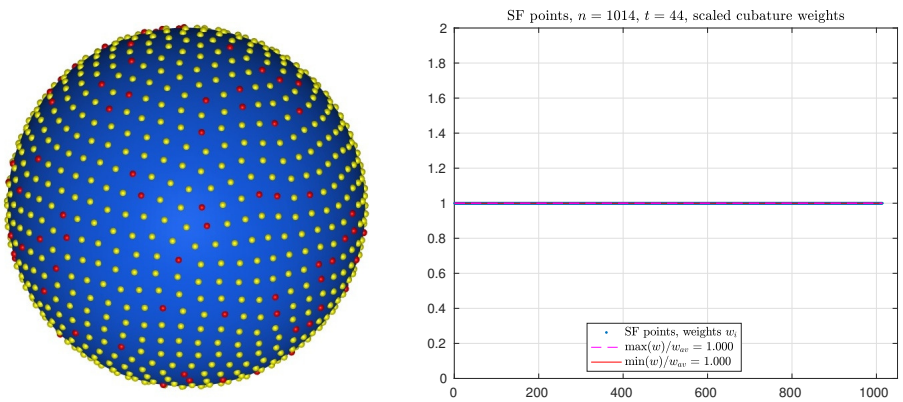

Figure 3. *SF points and cubature weights.*

In Figures 3–6, we show the distribution of these nodes on the sphere and the corresponding

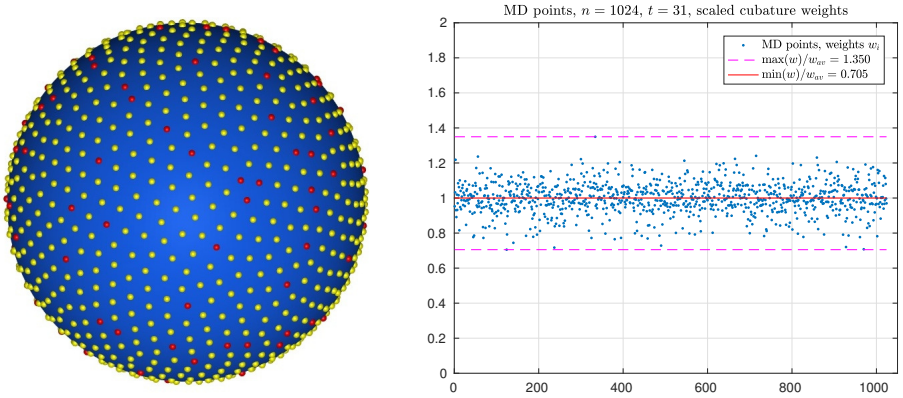

Figure 4. *MD points and cubature weights.*

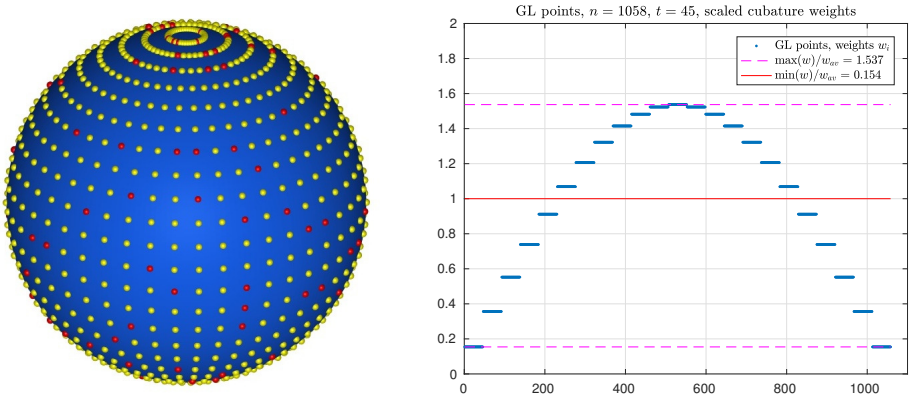

Figure 5. *GL points and cubature weights.*

cubature weights.

Examples of spherical t_ϵ-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 ℓ_1 basis pursuit denoising problem (7) is a convex problem and could be solved by a variety of other methods, including CVX [31], a 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.

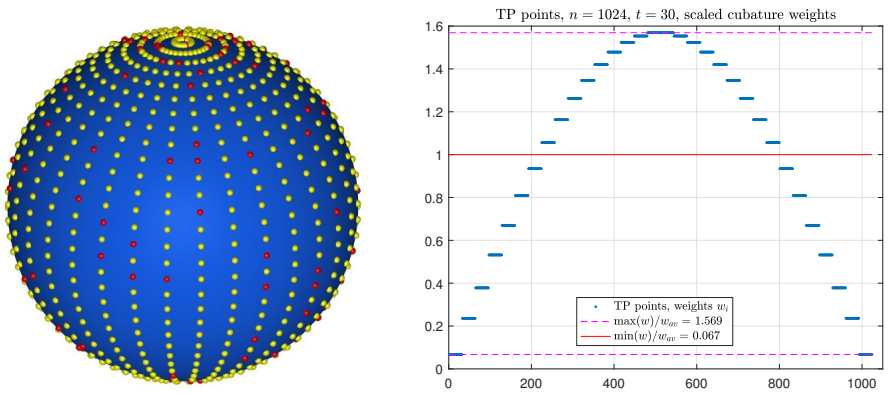

Figure 6. *TP points and cubature weights.*

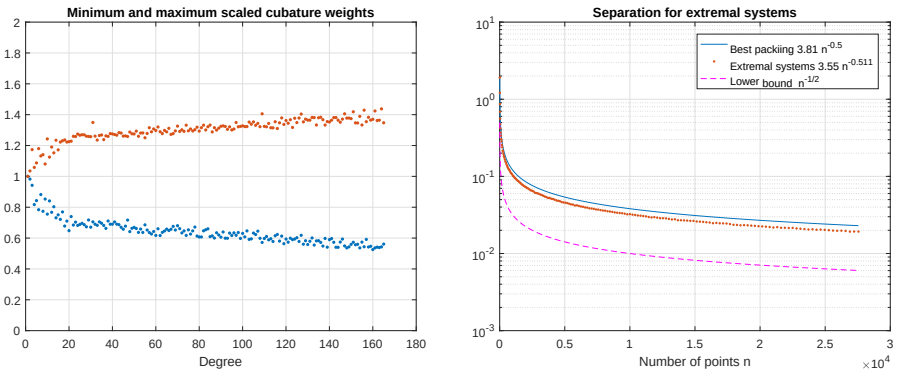

Figure 7. *Scaled cubature weights and separation for maximum determinant systems.*

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 ℓ_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, and 5. Other parameters are the same as those in [13, 42].

We choose 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 ξ following the standard Gaussian distribution. This corresponds to an additive noise 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, and 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 $\varphi(v) \le 0$ for a feasible solution v. While $\|\widehat{v}\|_0$ is the number of nonzeros in the calculated solution, $\|v^* \,\&\, \widehat{v}\|_0$ is the number of nonzero values that v^* and $\widehat{v}$ have in common, so the number of "false positives" where $\widehat{v}_i$ is nonzero, but v^*_i is zero, is $\|\widehat{v}\|_0 - \|v^* \,\&\, \widehat{v}\|_0$. These tables show that ℓ_q, RWL1-8 (Algorithm 4.1 with $k^* = 8$) performs best, which solves the ℓ_q problem (8) using the penalty method

Table 2

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

Method	$\|\widehat{v}\|_q^q$	$\varphi(\widehat{v})$	$\|\|v^* - \widehat{v}\|\|_2$	$\|\widehat{v}\|_0$	$\|v^* \,\&\, \widehat{v}\|_0$	False
SPGL1	99.688	-1.1e-08	3.21	217.4	101.7	115.7
ℓ_q, ones	94.715	2.1e-17	7.65	150.7	42.8	107.9
ℓ_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
ℓ_q, RWL1-8	83.999	5.9e-19	3.19	119.9	87.7	**32.2**

(40) with a solution from the reweighted ℓ_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 behavior on individual trials is consistent with the sample mean data in Table 2. From Figure 8 and Table 2, we can see that ℓ_q, RWL1-8 performs best regarding the values of the objective function at the calculated solutions.

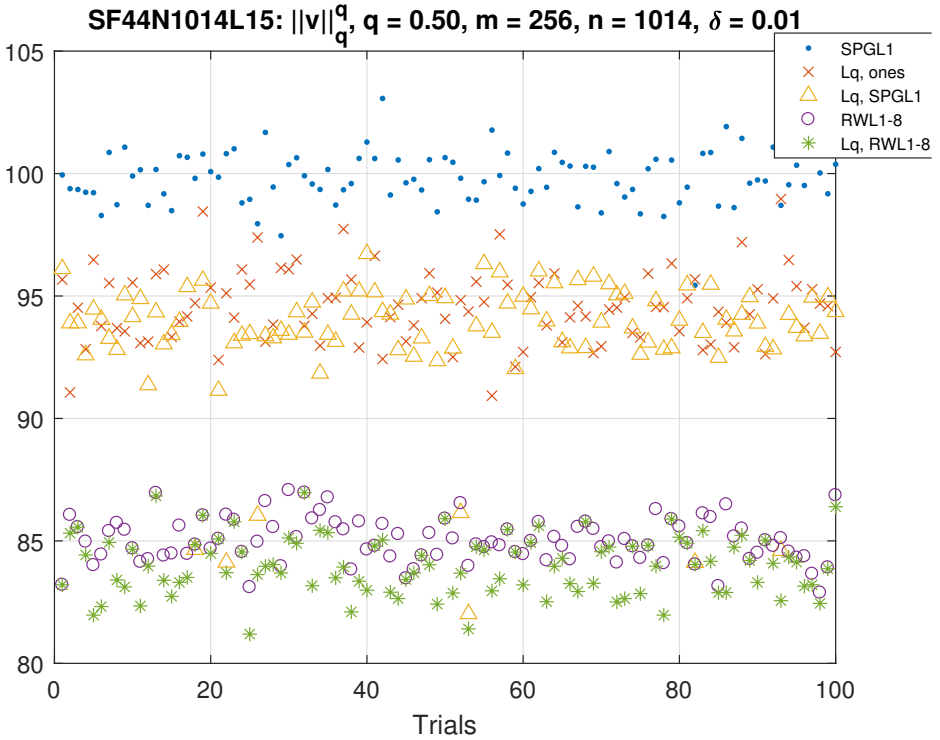

Figure 8. *Values of* $\|v\|_q^q$ *at the calculated solutions using SF nodes:* $m = 256$, $n = 1014$, $q = 0.5$, $|J^*| = 120$ *for* 100 *trials with* $\delta = 0.01$. *The value of* $\|v\|_q^q$ *at the randomly generated sparse solution* v^* *is* $\|v^*\|_q^q = 84.639$.

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

Table 3

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

Method	$\|\widehat{v}\|_q^q$	$\varphi(\widehat{v})$	$\|\|v^* - \widehat{v}\|\|_2$	$\|\widehat{v}\|_0$	$\|v^* \,\&\, \widehat{v}\|_0$	False
SPGL1	104.964	-7.1e-10	4.65	230.5	88.6	141.9
ℓ_q, ones	94.905	1.8e-17	7.59	151.6	45.6	106
ℓ_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
ℓ_q, RWL1-8	88.779	1.2e-17	4.77	133.1	75.4	**57.7**

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	$\|\widehat{v}\|_q^q$	$\varphi(\widehat{v})$	$\|\|v^* - \widehat{v}\|\|_2$	$\|\widehat{v}\|_0$	$\|v^* \,\&\, \widehat{v}\|_0$	False
SPGL1	92.310	-1.2e-07	3.15	213.8	87.3	126.5
ℓ_q, ones	85.952	5.0e-18	6.76	134.3	48.9	85.4
ℓ_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
ℓ_q, RWL1-8	74.757	4.6e-18	3.87	98.2	75.1	**23.1**

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

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

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 reweighted ℓ_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, and $\text{cond}(W^{-\frac{1}{2}}) \leq (1-\epsilon)^{-1}$ for nodes based on spherical t_ϵ-designs. For nodes based on tensor product point sets, $\text{cond}(W^{-\frac{1}{2}})$ may be large, especially for the TP nodes based on equally spaced points in θ and ϕ. This accounts for the significantly larger errors in Figure 12.

Table 5

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

Method	$\|\widehat{v}\|_q^q$	$\varphi(\widehat{v})$	$\|\|v^* - \widehat{v}\|\|_2$	$\|\widehat{v}\|_0$	$\|v^* \,\&\, \widehat{v}\|_0$	False
SPGL1	92.275	-1.7e-07	3.57	206.1	86.6	119.5
ℓ_q, ones	84.746	1.5e-18	7.04	129.5	48.1	81.4
ℓ_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
ℓ_q, RWL1-8	75.741	6.9e-20	4.20	100.0	73.3	**26.7**

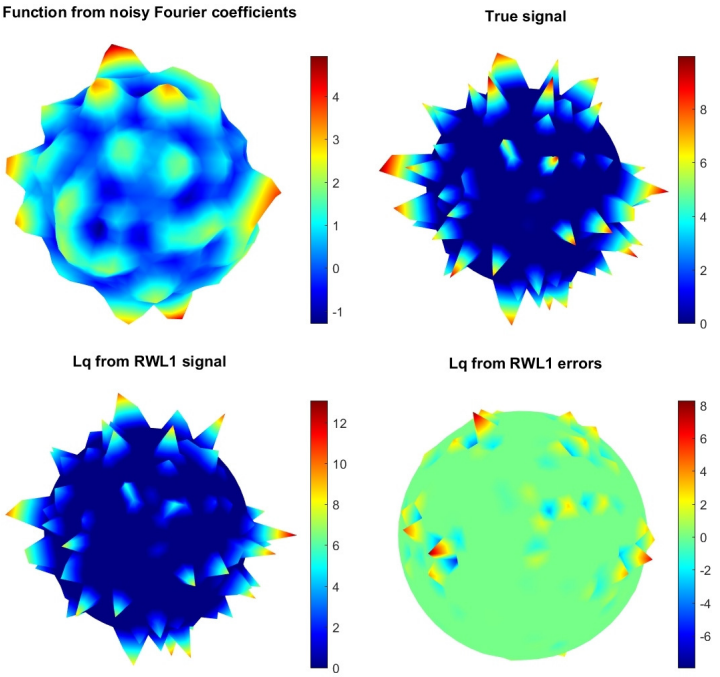

Figure 9. *SF nodes: Function from noisy Fourier coefficients, true signal, signal minimizing* $\|v\|_q^q$ *from solution of Algorithm* 4.1 *with* $k^* = 8$, *and error.*

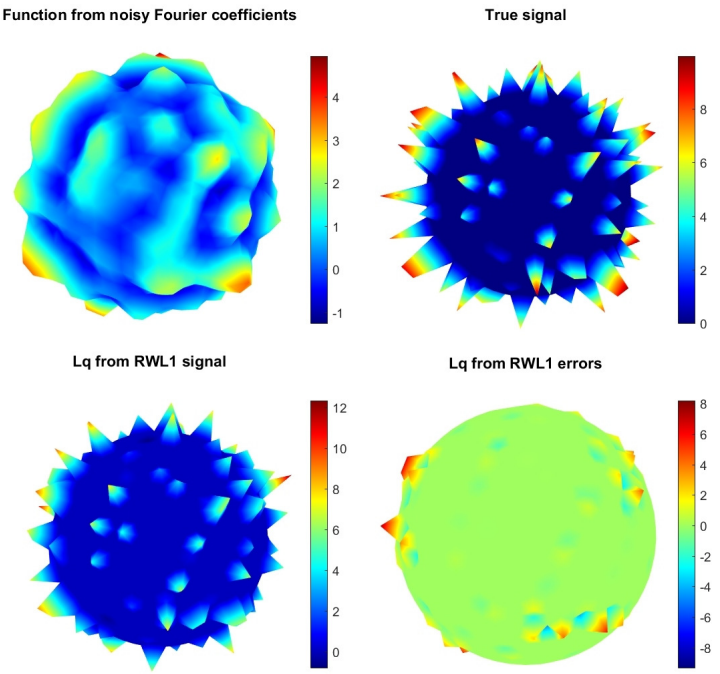

Figure 10. *MD nodes: Function from noisy Fourier coefficients, true signal, signal minimizing* $\|v\|_q^q$ *from solution of Algorithm 4.1 with* $k^* = 8$*, and error.*

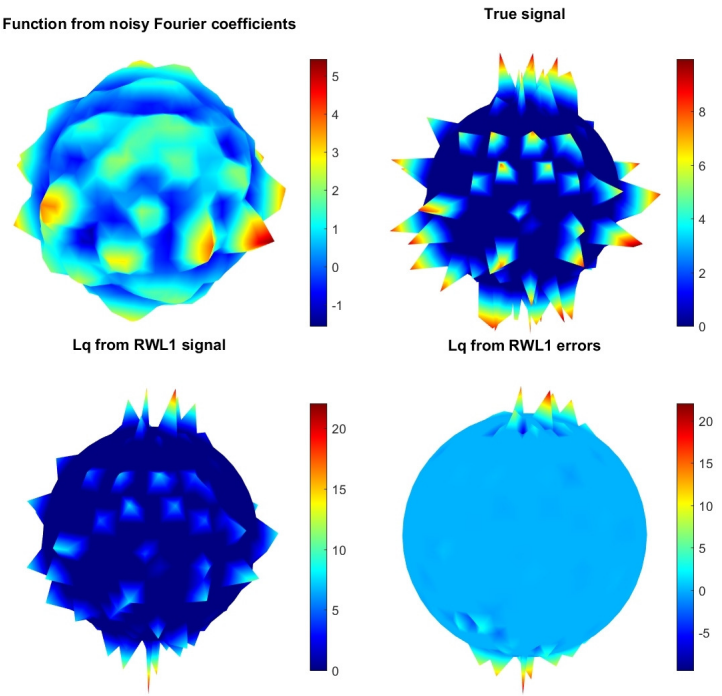

Figure 11. *GL nodes: Function from noisy Fourier coefficients, true signal, signal minimizing* $\|v\|_q^q$ *from solution of Algorithm 4.1 with* $k^* = 8$*, and error.*

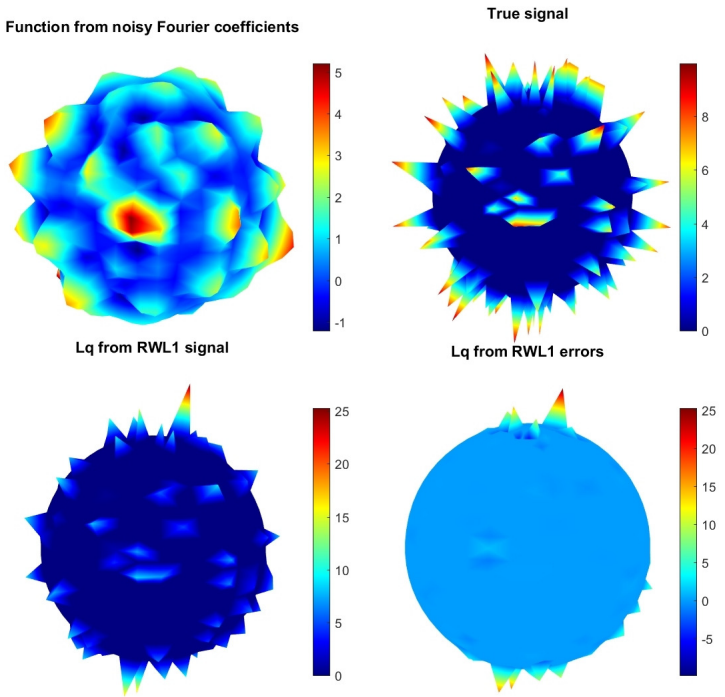

Figure 12. *TP nodes: Function from noisy Fourier coefficients, true signal, signal minimizing* $\|v\|_q^q$ *from solution of Algorithm 4.1 with* $k^* = 8$*, and error.*

6. Concluding remarks. Numerical integration and optimization methods are two important factors in signal processing. In this paper, we comprehensively investigate 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 optimization model (8) and then using Algorithm 4.1 to solve (8) can find better sparse solutions than using other cubature rules and other optimization methods for the ℓ_1 optimization model (7) and the ℓ_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 at The Hong Kong Polytechnic University. Their hospitality is gratefully acknowledged.

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, https://doi.org/10.1137/100795140.

[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, https://doi.org/10.1007/s00365-014-9263-1.

[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, https://doi.org/10.1109/TSP.2015.2399861.

[4] A. Bondarenko, D. Radchenko, and M. Viazovska, *Optimal asymptotic bounds for spherical designs*, Ann. of Math. (2), 178 (2013), pp. 443–452, https://doi.org/10.4007/annals.2013.178.2.2.

[5] A. Bondarenko, D. Radchenko, and M. Viazovska, *Well-separated spherical designs*, Constr. Approx., 41 (2015), pp. 93–112, https://doi.org/10.1007/s00365-014-9238-2.
[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, https://doi.org/10.1137/060657704.
[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, https://doi.org/10.1002/cpa.20124.
[9] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215, https://doi.org/10.1109/TIT.2005.858979.
[10] E. J. Candès, M. B. Wakin, and S. P. Boyd, *Enhancing sparsity by reweighted l_1 minimization*, J. Fourier Anal. Appl., 14 (2008), pp. 877–905, https://doi.org/10.1007/s00041-008-9045-x.
[11] X. Chen, A. Frommer, and B. Lang, *Computational existence proofs for spherical t-designs*, Numer. Math., 117 (2011), pp. 289–305, https://doi.org/10.1007/s00211-010-0332-5.
[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, https://doi.org/10.1007/s10107-012-0613-0.
[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, https://doi.org/10.1137/15M1028054.
[14] X. Chen, F. Xu, and Y. Ye, *Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization*, SIAM J. Sci. Comput., 32 (2010), pp. 2832–2852, https://doi.org/10.1137/090761471.
[15] X. Chen and W. Zhou, *Convergence of the reweighted ℓ_1 minimization algorithm for ℓ_2-ℓ_p minimization*, Comput. Optim. Appl., 59 (2014), pp. 47–61, https://doi.org/10.1007/s10589-013-9553-8.
[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*, 3rd ed., Springer-Verlag, New York, 1999, https://doi.org/10.1007/978-1-4757-6568-7.
[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, https://doi.org/10.1002/cpa.20303.
[19] M. E. Davies and R. Gribonval, *Restricted isometry constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$*, IEEE Trans. Inform. Theory, 55 (2009), pp. 2203–2214, https://doi.org/10.1109/TIT.2009.2016030.
[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, https://doi.org/10.1073/pnas.0437847100.
[23] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory, 47 (2001), pp. 2845–2862, https://doi.org/10.1109/18.959265.
[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, https://doi.org/10.1016/j.acha.2008.09.001.
[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, https://doi.org/10.1007/s10107-011-0470-2.
[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*, Astrophys. J., 622 (2005), pp. 759–771.
[28] M. Gräf, *Efficient Algorithms for the Computation of Optimal Quadrature Points on Riemannian Man-*

ifolds, Dissertation, Universitätsverlag Chemnitz, http://www.qucosa.de/fileadmin/data/qucosa/documents/11528/Dissertation_Gr%C3%A4f.pdf, 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* 2.1, http://cvxr.com/cvx, 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, https://doi.org/10.1016/j.acha.2006.09.003.

[33] S. Kunis, H. Möller, and U. von der Ohe, *Prony's Method on the Sphere*, preprint, https://arxiv.org/abs/1603.02020, 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] M. Nikolova, *Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares*, Multiscale Model. Simul., 4 (2005), pp. 960–991, https://doi.org/10.1137/040619582.

[36] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998, https://doi.org/10.1007/978-3-642-02431-3.

[37] E. B. Saff and A. B. J. Kuijlaars, *Distributing many points on a sphere*, Math. Intelligencer, 19 (1997), pp. 5–11, https://doi.org/10.1007/BF03024331.

[38] R. Sala Llonch, E. Kokiopoulou, I. Tosić, and P. Frossard, 3*D face recognition with sparse spherical representations*, Pattern Recogn., 43 (2010), pp. 824–834.

[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, https://doi.org/10.1023/B:ACOM.0000016428.25905.da.

[40] A. H. Stroud, *Approximate Calculation of Multiple Integrals*, Prentice-Hall Ser. Automat. Comput., Prentice-Hall, Englewood Cliffs, NJ, 1971. .

[41] E. van den Berg and M. P. Friedlander, *SPGL1: A Solver for Large-Scale Sparse Reconstruction*, http://www.cs.ubc.ca/labs/scl/spgl1, 2007.

[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, https://doi.org/10.1137/080714488.

[43] E. van den Berg and M. P. Friedlander, *Sparse optimization with least-squares constraints*, SIAM J. Optim., 21 (2011), pp. 1201–1229, https://doi.org/10.1137/100785028.

[44] M. Wang, W. Xu, and A. Tang, *On the performance of sparse recovery via ℓ_p-minimization* ($0 \le p \le 1$), IEEE Trans. Inform. Theory, 57 (2011), pp. 7255–7278, https://doi.org/10.1109/TIT.2011.2159959.

[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 Contemporary Computational Mathematics—A Celebration of the 80th Birthday of Ian Sloan, J. Dick, F. Y. Kuo, and H. Wozniakowski, eds., Springer, New York, 2018, to appear.

[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, https://doi.org/10.1109/TSP.2009.2016892.

[48] Y. Zhang, *Theory of compressive sensing via ℓ_1 minimization: A non-RIP analysis and extensions*, J. Oper. Res. Soc. China, 1 (2013), pp. 79–105, https://doi.org/10.1007/s40305-013-0010-2.

[49] Y.-B. Zhao and Z.-Q. Luo, *Constructing new weighted l_1-algorithms for the sparsest points of polyhedral sets*, Math. Oper. Res., 42 (2017), pp. 57–76, https://doi.org/10.1287/moor.2016.0791.

[50] Y. Zhou and X. Chen, *Spherical t_ϵ-designs for applications on the sphere*, Math. Comp., to appear, https://doi.org/10.1090/mcom/3306.

[51] J. Zhu, X. Li, F. Arroyo, and E. Arroyo, *Error analysis of reweighted l_1 greedy algorithm for noisy reconstruction*, J. Comput. Appl. Math., 286 (2015), pp. 93–101, https://doi.org/10.1016/j.cam.2015.02.038.