Spherical designs and nonconvex minimization for recovery of sparse signals on the sphere

Xiaojun Chen
Department of Applied Mathematics
The Hong Kong Polytechnic University

Joint work with Rob Womersley
School of Mathematics and Statistics
University of New South Wales

SIAM J. Imaging Science, 2018

Workshop on Modern Optimization and Applications 2018, Beijing

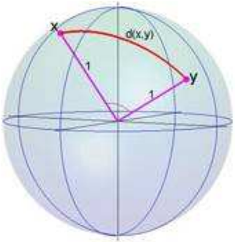
Outline

Recovery of sparse signals on the unit sphere

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

Applications

- astrophysics, geophysics, climate modelling
- gravitational sensing, global navigation
- 3D face recognition



For polar angle $\theta \in [0, \pi]$, azimuthal angle $\phi \in [0, 2\pi)$

$$\boldsymbol{x} = (\sin(\theta)\cos(\phi),\ \sin(\theta)\sin(\phi),\ \cos(\theta))^T, \quad \boldsymbol{x} \in \mathbb{S}^2$$

Fourier coefficients

Let $\mathbb{P}_L(\mathbb{S}^2)$ denote the space of all spherical polynomials of degree at most L on $\mathbb{S}^2$. Let $Y_{\ell,k}, \ell = 0, \ldots, L, k = 1, \ldots, 2\ell + 1$ be an orthonormal basis for $\mathbb{P}_L(\mathbb{S}^2)$.

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

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

$\ell = 0, \ldots, L, \quad k = 1, \ldots, 2\ell + 1$, with noise $\eta_{\ell,k}$.

Use a cubature rule for $\mathbb{S}^2$ with weights and nodes

$$w_j > 0, \quad \mathcal{X}_n := \{\boldsymbol{x}_j \in \mathbb{S}^2, j = 1, \ldots, n\},$$

which is exact for all spherical polynomials of degree at most $2L$. Applying such a cubature rule to (1), we have the discrete approximation

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

for $F \in \mathbb{P}_L(\mathbb{S}^2)$.

In matrix notation, (2) is

$$\boldsymbol{c} = YWf + \eta \tag{3}$$

- $\boldsymbol{c} \in \mathbb{R}^m$ is the noisy Fourier coefficient vector
- $Y := Y(\mathcal{X}_n) \in \mathbb{R}^{m \times n}$ 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
- $f = (F(\boldsymbol{x}_1), \ldots, F(\boldsymbol{x}_n))^T \in \mathbb{R}^n$ is the signal vector which we want to recover.
- $m = (L+1)^2$ is the dimension of $\mathbb{P}_L(\mathbb{S}^2)$

Let

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

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 $2L$, the matrix $\boldsymbol{A}$ satisfies

$$\boldsymbol{A}\boldsymbol{A}^T = YWY^T = I \in \mathbb{R}^{m \times m} \tag{4}$$

Optimization model

- Optimization problem

$$(P) \qquad \text{Minimize} \quad \|\boldsymbol{v}\|_q^q$$
$$\text{Subject to} \quad \|\boldsymbol{A}\boldsymbol{v} - \boldsymbol{c}\|_2 \leq \sigma,$$

- $0 < q \leq 1$
- Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\quad \boldsymbol{c} \in \mathbb{R}^m$, $\quad \sigma \geq 0$
- $\sigma = 0 \iff \boldsymbol{A}\boldsymbol{v} = \boldsymbol{c}$ (Basis Pursuit)
- $\sigma > 0$ (Basis Pursuit De-Noising)

Quadrature rules

- Quadrature rule Q_n for $\mathbb{S}^2$:
  - nodes $\mathcal{X}_n = \{\boldsymbol{x}_j \in \mathbb{S}^2, j = 1, \ldots, n\}$
  - weights $w_j > 0$, $j = 1, \ldots, n$
  - $$Q_n(f) := \sum_{j=1}^{n} w_j f(\boldsymbol{x}_j) \approx \int_{\mathbb{S}^2} f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}),$$
- $\mathbb{P}_t(\mathbb{S}^2)$ = spherical polynomials of degree at most t on $\mathbb{S}^2$
- Degree of precision t:
  - Q_n exact for polynomials $p \in \mathbb{P}_t(\mathbb{S}^2)$
  - not exact for some polynomial $\bar{p} \in \mathbb{P}_{t+1}(\mathbb{S}^2)$
- Degree of precision $0 \iff$ exact for constants

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

Spherical t-design

- $\mathcal{X}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ spherical t-design Delsarte, Goethals, Seidel[1977]

$$\frac{1}{n}\sum_{j=1}^{n} p(\boldsymbol{x}_j) = \frac{1}{|\mathbb{S}^2|}\int_{\mathbb{S}^2} p(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{x}) \qquad \forall\, p \in \mathbb{P}_t(\mathbb{S}^2),$$

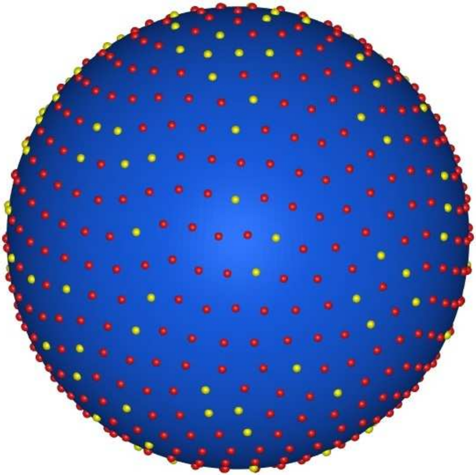
  - Equal weight $w_j = 4\pi/n$, $j = 1, \ldots, n$, degree of precision t
- $Y_{\ell,k}, k = 1, \ldots, 2\ell+1, \quad \ell = 0, \ldots, t$ orthonormal basis $\mathbb{P}_t(\mathbb{S}^2)$
- Spherical harmonic basis matrix $\boldsymbol{Y}(\mathcal{X}_n) \in \mathbb{R}^{(t+1)^2 \times n}$
- Spherical t-design $\iff$ $\mathcal{X}_n$ is a solution of

$$\boldsymbol{Y}(\mathcal{X}_n)\boldsymbol{e} = \frac{n}{\sqrt{4\pi}}\boldsymbol{e}_1$$

  [Chen-Womersley, SINUM2006]
- Well-separated t-designs [An-Chen-Sloan-Womersley, SINUM2010,2012], [Chen-Frommer-Lang, Numer. Math 2011]

Point sets on $\mathbb{S}^2$

- Spherical t-design: $t = 31$, $n = 1024$, $|J| = 120$

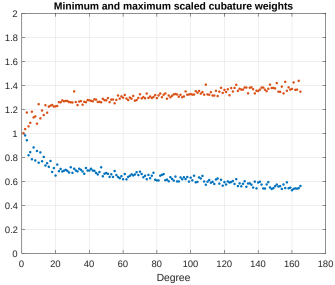


Spherical t_ϵ designs

- Yang and Chen [Math. Comp 2017], Degree of precision t
- Relaxed weights: For $0 < \epsilon < 1$

$$\frac{4\pi}{n}(1-\epsilon) \leq w_i \leq \frac{4\pi}{n}(1-\epsilon)^{-1}.$$

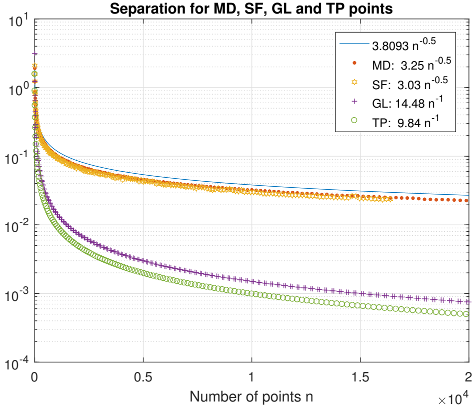
- Extremal points Sloan and Womersley [2004])
  - $n = (t+1)^2 = \dim \mathbb{P}_t(\mathbb{S}^2)$, Degree of precision t
  - nodes chosen to maximize det $(\boldsymbol{G})$, gram matrix $\boldsymbol{G} = \boldsymbol{Y}^T \boldsymbol{Y}$



Well Separated

- Minimal separation $\delta(\mathcal{X}_n) = \min_{j \neq k} \text{dist}\,(\boldsymbol{x}_j, \boldsymbol{x}_k) := \arccos(\boldsymbol{x}_j \cdot \boldsymbol{x}_k)$
- Sequence of point sets $\{\mathcal{X}_n\}$ is well-separated if there is c such that

$$\delta(\mathcal{X}_n) \geq cn^{-1/2} \quad \text{for all } n \geq n_0$$



Stability

Theorem

Let $\widehat{\boldsymbol{v}}$ *be a solution of problem (P). If there exists* $\gamma \in (0,1)$ *with* $\|\boldsymbol{r}_J\|_q \leq \gamma \|\boldsymbol{r}_{J^c}\|_q$, *then*

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

- J support set of oracle solution $\boldsymbol{v}^*$
- $J^c = \{1, \ldots, n\} \backslash J$, complement of J
- $\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{Q})(\widehat{\boldsymbol{v}} - \boldsymbol{v}^*)$, $\quad \boldsymbol{Q} = \boldsymbol{A}^T\boldsymbol{A}$ idempotent ($\boldsymbol{Q}\boldsymbol{Q} = \boldsymbol{Q}$)
  - Key: Degree of precision $t \geq 2L$, orthonormal spherical harmonics
- $\beta = n^{\frac{2-q}{2q}}$
- $\boldsymbol{v}^*$ oracle solution, the unique solution of

$$\begin{aligned} &\text{Minimize} && \|\boldsymbol{v}\|_2^2 \\ &\text{Subject to} && \|\boldsymbol{A}\boldsymbol{v} - \boldsymbol{c}\|_2 \leq \sigma, \qquad \boldsymbol{v}_{J^c} = 0 \end{aligned}$$

Sufficient conditions to recover 0-norm solution

- Consider basis pursuit, $\sigma = 0 \implies \boldsymbol{A}\boldsymbol{v} = \boldsymbol{c}$
- Want conditions:

$$\operatorname{argmin}\{\|\boldsymbol{v}\|_1 : \boldsymbol{A}\boldsymbol{v} = \boldsymbol{c}\} \subseteq \operatorname{argmin}\{\|\boldsymbol{v}\|_0 : \boldsymbol{A}\boldsymbol{v} = \boldsymbol{c}\}$$

- spark$(\boldsymbol{A}) = \min\{\|\boldsymbol{v}\|_0 : \boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}, \boldsymbol{v} \neq \boldsymbol{0}\}$
- Mutual coherence: Donoho, Huo [2001], columns $\boldsymbol{a}_j$ of $\boldsymbol{A}$ length 1

$$M(A) = \max_{i \neq j} |\boldsymbol{a}_i^T \boldsymbol{a}_j|$$

- RIP: Candes, Romberg, Tao [2004] Smallest δ_s s.t.

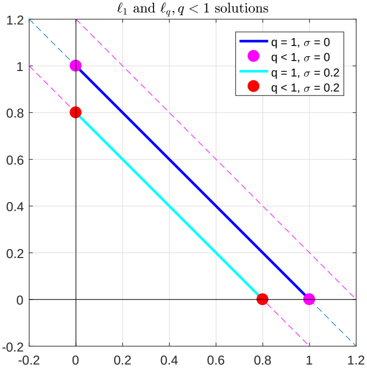
$$(1-\delta_s)\|\boldsymbol{v}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{v}\|_2^2 \leq (1+\delta_s)\|\boldsymbol{v}\|_2^2, \quad \forall \boldsymbol{v} : \|\boldsymbol{v}\|_0 \leq s$$

- Null space (Non-RIP): Zhang [2008, 2012],

$$(\|\boldsymbol{v}\|_0)^{\frac{1}{2}} < \tfrac{1}{2} \min\left\{ \frac{\|\boldsymbol{v}\|_1}{\|\boldsymbol{v}\|_2} : \boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}, \boldsymbol{v} \neq \boldsymbol{0} \right\}$$

Example

- Chen, Ge, Wang and Ye [MP 2014]
- $\boldsymbol{A} = [1\ 1]$, $\quad \boldsymbol{c} = 1$
- $0 \leq \sigma < 1$



- $q = 1$ solution set

$$\left\{ \begin{bmatrix} v_1 \\ 1 - \sigma - v_1 \end{bmatrix},\ v_1 \in [0, 1 - \sigma] \right\}$$

- $q < 1$ two global minimizers

$$\left\{ \begin{bmatrix} 0 \\ 1 - \sigma \end{bmatrix}, \begin{bmatrix} 1 - \sigma \\ 0 \end{bmatrix} \right\}$$

Regularization vs penalty formulation

- Constrained problem

$$(P) \qquad \text{Minimize} \quad \|\boldsymbol{v}\|_q^q$$
$$\text{Subject to} \quad \|\boldsymbol{A}\boldsymbol{v} - \boldsymbol{c}\|_2 \le \sigma,$$

- Regularization

$$(P_R) \qquad \text{Minimize} \quad \tau\|\boldsymbol{A}\boldsymbol{v} - \boldsymbol{c}\|_2^2 + \|\boldsymbol{v}\|_q^q$$

- Exact penalty formulation

$$(\mathrm{P_E}) \qquad \text{Minimize} \quad \tau\left(\|\boldsymbol{A}\boldsymbol{v} - \boldsymbol{c}\|_2^2 - \sigma^2\right)_+ + \|\boldsymbol{v}\|_q^q$$

  - $q = 1$: (P_R) is equivalent to (P) some $\tau > 0$
  - $q < 1$: there is no $\tau > 0$ so (P_R), (P) equivalent
  - $q < 1$: (P_E) is equivalent to (P) some $\tau > 0$
  - Chen, Lu, Pong [SIOPT 2016]

Stationary points

For $\mu > 0$, consider

$$\begin{array}{ll} \text{Minimize} & \||\boldsymbol{v}| + \mu e\|_q^q = \sum_{i=1}^n (|v_i| + \mu)^q \\ \text{Subject to} & \|\boldsymbol{A}\boldsymbol{v} - \boldsymbol{c}\|_2 \leq \sigma \end{array} \tag{5}$$

Theorem 2 (i) Let $\boldsymbol{v}^*$ and $\boldsymbol{v}_\mu$ be global minimizers of problems (P) and (5), respectively. Then

$$\||\boldsymbol{v}^*| + \mu e\|_q^q \leq \||\boldsymbol{v}_\mu| + \mu e\|_q^q + n\mu^q$$
$$\|\boldsymbol{v}_\mu\|_q^q \leq \|\boldsymbol{v}^*\|_q^q + n\mu^q.$$

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

$$0 \in (I - \boldsymbol{Q})\mathcal{W}(\boldsymbol{v}, \mu), \qquad \text{where}$$

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

Algorithm

Choose $k^* > 1$ and $\mu > 0$. Let $\alpha_i^0 = 1, i = 1, \ldots, n$ and $k = 0$.
1. Using SPGL1 (Berg and Friedlander 2011), solve

$$\begin{aligned} \boldsymbol{v}^k \in \operatorname{Argmin} \quad & \sum_{i=1}^{n} \alpha_i^k |v_i| \\ \text{s.t.} \quad & \|\boldsymbol{A}\boldsymbol{v} - \boldsymbol{c}\|_2 \le \sigma. \end{aligned}$$

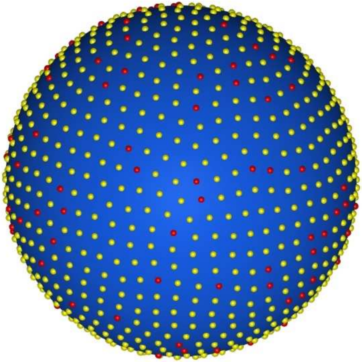
2. If $k + 1 = k^*$, go to Step 3. Otherwise let

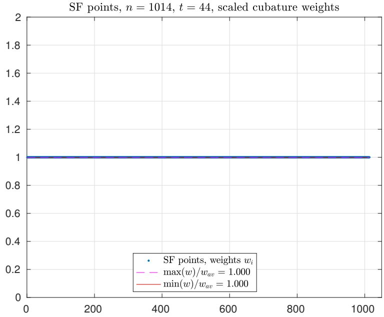
$$\alpha_i^{k+1} = q(|v_i^k| + \mu)^{q-1}, \ i = 1, \ldots, n$$

and go to Step 1.
3. Using $\boldsymbol{v}^{k^*}$ as an initial point, solve problem (P) by the penalty method (Chen, Lu and Pong 2016).

Spherical t-designs

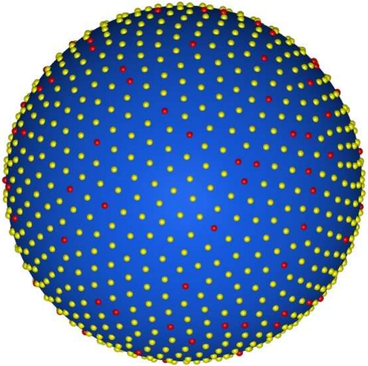
- Spherical t-design, equal weight, Efficient $n = t^2/2 + O(t)$
- $t = 44$, $n = 1014$, $L = 15$, $|J| = 120$

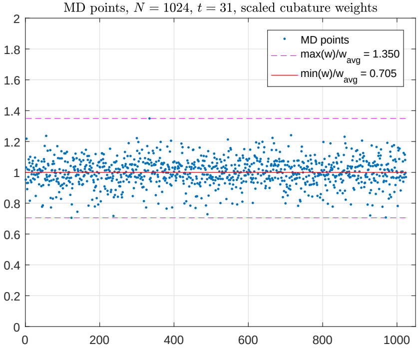




Extremal, spherical t_ϵ design

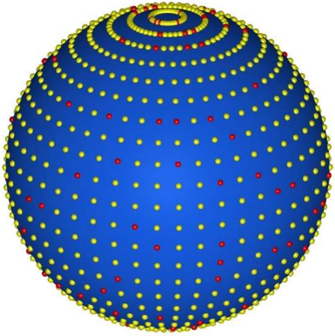
- t_ϵ designs from extremal points:
- $n = (t+1)^2$, points chosen to maximize det $\left(\boldsymbol{Y}\boldsymbol{Y}^T\right)$
- Computed weights bounded, provably well separated
- $t = 31$, $n = 1024$, $L = 15$, $|J| = 120$

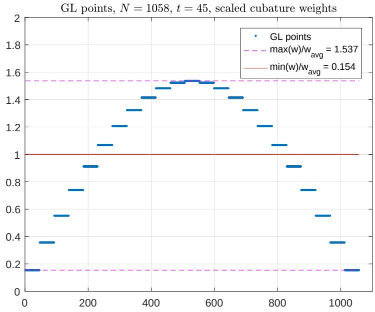




Tensor product Gauss, equal

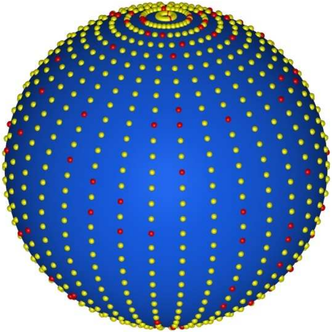
- Gauss quadrature on $[-1,1] = [\cos(\pi), \cos(0)]$, Equally spaced azimuthal $[0, 2\pi)$
- $t = 45$, $n = 1058$, $L = 15$, $|J| = 120$

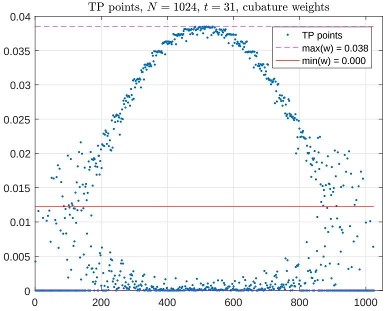




θ, ϕ tensor product

- Equal spacing in polar angle $\theta \in [0, \pi]$, azimuthal angle $\phi \in [0, 2\pi)$
- $t = 31$, $n = 1024$, $L = 15$, $|J| = 120$
- full $\boldsymbol{Y} \in \mathbb{R}^{n \times n}$ but $\operatorname{rank}(\boldsymbol{Y}) = 768$
- Unscaled quadrature weights, forced to be non-negative





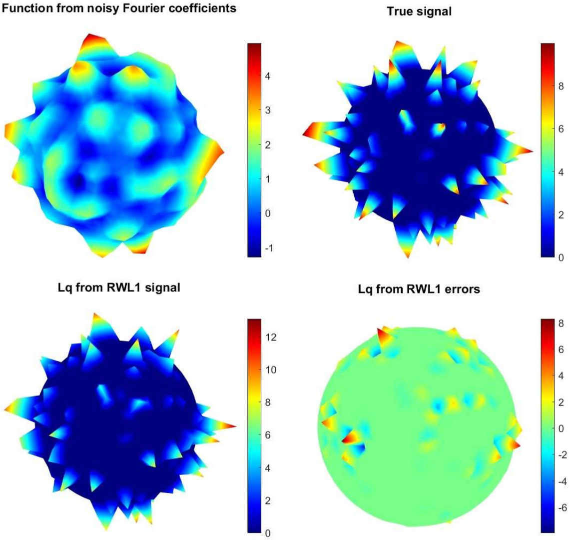


Figure: SF nodes: function from noisy Fourier coefficients, true signal, signal minimizing $\|v\|_q^q$ from solution of Algorithm 4.1 with $k^* = 8$, and error

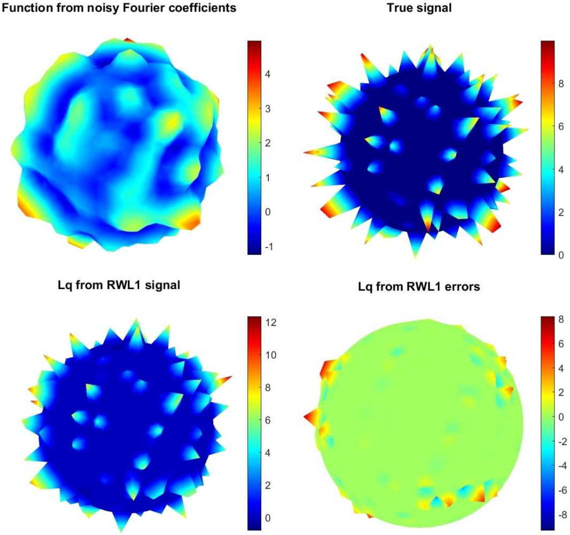


Figure: MD nodes: function from noisy Fourier coefficients, true signal, signal minimizing $\|v\|_q^q$ from solution of Algorithm 4.1 with $k^* = 8$, and error

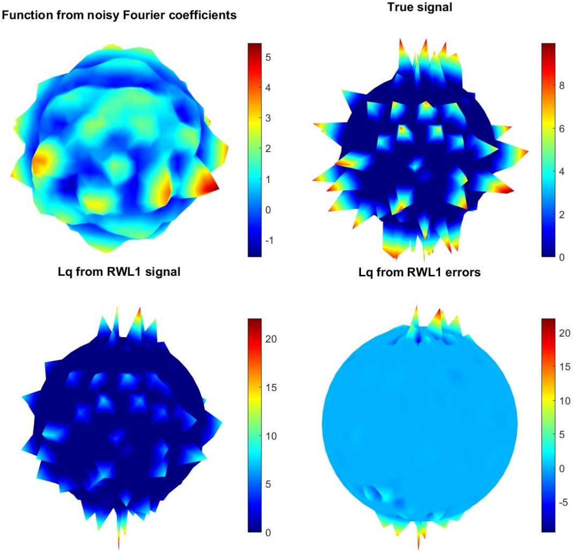


Figure: `GL` nodes: function from noisy Fourier coefficients, true signal, signal minimizing $\|v\|_q^q$ from solution of Algorithm 4.1 with $k^* = 8$, and error

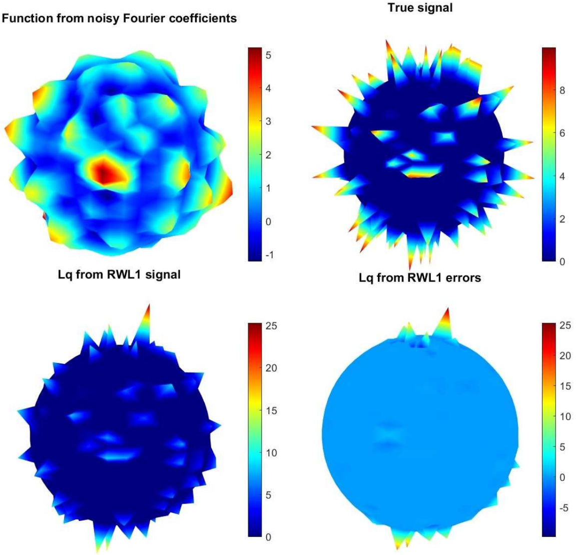


Figure: TP nodes: function from noisy Fourier coefficients, true signal, signal minimizing $\|v\|_q^q$ from solution of Algorithm 4.1 with $k^* = 8$, and error

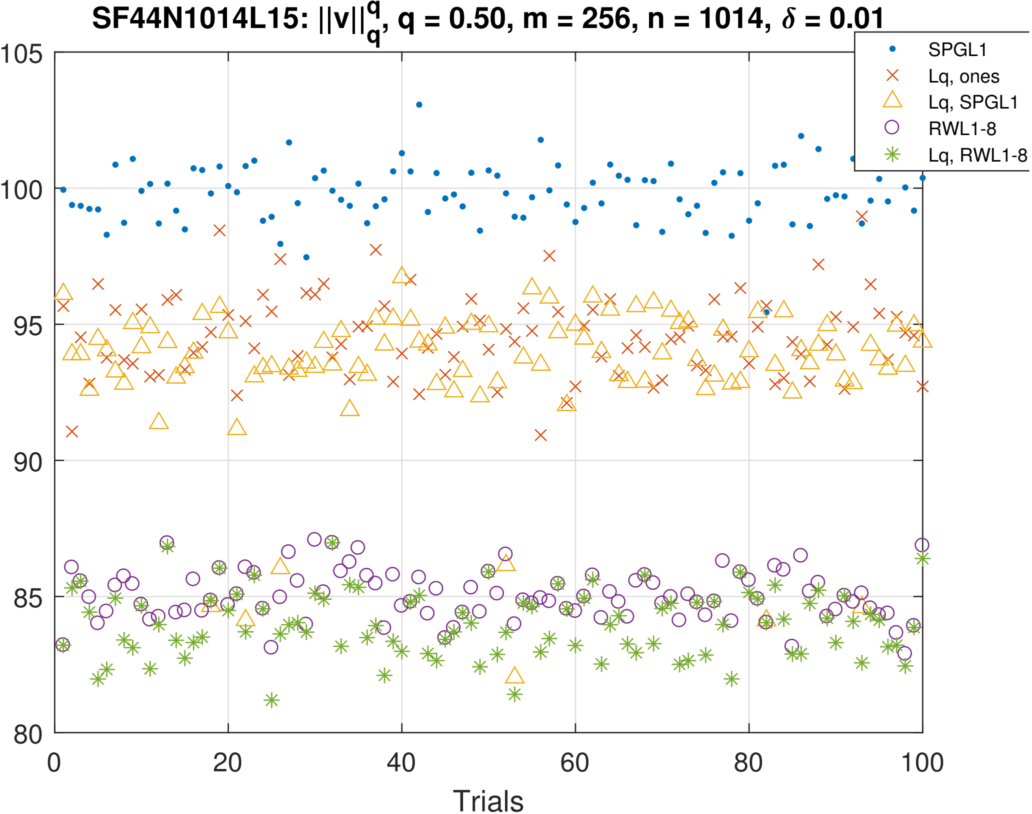
Method	$\|\hat{\boldsymbol{v}}\|_q^q$	$\varphi(\hat{\boldsymbol{v}})$	$\|\boldsymbol{v}^* - \hat{\boldsymbol{v}}\|_2$	$\|\hat{\boldsymbol{v}}\|_0$	$\|\boldsymbol{v}^* \& \hat{\boldsymbol{v}}\|_0$	false
SPGL1	99.688	-1.1e-08	3.21	217.4	101.7	115.7
ℓ_q, ones	94.715	2.1e-17	7.65	150.7	42.8	107.9
ℓ_q, SPGL1	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**

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

$$\varphi(\boldsymbol{v}) = \|\boldsymbol{A}\boldsymbol{v} - \boldsymbol{c}\| - \sigma$$

X. Chen and R. Womersley, SIAM J. Imag. Sci. (2018)

1. Y. Zhou amd X. Chen, Spherical t_ϵ-designs for approximations on the sphere, Math. Comput. (2018)
2. X. Chen, Z. Lu and T.K. Pong, Penalty methods for a class of non-Lipschitz optimization problems, SIAM J. Optim. (2016).
3. X. Chen, D. Ge, Z. Wang and Y. Ye, Complexity of unconstrained L_2-L_p minimization, Math. Program. (2014).
4. C. An, X. Chen, I.H. Sloan, R. S. Womersley, Regularized least squares approximations on the sphere using spherical designs, SIAM J. Numer. Anal. (2012).
5. X. Chen, A. Frommer and B. Lang, Computational existence proofs for spherical t-designs, Numer. Math. (2011).
6. C. An, X. Chen, I.H. Sloan, R. S. Womersley, Well conditioned spherical designs for integration and interpolation on the two-sphere, SIAM J. Numer. Anal. (2010).
7. X. Chen and R. Womersley, Existence of solutions to systems of underdetermined equations and spherical designs, SIAM J. Numer. Anal. (2006)



SF44N1014L15: $||v||_q^q$, q = 0.50, m = 256, n = 1014, δ = 0.01
SPGL1
Lq, ones
Lq, SPGL1
RWL1-8
Lq, RWL1-8
105
100
95
90
85
80
0
20
40
60
80
100
Trials