Uncertainty Quantification and Approximation Theory for Parameterized PDEs

Part IV: Discrete least square and compressed sensing techniques

Hoang A. Tran

Clayton G. Webster, Guannan Zhang

Computer Science and Mathematics Division
Oak Ridge National Laboratory, USA

Supported by: **DOE** (ASCR)

Vietnam Institute for Advanced Study in Mathematics
November 15, 2016

Outline

1. Discrete least square
2. Compressed sensing: Introduction and motivation
3. Restricted isometry property
4. Best lower s-term reconstruction
5. Optimize sample complexity estimate using envelope bound

Discrete least square

Objective: Approximate $h \in L^2(\mathcal{U}, d\varrho)$

$$h(\boldsymbol{y}) = \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d} \widehat{h}_{\boldsymbol{\nu}} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}), \quad \text{with } \widehat{h}_{\boldsymbol{\nu}} = \langle h, \boldsymbol{\Psi}_{\boldsymbol{\nu}} \rangle.$$

Parametric discretization: global polynomial space

$$\mathcal{P}_\Lambda(\mathcal{U}) = \operatorname{span}\left\{ \prod_{i=1}^d {y_i}^{\nu_i}, \text{ with } \boldsymbol{\nu} \in \Lambda \right\} \subset L^2(\mathcal{U}, d\varrho).$$

The best approximation of h on $\mathcal{P}_\Lambda(\mathcal{U})$ is $h_\Lambda(\boldsymbol{y}) := \sum_{\boldsymbol{\nu} \in \Lambda} \widehat{h}_{\boldsymbol{\nu}} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y})$:

$$\|h - h_\Lambda\|_\varrho = \min_{q \in \mathcal{P}_\Lambda(\mathcal{U})} \|h - q\|_\varrho.$$

- In general, we can only access h from the observations at the points $(\boldsymbol{y}_i)_{i=1}^m$.

Discrete least square (DLS) problem:

$$h^{LS} := \operatorname*{arg\,min}_{q \in \mathcal{P}_\Lambda(\mathcal{U})} \sum_{i=1}^m |h(\boldsymbol{y}_i) - q(\boldsymbol{y}_i)|^2.$$

Discrete least square

- Recall: $s = \#\Lambda = \dim[\mathcal{P}_\Lambda(\mathcal{U})]$. Assume $h^{LS} = \sum_{\boldsymbol{\nu}\in\Lambda} c_{\boldsymbol{\nu}} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y})$, then

$$(c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda} := \operatorname*{arg\,min}_{\boldsymbol{z}=(z_{\boldsymbol{\nu}})\in\mathbb{C}^s} \sum_{i=1}^m \left| h(\boldsymbol{y}_i) - \sum_{\boldsymbol{\nu}\in\Lambda} z_{\boldsymbol{\nu}} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i) \right|^2 .$$

- Taking derivative with respect to $\boldsymbol{z}_{\boldsymbol{\nu}}$ yields

$$0 = \sum_{i=1}^m \left(h(\boldsymbol{y}_i) - \sum_{\boldsymbol{\nu}'\in\Lambda} z_{\boldsymbol{\nu}'} \boldsymbol{\Psi}_{\boldsymbol{\nu}'}(\boldsymbol{y}_i) \right) \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i)$$

$$= \sum_{i=1}^m h(\boldsymbol{y}_i)\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i) - \sum_{\boldsymbol{\nu}'\in\Lambda} z_{\boldsymbol{\nu}'} \sum_{i=1}^m \boldsymbol{\Psi}_{\boldsymbol{\nu}'}(\boldsymbol{y}_i)\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i)$$

- $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ is the solution of

$$\boldsymbol{G}\boldsymbol{c} = \boldsymbol{h}$$

where $\boldsymbol{G}$ is an $s \times s$ matrix and $\boldsymbol{h}$ is an $s \times 1$ vector given by

$$\boldsymbol{G}_{\boldsymbol{\nu},\boldsymbol{\nu}'} = \frac{1}{m}\sum_{i=1}^m \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i)\boldsymbol{\Psi}_{\boldsymbol{\nu}'}(\boldsymbol{y}_i), \qquad \boldsymbol{h}_{\boldsymbol{\nu}} = \frac{1}{m}\sum_{i=1}^m h(\boldsymbol{y}_i)\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i)$$

Discrete least square

Consider the least square problem

$$\boldsymbol{G}\boldsymbol{c} = \boldsymbol{h}$$

where $\boldsymbol{G}$ is an $s \times s$ matrix and $\boldsymbol{h}$ is an $s \times 1$ vector given by

$$\boldsymbol{G}_{\boldsymbol{\nu},\boldsymbol{\nu}'} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i)\boldsymbol{\Psi}_{\boldsymbol{\nu}'}(\boldsymbol{y}_i), \qquad \boldsymbol{h}_{\boldsymbol{\nu}} = \frac{1}{m}\sum_{i=1}^{m} h(\boldsymbol{y}_i)\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i)$$

- For the stability, $\boldsymbol{G}$ needs to be well-conditioned.

Observation: assume $\boldsymbol{y}_i$ is randomly sampled according to the measure ϱ,

$$\text{for } m \to \infty, \qquad \boldsymbol{G}_{\boldsymbol{\nu},\boldsymbol{\nu}'} \to \int_{\mathcal{U}} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y})\boldsymbol{\Psi}_{\boldsymbol{\nu}'}(\boldsymbol{y})\varrho(\boldsymbol{y})d\boldsymbol{y} = \delta_{\boldsymbol{\nu},\boldsymbol{\nu}'}$$

$$\mathbb{E}(\boldsymbol{G}) = \boldsymbol{I}.$$

- how to quantify the proximity of the matrices $\boldsymbol{G}$ and $\boldsymbol{I}$?

Discrete least square

Introduce the quantity: $K(\Lambda) = \sup_{\boldsymbol{y}\in\mathcal{U}} \sum_{\boldsymbol{\nu}\in\Lambda} |\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y})|^2$

Spectral norm: $|||\boldsymbol{G}||| = \max_{\boldsymbol{z}\neq 0} \frac{|\langle \boldsymbol{G}\boldsymbol{z}, \boldsymbol{z}\rangle|}{\|\boldsymbol{z}\|^2}$

Theorem [Cohen, Davenport, Leviatan '13]

For $0 < \delta < 1$:

$$\mathbb{P}(|||\boldsymbol{G} - \boldsymbol{I}||| \leq \delta) > 1 - 2s\exp\left(-\frac{c_\delta m}{K(\Lambda)}\right)$$

where $c_\delta := \delta + (1-\delta)\log(1-\delta) > 0$.

$$\begin{aligned} |||\boldsymbol{G} - \boldsymbol{I}||| \leq \delta &\iff \max_{\boldsymbol{z}\neq 0} \frac{|\langle \boldsymbol{G}\,\boldsymbol{z}, \boldsymbol{z}\rangle - \|\boldsymbol{z}\|^2|}{\|\boldsymbol{z}\|^2} \leq \delta \\ &\iff (1-\delta)\|\boldsymbol{z}\|^2 \leq \langle \boldsymbol{G}\,\boldsymbol{z}, \boldsymbol{z}\rangle \leq (1+\delta)\|\boldsymbol{z}\|^2, \quad \forall \boldsymbol{z} \in \mathbb{C}^s \\ &\iff (1-\delta)\|\boldsymbol{z}\|^2 \leq \|\boldsymbol{A}\boldsymbol{z}\|^2 \leq (1+\delta)\|\boldsymbol{z}\|^2, \quad \forall \boldsymbol{z} \in \mathbb{C}^s, \end{aligned}$$

where $\boldsymbol{A}$ is an $m \times s$ sampling matrix with $\boldsymbol{A}_{i,\boldsymbol{\nu}} = \frac{1}{\sqrt{m}}\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i)$.

Discrete least square

"Isometry property"

Theorem [CDL13] restated

Let $\boldsymbol{A}$ is a sampling matrix with size $m \times s$

$$\boldsymbol{A}_{i,\boldsymbol{\nu}} = \frac{1}{\sqrt{m}} \Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i).$$

For $0 < \delta < 1$ and $c_\delta = \delta + (1-\delta)\log(1-\delta) > 0$, with probability exceeding $1 - 2s\exp\left(-\frac{c_\delta m}{K(\Lambda)}\right)$ then

$$(1-\delta)\|\boldsymbol{z}\|^2 \leq \|\boldsymbol{A}\boldsymbol{z}\|^2 \leq (1+\delta)\|\boldsymbol{z}\|^2, \quad \forall \boldsymbol{z} \in \mathbb{C}^s. \tag{IP}$$

$\boldsymbol{A}$ satisfies the "isometry property":

- Set $\delta = \frac{1}{2}$ and m such that $\frac{m}{\log m} \geq \frac{K(\Lambda)(1+r)}{c_{1/2}}$: (IP) holds with prob. $\geq 1 - 2m^{-r}$.
- Set $m \geq \frac{K(\Lambda)}{c_\delta}\left(\log(2s) + \log(\frac{1}{\gamma})\right)$: (IP) holds with prob. $\geq 1 - \gamma$.

Discrete least square

Stability

Theorem

Assume $|h(\boldsymbol{y})| \leq L$, $\forall \boldsymbol{y} \in \mathcal{U}$. For any $r > 0$, if m satisfies

$$\frac{m}{\log m} \geq \frac{K(\Lambda)(1+r)}{c_{1/2}},$$

then

$$\mathbb{E}(\|h - h^{LS}\|^2) \lesssim \underbrace{\|h - h_\Lambda\|^2}_{\text{best approximation error on } \Lambda} + L^2 m^{-r}.$$

- Estimate $K(\Lambda)$?

Estimate $K(\Lambda)$

1d setting: $\Lambda = \{0, 1, \dots, s-1\}$

Recall from above

$$K(\Lambda) = \sup_{\boldsymbol{y}\in\mathcal{U}} \sum_{\boldsymbol{\nu}\in\Lambda} |\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y})|^2 = \sum_{\boldsymbol{\nu}\in\Lambda} \|\boldsymbol{\Psi}_{\boldsymbol{\nu}}\|^2_{L^\infty}.$$

Trigonometric polynomials: $\Psi_j(y) = e^{ijy}$

- $\|\Psi_j\|_{L^\infty} = 1 \;\Rightarrow\; K(\Lambda) = s.$ Stability condition: $\dfrac{m}{\log(m)} \gtrsim s.$

Legendre polynomials: $L_j(y)$

- $\|L_j\|_{L^\infty} = L_j(1) = \sqrt{2j+1}$

  $\Rightarrow\; K(\Lambda) = \sum_{j=0}^{s-1}(2j+1) = s^2.$ Stability condition: $\dfrac{m}{\log(m)} \gtrsim s^2.$

Chebyshev polynomials: $T_0(y) = 1$, $T_j(y) = \sqrt{2}\cos(j\arccos(y))$

- $\|T_j\|_{L^\infty} = \sqrt{2}$ if $j \geq 1$

  $\Rightarrow\; K(\Lambda) = 2s - 1.$ Stability condition: $\dfrac{m}{\log(m)} \gtrsim s.$

Estimate $K(\Lambda)$

Multidimensional setting

Lower set: An index set $\Lambda \subset \mathbb{N}_0^d$ is lower, a.k.a., downward closed, if

$$(\boldsymbol{\nu} \in \Lambda \text{ and } \boldsymbol{\nu}' \leq \boldsymbol{\nu}) \implies \boldsymbol{\nu}' \in \Lambda,$$

where $\boldsymbol{\nu}' \leq \boldsymbol{\nu}$ means that $\nu_i' \leq \nu_i$ for all $1 \leq i \leq d$.

- Generalization of the set $\Lambda = \{0, \dots, s\}$ in $1d$.
- For smooth functions, good index set is often lower set.

Theorem [Chkifa, Cohen, Migliorati, Nobile, Tempone '14]

Assume Λ is a lower set.

- Legendre: $K_L(\Lambda) = \sum_{\boldsymbol{\nu}\in\Lambda} \prod_{i=1}^d (2\nu_i + 1) \leq (\#\Lambda)^2$.
- Chebyshev: $K_T(\Lambda) = \sum_{\boldsymbol{\nu}\in\Lambda} 2^{\#(\mathrm{supp}(\boldsymbol{\nu}))} \leq (\#\Lambda)^{\frac{\log 3}{\log 2}}$.

Stability conditions:

$$\frac{m}{\log(m)} \gtrsim s^2 \text{ for Legendre systems and } \frac{m}{\log(m)} \gtrsim s^{\frac{\log 3}{\log 2}} \text{ for Chebyshev systems.}$$

Discrete least square

Application to parameterized PDEs

1. Generate m random samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ according to measure ϱ
2. For each $\boldsymbol{y}_i$, solve the parameterized PDEs $\mathcal{L}(u, \boldsymbol{y}_i) = f$ for solution $u(\boldsymbol{y}_i) \in \mathcal{V}^h$.
3. $u^{LS} = \sum_{\boldsymbol{\nu} \in \Lambda} c_{\boldsymbol{\nu}}(x) \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y})$, where

$$(c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda} = \underset{\boldsymbol{z}=(z_{\boldsymbol{\nu}}) \in (\mathcal{V}^h)^s}{\arg\min} \sum_{i=1}^{m} \left\| u(\boldsymbol{y}_i) - \sum_{\boldsymbol{\nu} \in \Lambda} z_{\boldsymbol{\nu}} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i) \right\|_{\mathcal{V}}^2$$

4. $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$ is the solution of

$$\boldsymbol{G}\boldsymbol{c} = \boldsymbol{u},$$

where $\boldsymbol{G}$ is an $s \times s$ matrix and $\boldsymbol{u}$ is an $s \times 1$ vector in $(\mathcal{V}^h)^s$ given by

$$\boldsymbol{G}_{\boldsymbol{\nu}, \boldsymbol{\nu}'} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i) \boldsymbol{\Psi}_{\boldsymbol{\nu}'}(\boldsymbol{y}_i), \qquad \boldsymbol{u}_{\boldsymbol{\nu}} = \frac{1}{m} \sum_{i=1}^{m} u(\boldsymbol{y}_i) \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i)$$

Discrete least square

Summary

$u^{LS} = \sum_{\boldsymbol{\nu}\in\Lambda} c_{\boldsymbol{\nu}}(x)\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y})$, where

$$(c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda} = \underset{\boldsymbol{z}=(z_{\boldsymbol{\nu}})\in(\mathcal{V}^h)^s}{\arg\min} \sum_{i=1}^{m} \left\| u(\boldsymbol{y}_i) - \sum_{\boldsymbol{\nu}\in\Lambda} z_{\boldsymbol{\nu}} \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i) \right\|_{\mathcal{V}}^2$$

Pros:

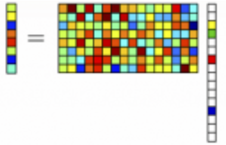
- Non-intrusive method.
- Mitigate Runge phenomena

Cons:

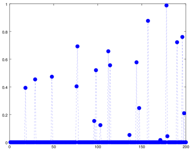
- Number of samples is bigger than the degree of freedom.
- Accuracy is sensitive to the a priori choice of polynomial space.

Compressed sensing

- Initially developed for signal recovery [Candès, Romberg, Tao '06; Donoho '06]:
- Recover signals/functions from an underdetermined system



- Sparsity assumption: only a few coordinates are non-zero.



- Sparse signals are recovered via sparsity-induced norm, i.e.,

$$\boldsymbol{c} = \arg\min \|\boldsymbol{z}\|_0 \text{ subject to } \boldsymbol{u} = \mathcal{F}\boldsymbol{z}$$

Sparsity-induced norms

$$\boldsymbol{c} = \arg\min \|\boldsymbol{z}\|_{?} \text{ subject to } \boldsymbol{u} = \mathcal{F}\boldsymbol{z}$$

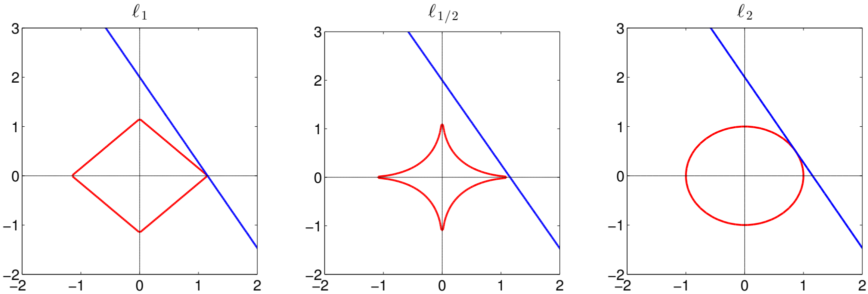
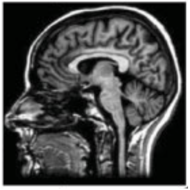
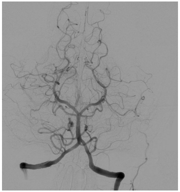
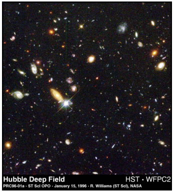


Figure : Illustrations of some types of regularization

Example: signal and image processing

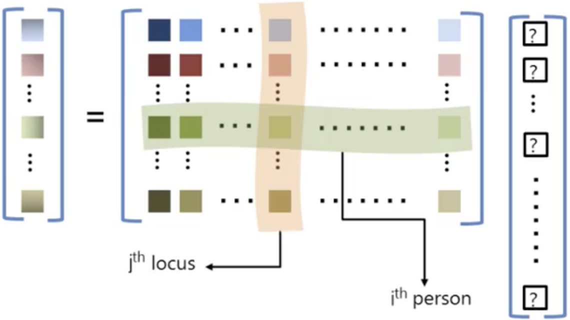




- Reconstruct image ($\boldsymbol{c}$) from the measurement ($\boldsymbol{u}$), obtained via a transform ($\mathcal{F}$)

$$\boldsymbol{u} = \mathcal{F}\boldsymbol{c}.$$

- Images often possess sparse structures.
- Less measurements are generally preferred.

Examples: Genetics



Source: Coursera

Compressed sensing

Application to parameterized PDEs

- Recall from above: approximate $u(x, \boldsymbol{y})$ by

$$u(x, \boldsymbol{y}) \simeq u^{\#}(x, \boldsymbol{y}) = \sum_{\boldsymbol{\nu} \in \Lambda} c_{\boldsymbol{\nu}}(x) \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}).$$

- Non-intrusive approach: compute $u(\,\cdot\,, \boldsymbol{y}_i)$ for a set of samples $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}$ on $\mathcal{U}$.
- $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda}$ satisfies

$$u(x, \boldsymbol{y}_i) = \sum_{\boldsymbol{\nu} \in \Lambda} c_{\boldsymbol{\nu}}(x) \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i), \quad \forall i = 1, \ldots, m$$

$$\iff \qquad \boldsymbol{u} = \boldsymbol{\Psi} \boldsymbol{c}.$$

where

$$\boldsymbol{\Psi} = (\boldsymbol{\Psi}_{i,\boldsymbol{\nu}}) = (\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i))_{\substack{i \in [m], \\ \boldsymbol{\nu} \in \Lambda}}, \qquad \boldsymbol{u} = (u(\,\cdot\,, \boldsymbol{y}_i))_{i \in [m]} \in \mathcal{V}^m.$$

Compressed sensing

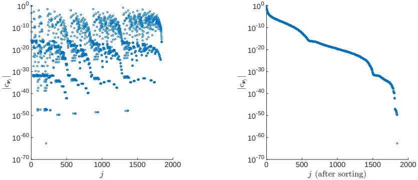
Application to parameterized PDEs

Solve $\boldsymbol{u} = \boldsymbol{\Psi c}$ where

$$\boldsymbol{\Psi} = (\boldsymbol{\Psi}_{i,\boldsymbol{\nu}}) = (\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_i))_{\substack{i\in[m],\\ \boldsymbol{\nu}\in\Lambda}}, \qquad \boldsymbol{u} = (u(\,\cdot\,,\boldsymbol{y}_i))_{i\in[m]} \in \mathcal{V}^m.$$

Observation:

1. each measurement $u(\boldsymbol{y}_i) \iff$ 1 PDE solve.
2. $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ decays fast ("approximately sparse").



3. important coefficients often has low indices.
4. however, we don't know the shape of the correct index set.

Compressed sensing

Application to parameterized PDEs

Main idea:

- approximate $\boldsymbol{u}$ on a big polynomial subspace $\mathbb{P}_{\Lambda_0}(\mathcal{U})$ with Λ_0 possibly far from optimal.
  - Denote $N := \#(\Lambda_0)$.
- undersampling: generate $m \ll N$ samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ and solve for $u(\boldsymbol{y}_1), \ldots, u(\boldsymbol{y}_m)$.
- reconstruct $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \Lambda_0}$ from the underdetermined system $\boldsymbol{u} = \boldsymbol{\Psi} \boldsymbol{c}$ using compressed sensing algorithm:

$$\boldsymbol{c} = \arg\min_{\boldsymbol{z} \in \mathcal{V}^N} \|\boldsymbol{z}\|_1 \ \text{ s.t. } \ \boldsymbol{u} = \boldsymbol{\Psi} \boldsymbol{z}.$$

  - $\|\boldsymbol{z}\|_1 := \sum_{\boldsymbol{\nu} \in \Lambda_0} \|z_{\boldsymbol{\nu}}\|_{\mathcal{V}}$.

Uniform vs. non-uniform recovery

Def. s-sparse vector: a vector with at most s nonzero entries.

$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are randomly sampled according to measure ϱ.

The reconstruction is guaranteed with high probability.

- Non-uniform recovery:

  the recovery is success with high probability for each fixed s-sparse vector.

- Uniform recovery:

  the recovery is success with high probability for all s-sparse vector.

Restricted isometry property (RIP)

Uniform recovery is guaranteed by the **restricted isometry property (RIP)** of the normalized matrix $\boldsymbol{A} = \frac{1}{\sqrt{m}}\boldsymbol{\Psi}$:

$\boldsymbol{A}$ satisfies the RIP if there exists small δ_s, s.t. for all $\boldsymbol{z} \in \mathbb{C}^N$ s-sparse vectors,

$$(1-\delta_s)\|\boldsymbol{z}\|_2^2 \le \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \le (1+\delta_s)\|\boldsymbol{z}\|_2^2.$$

- Intuition: $\ker(\boldsymbol{A})$ does not contain any non-zero s-sparse vectors.
- Comparison to stability condition of DLS:

$$(1-\delta_s)\|\boldsymbol{z}\|_2^2 \le \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \le (1+\delta_s)\|\boldsymbol{z}\|_2^2, \quad \forall \boldsymbol{z} \in \mathbb{C}^N.$$

RIP implies the recovery of best s-term.

Theorem: Assume $\boldsymbol{A}$ satisfies RIP. Then

$$\|u^{CS} - u\|_1 \lesssim \sigma_s(u)_1.$$

- Estimate the number of measurements such that the RIP is guaranteed?

Restricted isometry property (RIP)

Uniform recovery is guaranteed by the **restricted isometry property (RIP)** of the normalized matrix $\boldsymbol{A} = \frac{1}{\sqrt{m}}\boldsymbol{\Psi}$:

> $\boldsymbol{A}$ satisfies the RIP if there exists small δ_s, s.t. for all $\boldsymbol{z} \in \mathbb{C}^N$ s-sparse vectors,
>
> $$(1-\delta_s)\|\boldsymbol{z}\|_2^2 \le \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \le (1+\delta_s)\|\boldsymbol{z}\|_2^2.$$

- Intuition: $\ker(\boldsymbol{A})$ does not contain any non-zero s-sparse vectors.
- Comparison to stability condition of DLS:

$$(1-\delta_s)\|\boldsymbol{z}\|_2^2 \le \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \le (1+\delta_s)\|\boldsymbol{z}\|_2^2, \quad \forall \boldsymbol{z} \in \mathbb{C}^N.$$

RIP implies the recovery of best s-term.

Theorem: Assume $\boldsymbol{A}$ satisfies RIP. Then

$$\|u^{CS} - u\|_1 \lesssim \sigma_s(u)_1.$$

- Estimate the number of measurements such that the RIP is guaranteed?

RIP estimate

How many samples do we need?

Define the uniform bound of orthonormal system

$$\Theta = \sup_{\boldsymbol{\nu} \in \Lambda_0} \|\boldsymbol{\Psi}_{\boldsymbol{\nu}}\|_{\infty}.$$

Theorem

Let $\boldsymbol{\Psi} \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated with a BOS. Provided that

$$m \geq C\Theta^2 s \log^2(s) \log(N),$$

then $\boldsymbol{A}$ is satisfied the RIP with high probability.

- Developed through a series of papers [Candes, Tao '06; Rudelson, Vershynin '08; Rauhut '10; Cheraghchi, Guruswami, Velingker '13; Bourgain '14; Haviv, Regev '15; Chkifa, Dexter, T., Webster '15].

RIP estimate

How many samples do we need?

RIP estimate

$$m \geq C\Theta^2 s \log^2(s) \log(N),$$

- Very mild dependence on N.
- Signal processing: $\Theta = 1$ for Fourier, Hadamard, circulant, etc. matrices
  - to reconstruct best s-term, need $\simeq s$ samples.
- Polynomial approximations: Θ can be prohibitively high
  - Chebyshev basis: $\Theta = 2^{d/2}$.
  - Legendre basis: $\Theta \gtrsim N$.
  - preconditioned Legendre basis: $\Theta = 2^{d/2}$.

Reconstruction of best *lower* s-term approximations

Lower RIP

Plan: Reconstruct an approximation of u which is comparable to the **best lower s-term approximation**, i.e., best approximation by s-terms in a lower set.

$$u_{\rm bl} := \mathop{\arg\min}_{\substack{\#(\Lambda)=s \\ \Lambda \text{ lower}}} \|u - u_\Lambda\|, \qquad \sigma_s^\ell(u) := \|u - u_{\rm bl}\|.$$

Main advantages:

- Less demanding approximations, thus, the sample complexity is reduced.
- We can show that the best lower s-term is as good as best s-term approximation.
- Reduce the effect of Runge's phenomenon.
- We can choose the multi-index set Λ_0 as a hyperbolic cross $\mathcal{H}_s$, which is the union of all lower sets of cardinality s, i.e.,

$$\mathcal{H}_s = \left\{ \boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) \in \mathbb{N}_0^d : \prod_{i=1}^{d} (\nu_i + 1) \leq s \right\}.$$

Note: $N = \#(\mathcal{H}_s) \leq 2s^3 4^d$, [Chernov, Dũng '15].

Reconstruction of best *lower* s-term approximations

Lower RIP

For index sets $\Lambda \subset \mathbb{N}_0^d$ and $s \in \mathbb{N}$, recall

$$K(\Lambda) := \sum_{\boldsymbol{\nu}\in\Lambda} \|\Psi_{\boldsymbol{\nu}}\|_\infty^2 \quad \text{and} \quad K(s) = \sup_{\Lambda \text{ lower},\ |\Lambda|=s} K(\Lambda).$$

lower RIP:

There exists small $\delta_{l,s}$ such that

$$(1-\delta_{l,s})\|\boldsymbol{c}\|_2^2 \le \|\tilde{\boldsymbol{\Psi}}\boldsymbol{c}\|_2^2 \le (1+\delta_{l,s})\|\boldsymbol{c}\|_2^2, \quad \forall \boldsymbol{c}\ s\text{-sparse},\ supp(\boldsymbol{c}) \text{ lower}.$$

Theorem [Chkifa, Dexter, T., Webster '15]

Let $\boldsymbol{\Psi} \in \mathbb{C}^{m\times N}$ be the orthonormal random sampling matrix. If, for $\delta \in (0,1)$,

$$m \ge C_\delta K(s) \log^2(K(s)) \log(N),$$

then with high probability, $\boldsymbol{A}$ satisfies the lower RIP with $\delta_{l,s} \le \delta$.

- Legendre: $K_L(\Lambda) = (\#\Lambda)^2$, $\forall\Lambda$ lower set $\implies K(s) \le s^2$.
- Chebyshev: $K_T(\Lambda) = (\#\Lambda)^{\frac{\log 3}{\log 2}}$, $\forall\Lambda$ lower set $\implies K(s) \le s^{\frac{\log 3}{\log 2}}$.

Recovery of best lower s-term approximations

Implementation: Weighted ℓ_1 minimization

Weighted ℓ_1 minimization:

- Choose the specific weight $\omega_\nu = \|\Psi_\nu\|_\infty$,
- $\boldsymbol{\Psi} = (\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i))$ is an $m \times N$ sampling matrix,
- $\boldsymbol{u} = (u(\boldsymbol{y}_i))_{i=1,\ldots,m}$,

Find $u^{CS}(\boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\Lambda_0} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})$, where $\boldsymbol{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_0}$ is the solution of

$$\min \sum_{\boldsymbol{\nu}\in\Lambda_0} \omega_{\boldsymbol{\nu}}\|z_{\boldsymbol{\nu}}\|_{\mathcal{V}} \quad \text{subject to } \boldsymbol{u} = \boldsymbol{\Psi}\boldsymbol{z}.$$

Remark:

- The weight favors low index.
- The weight penalizes high index.

Recovery of best lower s-term approximations

Weighted ℓ_1 minimization

Let $\boldsymbol{\omega} = (\|\boldsymbol{\Psi}_{\boldsymbol{\nu}}\|_\infty)_{\boldsymbol{\nu}\in\Lambda_0}$ be a vector of weights. We define

- for $u(x,\boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\Lambda_0} \widehat{u}_{\boldsymbol{\nu}}(x)\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})$, $\|f\|_{\omega,1} := \sum_{\boldsymbol{\nu}\in\Lambda_0} \omega_{\boldsymbol{\nu}}\|\widehat{u}_{\boldsymbol{\nu}}\|_{\mathcal{V}}$,
- $\sigma_s^{(\ell)}(u)_{\omega,1} = \inf_{\substack{supp(u^\#) \text{ lower} \\ |supp(u^\#)|=s}} \|u - u^\#\|_{\omega,1}$.

Theorem [Chkifa, Dexter, T., Webster '15]

Assume that the number of samples satisfies

$$m \geq CK(s)\log^2(K(s))\log(N)$$

then, with high probability, there holds

$$\|u - u^{CS}\|_{\omega,1} \leq C\sigma_s^{(\ell)}(u)_{\omega,1}$$

We improve a weighted ℓ_1 minimization with a specific choice of weights which

- has a reduced sample complexity compared to unweighted ℓ_1 minimization,
- lead to approx. comparable to best s-term approx. in case of smooth solutions.

Example 1: $u(\boldsymbol{y}) = \frac{\prod_{i=1}^{d/2} \cos(8y_i/2^i)}{\prod_{i=d/2+1}^{d}(1-y_i/4^i)}$

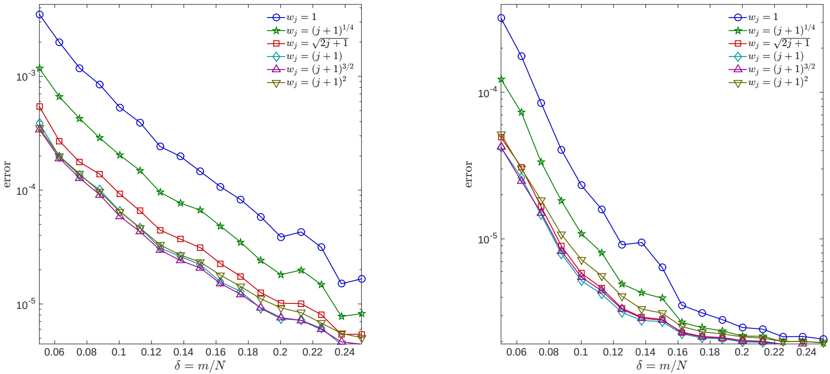


Figure : **Left:** $d = 8,\ N = 1843$. **Right:** $d = 16,\ N = 4129$.

Example 2: $u(\boldsymbol{y}) = \exp\left(-\frac{\sum_{i=1}^{d}\cos(y_i)}{8d}\right)$

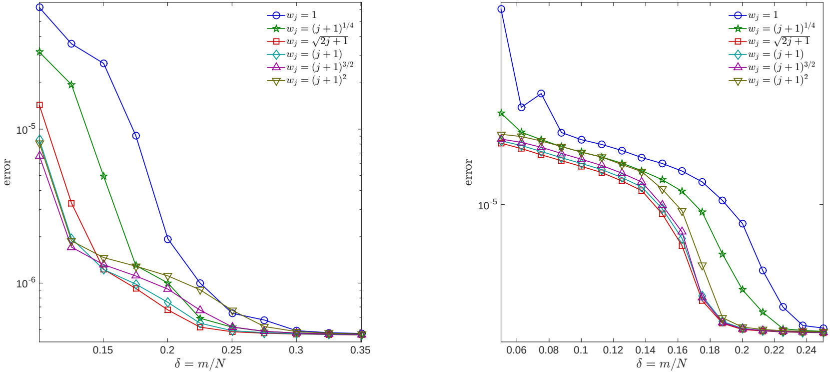


Figure : **Left:** $d = 8,\ N = 1843$. **Right:** $d = 16,\ N = 4129$.

Example 3: $u(\boldsymbol{y}) = \exp\left(-\frac{\sum_{i=1}^{d} y_i}{2d}\right)$

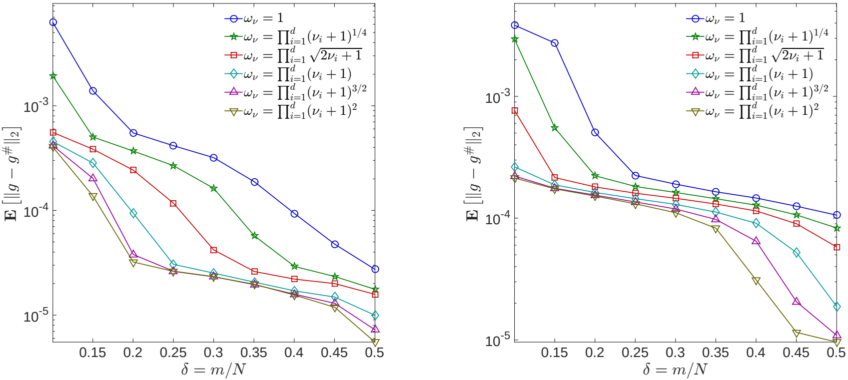


Figure : **Left:** $d = 8,\ N = 1843.$ **Right:** $d = 16,\ N = 4129.$

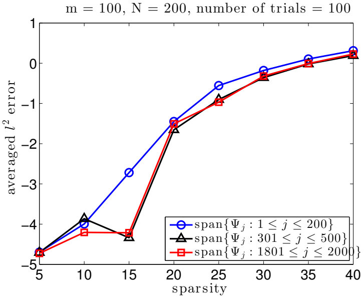
Optimize sample complexity estimates

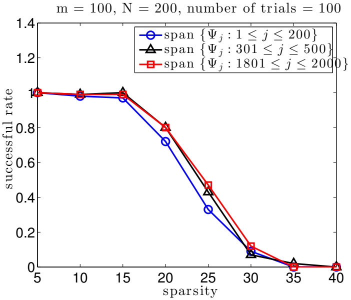
Sparse Legendre expansions

Consider $1d$ Legendre expansion on $\Lambda_0 = \{0, 1, \ldots, N-1\}$.

Current theoretical condition gives: $m \geq \Theta^2 s \times$ log factor, where $\Theta = \sqrt{2N-1}$.

Numerical experiment shows:





- There is some successful recovery with underdetermined Legendre systems.
- The number of measurements for recovery guarantee should not be depend on N, or maximum polynomial degree.

Optimize sample complexity estimates

Sparse Legendre expansions

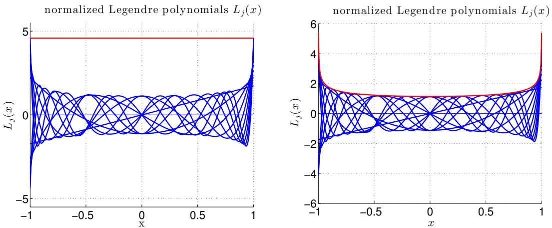
Theorem [T., Webster '16]

The sufficient condition for recovery of sparse standard Legendre expansions is

$$m \geq s^2 \times \text{log factor},$$

independent of polynomial degree.

- a new, improved estimate of the number of measurements using the envelope bound, independently of mutual coherence, for univariate sparse Legendre expansion.
- a simple criteria for selecting random sample sets, that helps to improve probability of reconstruction of sparse solutions.



Restricted eigenvalue condition

Uniform recovery is usually guaranteed by the **restricted isometry property (RIP)** of the normalized matrix $\boldsymbol{A} = \frac{1}{\sqrt{m}}\boldsymbol{\Psi}$:

$\boldsymbol{A}$ satisfies the RIP if there exists small δ, s.t. for all $\boldsymbol{z}$ s-sparse vectors,

$$(1-\delta)\|\boldsymbol{z}\|_2^2 \le \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \le (1+\delta)\|\boldsymbol{z}\|_2^2.$$

For reconstruction using ℓ_1 minimization, the upper bound of $\|\boldsymbol{A}\boldsymbol{z}\|_2^2$ is not necessary.

Restricted eigenvalue condition [Bickel, Ritov, Tsybakov '09; van de Geer, Bühlmann '09]

For $\alpha > 1$, define

$$C(S;\alpha) := \left\{\boldsymbol{z} \in \mathbb{C}^d : \|\boldsymbol{z}_{S^c}\|_1 \le \alpha\sqrt{s}\|\boldsymbol{z}_S\|_2\right\}.$$

$\boldsymbol{A}$ satisfies the **restricted eigenvalue condition (REC)** of order s if there exist $\alpha > 1$ and $0 < \delta < 1$ such that

$$\|\boldsymbol{A}\boldsymbol{z}\|_2^2 \ge (1-\delta)\|\boldsymbol{z}\|_2^2,$$

for all $\boldsymbol{z} \in C(s;\alpha) := \bigcup_{(\#S)=s} C(S;\alpha)$.

- $C(s;\alpha)$ contains all s-sparse vector.
- REC holds but RIP fails to hold for many random Gaussian and sub-Gaussian design matrices, e.g., [Raskutti, Wainwright, Yu '10; Rudelson, Zhou '13].

Restricted eigenvalue condition

Uniform recovery is usually guaranteed by the **restricted isometry property (RIP)** of the normalized matrix $\boldsymbol{A} = \frac{1}{\sqrt{m}}\boldsymbol{\Psi}$:

$\boldsymbol{A}$ satisfies the RIP if there exists small δ, s.t. for all $\boldsymbol{z}$ s-sparse vectors,

$$(1-\delta)\|\boldsymbol{z}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \leq (1+\delta)\|\boldsymbol{z}\|_2^2.$$

For reconstruction using ℓ_1 minimization, the upper bound of $\|\boldsymbol{A}\boldsymbol{z}\|_2^2$ is not necessary.

Restricted eigenvalue condition [Bickel, Ritov, Tsybakov '09; van de Geer, Bühlmann '09]

For $\alpha > 1$, define

$$C(S;\alpha) := \left\{\boldsymbol{z} \in \mathbb{C}^d : \|\boldsymbol{z}_{S^c}\|_1 \leq \alpha\sqrt{s}\|\boldsymbol{z}_S\|_2\right\}.$$

$\boldsymbol{A}$ satisfies the **restricted eigenvalue condition (REC)** of order s if there exist $\alpha > 1$ and $0 < \delta < 1$ such that

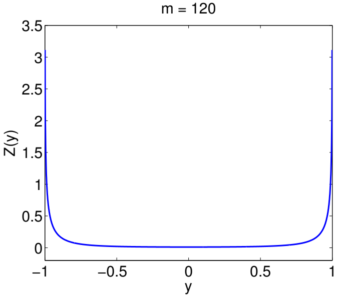
$$\|\boldsymbol{A}\boldsymbol{z}\|_2^2 \geq (1-\delta)\|\boldsymbol{z}\|_2^2,$$

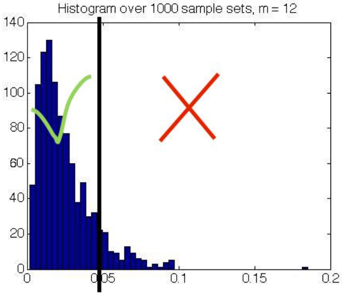
for all $\boldsymbol{z} \in C(s;\alpha) := \bigcup_{(\#S)=s} C(S;\alpha)$.

- $C(s;\alpha)$ contains all s-sparse vector.
- REC holds but RIP fails to hold for many random Gaussian and sub-Gaussian design matrices, e.g., [Raskutti, Wainwright, Yu '10; Rudelson, Zhou '13].

Indicator function and preferable sample sets

- Indicator function: $Z(y) := \Theta(y) \exp(-\frac{1}{4\Theta^2(y)}\sqrt{m})$.
- Preferable sample sets: $\{y_1, \ldots, y_m\}$ such that $\sum_{i=1}^{m} Z(y_i)$ is small.

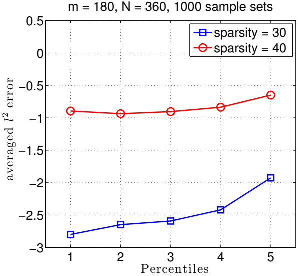


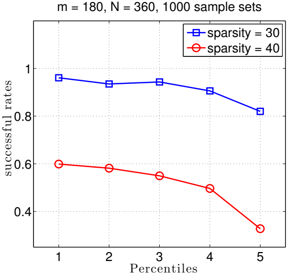


Numerical illustration

Sample sets corresponding to small $\sum_{i=1}^{m} Z(y_i)$ is better

- Generate 1000 random Legendre sampling matrices $\mathbf{\Psi}$ of size 180×360.
- Allocate these matrices to five percentiles according to ascending order of $\sum_{i=1}^{m} Z(y_i)$.
- Generate 20 manufactured solutions $\boldsymbol{c}$ with sparsity 30 and 40. The supports and values of components of c are chosen randomly.
- Reconstruct $\boldsymbol{c}$ by solving $\boldsymbol{u} = \mathbf{\Psi}\boldsymbol{c}$ with ℓ_1 minimization.
- Plot the averaged error and probability of success (TOL: error $< 10^{-3}$) for each percentile.
- Software: SPGL1 [van den Berg and M. P. Friedlander '07 '08].

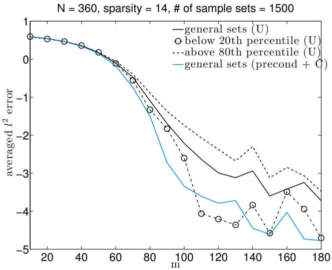


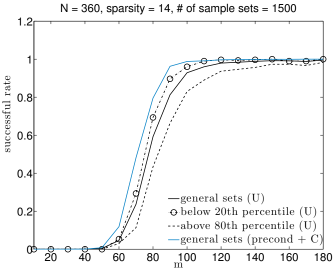


Numerical illustration

Sample sets corresponding to small $\sum_{i=1}^{m} Z(y_i)$ is better

- For each $m = 10, 20, \ldots, 180$, generate 1500 random Legendre sampling matrices $\boldsymbol{\Psi}$ of size $m \times 360$.
- For each m, allocate the matrices to five percentiles according to ascending order of $\sum_{i=1}^{m} Z(y_i)$.
- For each $\boldsymbol{\Psi}$, generate manufactured solutions $\boldsymbol{c}$ randomly with sparsity 14 and then reconstruct $\boldsymbol{c}$ by ℓ_1 minimization.
- For each m, compute the averaged error and probability of success (TOL: error $< 10^{-3}$) over the first and last percentile and all sampling matrices.
- Plot those quantities against m.

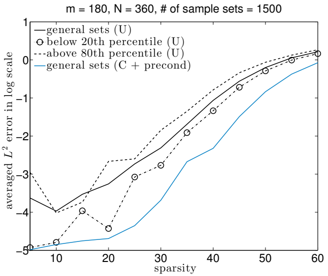


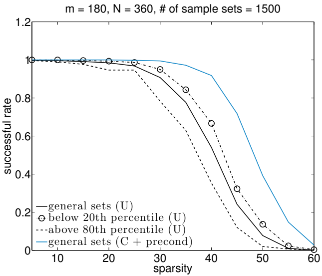


Numerical illustration

Sample sets corresponding to small $\sum_{i=1}^{m} Z(y_i)$ is better

- For each sparsity $s = 5, 10, \ldots, 60$, generate 1500 random Legendre sampling matrices $\boldsymbol{\Psi}$ of size 180×360.
- For each s, allocate the matrices to five percentiles according to ascending order of $\sum_{i=1}^{m} Z(y_i)$.
- For each s and $\boldsymbol{\Psi}$, generate manufactured solutions $\boldsymbol{c}$ randomly with sparsity s and then reconstruct $\boldsymbol{c}$ by ℓ_1 minimization.
- For fixed s, compute the averaged error and probability of success (TOL: error $< 10^{-3}$) over the first and last percentile and all sampling matrices.
- Plot those quantities against s.





References

- J. Bourgain, *An improved estimate in the restricted isometry problem*. In: Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, pp. 65–70 (2014)
- A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, R. Tempone, *Discrete least squares polynomial approximation with random evaluations - application to parametric and stochastic elliptic PDEs*. ESAIM: M2AN 49 (2015), no. 3, 815–837.
- A. Chkifa, N. Dexter, H. Tran, C. G. Webster, *Polynomial approximation via compressed sensing of high-dimensional functions on lower sets*. Mathematics of Computation, to appear (2016).
- A. Cohen, M. A. Davenport, D. Leviatan, *On the Stability and Accuracy of Least Squares Approximations*, Found Comput Math (2013) 13:819–834
- A. Chernov, D. Dũng, *New explicit-in-dimension estimates for the cardinality of high-dimensional hyperbolic crosses and approximation of functions having mixed smoothness*. Journal of Complexity 32 (2016) 92–121
- S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser (2013).
- I. Haviv, O. Regev, *The restricted isometry property of subsampled Fourier matrices*, SODA 2016.
- H. Rauhut, C. Schwab, *Compressive sensing Petrov-Galerkin approximation of high dimensional parametric operator equations*, Math. Comp. (2016), to appear.
- H. Rauhut, R. Ward, *Interpolation via weighted l1-minimization*. Applied and Computational Harmonic Analysis 40 (2016), no. 2, 321–351.
- H. Tran, C. G. Webster, *A uniform bound of sample complexity for sparse Legendre approximation via ℓ_1 minimization in arbitrary polynomial subspace*, preprint (2016).