OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Polynomial approximations via compressed sensing of high-dimensional functions

Clayton G. Webster[†⋆]

A. Chkifa[⋆], N. Dexter[†], H. Tran[⋆]

[†]Department of Mathematics, University of Tennessee

[⋆]Department of Computational & Applied Mathematics (CAM)
Oak Ridge National Laboratory

Supporting agencies: **DOE** (ASCR, BES), **DOD** (AFOSR, DARPA)

`equinox.ornl.gov` `formulate.ornl.gov` `tasmanian.ornl.gov`

Outline

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

1. Motivation - parameterized / stochastic equations
2. Compressed sensing approaches for recovering best s-term approximations
3. An improved complexity estimate for uniform polynomial recovery
4. lower-RIP: new weighted ℓ_1 minimization and lower hard thresholding for overcoming the curse of dimensionality
5. Compressed sensing for recovering parametric PDEs
6. A general theory for nonconvex regularizations
7. Concluding remarks

Motivation: Parameterized PDE models

Deterministic and stochastic coefficients

parameters $\boldsymbol{y} \in \mathcal{U} \subset \mathbb{R}^d$ $\longrightarrow$ **PDE model:** $\mathcal{F}(a(y))[u(\boldsymbol{y})] = 0$ in $D \subset \mathbb{R}^n$, $n = 1, 2, 3$ $\longrightarrow$ quantity of interest $Q[u(\boldsymbol{y})]$

- The operator $\mathcal{F}$, linear or nonlinear, depends on a vector of d parameters $\boldsymbol{y} = (y_1, y_2, \ldots, y_d) \in \mathcal{U} = \prod_{i=1}^d \mathcal{U}_i$, which can be deterministic or stochastic.
- **Deterministic setting**: $\boldsymbol{y}$ are known or controlled by the user.
  - **Goal**: a query $\boldsymbol{y} \in \mathcal{U}$, quickly approximation the solution map $\boldsymbol{y} \mapsto u(\boldsymbol{y}) \in \mathcal{V}$.
- **Stochastic setting**: $\boldsymbol{y}$ may be affected by uncertainty and are modeled as a random vector $\boldsymbol{y} : \Omega \to \mathcal{U}$ with joint PDF $\varrho : \mathcal{U} \to \mathbb{R}_+$ s.t. $\varrho(\boldsymbol{y}) = \prod_{i=1}^d \varrho_i(y_i)$.
  - **Goal**: Uncertainty quantification of u or some statistical QoI depending on u, i.e.,

$$\mathbb{E}[u],\ \mathbb{V}\mathrm{ar}[u],\ \mathbb{P}[u > u_0] = \mathbb{E}[\mathbb{1}_{\{u > u_0\}}].$$

UQ for parameterized PDE models

Some assumptions

Oak Ridge National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Continuity and coercivity **(CC)**

For all $x \in \overline{D}$ *and* $\boldsymbol{y} \in \mathcal{U}$, $0 < a_{\min} \leq a(x, \boldsymbol{y}) \leq a_{\max}$.

Analyticity **(AN)**

The complex continuation of a*, represented as the map* $a : \mathbb{C}^d \to L^\infty(D)$*, is an* $L^\infty(D)$*-valued analytic function on* $\mathbb{C}^d$.

Existence and uniqueness of solutions **(EU)**

For all $\boldsymbol{y} \in \mathcal{U}$ *the PDE problem admits an unique solution* $u \in \mathcal{V}$*, where* $\mathcal{V}$ *is a suitable finite or infinite dimensional Hilbert or Banach space. In addition*

$$\forall \boldsymbol{y} \in \mathcal{U}, \ \ \exists C(\boldsymbol{y}) > 0 \text{ such that } \|u(\boldsymbol{y})\|_{\mathcal{V}} \leq C(\boldsymbol{y})$$

Some simple consequences:

- The PDE induces a map $u = u(\boldsymbol{y}) : \mathcal{U} \to \mathcal{V}$.
- If $\int_{\mathcal{U}} C(\boldsymbol{y})^p \varrho(\boldsymbol{y}) d\boldsymbol{y} < \infty$ then $u \in L^p_\varrho(\mathcal{U}, \mathcal{V})$.

A simple illustrative example

Parameterized elliptic problems: $\mathcal{U} = [-1,1]^d$, $\mathcal{V} = H_0^1(D)$, $\varrho = 1/2^d$

$$\begin{cases} -\nabla \cdot (a(x,\boldsymbol{y})\nabla u(x,\boldsymbol{y})) & = f(x) \quad x \in D,\ \boldsymbol{y} \in \mathcal{U} \\ u(x,\boldsymbol{y}) & = 0 \quad\quad x \in \partial D,\ \boldsymbol{y} \in \mathcal{U} \end{cases}$$

Assume $a(x,\boldsymbol{y})$ satisfies **(CC)** and **(AN)**, and that $f \in L^2(D)$, then:

$$\forall y \in \mathcal{U}, \quad u(\boldsymbol{y}) \in H_0^1(D) \equiv \mathcal{V} \quad \text{and} \quad \|u(\boldsymbol{y})\|_{\mathcal{V}} \leq \frac{C_P}{a_{\min}} \|f\|_{L^2(D)}$$

- Lax-Milgram ensures the existence and uniqueness of solution $u \in L^2_\varrho(\mathcal{U},\mathcal{V})$.

Affine and non-affine coefficients:

1. $a(x,\boldsymbol{y}) = a_0(x) + \sum_{i=1}^d y_i \psi_i(x)$.
2. $a(x,\boldsymbol{y}) = a_0(x) + \left(\sum_{i=1}^d y_i \psi_i(x)\right)^q$, $q \in \mathbb{N}$.
3. $a(x,\boldsymbol{y}) = a_0(x) + \exp\left(\sum_{i=1}^d y_i \psi_i(x)\right)$ (e.g., truncated KL expansion in the log scale).

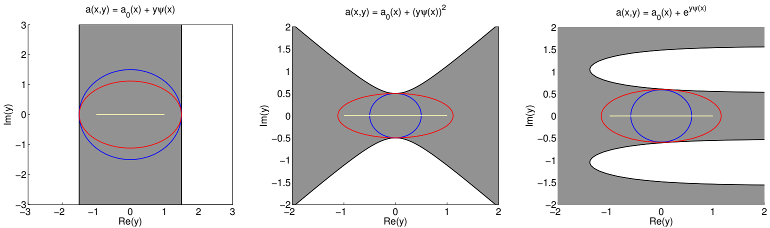
Remark. In what follows - can be extended to nonlinear elliptic (u^k), parabolic, and some hyperbolic PDEs, all defined on unbounded high-dimensional domains.

Analyticity of the solution

$\boldsymbol{\rho} = (\rho_i)_{1\le i\le d},\ \rho_i > 1\ \forall i$

- Polydisc: $\mathcal{O}_{\boldsymbol{\rho}} = \bigotimes_i \{z_i \in \mathbb{C}; |z_i| \le \rho_i\}$.
- Polyellipse: $\mathcal{E}_{\boldsymbol{\rho}} = \bigotimes_i \left\{ \frac{z_i + z_i^{-1}}{2}; z_i \in \mathbb{C}, |z_i| = \rho_i \right\}$.

Theorem. [Tran, W., Zhang '16]
Assume $a(x, \boldsymbol{y})$ satisfies **CC** and **AN**. Then the function $\boldsymbol{z} \mapsto u(\boldsymbol{z})$ is well-defined and analytic in an open neighborhood of some polyellipse $\mathcal{E}_{\boldsymbol{\rho}}$ (or polydisc $\mathcal{O}_{\boldsymbol{\rho}}$).



Domain of complex uniform ellipticity for some random fields.

Remark. The high-dimensional discontinuous case is analyzed in:
[Gunzburger, W., Zhang '14], [Burkardt, Gunzburger, W., Zhang '15 (SINUM), '16 (SIREV)]

Brief taxonomy of numerical strategies

Stochastic FEMs [Gunzburger, W., Zhang, '14 (Acta Numerica)]

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

- **Monte Carlo methods:** Let $\{\boldsymbol{y}_k \in \mathcal{U}\}_{k=1}^m$ denote a set of random sample points

$$\boxed{\mathbb{E}[u] = \frac{1}{m}\sum_{k=1}^{m} u(\boldsymbol{y}_k)}$$

  - Simple to implement, parallelize, and convergence rate is independent of d.
  - Asymptotic rate is $\mathcal{O}(1/\sqrt{m})$.
  - Unable to simultaneously approximate $\boldsymbol{y} \mapsto u(\boldsymbol{y})$.

- **Polynomial approximations**: Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) \in \Lambda \subset \mathbb{N}^d$ a multi-index set, and $\Psi_{\boldsymbol{\nu}}$ be multivariate polynomials in $\mathbb{P}_\Lambda(\mathcal{U}) = span\left\{\prod_{i=1}^d {y_i}^{\mu_i}, \mu_i \leq \nu_i \ \forall i\right\}$. Approximate the solution u by:

$$\boxed{u_\Lambda(x, \boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\Lambda} c_{\boldsymbol{\nu}}(x)\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}) \in \mathcal{V} \otimes \mathbb{P}_\Lambda(\mathcal{U})}$$

  - Takes advantage of the smoothness and/or the sparsity structure of u.
  - Can feature faster convergence than MC.
  - The evaluation of u_Λ requires the computation of $c_{\boldsymbol{\nu}}$ (in possibly) high-dimensions.

Brief taxonomy of numerical strategies

Stochastic FEMs [Gunzburger, W., Zhang, '14 (Acta Numerica)]

- **Monte Carlo methods:** Let $\{\boldsymbol{y}_k \in \mathcal{U}\}_{k=1}^m$ denote a set of random sample points

$$\boxed{\mathbb{E}[u] = \frac{1}{m}\sum_{k=1}^{m} u(\boldsymbol{y}_k)}$$

  - Simple to implement, parallelize, and convergence rate is independent of d.
  - Asymptotic rate is $\mathcal{O}(1/\sqrt{m})$.
  - Unable to simultaneously approximate $\boldsymbol{y} \mapsto u(\boldsymbol{y})$.

- **Polynomial approximations**: Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) \in \Lambda \subset \mathbb{N}^d$ a multi-index set, and $\boldsymbol{\Psi}_{\boldsymbol{\nu}}$ be multivariate polynomials in $\mathbb{P}_\Lambda(\mathcal{U}) = span\left\{\prod_{i=1}^d {y_i}^{\mu_i}, \mu_i \leq \nu_i \ \forall i\right\}$. Approximate the solution u by:

$$\boxed{u_\Lambda(x, \boldsymbol{y}) = \sum_{\boldsymbol{\nu} \in \Lambda} c_{\boldsymbol{\nu}}(x) \boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}) \in \mathcal{V} \otimes \mathbb{P}_\Lambda(\mathcal{U})}$$

  - Takes advantage of the smoothness and/or the sparsity structure of u.
  - Can feature faster convergence than MC.
  - The evaluation of u_Λ requires the computation of $c_{\boldsymbol{\nu}}$ (in possibly) high-dimensions.

Multivariate polynomial approximations

Major challenge: *curse of dimensionality*

1. **Taylor approximations:** [Cohen et. al. '10, '11; Tran, W., Zhang '14, '15]
   - $\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}) = \boldsymbol{y}^{\boldsymbol{\nu}}$ and $c_{\boldsymbol{\nu}} = \frac{1}{\boldsymbol{\nu}!}\partial^{\boldsymbol{\nu}} u(\mathbf{0})$ can be computed recursively.
   - useful when ψ_i have non-overlapping supports (affine "inclusion problems")

2. **Galerkin projection methods:** [Wiener '38, Ghanem, Spanos '99; Xiu, Karniadakis '02; Babuška et. al. '02; Todor, Schwab '03; Tran, W., Zhang '14; Dexter, W. '15]
   - $\{\boldsymbol{\Psi}_{\boldsymbol{\nu}}\}$ is a multivariate orthonormal polynomial basis in $\boldsymbol{y}$, e.g., Legendre polynomials, Hermite polynomials, etc.
   - u_Λ is the L^2_ϱ projection of u on $\mathbb{P}_\Lambda(\mathcal{U})$, with $\dim(\mathbb{P}_\Lambda) = \#(\Lambda) \equiv N$.
   - Couples the parametric and physical degrees of freedom.

3. **Interpolation methods:** [Smolyak, '63; Griebel et. al '99,'04; Nobile, Tempone, W. '08a, b; Jantsch, W., Zhang '13, '15; Gunzburger, Jantsch, Teckentrup, W., '15]
   - Given $m \geq \#(\Lambda)$ evaluations $\{u(\boldsymbol{y}_k)\}_{k=1}^m$, and $\{\boldsymbol{\Psi}_{\boldsymbol{\nu}}\}$ a Lagrange basis.
   - u_Λ is the interpolant of u over an associated grid (structured vs. unstructured).
   - Non-intrusive, sample-based approaches. Allow the use of legacy code.
   - May be unstable if the interpolation nodes are poorly chosen (i.e., $m >> \#(\Lambda)$).

Multivariate polynomial approximations

continued...

4. **Discrete least squares:** [Cohen et. al. '13; Migliorati et. al. '13, Narayan et. al. '13; Zhou et. al. '14; Chkifa et. al. '15]

   - Given m evaluations $\{u(\boldsymbol{y}_k)\}_{k=1}^m$, find $(c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda}$ by minimizing

     $$\sum_{k=1}^{m} \|u(\boldsymbol{y}_k) - u_\Lambda(\boldsymbol{y}_k)\|^2_{\mathcal{V},\ell^2}.$$

   - Mitigate Runge's phenomenon.
   - Reconstruct statistics of u, and stability of the design matrix requires $m >> \#(\Lambda)$.

5. **Compressed sensing:** [Doostan, Owhadi '11; Mathelin, Gallivan '12; Yang, Karniadakis '13; Rauhut, Schwab '14; Adcock '15, '16; Chkifa, Dexter, Tran, W. '15]

   - Given an enriched set Λ_0, and $m << \#(\Lambda_0)$ evaluations $\{u(\boldsymbol{y}_k)\}_{k=1}^m$, find $(c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\Lambda_0}$ by solving the following minimization problem:

     $$\text{argmin} \|\hat{c}_{\boldsymbol{\nu}}\|_{\mathcal{V},\ell^1(\Lambda_0)}, \text{ subject to } u(\boldsymbol{y}_k) = \sum_{\boldsymbol{\nu}\in\Lambda_0} \hat{c}_{\boldsymbol{\nu}}(x)\boldsymbol{\Psi}_{\boldsymbol{\nu}}(\boldsymbol{y}_k).$$

   - Number of samples to recover the best s-term scales linearly in s (up to log factors).
   - ℓ^1 minimization may be impractical in high dimensional problems.

Multivariate polynomial approximations

Selection of (**lower**) index sets in high-dimensions

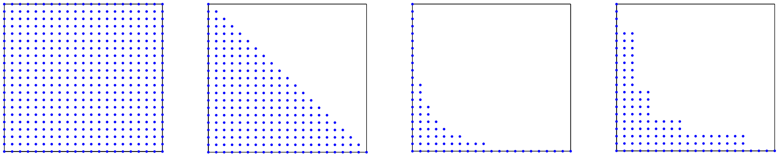
- The efficiency of polynomial approximations depends on the selection of $\Lambda \subset \mathbb{N}_0^d$.

- **Standard approaches:** impose lower, a.k.a., downward closed index sets Λ *a priori*,

$$\text{i.e.,} \qquad (\boldsymbol{\nu} \in \Lambda \ \text{ and } \ \boldsymbol{\nu}' \leq \boldsymbol{\nu}) \quad \Longrightarrow \quad \boldsymbol{\nu}' \in \Lambda.$$

  Challenge: $\#(\Lambda) = N$ can grow quickly w.r.t. the dimension d.

- Some most common choices of index sets Λ



Tensor Product $\Lambda(w) = \{\nu \in \mathbf{N}^N : \max_{1 \leq i \leq N} \nu_i \leq w\}$

Total Degree $\Lambda(w) = \{\nu \in \mathbf{N}^N : \sum \nu_i \leq w\}$

Hyperbolic Cross $\Lambda(w) = \{\nu \in \mathbf{N}^N : \prod (\nu_i + 1) \leq w + 1\}$

Smolyak $\Lambda(w) = \{\nu \in \mathbf{N}^N : \sum f(\nu_i) \leq f(w)\}$, with $f(\nu) = \lceil \log_2(\nu) \rceil$, $\nu \geq 2$.

- Best s-term approximations: the optimal set Λ_s^{best} of the s most effective indices.

Goals of this talk

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Present a random sampling polynomial strategy for recovering the s most effective indices (without imposing a subspace a priori), for both complex-valued, i.e., $c_{\boldsymbol{\nu}} \in \mathbb{C}^N$, and Hilbert-valued, i.e., $c_{\boldsymbol{\nu}} \in \mathcal{V}^N$, such as ℓ_1 minimization, and nonconvex optimization:

- Provide the optimal estimate of the minimum number of samples required by ℓ_1 minimization, to uniformly recover the best s-term approximation .
- Introduce a **specific choice of weights** for ℓ_1 minimization, and lower hard thresholding operators, that exploit the structure of the best s-term polynomial space to overcome the curse of dimensionality.

[Chkifa, Dexter, Tran, and W. '16]. *Polynomial approximation via compressed sensing of high-dimensional functions on lower sets*. Mathematics of Computation (arXiv:1602.05823), 2016.

- Provide a unified null space property (NSP)-based condition for a general class of nonconvex minimizations, proving they are at least as good as ℓ_1 minimization in exact recovery of sparse signals.

[Tran, and W. '16]. *New sufficient conditions for sparse recovery via nonconvex minimizations*, 2016.

Compressed sensing recovery of best s-term approximations

Novel ℓ_1-minimization and hard thresholding approaches for signal recovery

Goal. Given an index set Λ_0 with $\#(\Lambda_0) = N \gg s$ (could be far from optimal), $\{\Psi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu}\in\Lambda_0}$ an $L^2(\mathcal{U}, d\varrho)$-orthonormal basis, find and approximation of u:

$$\boxed{u(\boldsymbol{y}) \approx \sum_{\boldsymbol{\nu}\in\Lambda_0} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}(\boldsymbol{y}).}$$

CS was initially developed for signal recovery [Candès, Romberg, Tao '06; Donoho '06].

- generate m samples $\{\boldsymbol{y}_k\}_{k=1}^m$ according to the orthogonalization measure $\varrho(\boldsymbol{y})$.
- $m \times N$ sampling matrix and observations:

$$\boldsymbol{\Psi} := \left(\Psi_{\boldsymbol{\nu}_j}(\boldsymbol{y}_i)\right)_{\substack{1\le i\le m,\\ 1\le j\le N}} \quad \text{and} \quad \boldsymbol{u} := (u(\boldsymbol{y}_1), \ldots, u(\boldsymbol{y}_m)).$$

- the coefficient $\boldsymbol{c} = (c_{\boldsymbol{\nu}_j})_{1\le j\le N} \in \mathbb{C}^N$ is a solution of the linear system

$$\boldsymbol{u} = \boldsymbol{\Psi}\boldsymbol{c}. \tag{1}$$

- to construct approximation comparable to the best s-term approximation on Λ_s^{best} requires only $m \approx s \ll N$ samples, and thus the system (1) is **underdetermined**.
- (1) can be solved using optimization, greedy, or thresholding approaches.

Sparse recovery via compressed sensing

Problem. Given $\boldsymbol{\Psi} \in \mathbb{C}^{m \times N}$ $(m \ll N)$, reconstruct an unknown s-sparse vector $\boldsymbol{c} \in \mathbb{C}^N$ from the measurements $\boldsymbol{u} = \boldsymbol{\Psi}\boldsymbol{c}$:

$$\boldsymbol{c} \text{ is the unique sparest solution to } \boldsymbol{\Psi}\boldsymbol{c} = \boldsymbol{u}$$

$$\Updownarrow$$

$$\boldsymbol{c} = \underset{\boldsymbol{z} \in \mathbb{C}^N}{\text{argmin}} \|\boldsymbol{z}\|_{\ell_0} \text{ subject to } \boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u},$$

where $\|\boldsymbol{z}\|_{\ell_0} = \sum_{\nu \in \Lambda_0} |z_\nu|^0 = \#\{\Lambda_0 : z_\nu \neq 0\}$ is the sparsity of $\boldsymbol{c}$.

- When $\boldsymbol{c}$ is sparse, ℓ_0 minimization is often correct, but is computationally intractable (an NP-hard problem in general $\rightarrow$ in part due to nonconvex nature).
- Require $m \gg N$ (overdetermined system) to solve (1) with ℓ_2 (least squares).
- Can we split the difference?

Compressed sensing recovery using ℓ_1 minimization

Convex ℓ_1 regularizations: $\underset{\boldsymbol{z}\in\mathbb{C}^N}{\text{minimize}} \; \|\boldsymbol{z}\|_{\ell_1}$ subject to $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$.

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

$\tilde{\boldsymbol{\Psi}}$ satisfies the RIP if there exists small δ_s, s.t. for all $\mathbf{c}$ s-sparse vectors,

$$(1-\delta_s)\|\mathbf{c}\|_{\ell_2}^2 \le \|\tilde{\boldsymbol{\Psi}}\mathbf{c}\|_{\ell_2}^2 \le (1+\delta_s)\|\mathbf{c}\|_{\ell_2}^2.$$

Define $\Theta = \sup_{\boldsymbol{\nu}\in\Lambda_0} \|\Psi_{\boldsymbol{\nu}}\|_\infty$ (uniform bound of the orthonormal system).

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

$$m \ge C\,\Theta^2 s\,\log^3(s)\,\log(N) \qquad \text{then,}$$

with high probability, the RIP for $\tilde{\boldsymbol{\Psi}}$ is satisfied, and uniform recovery is guaranteed.

Original proof developed through a series of papers [Candes, Tao '06; Rudelson, Vershynin '08; Rauhut '10; Cheraghchi, Guruswami, Velingker '13].

RIP for bounded orthonormal systems (BOS)

Sample complexity: Previous available estimates

Define $\Theta = \sup_{\boldsymbol{\nu} \in \Lambda_0} \|\Psi_{\boldsymbol{\nu}}\|_\infty$ (uniform bound of the orthonormal system).

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

$$m \geq C\,\Theta^2 s\,\log^3(s)\log(N) \qquad \text{then,}$$

with high probability, the RIP for $\tilde{\boldsymbol{\Psi}}$ is satisfied, and uniform recovery is guaranteed.

Original proof developed through a series of papers [Candes, Tao '06; Rudelson, Vershynin '08; Rauhut '10; Cheraghchi, Guruswami, Velingker '13].

Room for improvement:

- Θ can be prohibitively BIG due to the curse of dimensionality: Chebyshev: $\Theta \sim 2^{d/2}$; Legendre: $\Theta \sim N$; Pre-conditioned Legendre [Rauhut, Ward '13]: $\Theta \sim 3^{d/2}$.
- Non-uniform recovery guarantees: $m \geq C\Theta^2 s \log(s) \log(N)$.
- Exploit any structure of the Λ_0 to reduce dependence on N.

RIP for bounded orthonormal systems (BOS)

Sample complexity: Previous available estimates

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Define $\Theta = \sup_{\boldsymbol{\nu} \in \Lambda_0} \|\Psi_{\boldsymbol{\nu}}\|_\infty$ (uniform bound of the orthonormal system).

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

$$m \geq C\,\Theta^2 s\,\log^3(s)\log(N) \qquad \text{then,}$$

with high probability, the RIP for $\tilde{\boldsymbol{\Psi}}$ is satisfied, and uniform recovery is guaranteed.

Original proof developed through a series of papers [Candes, Tao '06; Rudelson, Vershynin '08; Rauhut '10; Cheraghchi, Guruswami, Velingker '13].

Room for improvement:

- Θ can be prohibitively BIG due to the curse of dimensionality: Chebyshev: $\Theta \sim 2^{d/2}$; Legendre: $\Theta \sim N$; Pre-conditioned Legendre [Rauhut, Ward '13]: $\Theta \sim 3^{d/2}$.
- Non-uniform recovery guarantees: $m \geq C\Theta^2 s \log(s)\log(N)$.
- Exploit any structure of the Λ_0 to reduce dependence on N.

RIP for bounded orthonormal systems (BOS)

Sample complexity: Previous available estimates

Define $\Theta = \sup_{\boldsymbol{\nu} \in \Lambda_0} \|\Psi_{\boldsymbol{\nu}}\|_\infty$ (uniform bound of the orthonormal system).

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

$$m \geq C\,\Theta^2 s \log^3(s) \log(N) \qquad \text{then,}$$

with high probability, the RIP for $\tilde{\boldsymbol{\Psi}}$ is satisfied, and uniform recovery is guaranteed.

Original proof developed through a series of papers [Candes, Tao '06; Rudelson, Vershynin '08; Rauhut '10; Cheraghchi, Guruswami, Velingker '13].

Room for improvement:

- Θ can be prohibitively BIG due to the curse of dimensionality: Chebyshev: $\Theta \sim 2^{d/2}$; Legendre: $\Theta \sim N$; Pre-conditioned Legendre [Rauhut, Ward '13]: $\Theta \sim 3^{d/2}$.
- Non-uniform recovery guarantees: $m \geq C\Theta^2 s \log(s) \log(N)$.
- Exploit any structure of the Λ_0 to reduce dependence on N.

RIP for bounded orthonormal systems (BOS)

Sample complexity: Previous available estimates

Define $\Theta = \sup_{\boldsymbol{\nu} \in \Lambda_0} \|\Psi_{\boldsymbol{\nu}}\|_\infty$ (uniform bound of the orthonormal system).

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

$$m \geq C\,\Theta^2 s\,\log^3(s)\log(N) \qquad \text{then,}$$

with high probability, the RIP for $\tilde{\boldsymbol{\Psi}}$ is satisfied, and uniform recovery is guaranteed.

Original proof developed through a series of papers [Candes, Tao '06; Rudelson, Vershynin '08; Rauhut '10; Cheraghchi, Guruswami, Velingker '13].

Room for improvement:

- Θ can be prohibitively BIG due to the curse of dimensionality: Chebyshev: $\Theta \sim 2^{d/2}$; Legendre: $\Theta \sim N$; Pre-conditioned Legendre [Rauhut, Ward '13]: $\Theta \sim 3^{d/2}$.
- Non-uniform recovery guarantees: $m \geq C\Theta^2 s\log(s)\log(N)$.
- Exploit any structure of the Λ_0 to reduce dependence on N.

RIP for bounded orthonormal systems

An improved sample complexity estimate

Goal: Establish a reduced sample complexity, associated with novel CS strategies, that:

1. Improves the logarithmic factor by at least one unit.
2. Removes the constraint of m on $\Theta^2 s$.
3. Optimizes the cardinality N of Λ_0 (which also speeds up mat-vec multiplication).

Theorem [Chkifa, Dexter, Tran, W. '16]. Let $\Psi \in \mathbb{C}^{m\times N}$ be the random sampling matrix associated with a BOS. For $\delta \in (0,1)$,

$$\text{if} \qquad m \geq C_\delta \Theta^2 s \log^2(s)\log(N), \qquad \text{then,}$$

with high probability, $\tilde{\Psi}$ satisfies the restricted isometry property with $\delta_s \leq \delta$.

- Best available estimate for m that improves upon [Rudelson, Vershynin '08; Bourgain '14] in the power of the logarithm's.
- Extends the chaining arguments in [Bourgain '14; Haviv, Regev '15].

RIP for bounded orthonormal systems

An improved sample complexity estimate

Goal: Establish a reduced sample complexity, associated with novel CS strategies, that:

1. Improves the logarithmic factor by at least one unit.
2. Removes the constraint of m on $\Theta^2 s$.
3. Optimizes the cardinality N of Λ_0 (which also speeds up mat-vec multiplication).

Theorem [Chkifa, Dexter, Tran, W. '16]. Let $\mathbf{\Psi} \in \mathbb{C}^{m\times N}$ be the random sampling matrix associated with a BOS. For $\delta \in (0,1)$,

$$\text{if} \qquad m \geq C_\delta \Theta^2 s \log^2(s) \log(N), \qquad \text{then,}$$

with high probability, $\tilde{\mathbf{\Psi}}$ satisfies the restricted isometry property with $\delta_s \leq \delta$.

- Best available estimate for m that improves upon [Rudelson, Vershynin '08; Bourgain '14] in the power of the logarithm's.
- Extends the chaining arguments in [Bourgain '14; Haviv, Regev '15].

RIP for bounded orthonormal systems

Sketch of the proof

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Basic goal. To prove that that random matrix $\tilde{\mathbf{\Psi}}$ of size $m \times N$ satisfies with high probability that for all s-sparse $\mathbf{c} \in \mathbb{C}^N$,

$$\|\tilde{\mathbf{\Psi}}\mathbf{c}\|_2^2 \approx \|\mathbf{c}\|_2^2.$$

- Basic strategy is similar to [Baraniuk, Davenport, DeVore, Wakin '08; Bourdain '14; Haviv, Regev '15] with some fundamentally new tricks.

1. Construct a class $\mathcal{F}$ of "discrete" approximations of ψ:
   $\rightarrow$ for all $\mathbf{c}$ s-sparse, there exists $\widehat{\psi}_{\mathbf{c}} \in \mathcal{F}$ such that $\widehat{\psi}_{\mathbf{c}}(\,\cdot\,) \approx \psi(\,\cdot\,,\mathbf{c})$
   $\rightarrow$ $\widehat{\psi}_{\mathbf{c}}$ can be decomposed as $\widehat{\psi}_{\mathbf{c}} = \sum_j \widehat{\psi}_{\mathbf{c}}^j$, each $\widehat{\psi}_{\mathbf{c}}^j$ is an indicator function and represents a scale of $\widehat{\psi}_{\mathbf{c}}$.
2. Basic tail estimate gives: $\forall j$, $\forall \mathbf{c}$ s-sparse, with high probability

$$\frac{1}{m}\sum_{i=1}^{m} |\widehat{\psi}_{\mathbf{c}}^j(\boldsymbol{y}_i)|^2 \approx \int_{\mathcal{U}} |\widehat{\psi}_{\mathbf{c}}^j(\boldsymbol{y})|^2.$$

To prove $(\star)$, we apply union bound to show with high probability

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{j} |\widehat{\psi}_{\mathbf{c}}^j(\boldsymbol{y}_i)|^2 \approx \int_{\mathcal{U}} \sum_{j} |\widehat{\psi}_{\mathbf{c}}^j(\boldsymbol{y})|^2 \ \textbf{uniformly in } \mathbf{c}.$$

RIP for bounded orthonormal systems

Sketch of the proof

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

$\psi(\boldsymbol{y},\mathbf{c}) := \sum_{\boldsymbol{\nu}\in\Lambda_0} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})\ \forall \boldsymbol{y}\in\mathcal{U}$, $\mathcal{B}_s := \{\mathbf{c}\in\mathbb{C}^N : \mathbf{c}\ s\text{-sparse},\ \|\mathbf{c}\|_2 = 1\}$.

Goal: For randomly sampling points $\{\boldsymbol{y}_i\}_{i=1}^m$, prove that with high probability

$$\frac{1}{m}\sum_{i=1}^m |\psi(\boldsymbol{y}_i,\mathbf{c})|^2 \approx \int_{\mathcal{U}} |\psi(\boldsymbol{y},\mathbf{c})|^2,\quad \forall \mathbf{c}\in\mathcal{B}_s. \qquad (\star)$$

1. Construct a class $\mathcal{F}$ of "discrete" approximations of ψ:
   $\to$ for all $\mathbf{c}$ s-sparse, there exists $\widehat{\psi}_{\mathbf{c}}\in\mathcal{F}$ such that $\widehat{\psi}_{\mathbf{c}}(\cdot)\approx\psi(\cdot,\mathbf{c})$
   $\to$ $\widehat{\psi}_{\mathbf{c}}$ can be decomposed as $\widehat{\psi}_{\mathbf{c}} = \sum_j \widehat{\psi}_{\mathbf{c}}^j$, each $\widehat{\psi}_{\mathbf{c}}^j$ is an indicator function and represents a scale of $\widehat{\psi}_{\mathbf{c}}$.
2. Basic tail estimate gives: $\forall j$, $\forall \mathbf{c}$ s-sparse, with high probability

$$\frac{1}{m}\sum_{i=1}^m |\widehat{\psi}_{\mathbf{c}}^j(\boldsymbol{y}_i)|^2 \approx \int_{\mathcal{U}} |\widehat{\psi}_{\mathbf{c}}^j(\boldsymbol{y})|^2.$$

   To prove $(\star)$, we apply union bound to show with high probability

$$\frac{1}{m}\sum_{i=1}^m\sum_j |\widehat{\psi}_{\mathbf{c}}^j(\boldsymbol{y}_i)|^2 \approx \int_{\mathcal{U}}\sum_j |\widehat{\psi}_{\mathbf{c}}^j(\boldsymbol{y})|^2 \ \textbf{uniformly in } \mathbf{c}.$$

3. Control $\#\{\text{all } \widehat{\psi}_{\mathbf{c}}^j : \mathbf{c}\in\mathcal{B}_s\}$ by using a covering number argument.

RIP for bounded orthonormal systems

Sketch of the proof

$\psi(\boldsymbol{y},\mathbf{c}) := \sum_{\boldsymbol{\nu}\in\Lambda_0} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}(\boldsymbol{y})\ \forall \boldsymbol{y}\in\mathcal{U}$, $\mathcal{B}_s := \{\mathbf{c}\in\mathbb{C}^N : \mathbf{c}\ s\text{-sparse},\ \|\mathbf{c}\|_2 = 1\}$.

Goal: For randomly sampling points $\{\boldsymbol{y}_i\}_{i=1}^m$, prove that with high probability

$$\frac{1}{m}\sum_{i=1}^m |\psi(\boldsymbol{y}_i,\mathbf{c})|^2 \approx \int_{\mathcal{U}} |\psi(\boldsymbol{y},\mathbf{c})|^2, \quad \forall \mathbf{c}\in\mathcal{B}_s. \qquad (\star)$$

1. Construct a class $\mathcal{F}$ of "discrete" approximations of ψ:
   - $\rightarrow$ for all $\mathbf{c}$ s-sparse, there exists $\widehat{\psi}_{\mathbf{c}} \in \mathcal{F}$ such that $\widehat{\psi}_{\mathbf{c}}(\cdot) \approx \psi(\cdot,\mathbf{c})$
   - $\rightarrow$ $\widehat{\psi}_{\mathbf{c}}$ can be decomposed as $\widehat{\psi}_{\mathbf{c}} = \sum_j \widehat{\psi}^j_{\mathbf{c}}$, each $\widehat{\psi}^j_{\mathbf{c}}$ is an indicator function and represents a scale of $\widehat{\psi}_{\mathbf{c}}$.
2. Basic tail estimate gives: $\forall j$, $\forall \mathbf{c}$ s-sparse, with high probability

$$\frac{1}{m}\sum_{i=1}^m |\widehat{\psi}^j_{\mathbf{c}}(\boldsymbol{y}_i)|^2 \approx \int_{\mathcal{U}} |\widehat{\psi}^j_{\mathbf{c}}(\boldsymbol{y})|^2.$$

   To prove $(\star)$, we apply union bound to show with high probability

$$\frac{1}{m}\sum_{i=1}^m \sum_j |\widehat{\psi}^j_{\mathbf{c}}(\boldsymbol{y}_i)|^2 \approx \int_{\mathcal{U}} \sum_j |\widehat{\psi}^j_{\mathbf{c}}(\boldsymbol{y})|^2 \ \textbf{uniformly in } \mathbf{c}.$$

3. Control $\#\{\text{all } \widehat{\psi}^j_{\mathbf{c}} : \mathbf{c}\in\mathcal{B}_s\}$ by using a covering number argument.

RIP for bounded orthonormal systems

Sketch of the proof

- Straightforward extension from the unitary case in [Bourgain '14]:

  $\#\{\widehat{\psi}^j_{\mathbf{c}} : \mathbf{c} \in \mathcal{B}_s\}$ scales like the covering number of $\mathcal{B}_s$ under the pseudo-metric

  $$d(\mathbf{c}, \mathbf{c}') = \sup_{\boldsymbol{y} \in \mathcal{U}} |\psi(\boldsymbol{y}, \mathbf{c} - \mathbf{c}')|.$$

  However, in the covering number may grow **exponentially** with d in our case.

- We ignore "bad" $\boldsymbol{y} \in \mathcal{U}$ which may make $|\psi(\boldsymbol{y}, \mathbf{c} - \mathbf{c}')|$ big, and come up with following "distance" (not a proper pseudo-metric)

  $$d_\varrho(\mathbf{c}, \mathbf{c}') = \inf_{\substack{\widetilde{\mathcal{U}} \subset \mathcal{U} \\ |\widetilde{\mathcal{U}}| = 1-\varrho}} \sup_{\boldsymbol{y} \in \widetilde{\mathcal{U}}} |\psi(\boldsymbol{y}, \mathbf{c} - \mathbf{c}')|.$$

  - $\widehat{\psi}_{\mathbf{c}}(\,\cdot\,)$ now agrees with $\psi(\,\cdot\,, \mathbf{c})$ for all except the "bad" $\boldsymbol{y}$.
  - However, since such $\boldsymbol{y}$ only constitutes a small fraction of $\mathcal{U}$, their elimination does not affect the overall RIP estimate.
  - The result: the covering number is reduced by $\log(s)$ and does **not** depend on d.

Recovery of best lower s-term approximations

Reduce the dimension-dependence of m on $\Theta^2 s$

When the solutions of parameterized PDEs are smooth, the set of best s-terms are often defined on an (approximately) lower set.

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.

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.
- We can choose the enriched set Λ_0 as a hyperbolic cross $\mathcal{H}_s$, which is the union of all lower sets of cardinality s, i.e.,

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

Note: $N = \#(\mathcal{H}_s) \leq 2s^3 4^d$ [Dũng '13; Dũng, Griebel '15].

Recovery of best lower s-term approximations

lower-RIP $\equiv \ell_{\omega,1}$ with $\omega_{\boldsymbol{\nu}} = \|\Psi_{\boldsymbol{\nu}}\|_\infty$

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

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

lower-RIP: There exists small $\delta_{l,s}$ such that

$$(1-\delta_{l,s})\|\mathbf{c}\|_2^2 \le \|\tilde{\boldsymbol{\Psi}}\mathbf{c}\|_2^2 \le (1+\delta_{l,s})\|\mathbf{c}\|_2^2.$$

for all $\mathbf{c}$ satisfying $\tilde{K}(supp(\mathbf{c})) \le K(s)$ $(\dagger)$.

- A particular subclass of $(\dagger)$ is the set of all $\mathbf{c}$ s-sparse, $supp(\mathbf{c})$ lower.
- Lower-RIP can be considered a special case of weighted RIP [Rauhut, Ward '15] with weight $\omega_{\boldsymbol{\nu}} = \|\Psi_{\boldsymbol{\nu}}\|_\infty$ (see also [Adcock '15, '16]).

Theorem [Chkifa, Dexter, Tran, W. '16]. 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, $\tilde{\boldsymbol{\Psi}}$ satisfies the **lower-RIP** with $\delta_{l,s} \le \delta$.

Recovery of best lower s-term approximations

Summary - what have we achieved?

- Our sufficient condition for best lower s-term reconstruction

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

Estimates of $K(s)$ [Chkifa, Cohen, Migliorati, Nobile, Tempone '15], and $\#(\mathcal{H}_s)$ yield:

$$m \geq \begin{cases} Cs^{\frac{\log 3}{\log 2}} \log^2(s)(\log(s)+d), & \text{if } (\boldsymbol{\Psi_\nu}) \text{ is Chebyshev basis,} \\ Cs^2 \log^2(s)(\log(s)+d), & \text{if } (\boldsymbol{\Psi_\nu}) \text{ is Legendre basis.} \end{cases}$$

- Previous (well-known) sufficient condition for best s-term reconstruction

$$m \geq C\Theta^2 s\log^3(s)\log(N).$$

Estimates of Θ on $\#(\mathcal{H}_s)$ give:

$$\Theta^2 s \geq \begin{cases} \frac{1}{2}s^2, & \text{if } (\boldsymbol{\Psi_\nu}) \text{ is Chebyshev basis,} \\ \frac{1}{3}s^{\frac{\log 3}{\log 2}+1}, & \text{if } (\boldsymbol{\Psi_\nu}) \text{ is Legendre basis.} \end{cases}$$

Recovery of best lower s-term approximations

Implementation: Weighted ℓ_1 minimization

Let $\boldsymbol{\omega} = (\omega_j)_{j\in\Lambda_0}$ be a vector of weights. We define

- for $f(\boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\Lambda_0} t_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})$, $\|f\|_{\omega,1} := \sum_{\boldsymbol{\nu}\in\Lambda_0} \omega_{\boldsymbol{\nu}}|t_{\boldsymbol{\nu}}|$,
- $\sigma_s^{(\ell)}(f)_{\omega,1} = \inf\limits_{\substack{supp(g) \text{ lower} \\ |supp(g)|=s}} \|f-g\|_{\omega,1}$.

Weighted ℓ_1 minimization:

- Choose our specific weight $\omega_{\boldsymbol{\nu}} = \|\Psi_{\boldsymbol{\nu}}\|_\infty$,
- $\boldsymbol{\Psi} = (\Psi_{\boldsymbol{\nu}}(\boldsymbol{y}_i))$ is an $m\times N$ sampling matrix,
- $\mathbf{u} = (u(\boldsymbol{y}_i))_{i=1,\dots,m}$,
- η is some estimate of the tail expansion.

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

$$\min \sum_{\boldsymbol{\nu}\in\Lambda_0} \omega_{\boldsymbol{\nu}}|z_{\boldsymbol{\nu}}| \quad \text{subject to } \|\mathbf{u}-\boldsymbol{\Psi}\mathbf{z}\|_2 \le \eta\sqrt{m}.$$

Recovery of best lower s-term approximations

Weighted ℓ_1 minimization

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Theorem [Chkifa, Dexter, Tran, W. '16]. Assume that the number of samples satisfies

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

then, with high probability, there holds

$$\|u - u^{\#}\|_{\omega,1} \leq c_1\sigma_s^{(\ell)}(u)_{\omega,1} + c_2\eta\sqrt{K(s)}, \quad \text{if upper bound of the tail is available,}$$

$$\|u - u^{\#}\|_{\omega,1} \leq c_1\sigma_s^{(\ell)}(u)_{\omega,1}, \quad \text{if accurate estimate of the tail is available.}$$

- We propose a specific choice of weights which can lead to an approximation that:
  - has reduced sample complexity compared to unweighted ℓ_1 minimization.
  - is comparable to best s-term approximation (for smooth solutions).
- Best available estimate with specific improvements [Rauhut, Ward '15]:
  - similar estimate exists **but** only true for the **best weighted s-term** approximation, which is much weaker and incomparable to the required best s-term approximation.
  - our enriched set has minimized cardinality and **NOT** depends on the weight.

Numerical examples: weighted ℓ_1 recovery of functions

Spectral projected-gradient for basis pursuit

Find $u^{\#}(\boldsymbol{y}) = \sum_{\boldsymbol{\nu} \in \mathcal{H}_s} c_{\boldsymbol{\nu}} \Psi_{\boldsymbol{\nu}}(\boldsymbol{y})$, where $\mathbf{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu} \in \mathcal{H}_s}$ is the solution of

$$\min \sum_{\boldsymbol{\nu} \in \mathcal{H}_s} \omega_{\boldsymbol{\nu}} |z_{\boldsymbol{\nu}}| \quad \text{subject to } \|\mathbf{u} - \boldsymbol{\Psi}\mathbf{z}\|_2 \leq \eta\sqrt{m} = \sigma, \tag{2}$$

for several illustrative example functions.

To solve (2), we rely on our version of the code SPGL1 [van den Berg and Friedlander '07] - implementation of the spectral projected-gradient (SPG) algorithm:

- Fix a-priori the cardinality $N = \#(\mathcal{H}_s)$ of the Hyperbolic Cross subspace in which we wish to approximate our function.
- Increase the number of random samples m up to $m_{\max}$.
- Fix the seed of the random number generator for each choice of weight $\omega_{\boldsymbol{\nu}}$ and m so that we can compare the relative performance.
- Run 50 random trials for each pair $(m/N, \omega_{\boldsymbol{\nu}})$.

Note: In all examples we use a Legendre expansion $\Rightarrow$ our proposed weight is:

$$\omega_{\boldsymbol{\nu}_j} = \|\Psi_{\boldsymbol{\nu}_j}\|_\infty = \sqrt{2\boldsymbol{\nu}_j + 1}$$

Numerical examples: weighted ℓ_1 recovery of functions

Spectral projected-gradient for basis pursuit

Find $u^{\#}(\boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\mathcal{H}_s} c_{\boldsymbol{\nu}}\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})$, where $\mathbf{c} = (c_{\boldsymbol{\nu}})_{\boldsymbol{\nu}\in\mathcal{H}_s}$ is the solution of

$$\min \sum_{\boldsymbol{\nu}\in\mathcal{H}_s} \omega_{\boldsymbol{\nu}}|z_{\boldsymbol{\nu}}| \quad \text{subject to } \|\mathbf{u} - \boldsymbol{\Psi}\mathbf{z}\|_2 \leq \eta\sqrt{m} = \sigma, \tag{2}$$

for several illustrative example functions.

To solve (2), we rely on our version of the code SPGL1 [van den Berg and Friedlander '07] - implementation of the spectral projected-gradient (SPG) algorithm:

- Fix a-priori the cardinality $N = \#(\mathcal{H}_s)$ of the Hyperbolic Cross subspace in which we wish to approximate our function.
- Increase the number of random samples m up to $m_{\max}$.
- Fix the seed of the random number generator for each choice of weight $\omega_{\boldsymbol{\nu}}$ and m so that we can compare the relative performance.
- Run 50 random trials for each pair $(m/N, \omega_{\boldsymbol{\nu}})$.

Note: In all examples we use a Legendre expansion $\Rightarrow$ our proposed weight is:

$$\omega_{\boldsymbol{\nu}_j} = \|\Psi_{\boldsymbol{\nu}_j}\|_\infty = \sqrt{2\boldsymbol{\nu}_j + 1}$$

Example 1: $d = 8$

Sparsity of the coefficients

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

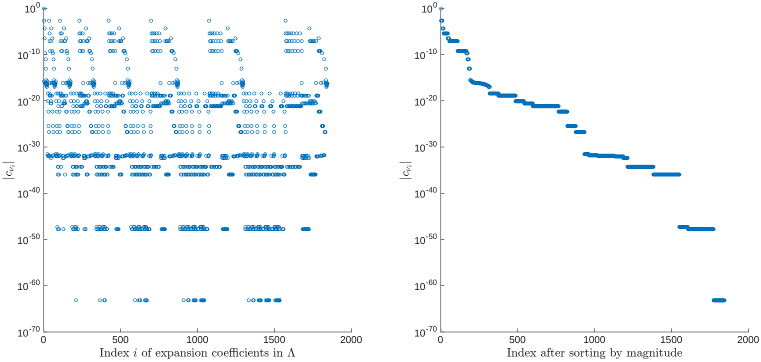


Figure In $d = 8$, $N = 1843$: **(left)** Magnitude of the Legendre expansion coefficients of $u(\boldsymbol{y})$. **(right)** Coefficients of $u(\boldsymbol{y})$ sorted by magnitude.

Example 1: $d = 8$

Convergence of the weighted ℓ_1 algorithm

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

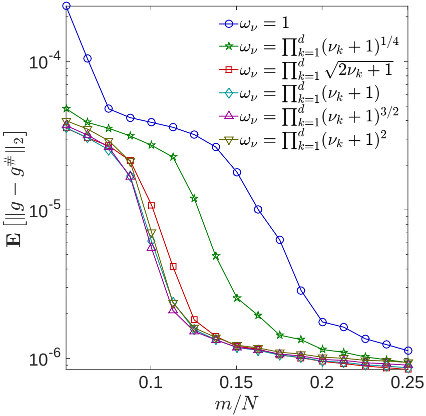


Figure In $d = 8$, $N = 1843$: average error measured in the L^2_ϱ norm.

Example 1: $d = 16$

Sparsity of the coefficients

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

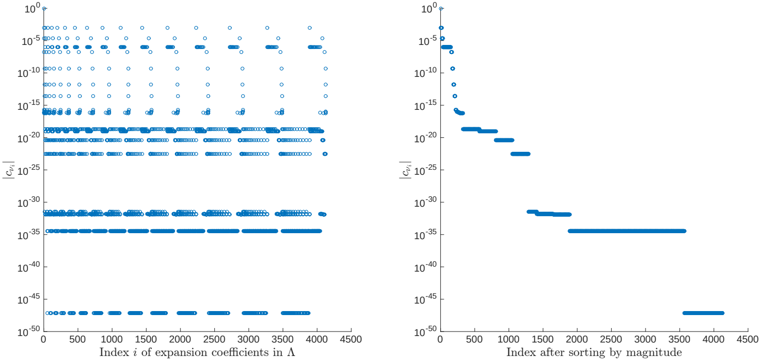


Figure In $d = 16$, $N = 4129$: **(left)** Magnitude of the Legendre expansion coefficients of $u(\boldsymbol{y})$. **(right)** Coefficients of $u(\boldsymbol{y})$ sorted by magnitude.

Example 1: $d = 16$

Convergence of the weighted ℓ_1 algorithm

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

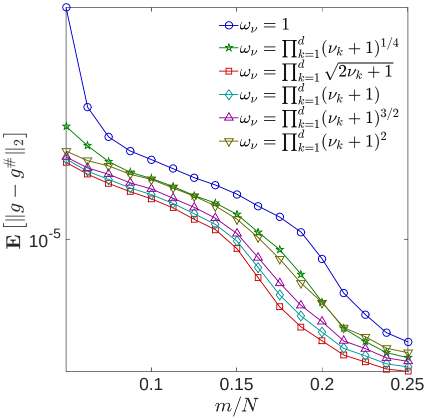


Figure In $d = 16$, $N = 4129$: average error measured in the L^2_ϱ norm.

Example 2: $d = 8$

Sparsity of the coefficients

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

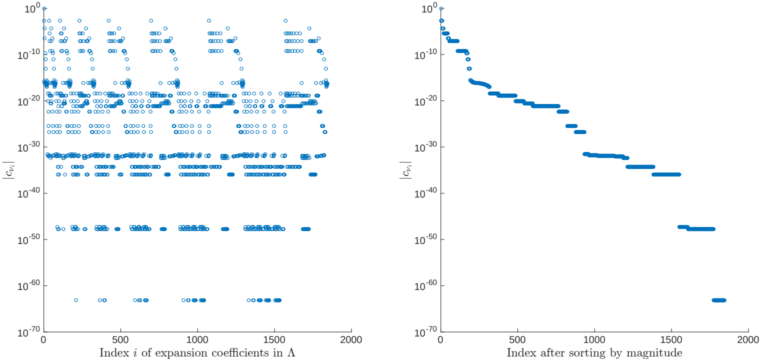


Figure In $d = 8$, $N = 1843$: **(left)** Magnitude of the Legendre expansion coefficients of $u(\boldsymbol{y})$. **(right)** Coefficients of $u(\boldsymbol{y})$ sorted by magnitude.

Example 2: $d = 8$

Convergence of the weighted ℓ_1 algorithm

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

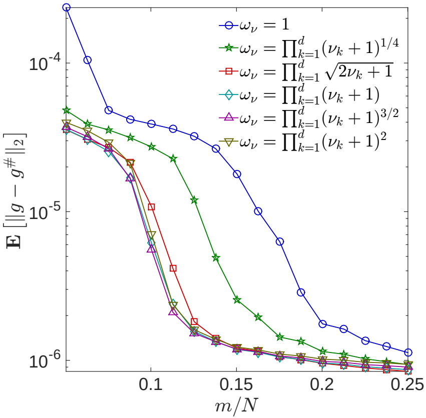


Figure In $d = 8$, $N = 1843$: average error measured in the L^2_ϱ norm.

Example 2: $d = 16$

Sparsity of the coefficients

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

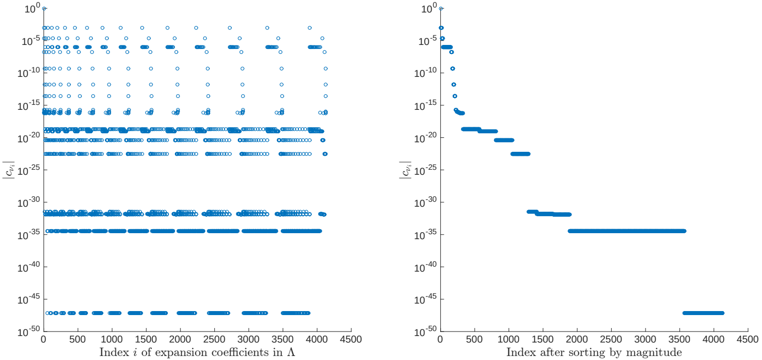


Figure In $d = 16$, $N = 4129$: **(left)** Magnitude of the Legendre expansion coefficients of $u(\boldsymbol{y})$. **(right)** Coefficients of $u(\boldsymbol{y})$ sorted by magnitude.

Example 2: $d = 16$

Convergence of the weighted ℓ_1 algorithm

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

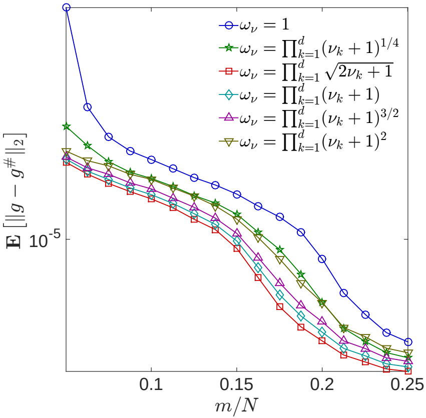


Figure In $d = 16$, $N = 4129$: average error measured in the L^2_ϱ norm.

Compressed sensing for parametric PDE recovery

Direct application of compressed sensing techniques to parameterized PDEs

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

"Uncoupled" approach

Given a point $x^* \in D$ in physical space, evaluate $\boldsymbol{u}_k^* = u(x^*, \boldsymbol{y}_k)$ (or some functional $G(u)$) at the points $\{\boldsymbol{y}_k\}_{k=1}^m$, and solve the basis pursuit denoising problem: find

$$\boldsymbol{c}^* = \operatorname{argmin}_{\boldsymbol{z}\in\mathbb{C}^N} \|\boldsymbol{z}\|_1 \quad \text{subject to } \|\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}^*\|_2 \leq \eta^*/\sqrt{m}. \tag{3}$$

The resulting solutions are an approximation to $u(x^*, \boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\Lambda_0} c_{\boldsymbol{\nu}}(x^*)\Psi_{\boldsymbol{\nu}}(\boldsymbol{y})$, [Doostan, Owhadi '11; Mathelin, Gallivan '12; Yan, Guo, Xiu '12; Yang, Karniadakis '13; Peng, Hampton, Doostan '14; Rauhut, Schwab '14; Hampton, Doostan '15; Narayan, Zhou '15].

- Construct the functions $\boldsymbol{c}(x)$ in the entire D using numerical methods such as piecewise polynomial interpolation, least square regression, etc.
- Here we require η^*, an estimate of the tail $\|u_{\Lambda_0^c}(x^*)\|_2$, a point-wise estimate.
- In parameterized PDEs, $c_\nu = c_\nu(x)$ is a function in D, and belongs to a Hilbert space $\mathcal{V} \Rightarrow \mathbf{c} \in \mathcal{V}^N$, equipped with the norm: $\|\boldsymbol{c}\|_{\mathcal{V},p} = (\sum_{i=1}^N \|c_i\|_{\mathcal{V}}^p)^{1/p}$.

Compressed sensing for parametric PDE recovery

Hilbert-valued approach for recovering the solution of a parameterized PDE

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

RIP for Hilbert-valued functions: For all $\mathbf{c} \in \mathcal{V}^N$ with $\|\mathbf{c}\|_{\mathcal{V},0} \leq s$, there exists $\delta_{\mathcal{V},s}$ s.t.

$$(1 - \delta_{\mathcal{V},s})\|\mathbf{c}\|^2_{\mathcal{V},2} \leq \|\tilde{\mathbf{\Psi}}\mathbf{c}\|^2_{\mathcal{V},2} \leq (1 + \delta_{\mathcal{V},s})\|\mathbf{c}\|^2_{\mathcal{V},2} \qquad (\mathcal{V}\text{-RIP})$$

Lemma. [Dexter, Tran, W. '16]

- A matrix $\tilde{\mathbf{\Psi}}$ satisfies RIP with δ_s iff it satisfies $\mathcal{V}$-RIP with $\delta_{\mathcal{V},s} = \delta_s$.
- Query complexity for complex-valued signal recovery carries over to this case.

"Coupled" (basis pursuit denoising) approach

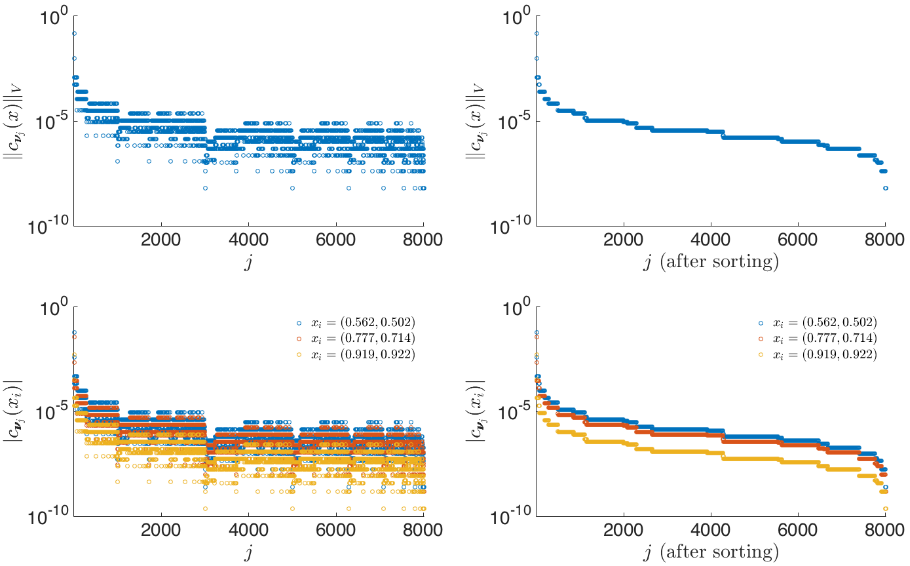
- Reconstruction algorithms that allow simultaneous approximation $\mathbf{c}$ in D.
- For instance, an extension of standard ℓ_1 minimization can be formulated as

$$\boldsymbol{c} = \text{argmin}_{\boldsymbol{z}\in\mathcal{V}^N} \|\boldsymbol{z}\|_{\mathcal{V},1} \;\text{ subject to }\; \|\mathbf{\Psi}\boldsymbol{z} - \boldsymbol{u}\|_{\mathcal{V},2} \leq \eta/\sqrt{m}.$$

- A priori information on the decay of $(\|c_{\boldsymbol{\nu}}\|_{\mathcal{V}})_{\boldsymbol{\nu}\in\Lambda_0}$ can be exploited to enhance the convergence of recovery algorithms.
- Global reconstruction only assumes a priori bounds of the tail expansion in energy norms $\eta = \|u_{\Lambda_0^c}\|_{\mathcal{V},2}$, which are much more realistic than pointwise bound

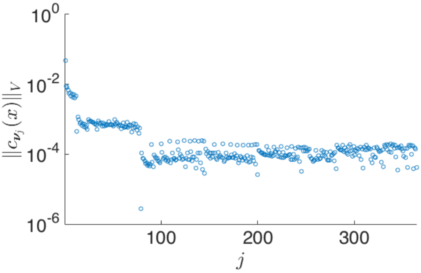
Coefficient decay

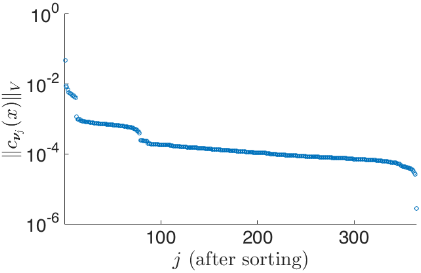
Affine coefficient example: $a(x, \boldsymbol{y}) = a_0(x) + \sum_{j=1}^{d} b_j(x) y_j$

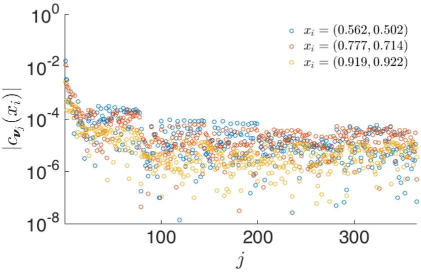


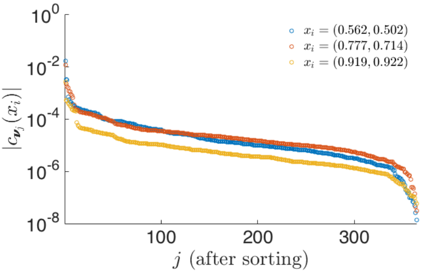
Coefficient decay

Log transformed KL example: $a(x, \mathbf{y}) \approx 0.5 + \exp(\varphi_0 + \sum_{k=1}^{d} \varphi_k y_k)$









Compressed sensing for parametric PDE recovery

Convergence of basis pursuit in the Hilbert-valued setting

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Theorem. [Dexter, Tran, W. '16]
Suppose that the $2s$ restricted isometry constant of the matrix $\mathbf{\Psi} \in \mathbb{C}^{m\times N}$ satisfies

$$\delta_{2s} < \frac{4}{\sqrt{41}} \approx 0.6246.$$

Then, for any $\boldsymbol{c} \in \mathcal{V}^N$ and $\boldsymbol{u} \in \mathcal{V}^m$ with $\|\mathbf{\Psi}\boldsymbol{c} - \boldsymbol{u}\|_{\mathcal{V},2} \le \eta/\sqrt{m}$, a solution $\boldsymbol{c}^\#$ of

$$\text{minimize}_{\boldsymbol{z}\in\mathcal{V}^N} \|\boldsymbol{z}\|_{\mathcal{V},1} \quad \text{subject to } \|\mathbf{\Psi}\boldsymbol{z} - \boldsymbol{u}\|_{\mathcal{V},2} \le \eta/\sqrt{m}$$

approximates the vector $\boldsymbol{c}$ with errors

$$\|\boldsymbol{c} - \boldsymbol{c}^\#\|_{\mathcal{V},1} \le C\sigma_s(\boldsymbol{c})_{\mathcal{V},1} + D\sqrt{s}\eta,$$

$$\|\boldsymbol{c} - \boldsymbol{c}^\#\|_{\mathcal{V},2} \le \frac{C}{\sqrt{s}}\sigma_s(\boldsymbol{c})_{\mathcal{V},1} + D\eta,$$

where the constants $C, D > 0$ depend only on δ_{2s}, and $\sigma_s(\boldsymbol{c})_{\mathcal{V},1}$ is the error of the best s-term approximation to $\boldsymbol{c}$ in the norm $\|\cdot\|_{\mathcal{V},1}$.

Given an exact estimate of the tail, it is possible to prove a rate which is independent of the tail bound η, for details see [Chkifa, Dexter, Tran, W. '16].

Example 3: Compressed sensing for parametric PDE recovery

Stochastic elliptic PDE with affine coefficient

Parameterized stochastic elliptic problem on and $D \times \mathcal{U} \subseteq \mathbb{R}^n \times \mathbb{R}^d$:

$$\begin{cases} -\nabla \cdot (a(x,\boldsymbol{y})\nabla u(x,\boldsymbol{y})) &= f(x) \quad \text{in } D \times \mathcal{U}, \\ u(x,\omega) &= 0 \quad \text{on } \partial D \times \mathcal{U}. \end{cases} \tag{4}$$

Here $a(x,\boldsymbol{y}) = a(x,\boldsymbol{y})$ is a random field parameterized by $\boldsymbol{y} \in \mathcal{U} \subset \mathbb{R}^d$. Specifically, we focus on the case that $y_n \sim \mathcal{U}(-\sqrt{3},\sqrt{3})$, and $a(x,\boldsymbol{y})$ is given by:

$$a(x,\boldsymbol{y}) = a_{\min} + y_1 \left(\frac{\sqrt{\pi}L}{2}\right)^{1/2} + \sum_{n=2}^{d} \zeta_n \varphi_n(x) y_n,$$

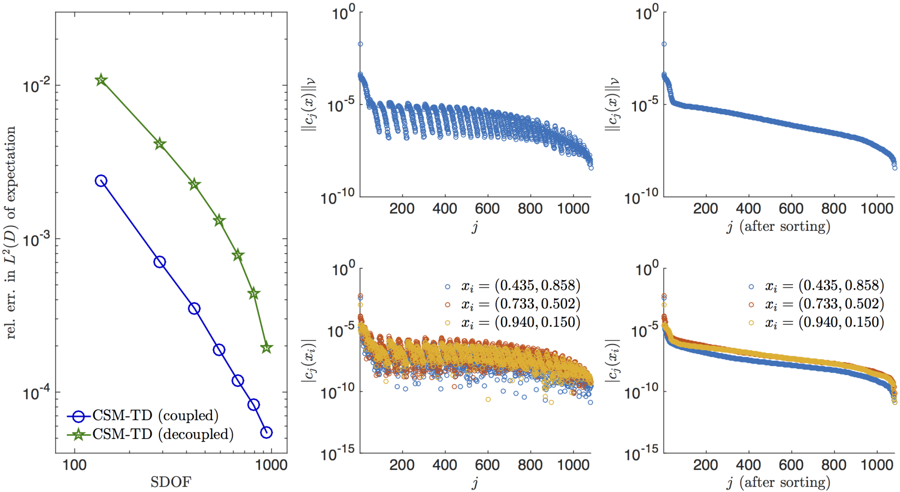
$$\zeta_n = (\sqrt{\pi}L)^{1/2} \exp\left(\frac{-\left(\left\lfloor \frac{n}{2} \right\rfloor \pi L\right)^2}{8}\right), \quad \text{for } n > 1,$$

$$\varphi_n(x) = \begin{cases} \sin\left(\left\lfloor \frac{n}{2} \right\rfloor \pi x_1 / L_p\right), & \text{if } n \text{ is even}, \\ \cos\left(\left\lfloor \frac{n}{2} \right\rfloor \pi x_1 / L_p\right), & \text{if } n \text{ is odd}, \end{cases}$$

which is the KL expansion associated with the squared exponential covariance kernel, $L_c = 1/4$ is the correlation length, and $a_{\min}$ is chosen so that $a(x,\boldsymbol{y}) > 0$ $\forall x \in D, \boldsymbol{y} \in \Gamma$.

Compressed sensing for parametric PDE recovery

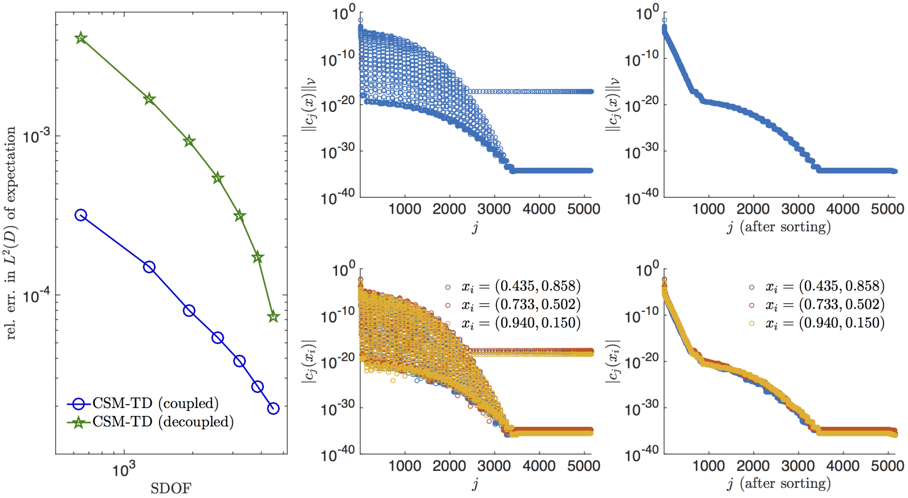
High-dimensional affine coefficient ($d = 45$)



Here $N = \#\Lambda = 1081$.

Compressed sensing for parametric PDE recovery

High-dimensional affine coefficient ($d = 100$)

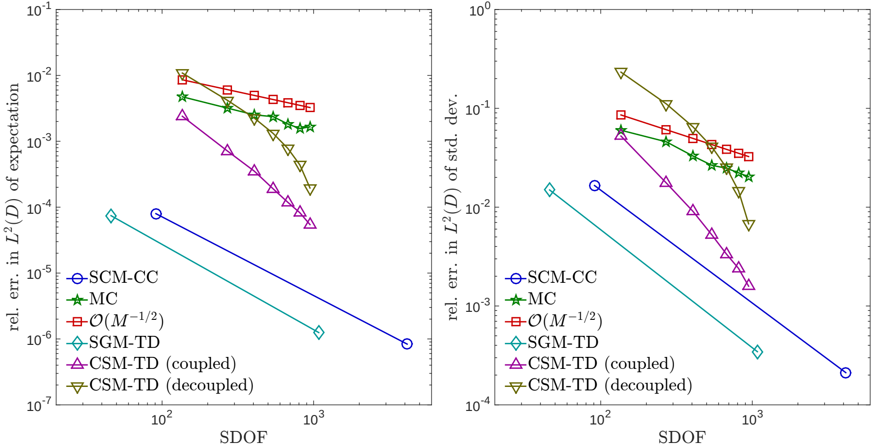


Here $N = \#\Lambda = 5151$.

Compressed sensing for parametric PDE recovery

Comparison to other methods ($d = 45$)

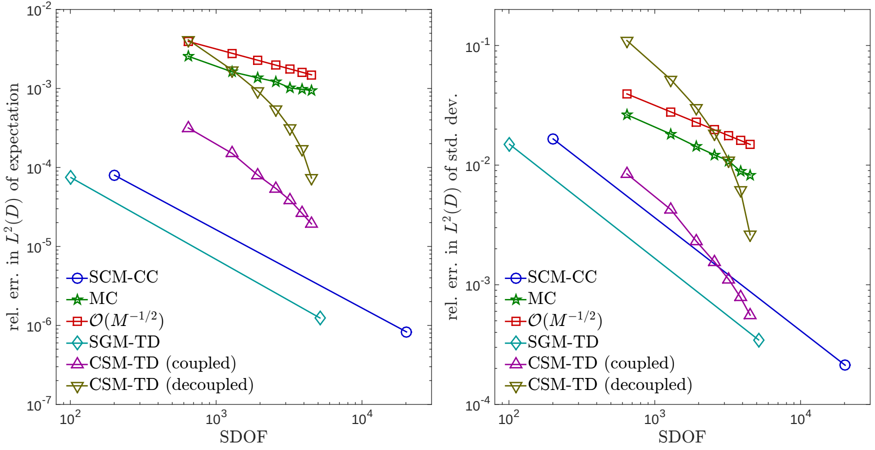
OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS



Here $N = \#\Lambda = 1081.$

Compressed sensing for parametric PDE recovery

Comparison to other methods ($d = 100$)



Here $N = \#\Lambda = 5151$.

Nonconvex regularizations

Problem revisted. Given $\boldsymbol{\Psi} \in \mathbb{R}^{m \times N}$ $(m \ll N)$, reconstruct an unknown sparse vector $\boldsymbol{c} \in \mathbb{R}^N$ from the measurements $\boldsymbol{u} = \boldsymbol{\Psi}\boldsymbol{c}$:

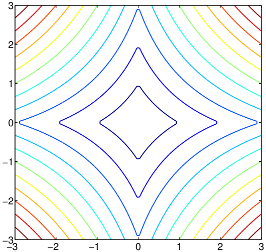
Nonconvex regularizations: $\underset{\boldsymbol{z} \in \mathbb{R}^N}{\text{minimize}} \;\; R(\boldsymbol{z})$ subject to $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$

1. ℓ_p norm with $0 \le p < 1$: $R(\boldsymbol{z}) = \|\boldsymbol{z}\|_p$ [Chartrand '07; Foucart, Lai '09]
2. $\ell_1 - \ell_2$: $R(\boldsymbol{z}) = \|\boldsymbol{z}\|_1 - \|\boldsymbol{z}\|_2$ [Esser, Lou, Xin '13; Yin, Lou, He, Xin '15]
3. capped ℓ_1: $R(\boldsymbol{z}) = \sum_{j=1}^{N} \min\{|z_j|, \alpha\}$ [Zhang '09; Shen, Pan, Zhu '12]
4. two-level ℓ_1: $R(\boldsymbol{z}) = \rho \sum_{j \in J(\boldsymbol{z})} |z_j| + \sum_{j \in J(\boldsymbol{z})^c} |z_j|$ [Huang, Liu, Shi '15]
   $\rightarrow$ *$\rho < 1$ and $J(\boldsymbol{z})$ is the set of largest components of $|z_j|$*
5. iterative support detection: $R(\boldsymbol{z}) = \|\boldsymbol{z}_T\|_1$ [Wang, Yin '10]
   $\rightarrow$ *T is a subset of $\{1, \dots, N\}$ s.t. $T^c \subset supp(\boldsymbol{u})$*
6. sorted ℓ_1: $R(\boldsymbol{z}) = \lambda_1 |\boldsymbol{z}|_{[1]} + \ldots + \lambda_N |\boldsymbol{z}|_{[N]}$ [Huang, Shi, Yan '15]

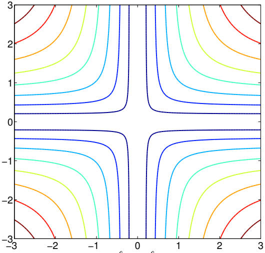
Here: $|\boldsymbol{z}|_{[1]} \ge \ldots \ge |\boldsymbol{z}|_{[N]}$ are the components ranked in decreasing order; $0 \le \lambda_1 \le \ldots \le \lambda_N$.

Nonconvex regularizations

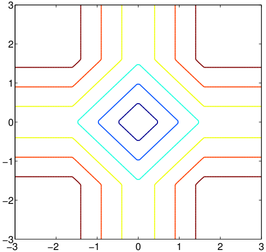
OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS



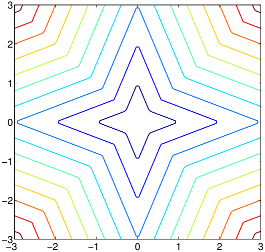
$\ell_{0.5}$



$\ell_1 - \ell_2$



capped ℓ_1



two level ℓ_1

Nonconvex regularizations

Sparse recovery in comparison with ℓ_1

COMPUTATIONAL & APPLIED MATHEMATICS

Intuitively, nonconvex minimization is better than ℓ_1 in enforcing sparsity.

In experiments, consistently outperforming ℓ_1, i.e., requiring less samples for exact reconstruction.

- These observations have not been fully supported in theory.

Theoretical recovery guarantee: often established via different versions of **null space property (NSP)** [Cohen, Dahmen, DeVore '09].

- Less restrictive than ℓ_1 ✓

  ℓ_p; iterative support detection

- More restrictive than ℓ_1 **?**

  two-level ℓ_1; $\ell_1 - \ell_2$

- Seemingly not available **?**

  capped ℓ_1; sorted ℓ_1

Our main result: a unified NSP based-condition for a general class of nonconvex minimizations (encompassing all listed above) showing that they are at least as good as ℓ_1 minimization in exact recovery of sparse signals.

Null space properties

- Necessary and sufficient condition for the exact recovery of every s-sparse vectors using ℓ_1:

$$\ker(\mathbf{\Psi}) \setminus \{0\} \subset \mathcal{N}_{\ell_1} := \left\{ \boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}_S\|_1 < \|\boldsymbol{v}_{S^c}\|_1, \ \forall S : \ |S| \leq s \right\}.$$

- Sufficient condition for two-level ℓ_1 [Huang, Liu, Shi et. al. '15]:

$$\ker(\mathbf{\Psi})\setminus\{0\} \subset \mathcal{N}_{\text{t-l } \ell_1} := \left\{ \boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}_S\|_1 < \|\boldsymbol{v}_{H\cap S^c}\|_1, \, \forall S\!: |S| \leq s, \, \forall H : |H| = \lfloor N/2 \rfloor \right\}.$$

- Sufficient condition for $\ell_1 - \ell_2$ [Yin, Lou, He, Xin '15]:

$$\ker(\mathbf{\Psi})\setminus\{0\} \subset \mathcal{N}_{\ell_1 - \ell_2} := \left\{ \boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}_S\|_1 + \|\boldsymbol{v}_S\|_2 + \|\boldsymbol{v}_{S^c}\|_2 < \|\boldsymbol{v}_{S^c}\|_1, \, \forall S\!: |S| \leq s \right\}.$$

$$\mathcal{N}_{\text{t-l } \ell_1} \subsetneq \mathcal{N}_{\ell_1} \quad \text{and} \quad \mathcal{N}_{\ell_1 - \ell_2} \subsetneq \mathcal{N}_{\ell_1}$$

Generalized recovery guarantee for nonconvex regularizations

$$\underset{\boldsymbol{z}\in\mathbb{R}^N}{\text{minimize}} \quad R(\boldsymbol{z}) \text{ subject to } \boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$$

Let $\mathcal{I}$ be the set of nonnegative real number and R be a mapping from $\mathbb{R}^N$ to $\mathcal{I}$ satisfying $R(z_1,\dots,z_N) = R(|z_1|,\dots,|z_N|),\ \forall \boldsymbol{z} = (z_1,\dots,z_N) \in \mathbb{R}^N$.

- R is called **symmetric** on $\mathcal{I}^N$ if for every $\boldsymbol{z} \in \mathcal{I}^N$ and every permutation $(j(1),\dots,j(N))$ of $(1,\dots,N)$:

$$R(z_{j(1)},\dots,z_{j(N)}) = R(\boldsymbol{z}).$$

- R is called **concave** on $\mathcal{I}^N$ if for every $\boldsymbol{z},\boldsymbol{z}' \in \mathcal{I}^N$ and $0 \le \lambda \le 1$:

$$R(\lambda\boldsymbol{z} + (1-\lambda)\boldsymbol{z}') \ge \lambda R(\boldsymbol{z}) + (1-\lambda)R(\boldsymbol{z}').$$

- R is called **increasing** on $\mathcal{I}^N$ if for every $\boldsymbol{z},\boldsymbol{z}' \in \mathcal{I}^N$, if $\boldsymbol{z} \ge \boldsymbol{z}'$ then

$$R(\boldsymbol{z}) \ge R(\boldsymbol{z}'),$$

  where $\boldsymbol{z} \ge \boldsymbol{z}'$ means $z_j \ge z'_j,\ \forall 1 \le j \le N$.

Generalized recovery guarantee for nonconvex regularizations

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

$$\underset{z \in \mathbb{R}^N}{\text{minimize}} \quad R(\boldsymbol{z}) \text{ subject to } \boldsymbol{\Psi z} = \boldsymbol{u} \qquad (P_R)$$

Theorem [Tran, W. '16]. Assume R is symmetric, concave, increasing in $\mathcal{I}^N$ and

$$R(\underbrace{\overline{z},\dots,\overline{z}}_{s}, z_1, \underbrace{0,\dots,0}_{N-s-1}) > R(\underbrace{\overline{z},\dots,\overline{z}}_{s}, \underbrace{0,\dots,0}_{N-s}), \ \forall\, 0 < z_1 \leq \overline{z}, \qquad (5)$$

then (P_R) is at least as good as ℓ_1:

every s-sparse vector $\boldsymbol{c} \in \mathbb{R}^N$ is the unique solution to (P_R) if $\ker(\boldsymbol{A}) \setminus \{0\} \subset \mathcal{N}_{\ell_1}$.

Applicable to

- ℓ_p $(0 < p < 1)$
- $\ell_1 - \ell_2$
- capped ℓ_1
- two-level ℓ_1
- sorted ℓ_1 if $\lambda_{s+1} > 0$

Generalized recovery guarantee for nonconvex regularizations

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Theorem [Tran, W. '16]. Assume R is symmetric, concave and increasing in $\mathcal{U}^N$. Replace (5) by the stronger condition

$$R(\underbrace{\bar{z},\dots,\bar{z}}_{s-1},\bar{z}-z_1,z_1,\underbrace{0,\dots,0}_{N-s-1}) > R(\underbrace{\bar{z},\dots,\bar{z}}_{s},\underbrace{0,\dots,0}_{N-s}),\ \forall\, 0<z_1<\bar{z}, \tag{6}$$

then (P_R) can be better than ℓ_1:

every s-sparse vector $\boldsymbol{c}\in\mathbb{R}^N$ (except equal-height signals[a]) is the unique solution to (P_R) if

$$\ker(\boldsymbol{A})\setminus\{0\}\subset\left\{\boldsymbol{v}\in\mathbb{R}^N:\|\boldsymbol{v}_S\|_1 \le \|\boldsymbol{v}_{S^c}\|_1,\ \forall S:|S|\le s\right\}.$$

[a] all nonzero coordinates of $\boldsymbol{c}$ have the same magnitude

Applicable to

- ℓ_p $(0<p<1)$
- $\ell_1-\ell_2$
- sorted ℓ_1 if $\lambda_{s+1}>\lambda_s$

Concluding remarks

- An unified NSP based-condition for a general class of nonconvex minimizations (encompassing all listed above) showing that they are at least as good as ℓ_1 minimization in exact recovery of sparse signals.
- Certified recovery guarantees that combat the curse of dimensionality through new weighted $\ell_1$minimization and iterative hard thresholding approaches:
  - Exploit the structure of the sets of best s-terms.
  - Established through a improved estimate of restricted isometry property (RIP), and proved for general bounded orthonormal systems.
  - Query complexity carries over to the Hilbert-valued reovery: currently extending the convergence analysis of Fixed-point continuation (FPC) [Hale, Yin, Zhang '08], and Bregman iterations [Yin, Osher, Goldfarb, Darbon '08] to Hilbert-valued signals in $\mathcal{V}^N$.
- "Let the RIP Rest In Peace"
  - Uniform recovery is usually guaranteed by the RIP of the normalized matrix $\tilde{\mathbf{\Psi}}$.
  - For reconstruction using ℓ_1 minimization, the upper bound of $\|\tilde{\mathbf{\Psi}}\boldsymbol{z}\|_2^2$ is **NOT necessary**.
  - A more natural formulation is given by the Restricted eigenvalue condition (REC) [Bickel, Ritov, Tsybakov '09; van de Geer, Bühlmann '09].

Concluding remarks

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

For $\alpha > 1$, define

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

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

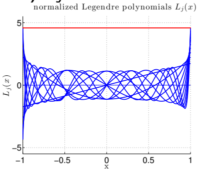
$$\|\tilde{\boldsymbol{\Psi}}\boldsymbol{z}\|_2^2 \geq (1-\delta)\|\boldsymbol{z}\|_2^2,$$

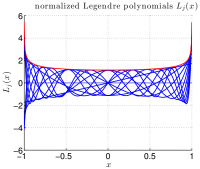
for all $\boldsymbol{z} \in C(s;\alpha) := \bigcup_{(\#S)=s} C(S;\alpha)$.

- The formulation with $C(S;\alpha) := \left\{ \boldsymbol{z} \in \mathbb{C}^N : \|\boldsymbol{z}_{S^c}\|_1 \leq \alpha\|\boldsymbol{z}_S\|_1 \right\}$ is more common.
- $C(s;\alpha)$ contains all s-sparse vectors.
- REC holds but RIP fails to hold for many random Gaussian and sub-Gaussian design matrices [Raskutti, Wainwright, Yu '10; Rudelson, Zhou '13].
- REC implies the robust null space property $\Rightarrow$ the upper bound of $\|\tilde{\boldsymbol{\Psi}}\boldsymbol{z}\|_2^2$ is not necessary [Tran, W. '16].

Concluding remarks

- Sufficient condition for successful reconstruction of s-sparse solution is $m \geq \Theta^2 s \times$ log factors, [Rudelson, Vershynin '08; Foucart, Rauhut '13].
- Several methods has been developed to overcome big Θ, i.e.,
  - preconditioning the basis [Rauhut, Ward '12];
  - asymptotic and equilibrium measure [Hampton, Doostan '15; Jakeman, Narayan, Zhou '16];
  - low degree subspaces [Guo, Yan, Xiu '12].
  - weighted ℓ_1 minimization [Adcock '15, '16; Rauhut, Ward '15; Chkifa, Dexter, Tran, W. '16].
- New idea: combine the REC with sharp estimates of the envelope bounds for bounded (i.e., Legendre) systems.





Let $\mathcal{E}_{s,\alpha} := \{\boldsymbol{z} \in C(s;\alpha) : \|\boldsymbol{z}\|_2 = 1\}$. We find the covering number of $\mathcal{E}_{s,\alpha}$ under the pseudo-metric

$$d(\boldsymbol{z}, \boldsymbol{z}') = \max_{1 \leq i \leq m} |\psi(y_i, \boldsymbol{z} - \boldsymbol{z}')|.$$

Concluding remarks

Improved sampling complexity for sparse Legendre systems

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

The covering number (and number of measurements) involves an upper bound of

$$\mathbb{E}(d(\boldsymbol{z},\overline{\boldsymbol{z}})) \lesssim \sqrt{\frac{s}{M}} \sum_{i=1}^{m} \Theta(y_i) \exp(-\frac{1}{4\Theta^2(y_i)}\sqrt{m}),$$

where $\Theta : \mathcal{U} \to \mathbb{R}$ is a bound of all Legendre polynomials $\{\Psi_j\}_{j\in\Lambda_0}$, [Foucart, Rauhut '13].

1. If setting $\Theta(y) \equiv K = \sup_{j\in\Lambda_0} \|\Psi_j\|_\infty$, then $\mathbb{E}(d(\boldsymbol{z},\overline{\boldsymbol{z}})) \lesssim \frac{K\sqrt{s}}{\sqrt{M}}\sqrt{\log(m)}$.

   Consequently, the condition $m \geq K^2 s \times$ log factors can be derived.

2. If setting $\Theta(y) = \frac{\sqrt{2/\pi}}{\sqrt[4]{1-y^2}}$, the envelope bound of all Legendre polynomials, $\mathbb{E}(d(\boldsymbol{z},\overline{\boldsymbol{z}}))$ is **unbounded**, however,

   with high probability of sample sets $\{y_1,\ldots,y_m\}$, $\mathbb{E}(d(\boldsymbol{z},\overline{\boldsymbol{z}})) \lesssim \frac{\sqrt{s}}{\sqrt{M}}\sqrt[4]{m}$.

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

   Consequently, $m \geq s\sqrt{m} \times$ log factors or $m \geq s^2 \times$ log factors independent of mutual coherence $\Theta^2 s$.

→ Multivariate expansions? Other incoherent systems? Nonuniform recovery? Other $Z(y)$?

References

- J. Bourgain, *An improved estimate in the restricted isometry problem*, Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, 65-70, 2014.
- M. Gunzburger, C. G. Webster, G. Zhang, *Stochastic finite element methods for PDEs with random input data*, Acta Numerica, 23:521-650, 2014.
- H. Tran, C. G. Webster, G. Zhang, *Analysis of quasi-optimal polynomial approximations for parameterized PDEs with deterministic and stochastic coefficients*. ORNL/TM-2014/468, Submitted: Numerische Mathematik, 2016.
- A. Chkifa, N. Dexter, H. Tran, C. G. Webster, *Polynomial approximation via compressed sensing of high-dimensional functions on lower sets*. Submitted: Mathematics of Computation, 2016.
- J. Burkardt, M. Gunzburger, C. G. Webster, G. Zhang, *A hyper-spherical sparse grid approach for high-dimensional discontinuity detection*. SIAM Journal on Numerical Analysis, 53(3):1508-1536, 2015.
- E. Candès, J. Romberg, T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*. IEEE Trans. Inform. Theory 52(1):489-509, 2006.
- A. Cohen, R. DeVore, C. Schwab, *Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs*, Found Comput Math (2010) 10: 615–646.
- A. Cohen, R. DeVore, C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs*, Analysis and Applications, Vol. 9, No. 1 (2011), 11–47.
- H. Rahut, R. Ward, *Interpolation via weighted ℓ_1-minimization*, Applied and Computational Harmonic Analysis. Submitted: 2015.
- M. Rudelson, R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*. Comm. Pure Appl. Math. 61:1025-1045, 2008.