Oak Ridge National Laboratory
Computational & Applied Mathematics

Foundations of compressed sensing for learning sparsity of high-dimensional problems

Clayton G. Webster[†*]

Special thanks to: B. Adcock[‡], S. Brugiapaglia[‡], N. Dexter[†], H. Tran[*]

[‡]Department of Mathematics, Simon Fraser University

[†]Department of Mathematics, University of Tennessee

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

Introduction

Why do we care about "sparse" signals?

Oak Ridge National Laboratory
Computational & Applied Mathematics

Example: We often represent images by expansions like

$$u(y) = \sum_{j=1}^{N} c_j \Psi_j(y)$$

where, e.g., $c = (c_1, \ldots, c_N) \in \mathbb{R}^N$ and $\{\Psi_j\}_{j=1}^{N}$ are wavelets.

Figure **Left:** Original image. **Right:** Image obtained after setting 99.98% of the coefficients c_j in the biorthogonal wavelet transform to 0. Preserves 97.87% of energy.

Takeaway: Sparse approximations can provide good solutions to real problems.

Motivation: Parameterized PDE models

Deterministic and stochastic coefficients

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

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

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

Motivation: Parameterized PDE models

Deterministic and stochastic coefficients

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

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

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

Motivation: Parameterized PDE models

Deterministic and stochastic coefficients

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

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

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

UQ for parameterized PDE models

Some assumptions

Continuity and coercivity (CC)

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

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 $y \in \mathcal{U}$ 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

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

Some simple consequences:

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

Motivation: Parameterized PDE models

Deterministic and stochastic coefficients

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

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

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

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 $y \in \mathcal{U}$, $0 < a_{\min} \leq a(x, y) \leq a_{\max}$.

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 $y \in \mathcal{U}$ 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

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

Some simple consequences:

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

Motivation: Parameterized PDE models

Deterministic and stochastic coefficients

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

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

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

UQ for parameterized PDE models

Some assumptions

Continuity and coercivity (CC)

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

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 $y \in \mathcal{U}$ 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

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

Some simple consequences:

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

Motivation: Parameterized PDE models

Deterministic and stochastic coefficients

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

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

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

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* $y \in \mathcal{U}$, $0 < a_{\min} \le a(x, y) \le a_{\max}$.

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 $y \in \mathcal{U}$ *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*

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

Some simple consequences:

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

A simple illustrative example

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

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

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

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

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

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

Affine and non-affine coefficients:

1. $a(x,y) = a_0(x) + \sum_{i=1}^d y_i \psi_i(x)$.
2. $a(x,y) = a_0(x) + \left(\sum_{i=1}^d y_i \psi_i(x)\right)^q$, $q \in \mathbb{N}$.
3. $a(x,y) = a_0(x) + \exp\left(\sum_{i=1}^d y_i \psi_i(x)\right)$ (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.

Asymptotic convergence analysis

The general abstract setting

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Main Theorem. [Tran, W., Zhang '16]

Let $b : [0,\infty)^d \to \mathbb{R}$ and Λ_s^{Qopt} be the set of indices corresponding to s largest $e^{-b(\nu)}$. Then, for any $\varepsilon > 0$, there exists $s_\varepsilon > 0$ s.t. for all $s > s_\varepsilon$:

$$\sum_{\nu \notin \Lambda_s^{\text{Qopt}}} e^{-b(\nu)} \leq C_u(\varepsilon)\, s \exp\left(- \left(\frac{s}{|\mathcal{P}|(1+\varepsilon)} \right)^{1/d} \right)$$

Here, $C_u(\varepsilon) = (4e + 4\varepsilon e - 2)\,\frac{e}{e-1}$ is independent of s and d.

- Achieve sub-exponential convergence rates $s\exp(-(\kappa s)^{1/d})$, with optimal κ.
- $|\mathcal{P}|$ can be determined computationally
  1. $\mathcal{P} = \left\{ \nu \in [0,\infty)^d : \sum_{i=1}^d \lambda_i \nu_i \leq 1 \right\}$, for $B(\nu) = \rho^{-\nu} \prod_{i=1}^d \sqrt{2\nu_i + 1}$.
  2. $\mathcal{P} = \left\{ \nu \in [0,\infty)^d : \sum_{i=1}^d \lambda_i \nu_i \leq 1 \ \ \forall (\rho,\delta) \in \mathcal{A} \right\}$, for $B(\nu) = \inf_{\rho,\delta} C_\delta \rho^{-\nu}$.
  3. $\mathcal{P} = \{ \nu \in (0,\infty)^d : \sum_{i=1}^d \lambda_i \nu_i - \log \frac{|\nu|^{|\nu|}}{\prod_{i=1}^d \nu_i^{\nu_i}} < 1 \}$, for $B(\nu) = \rho^{-\nu} \frac{|\nu|!}{\nu!}$.
- Faster rates are realized at larger cardinality.

Asymptotic convergence analysis

Comparisons to previous rates using Taylor polynomials in total degree subspaces

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Proposition. [Tran, W., Zhang '16]
Consider the Taylor series $\sum_{\nu \in \mathbb{N}^d} t_\nu y^\nu$ of u. Assume that

$$\|t_\nu\|_\mathcal{V} \le C\rho^{-\nu}, \quad \forall \nu \in \mathbb{N}^d. \tag{1}$$

Denote by Λ_s^{Qopt} the set of indices corresponding to s largest bounds in (1). For any $\varepsilon > 0$, there exists $s_\varepsilon > 0$ depending on ε such that, for all $s > s_\varepsilon$:

$$\sup_{y \in \mathcal{U}} \left\| u(y) - \sum_{\nu \in \Lambda_s^{\text{Qopt}}} t_\nu y^\nu \right\|_\mathcal{V} \le C_u(\varepsilon) s \exp\left(- \left(\frac{s\, d! \prod_{i=1}^d \lambda_i}{(1+\varepsilon)} \right)^{1/d} \right).$$

Previous rates:

1. Applying Stechkin estimate in [CDS '11] to our setting: $\left(\prod_{i=1}^d \frac{1}{1-e^{-p\lambda_i}}\right)^{1/p} s^{1-\frac{1}{p}}$.
   Rate is non-asymptotic and applicable for infinite dimensional parameter space.
2. Optimization of Stechkin rate [BNTT '14]: $s \exp\left(-\frac{1}{e} \left(s \prod_{i=1}^d \lambda_i \right)^{1/d} d\xi \right)$.
   $\xi \nearrow \frac{e-1}{e} \simeq 0.63$ as $s \nearrow \infty$.

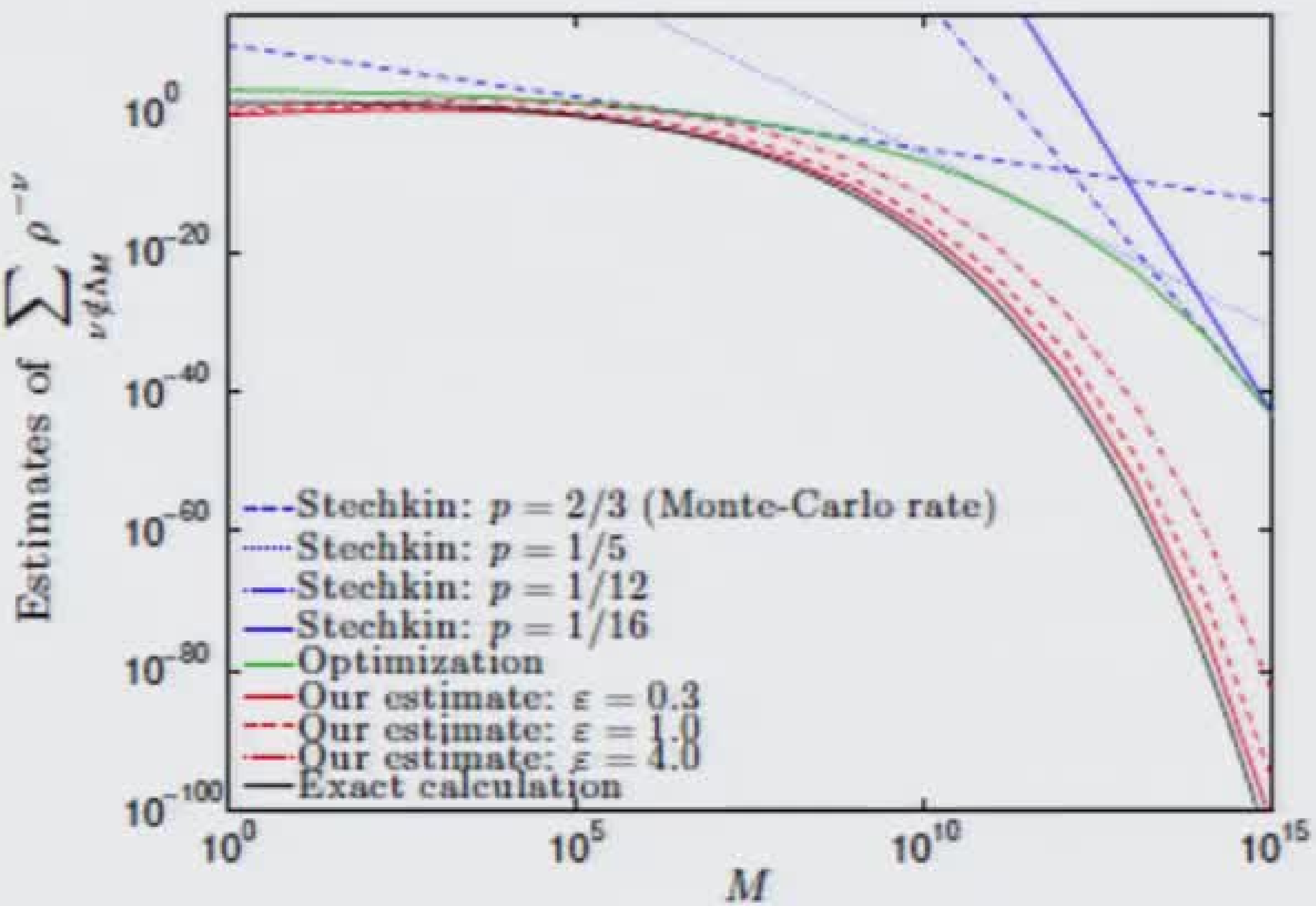
Asymptotic convergence analysis

Numerical illustration II

OAK RIDGE National Laboratory COMPUTATIONAL & APPLIED MATHEMATICS

Example 2: Isotropic 8-dimensional parametric domain

Estimate the truncation error of $\sum_{\nu \in \mathbb{R}^8} \rho^{-\nu}$, where $\rho_i = 2, \quad \forall 1 \leq i \leq 8$.



Forward-backward iterations for joint sparse recovery

A new theory to guarantee strong convergence

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

An operator $T : \mathcal{V} \to \mathcal{V}$ is said to be firmly nonexpansive (FNE) if

$$\|T\boldsymbol{x} - T\boldsymbol{y}\|_{2,2}^2 \le \|\boldsymbol{x} - \boldsymbol{y}\|_{2,2}^2 - \|(I - T)\boldsymbol{x} - (I - T)\boldsymbol{y}\|_{2,2}^2 \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}.$$

Lemma. [Bauschke, Combettes 2010]. Let $\tau > 0$. Then J_τ is row-wise firmly nonexpansive.

Proof.

Since $B_2(\mathbf{0}, 1)$ is a nonempty, closed, convex set, and $J_\tau = (I - \mathcal{P}_\tau)$ where $\mathcal{P}_\tau$ is a projection, we have

$$\langle \mathcal{P}_\tau \boldsymbol{v}_j - \mathcal{P}_\tau \boldsymbol{w}_j, \boldsymbol{w}_j - \mathcal{P}_\tau \boldsymbol{w}_j \rangle_2 \le 0$$
$$\langle \mathcal{P}_\tau \boldsymbol{w}_j - \mathcal{P}_\tau \boldsymbol{v}_j, \boldsymbol{v}_j - \mathcal{P}_\tau \boldsymbol{v}_j \rangle_2 \le 0$$

for every $\boldsymbol{v}_j, \boldsymbol{w}_j \in \mathbb{C}^\Omega$. Adding, we obtain $\langle \mathcal{P}_\tau \boldsymbol{v}_j - \mathcal{P}_\tau \boldsymbol{w}_j, \boldsymbol{v}_j - \boldsymbol{w}_j \rangle_2 \ge \|\mathcal{P}_\tau \boldsymbol{v}_j - \mathcal{P}_\tau \boldsymbol{w}_j\|_2^2$. It follows

$$\begin{aligned}\|J_\tau(\boldsymbol{v}_j) - J_\tau(\boldsymbol{w}_j)\|_2^2 &= \|(I - \mathcal{P}_\tau)\boldsymbol{v}_j - (I - \mathcal{P}_\tau)\boldsymbol{w}_j\|_2^2 \\ &= \|\boldsymbol{v}_j - \boldsymbol{w}_j\|_2^2 + \|\mathcal{P}_\tau \boldsymbol{v}_j - \mathcal{P}_\tau \boldsymbol{w}_j\|_2^2 - 2\langle \boldsymbol{v}_j - \boldsymbol{w}_j, \mathcal{P}_\tau \boldsymbol{v}_j - \mathcal{P}_\tau \boldsymbol{w}_j \rangle_2 \\ &\le \|\boldsymbol{v}_j - \boldsymbol{w}_j\|_2^2 - \|\mathcal{P}_\tau \boldsymbol{v}_j - \mathcal{P}_\tau \boldsymbol{w}_j\|_2^2,\end{aligned}$$

which implies J_τ is firmly nonexpansive since $(I - J_\tau) = (I - I + \mathcal{P}_\tau) = \mathcal{P}_\tau$. □

Forward-backward iterations for joint sparse recovery

A new theory to guarantee strong convergence

OAK RIDGE National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

Theorem. [Dexter, Tran, W. '17].
Let $0 < \tau < 2/\|H\|_2$. Then the iterations $x^{k+1} := J_\tau \circ G_\tau(x^k)$ converge strongly to an element $x^* \in X^*$ from any $x^0 \in \mathbb{C}^{N\times\Omega}$.

Sketch of proof: First-order optimality conditions imply $\|(A^*(Ax^* - u))_j\|_2 \leq 1$ for all $j \in [N]$ and $x^* \in X^*$. Therefore, we partition the index set into

$$L := \{\, j : \|(A^*(Ax^* - u))_j\|_2 < 1 \,\} \qquad E := \{\, j : \|(A^*(Ax^* - u))_j\|_2 = 1 \,\}.$$

Easy to see: $L \subset (\text{supp}(x^*))^c$, $\text{supp}(x^*) \subseteq E$, & $L \cup E = [N]$ $\forall\, x^* \in X^*$.

1. Finite convergence for $j \in L$ follows arguments from [Hale, Yin, Zhang '08]
2. We show "angular convergence" for $j \in E$ using the firmly nonexpansive property
3. Weak convergence has been shown in more general setting, see, e.g., [Daubechies, et al '04], [Combettes '04], via Opial's Theorem and "asymptotic regularity" of S_τ
4. Combine the weak and angular convergence to obtain strong convergence

Sum over $j \in [N]$ with $\bar{c}^k := \sum_{j=1}^N c_j^k$, apply the nonexpansiveness of G_τ and iterate:

$$\|x^{k+1} - x^*\|_2^2 \le \|G_\tau(x^k) - G_\tau(x^*)\|_2^2 - \bar{c}^k$$
$$\le \|x^k - x^*\|_2^2 - \bar{c}^k \le \underbrace{\cdots}_{k\text{-times}} \le \|x^0 - x^*\|_2^2 - \sum_{\ell=0}^{k} \bar{c}^\ell.$$

Rearrange: $\sum_{\ell=0}^k \bar{c}^\ell \le \underbrace{\|x^0 - x^*\|_2^2}_{\text{independent of } k} \implies \bar{c}^k \to 0,$ and hence $c_j^k \to 0$ as $k \to \infty$.

Collinearity & $c_j^k \to 0 \implies \theta_j^k := \sphericalangle(x_j^k, x_j^*) := \cos^{-1}\left(\frac{\langle x_j^k, x_j^* \rangle_2}{\|x_j^k\|_2 \|x_j^*\|_2}\right) \to 0$ as $k \to \infty$.

Weak convergence $\implies \|x_j^k\|_2 \cos\theta_j^k \equiv \langle x_j^k, x_j^* \rangle / \|x_j^*\|_2 \to \langle x_j^*, x_j^* \rangle / \|x_j^*\|_2 \equiv \|x_j^*\|_2$ as $k \to \infty$ (also works when $x_j^* = 0$, slight change) and x_j^k are bounded

Use weak and angular convergence to show

$$\|x_j^k\|_2 - \|x_j^*\|_2 = \underbrace{\left(\|x_j^k\|_2 - \|x_j^k\|_2 \cos\theta_j^k\right)}_{\text{angular convergence \& boundedness}} + \underbrace{\left(\|x_j^k\|_2 \cos\theta_j^k - \|x_j^*\|_2\right)}_{\text{weak convergence}} \to 0 \text{ as } k \to \infty$$

Strong convergence follows since

$$\|x_j^k - x_j^*\|_2^2 = \|x_j^k\|_2^2 + \|x_j^*\|_2^2 - 2\langle x_j^k, x_j^* \rangle_2 \to 2\|x_j^*\|_2^2 - 2\|x_j^*\|_2^2 = 0 \quad \text{as } k \to \infty.$$

Concluding remarks

Oak Ridge National Laboratory
COMPUTATIONAL & APPLIED MATHEMATICS

1. Certified recovery guarantees that combat the curse of dimensionality through new weighted ℓ_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.
   - Can recover the "true" best s-term approximation and not a best weighted s-term (which requires a weighted version of Stechkin's estimate).
2. Joint-sparse recovery enables the simultaneous reconstruction of a set of sparse vectors with common support, from measurements.
   - Derived the forward-backward splitting method in this setting
   - More work to be done in the convergence theory of these methods
     - Recently shown strong convergence for the forward-backward splitting method
     - Would like to show strong convergence for Bregman iterations
   - Showed connection between joint-sparse recovery problem and parameterized PDEs
   - Need more numerical experiments
     - Nonlinear parameterized PDEs
     - Linear vs. nonlinear stochastic parameterization
     - Can be improved with the introduction of a weighted ℓ_1 regularization (similar to previous work).
3. An unified NSP based-condition for a general class of nonconvex minimizations showing that they are at least as good as ℓ_1 minimization in exact recovery of sparse signals.