Compressive Sensing Approaches for High-Dimensional Function Approximation

Simone Brugiapaglia

E-mail: simone_brugiapaglia@sfu.ca

Web: sites.google.com/view/paglia

SFU SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Problem setting

Goal: To approximate a function

$$f : D = (-1,1)^d \to \mathbb{C}, \quad \text{with } d \gg 1,$$

from pointwise samples $f(t_1), \ldots, f(t_m)$.

Problem setting

Goal: To approximate a function

$$f : D = (-1,1)^d \to \mathbb{C}, \quad \text{with } d \gg 1,$$

from pointwise samples $f(t_1), \ldots, f(t_m)$.

Setting and assumptions (informal):

- We are **free to choose** the sampling points t_i;
- Samples $f(t_i)$ may be **expensive** to compute (e.g., involving PDE solve); and corrupted by **unknown sources of error**;
- f is **compressible** w.r.t. some orthogonal polynomials.

Main challenge: Curse of dimensionality. [Bellman, 1961]

Problem setting

Goal: To approximate a function

$$f : D = (-1,1)^d \to \mathbb{C}, \quad \text{with } d \gg 1,$$

from pointwise samples $f(t_1), \ldots, f(t_m)$.

Setting and assumptions (informal):

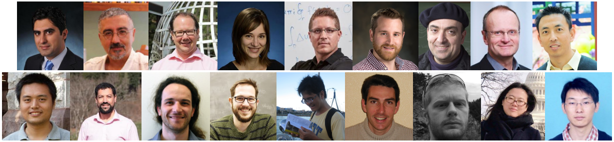
- We are **free to choose** the sampling points t_i;
- Samples $f(t_i)$ may be **expensive** to compute (e.g., involving PDE solve); and corrupted by **unknown sources of error**;
- f is **compressible** w.r.t. some orthogonal polynomials.

Main challenge: Curse of dimensionality. [Bellman, 1961]

Application: Uncertainty Quantification (UQ).

Function approximation and UQ meet CS

We will focus on a recent class of high-dimensional approximation techniques based on compressed sensing (CS).



(A subset of the) main references:

- Compressed sensing + orthogonal polynomials
  - [Rauhut, Ward, 2012], [Yan, Guo, Xiu, 2012];
- Weighted ℓ^1 minimization and function approximation
  - [Rauhut, Ward, 2016], [Adcock, 2017], [Chkifa, Dexter, Tran, Webster, 2018], [Adcock, B., Webster, 2018]
- CS + uncertainty quantification
  - [Doostan, Owhadi, 2011], [Mathelin, Gallivan, 2012], [Yang, Karniadakis, 2013], [Peng, Hampton, Doostan, 2014], [Rauhut, Schwab, 2017], [Bouchot, Rauhut, Schwab, 2017]
- Fast-growing literature, **very active community**!

The methodology (1/2)

Sparsity basis: We consider tensorized bases $\{\psi_j\}_{j \in \mathbb{N}_0^d}$ for $L^2(D)$

$$\psi_j = \phi_{j_1} \otimes \cdots \otimes \phi_{j_d},$$

where $\{\phi_j\}_{j \in \mathbb{N}_0}$ are 1D Chebyshev or Legendre orthogonal polynomials.

$$f = \sum_{j \in \mathbb{N}_0^d} x_j \psi_j.$$

Ambient set: Fixed a finite-dimensional set $\Lambda \subseteq \mathbb{N}_0^d$, with $|\Lambda| = N$, we truncate

$$f = \underbrace{\sum_{j \in \Lambda} x_j \psi_j}_{\text{Approximation}} + \underbrace{\sum_{j \notin \Lambda} x_j \psi_j}_{\text{Truncation error}} =: f_\Lambda + e_\Lambda.$$

The methodology (2/2)

Sampling: Evaluate f at m random sampling points

$$t_1, \dots, t_m \overset{\text{i.i.d.}}{\sim} \nu(t)$$

where ν is the orthogonality measure of $\{\psi_j\}_{j \in \mathbb{N}_0^d}$:

$$A = \left(\tfrac{1}{\sqrt{m}} \psi_j(t_i)\right)_{ij} \in \mathbb{C}^{m \times N}, \quad y = \left(\tfrac{1}{\sqrt{m}} f(t_i)\right)_i \in \mathbb{C}^m$$

Moreover, denoting $x_\Lambda = (x_i)_{i \in \Lambda} \in \mathbb{C}^N$, we have the linear system

$$y = A x_\Lambda + e,$$

where e is an unknown error corrupting the data.

Recovery: weighted quadratically-constrained basis pursuit (WQCBP)

$$\hat{x}_\Lambda := \arg\min_{z \in \mathbb{C}^N} \|z\|_{1,u} \quad \text{s.t. } \|Az - y\|_2 \le \eta \quad \rightsquigarrow \quad \hat{f} = \sum_{j \in \Lambda} \hat{x}_j \psi_j,$$

where $\|z\|_{1,u} = \sum_{j \in \Lambda} u_j |z_j|$ and the weights are defined as

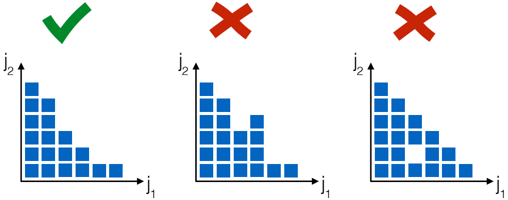
$$u_j := \|\psi_j\|_{L^\infty}.$$

Structured sparsity in lower sets

We study the recovery properties of the method using lower sets.

Definition (Lower or downward closed set)

A set $S \subseteq \mathbb{N}_0^d$ is lower if $\forall i, j:\ i \leq j$ and $j \in S \Longrightarrow i \in S$.



Goal: to find an approximation $\hat{x}_\Lambda$ to x s.t.

$$\|x - \hat{x}_\Lambda\|_{1,u} \approx \sigma_{s,L}(x)_{1,u} = \inf_{\substack{\|z\|_0 \leq s, \\ \text{supp}(z) \text{ lower}}} \|z - x\|_{1,u}.$$

Two key properties

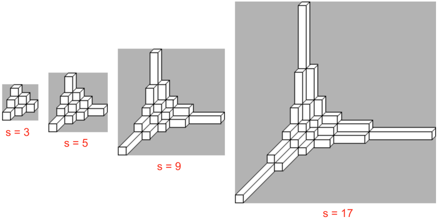
- **Compressibility:** In parametric PDEs, the best s-term approximation error in lower sets of the solution map has decay rate $s^{-\alpha}$, $\alpha > 0$ in L^2_ν or L^∞ for a large class of smooth PDE operators [Chkifa, Cohen, Schwab, 2015]
- The **union of all s-sparse lower sets** is the hyperbolic cross:

$$\Lambda^{\mathrm{HC}}_{d,s} = \left\{ i = (i_1, \dots, i_d) \in \mathbb{N}_0^d : \prod_{j=1}^{d} (i_j + 1) \leq s \right\},$$

resulting in a **controlled growth of N** with respect to d and s

$$N = |\Lambda^{\mathrm{HC}}_{d,s}| \lesssim \min \left\{ s^3 4^d, s^{2+\log_2(d)} \right\}.$$

[Kühn, Sickel, Ullrich, 2015; Chernov, Dũng, 2016]



[Image courtesy of Prof. Daniel Potts http://www-user.tu-chemnitz.de/~potts/nfft/nsfft.php]

Noise-aware recovery analysis

[Chkifa, Dexter, Tran, Webster, 2018]

Assuming to know an *a priori* upper bound of the form

$$\|e\|_2 \leq \eta,$$

and assuming

$$m \asymp s^{\gamma} \cdot \ln^2(s) \min\{d + \ln(s), \ln(2d)\ln(s)\} + \ln(s)\ln(\ln(s)/\varepsilon), \qquad (*)$$

where

$$\gamma = \begin{cases} 2 & \text{(Legendre)} \\ \frac{\ln(3)}{\ln(2)} \approx 1.58 & \text{(Chebyshev)}, \end{cases}$$

WQCBP recovers an approximation $\hat{f}$ to f such that

$$\|f - \hat{f}\|_{L^\infty} \lesssim \sigma_{s,L}(x)_{1,u} + s^{\gamma/2}\eta$$

with probability at least $1 - \varepsilon$. Similar bound holds w.r.t. the L^2_ν norm.

Good news! In $(*)$, *m depends logarithmically on d*.

As in the standard CS case, e may contain **truncation**, **numerical**, and **model error**. In particular, truncation error is unavoidable in this context. As a consequence,

$$\|e\|_2 \leq \eta, \qquad (*)$$

is usually not available.

As in the standard CS case, e may contain **truncation**, **numerical**, and **model error**. In particular, truncation error is unavoidable in this context. As a consequence,

$$\|e\|_2 \leq \eta, \qquad (*)$$

is usually not available.

Can we bridge this gap between theory and practice?

Noise-blind recovery analysis

Theorem [Adcock, S.B., Webster, 2018]

Let $\Lambda = \Lambda_{d,s}^{\mathrm{HC}}$ be the hyperbolic cross and

$$m \asymp s^{\gamma} \cdot \underbrace{\ln^2(s)\min\{d + \ln(s), \ln(2d)\ln(s)\} + \ln(s)\ln(\ln(s)/\varepsilon)}_{=: L(s,d,\varepsilon)}.$$

where $\gamma = 2$ (Legendre) or $\gamma = \frac{\ln(3)}{\ln(2)} \approx 1.58$ (Chebyshev).
Then, for any $f \in L^2_\nu(D) \cap L^\infty(D)$ and $\eta \geq 0$, WCQBP computes $\hat{f}$ s.t.

$$\|f - \hat{f}\|_{L^\infty(D)} \lesssim \sigma_{s,L}(x)_{1,u} + s^{\gamma/2}(\eta + \|e\|_2 + \mathcal{Q}_u(A) \cdot \max\{\|e\|_2 - \eta, 0\}),$$

with probability $1 - \varepsilon$. Moreover,

$$\mathcal{Q}_u(A) \leq s^{\alpha/2} \frac{\sqrt{L(s,d,\varepsilon)}}{\sigma_{\min}(\sqrt{\frac{m}{N}} A^*)},$$

where $\alpha = 2$ (Legendre) or $\alpha = 1$ (Chebyshev).

- Note: m depends logarithmically on d.
- An analogous result holds with respect to the $L^2_\nu(D)$ norm.

Good news:

- The assumption $\|e\|_2 \leq \eta$ is not needed (as opposed to previous results).
- The term $\max\{\|e\|_2 - \eta, 0\}$ suggests the choice $\eta \approx \|e\|_2$, theoretically justifying the use of **cross validation**.
- $\sigma_{\min}(\sqrt{\frac{m}{N}}A^*)$ behaves well in expectation.
- Numerics show that $\mathcal{Q}_u(A)$ has moderate size.
- Proof based on: (weighted versions of) **restricted isometry** and **null space properties**, **quotient property** [Wojtaszczyk, 2010]

Good news:

- The assumption $\|e\|_2 \leq \eta$ is not needed (as opposed to previous results).
- The term $\max\{\|e\|_2 - \eta, 0\}$ suggests the choice $\eta \approx \|e\|_2$, theoretically justifying the use of **cross validation**.
- $\sigma_{\min}(\sqrt{\frac{m}{N}} A^*)$ behaves well in expectation.
- Numerics show that $\mathcal{Q}_u(A)$ has moderate size.
- Proof based on: (weighted versions of) **restricted isometry** and **null space properties**, **quotient property** [Wojtaszczyk, 2010]

Problem:

- The choice of the tuning parameter η still depends on $\|e\|_2$...

Good news:

- The assumption $\|e\|_2 \le \eta$ is not needed (as opposed to previous results).
- The term $\max\{\|e\|_2 - \eta, 0\}$ suggests the choice $\eta \approx \|e\|_2$, theoretically justifying the use of **cross validation**.
- $\sigma_{\min}(\sqrt{\frac{m}{N}} A^*)$ behaves well in expectation.
- Numerics show that $\mathcal{Q}_u(A)$ has moderate size.
- Proof based on: (weighted versions of) **restricted isometry** and **null space properties**, **quotient property** [Wojtaszczyk, 2010]

Problem:

- The choice of the tuning parameter η still depends on $\|e\|_2$...

🤔

Can we get rid of the dependence of the decoder on e?

Alternative decoders

In [Adcock, Bao, S.B., 2017], we suggest and analyze the performance of two alternative decoders.

- **Weighted LASSO (WLASSO)**

$$\min_{z \in \mathbb{C}^N} \|Az - y\|_2^2 + \lambda \|z\|_{1,u},$$

The weighted version of the LASSO [Tibshirani, 1996], extremely popular in statistics, signal processing and, more recently, in function approximation.

- **Weighted square-root LASSO (WSR-LASSO)**

$$\min_{z \in \mathbb{C}^N} \|Az - y\|_2 + \lambda \|z\|_{1,u},$$

Introduced in [Belloni, Chernozhukov, Wangand, 2014] (in the unweighted version) and quite popular in statistics.

Its potential not fully exploited in CS (yet!).

Recovery guarantees

Theorem [Adcock, Bao, S.B., 2017]

Under the same setting as WQCBP, if $m \gtrsim s^\gamma \cdot L(s, d, \varepsilon)$ and

$$\lambda \asymp \frac{\|e\|_2}{s^{\gamma/2}} \text{ (WLASSO)}, \qquad \lambda \asymp \frac{1}{s^{\gamma/2}} \text{ (WSR-LASSO)}, \qquad (*)$$

where $\gamma = 2$ or $\frac{\ln(3)}{\ln(2)}$, for Legendre and Chebyshev polynomials, respectively, the approximation $\tilde{f}$ computed by WLASSO and WSR-LASSO satisfies

$$\|f - \tilde{f}\|_{L^\infty} \lesssim \sigma_{s,L}(x)_{1,u} + s^{\gamma/2}\|e\|_2,$$

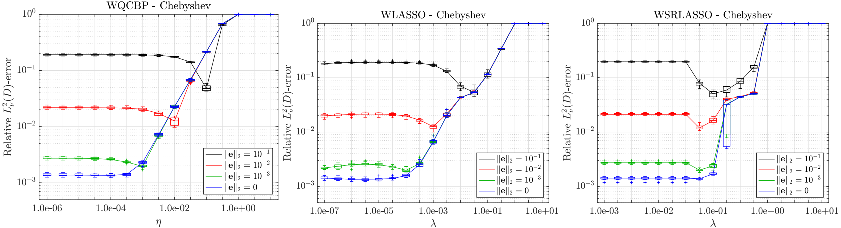
with probability at least $1 - \varepsilon$.

- ☺ The choice of tuning parameter $(*)$ is independent of e for WSR-LASSO.
- ☺ The term $\|e\|_2$ is not amplified by any log factor.

Numerics: function approximation

Dimension $d = 15$, function $f(t) = \exp\left(-\frac{1}{15}\sum_{\ell=1}^{15}\cos(t_\ell)\right)$,

$s = 10,\ N = 1432,\ m = 280,$ Gaussian noise, $1/s^{\gamma/2} \approx 1.6126e-01$



Some highlights:

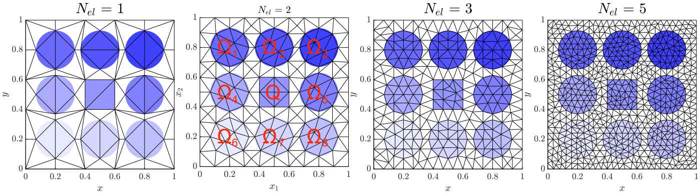
- Optimal tuning parameter strategies confirmed numerically.
- Highly-noisy data + right parameter choice
  $\rightsquigarrow$ **substantial error reduction** ($\approx$ **factor** 7)

An example from parametric PDEs

Consider the parametric diffusion equation

$$\begin{cases} -\nabla \cdot (a_t \nabla u_t) = 100 \cdot \mathbf{1}_Q, & \text{in } \Omega = (0,1)^2, \\ u_t = 0, & \text{on } \partial\Omega, \end{cases}$$

where $a_t = 1 - \sum_{\ell=1}^{8} \mathbf{1}_{\Omega_\ell}(0.595 + 0.395 t_\ell) \in [0.01, 0.99]$ and $t \in (-1,1)^8$.

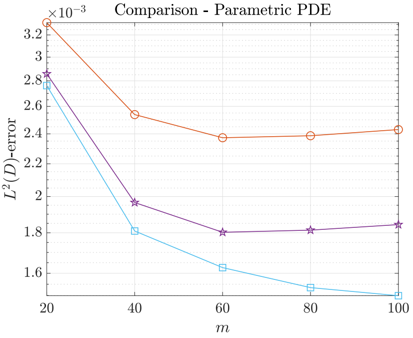


Quantity of interest: $f(t) = \int_Q u_t(x)\, dx$.

- Each sample $f(t_i)$ depends on a PDE solve;
- The samples are affected by discretization and numerical error.

Numerical results

The performance of WSR-LASSO is analogous to WQCBP and WLASSO without any *a priori* knowledge on e.



○	WQCBP	η tuned using a high-fidelity solution to estimate e
⋆	WLASSO	λ tuned using a high-fidelity solution to estimate e
□	WSR-LASSO	λ tuned according to the theory

Parameters: $s = 10$, $N = 353$.

The case of sparse corruptions

Assume an unknown subset of the samples to be corrupted arbitrarily.

$$e = e^{\text{bounded}} + e^{\text{sparse}},$$

where $\|e^{\text{sparse}}\|_0 \leq k$ has possibly unbounded entries.

Motivation:

- Large-scale UQ computations are performed on **big clusters**.
- **Node failures** can compromise these expensive computations.
- Need for **fault-tolerant methods**.

Decoder: $\min_{z \in \mathbb{C}^N} \|Az - y\|_1 + \lambda \|z\|_{1,u}$ (Weighted LAD-LASSO, [Xu, 2005])

If $\lambda \asymp \sqrt{\frac{k}{s^\gamma}}$ and $m \gtrsim s^\gamma \cdot L(s, d, \varepsilon)$, then [Adcock, Bao, S.B., 2017]

$$\|f - \tilde{f}\|_{L^\infty} \lesssim \sigma_{s,L}(x)_{1,u} + s^{\gamma/2} \|e^{\text{bounded}}\|_2.$$

Future challenges

Improved sampling strategies

In UQ, **sampling is the most computationally expensive part**. Hence, devising methods that need the less samples is crucial.

Improved sampling strategies

In UQ, **sampling is the most computationally expensive part**. Hence, devising methods that need the less samples is crucial.

1. **Can we improve the bound $m \asymp s^\gamma \cdot L(s, d, \varepsilon)$?**

Improved sampling strategies

In UQ, **sampling is the most computationally expensive part**. Hence, devising methods that need the less samples is crucial.

1. **Can we improve the bound $m \asymp s^{\gamma} \cdot L(s, d, \varepsilon)$?**
2. **Devise adaptive sampling strategies.**

   Promising preliminary results in 1D. [Adcock, Boyer, S.B., 2018]

Improved sampling strategies

In UQ, **sampling is the most computationally expensive part**. Hence, devising methods that need the less samples is crucial.

1. **Can we improve the bound $m \asymp s^{\gamma} \cdot L(s, d, \varepsilon)$?**
2. **Devise adaptive sampling strategies.**

   Promising preliminary results in 1D. [Adcock, Boyer, S.B., 2018]
3. **The quest for optimal sampling strategies.**

Faster recovery

CS is able to lessen the curse of dimensionality w.r.t. to the number of samples. However, for **large-scale UQ problems**, N may become too large to allow for convex minimization in $\mathbb{C}^N$.

Faster recovery

CS is able to lessen the curse of dimensionality w.r.t. to the number of samples. However, for **large-scale UQ problems**, N may become too large to allow for convex minimization in $\mathbb{C}^N$.

Future direction: Adopting **greedy strategies** to accelerate the recovery phase.

Preliminary numerical results show the potential of **weighted orthogonal matching pursuit** as an alternative to weighted ℓ^1 miminization. [Adcock, S.B., 2018]

Main advantages & opportunities:

- Number of iterations $O(s)$;
- Parallelizability;
- Easily adaptable to structured sparsity.

Bibliography:

- B. Adcock, S. B. Sparse approximation of multivariate functions from small datasets via weighted orthogonal matching pursuit. Submitted, 2018.
- B. Adcock, S. B., C.G. Webster. Compressed sensing approaches for polynomial approximation of high-dimensional functions. Applied and Numerical Harmonic Analysis, 2018.
- B. Adcock, A. Bao, S. B. Correcting for unknown errors in sparse high-dimensional function approximation. Submitted, 2017.

Thank you!