Recent advances in compressed sensing techniques for the numerical approximation of PDEs

Simone Brugiapaglia

Simon Fraser University, Canada

simone_brugiapaglia@sfu.ca

Joint work with
Ben Adcock (SFU), Stefano Micheletti (MOX), Fabio Nobile (EPFL),
Simona Perotto (MOX), Clayton G. Webster (ONL).

QUIET 2017
SISSA. Trieste, Italy – July 20, 2017

Compressed Sensing (CS)

Pioneering papers: [Donoho, 2006; Candès, Romberg, & Tao, 2006]

Main ingredients:

- Sparsity / Compressibility;
- Random measurements (sensing);
- Sparse recovery.

Sparsity: Let $\boldsymbol{s} \in \mathbb{C}^N$ be an **s-sparse** w.r.t. a basis $\boldsymbol{\Psi}$:

$$\boldsymbol{s} = \boldsymbol{\Psi}\boldsymbol{x} \quad \text{and} \quad \boldsymbol{x} \in \Sigma_s^N = \{\boldsymbol{z} \in \mathbb{C}^N : \|\boldsymbol{z}\|_0 \leq s\},$$

where $\|\boldsymbol{x}\|_0 := \#\{i : \boldsymbol{x}_i \neq 0\}$ and $s \ll N$.

Compressibility: fast decay of the best s-term approximation error

$$\sigma_s(\boldsymbol{x})_p = \inf_{\boldsymbol{z} \in \Sigma_s^N} \|\boldsymbol{x} - \boldsymbol{z}\|_p \leq C s^{-\alpha},$$

for some $C, \alpha > 0$, where .

Sensing

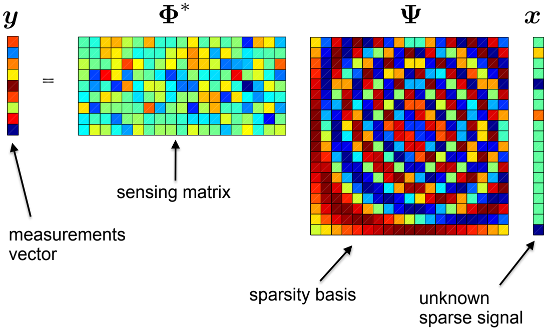
In order to acquire $\boldsymbol{s}$, we perform $m \sim s \cdot \mathrm{polylog}(N)$ **linear nonadaptive random measurements**

$$\langle \boldsymbol{s}, \boldsymbol{\varphi}_i \rangle =: y_i, \quad \text{for } i = 1, \dots, m.$$

If we consider the matrix $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_i] \in \mathbb{C}^{N \times m}$, we have

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y},$$

where $\boldsymbol{A} = \boldsymbol{\Phi}^* \boldsymbol{\Psi} \in \mathbb{C}^{m \times N}$ and $\boldsymbol{y} \in \mathbb{C}^m$. This system is **highly underdetermined**.



Sparse recovery

Thanks to the sparsity / compressibility of $\boldsymbol{s}$, we can resort to **sparse recovery techniques**. We aim at approximating the solution to

$$(\mathrm{P}_0) \quad \min_{\boldsymbol{z}\in\mathbb{C}^N} \|\boldsymbol{z}\|_0, \quad \text{s.t. } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}.$$

☹ In general, (P_0) is a **NP-hard** problem...

☺ There are **computationally tractable** strategies to approximate it!

In particular, it is possible to employ

- **greedy strategies**, e.g. **Orthogonal Matching Pursuit** (**OMP**);
- **convex relaxation**, e.g., the **quadratically-constrained basis pursuit (QCBP)** program:

$$\min_{\boldsymbol{z}\in\mathbb{C}^N} \|\boldsymbol{z}\|_1, \quad \text{s.t. } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \le \eta,$$

  referred to as Basis pursuit (BP) when $\eta = 0$.

Restricted isometry property

Many important recovery results in CS are based on the **Restricted Isometry Property** (**RIP**).

Definition (RIP)

A matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies the RIP(s, δ) with $\delta \in [0, 1)$ if

$$(1-\delta)\|\boldsymbol{z}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \leq (1+\delta)\|\boldsymbol{z}\|_2^2, \quad \forall \boldsymbol{z} \in \Sigma_s^N.$$

The RIP implies recovery results for:

- OMP [Zhang, 2011; Cohen, Dahmen, DeVore, 2015];
- QCBP [Candés, Romberg, Tao, 2006], [Foucart, Rauhut; 2013];

Optimal recovery error estimates (without noise) for a decoder Δ look like [Cohen, Dahmen, DeVore, 2009]

$$\|\boldsymbol{x} - \Delta(\boldsymbol{A}\boldsymbol{x})\|_2 \lesssim \frac{\sigma_s(\boldsymbol{x})_1}{\sqrt{s}}, \quad \forall \boldsymbol{x} \in \mathbb{C}^N,$$

and hold with high probability.

Compressed sensing

CS for (parametric) PDEs

Inside the black box

Outside the black box

Conclusions

CS as a tool to solve PDEs

Parametric PDEs' setting:

- $z \in D \subseteq \mathbb{R}^d$: parametric domain, $d \gg 1$;
- $L_z u_z = g$: PDE;
- $z \mapsto u_z$: solution map (the "black box");
- $u_z \mapsto Q(u_z)$: quantity of interest.

Can we take advantage of the CS paradigm in this setting?

CS as a tool to solve PDEs

Parametric PDEs' setting:

- $z \in D \subseteq \mathbb{R}^d$: parametric domain, $d \gg 1$;
- $L_z u_z = g$: PDE;
- $z \mapsto u_z$: solution map (the "black box");
- $u_z \mapsto Q(u_z)$: quantity of interest.

Can we take advantage of the CS paradigm in this setting?

YES! At least in two ways, addressed in this talk:

1. Inside the black box, to approximate $z \mapsto u_z$

2. Outside the black box, to approximate $z \mapsto f(z) = Q(u_z)$

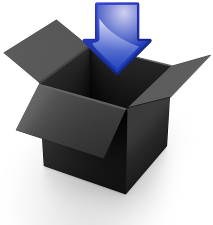
Compressed sensing

CS for (parametric) PDEs

Inside the black box

Outside the black box

Conclusions



CS inside the black box

Consider the **weak formulation** of a PDE

$$\text{find } u \in U : \quad a(u,v) = \mathcal{F}(v), \quad \forall v \in V,$$

and its **Petrov-Galerkin** (**PG**) discretization [Aziz, Babuška, 1972].

Motivation to apply CS:

- **reduce the computational cost** associated with a classical PG discretization;
- situations with a **limited budget** of evaluations of $\mathcal{F}(\cdot)$;
- deeper **theoretical understanding** of the PG method.

Case study:

✎ **Advection-diffusion-reaction (ADR) equation**, with $U = V = H_0^1(\Omega)$, $\Omega = [0,1]^d$, and

$$a(u,v) = (\eta \nabla u, \nabla v) + (\mathbf{b} \cdot \nabla u, v) + (\rho u, v), \quad \mathcal{F}(v) = (f, v).$$

Related literature

Ancestors: PDE solvers based on ℓ^1-minimization

1988 [J. Lavery, 1988; J. Lavery, 1989]
Inviscid Burgers' equation, conservation laws

2004 [J.-L. Guermond, 2004; J.-L. Guermond and B. Popov, 2009]
Hamilton-Jacobi, transport equation

CS techniques for PDEs

2010 [S. Jokar, V. Mehrmann, M. Pfetsch, and H. Yserentant, 2010]
Recursive mesh refinement based on CS (Poisson equation)

2015 [S. B., S. Micheletti, S. Perotto, 2015;
S. B., F. Nobile, S. Micheletti, S. Perotto, 2017]
CORSING for ADR problems

The Petrov-Galerkin method

Choose $U^N \subseteq H_0^1(\Omega)$ and $V^M \subseteq H_0^1(\Omega)$ with

$$U^N = \mathrm{span}\{\underbrace{\psi_1, \dots, \psi_N}_{\text{trials}}\}, \quad V^M = \mathrm{span}\{\underbrace{\varphi_1, \dots, \varphi_M}_{\text{tests}}\}$$

Then we can discretize the weak problem as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \quad A_{ij} = a(\psi_j, \varphi_i), \quad y_i = \mathcal{F}(\varphi_i)$$

with $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, $\boldsymbol{y} \in \mathbb{C}^M$.

The Petrov-Galerkin method

Choose $U^N \subseteq H_0^1(\Omega)$ and $V^M \subseteq H_0^1(\Omega)$ with

$$U^N = \mathrm{span}\{\underbrace{\psi_1, \dots, \psi_N}_{\text{trials}}\}, \quad V^M = \mathrm{span}\{\underbrace{\varphi_1, \dots, \varphi_M}_{\text{tests}}\}$$

Then we can discretize the weak problem as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}, \quad A_{ij} = a(\psi_j, \varphi_i), \quad y_i = \mathcal{F}(\varphi_i)$$

with $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, $\boldsymbol{y} \in \mathbb{C}^M$.

We can establish the following analogy:

Petrov-Galerkin method:		*Sampling:*
solution of a PDE	$\Longleftrightarrow$	signal
tests (bilinear form)		**measurements** (inner product)

Classical case: square matrices

When dealing with Petrov-Galerkin discretizations, one usually ends up with a **big square** matrix.

$$
\begin{array}{cc}
 & \begin{array}{ccccccc} \psi_1 & \psi_2 & \psi_3 & \psi_4 & \psi_5 & \psi_6 & \psi_7 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\
\begin{array}{c} \varphi_1 \rightarrow \\ \varphi_2 \rightarrow \\ \varphi_3 \rightarrow \\ \varphi_4 \rightarrow \\ \varphi_5 \rightarrow \\ \varphi_6 \rightarrow \\ \varphi_7 \rightarrow \end{array} &
\underbrace{\begin{bmatrix}
\times & \times & \times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times & \times & \times \\
\times & \times & \times & \times & \times & \times & \times
\end{bmatrix}}_{a(\psi_j,\varphi_i)}
\end{array}
\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{bmatrix}
=
\begin{bmatrix} \mathcal{F}(\varphi_1) \\ \mathcal{F}(\varphi_2) \\ \mathcal{F}(\varphi_3) \\ \mathcal{F}(\varphi_4) \\ \mathcal{F}(\varphi_5) \\ \mathcal{F}(\varphi_6) \\ \mathcal{F}(\varphi_7) \end{bmatrix}
$$

"Compressing" the discretization

We would like to use only m random tests instead of N, with $m \ll N$...

$$\begin{array}{cc}
 & \begin{array}{ccccccc} \psi_1 & \psi_2 & \psi_3 & \psi_4 & \psi_5 & \psi_6 & \psi_7 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\
\begin{array}{c} \varphi_1 \rightarrow \\ \varphi_2 \rightarrow \\ \varphi_3 \rightarrow \\ \varphi_4 \rightarrow \\ \varphi_5 \rightarrow \\ \varphi_6 \rightarrow \\ \varphi_7 \rightarrow \end{array} & \underbrace{\begin{bmatrix} \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times \end{bmatrix}}_{a(\psi_j,\varphi_i)}
\end{array}
\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{bmatrix} = \begin{bmatrix} \mathcal{F}(\varphi_1) \\ \mathcal{F}(\varphi_2) \\ \mathcal{F}(\varphi_3) \\ \mathcal{F}(\varphi_4) \\ \mathcal{F}(\varphi_5) \\ \mathcal{F}(\varphi_6) \\ \mathcal{F}(\varphi_7) \end{bmatrix}$$

Sparse recovery

...in order to obtain a **reduced discretization**.

$$
\begin{array}{cc}
 & \begin{array}{ccccccc} \psi_1 & \psi_2 & \psi_3 & \psi_4 & \psi_5 & \psi_6 & \psi_7 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\
\begin{array}{c} \varphi_2 \rightarrow \\ \varphi_5 \rightarrow \end{array} & \underbrace{\begin{bmatrix} \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times \end{bmatrix}}_{a(\psi_j, \varphi_i)}
\end{array}
\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{bmatrix} = \begin{bmatrix} \mathcal{F}(\varphi_2) \\ \mathcal{F}(\varphi_5) \end{bmatrix}
$$

The solution is then computed using **sparse recovery** techniques.

CORSING (COmpRessed SolvING)

First, we define the **local a-coherence**
[Krahmer, Ward, 2014; B., Nobile, Micheletti, Perotto, 2017]:

$$\mu_q^N := \sup_{j \in [N]} |a(\psi_j, \varphi_q)|^2, \quad \forall q \in \mathbb{N}.$$

COSRING algorithm:

1. Define a truncation level M and a number of measurements m;
2. Draw $\tau_1, \dots, \tau_m$ *independently* at random from $[M]$ according to the probability $\mathbf{p} \sim (\mu_1^N, \dots, \mu_M^N)$ (up to rescaling).
3. Build $\boldsymbol{A} \in \mathbb{R}^{m \times N}$, $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{D} \in \mathbb{R}^{m \times m}$, defined as:

$$A_{ij} := a(\psi_j, \varphi_{\tau_i}), \quad f_i := \mathcal{F}(\varphi_{\tau_i}), \quad D_{ik} := \frac{\delta_{ik}}{\sqrt{m p_{\tau_i}}}.$$

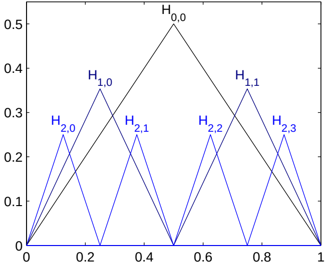
4. Use OMP to solve $\min_{\boldsymbol{z} \in \mathbb{R}^N} \|\boldsymbol{D}(\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y})\|_2^2$, s.t. $\|\boldsymbol{z}\|_0 \leq s$;

Sparsity + Sensing: How to choose $\{\psi_j\}$ and $\{\varphi_i\}$?

Heuristic criterion commonly used in CS: **space** *vs.* **frequency**.

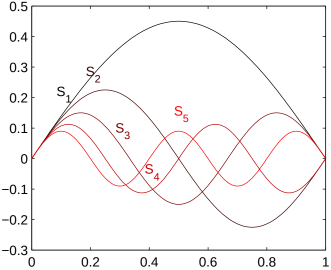
Hierarchical hat functions

[Smoliak, Dahmen, Griebel, Yserentant, Zienkiewicz, ...]



$\mathcal{H}$

Sine functions

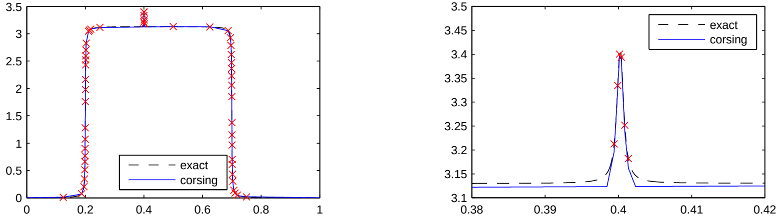


$\mathcal{S}$

We name the corresponding strategies CORSING $\mathcal{HS}$ and $\mathcal{SH}$.

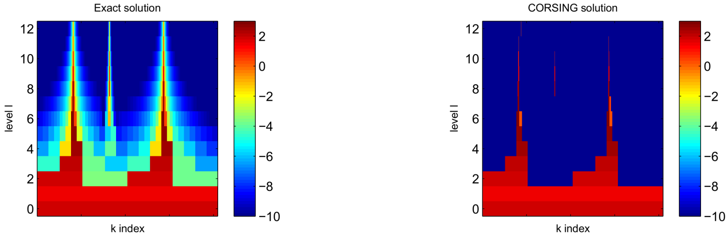
Homogeneous 1D Poisson problem CORSING $\mathcal{HS}$

$N = 8191$, $s = 50$, $m = 1200$. $\rightsquigarrow$ **Test Savings**: $\mathrm{TS} := \frac{N-m}{N} \cdot 100\% \approx 85\%$



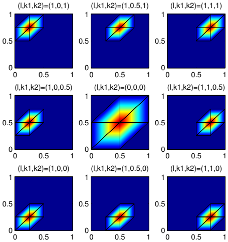
$\times$ = hat functions selected by OMP

Level-based ordering ($\log_{10} |\widehat{u}_{\ell,k}|$)

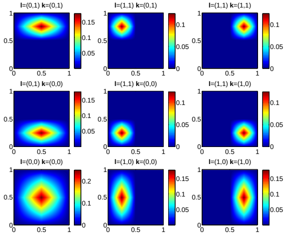


Sparsity + Sensing: 2D case

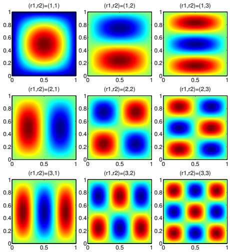
Hierarchical Pyramids ($\mathcal{P}$)



Tensor product of hat functions ($\mathcal{Q}$)



Tensor product of sine functions ($\mathcal{S}$)



An advection-dominated example

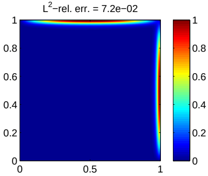
We consider a 2D advection-dominated problem

$$\begin{cases} -\mu\Delta u + \mathbf{b}\cdot\nabla u = f & \text{in } \Omega = (0,1)^2, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

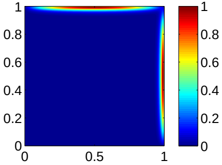
where $\mathbf{b} = [1,1]^\intercal$, $\mu = 0.01$.

CORSING $\mathcal{SP}$. Worst solution in the successful cluster over 50 runs:

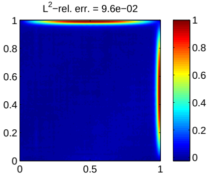
$N = 16129$
TS $= 85\%$
ESP $= 1.00$
L^2-rel. err. $= 7.1\text{e-}02$



Exact



$N = 16129$
TS $= 90\%$
ESP $= 0.94$
L^2-rel. err. $= 8.7\text{e-}02$



ESP = Empirical Success Probability

Cost reduction with respect to the "full" PG (m=N)

We compare the assembly/recovery times of "full" PG and CORSING.

"full" PG			CORSING $\mathcal{SP}$			
A	**f**	t_{rec} (\)	TS	**A**	**f**	t_{rec} (OMP)
2.5e+03	9.1e-01	7.1e+01	85%	3.8e+02	2.7e-01	8.1e+01
			90%	2.5e+02	2.0e-01	3.4e+01

- The assembly time reduction is proportional to TS.
- Also the RAM is reduced proportionally to TS.
- The recovery phase is cheaper for high TS rates.

The CORSING method can considerably reduce the computational cost associated with a "full" PG discretization.

Theoretical analysis

Theorem

Let $s, N \in \mathbb{N}$, with $s < N$. Suppose the truncation condition $\sum_{q>M} \mu_q^N \lesssim \frac{\alpha^2}{s}$ holds. Then, provided $\delta \in \left(1 - \frac{\alpha^2}{\beta^2}, 1\right)$, and

$$m \gtrsim \delta^{-2} \|\boldsymbol{\nu}^{N,M}\|_1 s \log^3(s) \log(N),$$

it holds

$$\mathbb{P}\{\beta^{-1}\mathbf{DA} \in \mathrm{RIP}(s,\delta)\} \geq 1 - N^{-\log^3(s)},.$$

where α and β are the inf-sup and the continuity constant of $a(\cdot,\cdot)$.

- Alternative analysis based on a **restricted inf-sup property** leads to suboptimal rate $m \sim s^2 \cdot$ (log factors).

Theoretical analysis

Theorem

Let $s, N \in \mathbb{N}$, with $s < N$. Suppose the truncation condition $\sum_{q>M} \mu_q^N \lesssim \frac{\alpha^2}{s}$ holds. Then, provided $\delta \in \left(1 - \frac{\alpha^2}{\beta^2}, 1\right)$, and

$$m \gtrsim \delta^{-2} \|\boldsymbol{\nu}^{N,M}\|_1 \, s \log^3(s) \log(N),$$

it holds

$$\mathbb{P}\{\beta^{-1}\mathbf{DA} \in \mathrm{RIP}(s,\delta)\} \geq 1 - N^{-\log^3(s)}, .$$

where α and β are the inf-sup and the continuity constant of $a(\cdot,\cdot)$.

- Alternative analysis based on a **restricted inf-sup property** leads to suboptimal rate $m \sim s^2 \cdot$ (log factors).

Algorithmic recovery guarantee:

CORSING recovers the best s-term approximation to u (up to a constant) using $\mathcal{O}(smN)$ flops with high probability.

Comparison with **adaptive wavelet methods**:

☹ Computational cost $O(smN)$ instead of $O(s)$;

☺ Easy parallelizability of OMP;

☺ No need for *a priori* error estimators.

CS outside the black box

We aim at approximating a function

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

of the form "(quantity of interest) $\circ$ (solution map)":

$$f(z) = Q(u_z), \quad \text{where } u_z \text{ solves } L_z u_z = g.$$

As sparsity basis, we consider the tensorized Chebyshev or Legendre orthogonal polynomials $\{\phi_j\}_{j \in \mathbb{N}_0^d}$. Then, we expand

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

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

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

Random sampling + weighted ℓ^1 minimization

We consider random evaluations of f at $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m$ drawn according to the orthogonality measure of $\{\phi_j\}_{j \in \mathbb{N}_0^d}$:

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

Moreover, denoting

$$\boldsymbol{x}_\Lambda = (x_{\boldsymbol{i}})_{\boldsymbol{i} \in \Lambda} \in \mathbb{C}^N, \quad \boldsymbol{e}_\Lambda = \frac{1}{\sqrt{m}} (f(\boldsymbol{z}_i) - f_\Lambda(\boldsymbol{z}_i)) \in \mathbb{C}^m,$$

we have the linear system

$$\boldsymbol{A}\boldsymbol{x}_\Lambda = \boldsymbol{y} + \boldsymbol{e}_\Lambda.$$

The solution is recovered by weighted QCBP

$$\min_{\boldsymbol{z} \in \mathbb{C}^N} \|\boldsymbol{z}\|_{1,\boldsymbol{u}} \quad \text{s.t. } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \eta,$$

where $\|\boldsymbol{z}\|_{1,\boldsymbol{u}} = \sum_{j \in [N]} u_j |z_j|$ the weights are chosen **intrinsically** as

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

Related literature

History of this idea:

- CS + orthogonal polynomials
  - [Rauhut, Ward, 2012], [Yau, Guo, Xiu, 2012];
- Weighted ℓ^1 minimization and function approximation
  - [Rauhut, Ward, 2016], [Adcock, 2017], [Chkifa, Dexter, Tran, Webster, 2017], [Adcock, B., Webster, 2017]
- CS + UQ with Polynomial Chaos expansion
  - [Doostan, Owhadi, 2011], [Mathelin, Gallivan, 2012], [Yang, Karniadakis, 2013], [Peng, Hampton, Doostan, 2014], [Rauhut, Schwab, 2017], [Bouchot, Rauhut, Schwab, 2017]

Lower sets and the choice of Λ

Definition (Lower or downward closed set)

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

Lower sets have been proved to be extremely effective for parametric PDEs: [Beck, Chkifa, Cohen, Dexter, DeVore, Griebel, Migliorati, Nobile, Schwab, Tamellini, Tempone, Tran, Webster, ...]

Why do they matter?

- Best s-term approximation in lower sets realizes the best s-term approximation for a large class of smooth operators, with decay rate $s^{-\alpha}$, $\alpha > 0$ in L^2 or L^∞. [Chkifa, Cohen, Schwab, 2015]
- The union of all s-sparse lower sets, is the hyperbolic cross:

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

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

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

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

Lower RIP and recovery guarantees

- Weighted cardinality of $S \subseteq \mathbb{N}_0^d$ is $|S|_{\boldsymbol{w}} := \sum_{\boldsymbol{i} \in \operatorname{supp}(S)} w_{\boldsymbol{i}}^2$
- $K(s) := \max\{|S|_{\boldsymbol{u}} : S \subseteq \mathbb{N}_0^d, S \text{ lower}\}$.

Definition (lower RIP [Chkifa, Dexter, Tran, Webster, 2017])

A matrix $\boldsymbol{A}$ fulfills the lower RIP of order s if $\exists \delta \in [0, 1)$ s.t.

$$(1-\delta)\|\boldsymbol{z}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \leq (1+\delta)\|\boldsymbol{z}\|_2^2, \quad \forall \boldsymbol{z} \in \mathbb{C}^N, \ |\operatorname{supp}(\boldsymbol{z})|_{\boldsymbol{u}} \leq K(s).$$

Assuming an a priori error bound $\|\boldsymbol{e}_\Lambda\|_2 \leq \eta$, the following uniform recovery error estimates hold [Chkifa, Dexter, Tran, Webster, 2017]:

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

$$\|f - \hat{f}\|_{L^2(D)} = \|\boldsymbol{x} - \hat{\boldsymbol{x}}_\Lambda\|_2 \lesssim \frac{\sigma_{s,L}(\boldsymbol{x})_{1,\boldsymbol{u}}}{s^{\gamma/2}} + \eta,$$

where

$$\sigma_{s,L}(\boldsymbol{x})_{1,\boldsymbol{u}} = \inf_{\boldsymbol{z} \in \Sigma_s^N, \operatorname{supp}(\boldsymbol{z}) \text{ lower}} \|\boldsymbol{z} - \boldsymbol{x}\|_{1,\mathbf{u}}.$$

Nonuniform recovery: optimality of the weights

Theorem [Adcock, 2017]

Let $0 < \epsilon < \mathrm{e}^{-1}$, $\eta \geq 0$, $\boldsymbol{w} = (w_{\boldsymbol{i}})_{\boldsymbol{i}\in\Lambda}$ be a set of weights, $\boldsymbol{x} \in \ell^2(\mathbb{N}_0^d)$ and $S \subseteq \Lambda$, $S \neq \emptyset$, be any fixed set. Suppose that $\|\boldsymbol{e}_\Lambda\|_2 \leq \eta$. Then, with probability at least $1-\epsilon$, any minimizer $\hat{\boldsymbol{x}}_\Lambda$ of

$$\min_{\boldsymbol{z}\in\mathbb{C}^N} \|\boldsymbol{z}\|_{1,\boldsymbol{w}} \quad \text{s.t.} \ \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \eta,$$

satisfies $\|\boldsymbol{x} - \hat{\boldsymbol{x}}_\Lambda\|_2 \lesssim \lambda\sqrt{|S|_{\boldsymbol{w}}}\,(\eta + \|\boldsymbol{x} - \boldsymbol{x}_\Lambda\|_{1,\boldsymbol{u}}) + \|\boldsymbol{x} - \boldsymbol{x}_S\|_{1,\boldsymbol{w}}$, provided

$$m \gtrsim \underbrace{\left(|S|_{\boldsymbol{u}} + \max_{\boldsymbol{i}\in\Lambda\setminus S}\{u_{\boldsymbol{i}}^2/w_{\boldsymbol{i}}^2\}|S|_{\boldsymbol{w}}\right)}_{=:\mathcal{M}(S;\boldsymbol{u},\boldsymbol{v})} L,$$

where $\lambda = 1 + \frac{\sqrt{\log(\epsilon^{-1})}}{\log(2N\sqrt{|S|_{\boldsymbol{w}}})}$ and $L = \log(\epsilon^{-1})\log\left(2N\sqrt{|S|_{\boldsymbol{w}}}\right)$.

- Seeking to minimize $\mathcal{M}(S;\boldsymbol{u},\boldsymbol{v})$, it is natural to choose $\boldsymbol{w} = \boldsymbol{u}$.
- This conclusion is supported by numerical evidence. [Adcock, B., Webster, 2017]

Robustness of $\ell^1_{\boldsymbol{u}}$-minimization to unknown error

Theorem [Adcock, B., Webster, 2017]

Let $\Lambda = \Lambda_s^{\mathrm{HC}}$ and assume

$$m \sim s^{\gamma} \cdot L,$$

where,

$$L = \ln^2(s) \min\{d + \ln(s), \ln(2d)\ln(s)\} + \ln(s)\ln(\ln(s)/\varepsilon).$$

Then, for every $\eta \geq 0$ and $f \in L^2(D) \cap L^\infty(D)$, the $\ell^1_{\boldsymbol{u}}$-minimization computes an approximation $\hat{f}$ s.t.

$$\|f - \hat{f}\|_{L^\infty(D)} \lesssim \sigma_{s,L}(\boldsymbol{x})_{1,\boldsymbol{u}} + s^{\gamma/2}(\eta + \|\boldsymbol{e}_\Lambda\|_2 + T_{\boldsymbol{u}}(\boldsymbol{A}, \Lambda, \boldsymbol{e}_\Lambda, \eta)),$$

$$\|f - \hat{f}\|_{L^2(D)} \lesssim \frac{\sigma_{s,L}(\boldsymbol{x})_{1,\boldsymbol{u}}}{s^{\gamma/2}} + \eta + \|\boldsymbol{e}_\Lambda\|_2 + T_{\boldsymbol{u}}(\boldsymbol{A}, \Lambda, \boldsymbol{e}_\Lambda, \eta),$$

with probability $1 - \varepsilon$, where $\gamma = 2$ or $\frac{\log(3)}{\log(2)}$, for Legendre and Chebyshev polynomials, respectively. Moreover,

$$T_{\boldsymbol{u}}(\boldsymbol{A}, \Lambda, \boldsymbol{e}_\Lambda, \eta) \lesssim \sqrt{\frac{|\Lambda|_{1,\boldsymbol{u}}}{N}} \frac{1}{\sigma_{\min}\left(\sqrt{\frac{m}{n}} \boldsymbol{A}^*\right)} \sqrt{L} \max\{\|\boldsymbol{e}_\Lambda\|_2 - \eta, 0\}.$$

The constant $\mathcal{Q}_{\boldsymbol{u}}(\boldsymbol{A})$

Consider the constant

$$\mathcal{Q}_{\boldsymbol{u}}(\boldsymbol{A}) := \sqrt{\frac{|\Lambda|_{1,\boldsymbol{u}}}{N}} \frac{1}{\sigma_{\min}(\sqrt{\frac{m}{n}}\boldsymbol{A}^*)}.$$

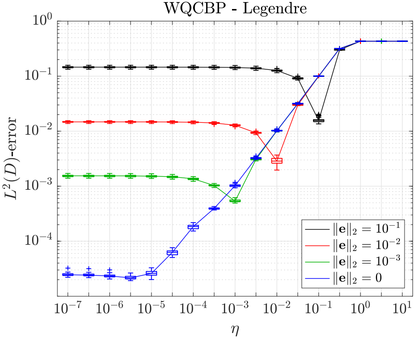
- Close link with the ℓ^1-quotient property of CS [Wojtaszczyk, 2010; Foucart, 2014; B., Adcock, 2017].
- Explicit bound of the form $\mathcal{Q}_{\boldsymbol{u}}(\boldsymbol{A}) \lesssim 1$ in probability can be proved in the 1D case. In general, we can estimate $\mathcal{Q}_{\boldsymbol{u}}(\boldsymbol{A})$ numerically:

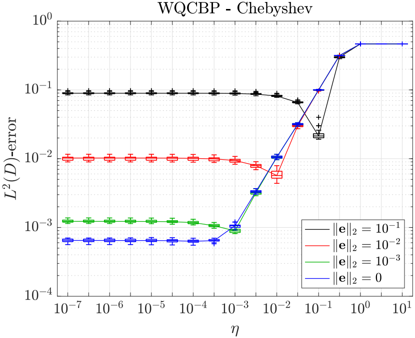
(d,s,N)	m	125	250	375	500	625	750	875	1000
$(8,22,1843)$	Che	2.65	3.07	3.53	3.95	4.46	5.03	5.78	6.82
	Leg	6.45	7.97	8.99	10.5	12.1	13.7	15.8	18.6
(d,s,N)	m	250	500	750	1000	1250	1500	1750	2000
$(16,13,4129)$	Che	2.64	2.93	3.30	3.63	3.99	4.41	4.95	5.62
	Leg	5.64	6.20	6.85	7.60	8.32	8.99	10.1	11.1

Table: The constant $Q_{\boldsymbol{u}}(A)$ (averaged over 50 trials).

The optimal choice of η

The term $\max\{\|\mathbf{e}_{\boldsymbol{\Lambda}}\|_{\mathbf{2}} - \eta, \mathbf{0}\}$ suggests that an optimal choice is $\eta = \|\boldsymbol{e}\|_2$. This is confirmed by numerical experiments, where random noise of a prescribed norm is added to the samples.





Approximation of $f(\boldsymbol{z}) = \exp(-\frac{1}{d}\sum_{i=1}^{d}\cos(z_i))$, with $d = 15$.

In practice, cross validation is employed to estimate the optimal η.

Summary

- CS is a useful tool for parametric PDEs inside / outside the black box.

Benefits:

- Exploit sparsity;
- Ability to capture local features (e.g., boundary layers);
- Easy parallelizability;
- No need for error estimators.

Challenges:

- Accelerate the recovery phase (improve $O(smN)$);
- High-dimensional physical domains;
- Complex geometries;
- Application to nonlocal problems.

Benefits:

- Low impact of the dimensionality d on the sample complexity ($\log(d)$);
- No need to fix the lower set in advance;
- Robustness to unknown error.

Challenges:

- Is it possible to achieve $m \sim s \cdot L$?
- Quantify the decay of $\sigma_{s,L}(\boldsymbol{x})_{1,\boldsymbol{u}}$ depending on the smoothness of f;
- Complex geometries of D;
- Different decoders (e.g., LASSO)

- S. Brugiapaglia. COmpRessed SolvING: sparse approximation of PDEs based on compressed sensing. PhD thesis, MOX - Politecnico di Milano, 2016.
- S. Brugiapaglia, S. Micheletti, and S. Perotto. Compressed solving: A numerical approximation technique for elliptic PDEs based on Compressed Sensing. Comput. Math. Appl., 70(6):1306-1335, 2015.
- S. Brugiapaglia, F. Nobile, S. Micheletti, and S. Perotto. A theoretical study of COmpRessed SolvING for advection-diffusion-reaction problems. Math. Comput., to appear, 2017.

- B. Adcock, C. Bao, and S. Brugiapaglia. Correcting for unknown errors in sparse high-dimensional function approximation. In preparation, 2017.
- B. Adcock, S. Brugiapaglia, and C. G. Webster. Compressed sensing approaches for polynomial approximation of high-dimensional functions. Chapter in "Compressed Sensing and its applications". To appear, 2017. (arXiv:1703.06987)
- S. B., B. Adcock. Robustness to unknown error in sparse regularization. Submitted, 2017. (arXiv:1705.10299)

- S. Brugiapaglia. COmpRessed SolvING: sparse approximation of PDEs based on compressed sensing. PhD thesis, MOX - Politecnico di Milano, 2016.
- S. Brugiapaglia, S. Micheletti, and S. Perotto. Compressed solving: A numerical approximation technique for elliptic PDEs based on Compressed Sensing. Comput. Math. Appl., 70(6):1306-1335, 2015.
- S. Brugiapaglia, F. Nobile, S. Micheletti, and S. Perotto. A theoretical study of COmpRessed SolvING for advection-diffusion-reaction problems. Math. Comput., to appear, 2017.

- B. Adcock, C. Bao, and S. Brugiapaglia. Correcting for unknown errors in sparse high-dimensional function approximation. In preparation, 2017.
- B. Adcock, S. Brugiapaglia, and C. G. Webster. Compressed sensing approaches for polynomial approximation of high-dimensional functions. Chapter in “Compressed Sensing and its applications”. To appear, 2017. (arXiv:1703.06987)
- S. B., B. Adcock. Robustness to unknown error in sparse regularization. Submitted, 2017. (arXiv:1705.10299)

Thank you!