Sparse Recovery of Hilbert-Valued Signals With Application to High-Dimensional Parametric PDEs

Nick Dexter[†], H. Tran[‡], C. G. Webster[†‡]

[†]Mathematics Department
University of Tennessee, Knoxville, TN

[‡]Computational and Applied Mathematics Group
Oak Ridge National Laboratory, Oak Ridge, TN

Supported by: **Department of Energy** (ASCR)

SISSA QUIET Workshop - Trieste, Italy
Thursday, July 19, 2017

Why do we care about "sparse" signals?

Example: We often represent images by expansions like

$$u(\boldsymbol{y}) = \sum_{j=1}^{N} \boldsymbol{z}_j \Psi_j(\boldsymbol{y})$$

where, e.g., $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathbb{R}^N$ and $\{\Psi_j\}_{j=1}^N$ are wavelets.

Figure : **Left:** Original image. **Right:** Image obtained after setting 99.00% of the coefficients $\boldsymbol{z}_j$ in the biorthogonal wavelet transform to 0.

Many practical problems have interesting solutions that are sparse.

Why do we care about "sparse" signals?

Example: We often represent images by expansions like

$$u(\boldsymbol{y}) = \sum_{j=1}^{N} \boldsymbol{z}_j \Psi_j(\boldsymbol{y})$$

where, e.g., $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathbb{R}^N$ and $\{\Psi_j\}_{j=1}^N$ are wavelets.

Figure : **Left:** Original image. **Right:** Image obtained after setting 99.75% of the coefficients $\boldsymbol{z}_j$ in the biorthogonal wavelet transform to 0.

Many practical problems have interesting solutions that are sparse.

Why do we care about "sparse" signals?

Example: We often represent images by expansions like

$$u(\boldsymbol{y}) = \sum_{j=1}^{N} \boldsymbol{z}_j \Psi_j(\boldsymbol{y})$$

where, e.g., $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathbb{R}^N$ and $\{\Psi_j\}_{j=1}^N$ are wavelets.

Figure : **Left:** Original image. **Right:** Image obtained after setting 99.95% of the coefficients $\boldsymbol{z}_j$ in the biorthogonal wavelet transform to 0.

Many practical problems have interesting solutions that are sparse.

Why do we care about "sparse" signals?

Example: We often represent images by expansions like

$$u(\boldsymbol{y}) = \sum_{j=1}^{N} \boldsymbol{z}_j \Psi_j(\boldsymbol{y})$$

where, e.g., $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathbb{R}^N$ and $\{\Psi_j\}_{j=1}^N$ are wavelets.

Figure : **Left:** Original image. **Right:** Image obtained after setting 99.98% of the coefficients $\boldsymbol{z}_j$ in the biorthogonal wavelet transform to 0.

Many practical problems have interesting solutions that are sparse.

Recovery of "sparse" Hilbert-valued signals

Question: Can we find sparse solutions when the vector is Hilbert-valued?

- Let $\mathcal{V}$ be a general Hilbert space, e.g., $L^2(D)$ or $H_0^1(D)$
- Hilbert-valued vector: $\boldsymbol{z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_N) \in \mathcal{V}^N = \bigoplus_{i=1}^N \mathcal{V}$
- $\boldsymbol{z}$ is s-sparse if all but s of its components are $\mathbf{0}$
- $\|\cdot\|$ a norm on $\mathcal{V}^N$, $\|\cdot\|$-error of best s-term approximation to $\boldsymbol{z}$:

$$\sigma_s(\boldsymbol{z}) := \inf\{\|\boldsymbol{z} - \boldsymbol{x}\| : \boldsymbol{x} \in \mathcal{V}^N \text{ is } s\text{-sparse}\}$$

- $\boldsymbol{z}$ is compressible if $\sigma_s(\boldsymbol{z}) \to 0$ quickly as s increases

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Recovery of “sparse” Hilbert-valued signals

Question: Can we find sparse solutions when the vector is Hilbert-valued?

- Let $\mathcal{V}$ be a general Hilbert space, e.g., $L^2(D)$ or $H_0^1(D)$
- Hilbert-valued vector: $\boldsymbol{z} = (z_1, \dots, z_N) \in \mathcal{V}^N = \bigoplus_{i=1}^N \mathcal{V}$
- $\boldsymbol{z}$ is s-sparse if all but s of its components are $\mathbf{0}$
- $\|\cdot\|$ a norm on $\mathcal{V}^N$, $\|\cdot\|$-error of best s-term approximation to $\boldsymbol{z}$:

$$\sigma_s(\boldsymbol{z}) := \inf\{\|\boldsymbol{z} - \boldsymbol{x}\| : \boldsymbol{x} \in \mathcal{V}^N \text{ is } s\text{-sparse}\}$$

- $\boldsymbol{z}$ is compressible if $\sigma_s(\boldsymbol{z}) \to 0$ quickly as s increases

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the “data”) with $m \ll N$, we want to solve $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the “signal”), when $\boldsymbol{z}$ is s-sparse (or compressible).

Recovery of "sparse" Hilbert-valued signals

Question: Can we find sparse solutions when the vector is Hilbert-valued?

- Let $\mathcal{V}$ be a general Hilbert space, e.g., $L^2(D)$ or $H_0^1(D)$
- Hilbert-valued vector: $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathcal{V}^N = \bigoplus_{i=1}^N \mathcal{V}$
- $\boldsymbol{z}$ is s-sparse if all but s of its components are $\mathbf{0}$
- $\|\cdot\|$ a norm on $\mathcal{V}^N$, $\|\cdot\|$-error of best s-term approximation to $\boldsymbol{z}$:

$$\sigma_s(\boldsymbol{z}) := \inf\{\|\boldsymbol{z} - \boldsymbol{x}\| : \boldsymbol{x} \in \mathcal{V}^N \text{ is } s\text{-sparse}\}$$

- $\boldsymbol{z}$ is compressible if $\sigma_s(\boldsymbol{z}) \to 0$ quickly as s increases

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Recovery of "sparse" Hilbert-valued signals

Question: Can we find sparse solutions when the vector is Hilbert-valued?

- Let $\mathcal{V}$ be a general Hilbert space, e.g., $L^2(D)$ or $H_0^1(D)$
- Hilbert-valued vector: $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathcal{V}^N = \bigoplus_{i=1}^N \mathcal{V}$
- $\boldsymbol{z}$ is s-sparse if all but s of its components are $\mathbf{0}$
- $\|\cdot\|$ a norm on $\mathcal{V}^N$, $\|\cdot\|$-error of best s-term approximation to $\boldsymbol{z}$:

$$\sigma_s(\boldsymbol{z}) := \inf\{\|\boldsymbol{z} - \boldsymbol{x}\| : \boldsymbol{x} \in \mathcal{V}^N \text{ is } s\text{-sparse}\}$$

- $\boldsymbol{z}$ is compressible if $\sigma_s(\boldsymbol{z}) \to 0$ quickly as s increases

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Recovery of "sparse" Hilbert-valued signals

Question: Can we find sparse solutions when the vector is Hilbert-valued?

- Let $\mathcal{V}$ be a general Hilbert space, e.g., $L^2(D)$ or $H_0^1(D)$
- Hilbert-valued vector: $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathcal{V}^N = \bigoplus_{i=1}^N \mathcal{V}$
- $\boldsymbol{z}$ is s-sparse if all but s of its components are $\mathbf{0}$
- $\|\cdot\|$ a norm on $\mathcal{V}^N$, $\|\cdot\|$-error of best s-term approximation to $\boldsymbol{z}$:

$$\sigma_s(\boldsymbol{z}) := \inf\{\|\boldsymbol{z} - \boldsymbol{x}\| : \boldsymbol{x} \in \mathcal{V}^N \text{ is } s\text{-sparse}\}$$

- $\boldsymbol{z}$ is compressible if $\sigma_s(\boldsymbol{z}) \to 0$ quickly as s increases

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Recovery of "sparse" Hilbert-valued signals

Question: Can we find sparse solutions when the vector is Hilbert-valued?

- Let $\mathcal{V}$ be a general Hilbert space, e.g., $L^2(D)$ or $H_0^1(D)$
- Hilbert-valued vector: $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathcal{V}^N = \bigoplus_{i=1}^N \mathcal{V}$
- $\boldsymbol{z}$ is s-sparse if all but s of its components are $\mathbf{0}$
- $\|\cdot\|$ a norm on $\mathcal{V}^N$, $\|\cdot\|$-error of best s-term approximation to $\boldsymbol{z}$:

$$\sigma_s(\boldsymbol{z}) := \inf\{\|\boldsymbol{z} - \boldsymbol{x}\| : \boldsymbol{x} \in \mathcal{V}^N \text{ is } s\text{-sparse}\}$$

- $\boldsymbol{z}$ is compressible if $\sigma_s(\boldsymbol{z}) \to 0$ quickly as s increases

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Recovery of "sparse" Hilbert-valued signals

Question: Can we find sparse solutions when the vector is Hilbert-valued?

- Let $\mathcal{V}$ be a general Hilbert space, e.g., $L^2(D)$ or $H_0^1(D)$
- Hilbert-valued vector: $\boldsymbol{z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_N) \in \mathcal{V}^N = \bigoplus_{i=1}^N \mathcal{V}$
- $\boldsymbol{z}$ is s-sparse if all but s of its components are $\mathbf{0}$
- $\|\cdot\|$ a norm on $\mathcal{V}^N$, $\|\cdot\|$-error of best s-term approximation to $\boldsymbol{z}$:

$$\sigma_s(\boldsymbol{z}) := \inf\{\|\boldsymbol{z} - \boldsymbol{x}\| : \boldsymbol{x} \in \mathcal{V}^N \text{ is } s\text{-sparse}\}$$

- $\boldsymbol{z}$ is compressible if $\sigma_s(\boldsymbol{z}) \to 0$ quickly as s increases

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Outline

1. Recovery of “sparse” Hilbert-valued signals
2. Motivating example: parameterized PDE models
3. Compressed sensing for parametric PDE recovery - *“Sparse Recovery of Hilbert-Valued Signals: Theory and Algorithms,” D., Tran, Webster, (in progress)*
   - Foundations
   - Algorithms for Hilbert-valued recovery
   - Numerical examples
4. Concluding remarks

Motivating example: parameterized PDE models

Recovery of solutions to high-dimensional PDEs

parameters
$\boldsymbol{y} \in \Gamma \subset \mathbb{R}^d$
d finite, but large

$\longrightarrow$

PDE model:
$\mathcal{L}(a(\cdot, \boldsymbol{y}))[u(\cdot, \boldsymbol{y})] = 0$
in $D \subset \mathbb{R}^n$, $n = 1, 2, 3$

$\longrightarrow$

quantity of interest
$Q[u(\cdot, \boldsymbol{y})]$

Example: Stochastic elliptic problem on $\overline{D} \times \Gamma$

$$\begin{cases} -\nabla \cdot (a(x, \boldsymbol{y}) \nabla u(x, \boldsymbol{y})) = f(x) & \text{in } D \times \Gamma, \\ u(x, \boldsymbol{y}) = 0 & \text{on } \partial D \times \Gamma. \end{cases} \tag{1}$$

- $a(x, \boldsymbol{y})$ is a random field and $f \in L^2(D)$
- $\boldsymbol{y} = (y_1, \ldots, y_d)$ with y_i i.i.d. bounded, e.g., $y_i \sim \mathcal{U}(-1, 1)$

Goal-2: Approximate the solution map $\boldsymbol{y} \mapsto u(\cdot, \boldsymbol{y})$ globally in D via an expansion

$$\boxed{u(\cdot, \boldsymbol{y}) \approx u_{\Lambda_0}(\cdot, \boldsymbol{y}) := \sum\nolimits_{j \in \Lambda_0} \boldsymbol{z}_j(\cdot) \Psi_j(\boldsymbol{y}).}$$

- $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathcal{V}^N$, $\mathcal{V}$ a Hilbert space, in this case $\mathcal{V} = H_0^1(D)$
- $\{\Psi_j\}_{j \in \Lambda_0}$ are, e.g., an orthonormal basis of $\mathcal{P}_{\Lambda_0}(\Gamma)$ or interpolating polynomials

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

Example: Stochastic elliptic problem on $\overline{D} \times \Gamma$

$$\begin{cases} -\nabla \cdot (a(x,\boldsymbol{y})\nabla u(x,\boldsymbol{y})) = f(x) & \text{in } D \times \Gamma, \\ u(x,\boldsymbol{y}) = 0 & \text{on } \partial D \times \Gamma. \end{cases} \tag{1}$$

- $a(x,\boldsymbol{y})$ is a random field and $f \in L^2(D)$
- $\boldsymbol{y} = (y_1, \ldots, y_d)$ with y_i i.i.d. bounded, e.g., $y_i \sim \mathcal{U}(-1,1)$

Goal-2: Approximate the solution map $\boldsymbol{y} \mapsto u(\,\cdot\,,\boldsymbol{y})$ globally in D via an expansion

$$\boxed{u(\,\cdot\,,\boldsymbol{y}) \approx u_{\Lambda_0}(\,\cdot\,,\boldsymbol{y}) := \sum_{j\in\Lambda_0} \boldsymbol{z}_j(\,\cdot\,)\Psi_j(\boldsymbol{y}).}$$

- $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathcal{V}^N$, $\mathcal{V}$ a Hilbert space, in this case $\mathcal{V} = H_0^1(D)$
- $\{\Psi_j\}_{j\in\Lambda_0}$ are, e.g., an orthonormal basis of $\mathcal{P}_{\Lambda_0}(\Gamma)$ or interpolating polynomials

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

Example: Stochastic elliptic problem on $\overline{D} \times \Gamma$

$$\begin{cases} -\nabla \cdot (a(x,\boldsymbol{y})\nabla u(x,\boldsymbol{y})) = f(x) & \text{in } D \times \Gamma, \\ u(x,\boldsymbol{y}) = 0 & \text{on } \partial D \times \Gamma. \end{cases} \tag{1}$$

- $a(x,\boldsymbol{y})$ is a random field and $f \in L^2(D)$
- $\boldsymbol{y} = (y_1, \ldots, y_d)$ with y_i i.i.d. bounded, e.g., $y_i \sim \mathcal{U}(-1,1)$

Goal-2: Approximate the solution map $\boldsymbol{y} \mapsto u(\,\cdot\,,\boldsymbol{y})$ globally in D via an expansion

$$\boxed{u(\,\cdot\,,\boldsymbol{y}) \approx u_{\Lambda_0}(\,\cdot\,,\boldsymbol{y}) := \sum\nolimits_{j\in\Lambda_0} \boldsymbol{z}_j(\,\cdot\,)\Psi_j(\boldsymbol{y}).}$$

- $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N) \in \mathcal{V}^N$, $\mathcal{V}$ a Hilbert space, in this case $\mathcal{V} = H_0^1(D)$
- $\{\Psi_j\}_{j\in\Lambda_0}$ are, e.g., an orthonormal basis of $\mathcal{P}_{\Lambda_0}(\Gamma)$ or interpolating polynomials

Motivating example: parameterized PDE models

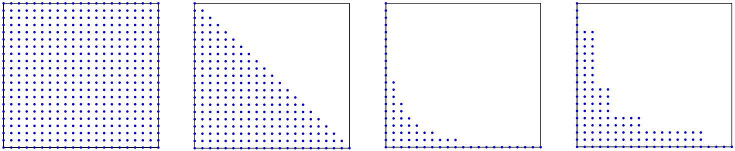
First steps → truncation (selection of the index sets in high-dimensions)

Goal-2: Approximate the solution map $\boldsymbol{y} \mapsto u(\cdot, \boldsymbol{y})$ globally in D via an expansion

$$\boxed{u(\cdot, \boldsymbol{y}) \approx u_{\Lambda_0}(\cdot, \boldsymbol{y}) := \sum_{j \in \Lambda_0} \boldsymbol{z}_j(\cdot)\Psi_j(\boldsymbol{y}).}$$

The first step is to truncate the infinite expansion $u = \sum_j \boldsymbol{z}_j \Psi_j$.

- Common choices of the multi-index set Λ_0:



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

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

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

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

- Each choice induces a truncation error $\eta_{\Lambda_0} := \|u - u_{\Lambda_0}\| = \left\| \sum_{j \notin \Lambda_0} \boldsymbol{z}_j \Psi_j \right\| = \|\boldsymbol{z}_{\Lambda_0^c}\|$
- Algorithm costs scale poorly with dimension, $N := \#(\Lambda_0)$ grows quickly
- Optimal choice: s most effective indices to minimize η (unknown in general)

Motivating example: parameterized PDE models

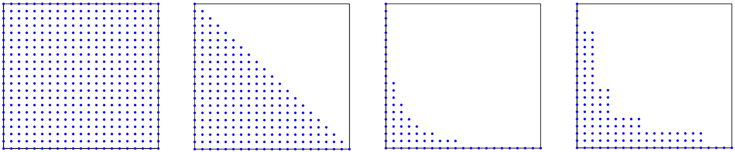
First steps → truncation (selection of the index sets in high-dimensions)

Goal-2: Approximate the solution map $\boldsymbol{y} \mapsto u(\cdot, \boldsymbol{y})$ globally in D via an expansion

$$\boxed{u(\cdot, \boldsymbol{y}) \approx u_{\Lambda_0}(\cdot, \boldsymbol{y}) := \sum\nolimits_{j \in \Lambda_0} \boldsymbol{z}_j(\cdot)\Psi_j(\boldsymbol{y}).}$$

The first step is to truncate the infinite expansion $u = \sum_j \boldsymbol{z}_j \Psi_j$.

- Common choices of the multi-index set Λ_0:



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

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

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

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

- Each choice induces a truncation error $\eta_{\Lambda_0} := \|u - u_{\Lambda_0}\| = \left\| \sum_{j \notin \Lambda_0} \boldsymbol{z}_j \Psi_j \right\| = \|\boldsymbol{z}_{\Lambda_0^c}\|$
- Algorithm costs scale poorly with dimension, $N := \#(\Lambda_0)$ grows quickly
- Optimal choice: s most effective indices to minimize η (unknown in general)

Motivating example: parameterized PDE models

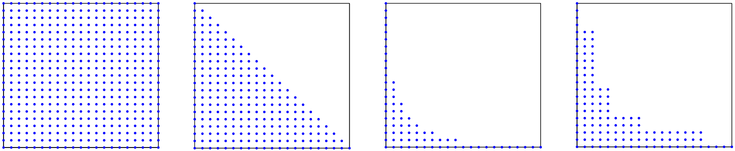
First steps $\to$ truncation (selection of the index sets in high-dimensions)

Goal-2: Approximate the solution map $\boldsymbol{y} \mapsto u(\cdot, \boldsymbol{y})$ globally in D via an expansion

$$\boxed{u(\cdot, \boldsymbol{y}) \approx u_{\Lambda_0}(\cdot, \boldsymbol{y}) := \sum_{j \in \Lambda_0} \boldsymbol{z}_j(\cdot) \Psi_j(\boldsymbol{y}).}$$

The first step is to truncate the infinite expansion $u = \sum_j \boldsymbol{z}_j \Psi_j$.

- Common choices of the multi-index set Λ_0:



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

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

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

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

- Each choice induces a truncation error $\eta_{\Lambda_0} := \|u - u_{\Lambda_0}\| = \left\| \sum_{j \notin \Lambda_0} \boldsymbol{z}_j \Psi_j \right\| = \|\boldsymbol{z}_{\Lambda_0^c}\|$
- Algorithm costs scale poorly with dimension, $N := \#(\Lambda_0)$ grows quickly
- Optimal choice: s most effective indices to minimize η (unknown in general)

Motivating example: parameterized PDE models

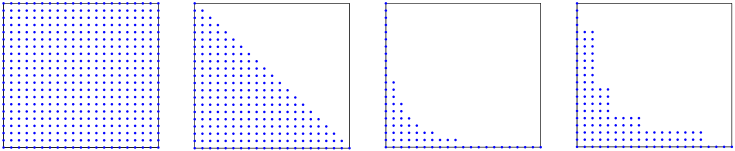
First steps $\rightarrow$ truncation (selection of the index sets in high-dimensions)

Goal-2: Approximate the solution map $\boldsymbol{y} \mapsto u(\cdot, \boldsymbol{y})$ globally in D via an expansion

$$u(\cdot, \boldsymbol{y}) \approx u_{\Lambda_0}(\cdot, \boldsymbol{y}) := \sum_{j \in \Lambda_0} \boldsymbol{z}_j(\cdot) \Psi_j(\boldsymbol{y}).$$

The first step is to truncate the infinite expansion $u = \sum_j \boldsymbol{z}_j \Psi_j$.

- Common choices of the multi-index set Λ_0:



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

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

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

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

- Each choice induces a truncation error $\eta_{\Lambda_0} := \|u - u_{\Lambda_0}\| = \left\| \sum_{j \notin \Lambda_0} \boldsymbol{z}_j \Psi_j \right\| = \|\boldsymbol{z}_{\Lambda_0^c}\|$
- Algorithm costs scale poorly with dimension, $N := \#(\Lambda_0)$ grows quickly
- Optimal choice: s most effective indices to minimize η (unknown in general)

Question: When can we hope that the solution to a parameterized PDE, e.g.,

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

is sparse (compressible)?

Assuming

1. Coercivity and continuity of a: there exists $0 < a_{\min} \leq a_{\max}$ such that $a_{\min} \leq a \leq a_{\max}$ uniformly in $\overline{D} \times \Gamma$.
2. Holomorphic parameter dependence: complex continuation, $a^* : \mathbb{C}^d \to L^\infty(D)$, is an $L^\infty(D)$-valued holomorphic function on $\mathbb{C}^d$.

Then the best s-term approximation u_{Λ_s} obeys

$$\|u - u_{\Lambda_s}\|^2_{\mathcal{V},2} = \underbrace{\sigma_s(\boldsymbol{z}_{\Lambda_s})_{\mathcal{V},2}}_{\text{finite part}} + \underbrace{\eta_{\Lambda_s}}_{\text{infinite part}} \lesssim s^2 \exp(-2(\kappa s)^{1/d}),$$

κ depending on the size and shape of Λ_s [Tran, Webster, Zhang '16].

Question: When can we hope that the solution to a parameterized PDE, e.g.,

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

is sparse (compressible)?

Assuming

1. Coercivity and continuity of a: there exists $0 < a_{\min} \leq a_{\max}$ such that $a_{\min} \leq a \leq a_{\max}$ uniformly in $\overline{D} \times \Gamma$.
2. Holomorphic parameter dependence: complex continuation, $a^* : \mathbb{C}^d \to L^\infty(D)$, is an $L^\infty(D)$-valued holomorphic function on $\mathbb{C}^d$.

Then the best s-term approximation u_{Λ_s} obeys

$$\|u - u_{\Lambda_s}\|^2_{\mathcal{V},2} = \underbrace{\sigma_s(\boldsymbol{z}_{\Lambda_s})_{\mathcal{V},2}}_{\text{finite part}} + \underbrace{\eta_{\Lambda_s}}_{\text{infinite part}} \lesssim s^2 \exp(-2(\kappa s)^{1/d}),$$

κ depending on the size and shape of Λ_s [Tran, Webster, Zhang '16].

Question: When can we hope that the solution to a parameterized PDE, e.g.,

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

is sparse (compressible)?

Assuming

1. Coercivity and continuity of a: there exists $0 < a_{\min} \leq a_{\max}$ such that $a_{\min} \leq a \leq a_{\max}$ uniformly in $\overline{D} \times \Gamma$.
2. Holomorphic parameter dependence: complex continuation, $a^* : \mathbb{C}^d \to L^\infty(D)$, is an $L^\infty(D)$-valued holomorphic function on $\mathbb{C}^d$.

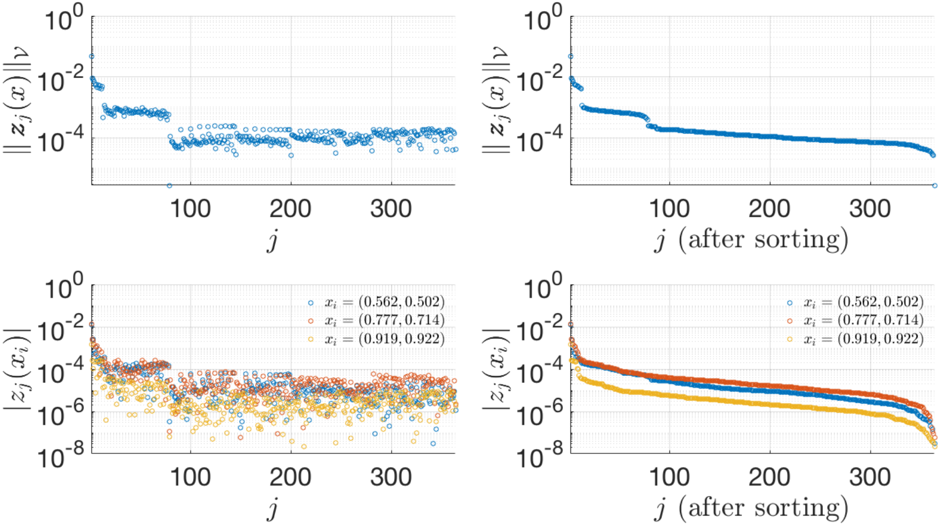
Then the best s-term approximation u_{Λ_s} obeys

$$\|u - u_{\Lambda_s}\|^2_{\mathcal{V},2} = \underbrace{\sigma_s(\boldsymbol{z}_{\Lambda_s})_{\mathcal{V},2}}_{\text{finite part}} + \underbrace{\eta_{\Lambda_s}}_{\text{infinite part}} \lesssim s^2 \exp(-2(\kappa s)^{1/d}),$$

κ depending on the size and shape of Λ_s [Tran, Webster, Zhang '16].

Motivating example: parameterized PDE models

Log transformed KL example: $a(x,\boldsymbol{y}) \approx 0.5 + \exp(\varphi_0 + \sum_{k=1}^{d} \sqrt{\lambda_k}\varphi_k y_k)$, $d = 11$, $L_c = 1/64$



Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Question: How can we set up this problem to approximate solutions to parametric PDEs?

Standard measurement scheme for compressed sensing (CS) requires:

- Multi-index set Λ_0 of size N
- Random samples $\{\boldsymbol{y}_i\}_{i=1}^m \subset \Gamma$ drawn from a measure $\varrho(\boldsymbol{y})$, e.g., Monte Carlo
- Bounded orthonormal system (BOS) $\{\Psi_j\}_{j\in\Lambda_0}$, i.e.,

$$\langle \Psi_j, \Psi_k \rangle_\varrho = \delta_{j,k} \quad \forall j,k \qquad \text{and} \qquad \sup_{j\in\Lambda_0} \|\Psi_j\|_{L^\infty_\varrho(\Gamma)} = \Theta < \infty$$

Set: $[\boldsymbol{\Psi}]_{i,j} = \Psi_j(\boldsymbol{y}_i)$ and $[\boldsymbol{u}]_i = u(\boldsymbol{y}_i) \in \mathcal{V}$ $\quad 1 \le i \le m, \ 1 \le j \le N$

Then our "operator" (matrix) $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ is defined by the action

$$[\boldsymbol{\Psi}\boldsymbol{z}]_i = \sum_{j=1}^N z_j \Psi_j(\boldsymbol{y}_i) \in \mathcal{V} \qquad \boldsymbol{z} \in \mathcal{V}^N, \ 1 \le i \le m.$$

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Question: How can we set up this problem to approximate solutions to parametric PDEs?

Standard measurement scheme for compressed sensing (CS) requires:

- Multi-index set Λ_0 of size N
- Random samples $\{\boldsymbol{y}_i\}_{i=1}^m \subset \Gamma$ drawn from a measure $\varrho(\boldsymbol{y})$, e.g., Monte Carlo
- Bounded orthonormal system (BOS) $\{\Psi_j\}_{j \in \Lambda_0}$, i.e.,

$$\langle \Psi_j, \Psi_k \rangle_\varrho = \delta_{j,k} \;\; \forall j,k \qquad \text{and} \qquad \sup_{j \in \Lambda_0} \|\Psi_j\|_{L^\infty_\varrho(\Gamma)} = \Theta < \infty$$

Set: $[\boldsymbol{\Psi}]_{i,j} = \Psi_j(\boldsymbol{y}_i)$ and $[\boldsymbol{u}]_i = u(\boldsymbol{y}_i) \in \mathcal{V}$ $\quad 1 \le i \le m, \; 1 \le j \le N$

Then our "operator" (matrix) $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ is defined by the action

$$[\boldsymbol{\Psi z}]_i = \sum_{j=1}^N z_j \Psi_j(\boldsymbol{y}_i) \in \mathcal{V} \qquad \boldsymbol{z} \in \mathcal{V}^N, \; 1 \le i \le m.$$

Goal-1: Given an operator $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ and $\boldsymbol{u} \in \mathcal{V}^m$ (the "data") with $m \ll N$, we want to solve $\boldsymbol{\Psi z} = \boldsymbol{u}$ for $\boldsymbol{z}$ (the "signal"), when $\boldsymbol{z}$ is s-sparse (or compressible).

Question: How can we set up this problem to approximate solutions to parametric PDEs?

Standard measurement scheme for compressed sensing (CS) requires:

- Multi-index set Λ_0 of size N
- Random samples $\{\boldsymbol{y}_i\}_{i=1}^m \subset \Gamma$ drawn from a measure $\varrho(\boldsymbol{y})$, e.g., Monte Carlo
- Bounded orthonormal system (BOS) $\{\Psi_j\}_{j\in\Lambda_0}$, i.e.,

$$\langle \Psi_j, \Psi_k \rangle_\varrho = \delta_{j,k} \ \ \forall j,k \qquad \text{and} \qquad \sup_{j\in\Lambda_0} \|\Psi_j\|_{L^\infty_\varrho(\Gamma)} = \Theta < \infty$$

Set: $\quad [\boldsymbol{\Psi}]_{i,j} = \Psi_j(\boldsymbol{y}_i) \qquad \text{and} \qquad [\boldsymbol{u}]_i = u(\boldsymbol{y}_i) \in \mathcal{V} \qquad 1 \le i \le m, \ \ 1 \le j \le N$

Then our "operator" (matrix) $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ is defined by the action

$$[\boldsymbol{\Psi z}]_i = \sum_{j=1}^N \boldsymbol{z}_j \Psi_j(\boldsymbol{y}_i) \in \mathcal{V} \qquad \boldsymbol{z} \in \mathcal{V}^N, \ \ 1 \le i \le m.$$

It is easy to see that:

- $(\mathcal{V}^N, \langle \cdot\,, \cdot \rangle_{\mathcal{V},2})$, with $\langle \boldsymbol{z}, \boldsymbol{z}' \rangle_{\mathcal{V},2} := \sum_{j\in\Lambda_0} \langle \boldsymbol{z}_j, \boldsymbol{z}_j' \rangle_{\mathcal{V}}$ is a Hilbert space
- $(\mathcal{V}^N, \|\cdot\|_{\mathcal{V},p})$, with $\|\boldsymbol{z}\|_{\mathcal{V},p} := \left(\sum_{j\in\Lambda_0} \|\boldsymbol{z}_j\|_{\mathcal{V}}^p\right)^{1/p}$ is a Banach space

For $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ we define the operator norm as

$$\|\boldsymbol{\Psi}\|_{p\to q} = \sup_{\|\boldsymbol{x}\|_{\mathcal{V},p}=1} \|\boldsymbol{\Psi}\boldsymbol{x}\|_{\mathcal{V},q} = \sup_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\|\boldsymbol{\Psi}\boldsymbol{x}\|_{\mathcal{V},q}}{\|\boldsymbol{x}\|_{\mathcal{V},p}} \quad \text{with} \quad \|\boldsymbol{\Psi}\|_p := \|\boldsymbol{\Psi}\|_{p\to p}.$$

The inner product also allows us to define adjoints

$$\langle \boldsymbol{\Psi}\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{V},2} = \langle \boldsymbol{x}, \boldsymbol{\Psi}^*\boldsymbol{y} \rangle_{\mathcal{V},2}$$

in the standard way.

These facts give us the necessary structure to establish many of the recovery guarantees and convergence results that hold in the real and complex-valued cases.

It is easy to see that:

- $(\mathcal{V}^N, \langle \cdot, \cdot \rangle_{\mathcal{V},2})$, with $\langle \boldsymbol{z}, \boldsymbol{z}' \rangle_{\mathcal{V},2} := \sum_{j\in\Lambda_0} \langle \boldsymbol{z}_j, \boldsymbol{z}'_j \rangle_{\mathcal{V}}$ is a Hilbert space
- $(\mathcal{V}^N, \|\cdot\|_{\mathcal{V},p})$, with $\|\boldsymbol{z}\|_{\mathcal{V},p} := \left(\sum_{j\in\Lambda_0} \|\boldsymbol{z}_j\|_{\mathcal{V}}^p\right)^{1/p}$ is a Banach space

For $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ we define the operator norm as

$$\|\boldsymbol{\Psi}\|_{p\to q} = \sup_{\|\boldsymbol{x}\|_{\mathcal{V},p}=1} \|\boldsymbol{\Psi}\boldsymbol{x}\|_{\mathcal{V},q} = \sup_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\|\boldsymbol{\Psi}\boldsymbol{x}\|_{\mathcal{V},q}}{\|\boldsymbol{x}\|_{\mathcal{V},p}} \quad \text{with} \quad \|\boldsymbol{\Psi}\|_p := \|\boldsymbol{\Psi}\|_{p\to p}.$$

The inner product also allows us to define adjoints

$$\langle \boldsymbol{\Psi}\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{V},2} = \langle \boldsymbol{x}, \boldsymbol{\Psi}^*\boldsymbol{y} \rangle_{\mathcal{V},2}$$

in the standard way.

These facts give us the necessary structure to establish many of the recovery guarantees and convergence results that hold in the real and complex-valued cases.

It is easy to see that:

- $(\mathcal{V}^N, \langle \cdot\,, \cdot \rangle_{\mathcal{V},2})$, with $\langle \boldsymbol{z}, \boldsymbol{z}' \rangle_{\mathcal{V},2} := \sum_{j\in\Lambda_0} \langle \boldsymbol{z}_j, \boldsymbol{z}_j' \rangle_{\mathcal{V}}$ is a Hilbert space
- $(\mathcal{V}^N, \|\cdot\|_{\mathcal{V},p})$, with $\|\boldsymbol{z}\|_{\mathcal{V},p} := \left(\sum_{j\in\Lambda_0} \|\boldsymbol{z}_j\|_{\mathcal{V}}^p \right)^{1/p}$ is a Banach space

For $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ we define the operator norm as

$$\|\boldsymbol{\Psi}\|_{p\to q} = \sup_{\|\boldsymbol{x}\|_{\mathcal{V},p}=1} \|\boldsymbol{\Psi x}\|_{\mathcal{V},q} = \sup_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\|\boldsymbol{\Psi x}\|_{\mathcal{V},q}}{\|\boldsymbol{x}\|_{\mathcal{V},p}} \qquad \text{with} \quad \|\boldsymbol{\Psi}\|_p := \|\boldsymbol{\Psi}\|_{p\to p}.$$

The inner product also allows us to define adjoints

$$\langle \boldsymbol{\Psi x}, \boldsymbol{y} \rangle_{\mathcal{V},2} = \langle \boldsymbol{x}, \boldsymbol{\Psi}^* \boldsymbol{y} \rangle_{\mathcal{V},2}$$

in the standard way.

These facts give us the necessary structure to establish many of the recovery guarantees and convergence results that hold in the real and complex-valued cases.

It is easy to see that:

- $(\mathcal{V}^N, \langle \cdot , \cdot \rangle_{\mathcal{V},2})$, with $\langle \boldsymbol{z}, \boldsymbol{z}' \rangle_{\mathcal{V},2} := \sum_{j \in \Lambda_0} \langle \boldsymbol{z}_j, \boldsymbol{z}'_j \rangle_{\mathcal{V}}$ is a Hilbert space
- $(\mathcal{V}^N, \| \cdot \|_{\mathcal{V},p})$, with $\|\boldsymbol{z}\|_{\mathcal{V},p} := \left(\sum_{j \in \Lambda_0} \|\boldsymbol{z}_j\|_{\mathcal{V}}^p \right)^{1/p}$ is a Banach space

For $\boldsymbol{\Psi} : \mathcal{V}^N \to \mathcal{V}^m$ we define the operator norm as

$$\|\boldsymbol{\Psi}\|_{p \to q} = \sup_{\|\boldsymbol{x}\|_{\mathcal{V},p} = 1} \|\boldsymbol{\Psi x}\|_{\mathcal{V},q} = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|\boldsymbol{\Psi x}\|_{\mathcal{V},q}}{\|\boldsymbol{x}\|_{\mathcal{V},p}} \qquad \text{with} \quad \|\boldsymbol{\Psi}\|_p := \|\boldsymbol{\Psi}\|_{p \to p}.$$

The inner product also allows us to define adjoints

$$\langle \boldsymbol{\Psi x}, \boldsymbol{y} \rangle_{\mathcal{V},2} = \langle \boldsymbol{x}, \boldsymbol{\Psi}^* \boldsymbol{y} \rangle_{\mathcal{V},2}$$

in the standard way.

These facts give us the necessary structure to establish many of the recovery guarantees and convergence results that hold in the real and complex-valued cases.

Compressed sensing for parametric PDE recovery

In compressed sensing, uniform recovery is guaranteed by the **restricted isometry property (RIP)** of the normalized matrix $\tilde{\boldsymbol{\Psi}} = \frac{1}{\sqrt{m}}\boldsymbol{\Psi}$.

RIP for $\mathcal{V}^N$: there exists a small $\delta_{\mathcal{V},s}$ such that for all s-sparse $\boldsymbol{z} \in \mathcal{V}^N$

$$(1-\delta_{\mathcal{V},s})\|\boldsymbol{z}\|_{\mathcal{V},2}^2 \le \|\tilde{\boldsymbol{\Psi}}\boldsymbol{z}\|_{\mathcal{V},2}^2 \le (1+\delta_{\mathcal{V},s})\|\boldsymbol{z}\|_{\mathcal{V},2}^2 \qquad (\mathcal{V}\text{-RIP})$$

Theorem [D., Tran, Webster '16]

- A matrix $\tilde{\boldsymbol{\Psi}}$ 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. Hence if, for $\delta \in (0,1)$,

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

then with high probability $\tilde{\boldsymbol{\Psi}}$ satisfies the $\mathcal{V}$-RIP with $\delta_s \le \delta$.

Compressed sensing for parametric PDE recovery

In compressed sensing, uniform recovery is guaranteed by the **restricted isometry property (RIP)** of the normalized matrix $\tilde{\mathbf{\Psi}} = \frac{1}{\sqrt{m}}\mathbf{\Psi}$.

RIP for $\mathcal{V}^N$: there exists a small $\delta_{\mathcal{V},s}$ such that for all s-sparse $\boldsymbol{z} \in \mathcal{V}^N$

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

Theorem [D., Tran, Webster '16]

- A matrix $\tilde{\Psi}$ 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. Hence if, for $\delta \in (0,1)$,

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

then with high probability $\tilde{\Psi}$ satisfies the $\mathcal{V}$-RIP with $\delta_s \leq \delta$.

In compressed sensing, uniform recovery is guaranteed by the **restricted isometry property (RIP)** of the normalized matrix $\tilde{\boldsymbol{\Psi}} = \frac{1}{\sqrt{m}}\boldsymbol{\Psi}$.

RIP for $\mathcal{V}^N$: there exists a small $\delta_{\mathcal{V},s}$ such that for all s-sparse $\boldsymbol{z} \in \mathcal{V}^N$

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

Theorem [D., Tran, Webster '16]

- A matrix $\tilde{\boldsymbol{\Psi}}$ 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. Hence if, for $\delta \in (0,1)$,

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

then with high probability $\tilde{\boldsymbol{\Psi}}$ satisfies the $\mathcal{V}$-RIP with $\delta_s \leq \delta$.

Compressed sensing for parametric PDE recovery

The recovery guarantees from the last theorem say that we can expect to find good approximations to $\boldsymbol{\Psi z} = \boldsymbol{u}$ when $\boldsymbol{z}$ is s-sparse (compressible), if we take enough samples.

Solving the following constrained optimization problem can give sparse approximations.

Basis pursuit denoising (BPDN) problem for $\mathcal{V}^N$:

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

- Can cast the BPDN problem as an unconstrained convex optimization problem
- **Long** history of research on iterative methods for fixed-point problems on Hilbert spaces dating back to the 1950-60's with *well-developed* convergence theory
- Easy to show that if δ_{2s} for $\tilde{\boldsymbol{\Psi}}$ satisfies $\delta_{2s} < 4/\sqrt{41}$, then solutions $\boldsymbol{z}^{\#}$ of the BPDN problem approx. the true solution $\boldsymbol{z}$ (satisfying the constraint) with error

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

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

Compressed sensing for parametric PDE recovery

The recovery guarantees from the last theorem say that we can expect to find good approximations to $\boldsymbol{\Psi z} = \boldsymbol{u}$ when $\boldsymbol{z}$ is s-sparse (compressible), if we take enough samples.

Solving the following constrained optimization problem can give sparse approximations.

Basis pursuit denoising (BPDN) problem for $\mathcal{V}^N$:

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

- Can cast the BPDN problem as an unconstrained convex optimization problem
- Long history of research on iterative methods for fixed-point problems on Hilbert spaces dating back to the 1950-60's with *well-developed* convergence theory
- Easy to show that if δ_{2s} for $\tilde{\boldsymbol{\Psi}}$ satisfies $\delta_{2s} < 4/\sqrt{41}$, then solutions $\boldsymbol{z}^{\#}$ of the BPDN problem approx. the true solution $\boldsymbol{z}$ (satisfying the constraint) with error

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

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

Compressed sensing for parametric PDE recovery

The recovery guarantees from the last theorem say that we can expect to find good approximations to $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ when $\boldsymbol{z}$ is s-sparse (compressible), if we take enough samples.

Solving the following constrained optimization problem can give sparse approximations.

Basis pursuit denoising (BPDN) problem for $\mathcal{V}^N$:

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

- Can cast the BPDN problem as an unconstrained convex optimization problem
- Long history of research on iterative methods for fixed-point problems on Hilbert spaces dating back to the 1950-60's with *well-developed* convergence theory
- Easy to show that if δ_{2s} for $\tilde{\boldsymbol{\Psi}}$ satisfies $\delta_{2s} < 4/\sqrt{41}$, then solutions $\boldsymbol{z}^{\#}$ of the BPDN problem approx. the true solution $\boldsymbol{z}$ (satisfying the constraint) with error

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

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

Compressed sensing for parametric PDE recovery

The recovery guarantees from the last theorem say that we can expect to find good approximations to $\boldsymbol{\Psi z} = \boldsymbol{u}$ when $\boldsymbol{z}$ is s-sparse (compressible), if we take enough samples.

Solving the following constrained optimization problem can give sparse approximations.

Basis pursuit denoising (BPDN) problem for $\mathcal{V}^N$:

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

- Can cast the BPDN problem as an unconstrained convex optimization problem
- **Long** history of research on iterative methods for fixed-point problems on Hilbert spaces dating back to the 1950-60's with *well-developed* convergence theory
- Easy to show that if δ_{2s} for $\tilde{\boldsymbol{\Psi}}$ satisfies $\delta_{2s} < 4/\sqrt{41}$, then solutions $\boldsymbol{z}^\#$ of the BPDN problem approx. the true solution $\boldsymbol{z}$ (satisfying the constraint) with error

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

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

Compressed sensing for parametric PDE recovery

The recovery guarantees from the last theorem say that we can expect to find good approximations to $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$ when $\boldsymbol{z}$ is s-sparse (compressible), if we take enough samples.

Solving the following constrained optimization problem can give sparse approximations.

Basis pursuit denoising (BPDN) problem for $\mathcal{V}^N$:

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

- Can cast the BPDN problem as an unconstrained convex optimization problem
- **Long** history of research on iterative methods for fixed-point problems on Hilbert spaces dating back to the 1950-60's with *well-developed* convergence theory
- Easy to show that if δ_{2s} for $\tilde{\boldsymbol{\Psi}}$ satisfies $\delta_{2s} < 4/\sqrt{41}$, then solutions $\boldsymbol{z}^\#$ of the BPDN problem approx. the true solution $\boldsymbol{z}$ (satisfying the constraint) with error

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

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

Goal-3: Find solutions to the basis pursuit denoising problem over $\mathcal{V}^N$:

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

Strategy: Extend algorithms for real-valued recovery to recovery in $\mathcal{V}^N$

- **Forward-backward splitting:** [Lions, Mercier '79], [Chen, Rockafeller '89], [Daubechies, Defrise, De Mol '04], [Combettes '04], and in [Hale, Yin, Zhang '08] was applied to compressed sensing problems with a continuation strategy (FPC),
- **Bregman iterations:** for Total Variation-based image restoration [Osher, Burger, Goldfarb, Xu, Yin '05] and applied to compressed sensing in [Yin, Osher, Goldfarb, Darbon '08]. Equivalent to the augmented Lagrangian method under certain parameterizations, and has nice error-forgetting properties [Yin, Osher '12].

Challenges:

- Proving strong convergence in this setting
- Implementing and parallelizing these algorithms

Goal-3: Find solutions to the basis pursuit denoising problem over $\mathcal{V}^N$:

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

Strategy: Extend algorithms for real-valued recovery to recovery in $\mathcal{V}^N$

- **Forward-backward splitting:** [Lions, Mercier '79], [Chen, Rockafeller '89], [Daubechies, Defrise, De Mol '04], [Combettes '04], and in [Hale, Yin, Zhang '08] was applied to compressed sensing problems with a continuation strategy (FPC),
- **Bregman iterations:** for Total Variation-based image restoration [Osher, Burger, Goldfarb, Xu, Yin '05] and applied to compressed sensing in [Yin, Osher, Goldfarb, Darbon '08]. Equivalent to the augmented Lagrangian method under certain parameterizations, and has nice error-forgetting properties [Yin, Osher '12].

Challenges:

- Proving strong convergence in this setting
- Implementing and parallelizing these algorithms

Goal-3: Find solutions to the basis pursuit denoising problem over $\mathcal{V}^N$:

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

Strategy: Extend algorithms for real-valued recovery to recovery in $\mathcal{V}^N$

- **Forward-backward splitting:** [Lions, Mercier '79], [Chen, Rockafeller '89], [Daubechies, Defrise, De Mol '04], [Combettes '04], and in [Hale, Yin, Zhang '08] was applied to compressed sensing problems with a continuation strategy (FPC),
- **Bregman iterations:** for Total Variation-based image restoration [Osher, Burger, Goldfarb, Xu, Yin '05] and applied to compressed sensing in [Yin, Osher, Goldfarb, Darbon '08]. Equivalent to the augmented Lagrangian method under certain parameterizations, and has nice error-forgetting properties [Yin, Osher '12].

Challenges:

- Proving strong convergence in this setting
- Implementing and parallelizing these algorithms

Compressed sensing for parametric PDE recovery - algorithms

Formulation of the forward-backward splitting method

Can solve (2) by solving the related problem for appropriately chosen values of μ:

$$\text{minimize}_{\boldsymbol{z}\in\mathcal{V}^N} \underbrace{\|\boldsymbol{z}\|_{\mathcal{V},1} + \frac{\mu}{2}\|\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}\|_{\mathcal{V},2}^2}_{=:\, F_\mu(\boldsymbol{z})}. \tag{3}$$

Recall: the subdifferential of a proper function $F : \mathcal{V}^N \to (-\infty, \infty]$ at a point $\boldsymbol{x} \in \mathcal{V}^N$ is the set-valued operator

$$\partial F(\boldsymbol{x}) = \left\{ \boldsymbol{v} \in \mathcal{V}^N : F(\boldsymbol{z}) \geq F(\boldsymbol{x}) + \langle \boldsymbol{v}, \boldsymbol{z} - \boldsymbol{x} \rangle \text{ for all } \boldsymbol{z} \in \mathcal{V}^N \right\}. \tag{4}$$

The elements of $\partial F(\boldsymbol{x})$ are called subgradients of F at $\boldsymbol{x}$. When the function F is convex and differentiable at $\boldsymbol{x}$, $\partial F(\boldsymbol{x}) = \{\nabla F(\boldsymbol{x})\}$, i.e. ∂F is single-valued.

Define the splitting $\partial F_\mu(\boldsymbol{z}) = T_1(\boldsymbol{z}) + T_2(\boldsymbol{z})$, where

- $T_1(\boldsymbol{z}) = \partial\|\boldsymbol{z}\|_{\mathcal{V},1} \leftarrow$ non-differentiable part of F_μ
- $T_2(\boldsymbol{z}) = \frac{\mu}{2}\nabla\|\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}\|_{\mathcal{V},2}^2 \leftarrow$ differentiable part of F_μ

T_1, T_2 are (sub)gradients of proper l.s.c. convex fcns., hence maximal monitone operators $\Rightarrow$ (3) is an instance of a monitone inclusion problem, vast literature on this topic.

Compressed sensing for parametric PDE recovery - algorithms

Formulation of the forward-backward splitting method

Can solve (2) by solving the related problem for appropriately chosen values of μ:

$$\text{minimize}_{\boldsymbol{z}\in\mathcal{V}^N} \underbrace{\|\boldsymbol{z}\|_{\mathcal{V},1} + \frac{\mu}{2}\|\boldsymbol{\Psi z} - \boldsymbol{u}\|_{\mathcal{V},2}^2}_{=:\, F_\mu(\boldsymbol{z})}. \tag{3}$$

Recall: the subdifferential of a proper function $F : \mathcal{V}^N \to (-\infty, \infty]$ at a point $\boldsymbol{x} \in \mathcal{V}^N$ is the set-valued operator

$$\partial F(\boldsymbol{x}) = \left\{\boldsymbol{v} \in \mathcal{V}^N : F(\boldsymbol{z}) \geq F(\boldsymbol{x}) + \langle \boldsymbol{v}, \boldsymbol{z} - \boldsymbol{x}\rangle \text{ for all } \boldsymbol{z} \in \mathcal{V}^N\right\}. \tag{4}$$

The elements of $\partial F(\boldsymbol{x})$ are called subgradients of F at $\boldsymbol{x}$. When the function F is convex and differentiable at $\boldsymbol{x}$, $\partial F(\boldsymbol{x}) = \{\nabla F(\boldsymbol{x})\}$, i.e. ∂F is single-valued.

Define the splitting $\partial F_\mu(\boldsymbol{z}) = T_1(\boldsymbol{z}) + T_2(\boldsymbol{z})$, where

- $T_1(\boldsymbol{z}) = \partial\|\boldsymbol{z}\|_{\mathcal{V},1} \leftarrow$ non-differentiable part of F_μ
- $T_2(\boldsymbol{z}) = \frac{\mu}{2}\nabla\|\boldsymbol{\Psi z} - \boldsymbol{u}\|_{\mathcal{V},2}^2 \leftarrow$ differentiable part of F_μ

T_1, T_2 are (sub)gradients of proper l.s.c. convex fcns., hence maximal monitone operators $\Rightarrow$ (3) is an instance of a monitone inclusion problem, vast literature on this topic.

Can solve (2) by solving the related problem for appropriately chosen values of μ:

$$\text{minimize}_{\boldsymbol{z}\in\mathcal{V}^N} \underbrace{\|\boldsymbol{z}\|_{\mathcal{V},1} + \frac{\mu}{2}\|\boldsymbol{\Psi}\boldsymbol{z}-\boldsymbol{u}\|_{\mathcal{V},2}^2}_{=:\, F_\mu(\boldsymbol{z})}. \tag{3}$$

Recall: the subdifferential of a proper function $F : \mathcal{V}^N \to (-\infty, \infty]$ at a point $\boldsymbol{x} \in \mathcal{V}^N$ is the set-valued operator

$$\partial F(\boldsymbol{x}) = \left\{\boldsymbol{v} \in \mathcal{V}^N : F(\boldsymbol{z}) \geq F(\boldsymbol{x}) + \langle \boldsymbol{v}, \boldsymbol{z}-\boldsymbol{x}\rangle \text{ for all } \boldsymbol{z} \in \mathcal{V}^N\right\}. \tag{4}$$

The elements of $\partial F(\boldsymbol{x})$ are called subgradients of F at $\boldsymbol{x}$. When the function F is convex and differentiable at $\boldsymbol{x}$, $\partial F(\boldsymbol{x}) = \{\nabla F(\boldsymbol{x})\}$, i.e. ∂F is single-valued.

Define the splitting $\partial F_\mu(\boldsymbol{z}) = T_1(\boldsymbol{z}) + T_2(\boldsymbol{z})$, where

- $T_1(\boldsymbol{z}) = \partial\|\boldsymbol{z}\|_{\mathcal{V},1} \leftarrow$ non-differentiable part of F_μ
- $T_2(\boldsymbol{z}) = \frac{\mu}{2}\nabla\|\boldsymbol{\Psi}\boldsymbol{z}-\boldsymbol{u}\|_{\mathcal{V},2}^2 \leftarrow$ differentiable part of F_μ

T_1, T_2 are (sub)gradients of proper l.s.c. convex fcns., hence maximal monitone operators $\Rightarrow$ (3) is an instance of a monitone inclusion problem, vast literature on this topic.

Theorem from Convex Analysis [Fermat's Rule]

A vector $\boldsymbol{x}$ is a minimum of the proper function F if and only if $\mathbf{0} \in \partial F(\boldsymbol{x})$.

Hence, if $X^* = \{\boldsymbol{z} \in \mathcal{V}^N : F_\mu(\boldsymbol{z}) \text{ is minimized}\}$ (the solution set), then since T_2 is single-valued, and $(I + \tau T_1)$ is invertible

$$\begin{aligned} \boldsymbol{z} \in X^* \iff \mathbf{0} \in \partial F_\mu(\boldsymbol{z}) &\iff \mathbf{0} \in (I + \tau T_1)\boldsymbol{z} - (I - \tau T_2)\boldsymbol{z} \\ &\iff (I - \tau T_2)\boldsymbol{z} \in (I + \tau T_1)\boldsymbol{z} \\ &\iff \boldsymbol{z} = (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{z}. \end{aligned}$$

Forward-backward iteration: $\boldsymbol{z}^{(k+1)} := (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{z}^{(k)}$, a fixed point alg.

In particular, since $T_2 = \nabla \phi_2(\boldsymbol{z})$ where $\phi_2 = \frac{1}{2}\|\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}\|_{\mathcal{V},2}^2$ is differentiable, we see that

$$G_\tau(\boldsymbol{z}) := (I - \tau T_2)\boldsymbol{z} = (I - \tau \nabla \phi_2)\boldsymbol{z} = \boldsymbol{z} - \tau \boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}),$$

which is a step of the gradient descent method, i.e., an (explicit) *forward* step.

T_2 is $\|\boldsymbol{\Psi}^*\boldsymbol{\Psi}\|_2$-Lipschitz $\Rightarrow$ G_τ is nonexpansive whenever $0 < \tau < 2/\lambda_{\max}(\boldsymbol{\Psi}^*\boldsymbol{\Psi})$.

Theorem from Convex Analysis [Fermat's Rule]

A vector $\boldsymbol{x}$ is a minimum of the proper function F if and only if $\mathbf{0} \in \partial F(\boldsymbol{x})$.

Hence, if $X^* = \{\boldsymbol{z} \in \mathcal{V}^N : F_\mu(\boldsymbol{z}) \text{ is minimized}\}$ (the solution set), then since T_2 is single-valued, and $(I + \tau T_1)$ is invertible

$$\begin{aligned} \boldsymbol{z} \in X^* \iff \mathbf{0} \in \partial F_\mu(\boldsymbol{z}) &\iff \mathbf{0} \in (I + \tau T_1)\boldsymbol{z} - (I - \tau T_2)\boldsymbol{z} \\ &\iff (I - \tau T_2)\boldsymbol{z} \in (I + \tau T_1)\boldsymbol{z} \\ &\iff \boldsymbol{z} = (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{z}. \end{aligned}$$

Forward-backward iteration: $\boldsymbol{z}^{(k+1)} := (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{z}^{(k)}$, a fixed point alg.

In particular, since $T_2 = \nabla \phi_2(\boldsymbol{z})$ where $\phi_2 = \frac{1}{2}\|\boldsymbol{\Psi z} - \boldsymbol{u}\|^2_{\mathcal{V},2}$ is differentiable, we see that

$$G_\tau(\boldsymbol{z}) := (I - \tau T_2)\boldsymbol{z} = (I - \tau \nabla \phi_2)\boldsymbol{z} = \boldsymbol{z} - \tau \boldsymbol{\Psi}^*(\boldsymbol{\Psi z} - \boldsymbol{u}),$$

which is a step of the gradient descent method, i.e., an (explicit) *forward* step.

T_2 is $\|\boldsymbol{\Psi}^*\boldsymbol{\Psi}\|_2$-Lipschitz $\Rightarrow$ G_τ is nonexpansive whenever $0 < \tau < 2/\lambda_{\max}(\boldsymbol{\Psi}^*\boldsymbol{\Psi})$.

Theorem from Convex Analysis [Fermat's Rule]

A vector $\boldsymbol{x}$ is a minimum of the proper function F if and only if $\mathbf{0} \in \partial F(\boldsymbol{x})$.

Hence, if $X^* = \{\boldsymbol{z} \in \mathcal{V}^N : F_\mu(\boldsymbol{z}) \text{ is minimized}\}$ (the solution set), then since T_2 is single-valued, and $(I + \tau T_1)$ is invertible

$$\begin{aligned} \boldsymbol{z} \in X^* \iff \mathbf{0} \in \partial F_\mu(\boldsymbol{z}) &\iff \mathbf{0} \in (I + \tau T_1)\boldsymbol{z} - (I - \tau T_2)\boldsymbol{z} \\ &\iff (I - \tau T_2)\boldsymbol{z} \in (I + \tau T_1)\boldsymbol{z} \\ &\iff \boldsymbol{z} = (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{z}. \end{aligned}$$

Forward-backward iteration: $\boldsymbol{z}^{(k+1)} := (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{z}^{(k)}$, a fixed point alg.

In particular, since $T_2 = \nabla\phi_2(\boldsymbol{z})$ where $\phi_2 = \frac{1}{2}\|\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}\|_{\mathcal{V},2}^2$ is differentiable, we see that

$$G_\tau(\boldsymbol{z}) := (I - \tau T_2)\boldsymbol{z} = (I - \tau\nabla\phi_2)\boldsymbol{z} = \boldsymbol{z} - \tau\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}),$$

which is a step of the gradient descent method, i.e., an (explicit) *forward* step.

T_2 is $\|\boldsymbol{\Psi}^*\boldsymbol{\Psi}\|_2$-Lipschitz $\Rightarrow$ G_τ is nonexpansive whenever $0 < \tau < 2/\lambda_{\max}(\boldsymbol{\Psi}^*\boldsymbol{\Psi})$.

Denote by $J_\tau := (I + \tau T_1)^{-1}$ the resolvent of τT_1. We can also characterize J_τ in terms of the Moreau proximity operator associated with $T_1 = \partial \|\cdot\|_{\mathcal{V},1}$:

$$\text{Prox}_\tau \boldsymbol{x} := (I + \tau \partial \|\cdot\|_{\mathcal{V},1})^{-1} \boldsymbol{x} = (I + \tau T_1)^{-1},$$

so that J_τ is a step of the proximal-point method.

A well-known result says that for $\boldsymbol{p}, \boldsymbol{x} \in \mathcal{V}^N$, $\quad \boldsymbol{p} = \text{Prox}_\tau \boldsymbol{x} \iff \boldsymbol{x} - \boldsymbol{p} \in \tau \partial \|\boldsymbol{p}\|_{\mathcal{V},1}$, so that J_τ can also be seen as an (implicit) subgradient step, i.e. a *backward* step.

This gives rise to both the *forward-backward* and *proximal-gradient* names for the composition $S_\tau(\boldsymbol{x}) := J_\tau \circ G_\tau(\boldsymbol{x}) = (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{x}$.

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

$$\|T\boldsymbol{x} - T\boldsymbol{y}\|_{\mathcal{V},2}^2 \leq \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathcal{V},2}^2 - \|(I - T)\boldsymbol{x} - (I - T)\boldsymbol{y}\|_{\mathcal{V},2}^2 \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}^N. \tag{5}$$

Another well-known result says: T_1 maximally monitone $\Rightarrow J_\tau$ is component-wise FNE.

All of these results can be found in "Convex Analysis and Monotone Operator Theory in Hilbert Spaces" by Bauschke & Combettes (2010).

Denote by $J_\tau := (I + \tau T_1)^{-1}$ the resolvent of τT_1. We can also characterize J_τ in terms of the Moreau proximity operator associated with $T_1 = \partial \| \cdot \|_{\mathcal{V},1}$:

$$\mathrm{Prox}_\tau \boldsymbol{x} := (I + \tau \partial \| \cdot \|_{\mathcal{V},1})^{-1} \boldsymbol{x} = (I + \tau T_1)^{-1},$$

so that J_τ is a step of the proximal-point method.

A well-known result says that for $\boldsymbol{p}, \boldsymbol{x} \in \mathcal{V}^N$, $\quad \boldsymbol{p} = \mathrm{Prox}_\tau \boldsymbol{x} \iff \boldsymbol{x} - \boldsymbol{p} \in \tau \partial \|\boldsymbol{p}\|_{\mathcal{V},1}$, so that J_τ can also be seen as an (implicit) subgradient step, i.e. a *backward* step.

This gives rise to both the *forward-backward* and *proximal-gradient* names for the composition $S_\tau(\boldsymbol{x}) := J_\tau \circ G_\tau(\boldsymbol{x}) = (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{x}$.

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

$$\|T\boldsymbol{x} - T\boldsymbol{y}\|_{\mathcal{V},2}^2 \le \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathcal{V},2}^2 - \|(I - T)\boldsymbol{x} - (I - T)\boldsymbol{y}\|_{\mathcal{V},2}^2 \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}^N. \tag{5}$$

Another well-known result says: T_1 maximally monitone $\Rightarrow J_\tau$ is component-wise FNE.

All of these results can be found in "Convex Analysis and Monotone Operator Theory in Hilbert Spaces" by Bauschke & Combettes (2010).

Denote by $J_\tau := (I + \tau T_1)^{-1}$ the resolvent of τT_1. We can also characterize J_τ in terms of the Moreau proximity operator associated with $T_1 = \partial \|\cdot\|_{\mathcal{V},1}$:

$$\mathsf{Prox}_\tau \boldsymbol{x} := (I + \tau \partial \|\cdot\|_{\mathcal{V},1})^{-1} \boldsymbol{x} = (I + \tau T_1)^{-1},$$

so that J_τ is a step of the proximal-point method.

A well-known result says that for $\boldsymbol{p}, \boldsymbol{x} \in \mathcal{V}^N$, $\quad \boldsymbol{p} = \mathsf{Prox}_\tau \boldsymbol{x} \iff \boldsymbol{x} - \boldsymbol{p} \in \tau \partial \|\boldsymbol{p}\|_{\mathcal{V},1}$, so that J_τ can also be seen as an (implicit) subgradient step, i.e. a *backward* step.

This gives rise to both the *forward-backward* and *proximal-gradient* names for the composition $S_\tau(\boldsymbol{x}) := J_\tau \circ G_\tau(\boldsymbol{x}) = (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{x}$.

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

$$\|T\boldsymbol{x} - T\boldsymbol{y}\|_{\mathcal{V},2}^2 \leq \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathcal{V},2}^2 - \|(I - T)\boldsymbol{x} - (I - T)\boldsymbol{y}\|_{\mathcal{V},2}^2 \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}^N. \tag{5}$$

Another well-known result says: T_1 maximally monitone $\Rightarrow J_\tau$ is component-wise FNE.

All of these results can be found in "Convex Analysis and Monotone Operator Theory in Hilbert Spaces" by Bauschke & Combettes (2010).

Denote by $J_\tau := (I + \tau T_1)^{-1}$ the resolvent of τT_1. We can also characterize J_τ in terms of the Moreau proximity operator associated with $T_1 = \partial\|\cdot\|_{\mathcal{V},1}$:

$$\mathrm{Prox}_\tau \boldsymbol{x} := (I + \tau\partial\|\cdot\|_{\mathcal{V},1})^{-1}\boldsymbol{x} = (I + \tau T_1)^{-1},$$

so that J_τ is a step of the proximal-point method.

A well-known result says that for $\boldsymbol{p}, \boldsymbol{x} \in \mathcal{V}^N$, $\quad \boldsymbol{p} = \mathrm{Prox}_\tau \boldsymbol{x} \iff \boldsymbol{x} - \boldsymbol{p} \in \tau\partial\|\boldsymbol{p}\|_{\mathcal{V},1}$, so that J_τ can also be seen as an (implicit) subgradient step, i.e. a *backward* step.

This gives rise to both the *forward-backward* and *proximal-gradient* names for the composition $S_\tau(\boldsymbol{x}) := J_\tau \circ G_\tau(\boldsymbol{x}) = (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{x}$.

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

$$\|T\boldsymbol{x} - T\boldsymbol{y}\|_{\mathcal{V},2}^2 \le \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathcal{V},2}^2 - \|(I - T)\boldsymbol{x} - (I - T)\boldsymbol{y}\|_{\mathcal{V},2}^2 \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}^N. \tag{5}$$

Another well-known result says: T_1 maximally monitone $\Rightarrow$ J_τ is component-wise FNE.

All of these results can be found in “Convex Analysis and Monotone Operator Theory in Hilbert Spaces” by Bauschke & Combettes (2010).

Denote by $J_\tau := (I + \tau T_1)^{-1}$ the resolvent of τT_1. We can also characterize J_τ in terms of the Moreau proximity operator associated with $T_1 = \partial \| \cdot \|_{\mathcal{V},1}$:

$$\mathrm{Prox}_\tau \boldsymbol{x} := (I + \tau \partial \| \cdot \|_{\mathcal{V},1})^{-1} \boldsymbol{x} = (I + \tau T_1)^{-1},$$

so that J_τ is a step of the proximal-point method.

A well-known result says that for $\boldsymbol{p}, \boldsymbol{x} \in \mathcal{V}^N$, $\quad \boldsymbol{p} = \mathrm{Prox}_\tau \boldsymbol{x} \iff \boldsymbol{x} - \boldsymbol{p} \in \tau \partial \|\boldsymbol{p}\|_{\mathcal{V},1}$, so that J_τ can also be seen as an (implicit) subgradient step, i.e. a *backward* step.

This gives rise to both the *forward-backward* and *proximal-gradient* names for the composition $S_\tau(\boldsymbol{x}) := J_\tau \circ G_\tau(\boldsymbol{x}) = (I + \tau T_1)^{-1}(I - \tau T_2)\boldsymbol{x}$.

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

$$\|T\boldsymbol{x} - T\boldsymbol{y}\|^2_{\mathcal{V},2} \leq \|\boldsymbol{x} - \boldsymbol{y}\|^2_{\mathcal{V},2} - \|(I - T)\boldsymbol{x} - (I - T)\boldsymbol{y}\|^2_{\mathcal{V},2} \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}^N. \tag{5}$$

Another well-known result says: T_1 maximally monitone $\Rightarrow$ J_τ is component-wise FNE.

All of these results can be found in "Convex Analysis and Monotone Operator Theory in Hilbert Spaces" by Bauschke & Combettes (2010).

Theorem [D., Tran, Webster '17]

Let $0 < \tau < 2/\lambda_{\max}(\boldsymbol{\Psi}^*\boldsymbol{\Psi})$. Then the iterations $\boldsymbol{x}^{(k+1)} := J_\tau \circ G_\tau(\boldsymbol{x}^{(k)})$ converge strongly to an element $\boldsymbol{x}^* \in X^*$ from any $\boldsymbol{x}^{(0)} \in \mathcal{V}^N$.

Sketch: Opial's Theorem $\Rightarrow$ weak convergence [Daubechies, et al '04], [Combettes '04].

Finite convergence is easily obtained for $j \in \Lambda_0$ s.t. $\|(\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_{\mathcal{V}} < 1$.
We focus on the set $j \in \Lambda_0$ s.t. $\|(\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_{\mathcal{V}} = 1$, i.e., the complement.

J_τ is component-wise given by $(I - \mathcal{P}_\tau)$, where $\mathcal{P}_\tau$ is metric projection onto $B_{\mathcal{V}}(\mathbf{0}, \tau)$
$\Rightarrow J_\tau(\boldsymbol{y})$, $\mathcal{P}_\tau(\boldsymbol{y})$, and $\boldsymbol{y}$ are colinear with the origin for any $\boldsymbol{y} \in \mathcal{V}^N$.

J_τ component-wise FNE and $(I - J_\tau) = (I - I + \mathcal{P}_\tau) = \mathcal{P}_\tau$, imply

$$\begin{aligned}\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 &\le \|G_\tau(\boldsymbol{x}_j^{(k)}) - G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2 - \|(I - J_\tau)G_\tau(\boldsymbol{x}_j^{(k)}) - (I - J_\tau)G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2 \\ &\le \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \underbrace{\|\mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^{(k)}) - \mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2}_{=:c_j^{(k)}},\end{aligned}$$

from the nonexpansiveness of G_τ, $\forall\ j \in \Lambda_0$ and $k \in \mathbb{N}$.

Theorem [D., Tran, Webster '17]

Let $0 < \tau < 2/\lambda_{\max}(\boldsymbol{\Psi}^*\boldsymbol{\Psi})$. Then the iterations $\boldsymbol{x}^{(k+1)} := J_\tau \circ G_\tau(\boldsymbol{x}^{(k)})$ converge strongly to an element $\boldsymbol{x}^* \in X^*$ from any $\boldsymbol{x}^{(0)} \in \mathcal{V}^N$.

Sketch: Opial's Theorem $\Rightarrow$ weak convergence [Daubechies, et al '04], [Combettes '04].

Finite convergence is easily obtained for $j \in \Lambda_0$ s.t. $\|(\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_{\mathcal{V}} < 1$.
We focus on the set $j \in \Lambda_0$ s.t. $\|(\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_{\mathcal{V}} = 1$, i.e., the complement.

J_τ is component-wise given by $(I - \mathcal{P}_\tau)$, where $\mathcal{P}_\tau$ is metric projection onto $B_{\mathcal{V}}(\mathbf{0}, \tau)$
$\Rightarrow J_\tau(\boldsymbol{y})$, $\mathcal{P}_\tau(\boldsymbol{y})$, and $\boldsymbol{y}$ are colinear with the origin for any $\boldsymbol{y} \in \mathcal{V}^N$.

J_τ component-wise FNE and $(I - J_\tau) = (I - I + \mathcal{P}_\tau) = \mathcal{P}_\tau$, imply

$$\begin{aligned}\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 &\le \|G_\tau(\boldsymbol{x}_j^{(k)}) - G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2 - \|(I - J_\tau)G_\tau(\boldsymbol{x}_j^{(k)}) - (I - J_\tau)G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2 \\ &\le \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \underbrace{\|\mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^{(k)}) - \mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2}_{=: c_j^{(k)}},\end{aligned}$$

from the nonexpansiveness of G_τ, $\forall\, j \in \Lambda_0$ and $k \in \mathbb{N}$.

Theorem [D., Tran, Webster '17]

Let $0 < \tau < 2/\lambda_{\max}(\boldsymbol{\Psi}^*\boldsymbol{\Psi})$. Then the iterations $\boldsymbol{x}^{(k+1)} := J_\tau \circ G_\tau(\boldsymbol{x}^{(k)})$ converge strongly to an element $\boldsymbol{x}^* \in X^*$ from any $\boldsymbol{x}^{(0)} \in \mathcal{V}^N$.

Sketch: Opial's Theorem $\Rightarrow$ weak convergence [Daubechies, et al '04], [Combettes '04].

Finite convergence is easily obtained for $j \in \Lambda_0$ s.t. $\|(\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_{\mathcal{V}} < 1$.
We focus on the set $j \in \Lambda_0$ s.t. $\|(\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_{\mathcal{V}} = 1$, i.e., the complement.

J_τ is component-wise given by $(I - \mathcal{P}_\tau)$, where $\mathcal{P}_\tau$ is metric projection onto $B_{\mathcal{V}}(\mathbf{0}, \tau)$
$\Rightarrow J_\tau(\boldsymbol{y})$, $\mathcal{P}_\tau(\boldsymbol{y})$, and $\boldsymbol{y}$ are colinear with the origin for any $\boldsymbol{y} \in \mathcal{V}^N$.

J_τ component-wise FNE and $(I - J_\tau) = (I - I + \mathcal{P}_\tau) = \mathcal{P}_\tau$, imply

$$\begin{aligned}\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 &\le \|G_\tau(\boldsymbol{x}_j^{(k)}) - G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2 - \|(I - J_\tau)G_\tau(\boldsymbol{x}_j^{(k)}) - (I - J_\tau)G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2 \\ &\le \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \underbrace{\|\mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^{(k)}) - \mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2}_{=:c_j^{(k)}},\end{aligned}$$

from the nonexpansiveness of G_τ, $\forall\ j \in \Lambda_0$ and $k \in \mathbb{N}$.

Theorem [D., Tran, Webster '17]

Let $0 < \tau < 2/\lambda_{\max}(\mathbf{\Psi}^*\mathbf{\Psi})$. Then the iterations $\boldsymbol{x}^{(k+1)} := J_\tau \circ G_\tau(\boldsymbol{x}^{(k)})$ converge strongly to an element $\boldsymbol{x}^* \in X^*$ from any $\boldsymbol{x}^{(0)} \in \mathcal{V}^N$.

Sketch: Opial's Theorem $\Rightarrow$ weak convergence [Daubechies, et al '04], [Combettes '04].

Finite convergence is easily obtained for $j \in \Lambda_0$ s.t. $\|(\mathbf{\Psi}^*(\mathbf{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_\mathcal{V} < 1$.
We focus on the set $j \in \Lambda_0$ s.t. $\|(\mathbf{\Psi}^*(\mathbf{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_\mathcal{V} = 1$, i.e., the complement.

J_τ is component-wise given by $(I - \mathcal{P}_\tau)$, where $\mathcal{P}_\tau$ is metric projection onto $B_\mathcal{V}(\mathbf{0}, \tau)$
$\Rightarrow J_\tau(\boldsymbol{y})$, $\mathcal{P}_\tau(\boldsymbol{y})$, and $\boldsymbol{y}$ are colinear with the origin for any $\boldsymbol{y} \in \mathcal{V}^N$.

J_τ component-wise FNE and $(I - J_\tau) = (I - I + \mathcal{P}_\tau) = \mathcal{P}_\tau$, imply

$$\begin{aligned}\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_\mathcal{V}^2 &\le \|G_\tau(\boldsymbol{x}_j^{(k)}) - G_\tau(\boldsymbol{x}_j^*)\|_\mathcal{V}^2 - \|(I - J_\tau)G_\tau(\boldsymbol{x}_j^{(k)}) - (I - J_\tau)G_\tau(\boldsymbol{x}_j^*)\|_\mathcal{V}^2 \\ &\le \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_\mathcal{V}^2 - \underbrace{\|\mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^{(k)}) - \mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^*)\|_\mathcal{V}^2}_{=:c_j^{(k)}},\end{aligned}$$

from the nonexpansiveness of G_τ, $\forall\ j \in \Lambda_0$ and $k \in \mathbb{N}$.

Theorem [D., Tran, Webster '17]

Let $0 < \tau < 2/\lambda_{\max}(\boldsymbol{\Psi}^*\boldsymbol{\Psi})$. Then the iterations $\boldsymbol{x}^{(k+1)} := J_\tau \circ G_\tau(\boldsymbol{x}^{(k)})$ converge strongly to an element $\boldsymbol{x}^* \in X^*$ from any $\boldsymbol{x}^{(0)} \in \mathcal{V}^N$.

Sketch: Opial's Theorem $\Rightarrow$ weak convergence [Daubechies, et al '04], [Combettes '04].

Finite convergence is easily obtained for $j \in \Lambda_0$ s.t. $\|(\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_{\mathcal{V}} < 1$.
We focus on the set $j \in \Lambda_0$ s.t. $\|(\boldsymbol{\Psi}^*(\boldsymbol{\Psi}\boldsymbol{x}^* - \boldsymbol{u}))_j\|_{\mathcal{V}} = 1$, i.e., the complement.

J_τ is component-wise given by $(I - \mathcal{P}_\tau)$, where $\mathcal{P}_\tau$ is metric projection onto $B_{\mathcal{V}}(\mathbf{0}, \tau)$
$\Rightarrow J_\tau(\boldsymbol{y})$, $\mathcal{P}_\tau(\boldsymbol{y})$, and $\boldsymbol{y}$ are colinear with the origin for any $\boldsymbol{y} \in \mathcal{V}^N$.

J_τ component-wise FNE and $(I - J_\tau) = (I - I + \mathcal{P}_\tau) = \mathcal{P}_\tau$, imply

$$
\begin{aligned}
\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 &\leq \|G_\tau(\boldsymbol{x}_j^{(k)}) - G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2 - \|(I - J_\tau)G_\tau(\boldsymbol{x}_j^{(k)}) - (I - J_\tau)G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2 \\
&\leq \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \underbrace{\|\mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^{(k)}) - \mathcal{P}_\tau \circ G_\tau(\boldsymbol{x}_j^*)\|_{\mathcal{V}}^2}_{=: c_j^{(k)}},
\end{aligned}
$$

from the nonexpansiveness of G_τ, $\forall\ j \in \Lambda_0$ and $k \in \mathbb{N}$.

Compressed sensing for parametric PDE recovery - algorithms

Iterate: $\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 \leq \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - c_j^{(k)} \leq \underbrace{\cdots}_{k\text{-times}} \leq \|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \sum_{\ell=0}^{k} c_j^{(\ell)}$

Rearrange: $\sum_{\ell=0}^{k} c_j^{(\ell)} \leq \underbrace{\|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2}_{\text{independent of } k} \implies c_j^{(k)} \to 0$ as $k \to \infty$.

Collinearity & $c_j^{(k)} \to 0 \Rightarrow$ angle $\theta_j^{(k)}$ between the iterates $\boldsymbol{x}_j^{(k)}$ and $\boldsymbol{x}_j^*$ is converging to 0.

Weak convergence $\implies$ $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \cos\theta_j^{(k)} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ (in cases $\boldsymbol{x}_j^* = \mathbf{0}$ and $\boldsymbol{x}_j^* \neq \mathbf{0}$).

Angular convergence $\implies$ $\cos\theta_j^{(k)} \to 1$.

Weak convergence & angular convergence imply $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ as $k \to \infty$ so that

$$\|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 = \|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}}^2 + \|\boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - 2\langle \boldsymbol{x}_j^{(k)}, \boldsymbol{x}_j^* \rangle_{\mathcal{V}} \to 0 \quad \text{as } k \to \infty.$$

Hence $\boldsymbol{x}_j^{(k)} \to \boldsymbol{x}_j^*$ as $k \to \infty$ for each $j \in \Lambda_0$.

Compressed sensing for parametric PDE recovery - algorithms

Iterate: $\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 \leq \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - c_j^{(k)} \leq \underbrace{\cdots}_{k\text{-times}} \leq \|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \sum_{\ell=0}^k c_j^{(\ell)}$

Rearrange: $\sum_{\ell=0}^k c_j^{(\ell)} \leq \underbrace{\|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2}_{\text{independent of } k} \implies c_j^{(k)} \to 0$ as $k \to \infty$.

Collinearity & $c_j^{(k)} \to 0 \Rightarrow$ angle $\theta_j^{(k)}$ between the iterates $\boldsymbol{x}_j^{(k)}$ and $\boldsymbol{x}_j^*$ is converging to 0.

Weak convergence $\implies$ $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \cos\theta_j^{(k)} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ (in cases $\boldsymbol{x}_j^* = \mathbf{0}$ and $\boldsymbol{x}_j^* \neq \mathbf{0}$).

Angular convergence $\implies$ $\cos\theta_j^{(k)} \to 1$.

Weak convergence & angular convergence imply $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ as $k \to \infty$ so that

$$\|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 = \|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}}^2 + \|\boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - 2\langle \boldsymbol{x}_j^{(k)}, \boldsymbol{x}_j^*\rangle_{\mathcal{V}} \to 0 \quad \text{as } k \to \infty.$$

Hence $\boldsymbol{x}_j^{(k)} \to \boldsymbol{x}_j^*$ as $k \to \infty$ for each $j \in \Lambda_0$.

Compressed sensing for parametric PDE recovery - algorithms

Iterate: $\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 \leq \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - c_j^{(k)} \leq \underbrace{\cdots}_{k\text{-times}} \leq \|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \sum_{\ell=0}^{k} c_j^{(\ell)}$

Rearrange: $\sum_{\ell=0}^{k} c_j^{(\ell)} \leq \underbrace{\|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2}_{\text{independent of } k} \implies c_j^{(k)} \to 0$ as $k \to \infty$.

Collinearity & $c_j^{(k)} \to 0 \Rightarrow$ angle $\theta_j^{(k)}$ between the iterates $\boldsymbol{x}_j^{(k)}$ and $\boldsymbol{x}_j^*$ is converging to 0.

Weak convergence $\implies$ $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \cos\theta_j^{(k)} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ (in cases $\boldsymbol{x}_j^* = \mathbf{0}$ and $\boldsymbol{x}_j^* \neq \mathbf{0}$).

Angular convergence $\implies$ $\cos\theta_j^{(k)} \to 1$.

Weak convergence & angular convergence imply $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ as $k \to \infty$ so that

$$\|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 = \|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}}^2 + \|\boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - 2\langle \boldsymbol{x}_j^{(k)}, \boldsymbol{x}_j^* \rangle_{\mathcal{V}} \to 0 \quad \text{as } k \to \infty.$$

Hence $\boldsymbol{x}_j^{(k)} \to \boldsymbol{x}_j^*$ as $k \to \infty$ for each $j \in \Lambda_0$.

Compressed sensing for parametric PDE recovery - algorithms

Iterate: $\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 \leq \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - c_j^{(k)} \leq \underbrace{\cdots}_{k\text{-times}} \leq \|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \sum_{\ell=0}^{k} c_j^{(\ell)}$

Rearrange: $\sum_{\ell=0}^{k} c_j^{(\ell)} \leq \underbrace{\|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2}_{\text{independent of } k} \implies c_j^{(k)} \to 0$ as $k \to \infty$.

Collinearity & $c_j^{(k)} \to 0 \Rightarrow$ angle $\theta_j^{(k)}$ between the iterates $\boldsymbol{x}_j^{(k)}$ and $\boldsymbol{x}_j^*$ is converging to 0.

Weak convergence $\implies$ $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \cos\theta_j^{(k)} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ (in cases $\boldsymbol{x}_j^* = \mathbf{0}$ and $\boldsymbol{x}_j^* \neq \mathbf{0}$).

Angular convergence $\implies$ $\cos\theta_j^{(k)} \to 1$.

Weak convergence & angular convergence imply $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ as $k \to \infty$ so that

$$\|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 = \|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}}^2 + \|\boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - 2\langle \boldsymbol{x}_j^{(k)}, \boldsymbol{x}_j^* \rangle_{\mathcal{V}} \to 0 \quad \text{as } k \to \infty.$$

Hence $\boldsymbol{x}_j^{(k)} \to \boldsymbol{x}_j^*$ as $k \to \infty$ for each $j \in \Lambda_0$.

Compressed sensing for parametric PDE recovery - algorithms

Iterate: $\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 \leq \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - c_j^{(k)} \leq \underbrace{\cdots}_{k\text{-times}} \leq \|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \sum_{\ell=0}^k c_j^{(\ell)}$

Rearrange: $\sum_{\ell=0}^k c_j^{(\ell)} \leq \underbrace{\|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2}_{\text{independent of } k} \implies c_j^{(k)} \to 0$ as $k \to \infty$.

Collinearity & $c_j^{(k)} \to 0 \Rightarrow$ angle $\theta_j^{(k)}$ between the iterates $\boldsymbol{x}_j^{(k)}$ and $\boldsymbol{x}_j^*$ is converging to 0.

Weak convergence $\implies \|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \cos\theta_j^{(k)} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ (in cases $\boldsymbol{x}_j^* = \mathbf{0}$ and $\boldsymbol{x}_j^* \neq \mathbf{0}$).

Angular convergence $\implies \cos\theta_j^{(k)} \to 1$.

Weak convergence & angular convergence imply $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ as $k \to \infty$ so that

$$\|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 = \|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}}^2 + \|\boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - 2\langle \boldsymbol{x}_j^{(k)}, \boldsymbol{x}_j^*\rangle_{\mathcal{V}} \to 0 \quad \text{as } k \to \infty.$$

Hence $\boldsymbol{x}_j^{(k)} \to \boldsymbol{x}_j^*$ as $k \to \infty$ for each $j \in \Lambda_0$.

Compressed sensing for parametric PDE recovery - algorithms

Iterate: $\|\boldsymbol{x}_j^{(k+1)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 \leq \|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - c_j^{(k)} \leq \underbrace{\cdots}_{k\text{-times}} \leq \|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - \sum_{\ell=0}^k c_j^{(\ell)}$

Rearrange: $\sum_{\ell=0}^k c_j^{(\ell)} \leq \underbrace{\|\boldsymbol{x}_j^{(0)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2}_{\text{independent of } k} \implies c_j^{(k)} \to 0$ as $k \to \infty$.

Collinearity & $c_j^{(k)} \to 0 \Rightarrow$ angle $\theta_j^{(k)}$ between the iterates $\boldsymbol{x}_j^{(k)}$ and $\boldsymbol{x}_j^*$ is converging to 0.

Weak convergence $\implies$ $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \cos\theta_j^{(k)} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ (in cases $\boldsymbol{x}_j^* = \mathbf{0}$ and $\boldsymbol{x}_j^* \neq \mathbf{0}$).

Angular convergence $\implies$ $\cos\theta_j^{(k)} \to 1$.

Weak convergence & angular convergence imply $\|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}} \to \|\boldsymbol{x}_j^*\|_{\mathcal{V}}$ as $k \to \infty$ so that

$$\|\boldsymbol{x}_j^{(k)} - \boldsymbol{x}_j^*\|_{\mathcal{V}}^2 = \|\boldsymbol{x}_j^{(k)}\|_{\mathcal{V}}^2 + \|\boldsymbol{x}_j^*\|_{\mathcal{V}}^2 - 2\langle \boldsymbol{x}_j^{(k)}, \boldsymbol{x}_j^* \rangle_{\mathcal{V}} \to 0 \quad \text{as } k \to \infty.$$

Hence $\boldsymbol{x}_j^{(k)} \to \boldsymbol{x}_j^*$ as $k \to \infty$ for each $j \in \Lambda_0$.

The Bregman distance w.r.t. $J(\cdot) := \|\cdot\|_{\mathcal{V},1}$ between the points $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{V}^N$ is defined as

$$D_J^{\boldsymbol{p}}(\boldsymbol{u},\boldsymbol{v}) = J(\boldsymbol{u}) - J(\boldsymbol{v}) - \langle \boldsymbol{p}, \boldsymbol{u}-\boldsymbol{v}\rangle_{\mathcal{V},2},$$

where $\boldsymbol{p} \in \partial J(\boldsymbol{v})$ is an element of the subdifferential of J at the point $\boldsymbol{v}$.

The Bregman iterative scheme can be written for $\mathcal{V}^N$:

$$\boldsymbol{u}^{(0)} \leftarrow \boldsymbol{0},\ \boldsymbol{z}^{(0)} \leftarrow \boldsymbol{0}, \tag{6}$$

$$\text{For } k = 0, 1, \dots \text{ do} \tag{7}$$

$$\boldsymbol{u}^{(k+1)} \leftarrow \boldsymbol{u} + (\boldsymbol{u}^{(k)} - \boldsymbol{\Psi}\boldsymbol{z}^{(k)}), \tag{8}$$

$$\boldsymbol{z}^{(k+1)} \leftarrow \operatorname*{arg\,min}_{\boldsymbol{z}\in\mathcal{V}^N} J(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}^{(k+1)}\|_{\mathcal{V},2}^2. \tag{9}$$

We apply the forward-backward splitting to find the intermediate solutions $\boldsymbol{z}^{(k)}$ in (9).

Adding residual back in step (8) gives nice error cancellation, allowing intermediate solns. (9) to be solved less accurately without affecting overall accuracy (error forgetting).

The Bregman distance w.r.t. $J(\cdot) := \|\cdot\|_{\mathcal{V},1}$ between the points $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{V}^N$ is defined as

$$D_J^{\boldsymbol{p}}(\boldsymbol{u}, \boldsymbol{v}) = J(\boldsymbol{u}) - J(\boldsymbol{v}) - \langle \boldsymbol{p}, \boldsymbol{u} - \boldsymbol{v} \rangle_{\mathcal{V},2},$$

where $\boldsymbol{p} \in \partial J(\boldsymbol{v})$ is an element of the subdifferential of J at the point $\boldsymbol{v}$.

The Bregman iterative scheme can be written for $\mathcal{V}^N$:

$$\boldsymbol{u}^{(0)} \leftarrow \mathbf{0}, \ \boldsymbol{z}^{(0)} \leftarrow \mathbf{0}, \tag{6}$$

$$\text{For } k = 0, 1, \dots \text{ do} \tag{7}$$

$$\boldsymbol{u}^{(k+1)} \leftarrow \boldsymbol{u} + (\boldsymbol{u}^{(k)} - \boldsymbol{\Psi} \boldsymbol{z}^{(k)}), \tag{8}$$

$$\boldsymbol{z}^{(k+1)} \leftarrow \underset{\boldsymbol{z} \in \mathcal{V}^N}{\arg\min}\, J(\boldsymbol{z}) + \frac{1}{2} \|\boldsymbol{\Psi} \boldsymbol{z} - \boldsymbol{u}^{(k+1)}\|_{\mathcal{V},2}^2. \tag{9}$$

We apply the forward-backward splitting to find the intermediate solutions $z^{(k)}$ in (9).

Adding residual back in step (8) gives nice error cancellation, allowing intermediate solns. (9) to be solved less accurately without affecting overall accuracy (error forgetting).

The Bregman distance w.r.t. $J(\cdot) := \|\cdot\|_{\mathcal{V},1}$ between the points $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{V}^N$ is defined as

$$D_J^{\boldsymbol{p}}(\boldsymbol{u},\boldsymbol{v}) = J(\boldsymbol{u}) - J(\boldsymbol{v}) - \langle \boldsymbol{p}, \boldsymbol{u}-\boldsymbol{v}\rangle_{\mathcal{V},2},$$

where $\boldsymbol{p} \in \partial J(\boldsymbol{v})$ is an element of the subdifferential of J at the point $\boldsymbol{v}$.

The Bregman iterative scheme can be written for $\mathcal{V}^N$:

$$\boldsymbol{u}^{(0)} \leftarrow \mathbf{0},\ \boldsymbol{z}^{(0)} \leftarrow \mathbf{0}, \tag{6}$$

$$\text{For } k = 0, 1, \dots \text{ do} \tag{7}$$

$$\boldsymbol{u}^{(k+1)} \leftarrow \boldsymbol{u} + (\boldsymbol{u}^{(k)} - \boldsymbol{\Psi}\boldsymbol{z}^{(k)}), \tag{8}$$

$$\boldsymbol{z}^{(k+1)} \leftarrow \underset{\boldsymbol{z}\in\mathcal{V}^N}{\arg\min}\, J(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}^{(k+1)}\|_{\mathcal{V},2}^2. \tag{9}$$

We apply the forward-backward splitting to find the intermediate solutions $\boldsymbol{z}^{(k)}$ in (9).

Adding residual back in step (8) gives nice error cancellation, allowing intermediate solns. (9) to be solved less accurately without affecting overall accuracy (error forgetting).

The Bregman distance w.r.t. $J(\cdot) := \|\cdot\|_{\mathcal{V},1}$ between the points $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{V}^N$ is defined as

$$D_J^{\boldsymbol{p}}(\boldsymbol{u},\boldsymbol{v}) = J(\boldsymbol{u}) - J(\boldsymbol{v}) - \langle \boldsymbol{p}, \boldsymbol{u}-\boldsymbol{v}\rangle_{\mathcal{V},2},$$

where $\boldsymbol{p} \in \partial J(\boldsymbol{v})$ is an element of the subdifferential of J at the point $\boldsymbol{v}$.

The Bregman iterative scheme can be written for $\mathcal{V}^N$:

$$\boldsymbol{u}^{(0)} \leftarrow \boldsymbol{0},\ \boldsymbol{z}^{(0)} \leftarrow \boldsymbol{0}, \tag{6}$$

$$\text{For } k = 0, 1, \dots \text{ do} \tag{7}$$

$$\boldsymbol{u}^{(k+1)} \leftarrow \boldsymbol{u} + (\boldsymbol{u}^{(k)} - \boldsymbol{\Psi}\boldsymbol{z}^{(k)}), \tag{8}$$

$$\boldsymbol{z}^{(k+1)} \leftarrow \operatorname*{arg\,min}_{\boldsymbol{z}\in\mathcal{V}^N} J(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{\Psi}\boldsymbol{z} - \boldsymbol{u}^{(k+1)}\|_{\mathcal{V},2}^2. \tag{9}$$

We apply the forward-backward splitting to find the intermediate solutions $\boldsymbol{z}^{(k)}$ in (9).

Adding residual back in step (8) gives nice error cancellation, allowing intermediate solns. (9) to be solved less accurately without affecting overall accuracy (error forgetting).

Forward-backward splitting:

- Finite convergence to the complement of the support of an element of X^*
- Strong convergence of the whole sequence to a fixed point
- Linear convergence, under minimum eigenvalue assumption, with an explicit bound of the constant

Bregman iterations:

- Monotonic decrease in the residual $\frac{1}{2}\|\boldsymbol{\Psi z}^{(k)} - \boldsymbol{u}\|^2_{\mathcal{V},2}$
- Monotonic decrease in the Bregman distance between iterates $D_J^{\boldsymbol{p}^{(k)}}(\boldsymbol{z}^{(k+1)}, \boldsymbol{z}^{(k)})$
- Existence of weak-* convergent subsequences in the Banach space $(\mathcal{V}^N, \|\cdot\|_{\mathcal{V},1})$ whose limit satisfy $\boldsymbol{\Psi z} = \boldsymbol{u}$

Main challenges: Infinite dimensions implies lack of compactness, some geometric arguments that work in $\mathbb{R}^N$ don't hold in $\mathcal{V}^N$.

Compressed sensing for parametric PDE recovery - algorithms

Summary of main convergence results shown for forward-backward splitting and Bregman iterations

Forward-backward splitting:

- Finite convergence to the complement of the support of an element of X^*
- Strong convergence of the whole sequence to a fixed point
- Linear convergence, under minimum eigenvalue assumption, with an explicit bound of the constant

Bregman iterations:

- Monotonic decrease in the residual $\frac{1}{2}\|\boldsymbol{\Psi}\boldsymbol{z}^{(k)} - \boldsymbol{u}\|_{\mathcal{V},2}^2$
- Monotonic decrease in the Bregman distance between iterates $D_J^{\boldsymbol{p}^{(k)}}(\boldsymbol{z}^{(k+1)}, \boldsymbol{z}^{(k)})$
- Existence of weak-* convergent subsequences in the Banach space $(\mathcal{V}^N, \|\cdot\|_{\mathcal{V},1})$ whose limit satisfy $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$

Main challenges: Infinite dimensions implies lack of compactness, some geometric arguments that work in $\mathbb{R}^N$ don't hold in $\mathcal{V}^N$.

Forward-backward splitting:

- Finite convergence to the complement of the support of an element of X^*
- Strong convergence of the whole sequence to a fixed point
- Linear convergence, under minimum eigenvalue assumption, with an explicit bound of the constant

Bregman iterations:

- Monotonic decrease in the residual $\frac{1}{2}\|\boldsymbol{\Psi}\boldsymbol{z}^{(k)} - \boldsymbol{u}\|^2_{\mathcal{V},2}$
- Monotonic decrease in the Bregman distance between iterates $D_J^{\boldsymbol{p}^{(k)}}(\boldsymbol{z}^{(k+1)}, \boldsymbol{z}^{(k)})$
- Existence of weak-* convergent subsequences in the Banach space $(\mathcal{V}^N, \|\cdot\|_{\mathcal{V},1})$ whose limit satisfy $\boldsymbol{\Psi}\boldsymbol{z} = \boldsymbol{u}$

Main challenges: Infinite dimensions implies lack of compactness, some geometric arguments that work in $\mathbb{R}^N$ don't hold in $\mathcal{V}^N$.

Stochastic elliptic problem on $D = [0,1]^2$:

$$\begin{cases} -\nabla \cdot (a(x,\boldsymbol{y})\nabla u(x,\boldsymbol{y})) &= f(x) \quad \text{in } \Gamma \times D, \\ u(x,\boldsymbol{y}) &= 0 \quad \text{on } \Gamma \times \partial D. \end{cases} \tag{10}$$

Specifically, we focus on the case that $y_j \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$, and $a(x,\boldsymbol{y})$ is given by:

$$a(x,\boldsymbol{y}) = a_{\min} + y_1 \left(\frac{\sqrt{\pi}L}{2}\right)^{1/2} + \sum_{j=2}^{d} \zeta_j \varphi_j(x) y_j,$$

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

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

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

Stochastic elliptic problem on $D = [0,1]^2$:

$$\begin{cases} -\nabla \cdot (a(x,\boldsymbol{y})\nabla u(x,\boldsymbol{y})) &= f(x) \quad \text{in } \Gamma \times D, \\ u(x,\boldsymbol{y}) &= 0 \quad \text{on } \Gamma \times \partial D. \end{cases} \tag{10}$$

Specifically, we focus on the case that $y_j \sim \mathcal{U}(-\sqrt{3},\sqrt{3})$, and $a(x,\boldsymbol{y})$ is given by:

$$a(x,\boldsymbol{y}) = a_{\min} + y_1 \left(\frac{\sqrt{\pi}L}{2}\right)^{1/2} + \sum_{j=2}^{d} \zeta_j \varphi_j(x) y_j,$$

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

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

which is the KL expansion associated with the squared exponential covariance kernel, L_c 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 - numerical examples

Fixed quasi-uniform triangulation of $D = [0,1]^2$ having 206 points ($h \approx 1/16$)

Compressed sensing setup:

- Fixed total degree subspace Λ_0 with $N = \#\Lambda_0$ large, increasing the number of samples m following $\lceil kN/8 \rceil$ for $k = 1, 2, \ldots, 7$
- Compute $\eta_{\Lambda_0} := \|\boldsymbol{u}_{\Lambda_0^c}\|_{\mathcal{V},2} = \|\boldsymbol{\Psi}\boldsymbol{z}_{\Lambda_0}^{\mathsf{SG}} - \boldsymbol{u}\|_{\mathcal{V},2}$ using stochastic Galerkin, and set $1.2 \cdot \eta_{\Lambda_0}$ as tolerance for the BPDN problem (choosing μ appropriately)
- Average the results over 24 trials

Compared against:

- "Decoupled approach", solve the same problem with compressed sensing pointwise.
- Stochastic Galerkin, with total degree of order $p = 2, 3$.
- Stochastic collocation, with Clenshaw-Curtis points with doubling, level $L = 2, 3$.
- Monte Carlo method, with uniform sampling.

Compressed sensing for parametric PDE recovery - numerical examples

Fixed quasi-uniform triangulation of $D = [0,1]^2$ having 206 points ($h \approx 1/16$)

Compressed sensing setup:

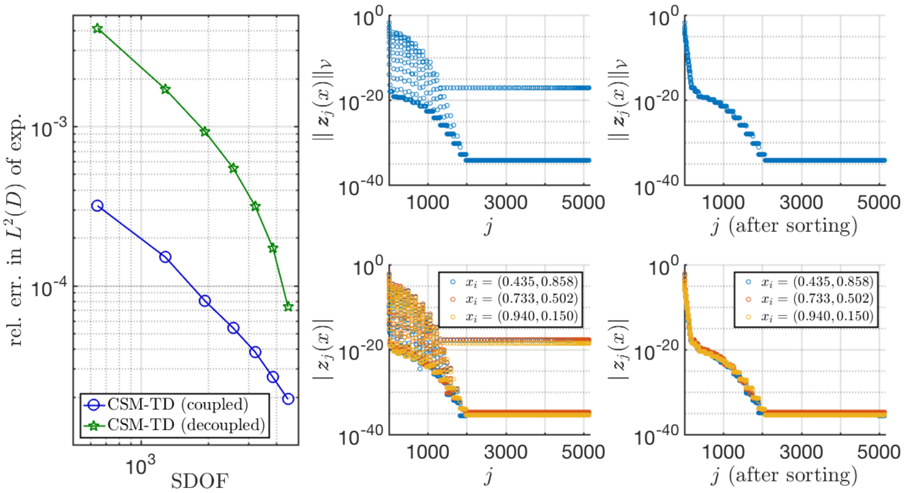
- Fixed total degree subspace Λ_0 with $N = \#\Lambda_0$ large, increasing the number of samples m following $\lceil kN/8 \rceil$ for $k = 1, 2, \ldots, 7$
- Compute $\eta_{\Lambda_0} := \|\boldsymbol{u}_{\Lambda_0^c}\|_{\mathcal{V},2} = \|\boldsymbol{\Psi} \boldsymbol{z}_{\Lambda_0}^{\mathsf{SG}} - \boldsymbol{u}\|_{\mathcal{V},2}$ using stochastic Galerkin, and set $1.2 \cdot \eta_{\Lambda_0}$ as tolerance for the BPDN problem (choosing μ appropriately)
- Average the results over 24 trials

Compared against:

- "Decoupled approach", solve the same problem with compressed sensing pointwise.
- Stochastic Galerkin, with total degree of order $p = 2, 3$.
- Stochastic collocation, with Clenshaw-Curtis points with doubling, level $L = 2, 3$.
- Monte Carlo method, with uniform sampling.

Compressed sensing for parametric PDE recovery - numerical examples

Comparison of Hilbert-valued and functional recovery strategies.

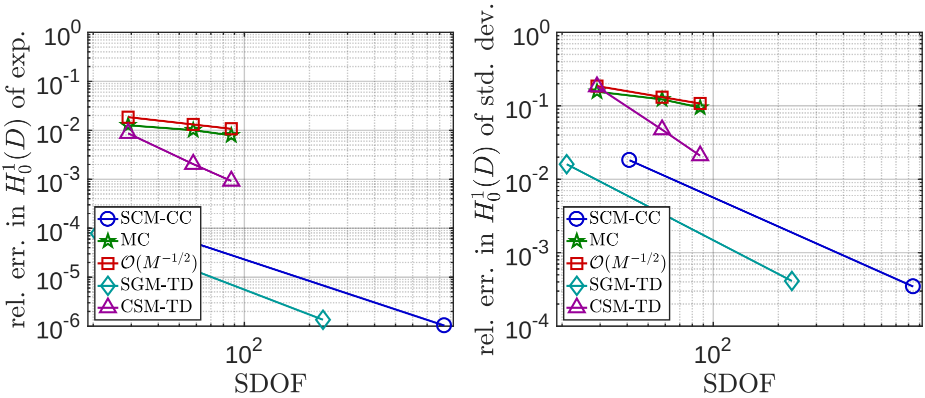


- $a(x, \boldsymbol{y})$ is the high-dimensional affine coefficient ($d = 100$, $L_c = 1/4$)
- Λ_0 the total degree space of order $p = 2$ with $N = \#\Lambda_0 = 5151$
- For the SGM, SDOF is N, for all other methods, SDOF is m, the number of samples

Compressed sensing for parametric PDE recovery - numerical examples

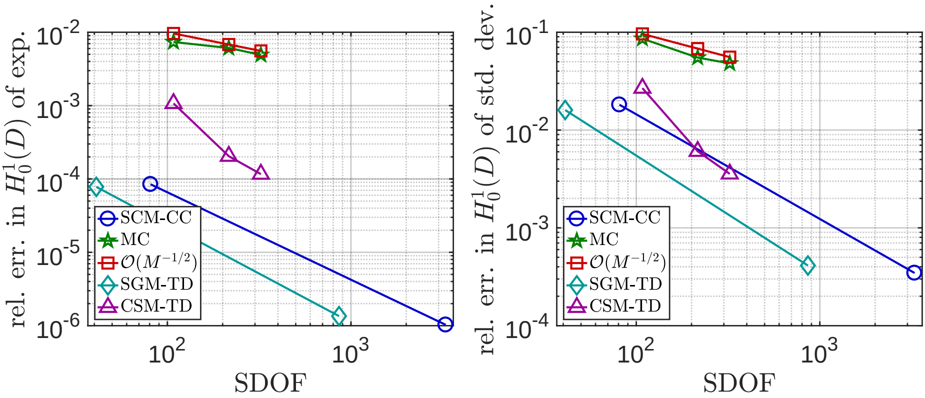
Comparison with stochastic Galerkin, stochastic collocation, and Monte Carlo methods

Here we compare against isotropic methods only to highlight the performance of all methods when no knowledge of the coefficient decay is known.



- $a(x, \boldsymbol{y})$ is the high-dimensional affine coefficient ($d = 20$, $L_c = 1/4$)
- Λ_0 the total degree space of order $p = 2$ with $N = \#\Lambda_0 = 231$
- For the SGM, SDOF is N, for all other methods, SDOF is m, the number of samples

Here we compare against isotropic methods only to highlight the performance of all methods when no knowledge of the coefficient decay is known.

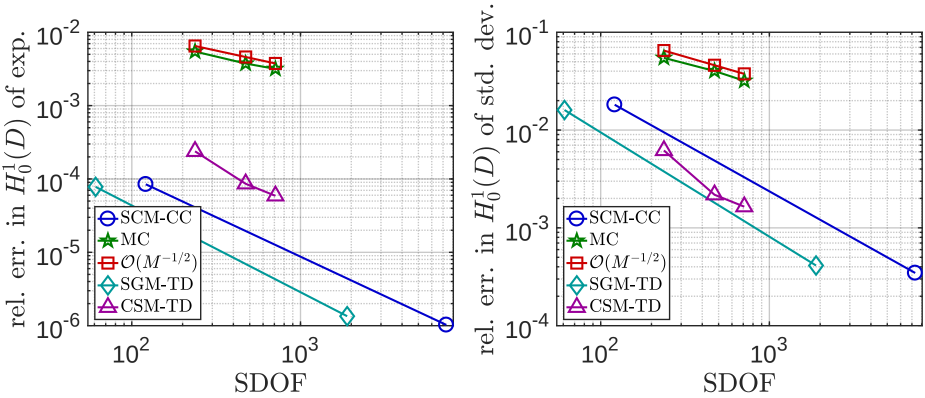


- $a(x, \boldsymbol{y})$ is the high-dimensional affine coefficient ($d = 40$, $L_c = 1/4$)
- Λ_0 the total degree space of order $p = 2$ with $N = \#\Lambda_0 = 861$
- For the SGM, SDOF is N, for all other methods, SDOF is m, the number of samples

Compressed sensing for parametric PDE recovery - numerical examples

Comparison with stochastic Galerkin, stochastic collocation, and Monte Carlo methods

Here we compare against isotropic methods only to highlight the performance of all methods when no knowledge of the coefficient decay is known.

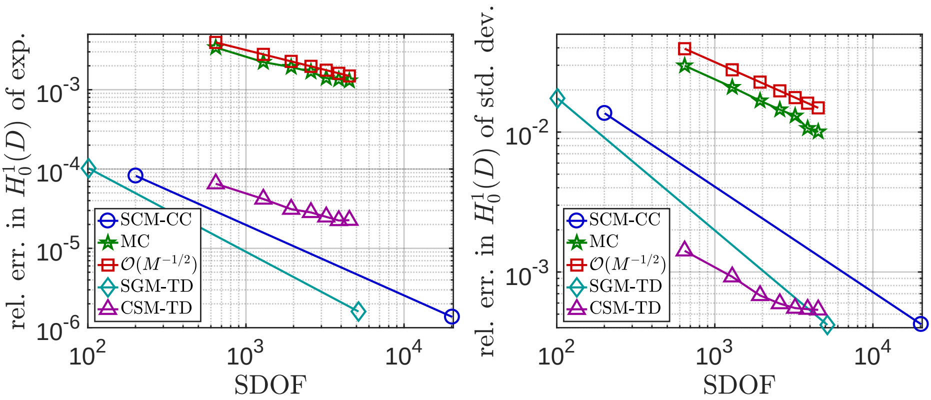


- $a(x, \boldsymbol{y})$ is the high-dimensional affine coefficient ($d = 60$, $L_c = 1/4$)
- Λ_0 the total degree space of order $p = 2$ with $N = \#\Lambda_0 = 1891$
- For the SGM, SDOF is N, for all other methods, SDOF is m, the number of samples

Compressed sensing for parametric PDE recovery - numerical examples

Comparison with stochastic Galerkin, stochastic collocation, and Monte Carlo methods

Here we compare against isotropic methods only to highlight the performance of all methods when no knowledge of the coefficient decay is known.



- $a(x, \boldsymbol{y})$ is the high-dimensional affine coefficient ($d = 100$, $L_c = 1/4$)
- Λ_0 the total degree space of order $p = 2$ with $N = \#\Lambda_0 = 5151$
- For the SGM, SDOF is N, for all other methods, SDOF is m, the number of samples

Concluding remarks

- Generalization of compressed sensing theory and algorithms to the Hilbert-valued case and connection to parameterized PDEs
    - Sparse approximation in the Hilbert-valued setting has been around for a **long time**
    - This approach puts approx. error estimates in terms of the best s-term w.r.t. Λ_0
- 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 the Bregman iterations
- Need more numerical experiments
    - Nonlinear parameterized PDEs
    - Linear vs. nonlinear stochastic parameterization

Some of the references discussed in this talk

- H. H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. In: CMS Books in Mathematics. Springer. (2010).
- A. Chkifa, N. Dexter, H. Tran, C. G. Webster, *Polynomial approximation via compressed sensing of high-dimensional functions on lower sets*. In: Mathematics of Computation. American Mathematical Society. (2017).
- P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*. In: Optimization. Taylor & Francis. (2004).
- I. Daubechies, M. Defrise, C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*. In: Communications on Pure and Applied Mathematics. Wiley. (2004).
- E. Hale, W. Yin, Y. Zhang, *Fixed-point continuation for l1-minimization: Methodology and convergence*, In: SIAM Journal on Optimization. (2008).
- H. Tran, C. G. Webster, G. Zhang, *Analysis of quasi-optimal polynomial approximations for parameterized PDEs with deterministic and stochastic coefficients*. In: Numerische Mathematik, (2016)
- W. Yin, S. Osher, D. Goldfarb, J. Darbon, *Bregman Iterative Algorithms for ℓ_1-Minimization with Applications to Compressed Sensing*, In: SIAM Journal on Imaging Sciences. (2008)
- W. Yin, S. Osher, *Error forgetting of Bregman iteration*, In: Journal of Scientific Computing. Springer. (2013)