Compressed sensing and application to uncertainty quantification

SIAM UQ 2016 – Minitutorial

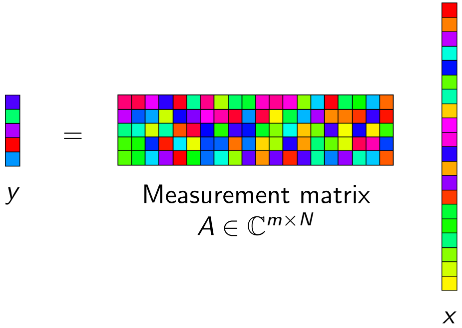
Ben Adcock
Department of Mathematics
Simon Fraser University

Slides are on my website: www.benadcock.ca/presentations

Outline

Underdetermined systems of linear equations

Let $x \in \mathbb{C}^N$ be an unknown vector. We consider $m \ll N$ measurements



Goal: Recover x from the underdetermined system of equations $Az = y$.

Compressed sensing: the highlights

Under appropriate conditions on x and A we can recover $x \in \mathbb{C}^N$ from the measurements $y = Ax \in \mathbb{C}^m$ in a stable and robust manner. Moreover, this can be done using efficient numerical algorithms.

- Condition on x: low-dimensionality $s \ll N$.
- Condition on A: E.g. Null Space Property, Restricted Isometry Property, incoherence,...
- Condition on m: It is possible to find matrices A such that only $m \approx C \cdot s \cdot \log(N)$ measurements suffice.
- Algorithms: convex optimization (ℓ^1 minimization), greedy methods, thresholding methods, message passing algorithms,...

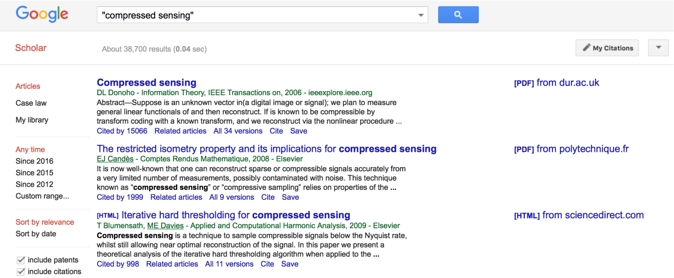
Compressed sensing: the highlights

Under appropriate conditions on x and A we can recover $x \in \mathbb{C}^N$ from the measurements $y = Ax \in \mathbb{C}^m$ in a stable and robust manner. Moreover, this can be done using efficient numerical algorithms.

- Condition on x: low-dimensionality $s \ll N$.
- Condition on A: E.g. Null Space Property, Restricted Isometry Property, incoherence,...
- Condition on m: It is possible to find matrices A such that only $m \approx C \cdot s \cdot \log(N)$ measurements suffice.
- Algorithms: convex optimization (ℓ^1 minimization), greedy methods, thresholding methods, message passing algorithms,...

A little history

- Initial developments ($\approx$ 2005): Candès, Romberg & Tao, Donoho
- Since then, the subject of thousands of papers, dozens of survey articles, and one textbook (Foucart & Rauhut, Birkhauser, 2013).
- a.k.a. compressive sensing, compressed sampling, compressive sampling



Origins: geophysics (1970s/80s), statistics, signal processing (1980s/90s), wavelets and nonlinear approximation (1980s/90s).

Why do we care?

In many applications, a key limitation is the amount of data available.

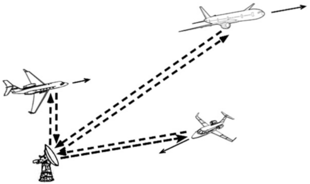
Examples:

- MRI: more measurements $\approx$ longer scan time.
- X-Ray CT: more measurements $\approx$ higher radiation doses.
- Microscopy: more measurements deteriorate/destroy the object.
- Seismic/infrarad/etc imaging: more measurements $\approx$ higher costs.
- Sensor networks: more measurements $\approx$ more power.

Why do we care?

But in many applications, the unknown x has a low-dimensional structure:

Examples: Radar, astronomical images, certain medical images,...



Radar

Astronomical images

Why do we care?

But in many applications, the unknown x has a low-dimensional structure:

Examples: Typical images are defined by edges $\Rightarrow$ x has a sparse representation in wavelets.

Image x

Wavelet coefficients

Compressed sensing for uncertainty quantification?

Big Picture:

1. In UQ one often faces the situation of limited measurements.
2. The solution/quantity of interest/etc typically lives in a high (perhaps infinite) dimensional space.
3. But there is often low-dimensional structure.

So compressed sensing is a good fit...?

Main example: solving parametric PDEs

Consider the parametrized PDE system

$$\mathcal{L}(u; x, z) = 0,$$

where $x \in \mathbb{R}^p$, $p = 1, 2, 3, 4$, is the physical variable and $z \in \mathbb{R}^d$ is a variable of parameters.

Goal: Compute the map $z \mapsto u(\cdot, z)$ or some functional $f : z \mapsto \mathcal{Q}u(\cdot, z)$.

Nonintrusive methods: Recover f from samples $\{f(z_i)\}_{i=1}^m$.

Generalized polynomial chaos: Approximate f using a basis of multivariate orthonormal polynomials $f(z) \approx \sum_{i \in I} x_i \phi_i(z)$.

Main example: solving parametric PDEs

Consider the parametrized PDE system

$$\mathcal{L}(u; x, z) = 0,$$

where $x \in \mathbb{R}^p$, $p = 1, 2, 3, 4$, is the physical variable and $z \in \mathbb{R}^d$ is a variable of parameters.

Goal: Compute the map $z \mapsto u(\cdot, z)$ or some functional $f : z \mapsto \mathcal{Q}u(\cdot, z)$.

Nonintrusive methods: Recover f from samples $\{f(z_i)\}_{i=1}^m$.

Generalized polynomial chaos: Approximate f using a basis of multivariate orthonormal polynomials $f(z) \approx \sum_{i \in I} x_i \phi_i(z)$.

Compressed sensing for parametric PDEs

Big Picture:

1. In UQ one often faces the situation of limited measurements.
   -
2. The solution/quantity of interest/etc etc typically lives in a high (perhaps infinite) dimensional space.
   -
3. But there is often low-dimensional structure.
   -
   -

Compressed sensing for parametric PDEs

Big Picture:

1. In UQ one often faces the situation of limited measurements.
   - Each sample $f(z_i)$ is expensive to acquire.
2. The solution/quantity of interest/etc etc typically lives in a high (perhaps infinite) dimensional space.
   -
3. But there is often low-dimensional structure.
   -
   -

Compressed sensing for parametric PDEs

Big Picture:

1. In UQ one often faces the situation of limited measurements.
   - Each sample $f(z_i)$ is expensive to acquire.
2. The solution/quantity of interest/etc etc typically lives in a high (perhaps infinite) dimensional space.
   - We want to include as many parameters as possible in the model, i.e. $d \gg 1$, and as many polynomials, i.e. $|I| \gg 1$.
3. But there is often low-dimensional structure.
   -
   -

Compressed sensing for parametric PDEs

Big Picture:

1. In UQ one often faces the situation of limited measurements.
   - Each sample $f(z_i)$ is expensive to acquire.
2. The solution/quantity of interest/etc etc typically lives in a high (perhaps infinite) dimensional space.
   - We want to include as many parameters as possible in the model, i.e. $d \gg 1$, and as many polynomials, i.e. $|I| \gg 1$.
3. But there is often low-dimensional structure.
   - The expansion coefficients $\{x_i\}_{i \in I}$ are often sparse.
   - See Albert Cohen's tutorial on Wednesday, for example.

Questions for the remainder of the talk

1. Given a polynomial (or nonpolynomial) basis, how should we sample?

2. What is a good low-dimensional model for such problems, and how do we properly exploit it?

3. What is the resulting sample complexity (= number of measurements m), and how does it depend on dimension d and sparsity s?

4. To what extent can the curse of dimensionality be broken?

5. The standard CS setup is finite-dimensional. How do we handle infinite-dimensionality of functions?

Existing work

Theory and techniques:

- Rauhut & Ward (2011, 2012), Yan, Guo & Xiu (2012), Tang & Iaccarino (2014), Hampton & Doostan (2014, 2015), Xu & Zhou (2014), Rauhut & Ward (2014), Adcock (2015), Chkifa, Dexter, Tran & Webster (2016), Guo, Narayan, Zhou & Chen (2016), Jakeman, Narayan & Zhou (2016) and others.

Applications:

- Doostan & Owhadi (2011), Mathelin & Gallivan (2012), Yang & Karniadakis (2013), Lei, Yang, Zheng, Lin & Baker (2014), Peng, Hampton & Doostan (2014), Rauhut & Schwab (2015), Yang, Lei, Baker & Lin (2015), Jakeman, Eldred & Sargsyan (2015), Karagiannis, Konomi & Lin (2015), Guo, Narayan, Xiu & Zhou (2015) and many others.

Also, many talks this week.

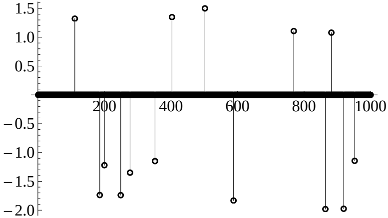
Outline

Sparsity

The standard low-dimensional model in CS:

Definition (Sparsity)

A vector $x \in \mathbb{C}^N$ is s-sparse if it has at most s nonzero entries.



Note: We may know s, but we do not know the locations on the nonzero coefficients of x.

ℓ^0 minimization

Let $x \in \mathbb{C}^N$ be s-sparse, $A \in \mathbb{C}^{m \times N}$ and $y = Ax$. To recover x from y, we can look for the sparsest solution:

$$\min_{z \in \mathbb{C}^N} \|z\|_0 \text{ subject to } Az = y, \qquad (\star)$$

where $\|z\|_0 = |\{j : z_j \neq 0\}|$ is the ℓ^0 'norm'.

Note: x is the unique s-sparse solution of $Az = y \iff x$ is the unique minimizer of $(\star)$.

Problem: $(\star)$ is NP-hard to solve in general.

ℓ^0 minimization

Let $x \in \mathbb{C}^N$ be s-sparse, $A \in \mathbb{C}^{m \times N}$ and $y = Ax$. To recover x from y, we can look for the sparsest solution:

$$\min_{z \in \mathbb{C}^N} \|z\|_0 \text{ subject to } Az = y, \qquad (\star)$$

where $\|z\|_0 = |\{j : z_j \neq 0\}|$ is the ℓ^0 'norm'.

Note: x is the unique s-sparse solution of $Az = y \iff x$ is the unique minimizer of $(\star)$.

Problem: $(\star)$ is NP-hard to solve in general.

ℓ^1 minimization

To obtain a computationally tractable problem, we make a convex relaxation. We replace

$$\min_{z \in \mathbb{C}^N} \|z\|_0 \text{ subject to } Az = y,$$

by

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \text{ subject to } Az = y, \qquad (\star)$$

where $\|z\|_1 = \sum_{i=1}^N |z_i|$ is the l^1-norm.

Many algorithms exist for solving the convex problem $(\star)$. E.g.

- homotopy methods, LARS, primal dual algorithms, pareto curve methods, iteratively reweighted least squares, splitting methods (e.g. split Bregman, ADMM),...

Note: Alternatives to ℓ^1: greedy methods, threshholding methods,...

The Restricted Isometry Property

A popular tool for the analysis of CS:

Definition

The restricted isometry constant δ_s of a matrix $A \in \mathbb{C}^{m \times N}$ is the smallest number such that

$$(1-\delta_s)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1+\delta_s)\|z\|_2^2, \qquad \forall\ s\text{-sparse } z.$$

We say A satisfies the Restricted Isometry Property (RIP) of order s with constant δ_s if $\delta_s \in (0,1)$.

Candès & Tao (2005,2006), Cohen, Dahmen & DeVore (2009)

Intuition

Suppose that the support

$$\Delta = \mathrm{supp}(x) = \{j : x_j \neq 0\}, \qquad |\Delta| = s,$$

were known. Let $A_\Delta = \{a_{ij} : i = 1, \ldots, m,\ j \in \Delta\} \in \mathbb{C}^{m \times s}$ be formed by the restriction of the columns of A to those with indices in Δ. If

$$\|A_\Delta^* A_\Delta - I\|_{2 \to 2} \in (0, 1),$$

then we can recover x stably and robustly via least-squares fitting:

$$x = \operatorname*{argmin}_{\mathrm{supp}(z) \subseteq \Delta} \|A_\Delta z - y\|_2 = A_\Delta^\dagger y.$$

However, note that

$$\delta_s = \max\left\{\|A_\Delta^* A_\Delta - I\|_{2 \to 2} : \Delta \subseteq \{1, \ldots, N\},\ |\Delta| \leq s\right\}.$$

$\Rightarrow$ the RIP ensures stable and robust recovery of this oracle for any Δ.

Intuition

Suppose that the support

$$\Delta = \mathrm{supp}(x) = \{j : x_j \neq 0\}, \qquad |\Delta| = s,$$

were known. Let $A_\Delta = \{a_{ij} : i = 1, \ldots, m,\ j \in \Delta\} \in \mathbb{C}^{m \times s}$ be formed by the restriction of the columns of A to those with indices in Δ. If

$$\|A_\Delta^* A_\Delta - I\|_{2\to2} \in (0,1),$$

then we can recover x stably and robustly via least-squares fitting:

$$x = \operatorname*{argmin}_{\mathrm{supp}(z)\subseteq\Delta} \|A_\Delta z - y\|_2 = A_\Delta^\dagger y.$$

However, note that

$$\delta_s = \max\left\{\|A_\Delta^* A_\Delta - I\|_{2\to2} : \Delta \subseteq \{1, \ldots, N\},\ |\Delta| \leq s\right\}.$$

$\Rightarrow$ the RIP ensures stable and robust recovery of this oracle for any Δ.

Stable and robust recovery with the RIP

Theorem

Suppose that the matrix $A \in \mathbb{C}^{m\times N}$ *satisfies the RIP of order* $2s$ *with constant* $\delta_{2s} < 1/\sqrt{2}$*. Then for any* $x \in \mathbb{C}^N$ *and* $y \in \mathbb{C}^m$ *with* $\|Ax - y\|_2 \leq \eta$*, any solution* $\hat{x}$ *of*

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \text{ subject to } \|Az - y\|_2 \leq \eta,$$

satisfies

$$\|x - \hat{x}\|_2 \lesssim \sigma_s(x)/\sqrt{s} + \eta, \qquad \|x - \hat{x}\|_1 \lesssim \sigma_s(x) + \sqrt{s}\eta,$$

where $\sigma_s(x) = \min\{\|x - z\|_1 : z \text{ is } s\text{-sparse}\}$.

Stability: x is recovered exactly up to an error proportional to its best s-term approximation $\sigma_s(x)$.

Robustness: For noisy measurements $y = Ax + e$ with noise bound $\|e\|_2 \leq \eta$, x is recovered up to an error proportional to η.

Candès (2008), Cohen, Dahmen & DeVore (2009), Cai & Zhang (2013, 2014) and others

Matrices that satisfy the RIP

Deterministic constructions of RIP matrices with m scaling linearly with s have proved elusive.

Key idea: Use randomness
Candès, Romberg & Tao (2005), Donoho (2005)

Early examples:

- Gaussian random matrices (great to analyze, but impractical).
- Subsampled Fourier transforms (harder to analyze, but more practical).

A general construction

Let F be a distribution of random vectors in $\mathbb{C}^N$.

Isometry condition: $\mathbb{E}(aa^*) = I$, $a \sim F$.

Construction of A: Draw $a_1, \ldots, a_m$ independently from F and define

$$A = \begin{bmatrix} a_1^* \\ \vdots \\ a_m^* \end{bmatrix} \in \mathbb{C}^{m \times N}.$$

Coherence: Let $\mu(F)$ be the smallest number such that

$$\|a\|_\infty^2 \leq \mu(F),$$

almost surely for $a \sim F$. Note that $\mu(F) \geq 1$.

Candès & Plan (2012), Gross & Kueng (2013), Adcock & Hansen (2013), Chun & Adcock (2016)

A general construction

Theorem

Let $0 < \delta, \epsilon < 1$. *If*

$$m \gtrsim \delta^{-2} \cdot \mu(F) \cdot s \cdot \left(\log^3(2s) \log(2N) + \log(\epsilon^{-1}) \right),$$

then the matrix $\frac{1}{\sqrt{m}} A$ *satisfies the RIP of order* s *with constant* $\delta_s \leq \delta$.

If F is incoherent, i.e. $\mu(F) \approx 1$, then $m \gtrsim s \times$ log factors.

- Similar to the *bounded orthonormal systems* approach, Rauhut (2010)
- Proof is based on arguments of Candès & Tao (2006), Rudelson & Vershynin (2008), Rauhut (2010)
- Variations/enhancements: Andersson & Stromberg (2014), Haviv & Regev (2016), Chkifa, Dexter, Tran & Webster (2016)

Outline

Example: one-dimensional Chebyshev polynomials

Consider $D = [-1, 1]$, $\nu(z) = \frac{1}{\pi\sqrt{1-z^2}}$ and the orthonormal basis of Chebyshev polynomials:

$$\phi_0(z) = 1, \quad \phi_i(z) = \sqrt{2}\cos\left(i\cos^{-1}(z)\right), \qquad i = 1, 2, \ldots .$$

Let

$$f(z) = \sum_{i=0}^{N-1} x_i \phi_i(z),$$

be a polynomial of degree N with coefficients $x = \{x_i\}_{i=0}^{N-1} \in \mathbb{C}^N$.

Measurements: Draw $z_1, \ldots, z_m$ independently from ν and set

$$y = \{f(z_i)\}_{i=1}^{m} = Ax, \qquad A = \{\phi_j(z_i)\}_{i=1,j=0}^{m,N-1} .$$

Recovery of one-dimensional Chebyshev polynomials

Theorem

Let $0 < \epsilon < 1$, $1 \leq s \leq N$, $\eta \geq 0$ *and*

$$m \gtrsim s \cdot \left(\log^3(2s)\log(2N) + \log(\epsilon^{-1})\right).$$

Draw $z_1, \ldots, z_m$ *independently from* ν *and form* $A \in \mathbb{C}^{m\times N}$*. Let* $f(z) = \sum_{i=0}^{N} x_i \phi_i(z) \in \mathbb{P}_{N-1}$ *be arbitrary and set* $y = \{f(z_i)\}_{i=1}^{m} + e$*, where* $\|e\|_2 \leq \eta$*. Then for any minimizer* $\hat{x}$ *of*

$$\min_{v\in\mathbb{C}^N} \|v\|_1 \text{ subject to } \|Av - y\|_2 \leq \eta,$$

we have

$$\|x-\hat{x}\|_2 \lesssim \sigma_s(x)/\sqrt{s} + \eta/\sqrt{m}, \qquad \|x-\hat{x}\|_1 \lesssim \sigma_s(x) + \eta\sqrt{s/m}.$$

- Applies only to finite polynomials f (see later)
- Related to *sparse recovery of trigonometric polynomials*, Rauhut (2007)

Proof

Let F be the family

$$a = a(z) = [\phi_0(z), \phi_1(z), \ldots, \phi_{N-1}(z)]^\top, \qquad z \sim \nu.$$

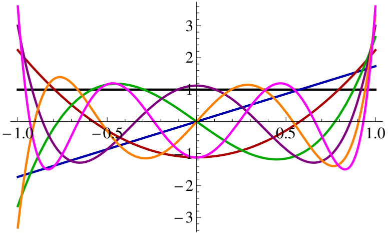
Then

- $\mathbb{E}\left(aa^*\right)_{ij} = \mathbb{E}\left(\phi_i \overline{\phi_j}\right) = \delta_{i,j} \implies F$ is isotropic.
- $\|a\|_\infty^2 \leq 2 \implies \mu(F) \leq 2$.

Hence $\frac{1}{\sqrt{m}}A$ satisfies the RIP with $m \gtrsim s \times \log$ factors.

What about Legendre polynomials?

Let $D = [-1, 1]$, $\nu(z) = \frac{1}{2}$ and $\phi_i(z)$, $i = 0, 1, 2, \ldots$, be the orthonormal Legendre polynomial basis.



Problem: $\|\phi_i\|_{L^\infty} = |\phi_i(1)| = \sqrt{2i + 1}$. Hence the coherence

$$\mu(F) = 2N + 1.$$

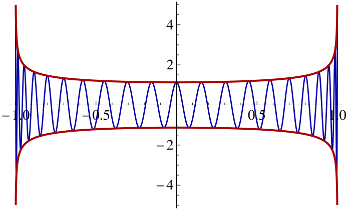
This gives the sample complexity $m \gtrsim N \cdot s \times \log$ factors, which is useless.

Rauhut & Ward (2012)

The preconditioning trick

Legendre polynomials possess an enveloping property:

$$|\phi_i(z)|(1-z^2)^{1/4} < 2/\sqrt{\pi}.$$



The preconditioned system

$$\Phi_i(z) = \sqrt{\pi/2}(1-z^2)^{1/4}\phi_i(z),$$

is orthonormal with respect to the measure $\nu(z) = \frac{1}{\pi\sqrt{1-z^2}}$ and satisfies

$$\|\Phi_i\|^2_{L^\infty} \leq 2.$$

Hence, if we draw samples $z_1, \dots, z_m$ from this measure, the resulting sample complexity is $m \gtrsim s \times$ log factors.

- Note: a similar approach can be used for any Jacobi polynomials.

Rauhut & Ward (2012)

Outline

Notation

Let

- $D \subseteq \mathbb{R}^d$ be a domain,
- $\rho(z)$ be a probability measure on D,
- $\{z_i\}_{i=1}^m \subseteq D$ be drawn independently from ρ,
- $\{\phi_i\}_{i \in I}$ be an orthonormal system in $L^2(D, \mathrm{d}\rho) \cap L^\infty(D)$, where I is a countable index set,
- $I_K \subseteq I$ be a finite index set and $N = |I_K|$.

Suppose that f is a finite polynomial in the ϕ_i:

$$f = \sum_{i \in I_K} x_i \phi_i, \qquad x_i = \int_D f(z) \overline{\phi_i(z)} \, \mathrm{d}\rho(z),$$

where $x = \{x_i\}_{i \in I_K}$ are the coefficients of f in the system $\{\phi_i\}_{i \in I}$.

Main example: tensor products of polynomials

Let ν be a density function on $(-1, 1)$ and $\{\psi_i\}_{i=0}^{\infty}$ be orthonormal polynomials with respect to ν. Set

- $D = (-1, 1)^d$,
- $\rho(z) = \prod_{j=1}^{d} \nu(z_j) \, \mathrm{d}z$,
- $I = \mathbb{N}_0^d$,
- $\phi_i(z) = \prod_{j=1}^{d} \psi_{i_j}(z_j)$ for $i = (i_1, \ldots, i_d) \in I$.

Note: unbounded domains – see later.

Choices for the truncated index set I_K

Various options, including:

1. Tensor product: $I_K^{TP} = \{ i = (i_1, \ldots, i_d) : 0 \leq i_j \leq K, \ j = 1, \ldots, d \}$.
   - $|I_K^{TP}| = (K+1)^d$ – often too large in practice.
2. Total degree: $I_K^{TD} = \left\{ i = (i_1, \ldots, i_d) : \sum_{j=1}^d i_j \leq K \right\}$.
   - $|I_K^{TD}| = \begin{pmatrix} K+d \\ d \end{pmatrix}$ – more manageable.
3. Hyperbolic cross: $I_K^{HC} = \left\{ i = (i_1, \ldots, i_d) : \prod_{j=1}^d (i_j + 1) \leq K + 1 \right\}$.
   - $|I_K^{HC}| \leq CK \min \left\{ \log(K)^{d-1}, d^{\log(K)} \right\}$ – even more manageable.

Considerations:

- Computational cost: $|I_K| = N$ is the number of matrix columns
- Smaller index sets I_K may miss important features.

Recovery of tensor Chebyshev polynomials

1D basis: $\psi_0(z) = 1$, $\psi_i(z) = \sqrt{2}\cos\left(i \cos^{-1}(z)\right)$ otherwise.

Observe that

$$\|\phi_i\|_\infty^2 = \prod_{j=1}^{d} \|\psi_{i_j}\|_{L^\infty}^2 = 2^{|i|_0},$$

where $|i|_0 = |\{j : i_j \neq 0\}|$. Hence

$$\mu(F) = 2^q, \quad q = \max\{|i|_0 : i \in I_K\}.$$

Recovery guarantees: Consider the total degree space I_K^{TD}.

Low to moderate dimensions	$d < K$	$m \gtrsim 2^d \cdot s \cdot L$
High dimensions	$d \geq K$	$m \gtrsim 2^K \cdot s \cdot L$

Here $L =$ log factors.

Recovery of tensor Legendre polynomials

Case 1: We sample from the uniform measure. Since

$$\|\phi_i\|_\infty^2 \leq \prod_{j=1}^{d}(2i_j + 1),$$

we get

Low to moderate dimensions	$d < K$	$m \gtrsim (2K/d + 1)^d \cdot s \cdot L$
High dimensions	$d \geq K$	$m \gtrsim 3^K \cdot s \cdot L$

Case 2: We sample from the Chebyshev measure and precondition. Since

$$\|\phi_i\|_\infty^2 \leq (\pi/2)^d (4/\pi)^{\|i\|_0},$$

we get

Low to moderate dimensions	$d < K$	$m \gtrsim 2^d \cdot s \cdot L$
High dimensions	$d \geq K$	$m \gtrsim (\pi/2)^d (4/\pi)^K \cdot s \cdot L$

Yan, Guo & Xiu (2012)

Recovery of tensor Legendre polynomials

Case 1: We sample from the uniform measure. Since

$$\|\phi_i\|_\infty^2 \leq \prod_{j=1}^{d}(2i_j+1),$$

we get

Low to moderate dimensions	$d < K$	$m \gtrsim (2K/d+1)^d \cdot s \cdot L$
High dimensions	$d \geq K$	$m \gtrsim 3^K \cdot s \cdot L$

Case 2: We sample from the Chebyshev measure and precondition. Since

$$\|\phi_i\|_\infty^2 \leq (\pi/2)^d (4/\pi)^{|i|_0},$$

we get

Low to moderate dimensions	$d < K$	$m \gtrsim 2^d \cdot s \cdot L$
High dimensions	$d \geq K$	$m \gtrsim (\pi/2)^d (4/\pi)^K \cdot s \cdot L$

Yan, Guo & Xiu (2012)

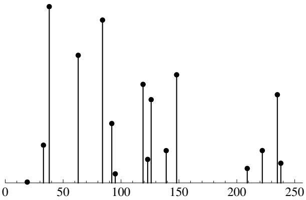
Problem

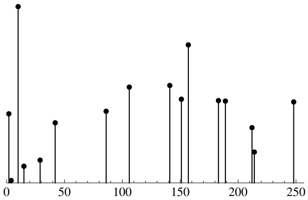
In all cases, there is exponential blow-up of the sample complexity with either dimension d or degree K.

Outline

Sparsity?

Sparsity permits the s non-zero coefficients to have arbitrary locations:





Bad news: recovering coefficients corresponding to high polynomial degrees requires more samples, due to the growth of $\|\phi_i\|_{L^\infty}$ with i.

Good news: For smooth functions, the nonzero polynomial coefficients typically occur at lower indices.

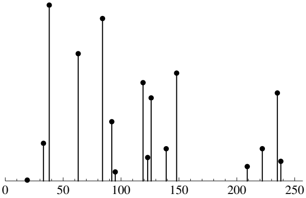
- Sparsity alone is too crude to capture this behaviour.

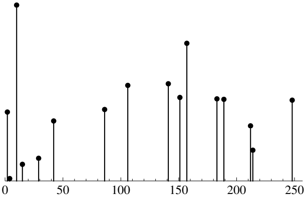
Solution:

1. Penalize high-degree coefficients in the regularization term.
2. Seek recovery guarantees for a fixed support set, not all supports.

Sparsity?

Sparsity permits the s non-zero coefficients to have arbitrary locations:





Bad news: recovering coefficients corresponding to high polynomial degrees requires more samples, due to the growth of $\|\phi_i\|_{L^\infty}$ with i.

Good news: For smooth functions, the nonzero polynomial coefficients typically occur at lower indices.

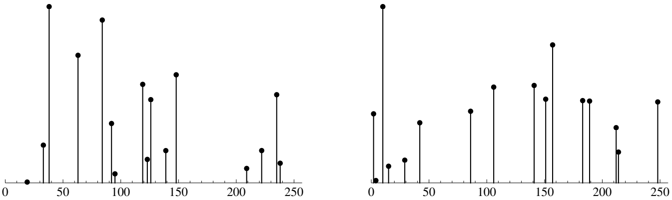
- Sparsity alone is too crude to capture this behaviour.

Solution:

1. Penalize high-degree coefficients in the regularization term.
2. Seek recovery guarantees for a fixed support set, not all supports.

Sparsity?

Sparsity permits the s non-zero coefficients to have arbitrary locations:



Bad news: recovering coefficients corresponding to high polynomial degrees requires more samples, due to the growth of $\|\phi_i\|_{L^\infty}$ with i.

Good news: For smooth functions, the nonzero polynomial coefficients typically occur at lower indices.

- Sparsity alone is too crude to capture this behaviour.

Solution:

1. Penalize high-degree coefficients in the regularization term.
2. Seek recovery guarantees for a fixed support set, not all supports.

Weighted ℓ^1 minimization

We solve

$$\min_{v \in \mathbb{C}^N} \|v\|_{1,w} \text{ subject to } \|Av - y\|_2 \leq \eta,$$

where

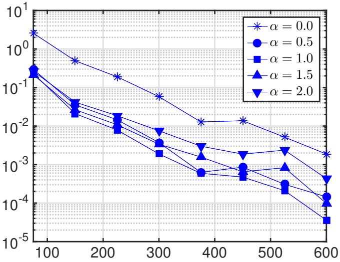
- $A = \{\phi_j(z_i) : i = 1, \ldots, m, j \in I_K\} \in \mathbb{C}^{m \times N}$, $N = |I_K|$,
- $y = \{f(z_i)\}_{i=1}^m + e$, $\|e\|_2 \leq \eta$,
- $w = \{w_i\}_{i \in I_K}$ are positive weights,
- $\|\cdot\|_{1,w}$ is the weighted ℓ^1 norm: $\|v\|_{1,w} = \sum_{i \in I_K} w_i |v_i|$.

It has been observed empirically that weights often give superior performance over unweighted ℓ^1 minimization.

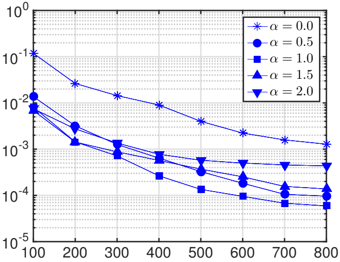
- See: Yang & Karniadakis (2013), Peng, Hampton & Doostan (2014), Rauhut & Ward (2015), Adcock (2015).

Weighting strategies

Example: error versus m for (unpreconditioned) Legendre polynomials.



$d = 2,\ f(z) = \exp(2z_1)\cos(3z_2)$
$w_i = (1 + i_1 + \ldots + i_d)^\alpha$



$d = 10,\ f(z) = \mathrm{e}^{-\frac{z_1 + \cdots + z_{10}}{20}}$
$w_i = (i_1 \cdots i_d)^\alpha$

Questions

How does the recovery error depend on the weights? Is there an optimal choice of weights? Does this overcome the curse of dimensionality?

Towards a theorem

We now focus on recovering a fixed support set $\Delta \subseteq I_K$.

- In particular, we must avoid the RIP.
- Follow ideas of nonuniform recovery in CS (e.g. RIPless CS).

Notation:

- Let $P_\Delta x$ be such that $(P_\Delta x)_j = x_j$, $j \in \Delta$ and 0 otherwise.
- Define the weighted cardinality of a set Δ as $|\Delta|_w = \sum_{i \in \Delta} w_i^2$.

Goal: Prove error estimates in terms of $\|x - P_\Delta x\|$ with sample complexities depending on Δ, not just $s = |\Delta|$.

Nonuniform recovery in CS: Candès & Plan (2012), Adcock & Hansen (2013), Boyer, Bigot & Weiss (2015), Chun & Adcock (2016).

Recovery guarantee

Theorem (BA, 2015)

Let $w = \{w_i\}_{i \in I}$ *be weights,* $x \in \mathbb{C}^N$ *and* $\Delta \subseteq I_K$ *be such that* $\min_{i \in \{1,\ldots,K\} \setminus \Delta} \{w_i\} \geq 1$. *Let*

$$m \gtrsim \left(|\Delta|_u + \max_{i \in I_K \setminus \Delta} \{u_i^2 / w_i^2\} \max\{|\Delta|_w, 1\} \right) \cdot L,$$

where $u_i = \|\phi_i\|_{L^\infty}$ *and* $L = \log(2\epsilon^{-1}) \cdot \log(2N\sqrt{\max\{|\Delta|_w, 1\}})$. *Draw* $z_1, \ldots, z_m$ *independently from* ν. *Then with probability at least* $1 - \epsilon$, *any minimizer* $\hat{x}$ *of*

$$\min_{v \in \mathbb{C}^N} \|v\|_{1,w} \text{ subject to } \|Av - y\|_2 \leq \eta,$$

satisfies

$$\|x - \hat{x}\|_2 \lesssim \|x - P_\Delta x\|_{1,w} + \eta\sqrt{|\Delta|_w / m}.$$

- The ℓ^2/ℓ^1_w error bound is worse than those implied by the RIP. For ℓ^1_w/ℓ^1_w bounds ($w = u$ only), see Chkifa, Dexter, Tran & Webster (2016).
- Earlier work (weighted RIP): Rauhut & Ward (2015).

Optimal non-adapted weights

Consider the main estimate:

$$m \gtrsim \left(|\Delta|_u + \max_{i \in I_K \setminus \Delta} \{u_i^2 / w_i^2\} \max\{|\Delta|_w, 1\} \right) \cdot L.$$

For generic choices of Δ, this is minimized by the choice

$$w_i = u_i = \|\phi_i\|_{L^\infty}.$$

Comparison of recovery guarantees:

ℓ^1 minimization	$m \gtrsim \max_{i \in I_K} \|\phi_i\|_{L^\infty}^2 \cdot |\Delta| \cdot L$ (1)
ℓ^1_u minimization	$m \gtrsim \left(\sum_{i \in \Delta} \|\phi_i\|_{L^\infty}^2\right) \cdot L$ (2)

For suitable Δ, we next show that (2) is substantially smaller than (1).

Optimal non-adapted weights

Consider the main estimate:

$$m \gtrsim \left(|\Delta|_u + \max_{i \in I_K \setminus \Delta} \{u_i^2 / w_i^2\} \max\{|\Delta|_w, 1\} \right) \cdot L.$$

For generic choices of Δ, this is minimized by the choice

$$w_i = u_i = \|\phi_i\|_{L^\infty}.$$

Comparison of recovery guarantees:

ℓ^1 minimization	$m \gtrsim \max_{i \in I_K} \|\phi_i\|_{L^\infty}^2 \cdot |\Delta| \cdot L$ (1)
ℓ_u^1 minimization	$m \gtrsim \left(\sum_{i \in \Delta} \|\phi_i\|_{L^\infty}^2 \right) \cdot L$ (2)

For suitable Δ, we next show that (2) is substantially smaller than (1).

Polynomial expansions and lower sets

Question: which types of support sets Δ do we encounter in practice?

Answer: In high dimensions, polynomial coefficients tend to concentrate on lower sets (see e.g. Chkifa, Cohen & Schwab, 2014).

$d = 2$, $s = 16$ $d = 2$, $s = 32$

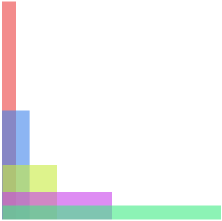
Definition (Lower/Downwards closed set)

A set $\Delta \subseteq \mathbb{N}^d$ is lower if, for any $i = (i_1, \ldots, i_d) \in \Delta$ and $j = (j_1, \ldots, j_d)$ with $j_k \leq i_k$, $\forall k$, it holds that $j \in \Delta$.

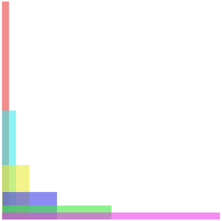
Polynomial expansions and lower sets

Question: which types of support sets Δ do we encounter in practice?

Answer: In high dimensions, polynomial coefficients tend to concentrate on lower sets (see e.g. Chkifa, Cohen & Schwab, 2014).



$d = 2,\ s = 16$



$d = 2,\ s = 32$

Definition (Lower/Downwards closed set)

A set $\Delta \subseteq \mathbb{N}^d$ is lower if, for any $i = (i_1, \ldots, i_d) \in \Delta$ and $j = (j_1, \ldots, j_d)$ with $j_k \leq i_k$, $\forall k$, it holds that $j \in \Delta$.

Optimal recovery of lower sets

The case $d < K$:

Basis	Samples	Measurements m $w_i = 1$	Measurements m $w_i = u_i$
Chebyshev	Chebyshev	$2^d \cdot s \cdot L$	$s^{\frac{\log(3)}{\log(2)}} \cdot L$
Legendre	Uniform	$\left(\frac{2K}{d}+1\right)^d \cdot s \cdot L$	$s^2 \cdot L$
Legendre	Chebyshev	$2^d \cdot s \cdot L$	$(\pi/2)^d s^{\frac{\log(1+4/\pi)}{\log(2)}} \cdot L$

The case $d \geq K$:

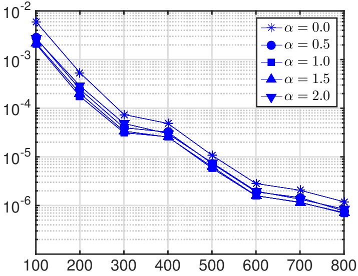
Basis	Samples	Measurements m $w_i = 1$	Measurements m $w_i = u_i$
Chebyshev	Chebyshev	$2^K \cdot s \cdot L$	$s^{\frac{\log(3)}{\log(2)}} \cdot L$
Legendre	Uniform	$3^K \cdot s \cdot L$	$s^2 \cdot L$
Legendre	Chebyshev	$(\pi/2)^d (4/\pi)^K \cdot s \cdot L$	$(\pi/2)^d s^{\frac{\log(1+4/\pi)}{\log(2)}} \cdot L$

Adcock (2015), Chkifa, Dexter, Tran & Webster (2016)

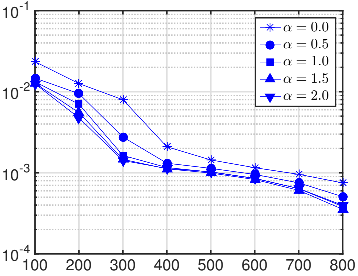
Numerical examples

Example 1: polynomials = Chebyshev, sampling = Chebyshev measure

- intrinsic weights $u_i = 2^{|i|_0/2}$
- optimization weights $w_i = (u_i)^\alpha$
- I_K is the total degree set of degree K
- $f(z) = \log(2 + d^{-1}(z_1 + \ldots + z_d))$



$(d, K) = (3, 24)$

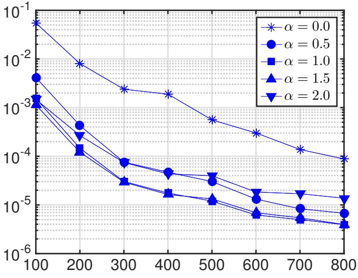


$(d, K) = (10, 5)$

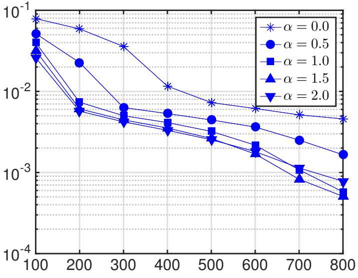
Numerical examples

Example 2: polynomials = Legendre, sampling = uniform measure

- intrinsic weights $u_i = \prod_{j=1}^{d} \sqrt{2i_j + 1}$
- optimization weights $w_i = (u_i)^\alpha$
- I_K is the total degree set of degree K
- $f(z) = \exp(-(z_1 + \ldots + z_d)/(2d))$



$(d, K) = (3, 24)$



$(d, K) = (10, 5)$

Comparison to least-squares fitting

Least-squares fitting: Preselect a support set Δ. Compute

$$\check{x} = \underset{\mathrm{supp}(v) \subseteq \Delta}{\operatorname{argmin}} \|A_\Delta v - y\|_2$$

- Theoretical guarantees: Cohen, Davenport & Leviatan (2013), Chkifa, Cohen, Migliorati, Nobile & Tempone (2015), Migliorati (2015) and others.

The sample complexity for the recovery of any set Δ is identical (up to possible log factors) to those of weighted ℓ^1 minimization with weights $w = u$ for Chebyshev/uniform sampling on bounded domains.

However, weighted ℓ^1 minimization requires no prior knowledge of Δ.

Adapted weights

In some scenarios, we may have some a priori knowledge about which coefficients in the expansion

$$f = \sum_{i \in I} x_i \phi_i,$$

are the largest. E.g. theoretical estimates, prior computations, etc.

Adapted weights: Use weights to penalize the expansion coefficients that are expected to be small.

Peng, Hampton & Doostan (2014), Yang & Karniadakis (2013), and others

Adapted weights

Corollary (BA, 2015)

Assume $u_i = 1$ for simplicity and let $x \in \mathbb{C}^N$ be s-sparse with support $\Delta = \{j : x_j \neq 0\}$. Let $\Gamma \subseteq I_K$ and suppose that $w_i = \sigma < 1$, $i \in \Gamma$, and $w_i = 1$, $i \notin \Gamma$. Then we require

$$m \gtrsim (2(1-\rho\alpha) + (1+\gamma)\rho) \cdot s \cdot L, \qquad L = \log(2\epsilon^{-1}) \cdot \log(2N\sqrt{s}),$$

measurements, where $\alpha = |\Delta \cap \Gamma|/|\Gamma|$ and $|\Gamma|/|\Delta| = \rho$.

- If $w_i = 1$ then we require $m \gtrsim 2 \cdot s \cdot L$ measurements.
- Hence we see an improvement whenever $\alpha > \frac{1}{2}(1+\gamma)$.
- That is, we estimate $\approx 50\%$ of the support correctly, for small γ.

Similar results: Friedlander, Mansour, Saab & Yilmaz (2012), Yu & Baek (2013), Mansour & Saab (2015) (random Gaussian measurements).

Ideas behind the proof of the main result

The proof is based on constructing an approximate dual certificate:

Lemma (BA, 2015)

Let $\Delta \subseteq \{1,\ldots,N\}$, $|\Delta| = s$. Suppose that A is such that

(i) $\|P_\Delta A^* A P_\Delta - P_\Delta\|_{2\to 2} \leq \alpha$ – *local isometry,*

(ii) $\max_{i\notin\Delta}\{\|Ae_i\|_2/w_i\} \leq \beta$ – *off-support incoherence,*

and that there exists a vector $\rho = W^{-1}A^*\xi$ for some $\xi \in \mathbb{C}^m$ such that

(iii) $\|W(P_\Delta\rho - \mathrm{sign}(P_\Delta x))\|_2 \leq \gamma$ – *approximate sign matching on Δ,*

(iv) $\|P_\Delta^\perp \rho\|_\infty \leq \theta$ – *strictly less than one off Δ,*

(v) $\|\xi\|_2 \leq \lambda\sqrt{|\Delta|_w}$ – *bounded growth,*

for $0 \leq \alpha, \theta < 1$ and $\beta, \gamma, \lambda \geq 0$ satisfying $\frac{\sqrt{1+\alpha}\beta\gamma}{(1-\alpha)(1-\theta)} < 1$. Then the conclusions of the theorem hold with $L = \lambda$.

Note that (i) and (ii) follow from standard concentration estimates.

Constructing the dual certificate

The construction of the dual certificate ρ uses an iterative approach known as the golfing scheme and due to D. Gross.

- First, one divides the rows of A into L bins, of sizes $m_1, \dots, m_L$.
- Set $\rho^{(0)} = 0$.
- For $l = 1, \dots, L$ perform the iterative update

$$\rho^{(l)} = m_l^{-1} W^{-1} (A^{(l)})^* A^{(l)} \left(\mathrm{sign}(P_\Delta x) - P_\Delta \rho^{(l-1)} \right) + \rho^{(l-1)},$$

  provided
  - $\|(P_\Delta - m_l^{-1} P_\Delta (A^{(l)})^* A^{(l)} P_\Delta) v^{(l-1)}\|_2 \leq a_l \|v^{(l-1)}\|_2$,
  - $\|m^{-1} P_\Delta^\perp W^{-1} (A^{(l)})^* A^{(l)} P_\Delta v^{(l-1)}\|_\infty \leq b_l \|v^{(l-1)}\|_2$,

  where $v^{(l)} = W \left(\mathrm{sign}(P_\Delta x) - P_\Delta \rho^{(l)} \right)$.
- The parameters $m_1, \dots, m_L$, L, a_l, b_l are carefully tuned to get the correct recovery guarantee.

Off versus on-support terms

Main estimate:

$$m \gtrsim \left(|\Delta|_u + \max_{i \in I_K \setminus \Delta} \{u_i^2 / w_i^2\} \max\{|\Delta|_w, 1\} \right) \cdot L = (T_1 + T_2) \cdot L.$$

Roughly speaking:

- T_1 comes from estimating the on-support terms.
- E.g. the local isometry property $\|P_\Delta A^* A P_\Delta - P_\Delta\|_{2\to 2}$.

- T_2 comes from estimating the off-support terms.
- E.g. the off-support coherence $\max_{i \notin \Delta} \{\|Ae_i\|_2 / w_i\} \leq \beta$.

Outline

Recovery of functions

Usually, a function f is not exactly a polynomial of finite degree. Instead it has an infinite expansion:

$$f(z) = \sum_{i \in I} x_i \phi_i(z).$$

Typical approach: Let $\eta \geq 0$ be chosen so that the expansion tail satisfies $\left\| f - \sum_{i \in I_K} x_i \phi_i \right\|_{L^\infty} \leq \eta$. Solve the problem

$$\min_{v \in \mathbb{C}^N} \|v\|_{1,w} \text{ subject to } \|Av - y\|_2 \leq \eta. \qquad (\star)$$

- Note that this condition ensures the vector $v^* = \{x_i\}_{i \in I_K}$ of the first $N = |I_K|$ exact coefficients is feasible for $(\star)$.
- The expansion tail is treated as noise on the samples.

Problems

In practice, the tail error $\left\|f - \sum_{i \in I_K} x_i \phi_i\right\|_{L^\infty}$ is unknown.

- Empirical solution: use cross validation.
- See, for example: Doostan & Owhadi (2011), Yang & Karniadakis (2013), Peng, Hampton & Doostan (2014).
- However, time-consuming to compute (multiple ℓ^1 solves).

Moreover, even if η can be estimated, the majority of existing theoretical results require η to satisfy

$$\eta \geq \left\| f - \sum_{i \in I_K} x_i \phi_i \right\|_{L^\infty}.$$

Recover without tail bounds

Suppose that $\eta \geq 0$ is arbitrary, and consider

$$\min_{v \in \mathbb{C}^N} \|v\|_{1,w} \text{ subject to } \|Av - y\|_2 \leq \eta. \qquad (\star)$$

Theorem (BA, 2015)

Let $w = \{w_i\}_{i \in \mathbb{N}}$ be weights, $x \in \ell^1_w(\mathbb{N})$ and $\Delta \subseteq \{1, \ldots, K\}$ be such that $\min_{i \in \{1,\ldots,K\} \setminus \Delta} \{w_i\} \geq 1$. Let

$$m \gtrsim \left(|\Delta|_u + \max_{i \in I_K \setminus \Delta} \{u_i^2 / w_i^2\} \max\{|\Delta|_w, 1\} \right) \cdot L,$$

where $L = \log(\epsilon^{-1}) \cdot \log(2N\sqrt{\max\{|\Delta|_w, 1\}})$ and draw $z_1, \ldots, z_m$ independently from ν. Then, with probability at least $1 - \epsilon$, any minimizer of $(\star)$ satisfies

$$\|x - \hat{x}\|_2 \lesssim \|x - P_\Delta x\|_{1,w} + \eta\sqrt{|\Delta|_w / m} + T_K(x),$$

where $T_K(x) = \min\left\{ \|x - v\|_{1,w} : v \in \mathbb{C}^N, \|Av - y\|_2 \leq \eta \right\}$.

Remarks

1. The measurement condition

$$m \gtrsim \left(|\Delta|_u + \max_{i \in I_K \setminus \Delta} \{u_i^2 / w_i^2\} \max\{|\Delta|_w, 1\} \right) \cdot L,$$

is the same as before.

2. The effect of the unknown expansion tail is the additional term

$$T_K(x) = \min \left\{ \|x - v\|_{1,w} : v \in \mathbb{C}^N, \|Av - y\|_2 \leq \eta \right\},$$

i.e. the error of best approximation of x from the feasible set.

3. If $\eta = 0$, then the overall approximation $\tilde{f} = \sum_{i \in I_K} \hat{x}_i \phi_i$ interpolates f.

Remarks

1. The measurement condition

$$m \gtrsim \left(|\Delta|_u + \max_{i \in I_K \setminus \Delta} \{u_i^2/w_i^2\} \max\{|\Delta|_w, 1\} \right) \cdot L,$$

is the same as before.

2. The effect of the unknown expansion tail is the additional term

$$T_K(x) = \min \left\{ \|x - v\|_{1,w} : v \in \mathbb{C}^N, \|Av - y\|_2 \le \eta \right\},$$

i.e. the error of best approximation of x from the feasible set.

3. If $\eta = 0$, then the overall approximation $\tilde{f} = \sum_{i \in I_K} \hat{x}_i \phi_i$ interpolates f.

Remarks

1. The measurement condition

$$m \gtrsim \left(|\Delta|_u + \max_{i \in I_K \setminus \Delta} \{u_i^2 / w_i^2\} \max\{|\Delta|_w, 1\} \right) \cdot L,$$

is the same as before.

2. The effect of the unknown expansion tail is the additional term

$$T_K(x) = \min \left\{ \|x - v\|_{1,w} : v \in \mathbb{C}^N, \|Av - y\|_2 \leq \eta \right\},$$

i.e. the error of best approximation of x from the feasible set.

3. If $\eta = 0$, then the overall approximation $\tilde{f} = \sum_{i \in I_K} \hat{x}_i \phi_i$ interpolates f.

Estimates for $T_K(x)$

Case 1 (known tail): If $\eta \geq \left\| f - \sum_{i \in I_K} x_i \phi_i \right\|_{L^\infty}$ then

$$T_K(x) \leq \sum_{i \notin I_K} w_i |x_i|.$$

Case 2 (unknown tail): If $0 \leq \eta < \left\| f - \sum_{i \in I_K} x_i \phi_i \right\|_{L^\infty}$, then

$$T_K(x) \leq \left(1 + \sigma^{-1} \|P_{I_K} w\|_2\right) \sum_{i \notin I_K} w_i |x_i|,$$

where $\sigma = \sigma_{\min}(A)$.

Note that $\sum_{i \notin I_K} w_i |x_i|$ is the ℓ^1_w-norm of the coefficients not included in the optimization problem.

Outline

Conclusions

1. Sparsity and limited measurements often arise in UQ problems. In particular, computing high-dimensional polynomial approximations to solutions of parametric PDEs.

2. CS can be a useful tool in such problems.

3. The sample complexity for CS is comparable to that of an oracle least squares and mitigates the curse of dimensionality to the same extent.

4. The standard formulation of CS is finite-dimensional. To incorporate functions requires both additional theory and numerical tools to get good tail estimates.

Recent enhancements

Sampling: Sampling from the orthogonality measure may not be the best in practice. Some alternatives:

- Randomized quadratures (Tang & Iaccarino (2014), Guo, Narayan, Zhou & Chen (2016))
- Coherence-optimal sampling (Hampton & Doostan (2014))
- Christoffel/Equipotential sampling (Jakeman, Narayan & Zhou (2016))

Recovery algorithms: ℓ^1 minimization is a convex relaxation of ℓ^0 minimization. Nonconvex alternatives:

- Reweighted ℓ^1 (Yang & Karniadakis (2013))
- $\ell^1 - \ell^2$ (Guo, Narayan, Zhou (2016))

Also:

- Sparsity enhancement via rotations (Jakeman, Eldred & Sargsyan (2015), Lei, Yang, Zheng, Lin & Baker (2014))
- Gradient enhancement (Hampton & Doostan (2016) and others)

Recent enhancements

Sampling: Sampling from the orthogonality measure may not be the best in practice. Some alternatives:

- Randomized quadratures (Tang & Iaccarino (2014), Guo, Narayan, Zhou & Chen (2016))
- Coherence-optimal sampling (Hampton & Doostan (2014))
- Christoffel/Equipotential sampling (Jakeman, Narayan & Zhou (2016))

Recovery algorithms: ℓ^1 minimization is a convex relaxation of ℓ^0 minimization. Nonconvex alternatives:

- Reweighted ℓ^1 (Yang & Karniadakis (2013))
- $\ell^1 - \ell^2$ (Guo, Narayan, Zhou (2016))

Also:

- Sparsity enhancement via rotations (Jakeman, Eldred & Sargsyan (2015), Lei, Yang, Zheng, Lin & Baker (2014))
- Gradient enhancement (Hampton & Doostan (2016) and others)

Recent enhancements

Sampling: Sampling from the orthogonality measure may not be the best in practice. Some alternatives:

- Randomized quadratures (Tang & Iaccarino (2014), Guo, Narayan, Zhou & Chen (2016))
- Coherence-optimal sampling (Hampton & Doostan (2014))
- Christoffel/Equipotential sampling (Jakeman, Narayan & Zhou (2016))

Recovery algorithms: ℓ^1 minimization is a convex relaxation of ℓ^0 minimization. Nonconvex alternatives:

- Reweighted ℓ^1 (Yang & Karniadakis (2013))
- $\ell^1 - \ell^2$ (Guo, Narayan, Zhou (2016))

Also:

- Sparsity enhancement via rotations (Jakeman, Eldred & Sargsyan (2015), Lei, Yang, Zheng, Lin & Baker (2014))
- Gradient enhancement (Hampton & Doostan (2016) and others)

Outlook

Solvers: Solving an ℓ^1 problem is computationally more intensive than solving a least squares problem. General purpose solvers may not be optimal for these problems.

Adaptivity: It is not clear how to adaptively sample in the CS setting.

Sparsity enhancement: learning a better expansion basis from the data (sparse learning). E.g. combining ideas from active subspaces.

Unbounded domains: less clear what are good sampling measures for Hermite or Laguerre expansions.

Recovering the whole solution: Rather than some functional $f(z) = \mathcal{Q}u(\cdot, z)$, devise efficient techniques to approximate $u(x, z)$.

Worst-case guarantees: Most existing results assume an ideal sampling. Few theoretical results exist for fixed (e.g. legacy) data.

Dealing with infinity: Better estimates for unknown expansion tails. Provable guarantees for estimation techniques, e.g. cross validation.

Robustness: Better estimates are need for robustness towards solver errors, including failures, data corruption, etc.

Outlook

Solvers: Solving an ℓ^1 problem is computationally more intensive than solving a least squares problem. General purpose solvers may not be optimal for these problems.

Adaptivity: It is not clear how to adaptively sample in the CS setting.

Sparsity enhancement: learning a better expansion basis from the data (sparse learning). E.g. combining ideas from active subspaces.

Unbounded domains: less clear what are good sampling measures for Hermite or Laguerre expansions.

Recovering the whole solution: Rather than some functional $f(z) = \mathcal{Q}u(\cdot, z)$, devise efficient techniques to approximate $u(x, z)$.

Worst-case guarantees: Most existing results assume an ideal sampling. Few theoretical results exist for fixed (e.g. legacy) data.

Dealing with infinity: Better estimates for unknown expansion tails. Provable guarantees for estimation techniques, e.g. cross validation.

Robustness: Better estimates are need for robustness towards solver errors, including failures, data corruption, etc.