Function interpolation and compressed sensing

Ben Adcock
Department of Mathematics
Simon Fraser University

Outline

Outline

Introduction

Infinite-dimensional framework

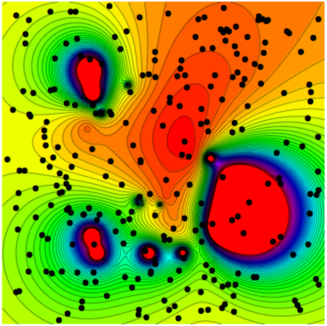
New recovery guarantees for weighted ℓ^1 minimization

References

High-dimensional approximation

Let

- $D \subseteq \mathbb{R}^d$ be a domain, $d \gg 1$
- $f : D \to \mathbb{C}$ be a (smooth) function
- $\{t_i\}_{i=1}^m$ be a set of sample points



Goal: Approximate f from $\{f(t_i)\}_{i=1}^m$.

Applications: Uncertainty Quantification (UQ), scattered data approximation, numerical PDEs,....

Main issue: curse of dimensionality (exponential blow-up with d).

Quantifying uncertainty via polynomial chaos expansions

Uncertainty Quantification: Understand how output f (the quantity of interest) of a physical system behave as functions of the inputs t (the parameters).

Polynomial Chaos Expansions: (Xiu & Karniadakis, 2002). Expand $f(t)$ using multivariate orthogonal polynomials

$$f(t) \approx \sum_{i=1}^{M} x_i \phi_i(t).$$

Non-intrusive methods: Recover $\{x_i\}_{i=1}^{M}$ from samples $\{f(t_i)\}_{i=1}^{m}$.

Stochastic Collocation

Two widely-used approaches:

Structured meshes and interpolation ($M = m$): E.g. Sparse grids.

- Efficient interpolation schemes in moderate dimensions
- But may be too structured for very high dimensions, or miss certain features (e.g. anisotropic behaviour).

Unstructured meshes and regression ($m > M$): Random sampling combined with least-squares fitting.

- For the right distributions, can obtain stable approximation with d-independent scaling of m and M.
- But still inefficient, especially in high dimensions.

Question

Can compressed sensing techniques be useful here?

Stochastic Collocation

Two widely-used approaches:

Structured meshes and interpolation ($M = m$): E.g. Sparse grids.

- Efficient interpolation schemes in moderate dimensions
- But may be too structured for very high dimensions, or miss certain features (e.g. anisotropic behaviour).

Unstructured meshes and regression ($m > M$): Random sampling combined with least-squares fitting.

- For the right distributions, can obtain stable approximation with d-independent scaling of m and M.
- But still inefficient, especially in high dimensions.

Question

Can compressed sensing techniques be useful here?

Compressed sensing in UQ

Theoretical work:

- Rauhut & Ward (2011), 1D Legendre polynomials
- Yan, Guo & Xiu (2012), dD Legendre polynomials
- Tang & Iaccarino (2014), randomized quadratures
- Hampton & Doostan (2014), coherence-optimized sampling
- Xu & Zhou (2014), deterministic sampling
- Rauhut & Ward (2014), weighted ℓ^1 minimization
- Chkifa, Dexter, Tran & Webster (2015), weighted ℓ^1 minimization

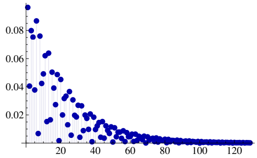
Applications to UQ:

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

Are polynomial coefficient sparse?

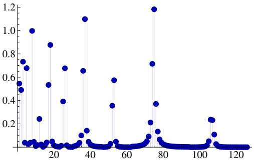
Low dimensions: polynomial coefficients exhibit decay, not sparsity:

Polynomial coefficients



Decay

Wavelet coefficients



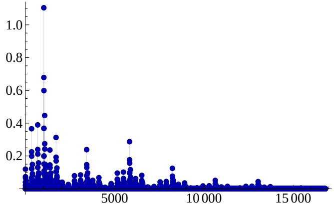
Sparsity

Nonlinear approximation error $\approx$ Linear approximation error

We may as well use interpolation/least squares.

Are polynomial coefficient sparse?

Higher dimensions: polynomial coefficients are increasingly sparse (Doostan et al., Schwab et al., Webster et al.,....).



Polynomial coefficients, $d = 10$

Nonlinear approximation error $\ll$ Linear approximation error

Sparsity and lower sets

In high dimensions, polynomial coefficients concentrate on lower sets:

Definition (Lower set)

A set $\Delta \subseteq \mathbb{N}^d$ is lower if, for any $i = (i_1, \dots, i_d)$ and $j = (j_1, \dots, j_d)$ with $j_k \leq i_k$, $\forall k$, we have

$$i \in \Delta \quad \Rightarrow \quad j \in \Delta.$$

Note: The number of lower sets of size s is $\mathcal{O}\left(s \log(s)^{d-1}\right)$.

Outline

Introduction

Infinite-dimensional framework

New recovery guarantees for weighted ℓ^1 minimization

References

Setup

Let

- ν be a measure on D with $\int_D \mathrm{d}\nu = 1$,
- $T = \{t_i\}_{i=1}^m \subseteq D$, $m \in \mathbb{N}$ be drawn independently from ν,
- $\{\phi_j\}_{j\in\mathbb{N}}$ be an orthonormal system in $L^2_\nu(D) \cap L^\infty(D)$ (typically, tensor algebraic polynomials).

Suppose that

$$f = \sum_{j\in\mathbb{N}} x_j \phi_j, \qquad x_j = \langle f, \phi_j \rangle_{L^2_\nu},$$

where $\{x_j\}_{j\in\mathbb{N}}$ are the coefficients of f in the system $\{\phi_j\}_{j\in\mathbb{N}}$.

Current approaches – discretize first

Most existing approaches follow a 'discretize first' approach.

Choose $M \geq m$ and solve the finite-dimensional problem

$$\min_{z \in \mathbb{C}^M} \|z\|_{1,w} \text{ subject to } \|Az - y\|_2 \leq \delta, \qquad (\star)$$

for some $\delta \geq 0$, where $\|z\|_{1,w} = \sum_{i=1}^{M} w_i |z_i|$, $\{w_i\}_{i=1}^{M}$ are weights and

$$A = \{\phi_j(t_i)\}_{i=1,j=1}^{m,M}, \quad y = \{f(t_i)\}_{i=1}^{m}.$$

If $\hat{x} \in \mathbb{C}^M$ is a minimizer, set $f \approx \tilde{f} = \sum_{i=1}^{M} \hat{x}_i \phi_i$.

Current approaches – discretize first

Most existing approaches follow a ‘discretize first’ approach.

Choose $M \geq m$ and solve the finite-dimensional problem

$$\min_{z \in \mathbb{C}^M} \|z\|_{1,w} \text{ subject to } \|Az - y\|_2 \leq \delta, \qquad (\star)$$

for some $\delta \geq 0$, where $\|z\|_{1,w} = \sum_{i=1}^{M} w_i |z_i|$, $\{w_i\}_{i=1}^{M}$ are weights and

$$A = \{\phi_j(t_i)\}_{i=1,j=1}^{m,M}, \quad y = \{f(t_i)\}_{i=1}^{m}.$$

If $\hat{x} \in \mathbb{C}^M$ is a minimizer, set $f \approx \tilde{f} = \sum_{i=1}^{M} \hat{x}_i \phi_i$.

The choice of δ

The parameter δ is chosen so that the best approximation $\sum_{i=1}^{M} x_i \phi_i$ to f from $\mathrm{span}\{\phi_i\}_{i=1}^{M}$ is feasible for $(\star)$.

In other words, we require

$$\delta \geq \left\| f - \sum_{i=1}^{M} x_i \phi_i \right\|_{L^\infty} = \left\| \sum_{i>M} x_i \phi_i \right\|_{L^\infty}.$$

Equivalently, we treat the expansion tail as noise in the data.

Problems

- This tail error is unknown in general.
- A good estimation is necessary in order to get good accuracy.
- Empirical estimation via cross validation is tricky and wasteful.
- Solutions of $(\star)$ do not interpolate the data.

The choice of δ

The parameter δ is chosen so that the best approximation $\sum_{i=1}^{M} x_i \phi_i$ to f from $\mathrm{span}\{\phi_i\}_{i=1}^{M}$ is feasible for $(\star)$.

In other words, we require

$$\delta \geq \left\| f - \sum_{i=1}^{M} x_i \phi_i \right\|_{L^\infty} = \left\| \sum_{i>M} x_i \phi_i \right\|_{L^\infty}.$$

Equivalently, we treat the expansion tail as noise in the data.

Problems

- This tail error is unknown in general.
- A good estimation is necessary in order to get good accuracy.
- Empirical estimation via cross validation is tricky and wasteful.
- Solutions of $(\star)$ do not interpolate the data.

New approach

We propose the infinite-dimensional ℓ^1 minimization

$$\inf_{z \in \ell^1_w(\mathbb{N})} \|z\|_{1,w} \text{ subject to } Uz = y,$$

where $y = \{f(t_i)\}_{i=1}^m$, $\{w_i\}_{i \in \mathbb{N}}$ are weights and

$$U = \{\phi_j(t_i)\}_{i=1,j=1}^{m,\infty} \in \mathbb{C}^{m \times \infty},$$

is an infinitely fat matrix.

Advantages

- Solutions are interpolatory.
- No need to know the expansion tail.
- Agnostic to the ordering of the functions $\{\phi_i\}_{i \in \mathbb{N}}$.

Note: a similar setup can also handle noisy data.

New approach

We propose the infinite-dimensional ℓ^1 minimization

$$\inf_{z \in \ell^1_w(\mathbb{N})} \|z\|_{1,w} \text{ subject to } Uz = y,$$

where $y = \{f(t_i)\}_{i=1}^m$, $\{w_i\}_{i\in\mathbb{N}}$ are weights and

$$U = \{\phi_j(t_i)\}_{i=1,j=1}^{m,\infty} \in \mathbb{C}^{m\times\infty},$$

is an infinitely fat matrix.

Advantages

- Solutions are interpolatory.
- No need to know the expansion tail.
- Agnostic to the ordering of the functions $\{\phi_i\}_{i\in\mathbb{N}}$.

Note: a similar setup can also handle noisy data.

Discretization

We cannot numerically solve the problem

$$\inf_{z \in \ell^1_w(\mathbb{N})} \|z\|_{1,w} \text{ subject to } Uz = y. \tag{1}$$

Discretization strategy: Introduce a parameter $K \in \mathbb{N}$ and solve the finite-dimensional problem

$$\min_{z \in P_K(\ell^1_w(\mathbb{N}))} \|z\|_{1,w} \text{ subject to } UP_K z = y, \tag{2}$$

where P_K is defined by $P_K z = \{z_1, \ldots, z_K, 0, 0, \ldots\}$.

- Note: UP_K is equivalent to a fat $m \times K$ matrix.

Main Idea

Choose K suitably large, and independent of f, so that solutions of (2) are close to solutions of (1).

How to choose K

Let $T_K(x)$ be the additional error introduced by this discretization.

Theorem (BA)

Let $x \in \ell^1_{\tilde{w}}(\mathbb{N})$, where $\tilde{w}_i \geq \sqrt{i} w_i^2$, $\forall i$. Suppose that K is sufficiently large so that $\sigma_r = \sigma_r(P_K U^*) > 0$, where $r = \mathrm{rank}(U)$. Then

$$T_K(x) \leq \|x - P_K x\|_{1,w} + 1/\sigma_r \|x - P_K x\|_{1,\tilde{w}}.$$

The truncation condition $\sigma_r \approx 1$ depends only on T and $\{\phi_i\}_{i\in\mathbb{N}}$ and is independent of the function f to recover.

Example: Let $D = (-1,1)^d$ with tensor Jacobi polynomials or the Fourier basis and equispaced data. Then $K = \mathcal{O}\left(m^{1+\epsilon}\right)$, $\epsilon > 0$, suffices.

Rule-of-thumb

Letting $K \approx 4m$ works fine in most settings.

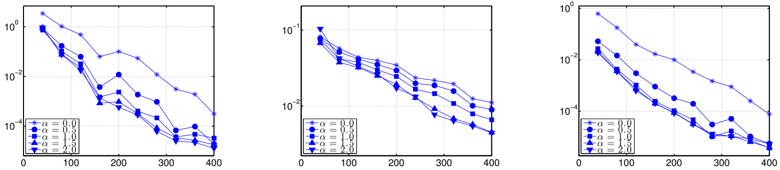
Outline

Background

Unweighted ℓ^1 minimization:

- Recovery guarantees: Rauhut & Ward (2011), Yan, Guo & Xiu (2012).
- Applications to UQ: Doostan & Owhadi (2011), Mathelin & Gallivan (2012), Hampton & Doostan (2014), Tang & Iaccarino (2014), Guo, Narayan, Xiu & Zhou (2015).

Weighted ℓ^1 minimization: Observed empirically to give superior results.



$f(t) = \mathrm{e}^{2t_1}\cos(3t_2)$ $\quad$ $f(t) = \sin(\mathrm{e}^{t_1 t_2 t_3}/2)$ $\quad$ $f(t) = \mathrm{e}^{-\frac{t_1+t_2+t_3+t_4}{6}}$

Plot of error versus m with algebraic weights: $w_i = (i_1 \cdots i_d)^\alpha$, $\alpha \geq 0$.

Standard weighting strategies

Non-adapted weights: Slowly-growing (e.g. algebraic) weights used to alleviate aliasing/overfitting.

- Rauhut & Ward (2014), Rauhut & Schwab (2015), BA (2015).

Adapted weights: Weights chosen according to support estimates.

- A priori estimates: Peng, Hampton & Doostan (2014).
- Iterative re-weighting: Yang & Karniadakis (2014).
- See also: Bah & Ward (2015).

Goal

Find recovery guarantees that explain the effectiveness of both strategies.

Existing recovery guarantees

Rauhut & Ward (2014):

- Weights: $w_i \geq \|\phi_i\|_{L^\infty}$
- Weighted sparsity: $s = |\Delta|_w = \sum_{i \in \Delta} w_i^2$, where $\Delta = \mathrm{supp}(x)$.
- Recovery guarantee: $m \gtrsim s \times$ log factors.

Problem: This is not sharp. Let $w_i = i^\alpha$ and suppose that f is such that

$$x_j \neq 0, \quad 1 \leq j \leq k, \qquad x_j \approx 0, \quad j > k.$$

This is reasonable for oscillatory functions, for example. Then

$$m \gtrsim k^{2\alpha+1} \times \text{log factors.}$$

This estimate deteriorates with increasing α.

- Note: The same argument generalizes to any dimension when the coefficients lie on a hyperbolic cross, BA (2015).

Existing recovery guarantees

Rauhut & Ward (2014):

- Weights: $w_i \geq \|\phi_i\|_{L^\infty}$
- Weighted sparsity: $s = |\Delta|_w = \sum_{i \in \Delta} w_i^2$, where $\Delta = \mathrm{supp}(x)$.
- Recovery guarantee: $m \gtrsim s \times$ log factors.

Problem: This is not sharp. Let $w_i = i^\alpha$ and suppose that f is such that

$$x_j \neq 0, \quad 1 \leq j \leq k, \qquad x_j \approx 0, \quad j > k.$$

This is reasonable for oscillatory functions, for example. Then

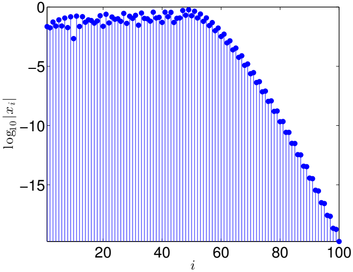
$$m \gtrsim k^{2\alpha+1} \times \text{log factors.}$$

This estimate deteriorates with increasing α.

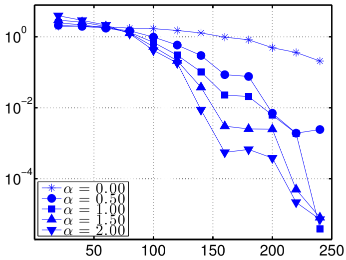
- Note: The same argument generalizes to any dimension when the coefficients lie on a hyperbolic cross, BA (2015).

Example

Take $f(t) = \cos(45\sqrt{2}t + 1/3)$ and consider Chebyshev polynomials with random samples drawn from the Chebyshev measure.



Coefficients x_j



Error versus m

A new recovery guarantee

Theorem (BA)

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 $t_1, \ldots, t_m$ be drawn independently from ν. Then

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

with probability at least $1 - \epsilon$, provided

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

where $u_i = \max\{\|\phi_i\|_{L^\infty}, 1\}$ and $L = \log(\epsilon^{-1}) \cdot \log(2N\sqrt{\max\{|\Delta|_w, 1\}})$.

Remarks:

- The weights u_i are intrinsic to the problem.
- This is a nonuniform guarantee – $(\star)$ relies heavily on this approach.
- As is typical, the error bound is weaker (ℓ^2/ℓ^1_w).

Consequence I: Sharpness for linear models

Consider the main estimate:

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

Sharpness for linear models: Let $\Delta = \{1, \dots, k\}$. Suppose that $u_i = \mathcal{O}\left(i^\gamma\right)$ and $w_i = \mathcal{O}\left(i^\alpha\right)$ for $\alpha > \gamma \geq 0$. Then

$$m \gtrsim k^{2\gamma+1} \times \text{log factors.}$$

- This is independent of the weights and optimal, up to log factors.
- Extends to any dimension for coefficients lying on a hyperbolic cross.

Consequence II: Optimal non-adapted weights

For non-adapted weights, the estimate

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

is minimized by setting $w_i = u_i$.

Example 1: Legendre polynomials, uniform measure.

- $w_i = 1$: $m \gtrsim 3^d \cdot s \cdot L$, where $s = |\Delta|$.
- $w_i = u_i$: $m \gtrsim s^2 \cdot L$ provided Δ is a lower set.
- Note that s^2 is sharp and avoids the curse of dimensionality.

Example 2: Chebyshev polynomials, Chebyshev measure.

- $w_i = 1$: $m \gtrsim 2^d \cdot s \cdot L$.
- $w_i = u_i$: $m \gtrsim s^{\log(3)/\log(2)} \cdot L$ provided Δ is a lower set.

Consequence II: Optimal non-adapted weights

For non-adapted weights, the estimate

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

is minimized by setting $w_i = u_i$.

Example 1: Legendre polynomials, uniform measure.

- $w_i = 1$: $m \gtrsim 3^d \cdot s \cdot L$, where $s = |\Delta|$.
- $w_i = u_i$: $m \gtrsim s^2 \cdot L$ provided Δ is a lower set.
- Note that s^2 is sharp and avoids the curse of dimensionality.

Example 2: Chebyshev polynomials, Chebyshev measure.

- $w_i = 1$: $m \gtrsim 2^d \cdot s \cdot L$.
- $w_i = u_i$: $m \gtrsim s^{\log(3)/\log(2)} \cdot L$ provided Δ is a lower set.

Consequence II: Optimal non-adapted weights

For non-adapted weights, the estimate

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

is minimized by setting $w_i = u_i$.

Example 1: Legendre polynomials, uniform measure.

- $w_i = 1$: $m \gtrsim 3^d \cdot s \cdot L$, where $s = |\Delta|$.
- $w_i = u_i$: $m \gtrsim s^2 \cdot L$ provided Δ is a lower set.
- Note that s^2 is sharp and avoids the curse of dimensionality.

Example 2: Chebyshev polynomials, Chebyshev measure.

- $w_i = 1$: $m \gtrsim 2^d \cdot s \cdot L$.
- $w_i = u_i$: $m \gtrsim s^{\log(3)/\log(2)} \cdot L$ provided Δ is a lower set.

Consequence III: The benefits of adapted weights

Corollary (BA)

Assume $u_i = 1$ for simplicity. Let x be s-sparse with support Δ. Let $\Gamma \subseteq \{1, \ldots, K\}$ and suppose that $w_i = \sigma < 1$, $i \in \Gamma$, and $w_i = 1$, $i \notin \Gamma$. Then we require

$$m \gtrsim (2(1 - \rho\alpha) + (1 + \gamma)\rho) \cdot s \cdot L,$$

measurements, where

$$\alpha = |\Delta \cap \Gamma| / |\Gamma|, \qquad |\Gamma| / |\Delta| = \rho.$$

- Recall that $m \gtrsim 2 \cdot s \cdot L$ in the unweighted case.
- 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.

Related work:

- Friedlander, Mansour, Saab & Yilmaz (2012), Yu & Baek (2013) (random Gaussian measurements).

Outline

Thanks!

For more info, see:

B. Adcock, *Infinite-dimensional weighted ℓ^1 minimization and function approximation from pointwise data*, arXiv:1503.02352 (2015).

B. Adcock, *Infinite-dimensional compressed sensing and function interpolation*, arXiv:1509.06073 (2015).