ℓ^1 minimization and function interpolation

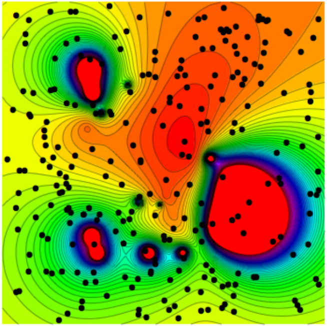
Ben Adcock
Department of Mathematics
Simon Fraser University

Outline

Setup

Let

- $D \subseteq \mathbb{R}^d$ be a domain
- $f : D \to \mathbb{C}$ be a (smooth) function
- $T = \{t_i\}_{i=1}^m \subseteq D$ a set of points
- $\{\phi_j\}_{j\in\mathbb{N}}$ an orthonormal system of functions (e.g. polynomials)



Problem: Recover f in the basis $\{\phi_j\}_{j\in\mathbb{N}}$ from the samples $\{f(t_i)\}_{i=1}^m$.

Motivations

Applications:

- Uncertainty Quantification (UQ)
- Scattered data approximation
- Numerical PDEs
- ....

Issues:

1. Amount of data is severely limited.
2. Dimension may be high (curse of dimensionality).
3. In some cases the data points T may be fixed and unstructured, in others they can be chosen to get the best approximation.

Questions

Over the last decade, ℓ^1 minimization has been shown to be effective for recovering sparse vectors from limited data (i.e. compressed sensing).

1. How does one properly use ℓ^1 minimization techniques for the function approximation problem?

2. What can we say about the approximation error?

3. What are the advantages over standard techniques, e.g. least squares?

This talk

(1) An overview of recent progress on using ℓ^1 minimization techniques for function interpolation.

(2) A new infinite-dimensional framework for this problem.

(3) Discussion of the role weights play in the optimization.

(4) Approximation theory for ℓ^1 minimization in various settings.

Outline

Polynomial approximation

Univariate orthogonal polynomials:

- Bounded domains, $D = (-1, 1)$
    - Legendre, $\nu(t) \propto 1$,
    - Chebyshev, $\nu(t) \propto \frac{1}{\sqrt{1-t^2}}$,
    - Jacobi, $\nu(t) \propto (1-t)^{\alpha}(1+t)^{\beta}$.
- Unbounded domains
    - Laguerre, $D = (0, \infty)$, $\nu(t) \propto \mathrm{e}^{-t}$
    - Hermite, $D = (-\infty, \infty)$, $\nu(t) \propto \mathrm{e}^{-t^2}$.

Multivariate orthogonal polynomials:

- Extension via tensor products.
- Truncated spaces: tensor product (too large in high dimensions), total degree, hyperbolic cross,...

Sampling: typically, points t_i are chosen randomly from the measure $\nu(t)$.

Polynomial approximation with ℓ^1 minimization

Rauhut & Ward (2011)

- One-dimensional Legendre polynomials with preconditioning trick.
- Compressed sensing recovery guarantees for sparse coefficients.
- Sampling points drawn randomly from the Chebyshev measure.

Yan, Guo & Xiu (2012)

- Generalize Rauhut & Ward to d-dimensions. Exponential growth in recovery guarantee with dimension.
- Uniform sampling points: d-independent recovery guarantee for large d. Holds whenever the total degree polynomial space is used, and for $d \geq P$.

See also:

- Doostan & Owhadi (2011), Mathelin & Gallivan (2012),...

ℓ^1 minimization and sampling strategies

Hampton & Doostan (2014)

- Random sampling from continuous measures, based on analytical estimates for coherence.

Xu & Zhou (2014)

- Deterministic sampling based on Weyl points.
- Quadratic bottleneck.

Tang & Iaccarino (2014)

- Legendre polynomials, random subsampling from deterministic Gauss–Legendre nodes.
- Enhanced performance over Chebyshev sampling in some situations.

Guo, Narayan, Xiu & Zhou (2015)

- General polynomials, random subsampling from Gaussian nodes.

Weighted ℓ^1 minimization

Yang & Karniadakis (2013)

- Sparsity enhancement via iteratively reweighted ℓ^1 minimization (based on a general technique of Candès et al.).

Peng, Hampton & Doostan (2014)

- Weights chosen according to *a priori* estimates for expansion coefficients.
- Moderate improvements over unweighted case.

Rauhut & Ward (2014)

- Compressed sensing recovery estimates for fixed weights.
- Based on weighted sparsity and a weighted version of the RIP.
- Weights remove exponentially-large dimension dependence in some circumstances (caveat: weighted sparsity, not sparsity).

See also:

- Jo (2014), Rauhut & Schwab (2014), Bah & Ward (2015),...

First issue: dealing with infinity

Let $\{\phi_j\}_{j\in\mathbb{N}}$ be an orthonormal system and write

$$f = \sum_{j\in\mathbb{N}} x_j \phi_j, \qquad x_j = \langle f, \phi_j \rangle,$$

where $\{x_j\}_{j\in\mathbb{N}}$ are the coefficients of f in the system $\{\phi_j\}_{j\in\mathbb{N}}$. Even though f may be highly compressible, its expansion is typically infinite.

As we shall see, most current approaches do not deal with infinite expansions in a rigorous way.

Question 1

How do we deal with infinite expansions faithfully?

First issue: dealing with infinity

Let $\{\phi_j\}_{j\in\mathbb{N}}$ be an orthonormal system and write

$$f = \sum_{j\in\mathbb{N}} x_j \phi_j, \qquad x_j = \langle f, \phi_j \rangle,$$

where $\{x_j\}_{j\in\mathbb{N}}$ are the coefficients of f in the system $\{\phi_j\}_{j\in\mathbb{N}}$. Even though f may be highly compressible, its expansion is typically infinite.

As we shall see, most current approaches do not deal with infinite expansions in a rigorous way.

Question 1

How do we deal with infinite expansions faithfully?

Second issue: best and worst case guarantees

Good sampling: In some applications, we can sample in the right way to ensure the best CS recovery guarantee.

- Typically, this is random sampling from an appropriate measure.
- But is this empirically optimal?

Bad sampling: But in other problems, the samples may be fixed and unstructured.

- Does ℓ^1 minimization still work well here?
- How does it compare to classical techniques, i.e. least squares?

Question 2

Can we provide recovery guarantees for a variety of scenarios?

Second issue: best and worst case guarantees

Good sampling: In some applications, we can sample in the right way to ensure the best CS recovery guarantee.

- Typically, this is random sampling from an appropriate measure.
- But is this empirically optimal?

Bad sampling: But in other problems, the samples may be fixed and unstructured.

- Does ℓ^1 minimization still work well here?
- How does it compare to classical techniques, i.e. least squares?

Question 2

Can we provide recovery guarantees for a variety of scenarios?

Second issue: best and worst case guarantees

Good sampling: In some applications, we can sample in the right way to ensure the best CS recovery guarantee.

- Typically, this is random sampling from an appropriate measure.
- But is this empirically optimal?

Bad sampling: But in other problems, the samples may be fixed and unstructured.

- Does ℓ^1 minimization still work well here?
- How does it compare to classical techniques, i.e. least squares?

Question 2

Can we provide recovery guarantees for a variety of scenarios?

Third issue: weighted ℓ^1 minimization

A number of works have suggested to consider weighted ℓ^1 minimization:

- Yang and Karniadakis (2013)
- Peng, Hampton & Doostan (2014)
- Rauhut & Ward (2014)
- Jo (2014)
- Rauhut & Schwab (2014)
- Bah & Ward (2015)

Question 3

What role do the weights play?

Third issue: weighted ℓ^1 minimization

A number of works have suggested to consider weighted ℓ^1 minimization:

- Yang and Karniadakis (2013)
- Peng, Hampton & Doostan (2014)
- Rauhut & Ward (2014)
- Jo (2014)
- Rauhut & Schwab (2014)
- Bah & Ward (2015)

Question 3

What role do the weights play?

Outline

Setup

Let

- $T = \{t_i\}_{i=1}^m \subseteq D$, $m \in \mathbb{N}$ be a set of m points in D,
- ν be a measure on D with $\int_D \mathrm{d}\nu = 1$,
- $\{\phi_j\}_{j\in\mathbb{N}}$ 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 $\{\phi_j\}_{j\in\mathbb{N}}$.

The current state-of-the-art

Roughly speaking, all existing approaches use the following discretize first technique.

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

All current approaches pick δ so that the best approximation $\sum_{i=1}^{M} x_i \phi_i$ to f from $\mathrm{span}\{\phi_1, \ldots, \phi_M\}$ 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

(1) This tail error is unknown in general.

(2) A good estimation is necessary in order to get good accuracy (Yang & Karniadakis).

(3) Empirical estimation via cross validation (Yang & Karniadakis, Doostan & Owhadi,...) is expensive and wasteful.

(4) Solutions of $(\star)$ do not interpolate the data.

(5) All existing theoretical recovery guarantees (Rauhut & Ward, Yan, Guo & Xiu, Hampton & Doostan,...) assume the tail error is known.

Computations in infinite dimensions

Principle

Formulate the problem in infinite dimensions first and then discretize.

Examples: Bayesian inverse problems, computational spectral theory, numerical PDEs,...

Quoting A. Stuart: *The list of problems where it is beneficial to defer discretization to the very end of the algorithmic formulation is almost endless.* (Acta Numerica, 2010).

Most closely related to this talk:

- BA & Hansen, *A generalized sampling theorem for stable reconstructions in arbitrary bases*, J. Fourier Anal. Appl. 18(4):685–716 (2012).
- BA & Hansen, *Generalized sampling and infinite-dimensional compressed sensing,* Found. Comput. Math. (to appear) (2015).

A 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.

A 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.

Key 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.

Examples: 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.

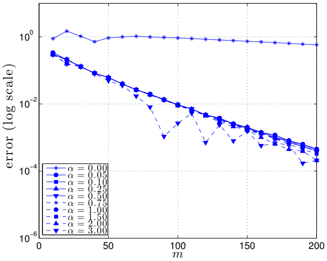
Rule-of-thumb

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

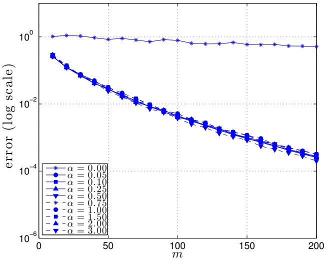
Outline

First experiment: deterministic samples in 1D

Example: deterministic equispaced samples and Chebyshev polynomials with weights $w_i = i^{\alpha}$.



$f(t) = \frac{1}{1+25t^2}$

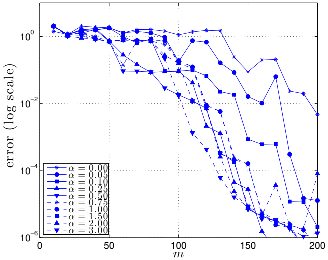


$f(t) = \frac{1}{35-34t}$

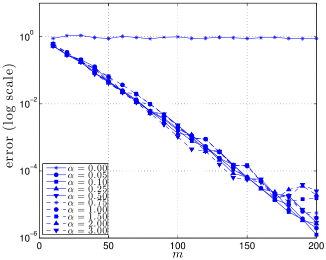
The error $\|f - \tilde{f}\|_{L^\infty}$ against m for different weights.

First experiment: deterministic samples in 1D

Example: deterministic equispaced samples and Chebyshev polynomials with weights $w_i = i^{\alpha}$.



$f(t) = \cos(30t)$

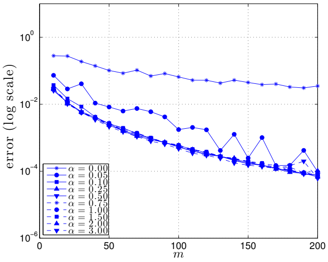


$f(t) = \frac{\cosh(30t^2)}{\cosh(30)}$

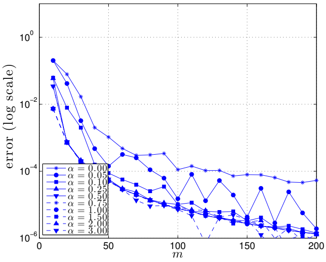
The error $\|f - \tilde{f}\|_{L^\infty}$ against m for different weights.

First experiment: deterministic samples in 1D

Example: deterministic equispaced samples and Chebyshev polynomials with weights $w_i = i^{\alpha}$.



$f(t) = \sqrt{1.01 + t}$



$f(t) = t^5 \log(t^2)$

The error $\|f - \tilde{f}\|_{L^\infty}$ against m for different weights.

Conclusions

(1) Unweighted ℓ^1 minimization ($\alpha = 0$) may not work in general.

Claim

This is due to an aliasing phenomenon in ℓ^1 minimization. In general, one needs the weights to satisfy

$$w_i / \|\phi_i\|_{L^\infty} \to \infty, \quad i \to \infty.$$

(2) Once $\alpha > 0$ there is no further gain from increasing α.

Remark: The use of weights has often been motivated by matching the decay rate of polynomial coefficients.

- See Peng, Hampton & Doostan, Rauhut & Ward.

Conclusions

(1) Unweighted ℓ^1 minimization ($\alpha = 0$) may not work in general.

Claim

This is due to an aliasing phenomenon in ℓ^1 minimization. In general, one needs the weights to satisfy

$$w_i / \|\phi_i\|_{L^\infty} \to \infty, \quad i \to \infty.$$

(2) Once $\alpha > 0$ there is no further gain from increasing α.

Remark: The use of weights has often been motivated by matching the decay rate of polynomial coefficients.

- See Peng, Hampton & Doostan, Rauhut & Ward.

Example: Fourier basis

Consider $D = (-1, 1)$, $\nu(t) = 1/2$ and the Fourier basis:

$$\phi_j(t) = \mathrm{e}^{\mathrm{i} j \pi t}, \quad j \in \mathbb{Z}.$$

In this case, $\|\phi_j\|_{L^\infty} = 1$, so we consider unweighted ℓ^1 minimization:

$$\inf_{z \in \ell^1(\mathbb{N})} \|z\|_1 \text{ subject to } Uz = y. \qquad (\star)$$

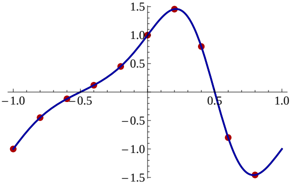
Proposition (Aliasing phenomenon)

Suppose that there exists a $P \in \mathbb{Z}$ such that

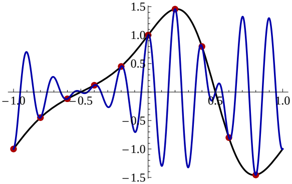
$$t_i P \in \mathbb{Z}, \quad i = 1, \ldots, m.$$

If $\hat{x}$ is a solution of $(\star)$ then so is every shift of $\hat{x}$ by a multiple of $2P$.

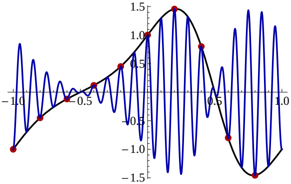
Aliased solutions are poor approximations



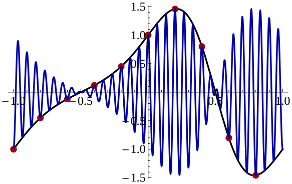
Good minimizer $\hat{x}$



Bad minimizer $(2P)$



Bad minimizer $(4P)$



Bad minimizer $(6P)$

Besides the first, none of these minimizers approximate f to any accuracy.

Example: Fourier basis

Now consider the weighted problem

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

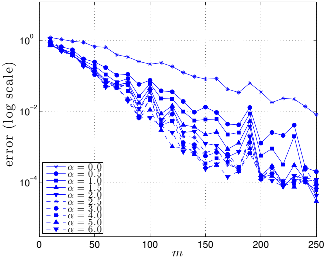
with any monotonically growing weights $w_i \to \infty$ as $|i| \to \infty$. Aliased solutions will generally no longer be minimizers, since they have larger weighted norm.

Summary

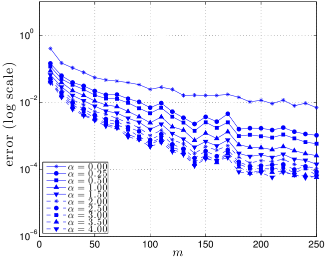
Growing weights regularize the minimization problem by removing (bad) aliased solutions.

Experiment 2: random sampling in 1D and 2D

Example: random samples from Chebyshev (C) or Uniform (U) measures with (tensor) Chebyshev (C) or Legendre (L) polynomials.



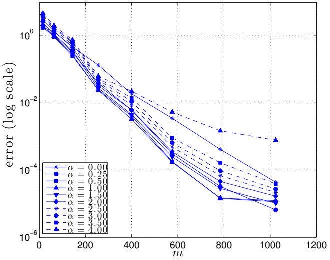
1D LC, $f(t) = \frac{1+3t}{1+50t^2}$



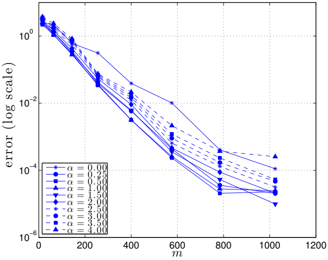
1D LU, $f(t) = \sqrt{1.05 + t}$

Experiment 2: random sampling in 1D and 2D

Example: random samples from Chebyshev (C) or Uniform (U) measures with (tensor) Chebyshev (C) or Legendre (L) polynomials.



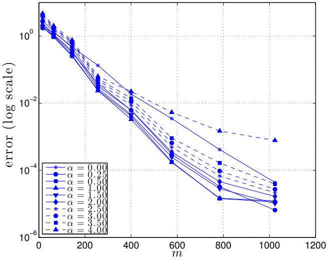
2D CC, $f(t_1, t_2) = \sin(\exp(2t_1 t_2))$



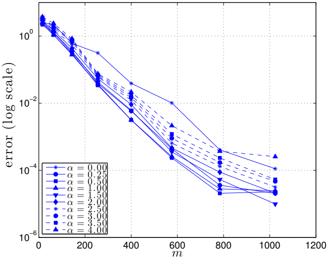
2D CC, $f(t_1, t_2) = \sin(\exp(2t_1 t_2))$

Experiment 2: random sampling in 1D and 2D

Example: random samples from Chebyshev (C) or Uniform (U) measures with (tensor) Chebyshev (C) or Legendre (L) polynomials.



2D CC, $f(t_1, t_2) = \sin(\exp(2t_1 t_2))$



2D CC, $f(t_1, t_2) = \sin(\exp(2t_1 t_2))$

Conclusion: Although convergence occurs in the unweighted case, weights appear to offer some moderate benefits.

Outline

Worst-case scenario

Data points $T = \{t_i\}_{i=1}^m$ are fixed, deterministic and unstructured.

Scattered data approximation: Quantify data in terms of the density

$$h = \sup_{t \in D} \min_{i=1,\ldots,m} |t - t_i|.$$

(also known as the fill distance).

Goal

We cannot expect to achieve the best s-term approximation rate in this setting. Instead, we aim to show near-optimal linear approximation (first s term) rates as $h \to 0$.

Example result

Theorem (BA)

Let $D = (-1, 1)$ and consider a Jacobi (e.g. Legendre, Chebyshev, Gegenbauer) polynomial basis $\{\phi_i\}_{i\in\mathbb{N}}$. Suppose that $w_i \sim ci^{\alpha}$, $i \to \infty$, for $\alpha > 1$. Then

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

where $\|x - P_s x\|_{1,w}$ is the linear approximation error, provided

$$h^{-1} \gtrsim s^2 \log s.$$

This scaling is optimal, up to the log factor in s.

Remarks

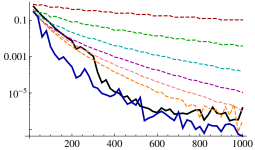
(1) Weighted ℓ^1 minimization achieves the optimal linear approximation rate as $h \to 0$, up to a log factor.

- Optimality is due to Platte, Trefethen & Kuijlaars.
- In particular, it is guaranteed to never perform worse than classical least-squares fitting.
- Note: this result extends to higher dimensions.

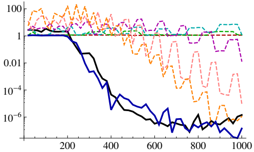
(2) The condition $h^{-1} \gtrsim s^2 \log s$ is independent of the weights used.

- Recall the earlier experiment.

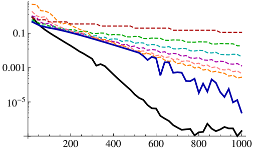
How well does ℓ^1 perform in bad scenarios?



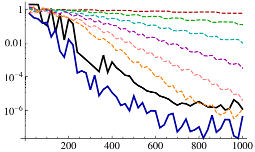
$f(t) = \frac{1}{50/49 - \sin(\pi t)}$



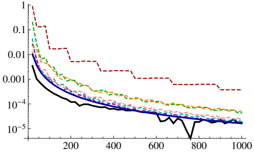
$f(t) = \sin(50t^2)$



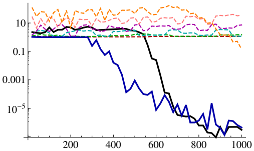
$f(t) = \frac{1}{1+50t^2}$



$f(t) = \frac{\cosh(100t^2)}{\cosh(100)}$



$f(t) = |t|^3$



$f(t) = \sin(80t)$

Black line is weighted ℓ^1 minimization. Dashed lines are least squares with $M = c\sqrt{m}$ and $c = 0.5, 1.0, 1.5, 2, 2.5, 3.0$. Blue line is oracle least squares based on choosing the aspect ratio to minimize the error for a given m and f. Random noise of magnitude 10^{-8} was added to the data.

Ideal scenario

Data points: The points $T = \{t_1, \ldots, t_m\}$ will now be drawn randomly from the orthogonality measure $\nu(t)$ of the functions $\{\phi_i\}_{i\in\mathbb{N}}$.

Goal

Show near-optimal s-term approximation rates.

However.....weighted sparsity

We solve a weighted ℓ^1 minimization problem, so it is more natural to consider weighted cardinality:

$$|\Delta|_w := \sum_{i \in \Delta} w_i^2,$$

and the weighted s-term approximation error

$$\sigma_{s,w}(x) = \min \left\{ \|x - P_\Delta x\|_{1,w} : |\Delta|_w \leq s \right\}.$$

Note that $s \in (0, \infty)$ in the weighted setup.

- See Rauhut & Ward (2014).

Weighted sparsity recovery guarantee

Theorem (BA)

Let $w = \{w_i\}_{i \in \mathbb{N}}$ be weights with $w_i \geq \|\phi_i\|_{L^\infty}$ and $\Delta \subseteq \{1, \ldots, K\}$. Let $x \in \ell^1_w(\mathbb{N})$ and suppose that $t_1, \ldots, t_m$ are drawn independently from ν. If $\hat{x}$ is any minimizer, then

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

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

$$m \gtrsim |\Delta|_w \cdot \log(\epsilon^{-1}) \cdot \log(2N\sqrt{|\Delta|_w}).$$

Earlier work: Rauhut & Ward (2014).

- Require knowledge of the tail bound δ.
- Provide uniform recovery guarantees (with additional log factors).

Is this good enough?

Let $w_i = i^\alpha$ and suppose that f is such that

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

This is reasonable for oscillatory functions, for example.

Question: How many samples m do we need to recover f exactly?

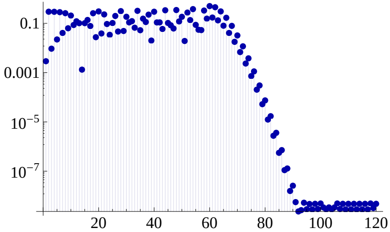
- According to the previous theorem, we set $\Delta = \{1, \ldots, k\}$.
- Then we need $m \gtrsim |\Delta|_w \times$ log factors, i.e.

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

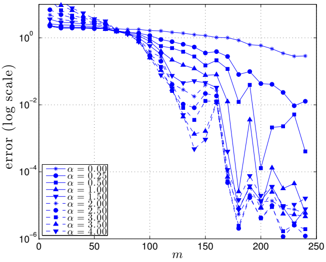
- This estimate deteriorates with increasing α.

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



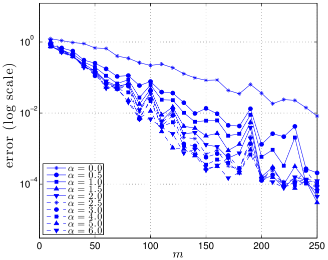
Error versus m

Infinite expansions and weighted sparsity

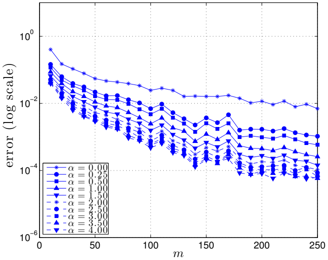
Proposition

Let $w_i = i^\alpha$ and suppose that $|x_j| = j^{-\alpha-\beta-1}$ for some $\beta > 0$. Then $\sigma_{s,w}(x) = \mathcal{O}\left(s^{-\frac{\beta}{2\alpha+1}}\right)$ as $s \to \infty$.

Thus the predicted convergence rate of the approximation in terms of s (equivalently, m) deteriorates with increasing α.



1D LC, $f(t) = \frac{1+3t}{1+50t^2}$



1D LU, $f(t) = \sqrt{1.05 + t}$

An improved recovery guarantee

Theorem (BA)

Let $w = \{w_i\}_{i\in\mathbb{N}}$ *be weights,* $x \in \ell^1_w(\mathbb{N})$ *and* $\Delta \subseteq \{1,\dots,K\}$ *be such that* $\min_{i\in\{1,\dots,K\}\setminus\Delta}\{w_i\} \geq 1$*. Let* $t_1,\dots,t_m$ *be drawn independently from* ν*. Then*

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

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

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

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

Note: All constants in the $\lesssim$ and $\gtrsim$ are independent of the weights w_i.

Consequences

Consider the main estimate:

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

Linear model: Let $\Delta = \{1, \dots, k\}$. Suppose that $u_i = \mathcal{O}(i^\gamma)$ and $w_i = \mathcal{O}(i^\alpha)$ 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.
- It addresses both examples considered previously.

Towards establishing the benefits of weights

The case $w_i = 1$. We get the estimate

$$m \gtrsim \left(|\Delta|_u + \max_{1 \le i \le K} \{u_i^2\} |\Delta| \right) \cdot L. \tag{1}$$

The case $w_i = u_i$. We get the estimate

$$m \gtrsim |\Delta|_u \cdot L. \tag{2}$$

Note: In general, the estimate (2) is no worse than (1). Hence, it makes sense to use weights with w_i at least as large as u_i.

Examples: the benefits of weights $w_i = u_i$

Example 1: Consider Legendre polynomials with points drawn from the uniform measure.

- If $w_i = 1$ then $m \gtrsim 3^{\min\{p,d\}} \cdot s \cdot L$, where $s = |\Delta|$, provided the index set $\{1, \ldots, K\}$ corresponds to a total degree space of degree p.
- If $w_i = u_i$ then $m \gtrsim s^2 \cdot L$ provided Δ is a lower set.
- Note that s^2 is sharp.

Example 2: Consider Chebyshev polynomials with points drawn from the Chebyshev measure. Then

- If $w_i = 1$ then $m \gtrsim 2^{\min\{p,d\}} \cdot s \cdot L$, provided the index set $\{1, \ldots, K\}$ corresponds to a total degree space of degree p.
- If $w_i = u_i$ then $m \gtrsim s^{\log(3)/\log(2)} \cdot L$ provided Δ is a lower set.

BA thanks A. Chkifa, H. Tran, C. Webster & G. Zhang for the observations about lower sets.

Examples: the benefits of weights $w_i = u_i$

Example 1: Consider Legendre polynomials with points drawn from the uniform measure.

- If $w_i = 1$ then $m \gtrsim 3^{\min\{p,d\}} \cdot s \cdot L$, where $s = |\Delta|$, provided the index set $\{1, \ldots, K\}$ corresponds to a total degree space of degree p.
- If $w_i = u_i$ then $m \gtrsim s^2 \cdot L$ provided Δ is a lower set.
- Note that s^2 is sharp.

Example 2: Consider Chebyshev polynomials with points drawn from the Chebyshev measure. Then

- If $w_i = 1$ then $m \gtrsim 2^{\min\{p,d\}} \cdot s \cdot L$, provided the index set $\{1, \ldots, K\}$ corresponds to a total degree space of degree p.
- If $w_i = u_i$ then $m \gtrsim s^{\log(3)/\log(2)} \cdot L$ provided Δ is a lower set.

BA thanks A. Chkifa, H. Tran, C. Webster & G. Zhang for the observations about lower sets.

Towards establishing the benefits of weights

Related work:

- Peng, Hampton & Doostan, Yang & Karniadakis: Empirical improvements for weights based on prior support information.
- Rauhut & Ward: Error is bounded in a stronger norm. However, guarantee deteriorates with w_i.
- Bah & Ward: Sample complexity of weighted minimization. But consider weighted cardinality.

Support estimation

Corollary (BA)

Let $u_i = 1$. Assume x is 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

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

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.
- Caveat: comparing sufficient conditions.

Related work:

- Friedlander et al., Yu & Baek (random Gaussian measurements).

Thanks!

For more info, see the paper:

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

Also, coming later in the summer:

B. Adcock, *Infinite-dimensional compressed sensing and function interpolation*, in preparation (2015).