Sparse Legendre expansions via $\ell_1$ minimization

Rachel Ward, Courant Institute, NYU
Joint work with Holger Rauhut, Hausdorff Center for Mathematics, Bonn, Germany.

June 8, 2010
Outline

Sparse recovery for bounded orthonormal systems

Sparse recovery for (certain) unbounded orthonormal systems

From compressive sensing to function approximation

Applications
Outline

Sparse recovery for bounded orthonormal systems

Sparse recovery for (certain) unbounded orthonormal systems

From compressive sensing to function approximation

Applications
Problem set-up

▶ Consider the general problem: reconstruct an $s$-sparse vector $x \in \mathbb{C}^N$ or $x \in \mathbb{R}^N$ (or a compressible vector) from its vector of $m$ measurements $y = Ax$

▶ Suppose that the measurements correspond to underdetermined system: $s < m < N$, so that the system $y = Ax$ has infinitely many solutions. We want a sparse solution.

▶ Preferably we would like to have a fast algorithm that performs the sparse reconstruction.
**Algorithms for sparse recovery**

\[ \ell_0\text{-minimization: } \|x\|_0 := |\text{supp } x|, \]

\[ \min_{x \in \mathbb{C}^N} \|x\|_0 \quad \text{subject to} \quad Ax = y. \]

**Problem:** \( \ell_0\)-minimization is NP hard.

\( \ell_1 \) minimization:

\[ \min_x \|x\|_1 = \sum_{j=1}^{N} |x_j| \quad \text{subject to} \quad Ax = y \]

This is the convex relaxation of \( \ell_0\)-minimization problem, can be solved efficiently.
A sufficient condition for sparse recovery

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

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2$$

for all $s$-sparse $x \in \mathbb{C}^N$.

(Candes, Romberg, Tao 2005, Foucart/Lai 2009): If $A \in \mathbb{R}^{m \times N}$ satisfies RIP of order $2s$, and if $\delta_{2s} \leq .46...$, then any vector $x \in \mathbb{R}^N$ may be approximated from $y = Ax + \xi$ with $\|\xi\|_2 \leq \epsilon$ by

$$\hat{x} = \min_z \|z\|_1 \quad \text{subject to} \quad \|Az - y\|_2 \leq \epsilon$$

up to error

$$\|x - \hat{x}\|_2 \leq C_1 \frac{\|x - x_s\|_1}{\sqrt{s}} + C_2\epsilon.$$
Which matrices satisfy the restricted isometry property?

If $A \in \mathbb{R}^{m \times N}$ is a Gaussian or Bernoulli random matrix,

$$m = O(s \ln(N/s))$$

measurements ensure that $A$ satisfy the restricted isometry property of order $s$ with high probability.

In contrast, all available deterministic constructions run into quadratic bottleneck: require $m = O(s^2 \log N)$ measurements.

Compromise between completely random and deterministic: structured random matrices.
Structured random matrices: the random partial Fourier matrix

Consider the discrete Fourier matrix $\mathcal{F} \in \mathbb{C}^{N \times N}$ with entries

$$F_{\ell,k} = \frac{1}{\sqrt{N}} e^{2\pi i \ell k / N}, \quad \ell, k = 1, \ldots, N,$$

Randomly subsample $m < N$ rows from $\mathcal{F}$ to get the $m \times N$ random partial Fourier matrix. The measurements $y_\ell = (\mathcal{F} x)_\ell$ are a random subsample of the DFT of $x$.

Theorem (Candes, Romberg, Tao 2005; Rudelson, Vershynin 2008)

If $m \geq O(s \ln^4(N))$, then with high probability, the $m \times N$ random partial Fourier matrix satisfies the restricted isometry property of order $s$. 

The random partial Fourier matrix: Function approximation interpretation

Random partial Fourier measurements correspond to samples of a trigonometric polynomial:

\[ y_\ell = (\mathcal{F}x)_\ell = g_x(t_\ell), \quad t_\ell = \frac{2\pi j_\ell}{N}, \]

where \( g_x(t) := \sum_{k=0}^{N-1} x_k e^{ikt} \).

We say that \( g_x(t) \) is an \( s \)-sparse trigonometric polynomial of degree \( N - 1 \) if the coefficient vector \( (x_k)_{k=0}^{N-1} \) is \( s \)-sparse.

RIP of the random partial Fourier matrix means that all \( s \)-sparse trigonometric polynomials of degree \( N \) can be recovered from their values on a fixed subset of \( m = O(s \log^4 N) \) measurements of the form \( t_\ell = \frac{2\pi \ell}{N} \).
Outline

Sparse recovery for bounded orthonormal systems

Sparse recovery for (certain) unbounded orthonormal systems

From compressive sensing to function approximation

Applications
Sparse recovery for more general orthonormal systems?

We may consider the recovery of functions

$$f(u) = \sum_{k=0}^{N-1} x_k \psi_k(u), \quad u \in \mathcal{D} \subset \mathbb{R}^d$$

that are \textit{sparse} relative to a function system $\psi_1, \ldots, \psi_N : \mathcal{D} \to \mathbb{C}$ that is \textit{orthonormal} with respect to measure $\nu$ on $\mathcal{D}$, from a number $m < N$ of point samples

$$y_\ell = f(u_\ell) = \sum_{k=0}^{N} x_k \psi_k(u_\ell), \quad u_1, \ldots, u_m \sim \nu.$$

where $m$ is proportional to the sparsity level $s$.

Whether sparse recovery is possible depends on the structure of the sampling matrix $A_{\ell,k} = \psi_k(u_\ell)$ associated to the orthonormal system.
Sparse recovery for general bounded orthonormal systems

**Sampling matrix:** $A_{\ell,k} = \psi_k(u_{\ell}), \quad \ell \in [m], k \in [N]$

**Theorem (Rauhut ’09)**

Suppose that $\|\psi_k\|_\infty \leq K$ for all $k \in [N]$. Let $A \in \mathbb{C}^{m \times N}$ be the sampling matrix associated to the bounded orthonormal system $\{\psi_j\}_{j=0}^{N-1}$ with sampling points $u_1, \ldots, u_m \in \mathcal{D}$ chosen i.i.d. from the orthogonalization measure $\nu$. If $m \geq O(K^2 s \ln^4(N))$, then with probability exceeding $1 - N^{-\log^3 N}$, the matrix $\frac{1}{\sqrt{m}} A$ satisfies RIP of order $s$.

**Examples of uniformly bounded orthonormal systems:** The complex exponential basis (with uniform orthogonalization measure) and Chebyshev polynomial basis, orthonormal on $[-1, 1]$ w.r.t. the Chebyshev measure.
The Legendre polynomials $P_\ell$ are orthonormal on $D = [-1, 1]$ with respect to Lebesgue measure.

- They are generated via Gram-Schmidt orthogonalization on the monomial system $1, x, x^2, \ldots$
- They are used in schemes for solving several classical PDEs, such as Laplace’s equation
- **Fast Legendre polynomial transform:** requires only $O(N \log N)$ arithmetic operations
The Legendre polynomials satisfy \( \| P_\ell \|_\infty = \sqrt{2\ell + 1} \), so 
\[
K = \sup_{0 \leq \ell \leq N - 1} \| P_\ell \|_\infty = \sqrt{2N - 1}.
\]

From existing theory: \( s \)-sparse polynomials of the form 
\[
f(t) = \sum_{k=0}^{N} x_k P_k(t)
\]
may be recovered from its values at 
\[
m = O(sK^2 \ln^4(N)) = O(sN \ln^4 N)
\]
sampling points.

this is trivial!
The Legendre polynomials satisfy $\|P_\ell\|_\infty = \sqrt{2\ell + 1}$, so $K = \sup_{0 \leq \ell \leq N - 1} \|P_\ell\|_\infty = \sqrt{2N - 1}$.

From existing theory: $s$-sparse polynomials of the form $f(t) = \sum_{k=0}^{N} x_k P_k(t)$ may be recovered from its values at $m = O(sK^2 \ln^4(N)) = O(sN \ln^4 N)$ sampling points.

this is trivial!
Theorem (Rauhut, W '10)

Let \( N \) and \( s \) be given. Suppose that \( m = O(s \log^4 N) \) sampling points \( t_1, \ldots, t_m \) are drawn from the Chebyshev measure \( d\nu = \pi^{-1}(1 - t^2)^{-1/2} dt \) on \([-1, 1]\). Then with probability exceeding \( 1 - N^{-\log^3 N} \), the following holds for all polynomials \( f(t) = \sum_{k=0}^{N-1} x_k P_k(t) \):

Suppose that noisy sample values \((f(t_1) + \eta_1, \ldots, f(t_m) + \eta_m)\) are observed, and \( \|\eta\|_2 \leq \sqrt{m\epsilon} \). Then the coefficient vector \( x = (x_0, x_1, \ldots, x_{N-1}) \) is approximated by

\[
\hat{x} := \min_z \|z\|_1 \quad \text{subject to} \quad \|Az - y\|_2 \leq \epsilon
\]

up to error

\[
\|x - \hat{x}\|_2 \leq C_1 \frac{\|x - x_s\|_1}{\sqrt{s}} + C_2 \epsilon.
\]
Numerical Example

A sparse Legendre polynomial with sparsity $s = 5$, maximal degree $N = 80$, and $n = 20$ i.i.d. sampling points from the Chebyshev measure. Reconstruction by $\ell_1$-minimization is exact!
The same sparse Legendre polynomial (black) subject to noisy random measurements, and the stable approximation obtained by $\ell_1$-minimization.
Key idea in proof: Premultiplication

The Legendre polynomials grow uniformly as they approach the endpoints:

\[ |P_n(t)| < 2\pi^{-1/2}(1 - t^2)^{-1/4}, \quad -1 \leq t \leq 1; \]

(Constant due to S. Bernstein)
Key idea in proof: Premultiplication

The Legendre polynomials grow uniformly as they approach the endpoints:

$$| P_n(t) | < 2\pi^{-1/2} 2\pi^{-1/2} (1 - t^2)^{-1/4}, \quad -1 \leq t \leq 1;$$

(Constant due to S. Bernstein)

So the functions $Q_n(t) = \sqrt{\frac{\pi}{2}} (1 - t^2)^{1/4} P_n(t)$ form a uniformly bounded system (and $\| Q_n \|_\infty \leq \sqrt{2}$).

Also, the $Q_n$’s are orthonormal with respect to the Chebyshev measure $d\nu = \pi^{-1} (1 - t^2)^{-1/2} dt$ on $[-1, 1]$:

$$\int_{-1}^{1} Q_n(t) Q_m(t) \pi^{-1} (1 - t^2)^{-1/2} dt = \int_{-1}^{1} P_n(t) P_m(t) dt = \delta_{n,m}.$$
Outline

Sparse recovery for bounded orthonormal systems

Sparse recovery for (certain) unbounded orthonormal systems

From compressive sensing to function approximation

Applications
Universality of the Chebyshev measure

Consider a probability measure $\nu(t)dt$ on $[-1, 1]$.

(Szego): If $\nu(t)$ satisfies a mild continuity condition, then the polynomials $p_n^{\nu}$ that are orthonormal w.r.t. $\nu$ will still satisfy a uniform growth condition:

$$|p_n^{\nu}(t)| \leq C_\nu (1 - t^2)^{-1/4} \nu(t)^{-1/2},$$

as before, $q_n^{\nu}(t) = (1 - t^2)^{1/4} \nu(t)^{1/2} p_n^{\nu}(t)$ is a uniformly bounded system and orthonormal w.r.t. Chebyshev measure $(1 - t^2)^{-1/2} dt$. 
From compressive sensing to function approximation

Consider the weighted norm on continuous functions in \([-1, 1]\),

\[
\|f\|_{\infty, w} = \sup_{t \in [-1, 1]} |f(t)| w(t), \quad w(t) = (1 - t^2)^{1/4} \sqrt{\nu(t)},
\]

and the error

\[
\sigma_{N,s}(f)_{\infty, w} = \inf \{ \|f - \sum_{k=0}^{N-1} x_k p_k^x\|_{\infty, w} : (x_k) \in \mathbb{R}^N, \|x\|_0 \leq s \}
\]

Theorem (Rauhut and W, '10)

Let \(N, m, s\) be given with \(m = O(s \log^4 N)\). Then there exist sampling points \(t_1, \ldots, t_m\) (chosen according to Chebyshev distribution) and an efficient reconstruction procedure (\(\ell_1\) minimization) such that for any continuous function \(f\), the polynomial \(P\) of degree at most \(N - 1\) reconstructed from \(f(t_1), \ldots, f(t_m)\) satisfies

\[
\|f - P\|_{\infty, w} \leq C_w \sqrt{s} \cdot \sigma_{N,s}(f)_{\infty, w}.
\]
From compressive sensing to function approximation

Consider the weighted norm on continuous functions in $[-1, 1]$,

$$
\|f\|_{\infty,w} = \sup_{t \in [-1, 1]} |f(t)| w(t), \quad w(t) = (1 - t^2)^{1/4} \sqrt{\nu(t)},
$$

and the error

$$
\sigma_{N,s}(f)_{\infty,w} = \inf \{ \| f - \sum_{k=0}^{N-1} x_k \varphi_k \|_{\infty,w} : (x_k) \in \mathbb{R}^N, \| x \|_0 \leq s \}
$$

**Theorem (Rauhut and W, ’10)**

Let $N, m, s$ be given with $m = O(s \log^4 N)$. Then there exist sampling points $t_1, \ldots, t_m$ (chosen according to Chebyshev distribution) and an efficient reconstruction procedure ($\ell_1$ minimization) such that for any continuous function $f$, the polynomial $P$ of degree at most $N - 1$ reconstructed from $f(t_1), \ldots, f(t_m)$ satisfies

$$
\| f - P \|_{\infty,w} \leq C_w \sqrt{s} \cdot \sigma_{N,s}(f)_{\infty,w}.
$$
Extension to spherical harmonics

\[ Y_{\ell}^m(\phi, \theta) = C_{\ell,m} P_{\ell}^m(\cos(\theta)) e^{im\phi}, \quad -\ell \leq m \leq \ell, \quad \ell = 0, 1, \ldots \]

\( P_{\ell}^m \)'s are the associated Legendre functions.

Fast Spherical harmonic transform: requires only \( O(N(\log N)^2) \) arithmetic operations
Extension to spherical harmonics

The associated Legendre functions $P^m_\ell$ may be written in terms of orthonormal polynomials with respect to the positive weights

$$\nu_\alpha(x) = (1 - x^2)^\alpha, \quad \alpha = 0, 1, ..., n.$$

Krasikov ’08: 

$$(1 - x^2)^{1/4} \nu_\alpha(x)^{1/2} |p_\alpha^n(x)| \leq C_\alpha \leq O(\alpha^{1/4})$$

from this one can derive ...

Corollary

Let $N, m, s$ be given with $m = O(sN^{1/4} \log^4 N)$. Then there exist sampling points $t_1, ..., t_m$ (chosen according to the spherical Chebyshev distribution) and an efficient reconstruction procedure ($\ell_1$ minimization) such that any function on the sphere which is an $s$-sparse expansion in the first $N$ spherical harmonic basis functions may be exactly recovered from these sampling points.
Outline

Sparse recovery for bounded orthonormal systems

Sparse recovery for (certain) unbounded orthonormal systems

From compressive sensing to function approximation

Applications
Application: Cosmic Microwave Background Radiation (CMB) map

\[ T(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell}^{m}(\theta, \phi) \] where the \( Y \)'s are the spherical harmonics.

- Red band: measurements are corrupted, interference with galactic foreground signal.
CMB map is compressible in spherical harmonics

Source: Lyman Page, http://ophelia.princeton.edu/

Consider the coefficient vector \( \mathbf{a} = (a_{\ell,m}) \) in
\[
T(\theta, \phi) \approx \sum_{\ell=0}^{n} \sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell}^{m}(\theta, \phi).
\]
This vector is predicted and observed to be compressible in the spherical harmonic basis.

\[ \| \mathbf{a} - \mathbf{a}_s \|_2 / \| \mathbf{a} \|_2, \ s = 1, \ldots, n^2 \]
CMB map is compressible in spherical harmonics

(J. Starck et. al., '08):\(^1\) Propose full-sky CMB map inpainting from partial CMB measurements \(T(\theta_j, \phi_k) = m_{j,k}\). Obtain the coefficients \(\mathbf{a} = (a_{\ell,m})\) by solving the \(\ell_1\) minimization problem,

\[
\mathbf{a} = \arg\min \sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} |z_{\ell,m}| \quad s.t. \quad \sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} z_{\ell,m} Y_{\ell,m}(\theta_j, \phi_k) = m_{j,k}
\]

Final remarks

Our results provide some theoretical justification for the good inpainting results obtained for the CMB map.

Open questions:

1. Sparse recovery in spherical harmonic basis: can we get better recovery guarantees?
2. Can we generalize these results to functions sparse in bases that are eigenfunction solutions to a more general class of PDE.
Final remarks

Our results provide some theoretical justification for the good inpainting results obtained for the CMB map.

Open questions:

1. Sparse recovery in spherical harmonic basis: can we get better recovery guarantees?
2. Can we generalize these results to functions sparse in bases that are eigenfunction solutions to a more general class of PDE

THANK YOU