Infinite-dimensional compressed sensing

Ben Adcock
Department of Mathematics
Purdue University (Simon Fraser University after August 1)

Joint work with Anders Hansen, Clarice Poon and Bogdan Roman (Cambridge)

Outline

Outline

Setup

Consider the linear system

$$y = \Psi^* x,$$

where

- $x = (x_1, x_2, \ldots, x_N)^\top \in \mathbb{C}^N$ is the unknown object,
- $y = (y_1, y_2, \ldots, y_N)^\top \in \mathbb{C}^N$ is the vector of measurements,
- $\Psi \in \mathbb{C}^{N \times N}$ is an measurement matrix (assumed to be an isometry).

Typically, we can access only a small subset of measurements

$$\{y_j,\ j \in \Omega\},$$

where $\Omega \subseteq \{1, 2, \ldots, N\}$, $|\Omega| = m \ll N$.

Problem: Recover x from the underdetermined system $P_\Omega \Psi^* w = P_\Omega y$, where P_Ω is the projection onto indices in Ω.

Compressed sensing (CS)

Under appropriate conditions on x, Ψ and Ω, we can recover x from $P_\Omega y$ in a stable and robust manner with efficient numerical algorithms.

- Origins: Candès, Romberg & Tao (2006), Donoho (2006).
- Since then, the subject of thousands of papers, dozens of survey articles, and one textbook (Foucart & Rauhut, Birkhauser, 2013).
- Applications: medical imaging, seismology, analog-to-digital conversion, microscopy, radar, sonar, communications,...

Key principles: sparsity, incoherence, uniform random subsampling

Principles

Sparsity: There exists an isometry $\Phi \in \mathbb{C}^{N \times N}$ (e.g. a wavelet transform) such that $x = \Phi z$, where the vector z is s-sparse:

$$|\{j : z_j \neq 0\}| \leq s.$$

Incoherence: The coherence of $U = (u_{ij}) = \Psi^* \Phi$ is

$$\mu(U) = \max_{i,j=1,\ldots,N} |u_{ij}|^2 \in [N^{-1}, 1]$$

The pair (Ψ, Φ) is incoherent if $\mu(U) \leq c/N$.

Uniform random subsampling: The index set

$$\Omega \subseteq \{1, \ldots, N\}, \quad |\Omega| = m,$$

is chosen uniformly at random.

Principles

Sparsity: There exists an isometry $\Phi \in \mathbb{C}^{N\times N}$ (e.g. a wavelet transform) such that $x = \Phi z$, where the vector z is s-sparse:

$$|\{j : z_j \neq 0\}| \leq s.$$

Incoherence: The coherence of $U = (u_{ij}) = \Psi^* \Phi$ is

$$\mu(U) = \max_{i,j=1,\dots,N} |u_{ij}|^2 \in [N^{-1}, 1]$$

The pair (Ψ, Φ) is incoherent if $\mu(U) \leq c/N$.

Uniform random subsampling: The index set

$$\Omega \subseteq \{1, \dots, N\}, \quad |\Omega| = m,$$

is chosen uniformly at random.

Principles

Sparsity: There exists an isometry $\Phi \in \mathbb{C}^{N\times N}$ (e.g. a wavelet transform) such that $x = \Phi z$, where the vector z is s-sparse:

$$|\{j : z_j \neq 0\}| \leq s.$$

Incoherence: The coherence of $U = (u_{ij}) = \Psi^*\Phi$ is

$$\mu(U) = \max_{i,j=1,\dots,N} |u_{ij}|^2 \in [N^{-1}, 1]$$

The pair (Ψ, Φ) is incoherent if $\mu(U) \leq c/N$.

Uniform random subsampling: The index set

$$\Omega \subseteq \{1, \dots, N\}, \quad |\Omega| = m,$$

is chosen uniformly at random.

Reconstruction algorithm

Typically, one solves the convex optimization problem

$$\min_{w \in \mathbb{C}^N} \|\Phi^* w\|_{l^1} \text{ subject to } \|P_\Omega \Psi^* w - y\|_{l^2} \leq \delta,$$

where $\|w\|_{l^1} = |w_1| + |w_2| + \ldots + |w_N|$ is the l^1-norm, and

$$y = P_\Omega \Psi^* x + e$$

are noisy measurements with $\|e\|_{l^2} \leq \delta$.

- Other approaches: greedy methods (e.g. OMP, CoSaMP), thresholding methods (e.g. IHT, HTP), message passing algorithms,....

Compressed sensing theorem

Theorem (see Candès & Plan, Adcock & Hansen)

Let $0 < \epsilon \leq \mathrm{e}^{-1}$ *and suppose that*

$$m \gtrsim s \cdot N \cdot \mu(U) \cdot \log(\epsilon^{-1}) \cdot \log N,$$

for some universal constant C*. Then with probability greater than* $1 - \epsilon$ *any minimizer* $\hat{x}$ *of the problem*

$$\min_{w \in \mathbb{C}^N} \|\Phi^* w\|_{l^1} \text{ subject to } \|P_\Omega \Psi^* w - y\|_{l^2} \leq \delta \sqrt{N/m},$$

satisfies

$$\|x - \hat{x}\|_{l^2} \lesssim \sigma_s(x) + \sqrt{s}\delta,$$

where $\sigma_s(x) = \min\{\|\Phi^* x - z\|_{l^1} : z \text{ is } s\text{-sparse}\}$.

$\Rightarrow$ If U is incoherent, i.e. $\mu(U) = \mathcal{O}\left(N^{-1}\right)$, then $m \approx s \log N \ll N$.

Outline

Recovery from the Fourier transform

Applications: Magnetic Resonance Imaging (MRI), X-ray Computed Tomography, Electron Microscopy, Seismology, Radio interferometry,....

Mathematically, all these problems can be reduced (possibly via the Fourier slice theorem) to the following:

Given $\{\hat{f}(\omega) : \omega \in \Omega\}$, recover the image f.

Here $\Omega \subseteq \hat{\mathbb{R}}^d$ is a finite set and $\hat{f}$ is the Fourier transform (FT).

However, f is a function (not a vector) and $\hat{f}$ is its continuous FT.

Standard CS approach

We approximate $f \approx x$ on a discrete grid, and let

- Ψ be the DFT,
- Φ be a discrete wavelet transform.

However, this setup is a discretization of the continuous model:

continuous FT $\approx$ discrete FT $\Rightarrow$ measurements mismatch

Issues:

1. If measurements are simulated via the DFT $\Rightarrow$ inverse crime.
   - In MRI, see Guerquin–Kern, Häberlin, Pruessmann & Unser (2012)
2. If measurements are simulated via the continuous FT. Minimization problem has no sparse solution $\Rightarrow$ poor reconstructions.

Standard CS approach

We approximate $f \approx x$ on a discrete grid, and let

- Ψ be the DFT,
- Φ be a discrete wavelet transform.

However, this setup is a discretization of the continuous model:

continuous FT $\approx$ discrete FT $\Rightarrow$ measurements mismatch

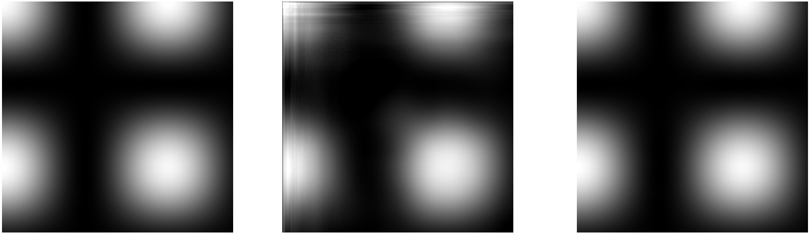
Issues:

1. If measurements are simulated via the DFT $\Rightarrow$ inverse crime.
   - In MRI, see Guerquin–Kern, Häberlin, Pruessmann & Unser (2012)

2. If measurements are simulated via the continuous FT. Minimization problem has no sparse solution $\Rightarrow$ poor reconstructions.

Poor reconstructions with standard CS

Example: Electron microscopy, $f(x,y) = \mathrm{e}^{-x-y}\cos^2(\frac{17\pi x}{2})\cos^2(\frac{17\pi y}{2})$, 6.15% Fourier measurements.



Original (zoomed) Fin. dim. CS, Err $= 12.7\%$ Inf. dim. CS, Err $= 0.6\%$

Infinite-dimensional formulation

Consider two orthonormal bases of a Hilbert space H (e.g. $\mathrm{L}^2(0,1)^d$):

- Sampling basis: $\{\psi_j\}_{j\in\mathbb{N}}$, e.g. the Fourier basis $\psi_j(x) = \exp(2\pi \mathrm{i} j \cdot x)$.
- Sparsity basis: $\{\phi_j\}_{j\in\mathbb{N}}$, e.g. a wavelet basis.

Let $f \in \mathrm{H}$ be the object to recover. Write

- $x_j = \langle f, \phi_j \rangle$ for the coefficients of f, i.e. $f = \sum_{j\in\mathbb{N}} x_j \phi_j$,
- $y_j = \langle f, \psi_j \rangle$ for the measurements of f.

Define the infinite matrix $U = \{\langle \phi_j, \psi_i \rangle\}_{i,j\in\mathbb{N}} \in B(\ell^2(\mathbb{N}))$ and note that

$$Ux = y.$$

New concepts

Uniform random subsampling: It is meaningless to draw $\Omega \subseteq \mathbb{N}$, $|\Omega| = m$ uniformly at random. It is also infeasible in practice due to bandwidth limitations. Hence, we fix the sampling bandwidth N and let

$$\Omega \subseteq \{1, \ldots, N\}, \quad |\Omega| = m,$$

be drawn uniformly at random.

Sparsity: Given finite sampling bandwidth, we cannot expect to recover any s-sparse infinite vector x stably. Let M be the sparsity bandwidth, and suppose that x is (s, M)-sparse:

$$|\{j = 1, \ldots, M : x_j \neq 0\}| \leq s, \qquad x_j = 0, \ j > M.$$

Coherence: Define $\mu(U) = \sup |u_{ij}|^2$ as before.

New concepts

Uniform random subsampling: It is meaningless to draw $\Omega \subseteq \mathbb{N}$, $|\Omega| = m$ uniformly at random. It is also infeasible in practice due to bandwidth limitations. Hence, we fix the sampling bandwidth N and let

$$\Omega \subseteq \{1, \ldots, N\}, \quad |\Omega| = m,$$

be drawn uniformly at random.

Sparsity: Given finite sampling bandwidth, we cannot expect to recover any s-sparse infinite vector x stably. Let M be the sparsity bandwidth, and suppose that x is (s, M)-sparse:

$$|\{j = 1, \ldots, M : x_j \neq 0\}| \leq s, \qquad x_j = 0, \; j > M.$$

Coherence: Define $\mu(U) = \sup |u_{ij}|^2$ as before.

New concepts

Uniform random subsampling: It is meaningless to draw $\Omega \subseteq \mathbb{N}$, $|\Omega| = m$ uniformly at random. It is also infeasible in practice due to bandwidth limitations. Hence, we fix the sampling bandwidth N and let

$$\Omega \subseteq \{1, \ldots, N\}, \quad |\Omega| = m,$$

be drawn uniformly at random.

Sparsity: Given finite sampling bandwidth, we cannot expect to recover any s-sparse infinite vector x stably. Let M be the sparsity bandwidth, and suppose that x is (s, M)-sparse:

$$|\{j = 1, \ldots, M : x_j \neq 0\}| \leq s, \qquad x_j = 0, \; j > M.$$

Coherence: Define $\mu(U) = \sup |u_{ij}|^2$ as before.

Uneven sections and the balancing property

Let $P_N : l^2(\mathbb{N}) \to l^2(\mathbb{N})$ be the projection onto the first N elements, i.e. $P_N x = \{x_1, \dots, x_N, 0, 0, \dots\}$, $x \in l^2(\mathbb{N})$.

Key idea: Given a sparsity bandwidth M, we need to take the sampling bandwidth N sufficiently large.

Definition (The balancing property)

$N \in \mathbb{N}$ and $K \geq 1$ satisfy the strong balancing property with respect to $s, M \in \mathbb{N}$ if

(i) $\|P_M U^* P_N U P_M - P_M\|_{l^\infty} \leq \frac{1}{8}\left(\log_2(4\sqrt{s}KM)\right)^{-1/2}$,

(ii) $\|P_M^\perp U^* P_N U P_M\|_{l^\infty} \leq \frac{1}{8}$.

Why: The uneven section $P_N U P_M$ dictates the stability of the mapping from sampling bandwidth N to sparsity bandwidth M.

- Typically cannot take $M = N$, i.e. the finite section $P_N U P_N$.

Uneven sections and the balancing property

Let $P_N : l^2(\mathbb{N}) \to l^2(\mathbb{N})$ be the projection onto the first N elements, i.e. $P_N x = \{x_1, \ldots, x_N, 0, 0, \ldots\}$, $x \in l^2(\mathbb{N})$.

Key idea: Given a sparsity bandwidth M, we need to take the sampling bandwidth N sufficiently large.

Definition (The balancing property)

$N \in \mathbb{N}$ and $K \geq 1$ satisfy the strong balancing property with respect to $s, M \in \mathbb{N}$ if

(i) $\|P_M U^* P_N U P_M - P_M\|_{l^\infty} \leq \frac{1}{8}\left(\log_2(4\sqrt{s}KM)\right)^{-1/2}$,

(ii) $\|P_M^\perp U^* P_N U P_M\|_{l^\infty} \leq \frac{1}{8}$.

Why: The uneven section $P_N U P_M$ dictates the stability of the mapping from sampling bandwidth N to sparsity bandwidth M.

- Typically cannot take $M = N$, i.e. the finite section $P_N U P_N$.

Infinite-dimensional CS theorem

Theorem (BA, Hansen)

Suppose that $N \in \mathbb{N}$ and $K = N/m \geq 1$ satisfy the strong balancing property with respect to $s, M \in \mathbb{N}$ and also, for some for $0 < \epsilon \leq \mathrm{e}^{-1}$,

$$m \gtrsim s \cdot N \cdot \mu(U) \cdot \log(\epsilon^{-1}) \cdot \log(K \tilde{M} \sqrt{s})$$

where $\tilde{M} = \min\{i \in \mathbb{N} : \max_{k \geq i} \|P_N U P_{\{i\}}\| \leq 1/(32K\sqrt{s})\}$. If $\hat{x}$ is any minimizer of

$$\inf_{z \in l^1(\mathbb{N})} \|z\|_{l^1} \text{ subject to } \|P_\Omega U z - y\|_{l^2} \leq \delta\sqrt{K},$$

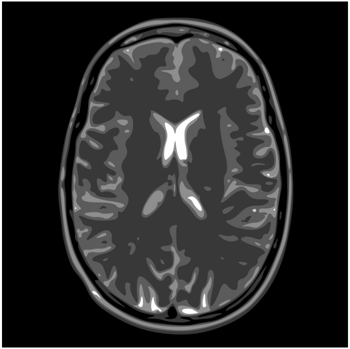
then

$$\|x - \hat{x}\|_{l^2} \lesssim \sigma_{s,M}(x) + \sqrt{s}\delta,$$

where $\sigma_{s,M}(x) = \min\{\|x - z\|_{l^1} : z \text{ is } (s, M)\text{-sparse}\}$.

Is this useful?

Consider the Fourier/wavelets problem.



Image

Reconstruction

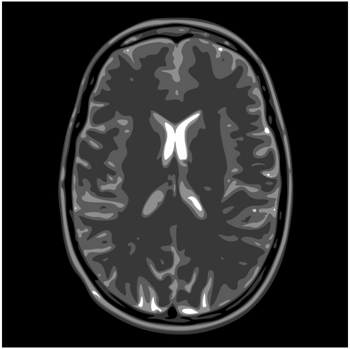
Unfortunately, for any wavelet basis with Fourier samples,

$$\mu(U) = \mathcal{O}(1), \quad N \to \infty.$$

$\Rightarrow$ one must take $m \approx s \cdot N \cdot \mu(U) = \mathcal{O}(N)$ measurements.

Is this useful?

Consider the Fourier/wavelets problem.



Image

Reconstruction

Unfortunately, for any wavelet basis with Fourier samples,

$$\mu(U) = \mathcal{O}(1), \quad N \to \infty.$$

$\Rightarrow$ one must take $m \approx s \cdot N \cdot \mu(U) = \mathcal{O}(N)$ measurements.

Lack of incoherence

Recall that

$$\mu(U) = \sup_{i,j \in \mathbb{N}} |u_{ij}|^2.$$

Hence $\mu(U)$ is a fixed quantity, independent of N.

This means that the bound

$$m \gtrsim s \cdot N \cdot \mu(U) \times \text{log factors},$$

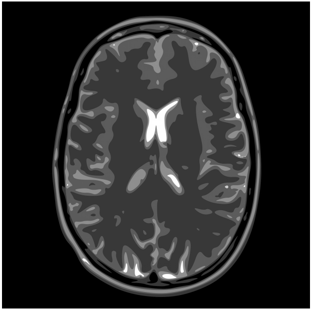
is $\gg s$ unless the sparsity bandwidth M, and therefore N, is small.

Hence, any such problem we run into the coherence barrier: *when subsampling uniformly at random, the number of samples required is typically much larger than the sparsity s.*

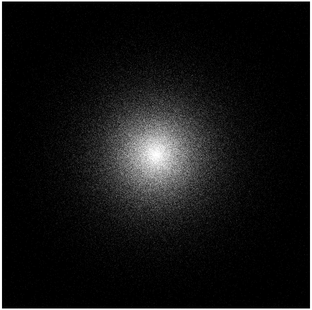
But CS is known to work in practice for MRI

To use CS here, one must sample according to a variable density (Lustig, Donoho & Pauli (2007)). Rather than choosing Ω uniformly at random, one oversamples at low frequencies.

Example: same image and same number of samples.



Image

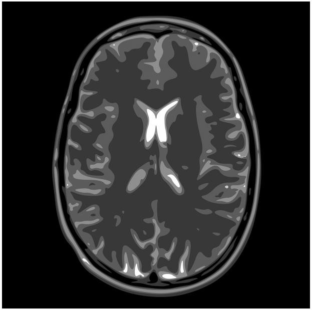


index set $\Omega \subseteq \mathbb{Z}^2$

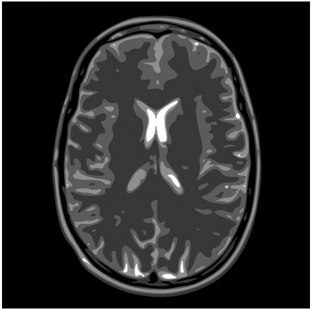
But CS is known to work in practice for MRI

To use CS here, one must sample according to a variable density (Lustig, Donoho & Pauli (2007)). Rather than choosing Ω uniformly at random, one oversamples at low frequencies.

Example: same image and same number of samples.



Image



Reconstruction

Infinite-dimensional CS requires new concepts

As shown by the previous example, neither

- Incoherence
- Uniform random subsampling

are applicable in this infinite-dimensional setting.

Claim: In the previous example, sparsity alone does not explain the reconstruction quality observed. In fact, the structure/ordering of the sparsity plays a crucial role.

The Flip Test

Recall: Sparsity means that there are s important coefficients, and their locations do not matter.

The Flip Test (BA, Hansen, Poon & Roman (2013)):

1. Take an image f with coefficients x. Form the measurements $y = P_\Omega U x$ and compute the approximation $f_1 \approx f$ by the usual CS reconstruction with appropriate Ω

$$\min_{z \in \mathbb{C}^N} \|z\|_{l^1} \text{ subject to } \|P_\Omega U z - y\|_{l^2} \leq \delta.$$

2. Permute the order of the wavelet coefficients by flipping the entries of x, to get a vector $\tilde{x}$.
3. Form measurements $\tilde{y} = P_\Omega U \tilde{x}$ and use exactly the same CS reconstruction to get the approximation $\tilde{x}_1 \approx \tilde{x}$.
4. Reverse the flipping operation to get the approximation $f_2 \approx f$.

The Flip Test

Recall: Sparsity means that there are s important coefficients, and their locations do not matter.

The Flip Test (BA, Hansen, Poon & Roman (2013)):

1. Take an image f with coefficients x. Form the measurements $y = P_\Omega U x$ and compute the approximation $f_1 \approx f$ by the usual CS reconstruction with appropriate Ω

$$\min_{z \in \mathbb{C}^N} \|z\|_{l^1} \text{ subject to } \|P_\Omega U z - y\|_{l^2} \leq \delta.$$

2. Permute the order of the wavelet coefficients by flipping the entries of x, to get a vector $\tilde{x}$.
3. Form measurements $\tilde{y} = P_\Omega U \tilde{x}$ and use exactly the same CS reconstruction to get the approximation $\tilde{x}_1 \approx \tilde{x}$.
4. Reverse the flipping operation to get the approximation $f_2 \approx f$.

The Flip Test

Recall: Sparsity means that there are s important coefficients, and their locations do not matter.

The Flip Test (BA, Hansen, Poon & Roman (2013)):

1. Take an image f with coefficients x. Form the measurements $y = P_\Omega U x$ and compute the approximation $f_1 \approx f$ by the usual CS reconstruction with appropriate Ω

$$\min_{z \in \mathbb{C}^N} \|z\|_{l^1} \text{ subject to } \|P_\Omega U z - y\|_{l^2} \leq \delta.$$

2. Permute the order of the wavelet coefficients by flipping the entries of x, to get a vector $\tilde{x}$.
3. Form measurements $\tilde{y} = P_\Omega U \tilde{x}$ and use exactly the same CS reconstruction to get the approximation $\tilde{x}_1 \approx \tilde{x}$.
4. Reverse the flipping operation to get the approximation $f_2 \approx f$.

The Flip Test

Recall: Sparsity means that there are s important coefficients, and their locations do not matter.

The Flip Test (BA, Hansen, Poon & Roman (2013)):

1. Take an image f with coefficients x. Form the measurements $y = P_\Omega U x$ and compute the approximation $f_1 \approx f$ by the usual CS reconstruction with appropriate Ω

$$\min_{z \in \mathbb{C}^N} \|z\|_{l^1} \text{ subject to } \|P_\Omega U z - y\|_{l^2} \leq \delta.$$

2. Permute the order of the wavelet coefficients by flipping the entries of x, to get a vector $\tilde{x}$.
3. Form measurements $\tilde{y} = P_\Omega U \tilde{x}$ and use exactly the same CS reconstruction to get the approximation $\tilde{x}_1 \approx \tilde{x}$.
4. Reverse the flipping operation to get the approximation $f_2 \approx f$.

The Flip Test

Recall: Sparsity means that there are s important coefficients, and their locations do not matter.

The Flip Test (BA, Hansen, Poon & Roman (2013)):

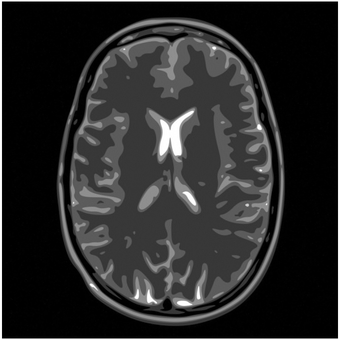
1. Take an image f with coefficients x. Form the measurements $y = P_\Omega U x$ and compute the approximation $f_1 \approx f$ by the usual CS reconstruction with appropriate Ω

$$\min_{z \in \mathbb{C}^N} \|z\|_{l^1} \text{ subject to } \|P_\Omega U z - y\|_{l^2} \leq \delta.$$

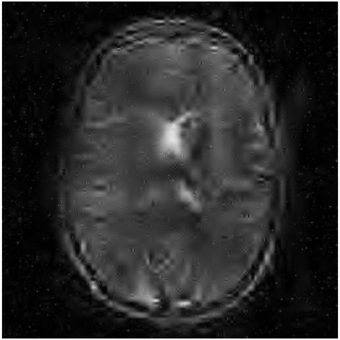
2. Permute the order of the wavelet coefficients by flipping the entries of x, to get a vector $\tilde{x}$.
3. Form measurements $\tilde{y} = P_\Omega U \tilde{x}$ and use exactly the same CS reconstruction to get the approximation $\tilde{x}_1 \approx \tilde{x}$.
4. Reverse the flipping operation to get the approximation $f_2 \approx f$.

Numerical results

Sparsity is unaffected by permuations, so f_1 and f_2 should give the same reconstructions:



unflipped reconstruction f_1



flipped reconstruction f_2

- 10% subsampling at 1024×1024 with a variable density strategy

Outline

New assumptions for compressed sensing

Conventional assumptions of CS:

- Incoherence
- Sparsity
- Uniform random subsampling

New assumptions:

- Local coherence in levels
- Sparsity in levels
- Multilevel random subsampling

New assumptions for compressed sensing

Conventional assumptions of CS:

- Incoherence
- Sparsity
- Uniform random subsampling

New assumptions:

- Local coherence in levels
- Sparsity in levels
- Multilevel random subsampling

Asymptotic incoherence

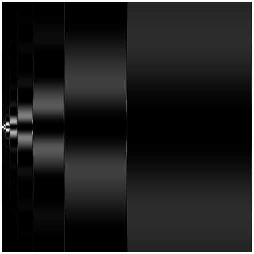
Definition

An infinite matrix U is asymptotically incoherent if

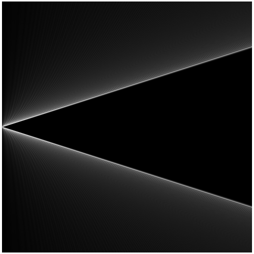
$$\mu(P_K^\perp U),\ \mu(UP_K^\perp) \to 0, \qquad K \to \infty.$$

- High coherence occurs only in the leading $K \times K$ submatrix of U.

Absolute values of the entries of U (both examples are coherent):



Fourier/wavelets, $\mathcal{O}(K^{-1})$



Fourier/polynomials, $\mathcal{O}(K^{-2/3})$

Local coherence in levels

We divide $P_N U P_M$ into rectangular blocks. Let

- $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ with $0 = N_0 < N_1 < \ldots < N_r = N$,
- $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $0 = M_0 < M_1 < \ldots < M_r = M$.

Notation: for $a, b \in \mathbb{N}$, let $P_a^b = P_a P_b^\perp$.

Definition

The $(k,l)^{\mathrm{th}}$ local coherence of U is given by

$$\mu(k,l) = \begin{cases} \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}})\mu(P_{N_k}^{N_{k-1}} U)} & l \neq r \\ \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_{l-1}}^{\perp})\mu(P_{N_k}^{N_{k-1}} U)} & l = r \end{cases} \qquad k, l = 1, \ldots, r.$$

Asymptotically incoherent matrices are globally coherent, but locally incoherent as k or l tends to infinity.

New assumptions for compressed sensing

Conventional assumptions of CS:

- Incoherence
- Sparsity
- Uniform random subsampling

New assumptions:

- Local coherence in levels
- Sparsity in levels
- Multilevel random subsampling

Multilevel random subsampling

We divide up the first N rows of U into the same levels indexed by $\mathbf{N}$, and define

- $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$ with $m_k \leq N_k - N_{k-1}$,
- $\Omega_k \subseteq \{N_{k-1}+1, \ldots, N_k\}$, $|\Omega_k| = m_k$ be chosen uniformly at random.

We call $\Omega_{\mathbf{N},\mathbf{m}} = \bigcup_k \Omega_k$ an $(\mathbf{N}, \mathbf{m})$-multilevel sampling scheme.

- Note that variable density strategies can be modelled by multilevel schemes with $m_k/(N_k - N_{k-1}) \to 0$ as $k \to \infty$.

New assumptions for compressed sensing

Conventional assumptions of CS:

- Incoherence
- Sparsity
- Uniform random subsampling

New assumptions:

- Local coherence in levels
- Sparsity in levels
- Multilevel random subsampling

Sparsity in levels

The flip test shows we must incorporate structure into the new sparsity assumption.

To do this, we divide $P_M x$ up into levels corresponding to the column blocks of U indexed by $\mathbf{M}$. Let

$$\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r, \quad s_k \leq M_k - M_{k-1}.$$

We say that $x = (x_1, \ldots, x_N)^\top$ is $(\mathbf{s}, \mathbf{M})$-sparse if

$$|\{j : x_j \neq 0\} \cap \{M_{k-1} + 1, \ldots, M_k\}| \leq s_k, \quad k = 1, \ldots, r.$$

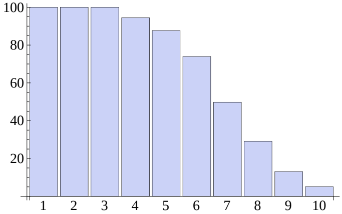
Write $\sigma_{\mathbf{s},\mathbf{M}}(x) = \min\{\|x - z\|_{l^1} : \text{z is } (\mathbf{s}, \mathbf{M})\text{-sparse}\}$

Images are asymptotically sparse in wavelets

Definition

An infinite vector x is asymptotically sparse in levels if x is sparse in levels with $s_k/(M_k - M_{k-1}) \to 0$ as $k \to \infty$.

Wavelet coefficients are not just sparse, but asymptotically sparse when the levels correspond to wavelet scales.



Left: image. Right: percentage of wavelet coefficients per scale $> 10^{-3}$.

Towards the main theorem

We need the concept of a relative sparsity.

Definition

Let $x \in \mathbb{C}^N$ be $(\mathbf{s}, \mathbf{M})$-sparse. Given $\mathbf{N}$, we define the relative sparsity

$$S_k = \max_{\eta \in \Theta} \| P_{N_k}^{N_{k-1}} U \eta \|^2,$$

where $\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1, \ \eta \text{ is } (\mathbf{s}, \mathbf{M})\text{-sparse}\}$.

This concept takes into account interference between different sparsity levels, i.e. the fact that U is not block diagonal.

Main theorem

Theorem (BA, Hansen, Poon & Roman (2013))

Suppose that $N = N_r$, $K = \max_{k=1,\ldots,r}\left\{\frac{N_k - N_{k-1}}{m_k}\right\}$ *satisfy the strong balancing property with respect to* $M = M_r$ *and* $s = s_1 + \ldots + s_r$, *and*

- *we have*

$$m_k \gtrsim (N_k - N_{k-1}) \left(\sum_{l=1}^{r} \mu(k,l) \cdot s_l \right) \cdot \log(\epsilon^{-1}) \cdot \log(K\tilde{M}\sqrt{s}),$$

- *we have* $m_k \gtrsim \hat{m}_k \cdot \log(\epsilon^{-1}) \cdot \log(K\tilde{M}\sqrt{s})$, *where* $\hat{m}_k$ *satisfies*

$$1 \gtrsim \sum_{k=1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1 \right) \cdot \mu(k,l) \cdot S_k, \quad l = 1, \ldots, r.$$

Then with probability at least $1 - s\epsilon$, *we have*

$$\|x - \hat{x}\|_{l^2} \lesssim \sigma_{\mathbf{s},\mathbf{M}}(x) + \sqrt{s}\delta.$$

Interpretation

The key parts of the theorem are the estimates

$$m_k \gtrsim (N_k - N_{k-1}) \left(\sum_{l=1}^{r} \mu(k,l) \cdot s_l \right) \left(\log(\epsilon^{-1}) + 1 \right) \cdot \log(\tilde{M}), \qquad (1)$$

and $m_k \gtrsim \hat{m}_k \cdot (\log(\epsilon^{-1})) \cdot \log(\tilde{M})$, where

$$1 \gtrsim \sum_{k=1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1 \right) \cdot \mu(k,l) \cdot S_k, \quad l = 1, \ldots, r. \qquad (2)$$

Main point: The local numbers of samples m_k depend on

- the local sparsities $s_1, \ldots, s_r$,
- the relative sparsities $S_1, \ldots, S_r$,
- the local coherences $\mu(k,l)$,

rather than the global sparsity s and global coherence μ.

Sharpness: Estimates reduce to inf.-theoretic limits in certain cases.

The Fourier/wavelets case

Recall this is the usual setup for CS in MRI and other applications.

Theorem (BA, Hansen, Poon & Roman (2013))

Let $\mathbf{M}$ *correspond to wavelet scales and* $\mathbf{s}$ *the sparsities within them. Let* $A > 1$ *be a constant depending on the smoothness and number of vanishing moments of the wavelet used. Then, subject to appropriate, but mild, conditions one can find* $N_k = \mathcal{O}(M_k)$ *such that it suffices to take*

$$m_k \gtrsim \left(s_k + \sum_{l \neq k} s_l A^{-|k-l|} \right) \cdot \log(\epsilon^{-1}) \cdot \log(\tilde{N}). \qquad (\star)$$

Remark: This theorem explains why CS works in such applications.

- $(\star)$ is in agreement with the flip test.
- In the presence of asymptotic sparsity, the subsampling fraction $m_k/(N_k - N_{k-1})$ decreases as $k \to \infty$.

Outline

Type II CS problems

Unlike in MRI, CT, etc, many CS applications are finite-dimensional, and there is substantial freedom to design the matrix Ψ.

- E.g. Single-pixel camera (Rice), lensless imaging (Bell Labs), fluorescence microscopy
- Hardware constraints: Ψ must be binary.

CS 'gospel': Gaussian random measurements are the 'optimal' choice.

- We should try to redesign MRI to produce random Gaussians.

Question: Is there a better choice when there is sparsity and structure?

Answer: No. Multilevel subsampled Fourier or Hadamard matrices lead to significant improvements.

Type II CS problems

Unlike in MRI, CT, etc, many CS applications are finite-dimensional, and there is substantial freedom to design the matrix Ψ.

- E.g. Single-pixel camera (Rice), lensless imaging (Bell Labs), fluorescence microscopy
- Hardware constraints: Ψ must be binary.

CS 'gospel': Gaussian random measurements are the 'optimal' choice.

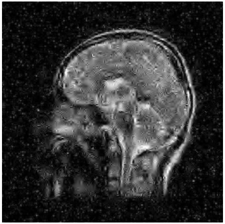
- We should try to redesign MRI to produce random Gaussians.

Question: Is there a better choice when there is sparsity and structure?

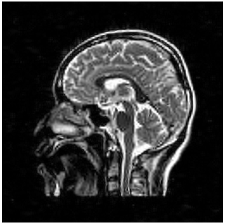
Answer: No. Multilevel subsampled Fourier or Hadamard matrices lead to significant improvements.

Example: 12.5% measurements using DB4 wavelets

256×256 | 512×512 | 1024×1024



Err = 41.6% | Err = 25.3% | Err = 11.6%



Err = 21.9% | Err = 10.9% | Err = 3.1%

Top row: Gaussian. Bottom row: Multilevel Fourier

Structured sampling vs. structured recovery

Multilevel subsampling with Fourier/Hadamard matrices

- Use standard recovery algorithm (l^1 minimization)
- Exploit asymptotic sparsity in levels structure in the sampling process, e.g. multilevel subsampled Fourier/Hadamard

Other structured CS algorithms: Model-based CS, Baraniuk et al. (2010), Bayesian CS, Ji, Xue & Carin (2008), He & Carin (2009), Turbo AMP, Som & Schniter (2012)

- Exploit the connected tree structure of wavelet coefficients
- Use standard measurements, e.g. random Gaussians/Bernoullis
- Modify the recovery algorithm (e.g. CoSaMP or IHT)

Comparison: 12.5% sampling at 256×256 resolution

Original

ℓ^1 Gauss., Err $= 15.7\%$

Model-CS, Err $= 17.9\%$

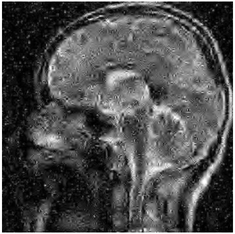
BCS, Err $= 12.1\%$

TurboAMP, Err $= 17.7\%$

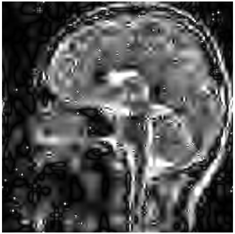
Mult. Four., Err $= 8.8\%$

Comparison: 12.5% sampling at 256×256 resolution

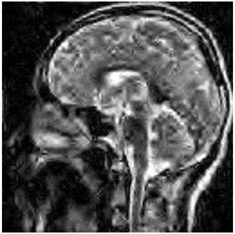
Original



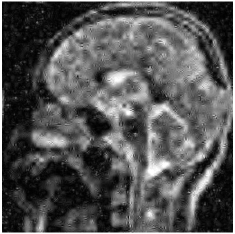
ℓ^1 Bern., Err $= 41.2\%$



Model-CS, Err $= 41.8\%$



BCS, Err $= 29.6\%$



TurboAMP, Err $= 39.3\%$

Mult. Four., Err $= 18.2\%$

Outline

Conclusions

1. Finite-dimensional CS is based on sparsity, incoherence and uniform random subsampling. However, in applications arising from continuous models, these are often not appropriate.

2. In infinite dimensions, more suitable properties are sparsity in levels, local coherence and multilevel random subsampling.

3. The mathematical framework introduced provides a theoretical explanation for the success of CS in some key applications.

4. Moreover, the insight gained from infinite dimensions about the key role played by structure leads to new and better approaches to finite-dimensional problems.

References

- Roman, BA & Hansen, *On asymptotic structure in compressed sensing*, arXiv:1406.4178, 2014.
- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: a new theory for compressed sensing*, arXiv:1302.0561, 2014.
- BA, Hansen & Roman, *The quest for optimal sampling: computationally efficient, structure-exploiting measurements for compressed sensing*, Compressed Sensing and its Applications, Springer (to appear), 2014.
- BA & Hansen, *Generalized sampling and infinite-dimensional compressed sensing,* Found. Comput. Math. (under revision), 2013.
- BA, Hansen, Roman & Teschke, *Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum,* Advances in Imaging and Electron Physics 182:187-279, 2014.
- BA & Hansen, *A generalized sampling theorem for stable reconstructions in arbitrary bases*, J. Fourier Anal. Appl. 18(4):685-716, 2012.