Getting even more from less: A new framework for compressed sensing

Ben Adcock
Department of Mathematics
Purdue University

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

Outline

Outline

Compressed sensing

The need for a new theory

A new framework for compressed sensing

Getting even more from less

Conclusions

Recovery problems

In many applications, we want to recover an object from a collection of measurements. If the sampling process is linear, then we may write this as

$$y = Ux,$$

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,
- $U \in \mathbb{C}^{N \times N}$ is the measurement matrix.

If U is invertible then we can recover x as $U^{-1}y$.

- For the remainder of the talk, U will be an isometry.

Compressed sensing

Typically, we do not have access to all the measurements

$$y = \{y_1, y_2, \ldots, y_N\}.$$

Instead, we only have a small subset

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

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

Problem: Recover x from the highly underdetermined linear system

$$P_\Omega U x = P_\Omega y.$$

Moreover, do this using efficient numerical algorithms.

Notation: write $P_\Omega \in \mathbb{C}^{N \times N}$ for the diagonal matrix with j^{th} entry 1 if $j \in \Omega$ and 0 otherwise.

Compressed sensing

Under appropriate conditions on x, U and Ω, we can recover x from $P_{\Omega}y$. Moreover, this can be achieved by efficient numerical algorithms.

Compressed Sensing (CS)

- 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,...
- Important philosophical shift in how we view the task of reconstruction/inference.

Key principles: sparsity, incoherence, uniform random subsampling

The condition on x: Sparsity

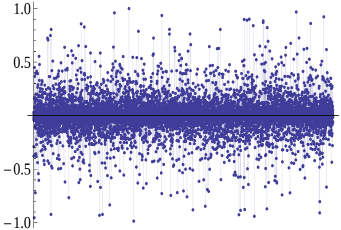
Definition

A vector $x = (x_1, \dots, x_N)^\top$ is *s-sparse* if $|\{j : x_j \neq 0\}| \leq s$.

- Typically $s \ll N$.
- The locations of the nonzero entries are unknown.

Often, $x = W^* z$ are the coefficients of the image z in some orthogonal sparsifying transformation $W = [w_1 | \cdots | w_N]$, e.g. wavelets.

z



x

The condition on U: Incoherence

Definition

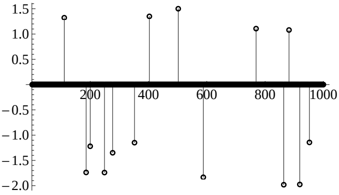
The (mutual) coherence of an isometry $U = (u_{ij}) \in \mathbb{C}^{N \times N}$ is

$$\mu(U) = \max |u_{ij}|^2 \in [N^{-1}, 1].$$

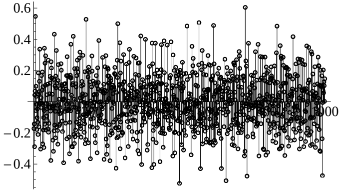
The matrix U is incoherent if $\mu(U) = \mathcal{O}\left(N^{-1}\right)$.

Intuition: if x is sparse, then $y = Ux$ cannot be sparse.

- Discrete uncertainty principle (Donoho & Starck, Elad & Bruckstein,...)



x



$y = Ux$

The condition on Ω: Uniform randomness

The index index set $\Omega \subseteq \{1, \ldots, N\}$, $|\Omega| = m$ should be taken uniformly at random.

Informal explanation:

- The sparse signal x has $2s$ information content:

$$s \text{ locations } + \ s \text{ coefficient values.}$$

- Incoherence means this information is distributed uniformly amongst the measurements $y_1, y_2, \ldots, y_N$.
- Hence, any $m = \mathcal{O}(s)$ 'representative' measurements should contain sufficient information to recover x.

Reconstruction algorithm

The system

$$P_\Omega U z = P_\Omega y,$$

has infinitely many solutions. We seek the one which coincides with the sparse vector x.

Most typically, one solves the convex optimization problem

$$\min_{z \in \mathbb{C}^N} \|z\|_{l^1} \text{ subject to } P_\Omega U z = P_\Omega y,$$

where $\|z\|_{l^1} = |z_1| + |z_2| + \ldots + |z_N|$ is the l^1-norm.

- Geometry of l^1 balls in high dimensions (Donoho & Tanner). The l^1-norm promotes sparsity.
- Other approaches: greedy methods (e.g. OMP, CoSaMP), thresholding methods (e.g. IHT, HTP), message passing algorithms,....

A compressed sensing theorem

Theorem (Candès & Plan (2011))

Let x be s-sparse and suppose that $\epsilon > 0$. Let $\Omega \subseteq \{1, \ldots, N\}$, $|\Omega| = m$ be chosen uniformly at random, where

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

for some universal constant C. Then with probability greater than $1 - \epsilon$ the problem

$$\min_{z \in \mathbb{C}^N} \|z\|_{l^1} \text{ subject to } P_\Omega U z = P_\Omega y,$$

has a unique solution and this solution coincides with x.

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

$$m \approx s \log N \ll N.$$

NB. No Restricted Isometry Property (RIP).

Outline

Compressed sensing

The need for a new theory

A new framework for compressed sensing

Getting even more from less

Conclusions

Compressed sensing in inverse problems

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

CS has been applied in/considered for all these problems.

- For MRI, see Lustig, Donoho & Pauli (2007), Lustig et al. (2008)

Let f be the image to recover. Mathematically, all these problems can be reduced (possibly via the Fourier slice theorem) to the following:

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

Here $\Omega \subseteq \hat{\mathbb{R}}^d$ is a finite set of frequencies, and $\hat{f}$ denotes the Fourier transform.

Compressed sensing in inverse problems

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

CS has been applied in/considered for all these problems.

- For MRI, see Lustig, Donoho & Pauli (2007), Lustig et al. (2008)

Let f be the image to recover. Mathematically, all these problems can be reduced (possibly via the Fourier slice theorem) to the following:

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

Here $\Omega \subseteq \hat{\mathbb{R}}^d$ is a finite set of frequencies, and $\hat{f}$ denotes the Fourier transform.

Standard compressed sensing setup

We form the measurement matrix

$$U = FW,$$

where $F \in \mathbb{C}^{N \times N}$ is the discrete Fourier transform, $W \in \mathbb{C}^{N \times N}$ is an appropriate sparsifying transform (e.g. wavelets), and solve

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

where $y = \{\hat{f}(\omega) : \omega \in \Omega\} + \eta$ is the vector of noisy measurements with $\|\eta\|_{l^2} \leq \delta$.

Warning

This setup is a discretization of the continuous model:

$$\text{continuous FT} \approx \text{discrete FT} \quad \Rightarrow \quad \text{samples mismatch}$$

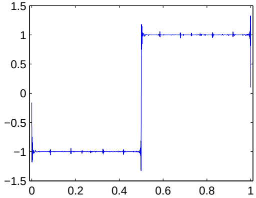
Similar to basis mismatch phenomenon

- Chi, Scharf, Pezeshki & Calderbank (2011), Herman & Strohmer (2010)

This mismatch has two primary effects:

1. If measurements are simulated via the DFT $\Rightarrow$ inverse crimes
   - Guerquin–Kern, Lejeune, Pruessman, Unser (2012)
2. Minimization problem has no sparse solution $\Rightarrow$ poor reconstructions

Example: Recovery of the 2nd Haar wavelet with $N = 256$ and $m = 128$.



How to avoid this: infinite-dimensional CS

To avoid these issues, one can formulate the reconstruction problem in the continuous domain first and then discretize.

Infinite-dimensional compressed sensing (BA & Hansen, 2012)

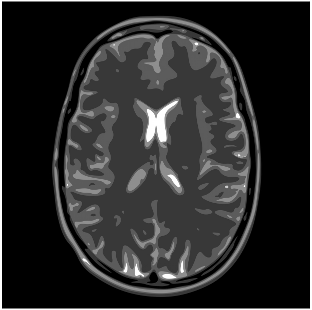
- Extends the standard CS theory:
  - Vector spaces $\rightarrow$ Separable Hilbert spaces
  - Matrices $U \in \mathbb{C}^{N \times N}$ $\rightarrow$ Operators $U : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$.
- Key issues: (i) truncation of U via uneven sections and balancing property (ii) dimension-independent probability bounds (iii) dealing with infinite, and unknown, tails

Implementation in CS MRI

- Guerquin-Kern, Häberlin, Pruessmann & Unser, 2011

The Fourier/wavelets recovery problem is coherent

Example: 6.25% random subsampling with $N = 2048 \times 2048$.



Image

Reconstruction

Unfortunately, for any wavelet basis

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

The theorem suggests that $m \approx N\mu(U)s \approx N$ is required.

Incoherence is rare in practice

Recall that

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

Any problem that arises from the combination of a

- continuous transform (e.g. Fourier, Radon,...),
- a countable orthonormal basis or frame,

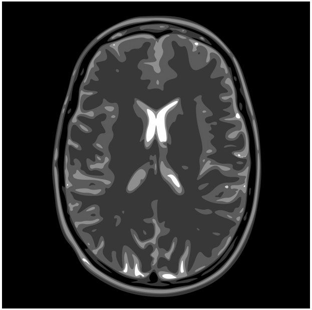
or a discretization thereof, will have a finite, and fixed coherence.

i.e. Most applications of CS in inverse problems.

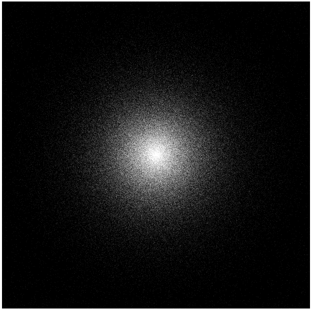
But CS is known to work for such problems...

To use CS in these applications, 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: 6.25% subsampling with $N = 2048 \times 2048$.



Image

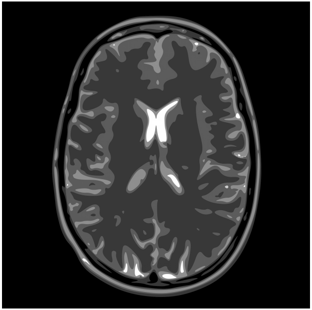


index set $\Omega \subseteq \mathbb{Z}^2$

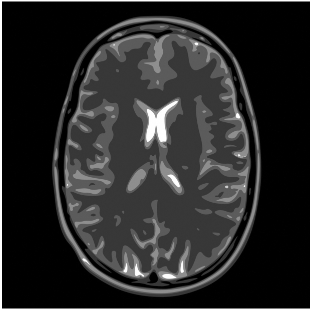
But CS is known to work for such problems...

To use CS in these applications, 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: 6.25% subsampling with $N = 2048 \times 2048$.



Image



Reconstruction

The need for a new framework

No existing CS theory fully explains why this works. In particular, the standard assumptions of

- Incoherence
- Uniform random subsampling

are clearly not relevant here.

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

The need for a new framework

No existing CS theory fully explains why this works. In particular, the standard assumptions of

- Incoherence
- Uniform random subsampling

are clearly not relevant here.

Claim: In such applications, 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)):

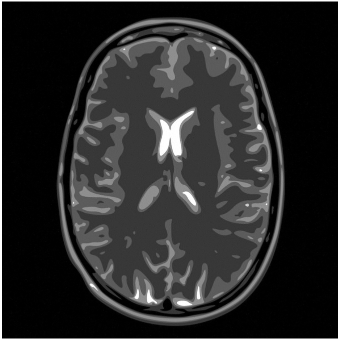
1. Take an image f with coefficients x. Form the measurements $y = P_\Omega U x$ and compute the approximation $f_1 \approx f$ by the usual CS reconstruction with appropriate Ω

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

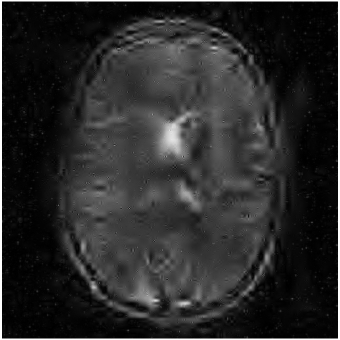
2. Permute the order of the wavelet coefficients by flipping the entries of x, to get a vector $\tilde{x}$.
3. Form measurements $\tilde{y} = P_\Omega U \tilde{x}$ and use exactly the same CS reconstruction to get the approximation $\tilde{x}_1 \approx \tilde{x}$.
4. Reverse the flipping operation to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



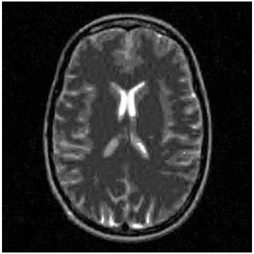
unflipped reconstruction f_1



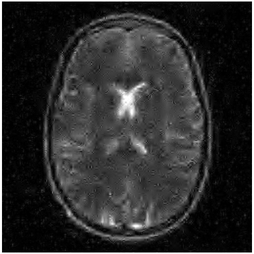
flipped reconstruction f_2

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

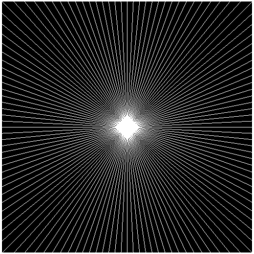
Numerical results



unflipped recon.



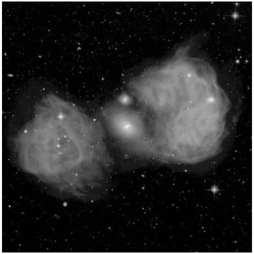
flipped recon.



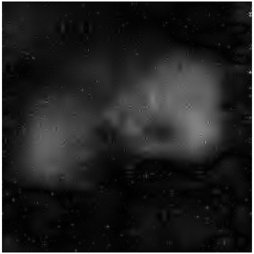
subsampling map

X-ray CT: $N = 512 \times 512$, $m/N = 12\%$, $U = FW$, F is the Fourier transform

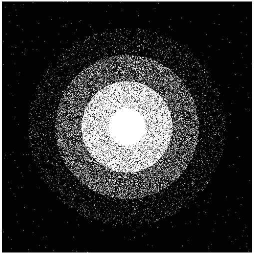
Numerical results



unflipped recon.



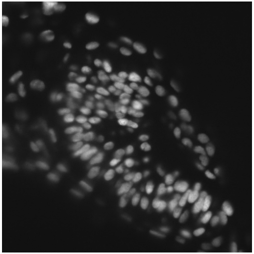
flipped recon.



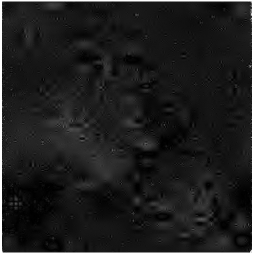
subsampling map

Radio interferometry: $N = 512 \times 512$, $m/N = 15\%$, $U = FW$, F is the Fourier transform

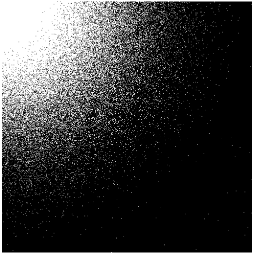
Numerical results



unflipped recon.



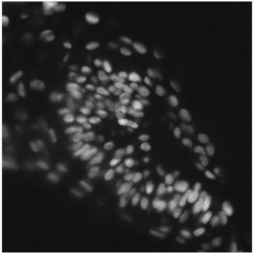
flipped recon.

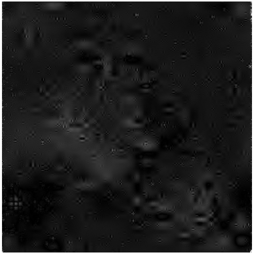


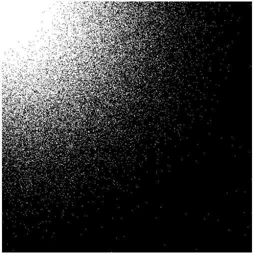
subsampling map

Fluorescence microscopy: $N = 512 \times 512$, $m/N = 20\%$, $U = HW$, H is the Hadamard transform

Numerical results


unflipped recon.


flipped recon.


subsampling map

Fluorescence microscopy: $N = 512 \times 512$, $m/N = 20\%$, $U = HW$, H is the Hadamard transform

The flip test also shows that: (i) there is no RIP (ii) the optimal subsampling strategy must depend on the image

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

Asymptotic incoherence

Definition

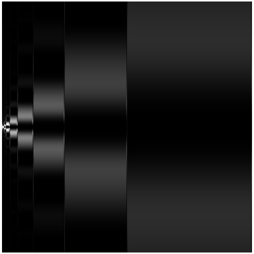
A matrix U is asymptotically incoherent if

$$\mu(P_K^{\perp} U),\ \mu(U P_K^{\perp}) \to 0, \qquad K, N \to \infty, K/N \leq c < 1.$$

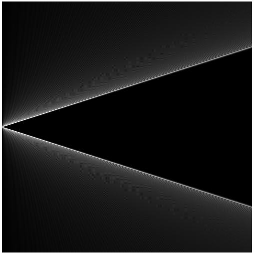
Here $P_K^{\perp}$ is the projection onto indices $\{K+1, \dots, N\}$.

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

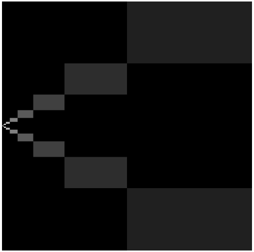
Abs. values of the entries of the matrix U (all examples are coherent):



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



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



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

Local coherence in levels

We divide the matrix U 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 $0 = M_0 < M_1 < \ldots < M_r = N$.

Notation: for $M, K \in \mathbb{N}$, let $P_M^K = P_M P_K^\perp$.

Definition (Local coherence)

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

$$\mu(k,l) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}}) \mu(P_{N_k}^{N_{k-1}} U)}, \quad k,l = 1, \ldots, r.$$

Asymptotically incoherent matrices are globally coherent, but locally incoherent away from their leading blocks.

Multilevel random subsampling

Asymptotic incoherence suggests a new sampling strategy:

1. Sample fully in coherent regions of U (first K rows).
2. Subsample elsewhere (remaining rows).

In general, we divide up the rows of U into the same levels indexed by $\mathbf{N}$, and let

- $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$ 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 $k \to \infty$.

Sparsity in levels

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

To do this, we divide our vector 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

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

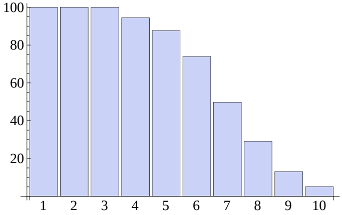
Note: The levels do not necessarily correspond to wavelet scales.

Images are asymptotically sparse in wavelelets

Definition

A 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 (Relative sparsity)

Let $x \in l^2(\mathbb{N})$ with $|\{j : x_j \neq 0\} \cap \{M_{k-1}+1, \dots, M_k\}| = s_k$ for $k = 1, \dots, r$. Define the relative sparsity

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

where $\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1,\ |\mathrm{supp}(P_{M_l}^{M_{l-1}}\eta)| = s_l,\ l = 1, \dots, r\}$.

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

Main theorem

Given $\mathbf{s}$ and $\mathbf{M}$, let $\sigma_{\mathbf{s},\mathbf{M}}(x)$ be the error of the best l^1 norm approximation of x using an $(\mathbf{s},\mathbf{M})$-sparse vector.

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

Let $\epsilon > 0$ be given. Suppose that:

- *we have*

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

- *we have* $m_k \gtrsim \hat{m}_k \cdot (\log(\epsilon^{-1}) + 1) \cdot \log(N)$, *where* $\hat{m}_k$ *satisfies*

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

Main theorem

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

Let $\hat{x}$ be any minimizer of

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

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

$$\|x - \hat{x}\|_{l^2} \leq C\left(\delta\sqrt{K}(1 + L\sqrt{s}) + \sigma_{\mathbf{s},\mathbf{M}}(x)\right),$$

for some universal constant $C > 0$, *where* $K = \max_{k=1,\ldots,r} \frac{N_k - N_{k-1}}{m_k}$ *and* $L = 1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4KM\sqrt{s})}$. *If* $m_k = N_k - N_{k-1}$, $k = 1, \ldots, r$, *then this holds with probability* 1.

- Generalization of standard CS result (Candès & Plan) to the case of more than one level.
- Extends to the infinite-dimensional CS setting (BA & Hansen).

Interpretation

The key parts of the theorem are the estimates

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

and $m_k \gtrsim \hat{m}_k \cdot (\log(\epsilon^{-1}) + 1) \cdot \log(N)$, 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 now 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 of the estimates

The estimates are complicated by two factors:

- The local sparsities S_k
- The nondiagonal terms $\mu(k,l)$, $k \neq l$.

Relative sparsities:

- Note that $S_k \leq s = s_1 + \ldots + s_r$ in general.
- Moreover, it is easy to construct examples where $S_k = s_{k'}$ for some $k' \neq k$, i.e. S_k may have no relation to s_k.

This means that we cannot in general get bounds of the form:

$$m_k \gtrsim s_k \times \text{log factors},$$

as one may intuitively have expected.

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

Subject to mild conditions, if $U = V \otimes W$ is a Kronecker product matrix then the estimates (1) and (2) reduce to known information-theoretic limits (up to a log factor in the failure probability ϵ).

Sharpness of the estimates

The estimates are complicated by two factors:

- The local sparsities S_k
- The nondiagonal terms $\mu(k,l)$, $k \neq l$.

Relative sparsities:

- Note that $S_k \leq s = s_1 + \ldots + s_r$ in general.
- Moreover, it is easy to construct examples where $S_k = s_{k'}$ for some $k' \neq k$, i.e. S_k may have no relation to s_k.

This means that we cannot in general get bounds of the form:

$$m_k \gtrsim s_k \times \text{log factors},$$

as one may intuitively have expected.

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

Subject to mild conditions, if $U = V \otimes W$ is a Kronecker product matrix then the estimates (1) and (2) reduce to known information-theoretic limits (up to a log factor in the failure probability ϵ).

Sharpness of the estimates

The estimates are complicated by two factors:

- The local sparsities S_k
- The nondiagonal terms $\mu(k,l)$, $k \neq l$.

Relative sparsities:

- Note that $S_k \leq s = s_1 + \ldots + s_r$ in general.
- Moreover, it is easy to construct examples where $S_k = s_{k'}$ for some $k' \neq k$, i.e. S_k may have no relation to s_k.

This means that we cannot in general get bounds of the form:

$$m_k \gtrsim s_k \times \text{log factors},$$

as one may intuitively have expected.

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

Subject to mild conditions, if $U = V \otimes W$ is a Kronecker product matrix then the estimates (1) and (2) reduce to known information-theoretic limits (up to a log factor in the failure probability ϵ).

The Fourier/wavelets case

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

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

Let $\mathbf{M}$ *correspond to wavelet scales. 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*

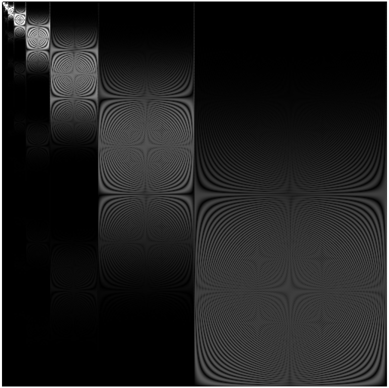
$$m_k \gtrsim \left(s_k + \sum_{l \neq k} A^{-|k-l|} s_l \right) \times \text{log factors}. \qquad (\star)$$

- The estimate $m_k \gtrsim s_k \times$ log factors is optimal.
- Thus, $(\star)$ is optimal up to exponentially decaying factors.

This is the first comprehensive proof that CS works in such applications.

- Krahmer & Ward (2013): sparsity-based estimates for Haar wavelets

Ideas behind the proof



Off-diagonal blocks have small contributions. In particular:

- $\mu(k,l) \lesssim A^{-|k-l|}\mu(k,k)$ for $l \neq k$
- $S_k \lesssim s_k + \sum_{l \neq k} A^{-|k-l|} s_l$

In other words, interference between levels is provably controllable.

Outline

Beyond sparsity: getting even more from less

Much of compressed sensing is based on sparsity.

- Most standard theory: incoherence, RIP, NSP, unversality...
- Standard 'designer' matrices: e.g. random Gaussian, Bernoulli

Inverse problems: the flip test shows that sparsity alone is not sufficient. Structure plays a vital role in the reconstruction quality.

Structured sparsity: Given that such structure is present in many CS applications, we should try to leverage it.

- Tsaig & Donoho (2006), Eldar (2009), He & Carin (2009), Baraniuk et al. (2010), Krzakala et al. (2011), Duarte & Eldar (2011), Som & Schniter (2012), Renna et al. (2013), Chen et al. (2013) + others

Compressive Imaging

Unlike in most inverse problems, in many CS applications we have substantial freedom in designing the sensing matrix U.

- E.g. Single-pixel camera (Rice), lenseless imaging (Bell Labs)
- Hardware constraints: U must be binary.

The usual CS approach is to use random Bernoulli matrices. However:

- Unstructured, must be stored. Infeasible for $> 256 \times 256$ resolution.
- Universal: work for any sparsity basis (e.g. wavelets), but cannot exploit sparsity structure.

There exist alternatives with fast transforms (e.g STOne, random convolutions), but these also cannot exploit structure.

Using asymptotic incoherence to exploit structure

Our theory shows that asymptotic incoherence is a blessing, as opposed to a curse.

- It is much more general than uniform incoherence (i.e. more flexibility in sensing matrix design).
- It allows one to exploit structure via multilevel random subsampling.

Question: Do their exist computationally efficient binary sensing matrices that are asymptotically incoherent with wavelet bases?

Answer: Yes! The Hadamard transform – $\mathcal{O}(N \log N)$ + no storage.

Using asymptotic incoherence to exploit structure

Our theory shows that asymptotic incoherence is a blessing, as opposed to a curse.

- It is much more general than uniform incoherence (i.e. more flexibility in sensing matrix design).
- It allows one to exploit structure via multilevel random subsampling.

Question: Do their exist computationally efficient binary sensing matrices that are asymptotically incoherent with wavelet bases?

Answer: Yes! The Hadamard transform – $\mathcal{O}(N \log N)$ + no storage.

Getting even more from less (and doing it efficiently)

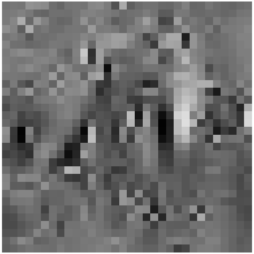
Example: The Berlin cathedral with 15% sampling at various resolutions using Daubechies-4 wavelets.

Experiments performed using SPGL1 on an Intel i7-3770K, 32 GB RAM and an Intel Xeon E7, 256 GB RAM.

Getting even more from less (and doing it efficiently)

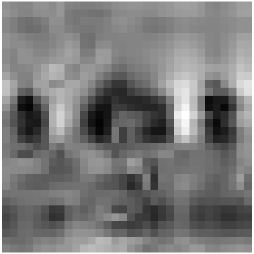
Resolution: 32×32

Random Bernoulli



RAM (GB): < 0.1
Speed (it/s): 55.7
Rel. Err. (%): 33.1
Time: 4.7s

Multilevel Hadamard



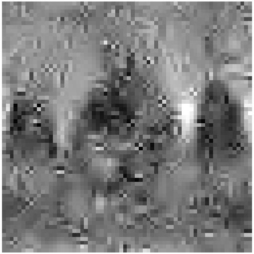
RAM (GB): < 0.1
Speed (it/s): 53.1
Rel. Err. (%): 20.5
Time: 5.7s

Original image

Getting even more from less (and doing it efficiently)

Resolution: 64×64

Random Bernoulli



RAM (GB): < 0.1
Speed (it/s): 39.8
Rel. Err. (%): 27.8
Time: 7.1s

Multilevel Hadamard

RAM (GB): < 0.1
Speed (it/s): 34.2
Rel. Err. (%): 19.3
Time: 7.9s

Original image

Getting even more from less (and doing it efficiently)

Resolution: 128×128

Random Bernoulli

RAM (GB): 0.3
Speed (it/s): 12.4
Rel. Err. (%): 26.4
Time: 25s

Multilevel Hadamard

RAM (GB): < 0.1
Speed (it/s): 26.4
Rel. Err. (%): 17.9
Time: 10.1s

Original image

Getting even more from less (and doing it efficiently)

Resolution: 256×256

Random Bernoulli	Multilevel Hadamard	Original image
		
RAM (GB): 4.8 Speed (it/s): 1.31 Rel. Err. (%): 22.4 Time: 4m27s	RAM (GB): < 0.1 Speed (it/s): 18.1 Rel. Err. (%): 14.7 Time: 18.6s	

Getting even more from less (and doing it efficiently)

Resolution: 512×512

Random Bernoulli	Multilevel Hadamard	Original image
		
RAM (GB): 76.8 Speed (it/s): 0.15 Rel. Err. (%): 19.0 Time: 42m	RAM (GB): < 0.1 Speed (it/s): 4.9 Rel. Err. (%): 12.2 Time: 1m13s	

Bernoulli only possible on the Xeon 256 GB RAM.

Getting even more from less (and doing it efficiently)

Resolution: 1024×1024

Random Bernoulli

?

RAM (GB): 1229
Speed (it/s): 0.0161
Rel. Err. (%): ?
Time: 6h36m

Multilevel Hadamard

RAM (GB): < 0.1
Speed (it/s): 1.07
Rel. Err. (%): 10.4
Time: 3m45s

Original image

Bernoulli not possible. Grey values are extrapolated.

Getting even more from less (and doing it efficiently)

Resolution: 2048×2048

Random Bernoulli

?

RAM (GB): 19661
Speed (it/s): 1.78e-3
Rel. Err. (%): ?
Time: 2d14h

Multilevel Hadamard

RAM (GB): < 0.1
Speed (it/s): 0.17
Rel. Err. (%): 8.5
Time: 28m

Original image

Bernoulli not possible. Grey values are extrapolated.

Getting even more from less (and doing it efficiently)

Resolution: 4096×4096

Random Bernoulli

RAM (GB): 314,573
Speed (it/s): 1.98e-4
Rel. Err. (%): ?
Time: 25d1h

Multilevel Hadamard

RAM (GB): < 0.1
Speed (it/s): 0.041
Rel. Err. (%): 6.6
Time: 1h37m

Original image

Bernoulli not possible. Grey values are extrapolated.

Getting even more from less (and doing it efficiently)

Resolution: 8192×8192

Random Bernoulli	Multilevel Hadamard	Original image
?		
RAM (GB): 5,033,165 Speed (it/s): 2.19e-5 Rel. Err. (%): ? Time: 238d1h	RAM (GB): < 0.1 Speed (it/s): 0.0064 Rel. Err. (%): 3.5 Time: 8h30m	

Bernoulli not possible. Grey values are extrapolated.

Comparison with other structured CS algorithms

Multilevel subsampling with Hadamard matrices

- Use standard recovery algorithm (l^1 minimization)
- Exploit structure in the sampling process

Other structured CS algorithms

- Use standard sensing matrices (random Gaussian/Bernoulli)
- Exploit structure in the recovery algorithm

Model-based CS: Baraniuk et al. (2010)

- Modification of CoSaMP based on connected tree structure

Turbo AMP: Som & Schniter (2012)

- Based on Bayesian CS – Ji, Xue & Carin (2008), He & Carin (2009)
- Modification of iterative thresholding algorithms

Comparison: 12.5% sampling at 256×256 resolution

Original

Model-CS, Err = 21.2%

TurboAMP, Err = 17.5%

Multilevel, Err = 7.1%

Comparison: 12.5% sampling at 256×256 resolution

Original

Model-CS, Err = 17.9%

TurboAMP, Err = 17.7%

Multilevel, Err = 8.8%

Outline

Compressed sensing

The need for a new theory

A new framework for compressed sensing

Getting even more from less

Conclusions

Conclusions

1. Standard compressed sensing is based on sparsity, incoherence and uniform random subsampling. However, in many applications these are not present.

2. We have introduced a new theory based on more realistic concepts: sparsity in levels, local coherence and multilevel random subsampling.

3. This provides the first theoretical explanation for the success of compressed sensing in many inverse problem applications.

4. Moreover, asymptotic incoherence allows one to exploit structure. This can be done with efficient sensing matrices, allowing one to get better reconstructions and go to much higher resolutions.

Paper:

- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: A new theory for compressed sensing.* Preprint on the arXiv, 2014.