Beyond incoherence and beyond sparsity

Compressed sensing in the real world

Ben Adcock
Department of Mathematics
Purdue University

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

Compressed sensing and -lets

Natural images are highly compressible in -lets:

Image

1.5% wavelet compressed image

$\Rightarrow$ -lets are widely used sparsifying transforms in compressed sensing for imaging problems.

Compressed sensing setup

Standard setup:

- $x = (x_1, \dots, x_N)^\top \in \mathbb{C}^N$ be an unknown image
- $A \in \mathbb{C}^{m \times N}$ is a sensing matrix
- $y = Ax + e$ are measurements, where $\|e\|_2 \leq \eta$
- $\Phi \in \mathbb{C}^{N \times N}$ is an orthonormal sparsifying transformation

Reconstruction algorithm:

$$\min_{z \in \mathbb{C}^N} \|\Phi^* z\|_1 \quad \text{subject to } \|Az - y\|_2 \leq \eta. \qquad (\star)$$

Recovery guarantees: Subject to appropriate conditions on A (e.g. RIP), if $\hat{x}$ is a minimizer of $(\star)$ and $\sigma_s(\cdot)$ is the best s-term approximation error,

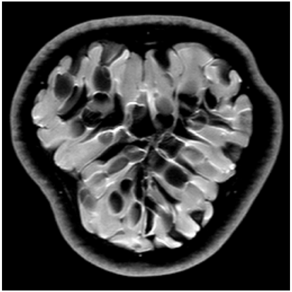
$$\|x - \hat{x}\|_2 \leq C_1 \frac{\sigma_s(\Phi^* x)}{\sqrt{s}} + C_2 \eta.$$

Sparsity is invariant under permutations

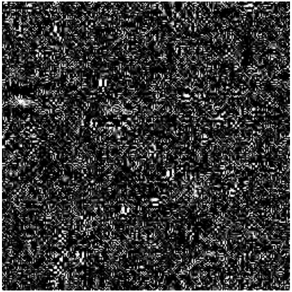
Let $P : \{1, \dots, N\} \to \{1, \dots, N\}$ be a permutation. Given $x \in \mathbb{C}^N$, define the permuted image

$$\tilde{x} = \Phi P \Phi^* x.$$

Example: CS reconstruction using DB4 wavelets and $m = 8192$ random Gaussian measurements with $N = 256 \times 256$



Original image x



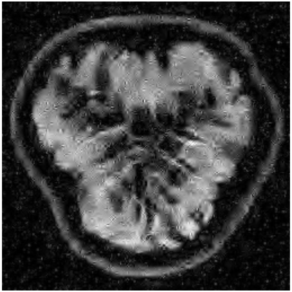
Permuted image $\tilde{x}$

Sparsity is invariant under permutations

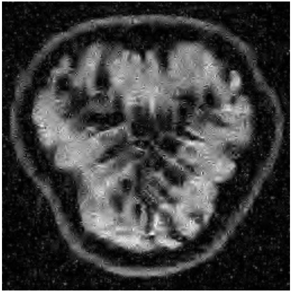
Let $P : \{1,\dots,N\} \to \{1,\dots,N\}$ be a permutation. Given $x \in \mathbb{C}^N$, define the permuted image

$$\tilde{x} = \Phi P \Phi^* x.$$

Example: CS reconstruction using DB4 wavelets and $m = 8192$ random Gaussian measurements with $N = 256 \times 256$



CS recon, Err=31.54%



Permuted CS recon, Err=31.51%

Beyond sparsity and incoherence

Conclusion: CS with incoherent measurements is suboptimal for natural images. It recovers all objects in the space

$$\Sigma_s = \left\{ x \in \mathbb{C}^N : \|\Phi^* x\|_0 \leq s \right\},$$

exactly, and this includes too many unphysical images.

New approach:

- Work with a smaller class of structured sparse objects.
- Don't use incoherent measurements. Modify the sensing matrix A to exploit structured sparsity.

Sparsity in levels

Wavelet bases can be decomposed into dyadic scales. Let

$$0 = M_0 < M_1 < \ldots < M_r = N,$$

be such that the k^{th} scale corresponds to indices $\{M_{k-1}+1, \ldots, M_k\}$. Define the sparsity in the k^{th} scale as

$$s_k = \left|\{j \in \{M_{k-1}+1, \ldots, M_k\} : c_j \neq 0\}\right|,$$

where $c = \Phi^* x$. Define the set of sparse in levels objects as

$$\Sigma_{\mathbf{s},\mathbf{M}} = \left\{x \in \mathbb{C}^N : |\{j \in \{M_{k-1}+1, \ldots, M_k\} : c_j \neq 0\}| \leq s_k\right\},$$

where $\mathbf{s} = (s_1, \ldots, s_r)$ and $\mathbf{M} = (M_1, \ldots, M_r)$.

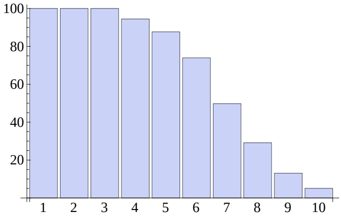
Objective: Find a sensing matrix A that promotes sparsity in levels when using l^1 minimization.

Natural images are asymptotically sparse in levels

We would a sensing matrix that will work for 'most' images. Fortunately, this is possible due to the property of asymptotic sparsity.

Wavelet coefficients are not just sparse, i.e. $s = \sum_{k=1}^{r} s_k \ll N$, but they are asymptotically sparse in levels, i.e. $x \in \Sigma_{\mathbf{s},\mathbf{M}}$, and

$$s_k/(M_k - M_{k-1}) \to 0, \quad k \to \infty.$$



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

Other work

Structured sparsity has been widely considered in CS.

- 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

Existing algorithms:

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

Although the particulars are different, these are based on similar ideas:

- Exploit the connected tree structure of wavelet coefficients
- Use standard measurements, e.g. random Gaussians
- Modify the recovery algorithm

However, there are issues with accuracy and efficiency. See later.

Designing measurements for sparsity in levels

Write $c = (c^{(1)}, \dots, c^{(r)})^\top$, where $c^{(k)}$ is the set of coefficients at the k^{th} scale. 'Ideal' measurements would be

$$y^{(k)} = B^{(k)} c^{(k)}, \qquad B^{(k)} \in \mathbb{C}^{m_k \times (M_k - M_{k-1})},$$

where $B^{(k)}$ is incoherent. Hence

$$B = A\Phi = \mathrm{diag}\left(B^{(1)}, \dots, B^{(r)}\right),$$

is block diagonal.

However, we are only allowed to design A, not B. Nevertheless, this suggests that good sensing matrices A for structured sparsity should yield approximate block diagonality of $B = A\Phi$ with incoherent blocks.

The Fourier transform of wavelets

Let $\phi_{l,k}$ be the Haar wavelet (for simplicity) with scale l and translation k. Then

$$|\mathcal{F}\phi_{l,k}(\omega)|^2 \lesssim 2^{-j}2^{-|j-l|}, \qquad 2^{j-1} \leq |\omega| \leq 2^j.$$

$\Rightarrow$ Wavelets give a natural division of Fourier space into dyadic bands

$$W_0 = \{0, 1\}, \quad W_j = \{-2^j + 1, \ldots, -2^{j-1}\} \cup \{2^{j-1} + 1, \ldots, 2^j\}.$$

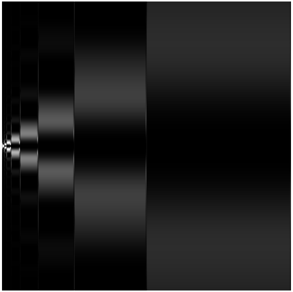
Only wavelets at scale $l \approx j$ have large spectrum within the band W_j.

Local incoherence of Fourier meaurements with wavelets

This means that the blocks $U^{(j,l)}$ of the matrix $U = \Psi^*\Phi$, where Ψ is the DFT, have the following structure:

- Diagonal incoherence: $\mu(U^{(j,j)}) \lesssim 2^{-j}$
- Exponential off-diagonal decay: $\mu(U^{(j,l)}) \lesssim 2^{-|j-l|}\mu(U^{(j,l)})$

Where $\mu(V) = \max_{i,j} |V_{ij}|^2$.



The absolute values of U

Multilevel random subsampling

Pick m_k indices from W_k uniformly at random:

$$\Omega_k \subseteq W_k, \qquad |\Omega_k| = m_k.$$

Let $A \in \mathbb{C}^{m \times N}$, where $m = \sum_k m_k$, be the subsampled DFT with rows corresponding to indices $\Omega_1 \cup \ldots \cup \Omega_r$.

Choosing the m_k's. One can show that to recover an $(\mathbf{s}, \mathbf{M})$-sparse image it suffices to take

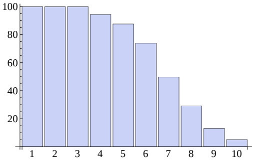
$$m_k \gtrsim s_k + \sum_{l \neq k} 2^{-\frac{|k-l|}{2}} s_l.$$

This is a consequence of a general theory for CS based on local coherence in levels, sparsity in levels and multilevel random subsampling:

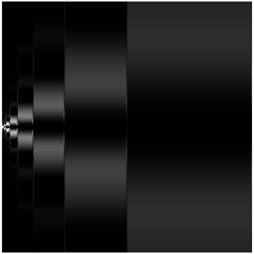
- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: a new theory for compressed sensing*, arXiv:1302.0561, 2013.

Multilevel random subsampling

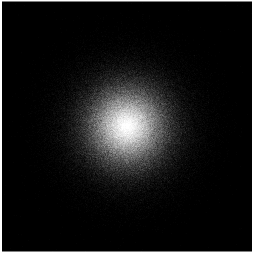
Image



Asymptotic sparsity of coefficients

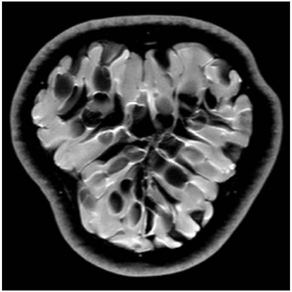


Matrix U

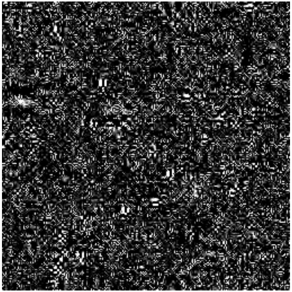


Subsampling map in 2D

Flip test revisited



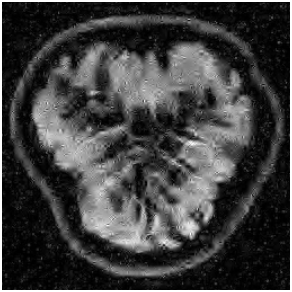
Original image x



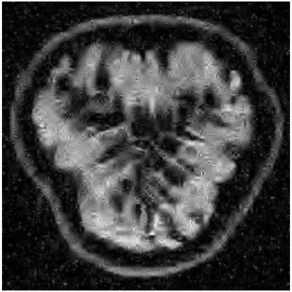
Permuted image $\tilde{x}$

Random Gaussian measurements: incoherent with wavelets, recover all sparse coefficients equally well.

Flip test revisited



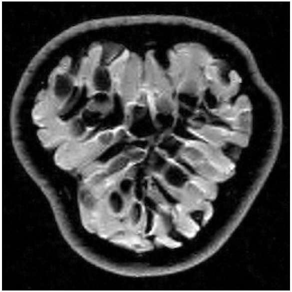
CS recon, Err=31.54%



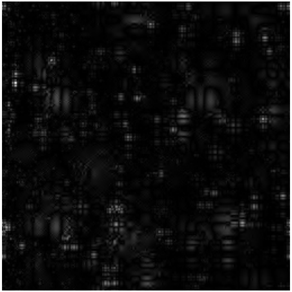
Permuted CS recon, Err=31.51%

Random Gaussian measurements: incoherent with wavelets, recover all sparse coefficients equally well.

Flip test revisited



CS recon, Err=10.96%



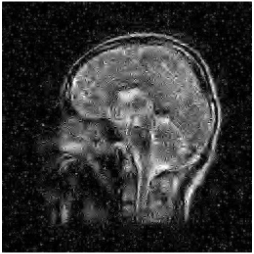
Permuted CS recon, Err=99.3%

Multilevel subsampled Fourier measurements: Asymptotically incoherent with wavelets, recover only asymptotically sparse coefficients.

Numerical example

Example: 12.5% measurements using DB4 wavelets.

256×256



Err = 41.6%

512×512

Err = 25.3%

1024×1024

Err = 11.6%

First case: Gaussian random measurements.

Numerical example

Example: 12.5% measurements using DB4 wavelets.

256×256

Err $= 21.9\%$
(41.6%)

512×512

Err $= 10.9\%$
(25.3%)

1024×1024

Err $= 3.1\%$
(11.6%)

Second case: Subsampled Fourier measurements.

Efficient compressive imaging

Example: The Berlin cathedral with 15% sampling at various resolutions using Daubechies-4 wavelets. Comparison between random Bernoulli and subsampled multilevel subsampled Hadamard measurements.

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

Efficient compressive imaging

Resolution: 128×128

Random Bernoulli	Hadamard	Original image
		
RAM (GB): 0.3 Speed (it/s): 12.4 Rel. Err. (%): 26.4 Time: 25s	RAM (GB): < 0.1 Speed (it/s): 26.4 Rel. Err. (%): 17.9 Time: 10.1s	

Efficient compressive imaging

Resolution: 256×256

Random Bernoulli

RAM (GB): 4.8
Speed (it/s): 1.31
Rel. Err. (%): 22.4
Time: 4m27s

Hadamard

RAM (GB): < 0.1
Speed (it/s): 18.1
Rel. Err. (%): 14.7
Time: 18.6s

Original image

Efficient compressive imaging

Resolution: 512×512

Random Bernoulli	Hadamard	Original image
		

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

Bernoulli only possible on the Xeon 256 GB RAM.

Efficient compressive imaging

Resolution: 1024×1024

Random Bernoulli	Hadamard	Original image
		
RAM (GB): 1229 Speed (it/s): 0.0161 Rel. Err. (%): ? Time: 6h36m	RAM (GB): < 0.1 Speed (it/s): 1.07 Rel. Err. (%): 10.4 Time: 3m45s	

Bernoulli not possible. Grey values are extrapolated.

Efficient compressive imaging

Resolution: 2048×2048

Random Bernoulli	Hadamard	Original image
?		
RAM (GB): 19661 Speed (it/s): 1.78e-3 Rel. Err. (%): ? Time: 2d14h	RAM (GB): < 0.1 Speed (it/s): 0.17 Rel. Err. (%): 8.5 Time: 28m	

Bernoulli not possible. Grey values are extrapolated.

Efficient compressive imaging

Resolution: 4096×4096

Random Bernoulli

?

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

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.

Efficient compressive imaging

Resolution: 8192×8192

Random Bernoulli

?

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

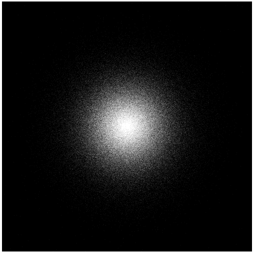
Bernoulli not possible. Grey values are extrapolated.

Example with other -lets

Example: 6.25% subsampling at 2048×2048 resolution. Comparing wavelets, curvelets, contourlets and shearlets.

2048×2048 image

256×256 crop



subsampling map

Note: The sampling pattern is not optimized to the sparsifying transformation.

Example with other -lets

wavelets

curvelets

contourlets

shearlets

Comparison with other structured CS algorithms

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. Fourier/Hadamard

Other structured CS algorithms

- E.g. model-CS, Bayesian CS, Turbo AMP
- Use standard sensing matrices (random Gaussian/Bernoulli)
- Exploit connected tree structure by modifying the recovery algorithm

Comparison: 12.5% sampling at 256×256 resolution

Original

ℓ^1 Gauss., Err $= 15.7\%$

Model-CS, Err $= 17.9\%$

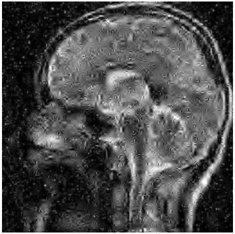
BCS, Err $= 12.1\%$

TurboAMP, Err $= 17.7\%$

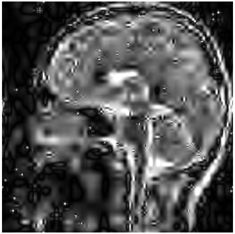
Mult. Four., Err $= 8.8\%$

Comparison: 12.5% sampling at 256×256 resolution

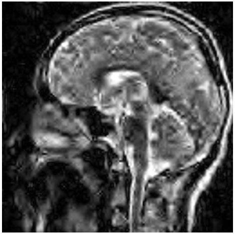
Original



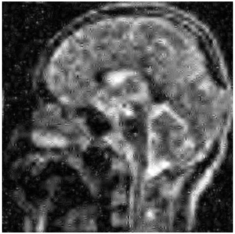
ℓ^1 Bern., Err $= 41.2\%$



Model-CS, Err $= 41.8\%$



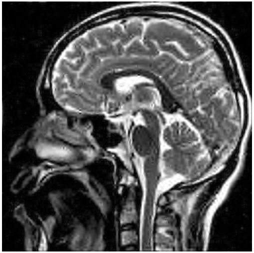
BCS, Err $= 29.6\%$



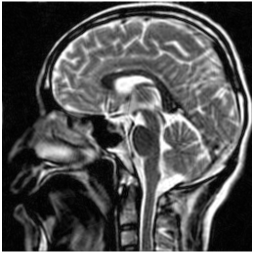
TurboAMP, Err $= 39.3\%$

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

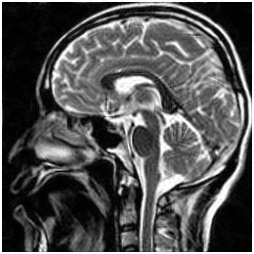
Multilevel Fourier with other sparsifying transformations



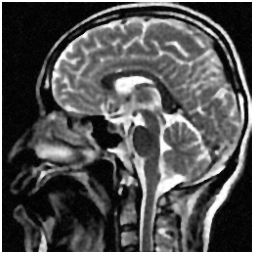
wavelets, Err $= 18.2\%$



curvelets, Err $= 17.4\%$



shearlets, Err $= 16.5\%$



TV, Err $= 17.6\%$

Summary

1. Standard CS using incoherent sensing matrices, e.g. random Gaussians, is highly suboptimal for imaging with -lets.

2. Images are not just sparse, but always possess a distinct asymptotic sparsity in levels structure.

3. Such structure can be exploited using multilevel subsampling of Fourier/Hadamard matrices.

4. These matrices are not incoherent with wavelets, but have a distinct asymptotic incoherence structure.

5. By doing so, one obtains substantial improvements in accuracy and computational efficiency over standard CS, and also outperforms other structured CS algorithms.

Final remarks

The majority of CS theory is based on sparsity and incoherence. This work suggests more general concepts of sparsity in levels and local coherence in levels are better suited in applications involving -lets.

- Moreover, these concepts naturally arise in many CS applications, due to the specific measurements (e.g. medical imaging, microscopy,...).

A new CS theory: based on sparsity in levels, local coherence and multilevel random subsampling. See

- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: a new theory for compressed sensing*, arXiv:1302.0561, 2013.
- As a corollary, provides the comprehensive recovery guarantees for CS in the above applications.

Open problems: Take your favourite CS concept, e.g. RIP, instance optimality, phase transitions, iterative algorithms,...., and generalize it to sparsity with levels. Also, optimal sampling strategies, optimal measurements, structured sampling + structured recovery,...