Getting even more from less

Structure-exploiting sampling strategies for compressive imaging

Ben Adcock
Department of Mathematics
Purdue University

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

Compressed sensing problems

Roughly speaking, compressed sensing problems fall into two categories:

Type I. Imposed sensing operators. Measurements are specified by the physical sensor.

- E.g. MRI, X-ray CT, electron microscopy, seismology, radio interferometry,...

Type II. Designed sensing operators. The sensing mechanism allows substantial freedom to design measurements so as to get the best CS reconstruction.

- E.g. compressive imaging (single pixel camera, lensless imaging), fluorescence microscopy,...
- The only constraint is the measurement matrix should be binary.

Summary of the previous talk

1. For type I problems, there is high global coherence between the measurements (e.g. Fourier) and the sparsifying system (e.g. wavelets).

2. However, there is asymptotic incoherence, which can be exploited by multilevel random subsampling.

3. Wavelet coefficients are not just sparse, but asymptotically sparse in levels.

4. Multilevel random subsampling recovers such coefficients using near-optimal numbers of measurements. This is why CS works in type I applications such as MRI, CT etc.

5. The flip test shows that the recovery quality depends crucially on this sparsity structure.

This talk

Goal: Show that the same principles also lead to improved CS approaches for type II problems whenever the sparsifying transform Φ is

- A wavelet, curvelet, shearlet, etc-let transform,
- Total variation.

The CS gospel

Gospel: *Random Gaussian/Bernoulli measurements are 'optimal' for CS.*

Near-optimal recovery guarantees: Taking

$$m \gtrsim s \log(N/s),$$

measurements ensures exact recovery of all s-sparse vectors with high probability.

- The term $s \log(N/s)$ is optimal.

Universality: Random (sub)Gaussians are also highly desirable since the recovery guarantees are independent of the sparsifying transformation.

The CS gospel

Gospel: *Random Gaussian/Bernoulli measurements are 'optimal' for CS.*

Near-optimal recovery guarantees: Taking

$$m \gtrsim s \log(N/s),$$

measurements ensures exact recovery of all s-sparse vectors with high probability.

- The term $s \log(N/s)$ is optimal.

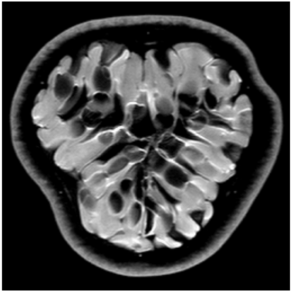
Universality: Random (sub)Gaussians are also highly desirable since the recovery guarantees are independent of the sparsifying transformation.

Sparsity is invariant under permutations

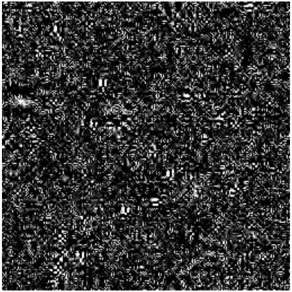
Let $P : \{1, \dots, N\} \to \{1, \dots, N\}$ be a permutation. Given $x \in \mathbb{C}^N$, define the permuted image

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

Example: CS reconstruction using DB4 wavelets, ℓ^1 minimization and $m = 8192$ random Gaussian measurements with $N = 256 \times 256$.



Original image x



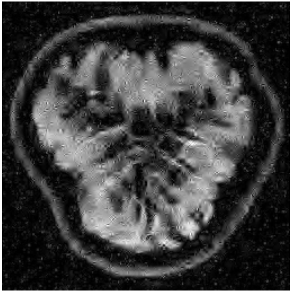
Permuted image $\tilde{x}$

Sparsity is invariant under permutations

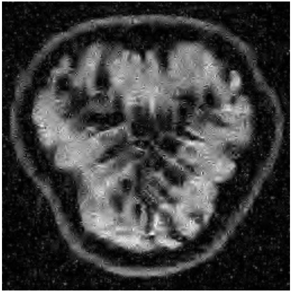
Let $P : \{1, \dots, N\} \to \{1, \dots, N\}$ be a permutation. Given $x \in \mathbb{C}^N$, define the permuted image

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

Example: CS reconstruction using DB4 wavelets, ℓ^1 minimization and $m = 8192$ random Gaussian measurements with $N = 256 \times 256$.

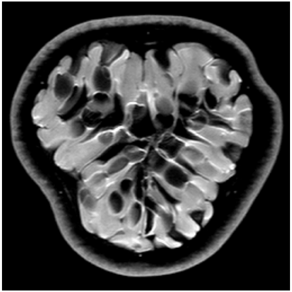


CS recon, Err=31.54%

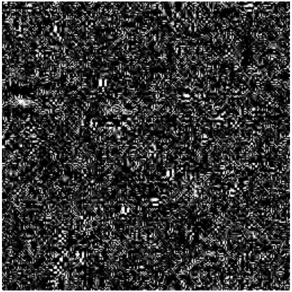


Permuted CS recon, Err=31.51%

Conclusion



Physical x



Unphysical $\tilde{x}$

CS with random Gaussian, or in general, incoherent measurements, is suboptimal for 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

Idea:

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

Choosing an appropriate structure. Some options:

- Linear+sparse (two-level)
- Asymptotic sparsity (multilevel)
- Connected trees
- .....

Need to balance efficiency of the structured representation with ease of measurement design and computational feasibility (cost and storage) of the sensing matrix.

Asymptotic sparsity in levels

For vectors $c \in \mathbb{C}^N$, let

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

denote r sparsity levels.

Definition

A vector $c \in \mathbb{C}^N$ is $(\mathbf{s}, \mathbf{M})$-sparse if

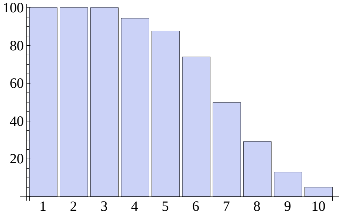
$$s_k = \left| \{ j \in \{ M_{k-1} + 1, \ldots, M_k \} : c_j \neq 0 \} \right|, \quad k = 1, \ldots, r.$$

Loosely speaking, $c \in \mathbb{C}^N$ is asymptotically sparse in levels if

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

Images are asymptotically sparse in levels

Suppose the levels are taken as the wavelet scales.



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

$\Rightarrow$ It is possible to find measurements that work well for 'most' images.

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, where $A \in \mathbb{C}^{m \times N}$ is the sensing matrix and Φ is the wavelet transform.

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.

Using Type I measurements for Type II problems

Type I insight: Fourier measurements with multilevel subsampling recover asymptotically sparse -let coefficients.

- If binary measurements are required, use the Hadamard transform instead, subsampled in a similar way.

Why Fourier measurement promote sparsity in levels

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

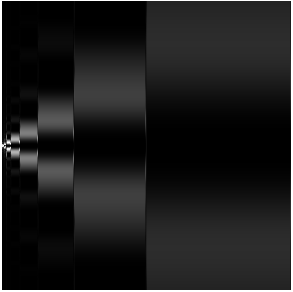
$$W_0 = \{0,1\}, \quad W_j = \{-2^j+1,\ldots,-2^{j-1}\} \cup \{2^{j-1}+1,\ldots,2^j\}.$$

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

Local incoherence of Fourier meaurements with wavelets

Let $U = F\Phi$ be the Fourier/wavelets matrix. Then U has a dyadic block structure:

$$U = \{U^{(j,l)}\}_{j,l=1}^{r}, \qquad U^{(j,l)} \in \mathbb{C}^{2^j \times 2^l}.$$



If $\mu(\cdot)$ denotes the coherence, then we have

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

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 $\Omega_1 \cup \ldots \cup \Omega_r$.

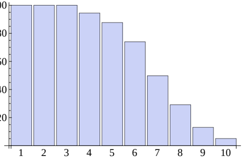
Choosing the m_k's. For Haar wavelets, one can show that to recover an $(\mathbf{s}, \mathbf{M})$-sparse vector it suffices to take

$$m_k \gtrsim \left(s_k + \sum_{l \neq k} (\sqrt{2})^{-|k-l|} s_l \right) \log(N).$$

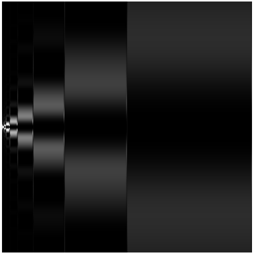
- Extends to general wavelets with $\sqrt{2} \to A$, where $A > 1$ depends on the smoothness and number of vanishing moments.

Multilevel random subsampling

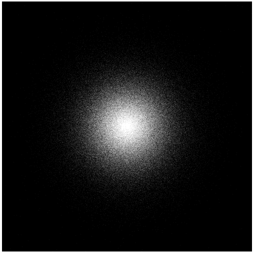
Image



Asymptotic sparsity of coefficients



Matrix $U = F\Phi$

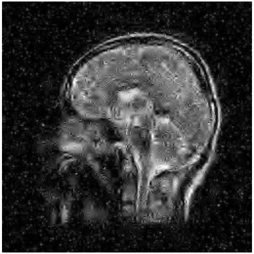


Subsampling map in 2D

Numerical example

Example: 12.5% measurements using DB4 wavelets.

256×256



$\text{Err} = 41.6\%$

512×512

$\text{Err} = 25.3\%$

1024×1024

$\text{Err} = 11.6\%$

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: Multilevel 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	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	

Efficient compressive imaging

Resolution: 512×512

Random Bernoulli

RAM (GB): 76.8
Speed (it/s): 0.15
Rel. Err. (%): 19.0
Time: 42m

Hadamard

RAM (GB): < 0.1
Speed (it/s): 4.9
Rel. Err. (%): 12.2
Time: 1m13s

Original image

Bernoulli only possible on the Xeon 256 GB RAM.

Efficient compressive imaging

Resolution: 1024×1024

Random Bernoulli

?

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

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.

Efficient compressive imaging

Resolution: 2048×2048

Random Bernoulli

?

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

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.

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

?

RAM (GB): 5,033,165
Speed (it/s): 2.19e-5
Rel. Err. (%): ?
Time: 238d1h

Hadamard

RAM (GB): < 0.1
Speed (it/s): 0.0064
Rel. Err. (%): 3.5
Time: 8h30m

Original image

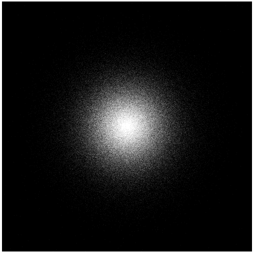
Bernoulli not possible. Grey values are extrapolated.

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

Other structured CS algorithms

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)

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. Fourier/Hadamard

Other structured CS algorithms

- 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 $= 17.9\%$

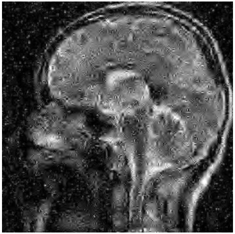
BCS, Err $= 12.1\%$

TurboAMP, Err $= 17.7\%$

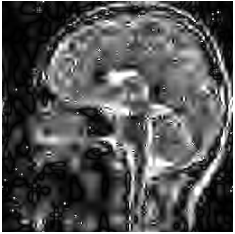
Mult. Four., Err $= 8.8\%$

Comparison: 12.5% sampling at 256×256 resolution

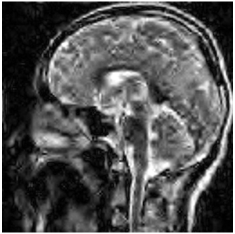
Original



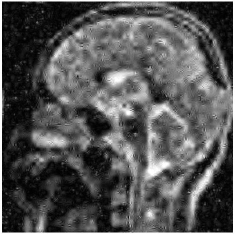
ℓ^1 Bern., Err $= 41.2\%$



Model-CS, Err $= 41.8\%$



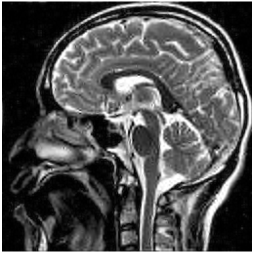
BCS, Err $= 29.6\%$



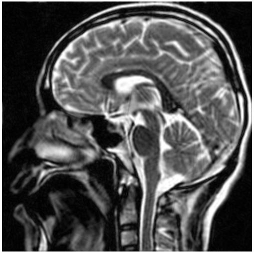
TurboAMP, Err $= 39.3\%$

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

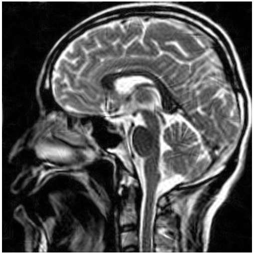
Multilevel Fourier with other sparsifying transformations



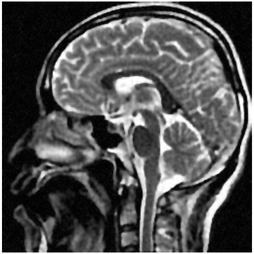
wavelets, Err $= 18.2\%$



curvelets, Err $= 17.4\%$



shearlets, Err $= 16.5\%$



TV, Err $= 17.6\%$

Remarks

Multilevel Fourier/Hadamard:

- substantial error reduction
- lower cost/storage
- uses black box optimization solvers
- East incorporation of other sparsifying transformations

Other approaches:

- Leverage more structure (connected trees)
- Asymptotically as $N \to \infty$, a better model for images
- However, for finite N may be too restrictive
- Algorithms used are less forgiving to model mismatch than l^1

Agnosticism of ℓ^1 minimization

The sampling levels $N_1, \ldots, N_r$ and number of samples $m_1, \ldots, m_r$ are chosen by the user. The sparsity levels $M_1, \ldots, M_r$ and sparsities $s_1, \ldots, s_r$ do not need to be specified.

Main estimates. Suppose that

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

and if $m_k \gtrsim \tilde{m}_k \cdot (\log(\epsilon^{-1}) + 1) \cdot \log(N)$, where $\tilde{m}_k$ satisfies

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

Then, in the absence of noise (for simplicity),

$$\|\hat{x} - x\| \lesssim \sigma_{\mathbf{s},\mathbf{M}}(x).$$

Agnosticism of ℓ^1 minimization

For the purposes of analysis, we choose the levels $M_1, \ldots, M_r$ to be wavelet scales. If $s_1, \ldots, s_r$, then the condition

$$\tilde{m}_k \gtrsim s_k + \sum_{l \neq k} A^{-|k-l|} s_l, \quad k = 1, \ldots, r.$$

This implies (1) and (2).

But this choice may not give the smallest $(\mathbf{s}, \mathbf{M})$-term approximation error $\sigma_{\mathbf{s},\mathbf{M}}(x)$. Other pairs $(\mathbf{s}', \mathbf{M}')$ may give smaller approximation errors.

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.

Open problems

The majority of CS is based on sparsity and incoherence. This work suggests that sparsity in levels and local coherence in levels are better suited in many cases.

- In type I problems (e.g. MRI), they explain why CS works.
- In type II problems (e.g. compressive imaging), they give substantial improvements over standard and structured CS algorithms.

Future work: Take your favourite CS concept and generalize it to sparsity with levels. E.g.

- RIP, instance optimality
- phase transitions
- iterative algorithms
- ....

References

- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: a new theory for compressed sensing,* arXiv:1302.0561, 2013.
- BA, Hansen & Roman, *The quest for optimal sampling: computationally efficient, structure-exploiting measurements for compressed sensing,* arXiv:1403.6540, 2014.
- BA, Hansen & Roman, *A note on compressed sensing of structured sparse wavelet coefficients from subsampled Fourier measurements,* arXiv:1403.6541, 2014.
- 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, *Generalized sampling and infinite-dimensional compressed sensing,* Technical Report, 2011.

Thank you!