Breaking the coherence barrier - A new theory for compressed sensing

Anders C. Hansen (Cambridge)

Joint work with:

B. Adcock (Simon Fraser University)
C. Poon (Cambridge)
B. Roman (Cambridge)

UCL-Duke Workshop, September 4, 2014

Compressed Sensing in Inverse Problems

Typical analog/infinite-dimensional inverse problem where compressed sensing is/can be used:

(i) Magnetic Resonance Imaging (MRI)
(ii) X-ray Computed Tomography
(iii) Thermoacoustic and Photoacoustic Tomography
(iv) Single Photon Emission Computerized Tomography
(v) Electrical Impedance Tomography
(vi) Electron Microscopy
(vii) Reflection seismology
(viii) Radio interferometry
(ix) Fluorescence Microscopy

Most of these problems are modelled by the Fourier transform

$$\mathcal{F}f(\omega) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \omega \cdot x}\, dx,$$

or the Radon transform $\mathcal{R}f : \mathbf{S} \times \mathbb{R} \to \mathbb{C}$ (where $\mathbf{S}$ denotes the circle)

$$\mathcal{R}f(\theta, p) = \int_{\langle x, \theta \rangle = p} f(x)\, dm(x),$$

where dm denotes Lebesgue measure on the hyperplane $\{x : \langle x, \theta \rangle = p\}$.

- Fourier slice theorem $\Rightarrow$ both problems can be viewed as the problem of reconstructing f from pointwise samples of its Fourier transform.

$$g = \mathcal{F}f, \quad f \in L^2(\mathbb{R}^d). \tag{1}$$

Compressed Sensing

- Given the linear system

$$Ux_0 = y.$$

- Solve

$$\min \|z\|_1 \quad \text{subject to } P_\Omega U z = P_\Omega y,$$

where P_Ω is a projection and $\Omega \subset \{1, \ldots, N\}$ is subsampled with $|\Omega| = m$.

If

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

then $\mathbb{P}(z = x_0) \geq 1 - \epsilon$, where

$$\mu(U) = \max_{i,j} |U_{i,j}|^2$$

is referred to as the incoherence parameter.

Pillars of Compressed Sensing

- Sparsity
- Incoherence
- Uniform Random Subsampling

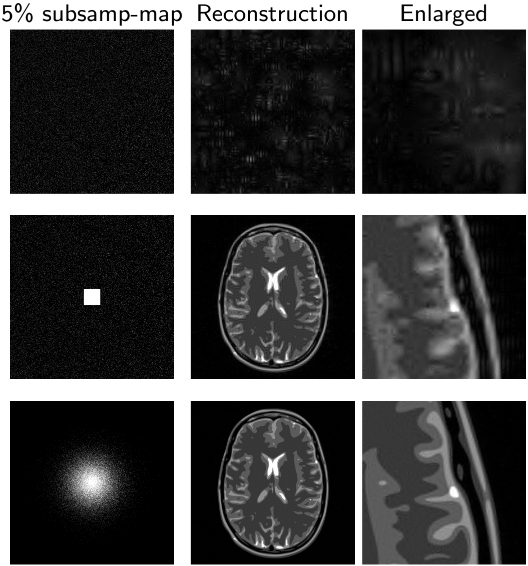
In addition: The Restricted Isometry Property + uniform recovery.

Problem: These concepts are absent in virtually all the problems listed above. Moreover, uniform random subsampling gives highly suboptimal results.

Compressed sensing is currently used with great success in many of these fields, however the current theory does not cover this.

Uniform Random Subsampling

$$U = U_{\mathrm{dft}} V_{\mathrm{dwt}}^{-1}.$$

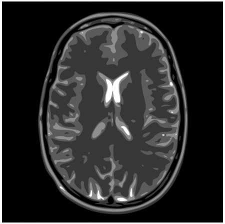


Sparsity

- The classical idea of sparsity in compressed sensing is that there are s important coefficients in the vector x_0 that we want to recover.
- The location of these coefficients is arbitrary.

Sparsity and the Flip Test

Let

$x =$ 

and

$$y = U_{\mathrm{df}} x, \qquad A = P_\Omega U_{\mathrm{df}} V_{\mathrm{dw}}^{-1},$$

where P_Ω is a projection and $\Omega \subset \{1, \dots, N\}$ is subsampled with $|\Omega| = m$. Solve

$$\min \|z\|_1 \quad \text{subject to } Az = P_\Omega y.$$

Sparsity - The Flip Test

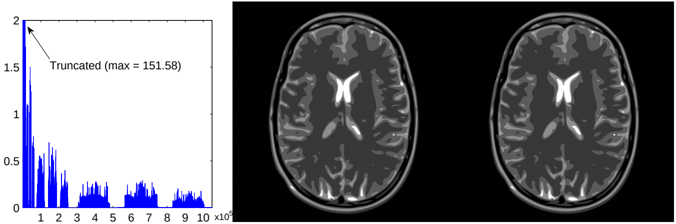
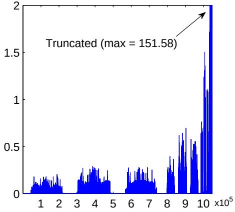


Figure : Wavelet coefficients and subsampling reconstructions from 10% of Fourier coefficients with distributions $(1+\omega_1^2+\omega_2^2)^{-1}$ and $(1+\omega_1^2+\omega_2^2)^{-3/2}$.

If sparsity is the right model we should be able to flip the coefficients. Let

$z_f =$



Sparsity - The Flip Test

- Let
$$\tilde{y} = U_{\mathrm{df}} V_{\mathrm{dw}}^{-1} z_f$$
- Solve
$$\min \|z\|_1 \quad \text{subject to } Az = P_\Omega \tilde{y}$$
  to get $\hat{z}_f$.
- Flip the coefficients of $\hat{z}_f$ back to get $\hat{z}$, and let $\hat{x} = V_{\mathrm{dw}}^{-1} \hat{z}$.
- If the ordering of the wavelet coefficients did not matter i.e. sparsity is the right model, then $\hat{x}$ should be close to x.

Sparsity- The Flip Test: Results

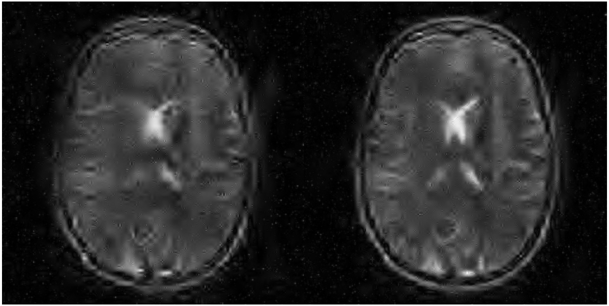


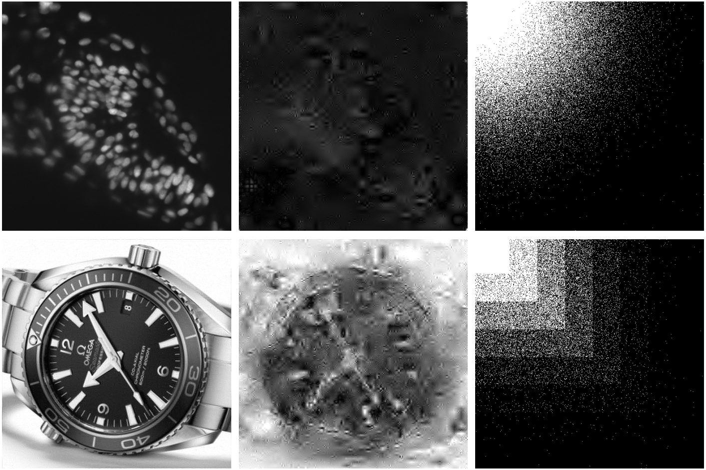
Figure : The reconstructions from the reversed coefficients.

Conclusion: The ordering of the coefficients did matter. Moreover, this phenomenon happens with all wavelets, curvelets, contourlets and shearlets and any reasonable subsampling scheme.

Question: Is sparsity really the right model?

Sparsity - The Flip Test

	CS reconstr.	CS reconstr, w/ flip coeffs.	Subsampling pattern
512, 20% $U_{\text{Had}} V_{\text{dwt}}^{-1}$ Fluorescence Microscopy			
1024, 12% $U_{\text{Had}} V_{\text{dwt}}^{-1}$ Compressive Imaging, Hadamard Spectroscopy			



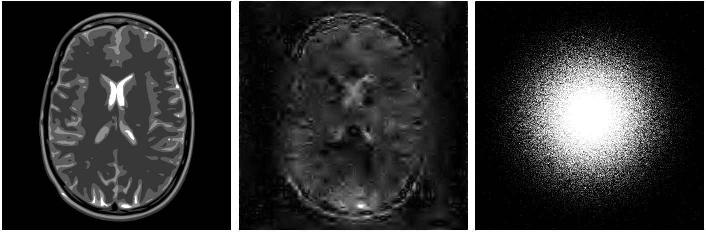
Sparsity - The Flip Test (contd.)

CS reconstr. | CS reconstr, w/ flip coeffs. | Subsampling pattern

1024, 20%

$U_{\mathrm{dft}} V_{\mathrm{dwt}}^{-1}$

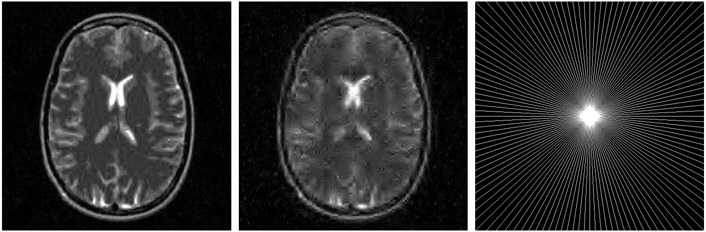
Magnetic Resonance Imaging



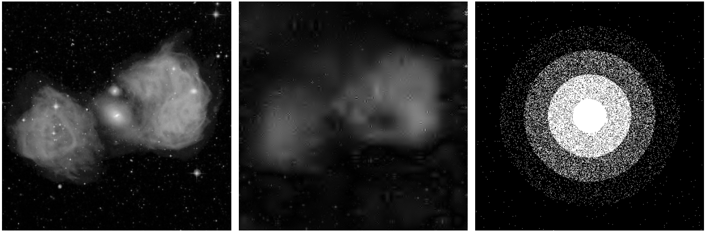
512, 12%

$U_{\mathrm{dft}} V_{\mathrm{dwt}}^{-1}$

Tomography, Electron Microscopy



Sparsity - The Flip Test (contd.)

	CS reconstr.	CS reconstr, w/ flip coeffs.	Subsampling pattern
1024, 10% $U_{\mathrm{dft}} V_{\mathrm{dwt}}^{-1}$ Radio interferometry			

Lesson Learned!

- The optimal sampling strategy depends on the sparsity structure of the signal.
- Compressed sensing theorems must therefore show how the optimal sampling strategy will depend on the structure.
- Such theorems currently do not exists.

What about the RIP?

- Did any of the matrices used in the examples satisfy the RIP?

How to measure asymptotic sparsity: Suppose

$$f = \sum_{j=1}^{\infty} \beta_j \varphi_j.$$

Let

$$\mathbb{N} = \bigcup_{k\in\mathbb{N}} \{M_{k-1}+1, \dots, M_k\},$$

where $0 = M_0 < M_1 < M_2 < \dots$ and $\{M_{k-1}+1, \dots, M_k\}$ is the set of indices corresponding to the k^{th} scale.
Let $\epsilon \in (0,1]$ and let

$$s_k := s_k(\epsilon) = \min\left\{K : \left\| \sum_{i=1}^{K} \beta_{\pi(i)}\varphi_{\pi(i)} \right\| \geq \epsilon \left\| \sum_{i=M_{k-1}+1}^{M_k} \beta_j \varphi_j \right\| \right\},$$

in order words, s_k is the effective sparsity at the k^{th} scale. Here $\pi : \{1, \dots, M_k - M_{k-1}\} \to \{M_{k-1}+1, \dots, M_k\}$ is a bijection such that $|\beta_{\pi(i)}| \geq |\beta_{\pi(i+1)}|$.

Images are not sparse, they are asymptotically sparse

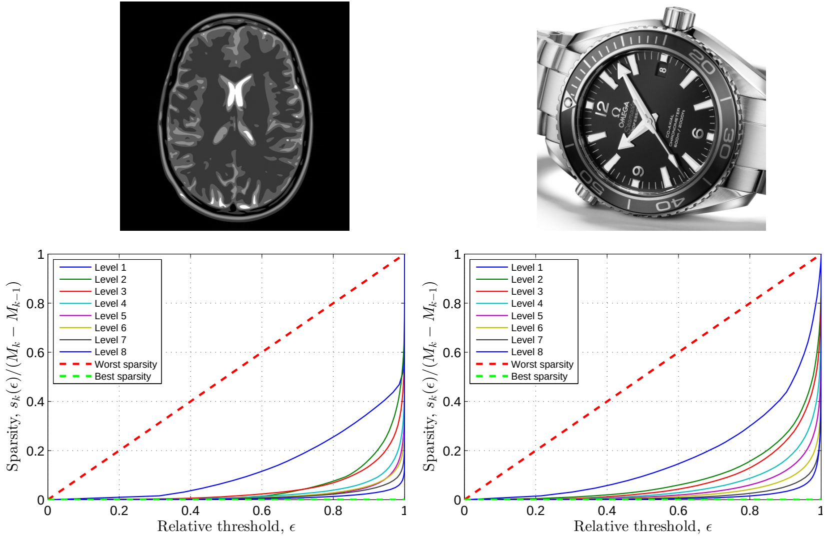


Figure : Relative sparsity of Daubechies 8 wavelet coefficients.

Analog inverse problems are coherent

Let

$$U_n = U_{\mathrm{df}} V_{\mathrm{dw}}^{-1} \in \mathbb{C}^{n \times n}$$

where U_{df} is the discrete Fourier transform and V_{dw} is the discrete wavelet transform. Then

$$\mu(U_n) = 1$$

for all n and all Daubechies wavelets!

Analog inverse problems are coherent, why?

Note that

$$\operatorname*{WOT-lim}_{n\to\infty} \; U_{\mathrm{df}} V_{\mathrm{dw}}^{-1} = U,$$

where

$$U = \begin{pmatrix} \langle \varphi_1, \psi_1 \rangle & \langle \varphi_2, \psi_1 \rangle & \cdots \\ \langle \varphi_1, \psi_2 \rangle & \langle \varphi_2, \psi_2 \rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Thus, we will always have

$$\mu(U_{\mathrm{df}} V_{\mathrm{dw}}^{-1}) \geq c.$$

Analog inverse problems are asymptotically incoherent

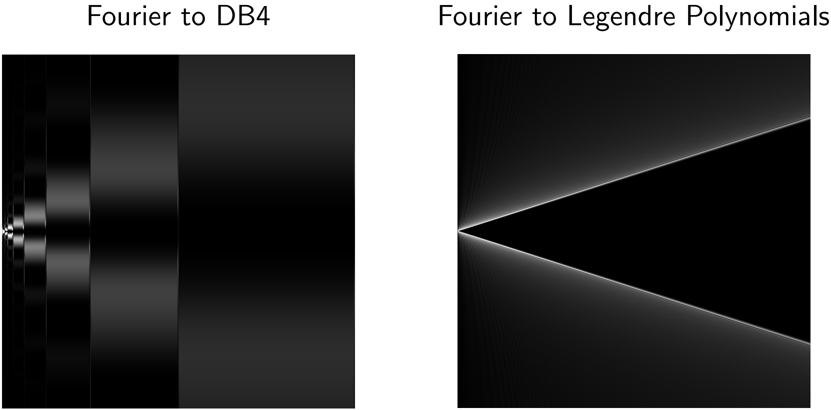


Figure : Plots of the absolute values of the entries of the matrix U

We need a new theory

- Such theory must incorporates asymptotic sparsity and asymptotic incoherence.
- It must explain the two intriguing phenomena observed in practice:
  - The optimal sampling strategy is signal structure dependent
  - The success of compressed sensing is resolution dependent
- The theory cannot be RIP based (at least not with the classical definition of the RIP)

New Pillars of Compressed Sensing

- Asymptotic Sparsity
- Asymptotic Incoherence
- Multi-level Subsampling

Definition

For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ with $1 \leq M_1 < \ldots < M_r$ and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \ldots, r$, where $M_0 = 0$. We say that $\beta \in l^2(\mathbb{N})$ is $(\mathbf{s}, \mathbf{M})$-sparse if, for each $k = 1, \ldots, r$,

$$\Delta_k := \mathrm{supp}(\beta) \cap \{M_{k-1} + 1, \ldots, M_k\},$$

satisfies $|\Delta_k| \leq s_k$. We denote the set of $(\mathbf{s}, \mathbf{M})$-sparse vectors by $\Sigma_{\mathbf{s},\mathbf{M}}$.

Definition

Let $f = \sum_{j\in\mathbb{N}} \beta_j \varphi_j \in \mathcal{H}$, where $\beta = (\beta_j)_{j\in\mathbb{N}} \in l^1(\mathbb{N})$. Let

$$\sigma_{\mathbf{s},\mathbf{M}}(f) := \min_{\eta\in\Sigma_{\mathbf{s},\mathbf{M}}} \|\beta - \eta\|_{l^1}. \tag{2}$$

Definition

Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \ldots < N_r$, $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$, with $m_k \leq N_k - N_{k-1}$, $k = 1, \ldots, r$, and suppose that

$$\Omega_k \subseteq \{N_{k-1}+1, \ldots, N_k\}, \quad |\Omega_k| = m_k, \quad k = 1, \ldots, r,$$

are chosen uniformly at random, where $N_0 = 0$. We refer to the set

$$\Omega = \Omega_{\mathbf{N},\mathbf{m}} := \Omega_1 \cup \ldots \cup \Omega_r.$$

as an $(\mathbf{N}, \mathbf{m})$-multilevel sampling scheme.

Definition

Let $U \in \mathbb{C}^{N\times N}$. If $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ and $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots N_r$ and $1 \leq M_1 < \dots < M_r$ we define the $(k, l)^{\mathrm{th}}$ local coherence of U with respect to $\mathbf{N}$ and $\mathbf{M}$ by

$$\mu_{\mathbf{N},\mathbf{M}}(k, l) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}}) \cdot \mu(P_{N_k}^{N_{k-1}} U)}, \quad k, l = 1, \dots, r,$$

where $N_0 = M_0 = 0$.

The optimization problem

$$\inf_{\eta \in \ell^1(\mathbb{N})} \|\eta\|_{\ell^1} \text{ subject to } \|P_\Omega U \eta - y\| \leq \delta. \tag{3}$$

Theoretical Results

Let $U \in \mathbb{C}^{N \times N}$ be an isometry and $\beta \in \mathbb{C}^N$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme, where $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ and $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$. Let $(\mathbf{s}, \mathbf{M})$, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_1 < \ldots < M_r$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, be any pair such that the following holds: for $\epsilon > 0$ and $1 \leq k \leq r$,

$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r} \mu_{\mathbf{N},\mathbf{M}}(k, l) \cdot s_l \right) \cdot \log(N). \tag{4}$$

Suppose that $\xi \in \mathbb{C}^N$ is a minimizer of (3) with $\delta = \tilde{\delta}\sqrt{K^{-1}}$ and $K = \max_{1 \leq k \leq r}\{(N_k - N_{k-1})/m_k\}$. Then, with probability exceeding $1 - s\epsilon$, where $s = s_1 + \ldots + s_r$, we have that

$$\|\xi - \beta\| \leq C \cdot \left(\tilde{\delta} \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f) \right),$$

for some constant C, where $\sigma_{\mathbf{s},\mathbf{M}}(f)$ is as in (2), $L = 1 + \frac{\sqrt{\log_2\left(6\epsilon^{-1}\right)}}{\log_2(4KM\sqrt{s})}$ and $K = \max_{k=1,\ldots,r}\left\{\frac{N_k - N_{k-1}}{m_k}\right\}$.

Fourier to wavelets

Number of samples in each level:

$$m_k \gtrsim \log(\epsilon^{-1}) \cdot \log(N) \cdot \frac{N_k - N_{k-1}}{N_{k-1}} \cdot \left(\hat{s}_k + \sum_{l=1}^{k-2} s_l \cdot 2^{-\alpha(k-l)} + \sum_{l=k+2}^{r} s_l \cdot 2^{-v(l-k)} \right),$$

where $\hat{s}_k = \max\{s_{k-1}, s_k, s_{k+1}\}$. Note that

$$m \gtrsim s \cdot \log(N), \qquad m = m_1 + \ldots + m_r, \quad s = s_1 + \ldots + s_r.$$

Note that this shows how the success of CS will be resolution dependent.

r-level Sampling Scheme

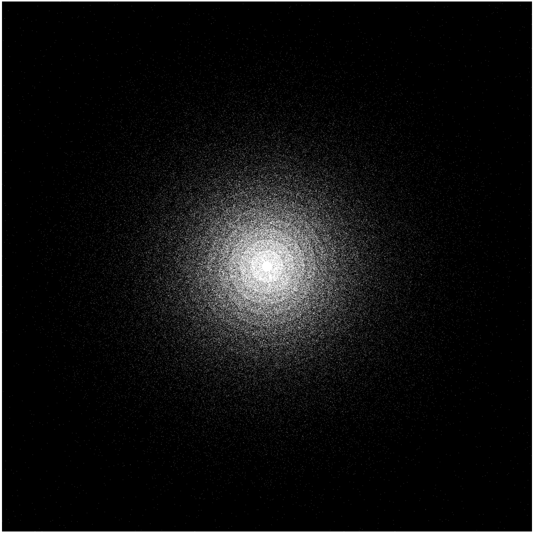
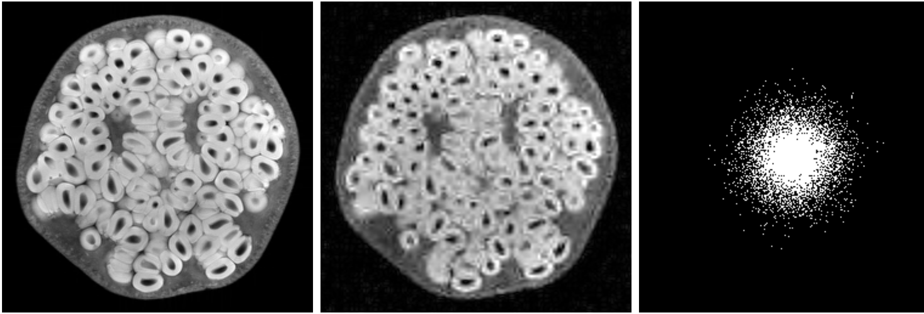


Figure : The typical sampling pattern that will be used.

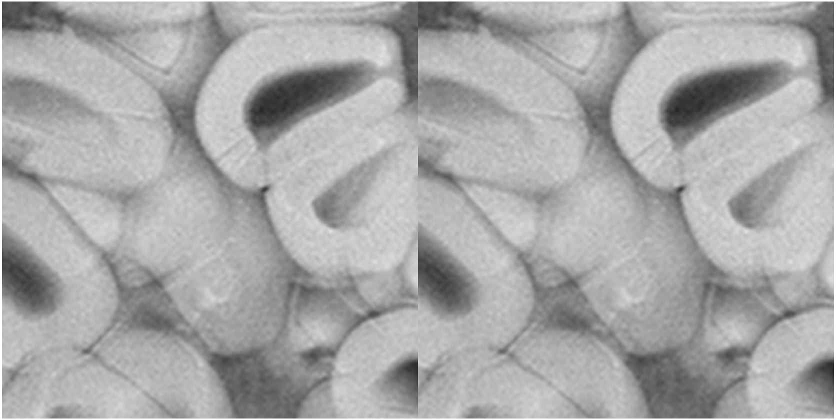
Resolution Dependence, 5% subsampling

Size: 256×256, Error $= 10.8\%$

Original | CS reconstruction | Subsamp. map

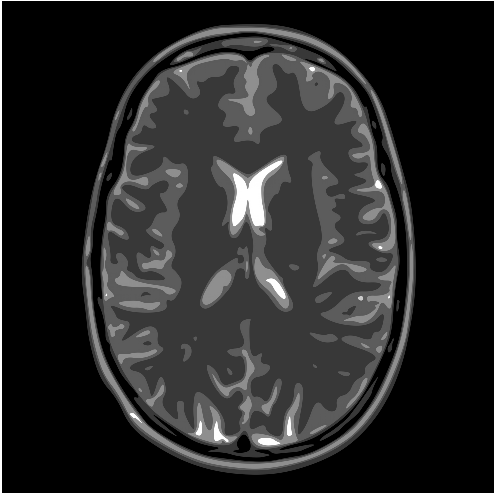


2048×2048 full sampling and 5% subsampling (DB4)



MRI Data courtesy of Andy Ellison, Boston University. Numerics taken from: On asymptotic structure in compressed sensing, *B. Roman, B. Adcock, A. C. Hansen*, arXiv:1406.4178

The GLPU-Phantom



The Guerquin-Kern,Lejeune, Pruessmann, Unser-Phantom (ETH and EPFL)

Seeing further with compressed sensing

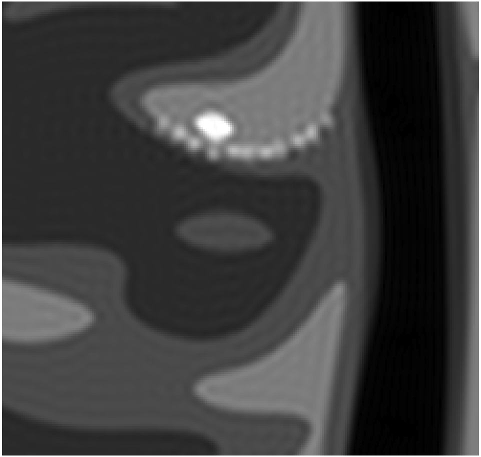


Figure : The figure shows 512 $\times$ 512 full sampling (= 262144 samples) with 2048 $\times$ 2048 zero padding.

Numerics taken from: On asymptotic structure in compressed sensing, *B. Roman, B. Adcock, A. C. Hansen*, arXiv:1406.4178

Seeing further with compressed sensing

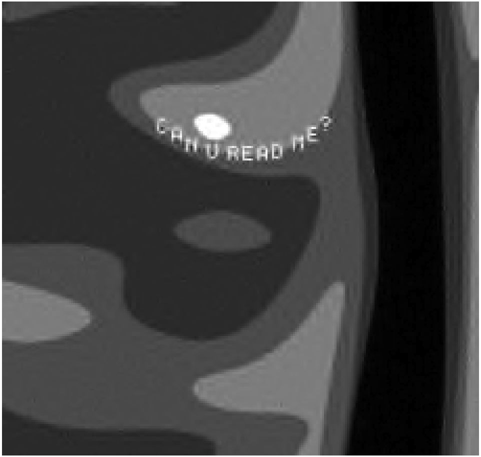


Figure : The figure shows 6.25% subsampling from 2048×2048 ($= 262144$ samples) and $DB4$.

Numerics taken from: On asymptotic structure in compressed sensing, *B. Roman, B. Adcock, A. C. Hansen*, arXiv:1406.4178

Key Question

- Question: Will compressed sensing ever be used in commercial MRI scanners?

Siemens

Siemens has implemented our experiments and verified our theory experimentally on their scanners.

- See the Siemens report: "Novel Sampling Strategies for Sparse MR Image Reconstruction," in Proceedings of the International Society for Magnetic Resonance in Medicine.

Siemens Conclusion:

- "Significant differences in the spatial resolution can be observed."
- "The image resolution has been greatly improved."
- "Current results practically demonstrated that it is possible to break the coherence barrier by increasing the spatial resolution in MR acquisitions. This likewise implies that the full potential of the compressed sensing is unleashed only if asymptotic sparsity and asymptotic incoherence is achieved. Therefore, compressed sensing might better be used to increase the spatial resolution rather than accelerating the data acquisition in the context of non-dynamic 3D MR imaging."

Key Question

- Question: Will Compressed sensing ever be used in commercial MRI scanners?
- Answer: We are closer than ever.

Other work

Structured sampling:

- Calderbank, Carin, Carson, Chen, Rodrigues (2012)

Structured sparsity:

- 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), Donoho & Montanari (2013) + others

Existing algorithms:

- Model-based CS: Baranuik, Cevher, Duarte, Hegde (2010)
- Bayesian CS: Ji, Xue & Carin (2008), He & Carin (2009)
- Turbo AMP: Som & Schniter (2012)

Variable density sampling:

- Krahmer & Ward (2013), Kutyniok & Lim (2014)

Related Papers

For details see [4, 5, 2, 1, 3]

B. Adcock and A. C. Hansen.
Generalized sampling and infinite-dimensional compressed sensing.
Found. Comp. Math., 2014.
(in revision).

B. Adcock, A. C. Hansen, C. Poon, and B. Roman.
Breaking the coherence barrier: A new theory for compressed sensing.
arXiv:1302.0561, 2014.

A. C. Hansen.
On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators.
J. Amer. Math. Soc., 24(1):81–124, 2011.

B. Roman, B. Adcock, and A. C. Hansen.
On asymptotic structure in compressed sensing.
arXiv:1406.4178, 2014.

Q. Wang, M. Zenge, H. E. Cetingul, E. Mueller, and M. S. Nadar.
Novel sampling strategies for sparse mr image reconstruction.
Proc. Int. Soc. Mag. Res. in Med., (22), 2014.