Compressed sensing in the real world - The need for a new theory

Anders C. Hansen (Cambridge)

Joint work with:

B. Adcock (Purdue)
C. Poon (Cambridge)
B. Roman (Cambridge)

Paris, January 13, 2014

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

Compressed Sensing in Inverse Problems

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.

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

- 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.

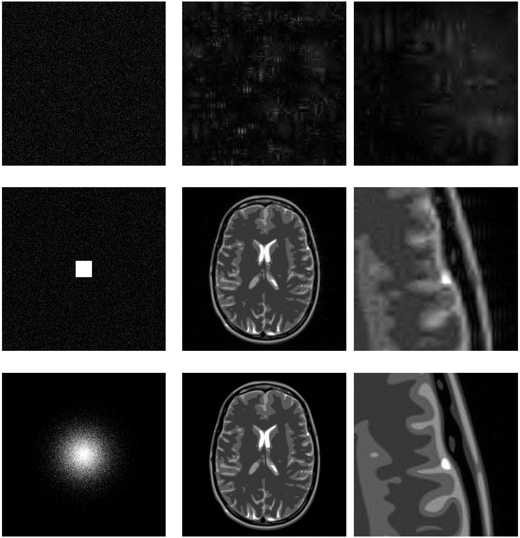
Problem: These concepts are absent in virtually all the problems listed above. Moreover, uniform random subsampling gives highly suboptimal results.

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

Uniform Random Subsampling

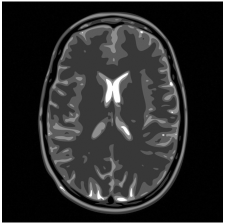
$$U = U_{\mathrm{dft}} V_{\mathrm{dwt}}^{-1}.$$

5% subsamp-map Reconstruction Enlarged



- 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.

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

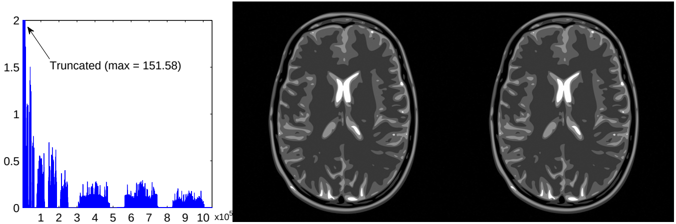
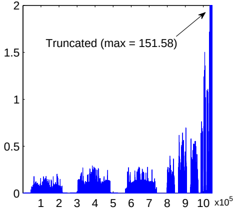


Figure: Wavelet coefficients and subsampling reconstructions from 10% of Fourier coefficients with distributions $(1+\omega_1^2+\omega_2^2)^{-1}$ and $(1+\omega_1^2+\omega_2^2)^{-3/2}$.

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

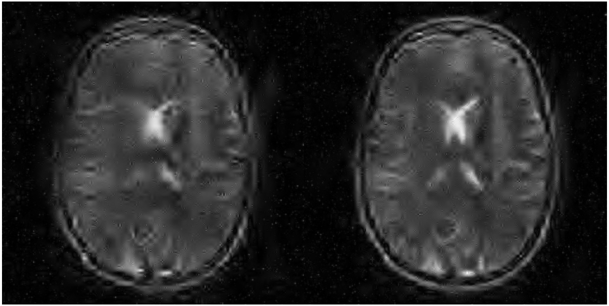


Figure: The reconstructions from the reversed coefficients.

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%

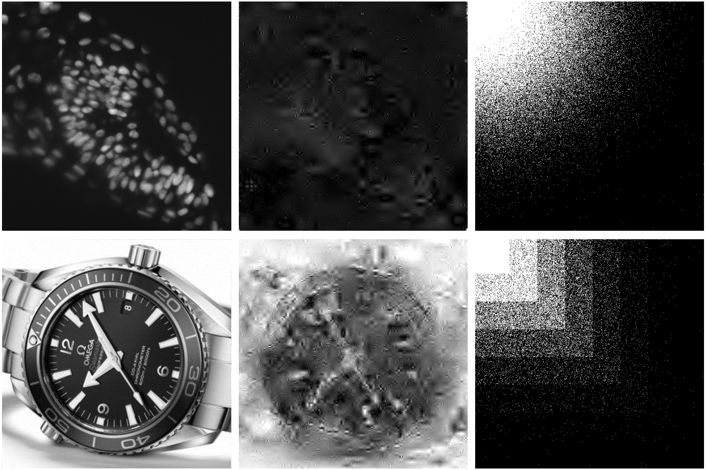
$U_{\mathrm{Had}} V_{\mathrm{dwt}}^{-1}$

Fluorescence Microscopy

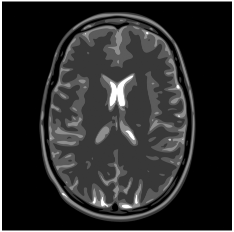
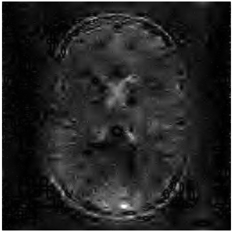
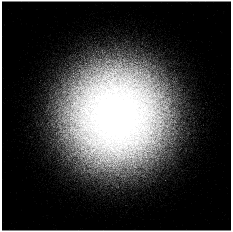
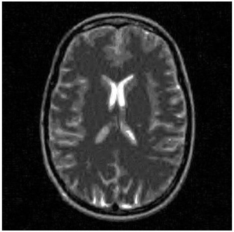
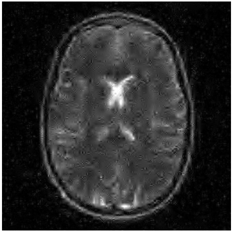
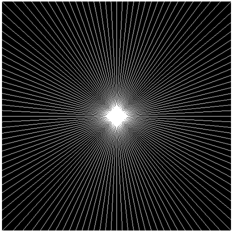
1024, 12%

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

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			

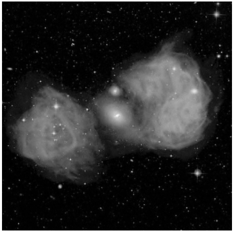
Sparsity - The Flip Test (contd.)

1024, 10%

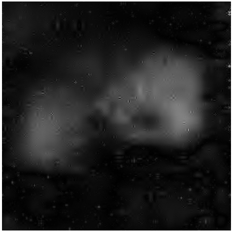
$U_{\mathrm{dft}} V_{\mathrm{dwt}}^{-1}$

Radio interferometry

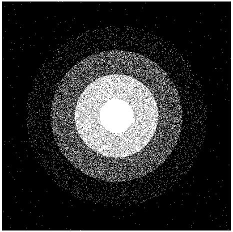
CS reconstr.



CS reconstr, w/ flip coeffs.



Subsampling pattern



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

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

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

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

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

Images are not sparse, they are asymptotically sparse

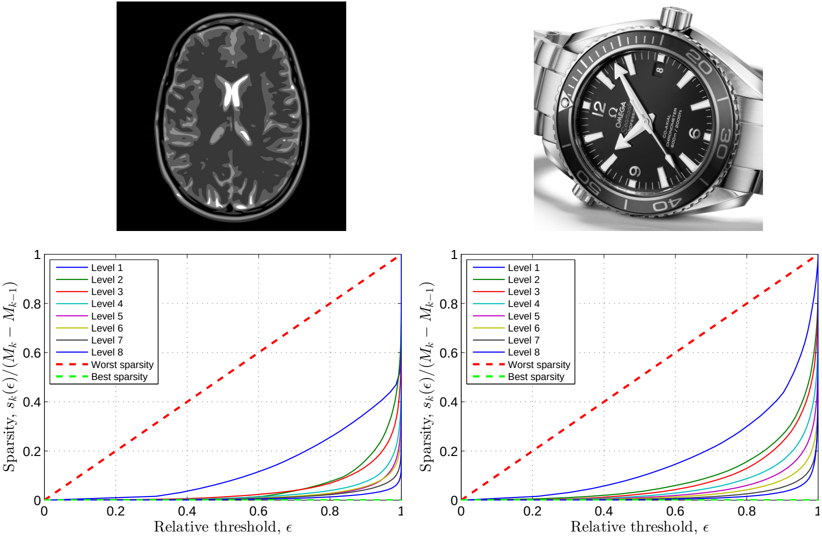
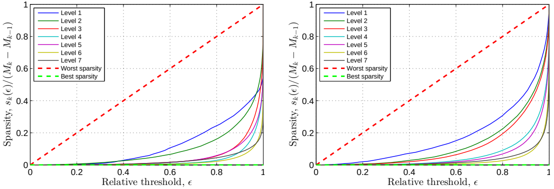


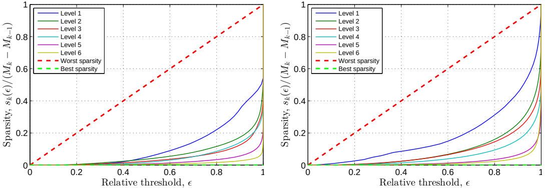
Figure: Relative sparsity of Daubechies 8 wavelet coefficients.

Images are not sparse, they are asymptotically sparse

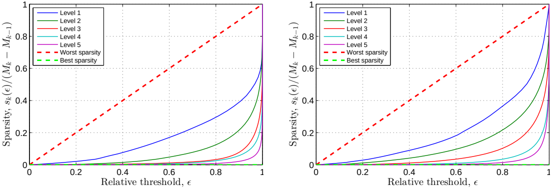
Curvelets



Contourlets



Shearlets



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

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

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

Fourier to DB4 Fourier to Legendre Polynomials

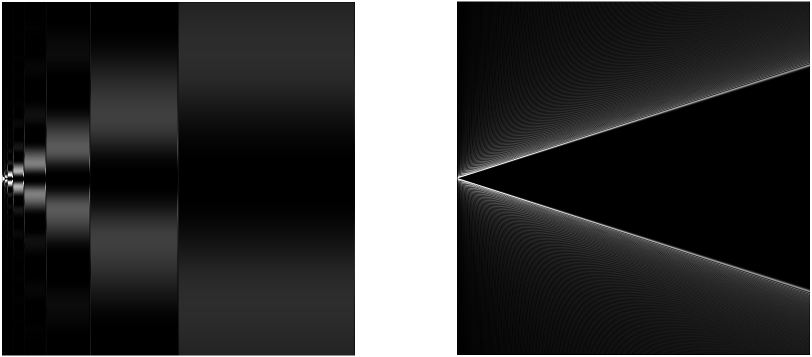


Figure: Plots of the absolute values of the entries of the matrix U

Hadamard and wavelets are coherent

Let

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

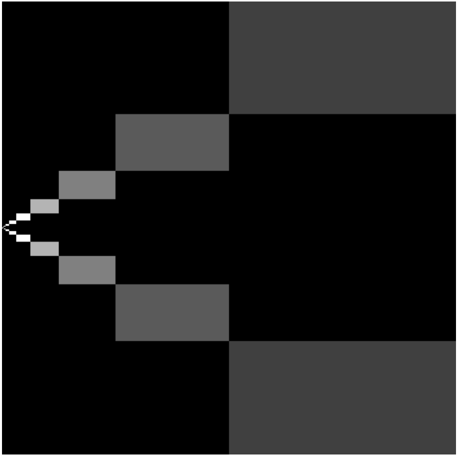
where H is a Hadamard matrix and V_{dw} is the discrete wavelet transform. Then

$$\mu(U_n) = 1$$

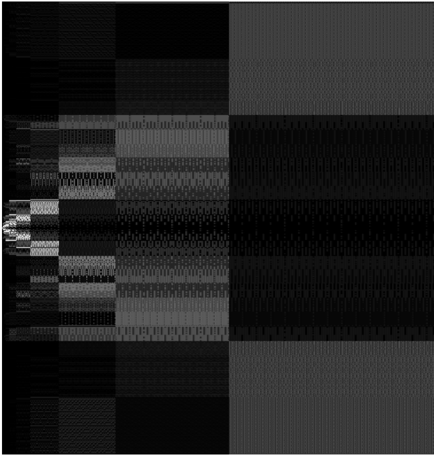
for all n and all Daubechies wavelets!

Hadamard and wavelets are asymptotically incoherent

Hadamard to Haar



Hadamard to DB8



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)

Definition

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

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

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

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

Multi-level sampling scheme

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.

Local coherence

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

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

Theorem

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* $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ *and* $\mathbf{m} = (m_1, \dots, m_r) \in \mathbb{N}^r$*. Let* $(\mathbf{s}, \mathbf{M})$*, where* $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$*,* $M_1 < \dots < M_r$*, and* $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$*, be any pair such that the following holds: for* $\epsilon > 0$ *and* $1 \leq k \leq r$*,*

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

and $m_k \gtrsim \hat{m}_k \cdot (\log(\epsilon^{-1}) + 1) \cdot \log(N)$*, with* $\hat{m}_k$ *satisfying*

$$1 \gtrsim \sum_{k=1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1 \right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot \tilde{s}_k, \qquad \forall\, l = 1, \dots, r, \qquad (5)$$

for all $\tilde{s}_1, \dots, \tilde{s}_r \in (0, \infty)$ *such that*

Theorem

$$\tilde{s}_1 + \ldots + \tilde{s}_r \leq s_1 + \ldots + s_r = s, \qquad \tilde{s}_k \leq S_k(\mathbf{N}, \mathbf{M}, \mathbf{s}).$$

Suppose that $\xi \in \ell^1(\mathbb{N})$ *is a minimizer of (3). Then, with probability exceeding* $1 - s\epsilon$, *we have that*

$$\|\xi - \beta\| \leq C \cdot \left(\delta \cdot \sqrt{K} \cdot \left(1 + L \cdot \sqrt{s}\right) + \sigma_{\mathbf{s},\mathbf{M}}(f)\right), \tag{6}$$

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* $K = \max_{k=1,\ldots,r}\left\{\frac{N_k - N_{k-1}}{m_k}\right\}$.

$$S_k = S_k(\mathbf{N}, \mathbf{M}, \mathbf{s}) = \max_{\eta \in \Theta} \|P_{N_k}^{N_{k-1}} U \eta\|^2,$$

where Θ is given by

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

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

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

r-level Sampling Scheme

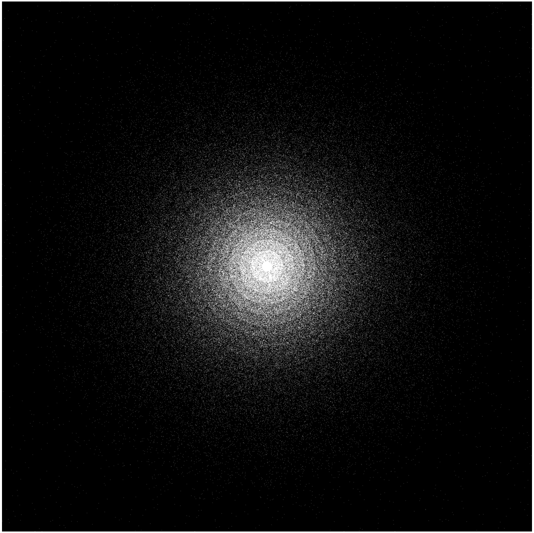


Figure: The typical sampling pattern that will be used.

The Berlin Cathedral (256x256)

15% Random Bernoulli to DB8	15% Multi level Hadamard to DB8	Fully sampled (original image)
		

	15% Random Bernoulli to DB8	15% Multi level Hadamard to DB8
RAM (GB):	**4.8**	< 0.1
Speed (it/s):	1.31	18.1
Conv. (sec):	($4m27s$) 267	18.6
Rel. err. (%):	22.4	14.7

The Berlin Cathedral (512x512)

15% Random Bernoulli to DB8

15% Multi level Hadamard to DB8

Fully sampled (original image)

	15% Random Bernoulli to DB8	15% Multi level Hadamard to DB8
RAM (GB):	**76.8**	< 0.1
Speed (it/s):	0.15	4.9
Conv. (sec):	$(42m)$ 2517	$(1m13s)$ 73.4
Rel. err. (%):	19.0	12.2

The Berlin Cathedral (1024x1024)

15% Random
Bernoulli to DB8

?

15% Multi level
Hadamard to DB8

Fully sampled
(original image)

	15% Random Bernoulli to DB8	15% Multi level Hadamard to DB8
RAM (GB):	**1229**	< 0.1
Speed (it/s):	0.0161	1.07
Conv. (sec):	$6h\,36m$	($3m45s$) 225.4
Rel. err. (%):		10.4

The Berlin Cathedral (2048x2048)

15% Random Bernoulli to DB8	15% Multi level Hadamard to DB8	Fully sampled (original image)
?		

	15% Random Bernoulli to DB8	15% Multi level Hadamard to DB8
RAM (GB):	**19661**	<0.1
Speed (it/s):		0.17
Conv. (sec):		(28m) 1687
Rel. err. (%):		8.5

The Berlin Cathedral (4096x4096)

	15% Random Bernoulli to DB8	15% Multi level Hadamard to DB8	Fully sampled (original image)
	?		
RAM (GB):	**314573**	< 0.1	
Speed (it/s):		0.041	
Conv. (sec):		$(1h37m)$ 5852	
Rel. err. (%):		6.56	

The Berlin Cathedral (8192x8192)

15% Random Bernoulli to DB8

15% Multi level Hadamard to DB8

Fully sampled (original image)

	15% Random Bernoulli to DB8	15% Multi level Hadamard to DB8
RAM (GB):	**5033165**	< 0.1
Speed (it/s):		0.0064
Conv. (sec):	238*d*	(8*h*30*m*) 30623
Rel. err. (%):		3.5

The Key Question

- Is universality really what we want?

Gaussian w/model based CS vs multi-level and Hadamard

30% Gaussian to wavelet 30% Hadamard to wavelet Original image

Figure: Gaussian w/model based CS vs Hadamard

Gaussian w/model based CS vs multi-level and Hadamard

30% Gaussian to wavelet | 30% Had. to Curvlts | Original image

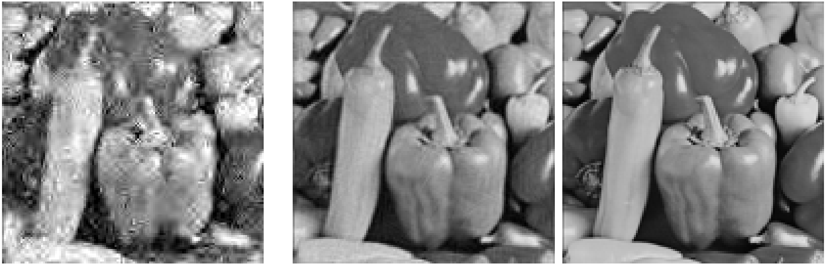
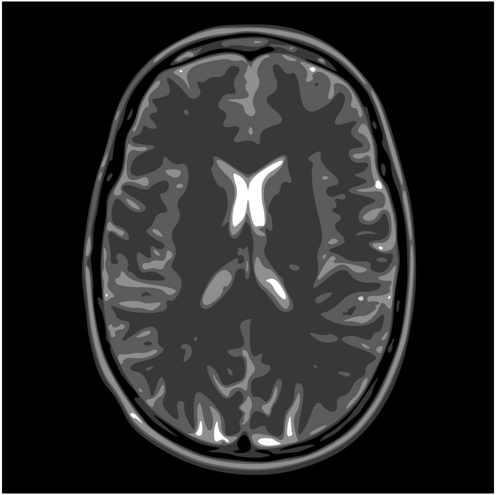


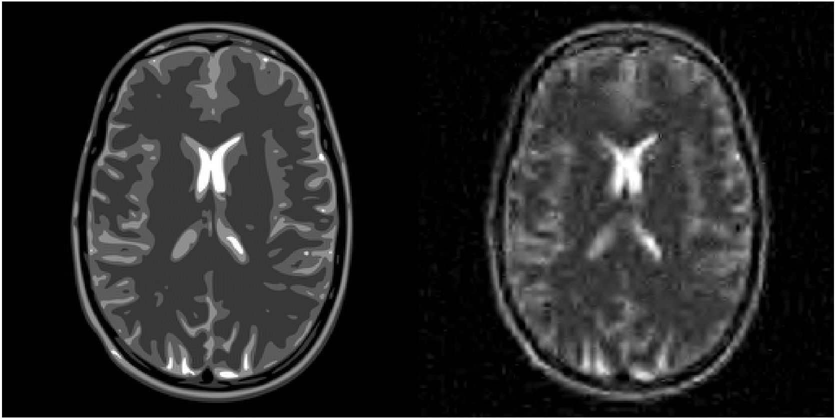
Figure: Gaussian w/model based CS vs Hadamard

The GLPU-Phantom



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

256×256 full sampling and 5% subsampling (DB4)



MSE is obviously different.

4096×4096 full sampling and 4% subsampling (DB4)

MSE is the same for both reconstructions.

Seeing further with compressed sensing

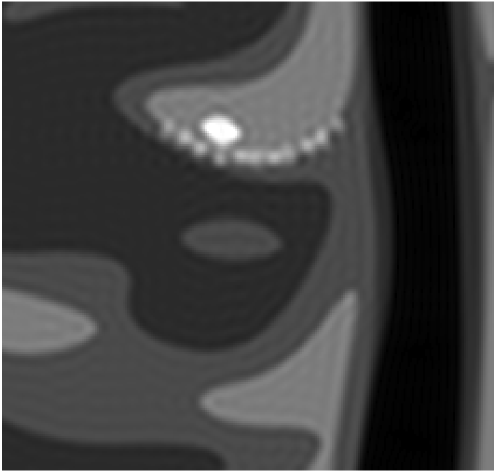


Figure: The figure shows 512 $\times$ 512 full sampling (= 262144 samples) with 2048 $\times$ 2048 zero padding.

Seeing further with compressed sensing

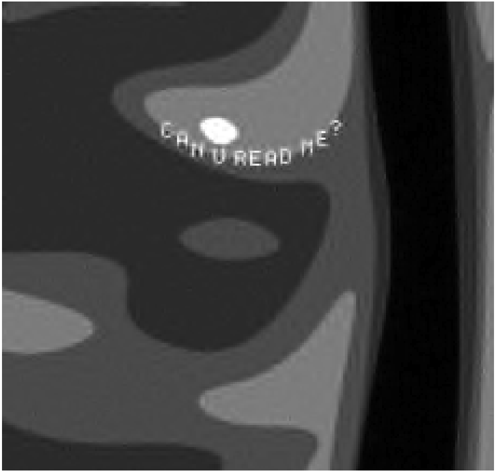


Figure: The figure shows 6.25% subsampling from 2048×2048 ($= 262144$ samples) and $DB4$.

Take home message: Compressed Sensing Vol. II

- The optimal sampling strategy depends on the structure of the signal (unless you have perfect incoherence).
- Real world problems are usually completely coherent, yet asymptotically incoherent. Thus, one must use multi-level sampling.
- We have covered the abstract orthonormal basis case, but there is tons of work to be done (frames, TV, curvelets, contourlets, shearlets, polynomials, etc).
- When building hardware one does not need to strive for incoherence, one only needs asymptotic incoherence.
- Speed and storage issues in compressive imaging can be solved by using multi-level sampling.
- Related work:
  - Krahmer and Ward
  - Baranuik, Cevher, Duarte, Hegde (model based CS)
  - Calderbank, Carin, Carson, Chen, Rodrigues