Infinite dimensional/continuous compressed sensing in physics

Anders C. Hansen (Cambridge)

Joint work with:

B. Adcock (Simon Fraser University)
I. Calvo-Almazan (Cambridge)
A. Jones (Cambridge)
C. Poon (Cambridge)
B. Roman (Cambridge)
T. Tamtogl (Cambridge)

XQCD 2016

Compressed Sensing in Physics

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) Nuclear Magnetic Resonance (NMR)
(vi) Electron Microscopy/Tomography
(vii) Reflection seismology
(viii) Radio interferometry
(ix) Helium Atom Scattering
(x) 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.

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

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

MRI Example

Suppose that

$$f = \mathcal{F}g, \qquad g \in L^2(\mathbb{R}),$$

and $\mathrm{supp}(g) \subset [-T, T]$ for some $T > 0$. If $\epsilon \leq \frac{1}{2T}$ (the Nyquist rate) then

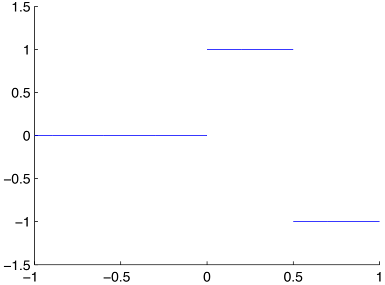
$$g = \epsilon \sum_{k=-\infty}^{\infty} f(k\epsilon) e^{2\pi \mathrm{i} \epsilon k \cdot}, \qquad L^2 \text{ convergence.} \tag{2}$$

In practice, one forms the approximation

$$g_N = \epsilon \sum_{k=-N}^{N} f(k\epsilon) e^{2\pi \mathrm{i} \epsilon k \cdot}.$$

MRI Example

Let g be



Approximating with the truncated Fourier series

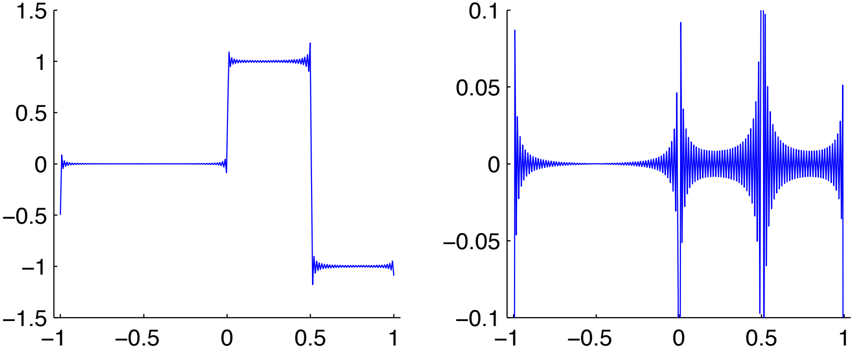


Figure : The figure displays g_N (left) as well as the error $g - g_N$ (right). $N = 128$

The Discrete Problem

Note that

$$g_N = \epsilon \sum_{k=-N+1}^{N} f(k\epsilon) e^{2\pi \mathrm{i} \epsilon k \cdot}.$$

can be written as

$$y = U_{df} x, \qquad U_{df} \in \mathbb{C}^{2N \times 2N}, \quad y, x \in \mathbb{C}^{2N},$$

where y represents a vector of the sampled values of f, x represents a vector of the point wise values of g_N (on a equidistant grid on [-1,1]), and U_{df} is a scalar multiple of the discrete Fourier transform.

The Finite Dimensional Compressed Sensing Problem

- If g is sparse in the Haar basis one could hope that

$$x_0 = V_{dw} x$$

is sparse, where V_{dw} is the discrete wavelet transform corresponding to the Haar wavelet.

- If that was the case we could randomly sample a set $\Omega \subset \{1, \ldots, 2N\}$ of size $|\Omega| = m < 2N$ and try to reconstruct x_0 (and hence x) from the subsampled vector $P_\Omega y$ by finding a minimizer ξ to

$$\min_{\eta \in \mathbb{C}^n} \|\eta\|_{l^1} \ : \ P_\Omega U_{df} V_{dw}^{-1} \eta = P_\Omega y, \tag{3}$$

where P_Ω denotes the projection onto $\mathrm{span}\{e_j\}_{j\in\Omega}$, and hope that $\xi = x_0$ with high probability.

Finite Dimensional Compressed Sensing Results

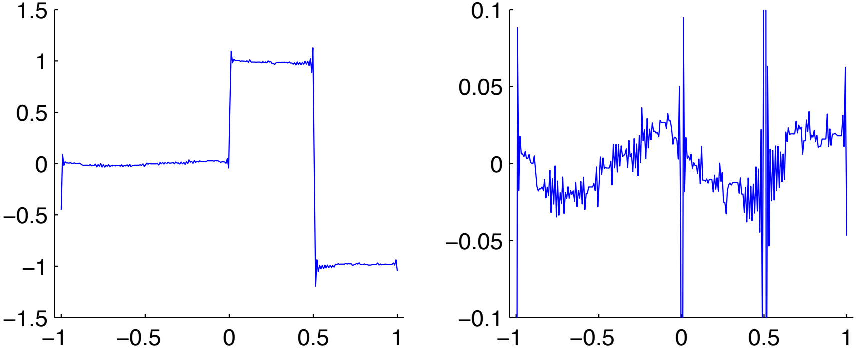


Figure : The left part displays the compressed sensing approximation $V_{dw}^{-1}\xi$ to g_N from solving (3) with $|\Omega| = 130$. The right part displays the error and $g - V_{dw}^{-1}\xi$.

Question

- Why does this happen? After all this is a super sparse problem?

Why not solve the true infinite-dimensional problem?

- Physical problems are often continuous/infinite-dimensional.

The Model

- Given a separable Hilbert space $\mathcal{H}$ with an orthonormal set $\{\varphi_k\}_{k\in\mathbb{N}}$.
- Given a vector

$$x_0 = \sum_{k=1}^{\infty} \beta_k \varphi_k, \qquad \beta = \{\beta_1, \beta_2, \dots\}.$$

- Suppose also that we are given a set of linear functionals $\{\zeta_j\}_{j\in\mathbb{N}}$ such that we can "measure" the vector x_0 by applying the linear functionals e.g. we can obtain $\{\zeta_j(x_0)\}_{j\in\mathbb{N}}$.

An Infinite System of Equations

With some appropriate assumptions on the linear functionals $\{\zeta_j\}_{j\in\mathbb{N}}$ we may view the full recovery problem as the infinite dimensional system of linear equations

$$\begin{pmatrix} \zeta_1(x_0) \\ \zeta_2(x_0) \\ \zeta_3(x_0) \\ \vdots \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots \\ u_{21} & u_{22} & u_{23} & \dots \\ u_{31} & u_{32} & u_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \end{pmatrix}, \qquad u_{ij} = \zeta_i(\varphi_j), \tag{4}$$

where we will refer to $U = \{u_{ij}\}_{i,j\in\mathbb{N}}$ as the "measurement matrix".

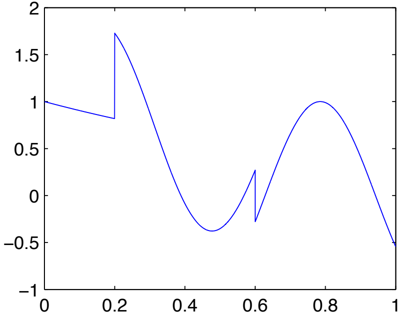
Let $\Omega \subset \mathbb{N}$ such that $|\Omega| = m < \infty$ be randomly chosen and let P_Ω denote the projection onto $\mathrm{span}\{e_j\}_{j\in\Omega}$. Now consider the convex (infinite-dimensional) optimization problem

$$\inf_{\eta \in l^1(\mathbb{N})} \|\eta\|_{l^1(\mathbb{N})} \ : \ P_\Omega \begin{pmatrix} \zeta_1(x_0) \\ \zeta_2(x_0) \\ \zeta_3(x_0) \\ \vdots \end{pmatrix} = P_\Omega \begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots \\ u_{21} & u_{22} & u_{23} & \dots \\ u_{31} & u_{32} & u_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \end{pmatrix}. \tag{5}$$

The solution to problem (5) cannot be computed explicitly because it is infinite-dimensional, and thus an approximation must be computed instead. For $R \in \mathbb{N}$, consider the optimization problem

$$\inf_{\eta \in P_R l^1(\mathbb{N})} \|\eta\|_{l^1(\mathbb{N})} \;:\; P_\Omega \begin{pmatrix} \zeta_1(x_0) \\ \zeta_2(x_0) \\ \zeta_3(x_0) \\ \vdots \end{pmatrix} = P_\Omega \begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots \\ u_{21} & u_{22} & u_{23} & \dots \\ u_{31} & u_{32} & u_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} P_R \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_R \end{pmatrix}. \tag{6}$$

Tests



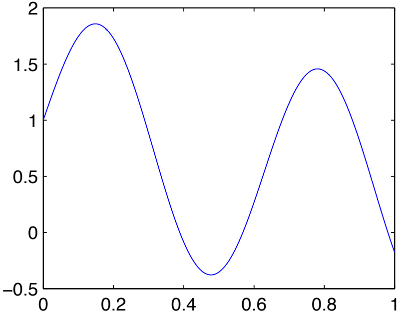


Figure : Left: the piecewise smooth test function f_1. Right: the smooth test function f_2

Tests

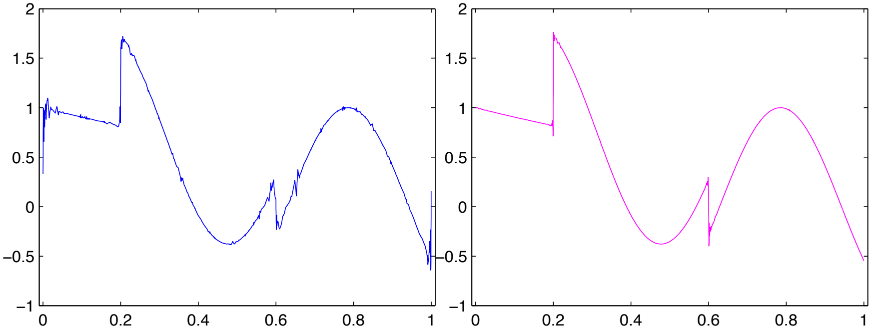


Figure : Left: Reconstruction error when reconstructing f_1 with compressed sensing using $U_{\mathrm{dft}} V_{\mathrm{dwt}^{-1}}$ using periodized DB6 wavelets. Right: Reconstruction error when reconstructing f_1 with compressed sensing using an "infinite-dimensional" discretization. Both methods use exactly the same samples.

Tests

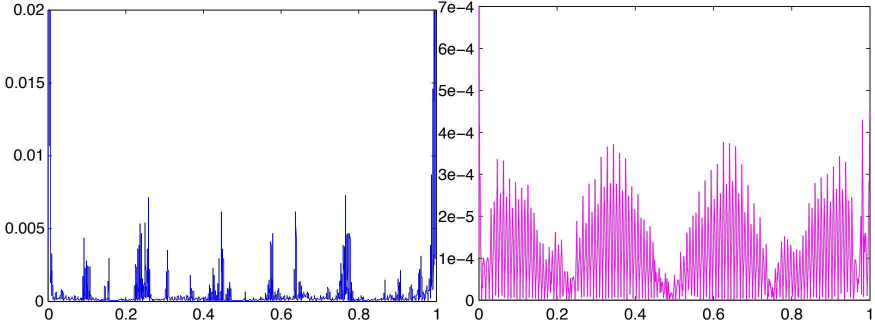
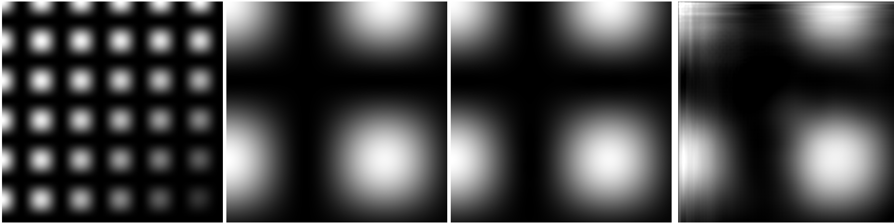


Figure : Left: errors for the reconstructions of f_2 with compressed sensing using $U_{\mathrm{dft}} V_{\mathrm{dwt}^{-1}}$. Righ: errors for the reconstructions of f_2 using an "infinite-dimensional" discretization. Both methods use exactly the same samples.

Electron Microscopy: Finte vs. infinite dimensions



Original | Original (zoom) | Inf-dim. CS (zoom) Err 0.6% | Fin-dim. CS (zoom) Err 12.7%

Figure : Subsampling 6.15%. Both reconstructions are based on identical sampling information.

Pillars of Compressed Sensing

- Sparsity
- Incoherence
- Uniform Random Subsampling

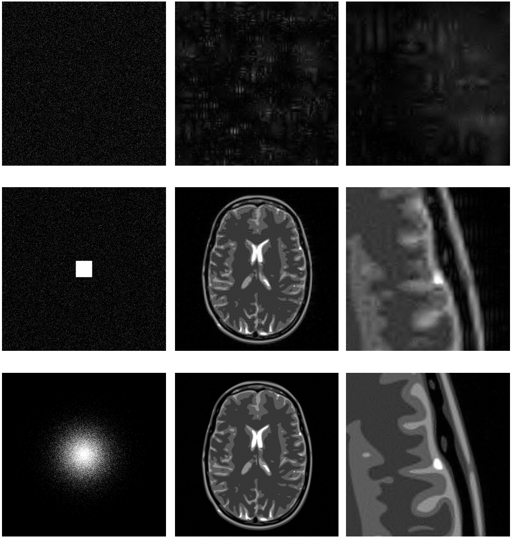
Problem: These concepts are absent in virtually all the problems listed above. Moreover, uniform random subsampling gives highly suboptimal results.

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

Uniform Random Subsampling

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

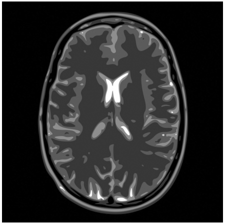
5% subsampling | Reconstruction | Enlarged



- 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.

Sparsity and the Flip Test

Let

$x =$ 

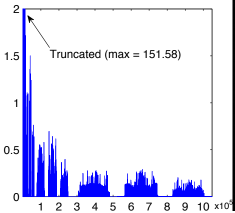
and

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

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



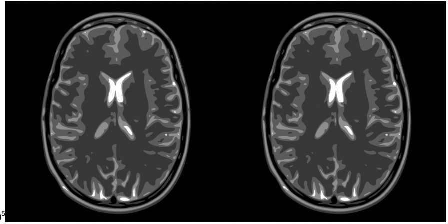
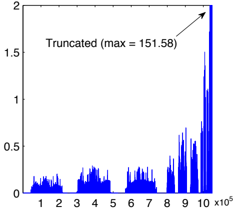


Figure : Wavelet coefficients and subsampling reconstructions from 10% of Fourier coefficients with distributions $(1+\omega_1^2+\omega_2^2)^{-1}$ and $(1+\omega_1^2+\omega_2^2)^{-3/2}$.

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

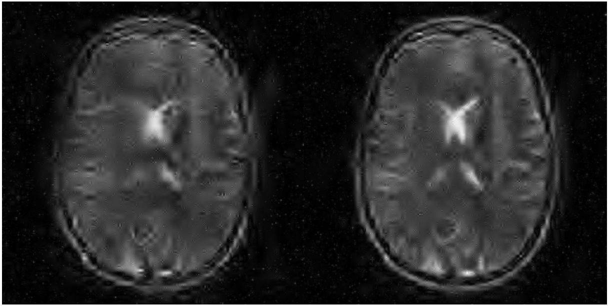


Figure : The reconstructions from the reversed coefficients.

Conclusion: The ordering of the coefficients did matter.

Question: Is sparsity really the right model?

Incoherence: 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!

Why?

- Physical problems are often continuous/infinite-dimensional.

Note that

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

where

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

where

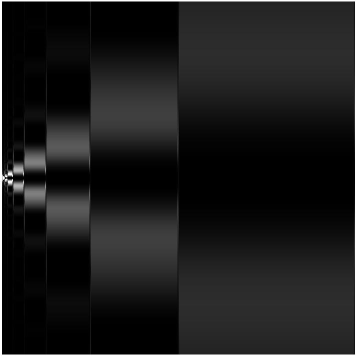
$$\{\varphi_j\}_{j\in\mathbb{N}} \quad \{\psi_j\}_{j\in\mathbb{N}}$$

are wavelets and complex exponentials respectively. 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

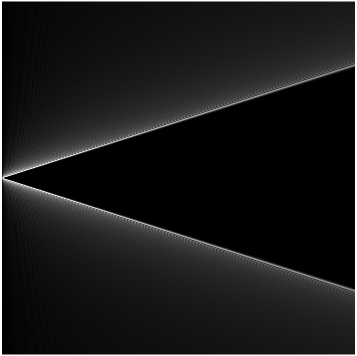


Figure : Plots of the absolute values of the entries of the matrix U

Images are asymptotically sparse

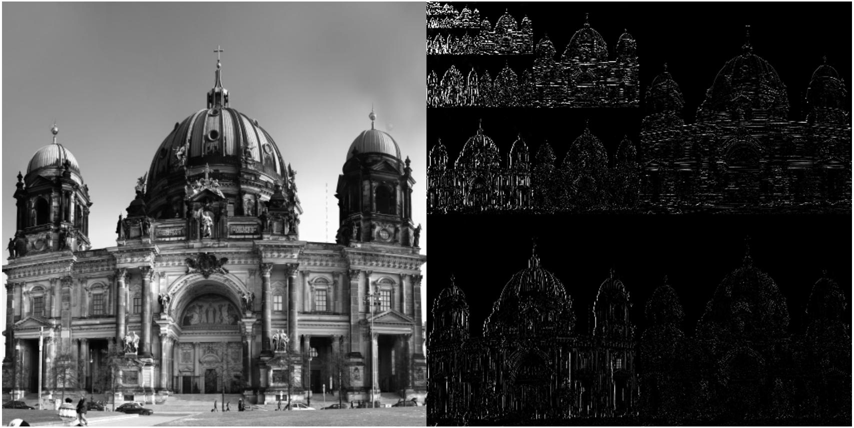
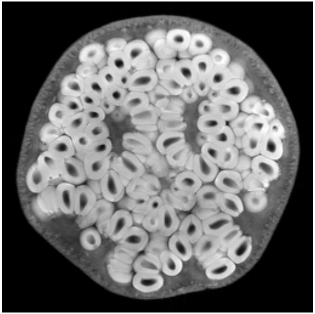


Figure : Left: original image x. Right: the wavelet coefficients of the image.

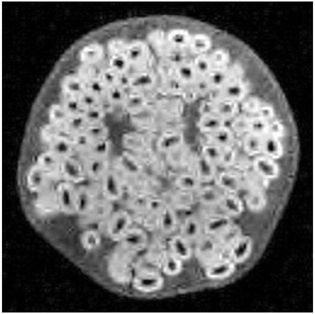
Resolution Dependence, 5% subsampling

Size: 256×256, Error $= 10.8\%$

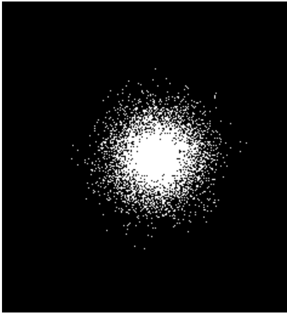
Original



CS reconstruction



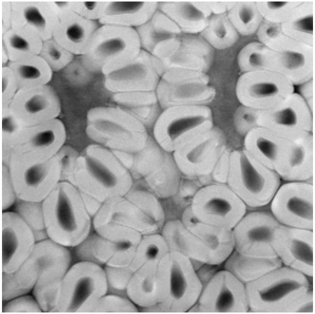
Subsamp. map



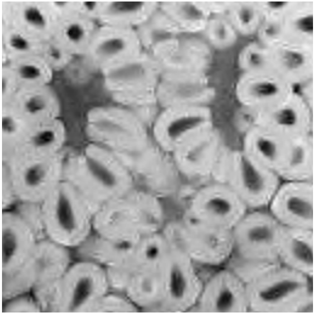
Resolution Dependence, 5% subsampling

Size: 512×512, Error $= 6.0\%$

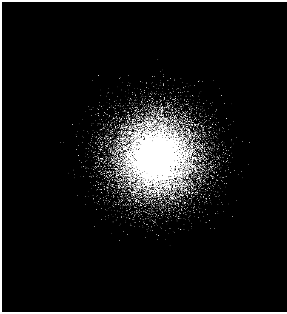
Original



CS reconstruction



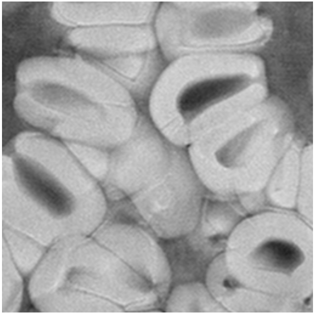
Subsamp. map



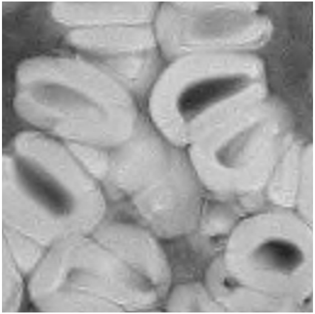
Resolution Dependence, 5% subsampling

Size: 1024×1024, Error $= 3.6\%$

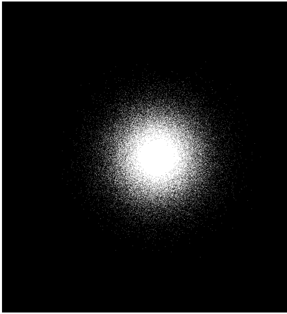
Original



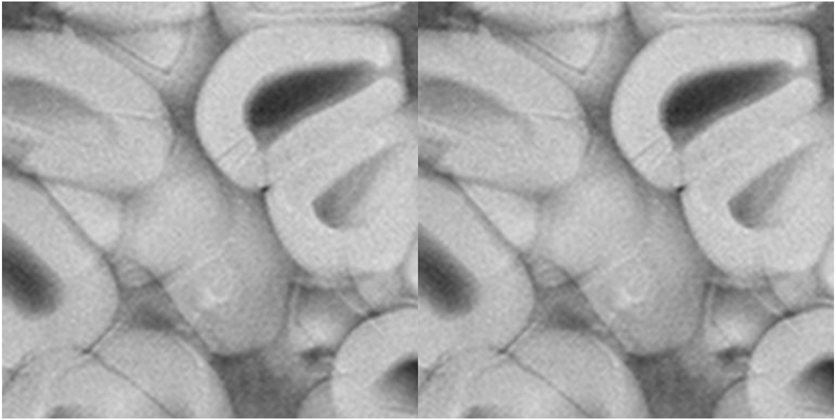
CS reconstruction



Subsamp. map

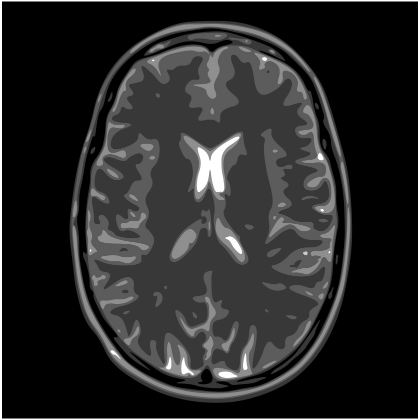


2048 × 2048 **full sampling and** 5% **subsampling (DB4)**



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

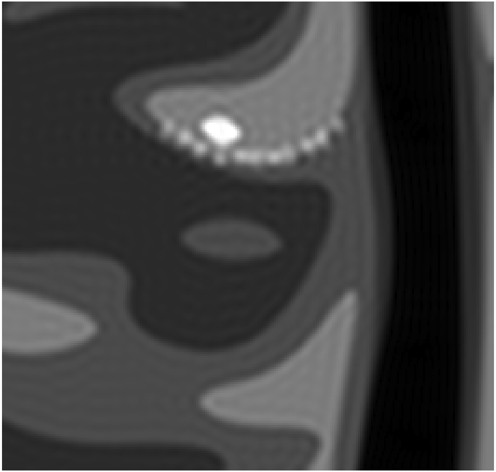
Resolution enhancing in MRI



The MRI machine samples the continuous Fourier transform of the brain.

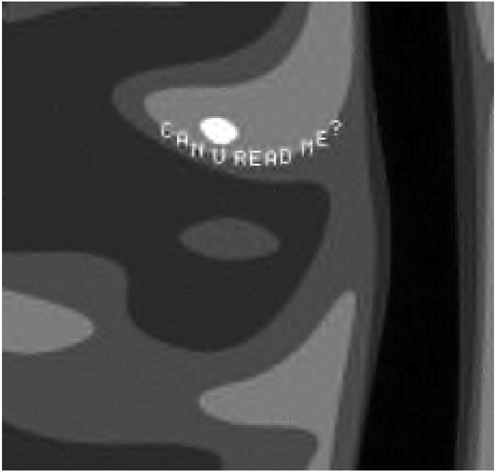
Test of compressed sensing in MRI

Classical MRI scanning with 512×512 full sampling ($= 262144$ samples) with 2048×2048 zero padding. Can you see the details?



Test of compressed sensing in MRI

Compressed sensing with 6.25% subsampling from 2048×2048 ($= 262144$ samples, the same number of samples as the previous example).



Siemens

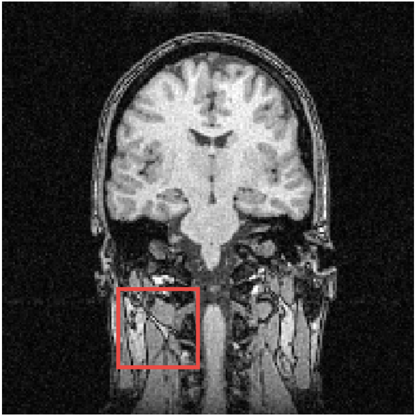
Siemens has implemented our experiments and verified our theory experimentally on their scanners.

- 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."

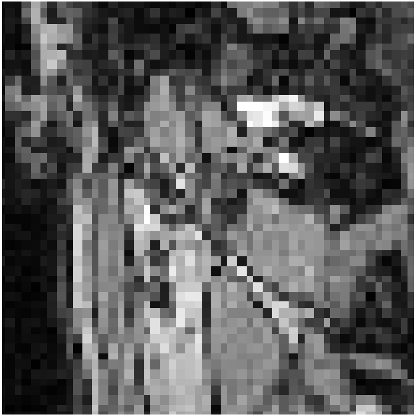
B.Roman, M.Graves, A.Hansen, D.Lomas, *"Improved Spatial Resolution and Targeted Sampling in MRI"* — University nomination for the Rosetrees Interdisciplinary Research Award 2016.



General Electric 1.5T MRI, Addenbrooke's. Slice from 3D scan (whole head).

B.Roman, M.Graves, A.Hansen, D.Lomas, *"Improved Spatial Resolution and Targeted Sampling in MRI"* — University nomination for the Rosetrees Interdisciplinary Research Award 2016.

192^3(1.2mm) **Full, 15 min**



512^3(0.4mm) **CS, 15 min**.

General Electric 1.5T MRI, Addenbrooke's. Slice from 3D scan (whole head).

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 $\Sigma_{\mathbf{s},\mathbf{M}}$.

Images are asymptotically sparse

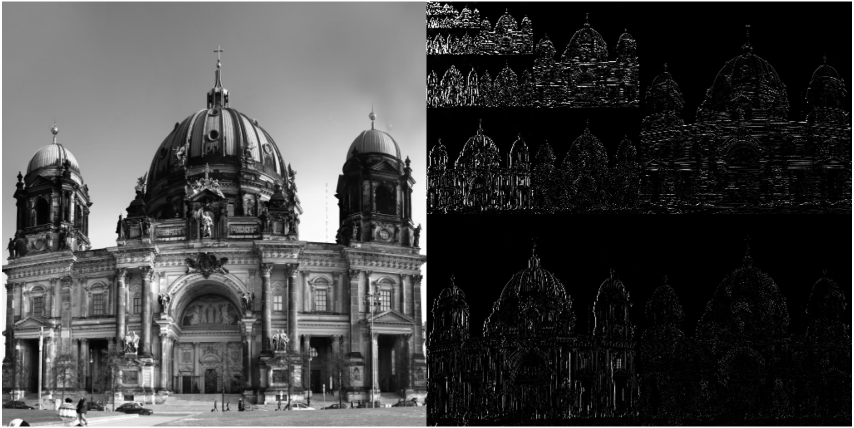


Figure : Left: original image x. Right: the wavelet coefficients of the image.

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

$$\sigma_{\mathbf{s},\mathbf{M}}(f) := \min_{\eta\in\Sigma_{\mathbf{s},\mathbf{M}}} \|\beta - \eta\|_{l^1}. \tag{7}$$

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.

r-level Sampling Scheme

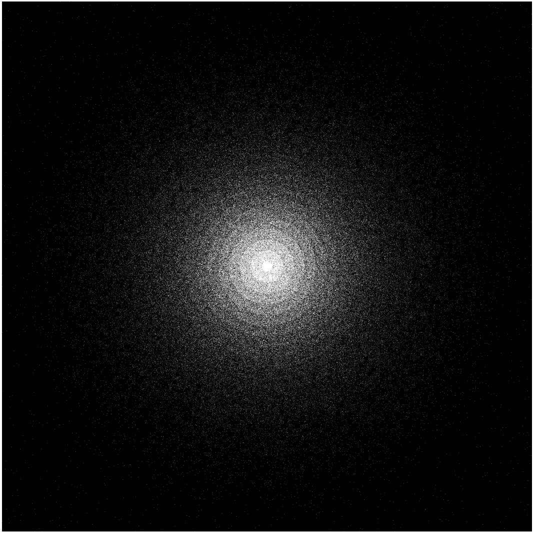


Figure : The typical sampling pattern that will be used.

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 $(k,l)^{\text{th}}$ 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 $N_0 = M_0 = 0$.

Analog inverse problems are asymptotically incoherent

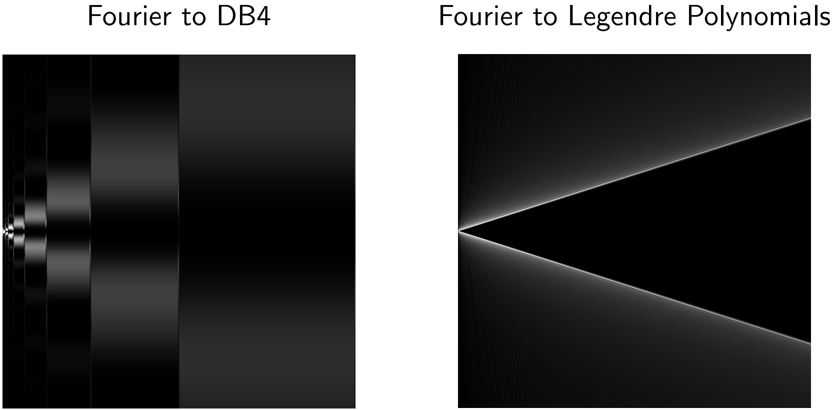


Figure : Plots of the absolute values of the entries of the matrix U

The optimization problem

$$\inf_{\eta \in \ell^1(\mathbb{N})} \|\eta\|_{\ell^1} \text{ subject to } \|P_\Omega U\eta - y\| \leq \delta. \tag{8}$$

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 $1 \leq k \leq r$,

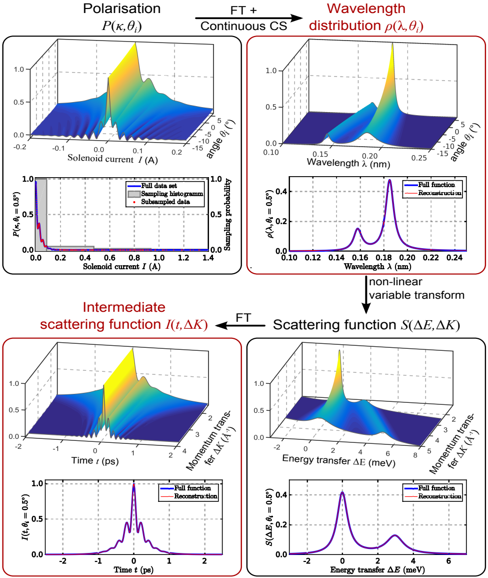
$$1 \gtrsim \frac{N_k - N_{k-1}}{m_k} \cdot \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r} \mu_{\mathbf{N},\mathbf{M}}(k,l) \cdot s_l \right) \cdot \log(N). \tag{9}$$

Suppose that $\xi \in \mathbb{C}^N$ is a minimizer of (8) 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 (7), $L = 1 + \frac{\sqrt{\log_2\left(6\epsilon^{-1}\right)}}{\log_2(4KM\sqrt{s})}$ and $K = \max_{k=1,\ldots,r}\left\{\frac{N_k - N_{k-1}}{m_k}\right\}$.

Continuous CS in Helium Atom Scattering



From *Nature, Sci. Rep.* July 2016