Compressed sensing and imaging

The effect and benefits of local structure

Ben Adcock
Department of Mathematics
Simon Fraser University

Overview

Previously: An introduction to compressed sensing.

- A (gentle) overview of the main principles of CS.

Now: Focus on the use of CS in imaging problems.

- In particular, my own research into explaining/enhancing CS performance in these applications.

Collaborators: Anders C. Hansen, Clarice Poon, Bogdan Roman.

Outline

Outline

Applications of CS

An inexhaustive list:

- Magnetic Resonance Imaging (MRI)
- Compressive imaging (single-pixel camiera, lensless, infrared)
- X-ray CT
- Seismic tomography
- Uncertainty Quantification
- Electron microscopy
- Fluorescence microscopy
- Radio interferometry
- Radar
- Analog-to-digital conversion
- ...

Applications of CS

An inexhaustive list:

- Magnetic Resonance Imaging (MRI)
- Compressive imaging (single-pixel camiera, lensless, infrared)
- X-ray CT
- Seismic tomography
- Uncertainty Quantification
- Electron microscopy
- Fluorescence microscopy
- Radio interferometry
- Radar
- Analog-to-digital conversion
- ...

Today's focus is imaging applications (in red).

Imaging applications of CS

Two broad categories:

Type I: Fixed sensing matrices/operators.

- Sensing matrix is dictated by the application.
- Typical examples: Fourier or Radon transforms.
- Applications: MRI, X-ray CT, seismic tomography, electron microscopy, radio interferometry...

Type II: Designed sensing matrices.

- Sensor can be designed to optimize the reconstruction quality.
- Applications: compressive imaging, fluorescence microscopy,...

This talk

Type I problems:

- Standard CS theory does not adequetly explain why CS actually works in these applications.
- We introduce a new CS framework which does this.
- The key element of this framework is local structure.

Type II problems:

- We introduce a new approach for these problems.
- This approach uses the new framework to exploit inherent structure through new sensing matrix design principles.

This talk

Type I problems:

- Standard CS theory does not adequetly explain why CS actually works in these applications.
- We introduce a new CS framework which does this.
- The key element of this framework is local structure.

Type II problems:

- We introduce a new approach for these problems.
- This approach uses the new framework to exploit inherent structure through new sensing matrix design principles.

Outline

Main principles

We consider the incoherence-based setup. Let $U = (u_{ij}) \in \mathbb{C}^{N \times N}$ be an isometry and $x \in \mathbb{C}^N$ be the object to recover.

Sparsity:

- x has s significant entries, $s \ll N$.
- Equivalently, $\sigma_s(x) = \inf\{\|x - z\|_1 : z \text{ is } s\text{-sparse}\}$ is small.

Incoherence:

- The coherence $\mu(U) = \max |u_{ij}|^2$ satisfies $\mu(U) \leq c/N$.

Uniform random subsampling:

- We select $\Omega \subseteq \{1, \ldots, N\}$, $|\Omega| = m$ uniformly at random.
- The measurements of x are $y = P_\Omega U x + e$, where P_Ω selects rows of U corresponding to Ω and $\|e\|_2 \leq \delta$ is noise.

Main principles

We consider the incoherence-based setup. Let $U = (u_{ij}) \in \mathbb{C}^{N \times N}$ be an isometry and $x \in \mathbb{C}^N$ be the object to recover.

Sparsity:

- x has s significant entries, $s \ll N$.
- Equivalently, $\sigma_s(x) = \inf\{\|x - z\|_1 : z \text{ is } s\text{-sparse}\}$ is small.

Incoherence:

- The coherence $\mu(U) = \max |u_{ij}|^2$ satisfies $\mu(U) \leq c/N$.

Uniform random subsampling:

- We select $\Omega \subseteq \{1, \ldots, N\}$, $|\Omega| = m$ uniformly at random.
- The measurements of x are $y = P_\Omega U x + e$, where P_Ω selects rows of U corresponding to Ω and $\|e\|_2 \leq \delta$ is noise.

Main theorem

Theorem (see Candès & Plan, BA & Hansen)

Let $0 < \epsilon \leq \mathrm{e}^{-1}$ *and suppose that*

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

Then with probability greater than $1 - \epsilon$ *any minimizer* $\hat{x}$ *of the problem*

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \text{ subject to } \|P_\Omega U z - y\|_2 \leq \delta \sqrt{N/m},$$

satisfies

$$\|x - \hat{x}\|_2 \leq C_1 \sigma_s(x) + C_2 L \sqrt{s} \delta,$$

where $L = 1 + \frac{\sqrt{\log(\epsilon^{-1})}}{\log(4N\sqrt{s})}$.

If U is incoherent, i.e. $\mu(U) \lesssim 1/N$, then $m \approx s \cdot \log(\epsilon^{-1}) \cdot \log(N)$.

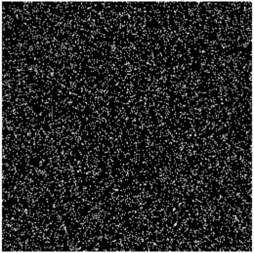
Outline

Type I problems

Typical setup: $U = \Psi^* \Phi$, where

- $\Psi \in \mathbb{C}^{N \times N}$ is the Fourier matrix,
- $\Phi \in \mathbb{C}^{N \times N}$ is a discrete wavelet transform.

Example: Recovery with $N = 256 \times 256$ and $m/N = 12.5\%$.



Subsampling map Ω

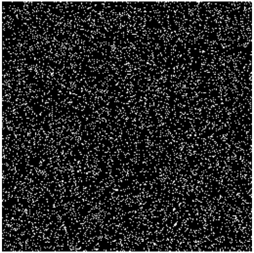
Original image

Type I problems

Typical setup: $U = \Psi^* \Phi$, where

- $\Psi \in \mathbb{C}^{N \times N}$ is the Fourier matrix,
- $\Phi \in \mathbb{C}^{N \times N}$ is a discrete wavelet transform.

Example: Recovery with $N = 256 \times 256$ and $m/N = 12.5\%$.



Subsampling map Ω

Reconstruction

High coherence

Explanation:

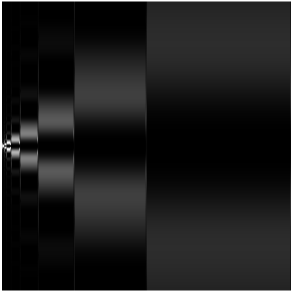
- $\mu(U) = \mathcal{O}(1)$ in this case, for any N and any wavelet.
- Hence the recovery guarantee saturates to $m \approx N$ in this case.

This phenomenon has been known since the earliest work in CS for applications such as MRI (see Lustig et al.).

Asymptotic incoherence

Although global coherence is high, there is a local incoherence structure:

- Coarse scale wavelets: coherent with low frequencies,
- Coarse scale wavelets: incoherent with high frequencies,
- Fine scale wavelets: incoherent with any frequencies.



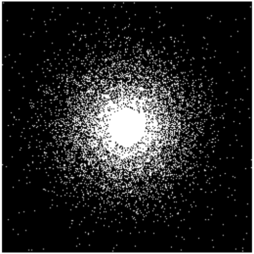
The absolute values of U

How to subsample the Fourier/wavelets matrix

Variable density sampling

- More samples at low frequencies (high coherence regions).
- Fewer samples at high frequencies (low coherence regions).

Example: Recovery with $N = 256 \times 256$ and $m/N = 12.5\%$.



Subsampling map Ω

Original image

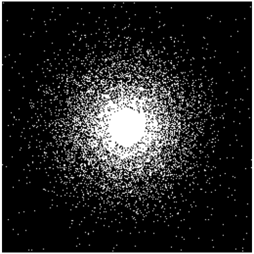
Conclusion: Local structure (coherence and sampling) matters.

How to subsample the Fourier/wavelets matrix

Variable density sampling

- More samples at low frequencies (high coherence regions).
- Fewer samples at high frequencies (low coherence regions).

Example: Recovery with $N = 256 \times 256$ and $m/N = 12.5\%$.



Subsampling map Ω

Reconstruction

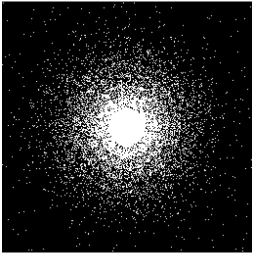
Conclusion: Local structure (coherence and sampling) matters.

How to subsample the Fourier/wavelets matrix

Variable density sampling

- More samples at low frequencies (high coherence regions).
- Fewer samples at high frequencies (low coherence regions).

Example: Recovery with $N = 256 \times 256$ and $m/N = 12.5\%$.



Subsampling map Ω

Reconstruction

Conclusion: Local structure (coherence and sampling) matters.

Related work

Variable density sampling:

- Lustig (2007), Lustig et al. (2007). Empirical observations, intuitive explanation.
- Wang & Arce (2010), Puy, Vandergheynst & Wiaux (2011),... Design of sampling strategies.

CS Theory (sparsity-based):

- Krahmer & Ward (2013), Boyer et al. (2012).

Sparsity?

Question: Does global sparsity explain the good reconstruction seen here?

The flip test

1. Given x, compute its wavelet coefficients $z = \Phi^* x$.
2. Permute the entries of z, giving z' .
3. Compute a new image $x' = \Phi z'$ with the same sparsity.
4. Run the same CS reconstruction on x and x', giving $\hat{x}$ and $\hat{x}'$.
5. Reverse the permutation on $\hat{x}'$ to get a new reconstruction $\check{x}$ of x.

Key point: Both z and z' have the same sparsity.

- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing*, arXiv:1302.0561 (2014).

Sparsity?

Question: Does global sparsity explain the good reconstruction seen here?

The flip test

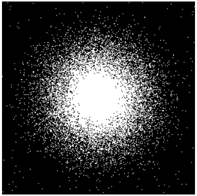
1. Given x, compute its wavelet coefficients $z = \Phi^* x$.
2. Permute the entries of z, giving z' .
3. Compute a new image $x' = \Phi z'$ with the same sparsity.
4. Run the same CS reconstruction on x and x', giving $\hat{x}$ and $\hat{x}'$.
5. Reverse the permutation on $\hat{x}'$ to get a new reconstruction $\check{x}$ of x.

Key point: Both z and z' have the same sparsity.

- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing*, arXiv:1302.0561 (2014).

The flip test

MRI example: $N = 256 \times 256$ and $m/N = 20\%$.

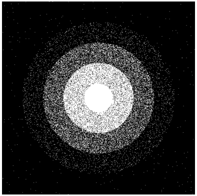


Subsampling map

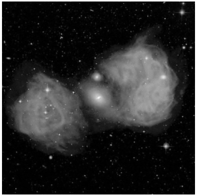
unflipped $\hat{x}$

flipped $\check{x}$

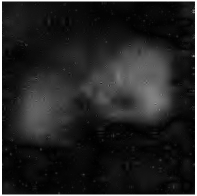
Radio interferometry example: $N = 512 \times 512$ and $m/N = 15\%$.



Subsampling map



unflipped $\hat{x}$

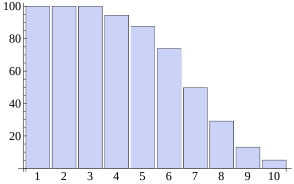


flipped $\check{x}$

Asymptotic sparsity

The flip test shows that sparsity is not the correct model: the ordering (local behaviour) of the coefficients matters.

Structured sparsity: Wavelet coefficients are asymptotically sparse.

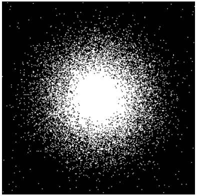


Left: image. Right: percentage of wavelet coefficients per scale $> 10^{-3}$.

At finer scales, more coefficients are negligible than at coarser scales.
The flip test destroys this structure, although it preserves overall sparsity.

Is this the correct model?

We perform a similar test, where the flipping is done within the scales.



Subsampling map

unflipped $\hat{x}$

flipped $\check{x}$

Conclusion: Sparsity within scales (i.e. a fixed number of nonzero per scale) appears to be the right model.

Outline

New concepts

Current global principles:

- Sparsity
- Incoherence
- Uniform random subsampling

New local principles:

- Sparsity in levels
- Local coherence in levels
- Multilevel random subsampling

Partitioning U

We first partition U into rectangular blocks indexed by levels

$$\mathbf{N} = (N_1, N_2, \ldots, N_r), \qquad \mathbf{M} = (M_1, M_2, \ldots, M_r),$$

where $N_r = M_r = n$ and $N_0 = M_0 = 0$.

$$U = \begin{pmatrix} U_{11} & U_{12} & \cdots & U_{1r} \\ U_{21} & U_{22} & \cdots & U_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ U_{r1} & U_{r2} & \cdots & U_{rr} \end{pmatrix}, \qquad U_{kl} \in \mathbb{C}^{(N_{k+1}-N_k)\times(M_{l+1}-M_l)}.$$

Note: The levels $\mathbf{M}$ need not be wavelet scales.

Sparsity in levels

Definition (Sparsity in levels)

A vector x is $(\mathbf{s}, \mathbf{M})$-sparse in levels, where $\mathbf{s} = (s_1, \ldots, s_r)$, if

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

- Models asymptotic sparsity of wavelet coefficients.
- Agrees with the flip test in levels.

Local coherence in levels

Definition (Local coherence in levels)

The $(k,l)^{\mathrm{th}}$ local coherence is $\mu(k,l) = \sqrt{\mu(U_{kl}) \max_t \mu(U_{kt})}$.

- Allows for varying coherence across U.
- E.g. the Fourier/wavelets matrix has $\mu(k,l) \to 0$ as $k,l \to \infty$.

Multilevel random subsampling

Definition (Multilevel random subsampling)

Let $\mathbf{m} = (m_1, \ldots, m_r)$ with $m_k \leq N_k - N_{k-1}$ and suppose that

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

is chosen uniformly at random. We call the set $\Omega = \Omega_1 \cup \cdots \cup \Omega_r$ an $(\mathbf{N}, \mathbf{m})$-multilevel subsampling scheme.

- Models variable density sampling by allowing varying m_k's.
- For Fourier/wavelets, we have $m_k/(N_k - N_{k-1}) \to 0$.

Interferences and relative sparsities

The matrix U is not block diagonal in general. Hence there may be interferences between sparsity levels.

To handle this, we need:

Definition

Let $x \in \mathbb{C}^N$ be $(\mathbf{s}, \mathbf{M})$-sparse. Given $\mathbf{N}$, we define the relative sparsity

$$S_k = S_k(\mathbf{s}, \mathbf{M}, \mathbf{N}) = \max_{\eta \in \Theta} \left\| \sum U_{kl} \eta_l \right\|^2,$$

where $\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1, \ \eta \text{ is } (\mathbf{s}, \mathbf{M})\text{-sparse}\}$.

Main result

Theorem (BA, Hansen, Poon & Roman)

Given $\mathbf{N}$ *and* $\mathbf{m}$ *suppose that* $\mathbf{s}$ *and* $\mathbf{M}$ *are such that*

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

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

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

If $\hat{x}$ *is a minimizer, then with probability at least* $1 - s\epsilon$ *we have*

$$\|x - \hat{x}\|_2 \lesssim \sigma_{\mathbf{s},\mathbf{M}}(x) + L\sqrt{s}\delta,$$

where $s = s_1 + \ldots + s_r$ *and* $L = 1 + \frac{\sqrt{\log_(\epsilon^{-1})}}{\log(4N\sqrt{s})}$.

- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: a new theory for compressed sensing*, arXiv:1302.0561 (2014).

Application to the Fourier/wavelets problem

For the discrete Fourier/Haar wavelet problem, one can show that

$$\mu(k,l) \lesssim 2^{-k} 2^{-|k-l|/2},$$

and

$$S_k \lesssim \sum_{l=1}^{r} 2^{-|k-l|/2} s_l,$$

provided the sampling levels are correspond to dyadic frequency bands. Hence the recovery guarantee reduces to

$$m_k \gtrsim \left(s_k + \sum_{l \neq k} 2^{-|k-l|/2} s_l \right) \cdot \log(\epsilon^{-1}) \cdot \log(N).$$

- BA, Hansen & Roman, *A note on compressed sensing of structured sparse wavelet coefficients from subsampled Fourier measurements*, arXiv:1403.6541 (2014).

Application to the Fourier/wavelets problem

The estimate

$$m_k \gtrsim \left(s_k + \sum_{l \neq k} 2^{-|k-l|/2} s_l \right) \cdot \log(\epsilon^{-1}) \cdot \log(N).$$

is optimal up to exponentially-decaying factors in $|k-l|$.

- Variable density sampling works because of asymptotic sparsity.
- As the sparsity increases, more subsampling is permitted in the corresponding high-frequency bands.
- This estimate also agrees with the flip test.

Note: The estimate generalizes to arbitrary wavelets, with $\sqrt{2}$ replaced by $A > 1$ depending on the smoothness and number of vanishing moments.

- BA, Hansen, Poon & Roman, *Breaking the coherence barrier: a new theory for compressed sensing*, arXiv:1302.0561 (2014).

Effect/benefits of this theory for type I problems

1. New framework explains why CS works in MRI, radio interferometry, X-ray CT,...

2. New insight into the design of sampling trajectories.
   - Nontrivial – must take into account physical limitations
   - Necessarily image-dependent – no one size fits all

3. Changes understanding on the benefits of CS in such applications.
   - Previous understanding: low(ish) resolution, scan time reduction
   - New understanding: higher resolution, increasing image quality
   - To quote Siemens (see Proc. Intl. Soc. Mag. Reson. Med., 2014):

     ...the full potential of the compressed sensing is unleashed only if asymptotic sparsity and asymptotic incoherence is achieved.

- Roman, BA & Hansen, *On asymptotic structure in compressed sensing*, arXiv:1406.4178, 2014.

Effect/benefits of this theory for type I problems

1. New framework explains why CS works in MRI, radio interferometry, X-ray CT,...

2. New insight into the design of sampling trajectories.
   - Nontrivial – must take into account physical limitations
   - Necessarily image-dependent – no one size fits all

3. Changes understanding on the benefits of CS in such applications.
   - Previous understanding: low(ish) resolution, scan time reduction
   - New understanding: higher resolution, increasing image quality
   - To quote Siemens (see Proc. Intl. Soc. Mag. Reson. Med., 2014):

> *...the full potential of the compressed sensing is unleashed only if asymptotic sparsity and asymptotic incoherence is achieved.*

- Roman, BA & Hansen, *On asymptotic structure in compressed sensing*, arXiv:1406.4178, 2014.

Effect/benefits of this theory for type I problems

1. New framework explains why CS works in MRI, radio interferometry, X-ray CT,...

2. New insight into the design of sampling trajectories.
   - Nontrivial – must take into account physical limitations
   - Necessarily image-dependent – no one size fits all

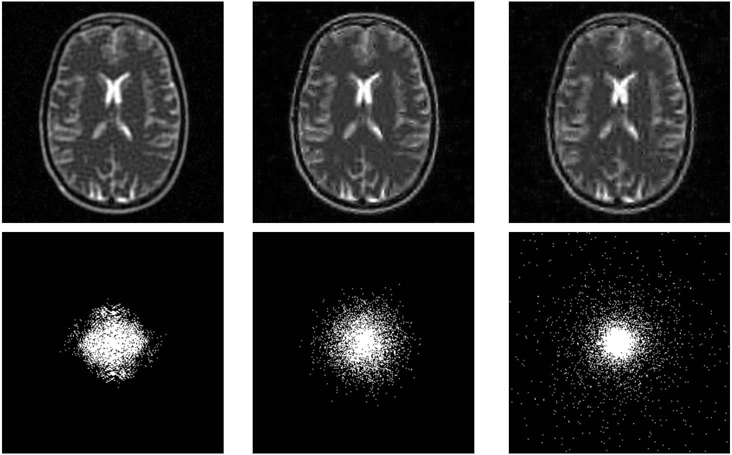
3. Changes understanding on the benefits of CS in such applications.
   - Previous understanding: low(ish) resolution, scan time reduction
   - New understanding: higher resolution, increasing image quality
   - To quote Siemens (see Proc. Intl. Soc. Mag. Reson. Med., 2014):

...the full potential of the compressed sensing is unleashed only if asymptotic sparsity and asymptotic incoherence is achieved.

- Roman, BA & Hansen, *On asymptotic structure in compressed sensing,* arXiv:1406.4178, 2014.

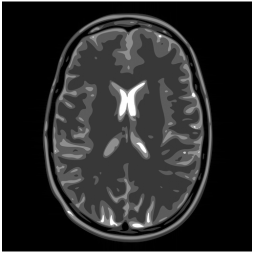
Resolution dependence – low resolution

5% samples at 256×256 resolution. Substantial subsampling is not possible, regardless of the scheme:

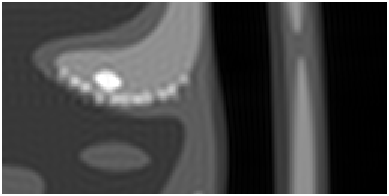


Oracle, Err = 18%   Multilevel, Err = 19%   Power law, Err = 22%

Resolution dependence – high resolution



At higher resolutions there is more asymptotic incoherence and sparsity.
Taking the same number of measurements, CS recovers the fine details.



512^2 lowest frequency coefficients

CS reconstruction

Outline

Introduction

Compressed sensing recap

Limitations of the current theory

A level-based theory of compressed sensing

A new compressive imaging paradigm

Conclusions

Type II problems

Unlike type I problems, in type II problems we have substantial freedom to design the sensing matrix Ψ.

Applications: compressive imaging (single-pixel, lensless), infrared imaging, fluorescence microscopy,...

Hardware constraint: Typically $\Psi \in \{0, 1\}^N$.

Sparsifying transform: We typically use a wavelet transform Φ as before.

Conventional CS approach

Use a Bernoulli random matrix and ℓ^1 minimization.

Limitations:

1. Ψ is dense and unstructured, i.e. computationally infeasible.
   - Solution: replace Ψ by fast transforms.
   - E.g. subsampled DCT with column randomization or random convolutions.

2. Only exploits the sparsity of the wavelet coefficients, and no further structure. Recovery quality is limited.

Enhancing reconstruction quality with structured recovery

Basic principle: wavelet coefficients lie on connected trees (persistence across scales model (Mallat)).

Structured recovery: Modify the recovery algorithm (typically a thresholding or greedy method) to enforce this type of structured sparsity. Use standard (i.e. incoherent) measurements.

State-of-the-art approaches:

- Model-based CS (Baraniuk et al.)
- HGL (Cevher et al.)
- TurboAMP (Som & Schniter)
- Bayesian CS (Chen & Carin)

New paradigm: structured sampling

Keep the standard recovery algorithm (ℓ^1 minimization) and modify the measurements to promote asymptotic sparsity in scales.

Practical implementation:

- Walsh–Hadamard transform Ψ (binary)
- Multilevel random subsampling according to wavelet scales

- Roman, BA & Hansen, *On asymptotic structure in compressed sensing,* arXiv:1406.4178, 2014.

Example (12.5% subsampling at 256×256 resolution)

ℓ^1 min., Bern.
Err = 16.0%

modelCS, Bern.
Err = 17.0%

TurboAMP, Bern.
Err = 13.1%

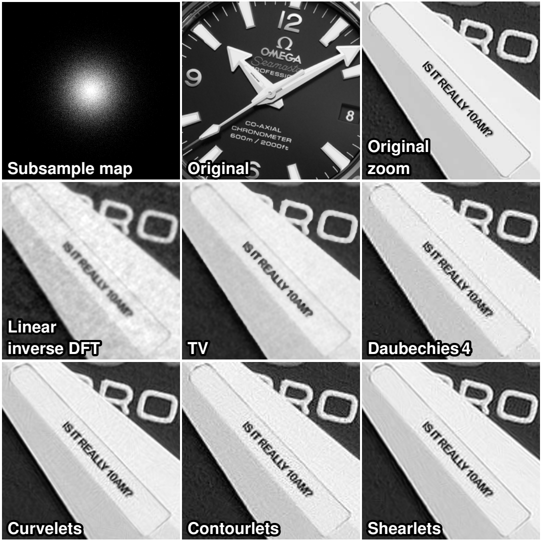
Bayesian, Bern.
Err = 12.6%

ℓ^1 min, Had., db4
Err = 9.5%

ℓ^1 min, Had., DT-CWT
Err = 8.6 %

Other advantages

It is also easy to change the sparsifying transform:



Other advantages

Fast transforms combined with efficient ℓ^1 algorithms (we use SPGL1 throughout) mean we can do high resolution imaging.

Example: The Berlin cathedral with 15% sampling at various resolutions using Daubechies-4 wavelets.

Efficient compressive imaging

Resolution: 128×128

Reconstruction (cropped)

RAM (GB): < 0.1
Speed (it/s): 26.4
Rel. Err. (%): 17.9
Time: 10.1s

Original image (cropped)

Efficient compressive imaging

Resolution: 256×256

Reconstruction (cropped)

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

Original image (cropped)

Efficient compressive imaging

Resolution: 512×512

Reconstruction (cropped)

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

Original image (cropped)

Efficient compressive imaging

Resolution: 1024×1024

Reconstruction (cropped)

RAM (GB): < 0.1
Speed (it/s): 1.07
Rel. Err. (%): 10.4
Time: 3m45s

Original image (cropped)

Efficient compressive imaging

Resolution: 2048×2048

Reconstruction (cropped)

RAM (GB): < 0.1
Speed (it/s): 0.17
Rel. Err. (%): 8.5
Time: 28m

Original image (cropped)

Efficient compressive imaging

Resolution: 4096×4096

Reconstruction (cropped)

RAM (GB): < 0.1
Speed (it/s): 0.041
Rel. Err. (%): 6.6
Time: 1h37m

Original image (cropped)

Efficient compressive imaging

Resolution: 8192×8192

Reconstruction (cropped)

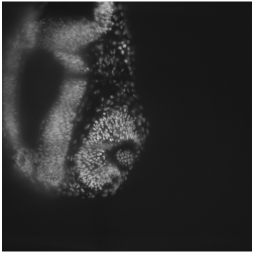
RAM (GB): < 0.1
Speed (it/s): 0.0064
Rel. Err. (%): 3.5
Time: 8h30m

Original image (cropped)

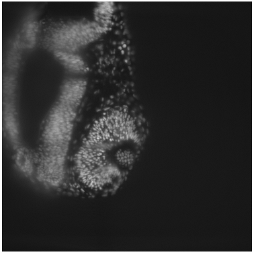
Application to fluorescence microscopy

We may apply this approach to fluorescence microscopy. This has to two key advantages:

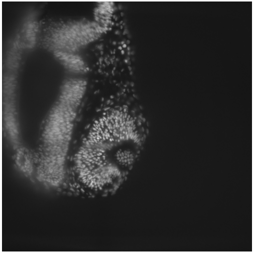
- Better inherent performance, due to structured sparsity.
- Mitigation of the point spread effect, since more of the measurements are taken at lower (Hadamard) frequencies.



Original image



Current CS*



New CS

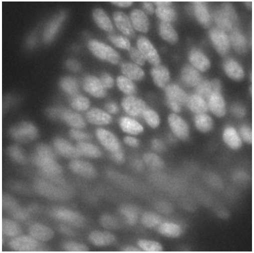
* See Studer, Bobin, Chahid, Mousavi, Candès & Dahan (2012).

Image of zebrafish cells, courtesy of the Cambridge Advanced Imaging Centre (CAIC). Practical CS fluorescence microscope under construction.

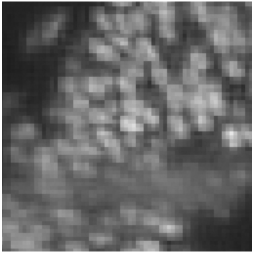
Application to fluorescence microscopy

We may apply this approach to fluorescence microscopy. This has to two key advantages:

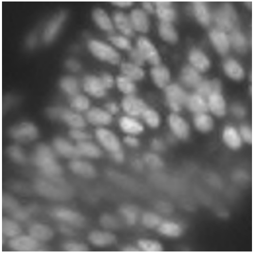
- Better inherent performance, due to structured sparsity.
- Mitigation of the point spread effect, since more of the measurements are taken at lower (Hadamard) frequencies.



Original image



Current CS*



New CS

* See Studer, Bobin, Chahid, Mousavi, Candès & Dahan (2012).

Image of zebrafish cells, courtesy of the Cambridge Advanced Imaging Centre (CAIC). Practical CS fluorescence microscope under construction.

Outline

Conclusions

- The standard CS principles are ill-suited to typical type I imaging problems (e.g. MRI).
- In these applications, local behaviour plays a crucial role.
- A new CS framework based on sparsity in levels, local coherence in levels and multilevel random subsampled was introduced. It establishes the key role of local structure in CS for type I problems.
- This not only explains the success of CS in many applications, it also provides new insights and techniques for maximizing its performance in type II problems.