Compressed sensing

Clarice Poon
University of Bath

April 3, 2020

Outline

Sparsity in imaging

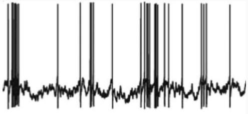
Compressed sensing

Uniform recovery of sparse vectors via ℓ^1 minimisation

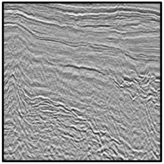
Recovery with incoherent bases

Compressed sensing with Fourier measurements

Sparse estimation

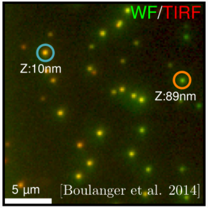


Neural spikes



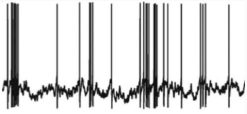
Seismic imaging

Astronomy

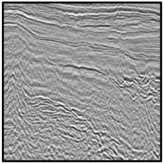


Fluorescence microscopy

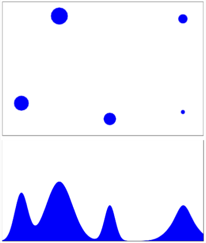
Sparse estimation



Neural spikes

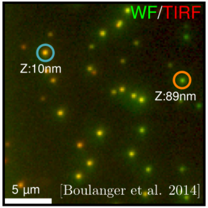


Seismic imaging



Mixture estimation

Astronomy



Fluorescence microscopy

Discrete representations

Task

Given a function $f(t)$ defined for $t \in \mathbb{R}$, (say, $f \in L^2(\mathbb{R})$), how can we transmit/store/analyse/reconstruct this function from finitely many values?

Discrete representations

Task

Given a function $f(t)$ defined for $t \in \mathbb{R}$, (say, $f \in L^2(\mathbb{R})$), how can we transmit/store/analyse/reconstruct this function from finitely many values?

To give some examples:

1. f is a voice signal and we want to transmit it over a telephone line.
2. f is the cross-section of a body whose image we want to reconstruct using finitely many samples.
3. f is an image that we want to store using finitely many values.

Task

Given a function $f(t)$ defined for $t \in \mathbb{R}$, (say, $f \in L^2(\mathbb{R})$), how can we transmit/store/analyse/reconstruct this function from finitely many values?

To give some examples:

1. f is a voice signal and we want to transmit it over a telephone line.
2. f is the cross-section of a body whose image we want to reconstruct using finitely many samples.
3. f is an image that we want to store using finitely many values.

Suppose we have an orthonormal basis $\{g_n : n \in \mathbb{Z}\}$ in $L^2(\mathbb{R})$, then we know that

$$f = \sum_{n\in\mathbb{Z}} c_n g_n, \qquad c_n = \langle f, g_n \rangle.$$

The coefficients $\{c_n\}_{n\in\mathbb{Z}}$ provides a discrete representation of f. In practice, we will choose some finite set $\Lambda \subset \mathbb{Z}$ and process only the coefficients $\{c_n\}_{n\in\Lambda}$. One would hope that

$$f \approx \sum_{n\in\Lambda} c_n g_n.$$

The Fourier basis

Recall that, the Fourier transform of $f \in L^1(\mathbb{R})$ is defined by

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-ix\xi} \mathrm{d}x, \qquad \xi \in \mathbb{R}$$

and this definition can be extended to $L^2(\mathbb{R})$ since $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$.

From classical Fourier analysis, we know that

$$\left\{ \frac{1}{\sqrt{2B\pi}} e^{iB^{-1}k\cdot} \ ; \ k \in \mathbb{Z} \right\}$$

is an orthonormal basis of $L^2([-B\pi, B\pi])$. So, given any $f \in L^2([-B\pi, B\pi])$,

$$f(x) = \frac{1}{2B\pi} \sum_{k \in \mathbb{Z}} \hat{f}(kB^{-1}) e^{ikB^{-1}x}. \tag{1.1}$$

The Shannon-Nyquist-Whittake sampling theorem (1950)

Theorem

Suppose $\hat{f}$ is piecewise smooth and continuous and $\hat{f}(\xi) = 0$ for all $|\xi| > B\pi$. Then,

$$f(x) = \sum_{k \in \mathbb{Z}} f\left(\frac{k}{B}\right) \varphi\left(x - \frac{k}{B}\right),$$

where $\varphi(x) = \frac{\sin(\pi B x)}{\pi B x}$. We also have that

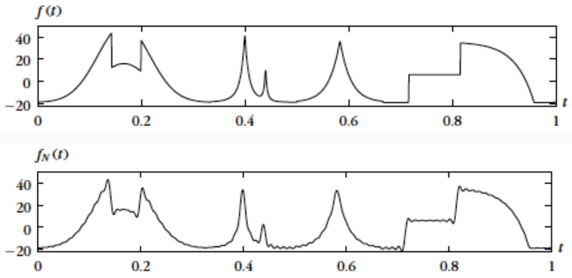
$$f_N = \sum_{|k| \leqslant N} f\left(\frac{k}{B}\right) \varphi\left(\cdot - \frac{k}{B}\right) \to f \qquad \text{in } L^\infty(\mathbb{R}).$$

The Shannon-Nyquist-Whittaker theorem provides a discrete representation of functions and describes how one may approximate f with finitely many values. Forms the basis of modern signal processing and communication theory.

Drawbacks

- Fourier approximations or Shannon approximations are better for approximating smooth signals. Natural images have discontinuities...
- Fourier representations have the drawback of requiring many samples or coefficients to represent localized events.

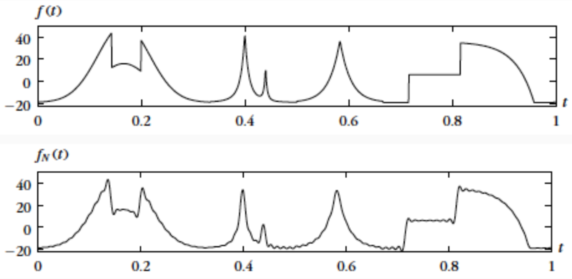
Approximation with $N = 128$ Fourier coefficients:



Drawbacks

- Fourier approximations or Shannon approximations are better for approximating smooth signals. Natural images have discontinuities...
- Fourier representations have the drawback of requiring many samples or coefficients to represent localized events.

Approximation with $N = 128$ Fourier coefficients:



Wavelets, developed between the 1990's and early 2000's form an alternative basis which are much better for approximating piecewise smooth signals.

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$\begin{cases} \text{Averages : } 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \end{cases}$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : -4, -46.5, 26, -1, \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : -4, -46.5, 26, -1, \end{cases} \qquad \begin{cases} A : 210.375, 124, \\ D : -21.25, 10, \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : -4, -46.5, 26, -1, \end{cases} \quad \begin{cases} A : 210.375, 124, \\ D : -21.25, 10, \end{cases} \quad \begin{cases} A : 167.188, \\ D : -86.375 \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : -4, -46.5, 26, -1, \end{cases} \qquad \begin{cases} A : 210.375, 124, \\ D : -21.25, 10, \end{cases} \qquad \begin{cases} A : 167.188, \\ D : -86.375 \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 0, 0, 0, -99, 0, 0, 0, 0, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : 0, -46.5, 26, 0, \end{cases} \quad \begin{cases} A : 210.375, 124, \\ D : -21.25, 0, \end{cases} \quad \begin{cases} A : 167.188, \\ D : -86.375 \end{cases}$$

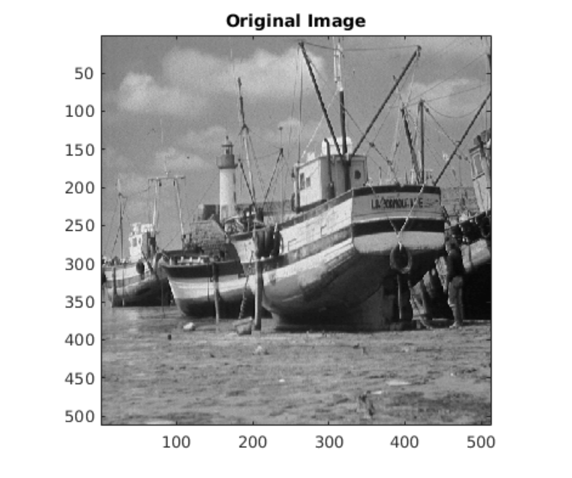
If we set all number $\leqslant 10$ to 0, then we retain only 6 values and the reconstructed sequence

$$221, 221, 221, 221, 223, 223, 226, 127, 111, 111, 137, 137, 124, 124, 124, 124$$

with absolute error:

$$3, 7, 6, 2, 2, 2, 0, 0, 5, 5, 5, 5, 5, 6, 5, 4$$

Wavelet decomposition



Wavelet decomposition

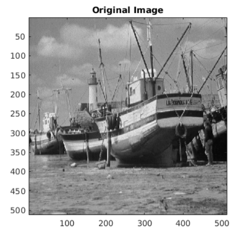
Wavelet decomposition

Wavelet decomposition

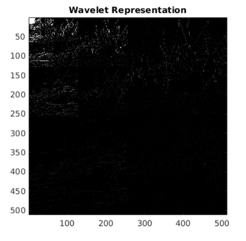
Wavelet decomposition

Wavelet decomposition

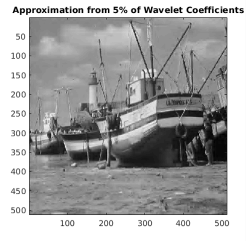
Sparse approximation with wavelets



I



$\{\langle I, \psi_{j,k}\rangle : k, j \in \mathbb{Z}\}$



$\sum_{k,j\in\Lambda}\langle I, \psi_{j,k}\rangle\psi_{j,k}$

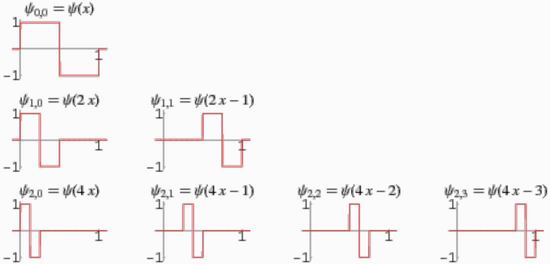
Wavelet definition

Wavelet

We say that a function $\psi \in L^2(\mathbb{R})$ is a wavelet for $L^2(\mathbb{R})$ if

$$\left\{\psi_{j,k}(t) := 2^{j/2}\psi(2^j t - k)\ ;\ j,k \in \mathbb{Z}\right\}$$

forms an orthonormal basis of $L^2(\mathbb{R})$.



$\psi = \mathbb{1}_{[-1,-1/2)} - \mathbb{1}_{[-1/2,0)}$,

A (very) brief history

- In 1910, Haar constructed the first wavelet basis (although it was not known as such).

A (very) brief history

- In 1910, Haar constructed the first wavelet basis (although it was not known as such).
- In 1980, Strömberg discovered a piecewise linear wavelet ψ which yields better approximation properties for smooth functions.

A (very) brief history

- In 1910, Haar constructed the first wavelet basis (although it was not known as such).
- In 1980, Strömberg discovered a piecewise linear wavelet ψ which yields better approximation properties for smooth functions.
- Unaware of this result, Meyer tried to prove that there does not exist a regular wavelet which generates an orthonormal basis. Instead of proving this, his attempt led to the construction of an entire family of orthonormal wavelet bases which are infinitely continuously differentiable.

A (very) brief history

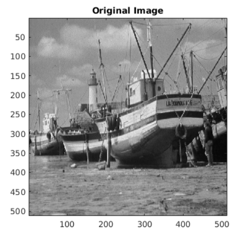
- In 1910, Haar constructed the first wavelet basis (although it was not known as such).
- In 1980, Strömberg discovered a piecewise linear wavelet ψ which yields better approximation properties for smooth functions.
- Unaware of this result, Meyer tried to prove that there does not exist a regular wavelet which generates an orthonormal basis. Instead of proving this, his attempt led to the construction of an entire family of orthonormal wavelet bases which are infinitely continuously differentiable.
- The work of Meyer led to a scurry of research on wavelets throughout the late 1980's and 1990's. In the following sections, we shall study the systematic approach of constructing orthonormal wavelet bases via multiresolution analysis, which was established by Meyer and Mallat.

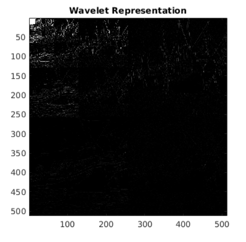
A (very) brief history

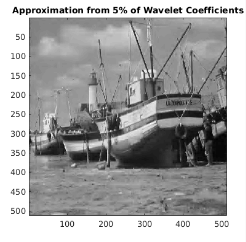
- In 1910, Haar constructed the first wavelet basis (although it was not known as such).
- In 1980, Strömberg discovered a piecewise linear wavelet ψ which yields better approximation properties for smooth functions.
- Unaware of this result, Meyer tried to prove that there does not exist a regular wavelet which generates an orthonormal basis. Instead of proving this, his attempt led to the construction of an entire family of orthonormal wavelet bases which are infinitely continuously differentiable.
- The work of Meyer led to a scurry of research on wavelets throughout the late 1980's and 1990's. In the following sections, we shall study the systematic approach of constructing orthonormal wavelet bases via multiresolution analysis, which was established by Meyer and Mallat.
- Today, JPEG 2000 standard provides compression architectures based on wavelets.

Key takeaway on wavelets

Typical images are sparse in wavelets.







I $\quad\quad$ $\{\langle I, \psi_{j,k}\rangle : k, j \in \mathbb{Z}\}$ $\quad\quad$ $\sum_{k,j\in\Lambda}\langle I, \psi_{j,k}\rangle\psi_{j,k}$

There are **fast algorithms** for numerical computation.
One attractive aspect of the Fourier transform is that the **Fast Fourier Transform** can be computed in $\mathcal{O}(n \log n)$ time.
For wavelets, we have the **Discrete Wavelet Transform** which can be computed in $\mathcal{O}(n)$ time.

Outline

Compressed sensing

Task: Given $y = Ax$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x.

In general, this is impossible, since we have more unknowns than knowns.

Compressed sensing

Task: Given $y = Ax$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x.

In general, this is impossible, since we have more unknowns than knowns.

But suppose that x is sparse with $s = |\mathrm{Supp}(x)| \ll n$?

Compressed sensing

Task: Given $y = Ax$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x.

In general, this is impossible, since we have more unknowns than knowns.

But suppose that x is sparse with $s = |\mathrm{Supp}(x)| \ll n$?

Still no... let $n = 100$ and let x be 1-sparse with its last entry being nonzero.

Even if we observe 10 entries uniformly at random, with high probability, we observe only zeros.

Task: Given $y = Ax$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x.

In general, this is impossible, since we have more unknowns than knowns.

But suppose that x is sparse with $s = |\mathrm{Supp}(x)| \ll n$?

Still no... let $n = 100$ and let x be 1-sparse with its last entry being nonzero.

Even if we observe 10 entries uniformly at random, with high probability, we observe only zeros.

Compressed sensing is about the design of matrices A such that s-sparse vectors are recovered from $\approx s$ (up to logarithmic factors) measurements.

- x is sparse.
- The rows of A do not correlate well with x... information is evenly spread over $\langle a_k, x \rangle$.
- A is a random matrix.

Compressed sensing [Candés, Romberg & Tao (2006); Donoho (2006)]

Key outcome of compressed sensing:

We can recover sparse vectors of length N from $m \ll N$ randomised linear measurements by solving the following convex optimisation problem:

$$\min_x \|x\|_1 \text{ subject to } Ax = y. \tag{BP}$$

Key outcome of compressed sensing:

We can recover sparse vectors of length N from $m \ll N$ randomised linear measurements by solving the following convex optimisation problem:

$$\min_x \|x\|_1 \text{ subject to } Ax = y. \tag{BP}$$

Practical remark: Suppose that $f \in \mathbb{C}^N$ is sparse in some orthonormal basis Ψ. That is, $f = \Psi x_0$ for some sparse vector x_0. Then,

$$y = Af = A \circ \Psi x_0 = \Phi x_0.$$

We therefore solve

$$\hat{x} \in \operatorname{argmin}_x \|x\|_1 \text{ subject to } \Phi x = y$$

and take the reconstruction to be $\hat{f} = \Psi \hat{x}$.

Compressed sensing [Candés, Romberg & Tao (2006); Donoho (2006)]

Key outcome of compressed sensing:

We can recover sparse vectors of length N from $m \ll N$ randomised linear measurements by solving the following convex optimisation problem:

$$\min_x \|x\|_1 \text{ subject to } Ax = y. \tag{BP}$$

Practical remark: Suppose that $f \in \mathbb{C}^N$ is sparse in some orthonormal basis Ψ. That is, $f = \Psi x_0$ for some sparse vector x_0. Then,

$$y = Af = A \circ \Psi x_0 = \Phi x_0.$$

We therefore solve

$$\hat{x} \in \operatorname{argmin}_x \|x\|_1 \text{ subject to } \Phi x = y$$

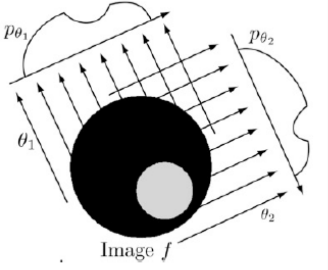
and take the reconstruction to be $\hat{f} = \Psi \hat{x}$.

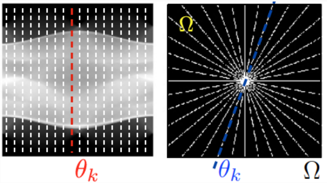
ℓ_1 has already been used extensively since the 1980's for promoting sparsity. However, the 2006 articles proved that ℓ_1 enables sparse recovery using far fewer measurements than previously thought possible.

Applications of compressed sensing – Fourier measurements

Many imaging devices can be seen as providing pointwise samples of the Fourier transform.

- Magnetic resonance imaging
- Radio interferometry
- Electron microscopy
- Tomography.





For tomography, if p_θ is the Radon projection of f at angle θ, then the Fourier splice theorem says:

$$\hat{p}_\theta(t) = \hat{f}(t\cos(\theta), t\sin(\theta)).$$

We therefore are interested in $y = P_\Omega \mathcal{F} W^{-1} x$.

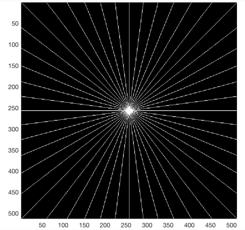
The matlab phantom experiment [Candès, Romberg and Tao '06]

Let $P_\Omega \mathcal{F} x = (\hat{x}_j)_{j\in\Omega}$. Given observations $y_0 = P_\Omega \mathcal{F} x_0$, take the reconstruction z as

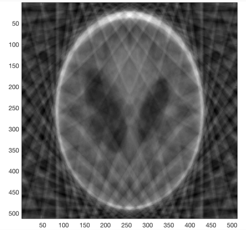
$$\operatorname*{argmin}_x \|Wx\|_1 \text{ subject to } P_\Omega \mathcal{F} x = y_0$$

If W is invertible, this is equivalent to

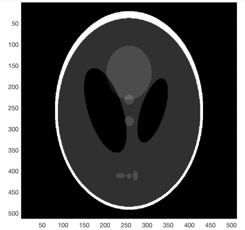
$$\operatorname*{argmin}_x \|x\|_1 \text{ subject to } P_\Omega \mathcal{F} W^{-1} x = y_0$$



Sampling map Ω



$\mathcal{F}^{-1} P_\Omega y_0$

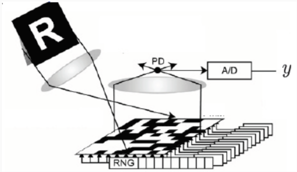


Sparse reconstruction

Let $z = \mathbb{R}^N$.

The single pixel camera is a microarray consisting of N mirrors, each of which can be switched on or off individually.

The light from the image is reflected on the micro array, and a lense then combines all reflected beams in one sensor.

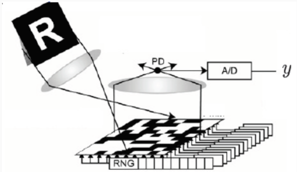


The single pixel camera [Duarte, Davenporte, Takhar, Laska, Sun, Kelly, Baraniuk '08]

Let $z = \mathbb{R}^N$.

The single pixel camera is a microarray consisting of N mirrors, each of which can be switched on or off individually.

The light from the image is reflected on the micro array, and a lense then combines all reflected beams in one sensor.



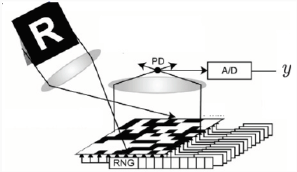
Each measurement is $\langle z, b \rangle$ where b is a vector consisting of 1's at locations where the mirrors are 'on' and 0 where the mirrors are 'off'.

The single pixel camera [Duarte, Davenporte, Takhar, Laska, Sun, Kelly, Baraniuk '08]

Let $z = \mathbb{R}^N$.

The single pixel camera is a microarray consisting of N mirrors, each of which can be switched on or off individually.

The light from the image is reflected on the micro array, and a lense then combines all reflected beams in one sensor.



Each measurement is $\langle z, b \rangle$ where b is a vector consisting of 1's at locations where the mirrors are 'on' and 0 where the mirrors are 'off'.

$m/N = 1$

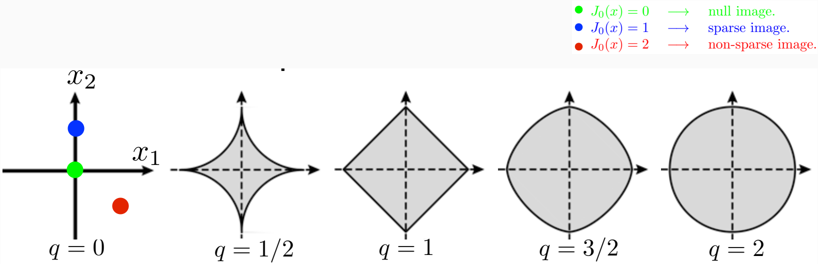
$m/N = 0.16$

$m/N = 0.02$

Outline

Let $\|x\|_q^q = \sum_j |x_j|^q$. Convex when $q \geqslant 1$ and "close to" ℓ_0 for small q.

$$\min_x \|x\|_p \text{ subject to } \Phi x = y$$



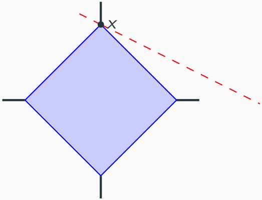
We have x is the unique solution to

$$\min_z \|z\|_1 \text{ subject to } Az = Ax$$

if and only if

$$\mathcal{F}_x \cap \mathcal{B}_x = \{x\}$$

where $\mathcal{F}_x \overset{\text{def.}}{=} \{z \,;\, Az = Ax\}$ and $\mathcal{B}_x \overset{\text{def.}}{=} \left\{z \,;\, \|z\|_1 \leqslant \|x\|_1\right\}$



Why does ℓ_1 work?

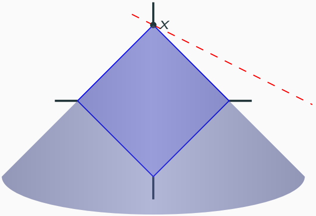
We have x is the unique solution to

$$\min_z \|z\|_1 \text{ subject to } Az = Ax$$

if and only if

$$\mathcal{F}_x \cap \mathcal{B}_x = \{x\}$$

where $\mathcal{F}_x \stackrel{\text{def.}}{=} \{z \,;\, Az = Ax\}$ and $\mathcal{B}_x \stackrel{\text{def.}}{=} \left\{z \,;\, \|z\|_1 \leqslant \|x\|_1\right\}$



In fact, this is equivalent to $\mathcal{N}(A) \cap \mathcal{C} = \{0\}$ where $\mathcal{C} \stackrel{\text{def.}}{=} \cup_{t>0} \left\{z \,;\, \|x + tz\|_1 \leqslant \|x\|_1\right\}$ is the descent cone.

In high dimensions, the cone at x is 'narrow' and $\mathcal{N}(A)$ which has codimension m misses $\mathcal{C}$ when m is large enough.

Null space property
$A \in \mathbb{C}^{m \times N}$ is said to satisfy the NSP relative to a set $S \subset [N]$ if

$$\|v_S\|_1 < \|v_{S^c}\|_1, \qquad \forall v \in \mathcal{N}(A) \setminus \{0\}$$

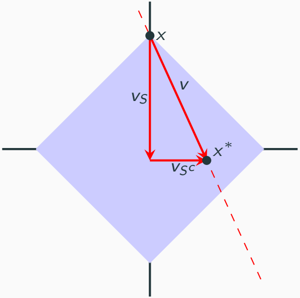
It is said to satisfy the NSP of order s if this holds for all $S \subset [N]$ with $|S| \leqslant s$.

Theorem

Given $A \in \mathbb{C}^{m \times N}$, every $x \in \mathbb{C}^N$ supported on $S \subset [N]$ is the unique solution to (BP) if and only if A satisfies the NSP relative to set S.

We say that A satisfies the null space property of order s if A satisfies the NSP relative to all S with $|S| \leqslant s$. Note that this implies unique recovery of s-sparse vectors via (BP).

Suppose that A satisfies the NSP on S and x is not the unique solution to (BP), i.e. there exists $x^* \neq x$ such that $Ax^* = Ax$ and $\|x^*\|_1 \leqslant \|x\|_1$.

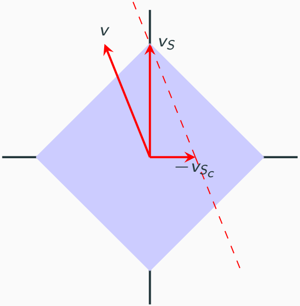


$v \overset{\text{def.}}{=} x^* - x \in \mathcal{N}(A) \setminus \{0\}$ satisfies:

$$\begin{aligned} \|v_{S^c}\|_1 &= \|x^*_{S^c}\|_1 = \|x^*_{S^c}\|_1 - \|x\|_1 + \|x - x^*_S + x^*_S\|_1 \\ &\leqslant \|x^*_{S^c}\|_1 - \|x\|_1 + \|v_S\|_1 + \|x^*_S\|_1 \leqslant \|v_S\|_1 \end{aligned}$$

Null space property

Suppose that x is the unique solution to (BP) and suppose that A does not satisfy the NSP. i.e. $\exists v \in \mathcal{N}(A) \setminus \{0\}$ s.t. $\|v_S\|_1 \geqslant \|v_{S^c}\|_1$



Let $x \stackrel{\text{def.}}{=} v_S$. Then, $Av_S = -Av_{S^c}$ but x is not the unique solution to (BP). $\mathcal{F}_x = \{z \; ; \; Az = Ax\}$ is the dotted red line.

Robust and stable recovery

Let $y = Ax + e$ with $\|e\| \leqslant \eta$. What conditions should we impose on A such that

$$\mathcal{R}^\eta(y) \stackrel{\text{def.}}{=} \operatorname{argmin} \|z\|_1 \text{ subject to } \|Az - y\|_2 \leqslant \eta.$$

satisfies $\|x - \mathcal{R}^\eta(y)\|_2 \leqslant \frac{C}{\sqrt{s}} \sigma_s(x)_1 + D\eta$ for some $C, D > 0$?

Let $y = Ax + e$ with $\|e\| \leqslant \eta$. What conditions should we impose on A such that

$$\mathcal{R}^\eta(y) \stackrel{\text{def.}}{=} \operatorname{argmin} \|z\|_1 \text{ subject to } \|Az - y\|_2 \leqslant \eta.$$

satisfies $\|x - \mathcal{R}^\eta(y)\|_2 \leqslant \frac{C}{\sqrt{s}} \sigma_s(x)_1 + D\eta$ for some $C, D > 0$?

Robust null space property

We say that A satisfies the robust NSP with constant $\rho, \tau > 0$ if

$$\|v_S\|_2 \leqslant \frac{\rho}{\sqrt{s}} \|v_{S^c}\|_1 + \tau \|Av\|_2, \qquad \forall v \in \mathbb{C}^N.$$

Robust and stable recovery

Let $y = Ax + e$ with $\|e\| \leqslant \eta$. What conditions should we impose on A such that

$$\mathcal{R}^\eta(y) \stackrel{\text{def.}}{=} \operatorname{argmin} \|z\|_1 \text{ subject to } \|Az - y\|_2 \leqslant \eta.$$

satisfies $\|x - \mathcal{R}^\eta(y)\|_2 \leqslant \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta$ for some $C, D > 0$?

Robust null space property

We say that A satisfies the robust NSP with constant $\rho, \tau > 0$ if

$$\|v_S\|_2 \leqslant \frac{\rho}{\sqrt{s}} \|v_{S^c}\|_1 + \tau \|Av\|_2, \qquad \forall v \in \mathbb{C}^N.$$

- robust NSP with $\rho < 1$ implies the NSP.
- If we have stable and robust recovery, then setting $x \stackrel{\text{def.}}{=} v \in \mathbb{C}^N$, $e = -Av$ and $\eta = \|Av\|_2$, we have $\mathcal{R}^\eta(Ax + e) = 0$ and $\|v\|_2 \leqslant \frac{C}{\sqrt{s}}\sigma_s(v)_1 + D\|Av\|_2$. So, this condition is necessary.
- The robust NSP with $\rho < 1$ is sufficient for robust and stable recovery.

This is one way to assess the quality of the matrix A for recovering s-sparse vectors.

The Restricted Isometry Property

The sth restricted isometry constant δ_s of a matrix A is the smallest $\delta > 0$ such that

$$(1-\delta)\|x\|_2^2 \leqslant \|Ax\|_2^2 \leqslant (1+\delta)\|x\|_2^2,$$

for all s-sparse vectors $x \in \mathbb{C}^N$.

- $\delta_s = \max_{|S| \leqslant s} \|A_S^* A_S - \mathrm{Id}\|$.
- All singular values of A_S are restricted to $[1-\delta_s, 1+\delta_s]$.

Theorem (RIP $\implies$ robust NSP $\implies$ robust and stable recovery)

If $\delta_{2s} < \frac{1}{\sqrt{2}}$, then A satisfies the robust NSP of order s with $\rho \in (0, 1)$ and $\tau > 0$ dependent only on δ_{2s}. So, the RIP implies that

$$\|x - \mathcal{R}^\eta(Ax + e)\|_2 \leqslant \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta$$

for some $C, D > 0$ which depend only on δ_{2s}.

Random matrices

Consider random matrices of the form $\frac{1}{\sqrt{m}}A$ where the entries of $A \in \mathbb{R}^{m\times N}$ are iid in accordance to some distribution Λ.

Random Gaussian: $\Lambda = \mathcal{N}(0,1)$. Note that $\mathbb{E}[\frac{1}{m}A^*A] = \mathrm{Id}$.

Random Bernoulli: entries take values $\{+1,-1\}$ with equal probability.

- For random Gaussian/Bernoulli matrices, there exists constants C_1, C_2 which depend only on δ such that for all s satisfying

$$m \geqslant C_1 s \ln(eN/s)$$

with probability at least $1 - \exp(-C_2 m)$, A has RIP constant $\delta_s \leqslant \delta$.

- If A is a random Gaussian/Bernoulli matrix and U is unitary, then the above is also true for AU^*.

Random matrices

Consider random matrices of the form $\frac{1}{\sqrt{m}} A$ where the entries of $A \in \mathbb{R}^{m\times N}$ are iid in accordance to some distribution Λ.

Random Gaussian: $\Lambda = \mathcal{N}(0,1)$. Note that $\mathbb{E}[\frac{1}{m} A^* A] = \mathrm{Id}$.

Random Bernoulli: entries take values $\{+1, -1\}$ with equal probability.

- For random Gaussian/Bernoulli matrices, there exists constants C_1, C_2 which depend only on δ such that for all s satisfying

$$m \geqslant C_1 s \ln(eN/s)$$

  with probability at least $1 - \exp(-C_2 m)$, A has RIP constant $\delta_s \leqslant \delta$.

- If A is a random Gaussian/Bernoulli matrix and U is unitary, then the above is also true for AU^*.

This sampling complexity is optimal, to ensure the stable recovery of all s sparse vectors from linear measurement, we need at least $m \geqslant Cs \ln(eN/s)$

Outline

Setup

Suppose that $V = [v_1|\cdots|v_N] \in \mathbb{C}^{N\times N}$ and $W = [w_1|\cdots|w_N] \in \mathbb{C}^{N\times N}$ are unitary matrices. Let $z \in \mathbb{C}^N$ be the signal of interest.

- Observe $\langle z, w_j\rangle$ for $j \in \Omega$ where $\Omega \subseteq [N]$ is a randomly chosen set of indices.
- z is s-sparse in V, that is, $z = Vx$ where $x \in \Sigma_s$.

Therefore, we want to recover x from

$$y = P_\Omega U x, \quad \text{where} \quad U = W^* V.$$

Coherence

The **coherence** of V and W is $\mu \overset{\text{def.}}{=} \max_{k,\ell} |\langle v_\ell,\, w_k \rangle|$. In the following, let $K \overset{\text{def.}}{=} \sqrt{N}\mu$.

The **coherence** of V and W is $\mu \overset{\text{def.}}{=} \max_{k,\ell} |\langle v_\ell, w_k \rangle|$. In the following, let $K \overset{\text{def.}}{=} \sqrt{N}\mu$.

Clearly, $\mu \leqslant 1$, and since W and V are unitary, we have

$$1 = \|w_k\|^2 = \sum_{\ell=1}^{N} |\langle w_k, v_\ell \rangle|^2 \leqslant N\mu^2$$

so $\mu \geqslant \frac{1}{\sqrt{N}}$. When $\mu = \frac{1}{\sqrt{N}}$, we say that V and W are maximally incoherent. So,

$$\mu \in [\frac{1}{\sqrt{N}}, 1] \quad \text{and} \quad K \in [1, \sqrt{N}]$$

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N} \right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N} \right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.
- The Hadamard transform is maximally incoherence with the canonical basis, where the Hadamard transform is $H \stackrel{\text{def.}}{=} H_n \in \mathbb{R}^{2^n \times 2^n}$ is defined recursively by

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \qquad H_0 = 1.$$

  It can be computed in $\mathcal{O}(N \log(N))$ time and is useful in modelling systems where there are 'on/off' measurements, such as the single-pixel camera, or Fluorescence microscopy.

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N} \right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.
- The Hadamard transform is maximally incoherence with the canonical basis, where the Hadamard transform is $H \stackrel{\text{def.}}{=} H_n \in \mathbb{R}^{2^n \times 2^n}$ is defined recursively by

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \qquad H_0 = 1.$$

  It can be computed in $\mathcal{O}(N \log(N))$ time and is useful in modelling systems where there are 'on/off' measurements, such as the single-pixel camera, or Fluorescence microscopy.

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N} \right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.
- The Hadamard transform is maximally incoherence with the canonical basis, where the Hadamard transform is $H \stackrel{\text{def.}}{=} H_n \in \mathbb{R}^{2^n \times 2^n}$ is defined recursively by

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \qquad H_0 = 1.$$

  It can be computed in $\mathcal{O}(N \log(N))$ time and is useful in modelling systems where there are 'on/off' measurements, such as the single-pixel camera, or Fluorescence microscopy.
- Any basis is maximally coherent with itself, as $\mu = 1$.

Uniform recovery guarantee

If

$$m \stackrel{\text{def.}}{=} |\Omega| \gtrsim K^2 \delta^{-2} s \ln^4(N),$$

then $\sqrt{\frac{N}{m}} U$ satisfies $\delta_s \leqslant \delta$ with probability at least $1 - N^{-\ln^3(N)}$. This guarantees uniform recovery of all s-sparse vectors.

Uniform recovery guarantee

Uniform recovery guarantee

If

$$m \stackrel{\text{def.}}{=} |\Omega| \gtrsim K^2 \delta^{-2} s \ln^4(N),$$

then $\sqrt{\frac{N}{m}} U$ satisfies $\delta_s \leqslant \delta$ with probability at least $1 - N^{-\ln^3(N)}$. This guarantees uniform recovery of all s-sparse vectors.

Note that $\log(N)^4$ is not so small... for $N = 1000$, $\log(N) \approx 6.9$ but $\log^4(N) > 2N$!.

Uniform vs nonuniform guarantees

So far, we have seen that NSP, robust NSP, RIP guarantee recovery of all s-sparse vectors. In particular, we have seen the following uniform recovery guarantee:

$$\mathbb{P}\left(\forall x \in \Sigma_s,\ \Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

Uniform vs nonuniform guarantees

So far, we have seen that NSP, robust NSP, RIP guarantee recovery of all s-sparse vectors. In particular, we have seen the following uniform recovery guarantee:

$$\mathbb{P}\left(\forall x \in \Sigma_s,\ \Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

However, we could ask for a weaker statement: given one vector $x \in \Sigma_s$, show that

$$\mathbb{P}\left(\Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

Uniform vs nonuniform guarantees

So far, we have seen that NSP, robust NSP, RIP guarantee recovery of all s-sparse vectors. In particular, we have seen the following uniform recovery guarantee:

$$\mathbb{P}\left(\forall x \in \Sigma_s,\ \Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

However, we could ask for a weaker statement: given one vector $x \in \Sigma_s$, show that

$$\mathbb{P}\left(\Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

Provided that

$$m \gtrsim K^2 s \ln(N) \ln(\varepsilon^{-1})$$

one can show that any minimizer x^* to $\min_z \|z\|_1$ subject to $\|Az-y\|_2 \leqslant \eta$ where $y = Ax+e$ with $\|e\| \leqslant \eta$ satisfies

$$\|x-x^*\|_2 \lesssim \sigma_s(x)_1 + \sqrt{s}\eta.$$

Outline

Fourier Sampling in Inverse Problems

Many imaging 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 $\mathcal{R}f : \boldsymbol{S}^{d-1} \times \mathbb{R} \to \mathbb{C}$ (where $\boldsymbol{S}^{d-1}$ denotes the sphere)

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

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.

$$g = \mathcal{F}f, \quad f \in L^2(\mathbb{R}^d). \tag{5.1}$$

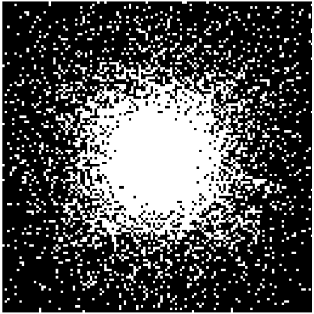
* Magnetic Resonance Imaging (MRI), X-ray Computed Tomography, Electron Microscopy, Radio interferometry, ...

Variable Density Sampling

We saw nonuniform recovery results with $m \gtrsim K^2 s \ln(N) \ln(\varepsilon^{-1})$. This result still does not account for the Fourier-wavelets case as $\mu(U_{dft} U_{dwt}^{-1}) = 1$.

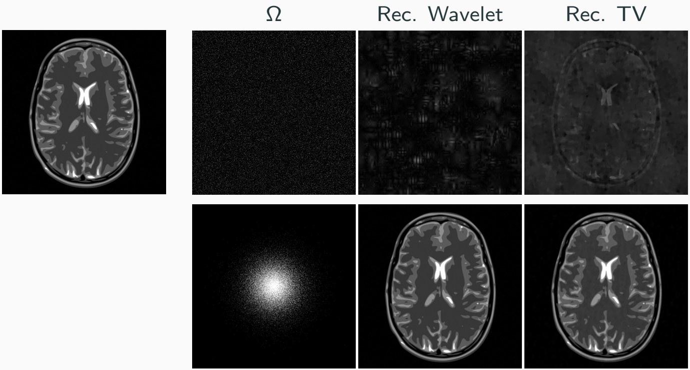
Lustig, Donoho & Pauli '07, Lustig et al. '08: Sample more densely at low Fourier frequencies and less at higher Fourier frequencies.

Ω (30%)



Reconstruction (128×128)

Variable Density Sampling



Why does VDS work?

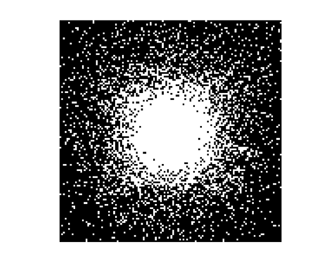
Test phantom constructed by Guerquin-Kern, Lejeune, Pruessmann, Unser, 2012

Sparsity and the Flip Test

In standard CS, the only signal structure considered is sparsity. In contrast, the flip test will demonstrate that we must look beyond sparsity.

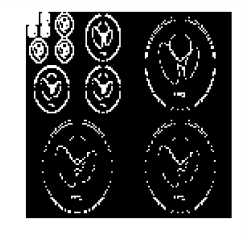
Consider the reconstruction of x from $P_\Omega U_{df} x$ by solving

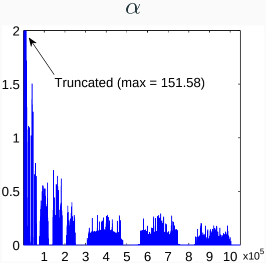
$$\min_z \|z\|_1 \text{ subject to } P_\Omega U_{df} U_{dw}^{-1} z = P_\Omega U_{df} x.$$

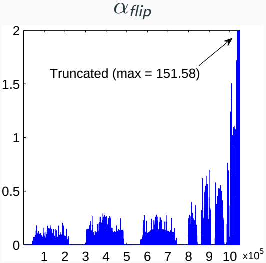


Sparsity and the Flip Test

Let α be the wavelet coefficients of x. Let $\alpha^{flip} = (\alpha_N, \dots, \alpha_1)$ and $x^{flip} = U_{dw}^{-1}\alpha^{flip}$.





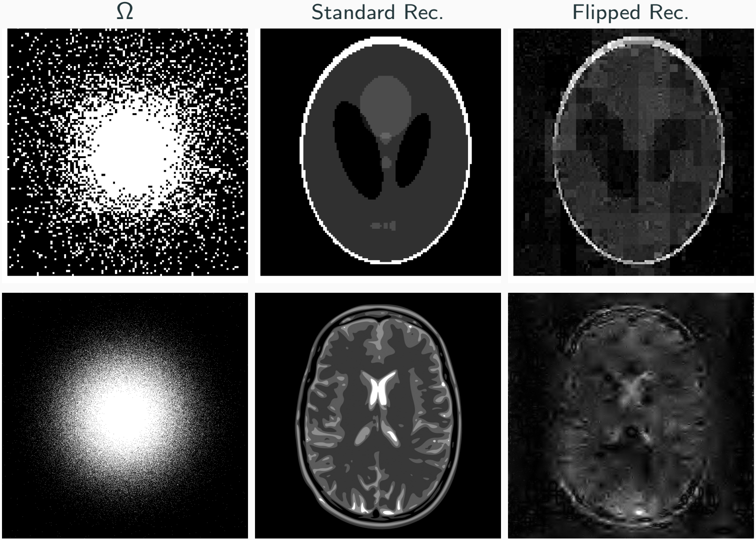


For the same Ω, let

$$\tilde{\alpha} \in \arg\min_{\beta} \|\beta\|_1 \text{ subject to } P_\Omega U_{df} U_{dw}^{-1}\beta = P_\Omega U_{df} x^{flip}$$

If sparsity was enough, then the wavelet coefficients $\tilde{\alpha}$ should be α^{flip}, we should be able to recover x as $U_{dw}^{-1}\tilde{\alpha}^{flip}$.

Sparsity and the Flip Test



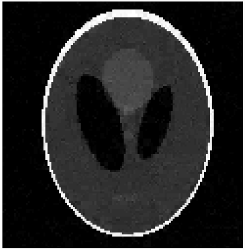
Flip test for random Gaussian measurements

In the Fourier-wavelets case, we saw that Ω cannot depend on sparsity alone.

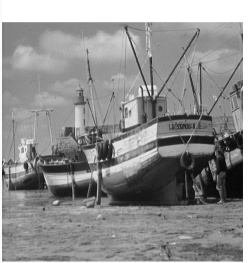
Actually, variable density sampling patterns exploit the fact that natural images are asymptotically sparse in wavelets.

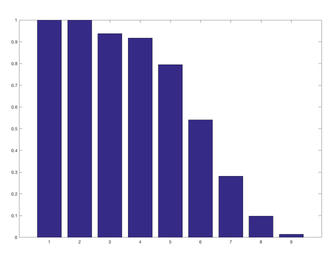
On the other hand, random Gaussian measurements are insensitive to sparsity structure:

Standard Rec.



Flipped Rec.





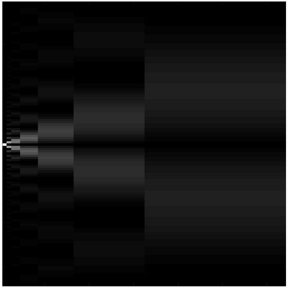
Variable density sampling patterns only give good reconstructions of asymptotically sparse wavelet coefficient sequences, and not sequences which are arbitrarily sparse.

Asymptotic Incoherence

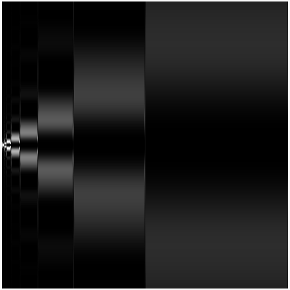
If $U = U_{df} U_{dw}^{-1}$ is the Fourier-wavelets matrix, then

$$\mu(P_K^{\perp} U), \mu(U P_K^{\perp}) = \mathcal{O}\left(K^{-1/2}\right).$$

Fourier to Haar



Fourier to DB4



Implication: Sample more at low Fourier frequencies where the local coherence is high and less at higher Fourier frequencies.

Recovery of Wavelet Coefficients from Partial Fourier Data

Theorem:

- $\{N_k\}_{k=1}^r$ and $\{M_k\}_{k=1}^r$ correspond to wavelet scales.
- The mother wavelet ψ has v vanishing moments.
- There exists $\alpha \geqslant 1,\ C > 0$ such that $\left|\hat{\psi}(\xi)\right| \leqslant \frac{C}{(1+|\xi|)^\alpha}$ for all $\xi \in \mathbb{R}$.

One is guaranteed recovery of a wavelet sequence which is s_k-sparse in the k^{th} wavelet scale by choosing $\Omega = \Omega_1 \cup \cdots \cup \Omega_r$ where $\Omega_k \subset \{N_{k-1}+1, \ldots, N_k\}$ are m_k samples chosen uniformly at random, with

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

where $\hat{s}_k = \max\{s_{k-1}, s_k, s_{k+1}\}$ and $\mathcal{L} = \log(\varepsilon^{-1}) \cdot \log\left(KN\sqrt{s}\right)$

Breaking the coherence barrier: A new theory for compressed sensing. Adcock, Hansen, Poon & Roman (2013)

Resolution Dependence (5% samples, varying resolution)

Asymptotic sparsity and asymptotic incoherence are only witnessed when N is large. Thus, V. D. sampling only reaps their benefits for large values of N and the success of compressed sensing is resolution dependent.

256x256

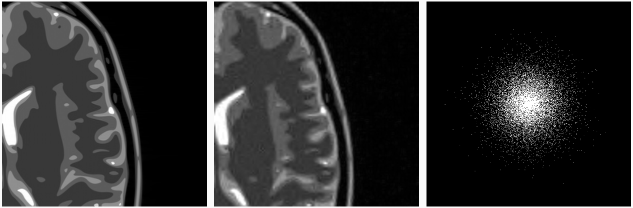
Error:

19.86%

512x512

Error:

10.69%

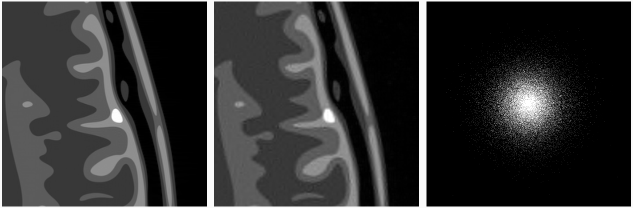


Resolution Dependence (5% samples, varying resolution)

1024x1024

Error:

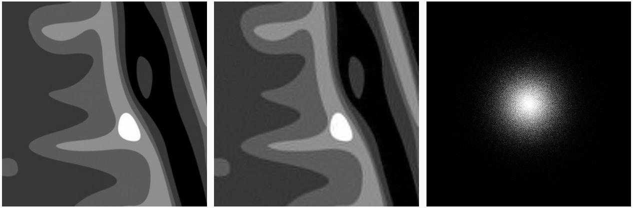
7.35%



2048x2048

Error:

4.87%



4096x4096

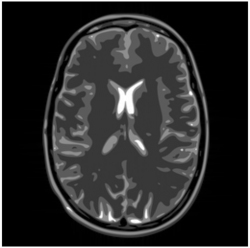
Error:

3.06%

Recovering Fine Details

At finer wavelet scales, the presence of sparsity and incoherence with Fourier samples allows us to subsample. Thus, compressed sensing allows one to enhance fine details without increasing the number of samples.

In the next example, consider the reconstruction of a 2048×2048 test phantom with details added at the finest wavelet scale.



Recovering Fine Details

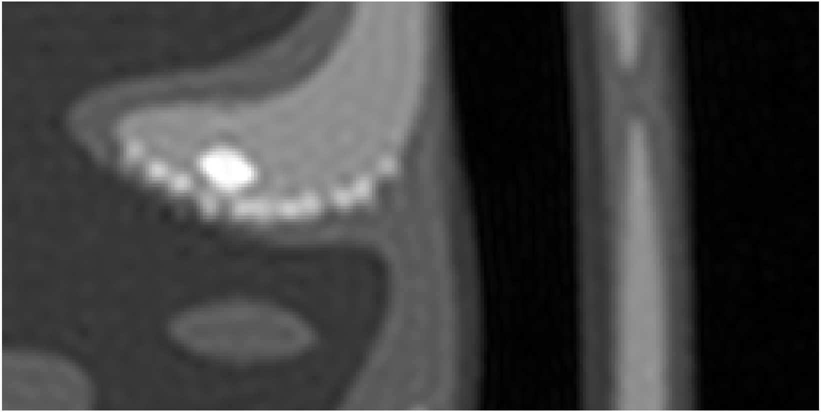


Figure 1: 2048×2048 linear reconstruction from the first 512×512 Fourier samples (6.25%)

Figure 2: 2048×2048 reconstruction from a multilevel scheme using 512×512 Fourier samples (6.25%)

Sources

- "A Mathematical Introduction to Compressive Sensing" by Simon Foucart & Holger Rauhut.
- "Flavors of Compressive Sensing" by Simon Foucart.