On the role of total variation in compressed sensing

Exact reconstructions from highly incomplete Fourier data

Clarice Poon

University of Cambridge

June 22nd, 2015

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC).

Sampling the Fourier transform

In many applications, we are required to recover some function $f \in L^2(\Omega)$, $\Omega \subset \mathbb{R}^d$, from pointwise evaluations of its Fourier transform:

$$\mathcal{F}f(\omega) = \int_{x \in \Omega} f(x) e^{-2\pi i \langle \omega, x \rangle} \mathrm{d}x.$$

- Medical imaging: Magnetic Resonance Imaging (MRI)/ Computed Tomography
- Astronomy: Radio Interferometry
- Biology: Electron/Fluorescence Microscopy

The Shannon Nyquist Sampling Theorem

Whittaker 1929, Kotelnikov 1933, Shannon 1949

If f has support included in $[-T, T]$, then for $\epsilon^{-1} \geq 2T$,

$$f = \epsilon \sum_{n \in \mathbb{Z}^2} \mathcal{F}f(\epsilon n) e^{2\pi i \epsilon n \cdot}, \qquad \text{with } L^2 \text{ convergence,}$$

$$\mathcal{F}f = \sum_{n \in \mathbb{Z}^2} \mathcal{F}f(\epsilon n) \operatorname{sinc}\left(\frac{\cdot + n\epsilon}{\epsilon}\right), \qquad \text{with } L^\infty, L^2 \text{ convergence.}$$

The Shannon Nyquist Sampling Theorem

Whittaker 1929, Kotelnikov 1933, Shannon 1949

If f has support included in $[-T, T]$, then for $\epsilon^{-1} \geq 2T$,

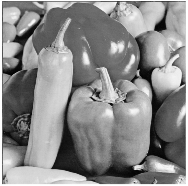
$$f = \epsilon \sum_{n \in \mathbb{Z}^2} \mathcal{F}f(\epsilon n) e^{2\pi i \epsilon n \cdot}, \qquad \text{with } L^2 \text{ convergence,}$$

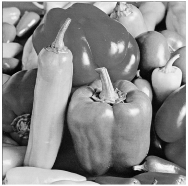
$$\mathcal{F}f = \sum_{n \in \mathbb{Z}^2} \mathcal{F}f(\epsilon n) \operatorname{sinc}\left(\frac{\cdot + n\epsilon}{\epsilon}\right), \qquad \text{with } L^\infty, L^2 \text{ convergence.}$$

Under no further assumptions, the sampling rate ϵ^{-1} must be at least $2T$, the Nyquist rate.

Sparsity

In the last few decades, sparsity has played a prominent role in image processing.

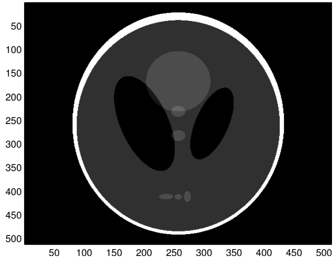
Wavelet:  $\overset{W}{\Longrightarrow}$ 

Gradient:  $\overset{D}{\Longrightarrow}$ 

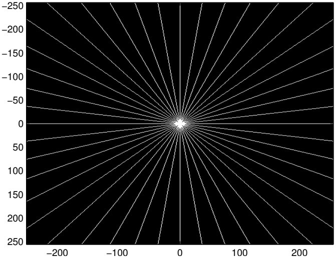
Total variation (1992), wavelets (1988), contourlets (2005), curvelets (2000), shearlets (2006), ...

An intriguing experiment

Candès, Romberg & Tao, 2006



x



Ω (4.24%)

$$\min_{z \in \mathbb{C}^{N \times N}} \|z\|_{TV} \text{ subject to } P_\Omega U z = P_\Omega U x$$

- $Uz = \left(\sum_{j_1=1}^{N} \sum_{j_2=1}^{N} z_{j_1,j_2} e^{i2\pi(k_1 j_1 + k_2 j_2)} \right)_{k_1,k_2 = -\lfloor N/2 \rfloor, \ldots, \lceil N/2 \rceil - 1}.$
- $Dz = D_1 z + i D_2 z$ where

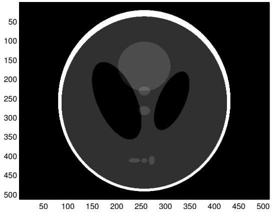
$$D_1 z = (x_{k+1,j} - z_{k,j})_{k,j=1}^{N}, \quad D_2 z = (z_{k,j+1} - z_{k,j})_{k,j=1}^{N}$$

with $z_{N+1,j} := z_{1,j}$ and $z_{k,N+1} := z_{k,1}$. Let $\|z\|_{TV} := \|Dz\|_1$.

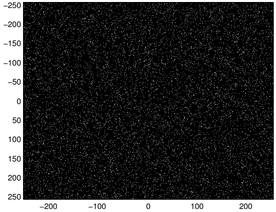
A theoretical explanation towards sub-Nyquist sampling

Candès, Romberg & Tao, 2006: Let $x \in \mathbb{C}^N$ be s-sparse in its discrete gradient and suppose we observe its 0^{th} Fourier coefficient plus $\mathcal{O}(s \log N)$ of its Fourier coefficients chosen uniformly at random. Then, with overwhelming probability, x is the unique solution to

$$\min_{z \in \mathbb{C}^N} \|z\|_{TV} \text{ subject to } P_\Omega U z = P_\Omega U x.$$



x

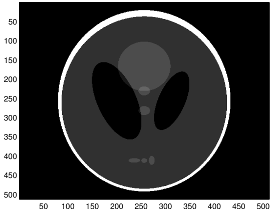


Ω (4.24%)

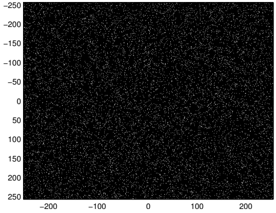
A theoretical explanation towards sub-Nyquist sampling

Candès, Romberg & Tao, 2006: Let $x \in \mathbb{C}^N$ be s-sparse in its discrete gradient and suppose we observe its 0^{th} Fourier coefficient plus $\mathcal{O}(s \log N)$ of its Fourier coefficients chosen uniformly at random. Then, with overwhelming probability, x is the unique solution to

$$\min_{z \in \mathbb{C}^N} \|z\|_{TV} \text{ subject to } P_\Omega U z = P_\Omega U x.$$



x



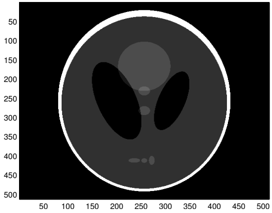
Ω (4.24%)

- $\mathcal{O}(s \log N)$ represents a substantial saving in the number of samples.

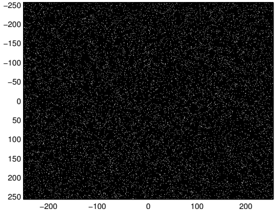
A theoretical explanation towards sub-Nyquist sampling

Candès, Romberg & Tao, 2006: Let $x \in \mathbb{C}^N$ be s-sparse in its discrete gradient and suppose we observe its 0^{th} Fourier coefficient plus $\mathcal{O}(s \log N)$ of its Fourier coefficients chosen uniformly at random. Then, with overwhelming probability, x is the unique solution to

$$\min_{z \in \mathbb{C}^N} \|z\|_{TV} \text{ subject to } P_\Omega U z = P_\Omega U x.$$



x



Ω (4.24%)

- $\mathcal{O}(s \log N)$ represents a substantial saving in the number of samples.
- The sampling cardinality of $\mathcal{O}(s \log N)$ is optimal for s-sparse vectors.

Compressed sensing*

How can we recover an s-sparse vector $x \in \mathbb{C}^N$ from Ux
where $U \in \mathbb{C}^{m \times N}$ and $m = \mathcal{O}(s \log N) \ll N$?

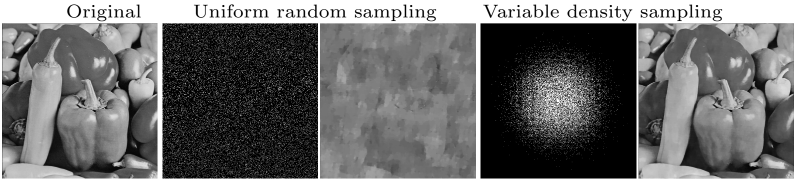
- Thousands of papers developing algorithms and random sensing matrices.
- But many applications where compressed sensing is of practical interest are constrained to Fourier sampling, and one of the most widely used sparsifying transform is the gradient operator.
    - Lustig et al. (2007) on MRI, Wiaux et al. (2009) on radio interferometry, Leary et al. (2013) on electron microscopy, ...

*Introduced in 2006 independently by Donoho and Candès, Romberg & Tao.

This talk:

In practice, signals are only approximately sparse and measurements are noisy. Given $y = P_\Omega U x + \eta \in \mathbb{C}^m$ with $\|\eta\|_2 \leq \delta\sqrt{m}$, we seek to solve

$$\min_z \|z\|_{TV} \text{ subject to } \|P_\Omega U z - y\|_2 \leq \delta\sqrt{m}.$$



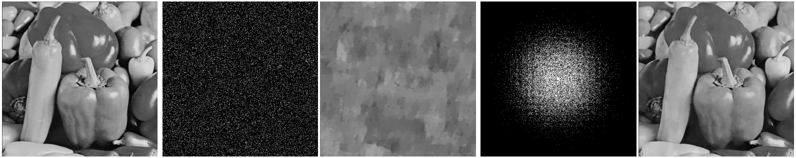
Reconstructions from sampling 10% of the Fourier coefficients.

This talk: 2 key questions

In practice, signals are only approximately sparse and measurements are noisy. Given $y = P_\Omega U x + \eta \in \mathbb{C}^m$ with $\|\eta\|_2 \leq \delta\sqrt{m}$, we seek to solve

$$\min_z \|z\|_{TV} \text{ subject to } \|P_\Omega U z - y\|_2 \leq \delta\sqrt{m}.$$

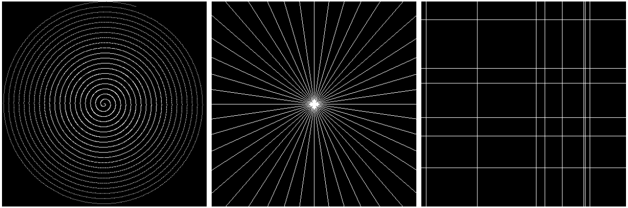
Original | Uniform random sampling | Variable density sampling



Reconstructions from sampling 10% of the Fourier coefficients.

1. What can we say about robustness to noise and stability to inexact sparsity under uniform random sampling at $\mathcal{O}(s \log N)$?
2. Why does variable density sampling outperform uniform random sampling?

Remark: Feasible sampling patterns



Spirals Radial lines Cartesian lines

In practical applications such as MRI, the hardware constraints mean that uniform random sampling cannot be implemented. Thus, there is a need to understand how one should sample along trajectories.

Last part of this talk: Sampling along Cartesian lines.

Notation

1D case: For $x \in \mathbb{C}^N$, $\Lambda \subset \mathbb{Z}$,

- $Ux = \left(\sum_{j=1}^N x_j e^{i2\pi kj} \right)_{k=-\lfloor N/2 \rfloor, \dots, \lceil N/2 \rceil - 1}$.
- $P_\Lambda : \mathbb{C}^N \to \mathbb{C}^N$, $(P_\Lambda x)_j = x_j$ if $j \in \Lambda$ and 0 otherwise.
- $Dx := (-x_j + x_{j+1})_{j=1}^N$ with $x_{N+1} := x_1$. Let $\|x\|_{TV} = \|Dx\|_1$.

2D case: For $x \in \mathbb{C}^{N \times N}$, $\Lambda \subset \mathbb{Z}^2$,

- $Ux = \left(\sum_{j_1=1}^N \sum_{j_2=1}^N x_{j_1,j_2} e^{i2\pi(k_1 j_1 + k_2 j_2)} \right)_{k_1,k_2=-\lfloor N/2 \rfloor, \dots, \lceil N/2 \rceil - 1}$
- $P_\Lambda : \mathbb{C}^{N \times N} \to \mathbb{C}^{N \times N}$, $(P_\Lambda x)_{k,j} = x_{k,j}$ if $(k,j) \in \Lambda$ and 0 otherwise.
- $Dx = D_1 x + i D_2 x$ where

$$D_1 x = (x_{k+1,j} - x_{k,j})_{k,j=1}^N, \quad D_2 x = (x_{k,j+1} - x_{k,j})_{k,j=1}^N$$

  with $x_{N+1,j} := x_{1,j}$ and $x_{k,N+1} := x_{k,1}$. Let $\|x\|_{TV} := \|Dx\|_1$.

Notation

1D case: For $x \in \mathbb{C}^N$, $\Lambda \subset \mathbb{Z}$,

- $Ux = \left(\sum_{j=1}^{N} x_j e^{i2\pi kj}\right)_{k=-\lfloor N/2\rfloor,\dots,\lceil N/2\rceil-1}$.
- $P_\Lambda : \mathbb{C}^N \to \mathbb{C}^N$, $(P_\Lambda x)_j = x_j$ if $j \in \Lambda$ and 0 otherwise.
- $Dx := (-x_j + x_{j+1})_{j=1}^{N}$ with $x_{N+1} := x_1$. Let $\|x\|_{TV} = \|Dx\|_1$.

2D case: For $x \in \mathbb{C}^{N\times N}$, $\Lambda \subset \mathbb{Z}^2$,

- $Ux = \left(\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} x_{j_1,j_2} e^{i2\pi(k_1j_1+k_2j_2)}\right)_{k_1,k_2=-\lfloor N/2\rfloor,\dots,\lceil N/2\rceil-1}$
- $P_\Lambda : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N}$, $(P_\Lambda x)_{k,j} = x_{k,j}$ if $(k,j) \in \Lambda$ and 0 otherwise.
- $Dx = D_1x + iD_2x$ where

$$D_1x = (x_{k+1,j} - x_{k,j})_{k,j=1}^{N}, \quad D_2x = (x_{k,j+1} - x_{k,j})_{k,j=1}^{N}$$

with $x_{N+1,j} := x_{1,j}$ and $x_{k,N+1} := x_{k,1}$. Let $\|x\|_{TV} := \|Dx\|_1$.

Given y and $\Omega \subset \mathbb{Z}$ (resp. $\mathbb{Z}^2$) of cardinality m, we will consider the solutions

$$\mathcal{R}(\Omega,\delta,y) = \operatorname*{argmin}_{z\in\mathbb{C}^N(\text{resp.}\mathbb{C}^{N\times N})} \|z\|_{TV} \text{ subject to } \|P_\Omega U - y\|_2 \le \sqrt{m}\cdot\delta.$$

Outline

Uniform + power law sampling

Uniform random sampling

Beyond sparsity – low frequency sampling

Sampling along Cartesian lines

Outline

Uniform + power law sampling

Uniform random sampling

Beyond sparsity – low frequency sampling

Sampling along Cartesian lines

Uniform + power law sampling (1D case)

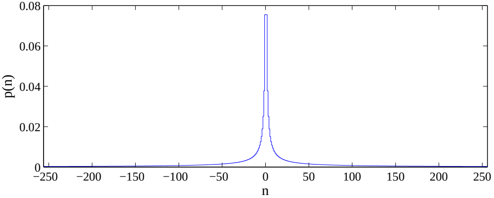
Definition

$\Omega = \Omega_1 \cup \Omega_2 \subset \{-\lfloor N/2 \rfloor + 1, \dots, \lceil N/2 \rceil\}$ is a uniform + power law sampling scheme of cardinality $2m$ if

- Ω_1 consists of m indices chosen uniformly at random.
- $\Omega_2 = \{k_1, \dots, k_m\}$ consist of m indices which are independent and identically distributed (i.i.d.) such that for each $j = 1, \dots, m$ and $n = -N/2 + 1, \dots, N/2$,

$$\mathbb{P}(k_j = n) = p(n), \quad p(n) = \frac{C \log(N)}{\max\{1, |n|\}},$$

where C is an appropriate constant such that p is a probability measure.



Plot of $p(n)$ for $N = 512$.

Uniform + power law sampling (1D case)

Theorem (P. 2015)

Let $N = 2^J$ with $J \in \mathbb{N}$, $\delta \geq 0$ and $\epsilon \in (0,1)$.

- Let $x \in \mathbb{C}^N$ and let Δ index the largest s coefficients of Dx.
- Suppose we are given $y = P_\Omega U x + \eta$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$ and Ω is a uniform + power law sampling scheme of cardinality $m = \mathcal{O}\left(s \log(N)(1 + \log(\epsilon^{-1})\right)$.

Then with probability exceeding $1 - \epsilon$, $\xi \in \mathcal{R}(\Omega, \delta, y)$ satisfies

$$\|Dx - D\xi\|_2 \lesssim \left(\delta\sqrt{s} + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}\right),$$

$$\frac{\|x - \xi\|_2}{\sqrt{N}} \lesssim \mathcal{L}_1 \cdot \left(\frac{\delta}{\sqrt{s}} + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{s}\right),$$

where $\mathcal{L}_1 = \log^2(s)\log(N)\log(m)$ and $\mathcal{L}_2 = \log(s)\log^{\frac{1}{2}}(m)$.

Uniform + power law sampling (1D case)

Theorem (P. 2015)

Let $N = 2^J$ with $J \in \mathbb{N}$, $\delta \geq 0$ and $\epsilon \in (0,1)$.

- Let $x \in \mathbb{C}^N$ and let Δ index the largest s coefficients of Dx.
- Suppose we are given $y = P_\Omega U x + \eta$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$ and Ω is a uniform + power law sampling scheme of cardinality $m = \mathcal{O}\left(s \log(N)(1 + \log(\epsilon^{-1})\right)$.

Then with probability exceeding $1 - \epsilon$, $\xi \in \mathcal{R}(\Omega, \delta, y)$ satisfies

$$\|Dx - D\xi\|_2 \lesssim \left(\delta\sqrt{s} + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}\right),$$

$$\frac{\|x - \xi\|_2}{\sqrt{N}} \lesssim \mathcal{L}_1 \cdot \left(\frac{\delta}{\sqrt{s}} + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{s}\right),$$

where $\mathcal{L}_1 = \log^2(s)\log(N)\log(m)$ and $\mathcal{L}_2 = \log(s)\log^{\frac{1}{2}}(m)$.

DeVore (1998): *The optimal error decay rate for any bounded variation function $f \in BV[0,1)$ by any type of nonlinear approximation $\tilde{f}$ from s samples is*

$$\|\tilde{f} - f\|_{L^2[0,1)} = \mathcal{O}\left(\|f\|_V \cdot s^{-1}\right).$$

Uniform + power law sampling (2D case)

Definition

$\Omega = \Omega_1 \cup \Omega_2 \subset \{-\lfloor N/2 \rfloor + 1, \dots, \lceil N/2 \rceil\}^2$ is a uniform + power law sampling scheme of cardinality $2m$ if

- Ω_1 consists of m indices chosen uniformly at random.
- $\Omega_2 = \{k_1, \dots, k_m\}$ consists of m i.i.d. indices such that for each $j = 1, \dots, m$, and $n_1, n_2 = -N/2 + 1, \dots, N/2$,

$$\mathbb{P}(k_j = (n_1, n_2)) = p(n_1, n_2),$$

$$p(n_1, n_2) = \frac{C \log(N)}{\max\{1, |n_1|^2 + |n_2|^2\}},$$

where $C > 0$ is such that p is a probability measure.

Uniform + power law sampling (2D case)

Theorem (P. 2015)

Let $N = 2^J$, $J \in \mathbb{N}$, $\epsilon \in (0,1)$ and $\delta \geq 0$.

- Let $x \in \mathbb{C}^{N \times N}$ and let Δ index the largest s coefficients of Dx.
- Suppose we are given $y = P_\Omega U x + \eta$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$ and Ω is a uniform + power law sampling scheme of cardinality

$$m = \mathcal{O}\left(s \log(N)(1 + \log(\epsilon^{-1}))\right).$$

Then, with probability exceeding $1 - \epsilon$, $\xi \in \mathcal{R}(\Omega, \delta, y)$ satisfies

$$\|Dx - D\xi\|_2 \lesssim \left(\delta \cdot \sqrt{s} + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}\right),$$

$$\|x - \xi\|_2 \lesssim \mathcal{L}_1 \cdot \left(\delta + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}\right),$$

where $\mathcal{L}_1 = \log(s)\log(\frac{N^2}{s})\log^{\frac{1}{2}}(N)\log^{\frac{1}{2}}(m)$, and $\mathcal{L}_2 = \log^{\frac{1}{2}}(m)\log(s)$.

Uniform + power law sampling (2D case)

Theorem (P. 2015)

Let $N = 2^J$, $J \in \mathbb{N}$, $\epsilon \in (0,1)$ and $\delta \geq 0$.

- Let $x \in \mathbb{C}^{N \times N}$ and let Δ index the largest s coefficients of Dx.
- Suppose we are given $y = P_\Omega U x + \eta$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$ and Ω is a uniform + power law sampling scheme of cardinality

$$m = \mathcal{O}\left(s \log(N)(1 + \log(\epsilon^{-1}))\right).$$

Then, with probability exceeding $1 - \epsilon$, $\xi \in \mathcal{R}(\Omega, \delta, y)$ satisfies

$$\|Dx - D\xi\|_2 \lesssim \left(\delta \cdot \sqrt{s} + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}\right),$$

$$\|x - \xi\|_2 \lesssim \mathcal{L}_1 \cdot \left(\delta + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}\right),$$

where $\mathcal{L}_1 = \log(s)\log(\frac{N^2}{s})\log^{\frac{1}{2}}(N)\log^{\frac{1}{2}}(m)$, and $\mathcal{L}_2 = \log^{\frac{1}{2}}(m)\log(s)$.

Candès & Tao (2006), Needell & Ward (2013):
The optimal error estimate from $\mathcal{O}\left(s\log(N^2/s)\right)$ *nonadaptive samples is*

$$\delta + \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}.$$

Uniform + power law sampling (2D case)

Theorem (P. 2015)

Let $N = 2^J$, $J \in \mathbb{N}$, $\epsilon \in (0,1)$ and $\delta \geq 0$.

- Let $x \in \mathbb{C}^{N \times N}$ and let Δ index the largest s coefficients of Dx.
- Suppose we are given $y = P_\Omega U x + \eta$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$ and Ω is a uniform + power law sampling scheme of cardinality

$$m = \mathcal{O}\left(s \log(N)(1 + \log(\epsilon^{-1})\right).$$

Then, with probability exceeding $1 - \epsilon$, $\xi \in \mathcal{R}(\Omega, \delta, y)$ satisfies

$$\|Dx - D\xi\|_2 \lesssim \left(\delta \cdot \sqrt{s} + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}\right),$$

$$\|x - \xi\|_2 \lesssim \mathcal{L}_1 \cdot \left(\delta + \mathcal{L}_2 \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}}\right),$$

where $\mathcal{L}_1 = \log(s)\log(\frac{N^2}{s})\log^{\frac{1}{2}}(N)\log^{\frac{1}{2}}(m)$, and $\mathcal{L}_2 = \log^{\frac{1}{2}}(m)\log(s)$.

Krahmer & Ward (2014): *Given* $\mathcal{O}\left(s \log^5(N) \log^3(s)\right)$ *Fourier coefficients distributed by a power law, TV regularization guarantees stable recovery up to gradient sparsity s.*

Outline

Uniform + power law sampling

Uniform random sampling

Beyond sparsity – low frequency sampling

Sampling along Cartesian lines

Uniform random sampling (1D case)

Theorem (P. 2015)

Let $N \in \mathbb{N}$, $\delta \geq 0$ and $\epsilon \in (0,1)$.

- Let $x \in \mathbb{C}^N$ and let Δ index the largest s coefficients of Dx.
- Suppose we are given $y = P_\Omega U x + \eta \in \mathbb{C}^m$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$ and Ω includes 0 and m indices chosen uniformly at random with

$$m = \mathcal{O}\left(s \cdot \log(N) \cdot \left(1 + \log(\epsilon^{-1})\right)\right).$$

Then, with probability exceeding $1-\epsilon$, $\xi \in \mathcal{R}(\Omega, \delta, y)$ satisfies

$$\|Dx - D\xi\|_2 \lesssim \delta \cdot \sqrt{s} + \mathcal{L} \cdot \frac{\|P_{\Delta^c} Dx\|_1}{\sqrt{s}},$$

$$\frac{\|x - \xi\|_2}{\sqrt{N}} \lesssim \delta \cdot \sqrt{s} + \mathcal{L} \cdot \|P_{\Delta^c} Dx\|_1,$$

where $\mathcal{L} = \log^{\frac{1}{2}}(m) \log(s)$.

Uniform random sampling (2D case)

Theorem (P. 2015)

Let $N \in \mathbb{N}$, $\delta \geq 0$ and $\epsilon \in (0,1)$.

- Let $x \in \mathbb{C}^{N \times N}$ and let Δ index the largest s coefficients of Dx.
- Suppose we are given $y = P_\Omega U x + \eta \in \mathbb{C}^m$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$ and Ω includes 0 and m indices chosen uniformly at random with

$$m = \mathcal{O}\left(s \cdot \log\left(N\right) \cdot \left(1 + \log(\epsilon^{-1})\right)\right).$$

Then, with probability exceeding $1 - \epsilon$, $\xi \in \mathcal{R}(\Omega, \delta, y)$ satisfies

$$\|Dx - D\xi\|_2 \lesssim \delta \cdot \sqrt{s} + \mathcal{L} \cdot \frac{\|P_{\Delta^c} D x\|_1}{\sqrt{s}},$$

$$\|x - \xi\|_2 \lesssim \delta \cdot \sqrt{s} + \mathcal{L} \cdot \|P_{\Delta^c} D x\|_1,$$

where $\mathcal{L} = \log^{\frac{1}{2}}(m) \log(s)$.

Question...

So, uniform random sampling does achieve robustness and stability...

Example: Reconstructions from 35% uniform random sampling:

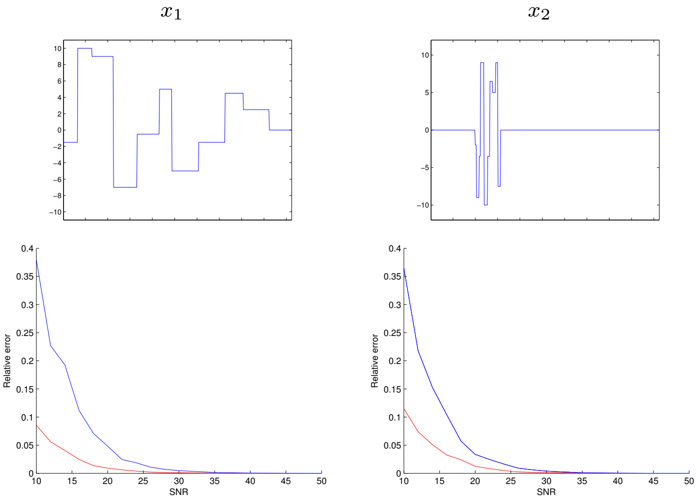
Original	No noise	SNR = 10	SNR = 5
	Err = 20%	Err = 22%	Err = 46%.

However, the error bounds obtained for the uniform random sampling strategy are sub-optimal, whereas, by adding the samples which concentrate on low frequencies, one can guarantee near-optimal error bounds.

Does dense sampling at low frequencies actually improve stability, or is the difference between the theorems simply an artefact of the proofs?

A numerical comparison

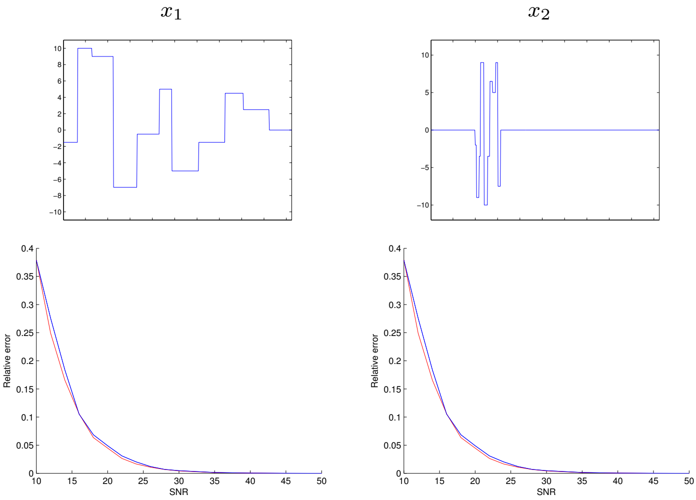
Consider the recovery of $x + h$ from 10% of its Fourier coefficients, with different $SNR = 10 \log_{10}(\|x\|_2 \,/\, \|h\|_2)$.



Relative errors of the recovered signals using Ω_P vs using Ω_U.

A numerical comparison

Consider the recovery of $x + h$ from 10% of its Fourier coefficients, with different $SNR = 10\log_{10}(\|x\|_2 \,/\, \|h\|_2)$.



Relative errors of the recovered gradients using Ω_P vs using Ω_U.

Outline

Uniform + power law sampling

Uniform random sampling

Beyond sparsity – low frequency sampling

Sampling along Cartesian lines

Detour: super resolution (discrete case)

The recovery of a super position of spikes, $x = \sum_j \alpha_j \delta_{t_j}$ *with* $t_j \in [0,1]$*, from low frequency samples only.*

Candès & Fernandez-Granda (2012):

Let $x \in \mathbb{C}^N$*, let* Δ *be its support and suppose that*

$$\min_{t,t' \in \Delta, t \neq t'} \frac{|t - t'|}{N} \geq \frac{2}{M}.$$

Then, given $y = P_{[M]} U x$ *(first* $2M+1$ *DFT coefficients of* x*) with* $[M] = \{-M, \ldots, M\}$*,* x *is the unique solution of*

$$\min_z \|z\|_1 \text{ subject to } P_{[M]} U z = y.$$

Detour: super resolution (discrete case)

The recovery of a super position of spikes, $x = \sum_j \alpha_j \delta_{t_j}$ *with* $t_j \in [0,1]$*, from low frequency samples only.*

Candès & Fernandez-Granda (2012):
Let $x \in \mathbb{C}^N$*, let* Δ *be its support and suppose that*

$$\min_{t,t' \in \Delta, t \neq t'} \frac{|t - t'|}{N} \geq \frac{2}{M}.$$

Then, given $y = P_{[M]} U x$ *(first* $2M+1$ *DFT coefficients of* x*) with* $[M] = \{-M, \ldots, M\}$*,* x *is the unique solution of*

$$\min_z \|z\|_1 \text{ subject to } P_{[M]} U z = y.$$

If instead, Dx has support Δ:

1. recover Dx by solving an ℓ^1 problem.
2. recover x from Dx by shifting by the mean of x (i.e. 0^{th} Fourier coefficient of x).

Detour: super resolution (discrete case)

The recovery of a super position of spikes, $x = \sum_j \alpha_j \delta_{t_j}$ *with* $t_j \in [0,1]$*, from low frequency samples only.*

Candès & Fernandez-Granda (2012):
Let $x \in \mathbb{C}^N$*, let* Δ *be its support and suppose that*

$$\min_{t,t' \in \Delta, t \neq t'} \frac{|t-t'|}{N} \geq \frac{2}{M}.$$

Then, given $y = P_{[M]}Ux$ *(first* $2M+1$ *DFT coefficients of* x*) with* $[M] = \{-M, \ldots, M\}$*,* x *is the unique solution of*

$$\min_z \|z\|_1 \text{ subject to } P_{[M]}Uz = y.$$

If instead, Dx has support Δ:

1. recover Dx by solving an ℓ^1 problem.
2. recover x from Dx by shifting by the mean of x (i.e. 0^{th} Fourier coefficient of x).

Tang & Bhaskar & Shah & Recht (2013): An analogous compressed sensing result.

Low frequency sampling

Theorem (P. 2015)

Let $N \in \mathbb{N}$, let $\epsilon \in [0,1]$ and let $M \in \mathbb{N}$ be such that $N/4 \geq M \geq 10$.

- Let $x \in \mathbb{C}^N$ and $\Delta \subset \{1,\ldots,N\}$ be of cardinality s and suppose that

$$\min_{k,j \in \Delta, k \neq j} \frac{|k-j|}{N} \geq \frac{2}{M}.$$

- Let $\Omega \subset \{-M,\ldots,M\}$ include 0 and m indices chosen uniformly at random with

$$m \gtrsim \max\left\{\log^2\left(\frac{M}{\epsilon}\right), \quad \log(N), \quad s \cdot \log\left(\frac{s}{\epsilon}\right) \cdot \log\left(\frac{M}{\epsilon}\right)\right\}.$$

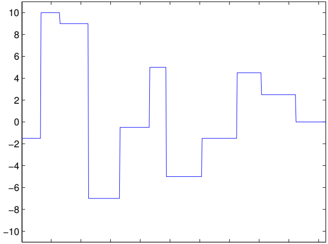
Then with probability exceeding $1-\epsilon$, given $y = P_\Omega A x + \eta$ and $\|\eta\|_2 \leq \delta \cdot \sqrt{m}$, any solution $\xi \in \mathcal{R}(\Omega, \delta, y)$ satisfies

$$\frac{\|x-\xi\|_2}{\sqrt{N}} \lesssim \frac{N^2}{M^2} \cdot \left(\delta \cdot s + \sqrt{s} \cdot \|P_{\Delta^c} D x\|_1\right).$$

If $m = 2M+1$, then the error bound holds with probability 1.

The price of randomness

Suppose that $x \in \mathbb{C}^N$ was s-gradient sparse with a minimum separation of $2/s$. Then, x can be exactly recovered from $2s+1$ Fourier coefficients. However, random sampling guarantees recovery only with $\mathcal{O}\left(s \log N\right)$ samples.



This signal can be recovered exactly from 3.9% of its Fourier coefficient of lowest frequencies, but uniform random sampling would require sampling at 10%.

What about the 2D case?

Candès & Fernandez-Granda (2012):

Let $x \in \mathbb{C}^{N \times N}$ have support Δ. If $\min_{\mathbf{k},\mathbf{j} \in \Delta, \mathbf{k} \neq \mathbf{j}} |\mathbf{k} - \mathbf{j}| \geq 2.38/M$ and we observe the Fourier coefficients of x up to frequency $M \in \mathbb{N}$, $y = P_{[M]} U x$. Then x is the unique solution of

$$\min_z \|z\|_1 \text{ subject to } P_{[M]} U z = y.$$

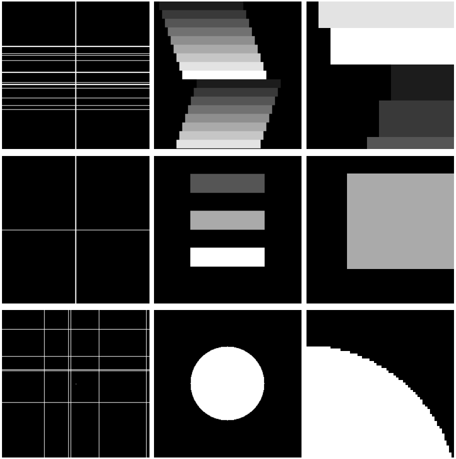
The difficulty with the recovery of images is that there is no separation in the edge set.

Outline

Sampling along Cartesian lines

In 1D, the *number* of samples depends on the *gradient sparsity*, and the *range* that we sample from depends on the *separation* of the support set. In 2D:

- The number of lines depends on the gradient sparsity along each direction.
- The sampling range depends on the separation along each direction.



Separation concepts

Definition

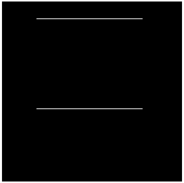
Let $N \in \mathbb{N}$ and let $\Delta \subset \{1, \ldots, N\}^2$. The minimum separation distance of

- its rows is defined to be

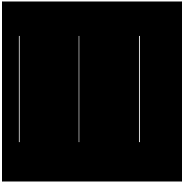
$$\nu_{\mathrm{row}}(\Delta, N) = \min_{n=1}^{N} \min \left\{ \frac{|j-k|}{N} : j, k \in \Delta \cap \{\{n\} \times \{1, \ldots, N\}\} \right\},$$

- its columns is defined to be

$$\nu_{\mathrm{col}}(\Delta, N) = \min_{n=1}^{N} \min \left\{ \frac{|j-k|}{N} : j, k \in \Delta \cap \{\{1, \ldots, N\} \times \{n\}\} \right\}.$$



$\nu_{\mathrm{col}}(\Delta, N) = 1/2$



$\nu_{\mathrm{row}}(\Delta, N) = 1/3$

Sparsity concepts

Definition

Let $\Delta \subset \{1, \ldots, N\}^2$. Δ is of cardinality s

- along its columns if $s = \max_{j=1}^{N} \left|\Delta_j^{[\mathrm{col}]}\right|$, where

$$\Delta_j^{[\mathrm{col}]} = \{(n_1, j) \in \Delta : n_1 = 1, \ldots, N\}.$$

- along its rows if $s = \max_{j=1}^{N} \left|\Delta_j^{[\mathrm{row}]}\right|$, where

$$\Delta_j^{[\mathrm{row}]} = \{(j, n_2) \in \Delta : n_2 = 1, \ldots, N\}.$$

Definition

Let $x \in \mathbb{C}^{N \times N}$. Let $x^{[\mathrm{col},j]}$ be the j^{th} column of x and $x^{[\mathrm{row},j]}$ be the j^{th} row of x. We say that x has T distinct supports

- along its columns if $T = \left|\left\{x^{[\mathrm{col},j]} : j = 1, \ldots, N\right\}\right|$.
- along its rows if $T = \left|\left\{x^{[\mathrm{row},j]} : j = 1, \ldots, N\right\}\right|$.

Our assumptions

Let $x \in \mathbb{C}^{N \times N}$ and let $\Delta_1, \Delta_2 \subset \{1, \dots, N\}^2$ index the largest s_1 coefficients of $D_1 x$ and the largest s_2 coefficients of $D_2 x$ resp.

- Along its columns, Δ_1 has a minimum separation of $2/M_1$ and is of cardinality s_1.
- Along its rows, Δ_2 has a minimum separation of $2/M_2$ and is of cardinality s_2.
- $P_{\Delta_1}\mathrm{sgn}(D_1 x)$ has T_1 distinct supports along its columns.
- $P_{\Delta_2}\mathrm{sgn}(D_2 x)$ has T_2 distinct supports along its rows.

For a Cartesian line sampling index set Ω with $|\Omega| = m$ and $y = P_\Omega x + \eta$ with $\|\eta\| \leq \delta\sqrt{m}$, we will consider the solutions of

$$\min_{z \in \mathbb{C}^{N \times N}} \|z\|_{TV,\mathrm{aniso}} \text{ subject to } \|P_\Omega U z - y\| \leq \delta\sqrt{m}.$$

Recall that D_1 performs finite differences along each column and D_2 performs finite differences along each row. The anisotropic total variation norm of $x \in \mathbb{C}^{N \times N}$ is

$$\|x\|_{TV,\mathrm{aniso}} := \|D_1 x\|_1 + \|D_2 x\|_1 = \sum_{\mathbf{j}} |(D_1 x)_{\mathbf{j}}| + |(D_2 x)_{\mathbf{j}}|,$$

as opposed to the isotropic total variation norm $\|x\|_{TV} = \sum_{\mathbf{j}} \sqrt{|(D_1 x)_{\mathbf{j}}|^2 + |(D_2 x)_{\mathbf{j}}|^2}$.

Sampling along Cartesian lines

Theorem (P. 2015)

Let $\epsilon \in (0,1)$. Let $\Omega = \{0\} \cup \{\Omega_1 \times [N]\} \cup \{[N] \times \Omega_2\}$, and $m = |\Omega|$, where

$$\Omega_1 \sim \mathrm{Unif}([M_1], m_1), \qquad \Omega_2 \sim \mathrm{Unif}([M_2], m_2),$$

$$m_1 \gtrsim \max\left\{\log^2(T_1 M_1/\epsilon), \quad \log(N), \quad s_1 \log(T_1 s_1/\epsilon)\log(T_1 M_1/\epsilon)\right\},$$

and

$$m_2 \gtrsim \max\left\{\log^2(T_2 M_2/\epsilon), \quad \log(N), \quad s_2 \log(T_2 s_2/\epsilon)\log(T_2 M_2/\epsilon)\right\}.$$

Then, with probability exceeding $1-\epsilon$, any minimizer $\hat{x}$ satisfies

$$\|D(x-\hat{x})\|_2 \lesssim \frac{N^2}{M_0^2}\left((m_0 N)^{-1/2}\sqrt{m}\delta + \left\|P_{\Delta_1}^{\perp} D_1 x\right\|_1 + \left\|P_{\Delta_2}^{\perp} D_2 x\right\|_1\right),$$

and

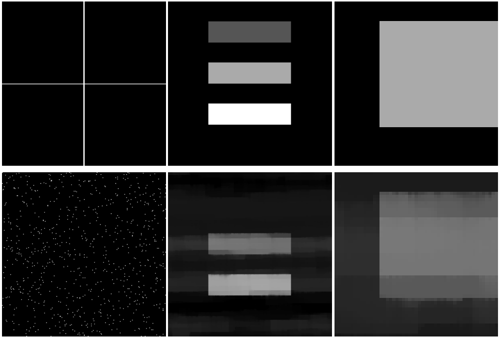
$$\|x-\hat{x}\| \lesssim \frac{N^2}{M_0^2}\sqrt{s}\left((m/m_0)^{1/2}\delta + \left\|P_{\Delta_1}^{\perp} D_1 x\right\|_1 + \left\|P_{\Delta_2}^{\perp} D_2 x\right\|_1\right),$$

where $s = \max\{s_1, s_2\}$, $m_0 = \min\{m_1, m_2\}$, and $M_0 = \min\{M_1, M_2\}$.

If $\Omega_1 = [M_1]$ and $\Omega_2 = [M_2]$, then these bounds hold with probability one.

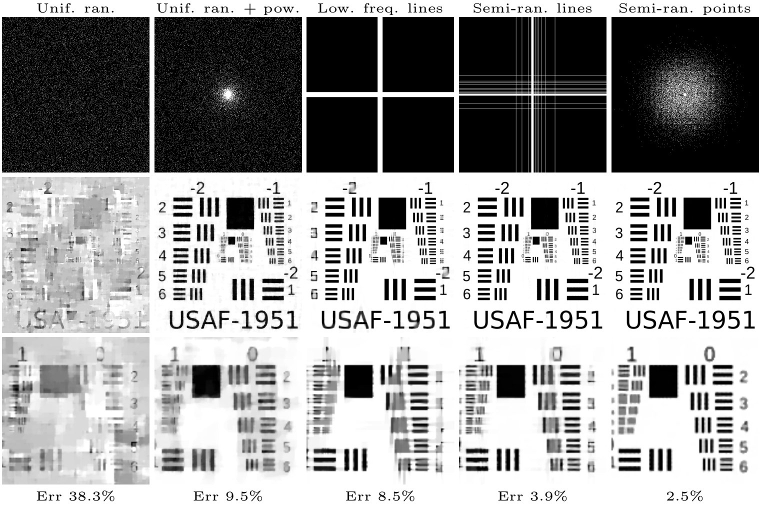
Example: Sampling 1.2% of the Fourier coefficients

If x has at most s_1 discontinuities along each of its columns with a minimum separation of $2/s_1$ and it has at most s_2 discontinuities along each of its rows with a minimum separation of $2/s_2$, then one is guaranteed exact recovery by sampling along $2(s_1 + s_2)$ Cartesian lines.



Sampling in accordance to sparsity structure allows for sub-$\mathcal{O}(s \log N)$ recovery.

Reconstruction of the 1951 USAF resolution test chart (6.5% sampling)



Conclusions

Although uniform random sampling is stable and robust...

- a uniform + power law sampling strategy achieves recovery guarantees which are optimal (for sparse vectors) up to log factors.
- in the 1D case where the discontinuities of the underlying signal are sufficiently far apart, one only needs to sample from low Fourier frequencies to ensure exact recovery.
- in the 2D case, recovery guarantees were presented for sampling along Cartesian lines. The sampling result is dependent both on sparsity and the sparsity separation in each direction.
- by accounting for sparsity structure, one can circumvent the $\mathcal{O}(s \log N)$ bound.
- variable density sampling schemes appear to combine the benefits of super resolution and compressed sensing: allows for a linear correspondence between the coarse features recovered and the number of samples, and also the recovery of fine features at the price of a log factor.

Conclusions

Although uniform random sampling is stable and robust...

- a uniform + power law sampling strategy achieves recovery guarantees which are optimal (for sparse vectors) up to log factors.
- in the 1D case where the discontinuities of the underlying signal are sufficiently far apart, one only needs to sample from low Fourier frequencies to ensure exact recovery.
- in the 2D case, recovery guarantees were presented for sampling along Cartesian lines. The sampling result is dependent both on sparsity and the sparsity separation in each direction.
- by accounting for sparsity structure, one can circumvent the $\mathcal{O}(s \log N)$ bound.
- variable density sampling schemes appear to combine the benefits of super resolution and compressed sensing: allows for a linear correspondence between the coarse features recovered and the number of samples, and also the recovery of fine features at the price of a log factor.

Thanks for listening!

On the role of total variation in compressed sensing. SIAM J. Imaging Sci., 8(1), 682-720, 2015.

Remark on the proofs

One can show that if $x \in \mathbb{C}^N$ and $\mathrm{supp}(Dx) = \Delta$, then x is the unique solution to

$$\min_z \|z\|_{TV} \text{ subject to } P_\Omega U x = P_\Omega U z,$$

provided that

1. $P_\Omega U P_\Delta$ is injective,
2. There exists $\eta \in \mathrm{ran}(U^* P_\Omega)$ such that $\eta_j = \mathrm{sgn}(Dx)_j$ for all $j \in \Delta$ and $\|\eta\|_\infty \leq 1$.
3. $\|h\| \leq C(N) \|h\|_{TV}$ whenever $P_\Omega U h = 0$.

- The second condition is simply asking if there exists a trigonometric polynomial

  $$p = \sum_{j \in \Omega} \alpha_j e^{2\pi i \langle j, \cdot \rangle}$$

  which interpolates the sign pattern of Dx and $\|p\|_\infty \leq 1$.
- The third condition is true if $0 \in \Omega$, and one can show that this condition holds with smaller constants $C(N)$ under power law sampling.