Recovery guarantees for TV regularized compressed sensing

Clarice Poon

University of Cambridge, UK

Abstract—**This paper considers the problem of recovering a one or two dimensional discrete signal which is approximately sparse in its gradient from an incomplete subset of its Fourier coefficients which have been corrupted with noise. The results show that in order to obtain a reconstruction which is robust to noise and stable to inexact gradient sparsity of order s with high probability, it suffices to draw $\mathcal{O}(s \log N)$ of the available Fourier coefficients uniformly at random. However, if one draws $\mathcal{O}(s \log N)$ samples in accordance to a particular distribution which concentrates on the low Fourier frequencies, then the stability bounds which can be guaranteed are optimal up to $\log$ factors. The final result of this paper shows that in the one dimensional case where the underlying signal is gradient sparse and its sparsity pattern satisfies a minimum separation condition, then to guarantee exact recovery with high probability, for some $M < N$, it suffices to draw $\mathcal{O}(s \log M \log s)$ samples uniformly at random from the Fourier coefficients whose frequencies are no greater than M.**

I. INTRODUCTION

This paper revisits one of the first examples of compressed sensing: the recovery of a gradient sparse signal from a highly incomplete subset of its Fourier coefficients. To motivate this paper, we first recall the key result from the seminal paper of Candès, Romberg and Tao [1]. Let $N \in \mathbb{N}$ and let the discrete gradient operator be $D : \mathbb{C}^N \to \mathbb{C}^N$ with $Dz = (z_j - z_{j+1})_{j=1}^N$ where $z_{N+1} := z_1$ for all $z \in \mathbb{C}^N$. Define the total variation (TV) norm by $\|z\|_{TV} = \|Dz\|_1$. Let $A \in \mathbb{C}^{N \times N}$ be the discrete Fourier transform on $\mathbb{C}^N$ such that given $z \in \mathbb{C}^N$,

$$Az = \left(\sum_{j=1}^N z_j e^{2\pi i k j / N} \right)_{k=-\lfloor N/2 \rfloor + 1}^{\lceil N/2 \rceil}. \quad (1)$$

The main result of [1] was that if $x \in \mathbb{C}^N$ was such that $|\{j : (Dx)_j \neq 0\}| = s$ and $\Omega = \Omega' \cup \{0\}$ where $\Omega' \subset \{-\lfloor N/2 \rfloor + 1, \ldots, \lceil N/2 \rceil\}$ consists of m indices chosen uniformly at random with

$$m \geq C \cdot s \cdot \left(\log(N) + \log(\epsilon^{-1}) \right)$$

for some numerical constant C, then, with probability exceeding $1 - \epsilon$, x is the unique solution to

$$\min_{z \in \mathbb{C}^N} \|z\|_{TV} \quad \text{subject to} \quad P_\Omega A z = P_\Omega A x. \quad (2)$$

Here, P_Ω denotes the projection matrix which restricts a vector to its entries indexed by Ω. In words, any gradient s-sparse signal can be perfectly recovered by $\mathcal{O}(s \log N)$ of its discrete Fourier coefficients. This represents a significant saving in the data acquisition process when $s << N$ and consequently, generated much research into the use of compressed sensing for imaging problems associated with the Fourier transform, such as magnetic resonance imaging, radio interferometry and electron tomography. (Note that the latter two applications are in fact associated with the Radon transform but can be seen as a Fourier sampling problem through the Fourier slice theorem). However, to fully understand the use of TV regularized compressed sensing for applications, it is imperative that we consider the following two questions.

1) Natural signals are generally not perfectly sparse but compressible in their gradient, and furthermore, measurements will likely be contaminated with noise. So, it is more realistic to consider the recovery of a signal $x \in \mathbb{C}^N$ from $y \in \mathbb{C}^\Omega$ such that $\|P_\Omega A x - y\|_2 \leq \delta\sqrt{m}$ where $m = |\Omega|$ and $\delta > 0$ is the noise level. This leads to the study of solutions of

$$\min_{z \in \mathbb{C}^N} \|z\|_{TV} \quad \text{subject to} \quad \|P_\Omega A z - y\|_2 \leq \delta\sqrt{m}, \quad (3)$$

and an immediate question is, to what the extent does choosing $\mathcal{O}(s \log N)$ samples uniformly at random guarantee reconstructions which are robust to noise and stable to inexact sparsity?

2) The work in [1] has generated much empirical work on the optimal choice of Ω and this has led to the proposal of variable density sampling schemes (see for example [2]) where one samples more densely at low Fourier frequencies and less densely at higher Fourier frequencies. So, why is uniform random sampling not used in practice and what advantages does dense sampling at low frequencies bring?

The results in this paper will partially address these two questions. Note that the second question has been of prominence recently [3], [4] and whilst it is now well understood in the case of wavelet regularization [4], the case of TV remains open.

A. Notation

This paper considers the recovery of both one dimensional and two dimensional signals. We have already defined the discrete Fourier operator A and the discrete gradient operator D for vectors in $\mathbb{C}^N$. To state the two dimensional results, we define the discrete Fourier transform and the discrete gradient operator for two dimensional vectors as follows. Let A be the discrete Fourier transform on $\mathbb{C}^{N \times N}$, such that given $z \in \mathbb{C}^{N \times N}$,

$$Az = \left(\sum_{j_1=1}^N \sum_{j_2=1}^N z_{j_1,j_2} e^{2\pi i (j_1 k_1 + j_2 k_2)/N} \right)_{k_1,k_2=-\lfloor N/2 \rfloor + 1}^{\lceil N/2 \rceil}. \quad (4)$$

Define the vertical gradient operator as

$$D_1 : \mathbb{C}^{N \times N} \to \mathbb{C}^{N \times N}, \quad x \mapsto (x_{j+1,k} - x_{j,k})_{j,k=1}^N$$

with $x_{N+1,k} = x_{1,k}$ for each $k = 1, \ldots, N$ and the horizontal gradient operator as

$$D_2 : \mathbb{C}^{N \times N} \to \mathbb{C}^{N \times N}, \quad x \mapsto (x_{j,k+1} - x_{j,k})_{j,k=1}^N$$

with $x_{j,N+1} = x_{j,1}$ for each $j = 1, \ldots, N$. Now define the gradient operator $D : \mathbb{C}^{N \times N} \to \mathbb{C}^{N \times N}$ as

$$Dx = D_1 x + i D_2 x, \quad (5)$$

and the isotropic total variation norm as

$$\|x\|_{TV} = \|Dx\|_1.$$

Given any $\Lambda \subset \mathbb{Z}^2$, and $x \in \mathbb{C}^{N \times N}$, $P_\Lambda : \mathbb{C}^{N \times N} \to \mathbb{C}^{N \times N}$ is the projection operator such that $P_\Lambda x$ is the restriction of x to its entries indexed by Λ.

Throughout, A denotes the discrete Fourier transform and D denotes the discrete gradient transform. It will be clear from the context whether the one dimensional or two dimensional definitions are being applied. The following three sections will present recovery guarantees for the 1D problem (3) and the 2D problem

$$\min_{z \in \mathbb{C}^{N \times N}} \|z\|_{TV} \text{ subject to } \|P_\Omega A z - y\|_2 \leq \delta\sqrt{m}. \quad (6)$$

II. UNIFORM + POWER LAW SAMPLING

This section presents results which offer near optimal error bounds when considering the recovery of gradient compressible signals by solving (3) and (6) with a uniform + power law sampling strategy.

Definition II.1. (i) We say that $\Omega \subset \{-\lfloor N/2 \rfloor + 1, \ldots, \lceil N/2 \rceil\}$ is a (one dimensional) uniform + power law sampling scheme of cardinality m if $\Omega = \Omega_1 \cup \Omega_2$, with Ω_1 and Ω_2 defined as follows. Let Ω_1 be m indices chosen uniformly at random, and let $\Omega_2 = \{k_1, \ldots, k_m\}$ consist of m indices which are independent and identically distributed (i.i.d.) such that for each $j = 1, \ldots, m$ and $n = -N/2 + 1, \ldots, N/2$,

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

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

(ii) We say that $\Omega \subset \{-\lfloor N/2 \rfloor + 1, \ldots, \lceil N/2 \rceil\}^2$ is a (two dimensional) uniform + power law sampling scheme of cardinality m if $\Omega = \Omega_1 \cup \Omega_2$ where Ω_1 and Ω_2 are defined as follows. Ω_1 consists of m indices chosen uniformly at random, and $\Omega_2 = \{k_1, \ldots, k_m\}$ consist of m i.i.d. indices such that for each $j = 1, \ldots, m$, and $n, m = -N/2 + 1, \ldots, N/2$,

$$\mathbb{P}(k_j = (n, m)) = p(n, m),$$

$$p(n, m) = C \left(\log(N) \max\{1, |n|^2 + |m|^2\}\right)^{-1},$$

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

We now present a result concerning TV regularization in the 1D case using a uniform + power law sampling scheme.

Theorem II.2. *[5] For $N = 2^J$ with $J \in \mathbb{N}$, let A be the discrete Fourier transform and let D be the discrete gradient operator on $\mathbb{C}^N$. Let $\epsilon \in (0, 1)$ and let $\Delta \subset \{1, \ldots, N\}$ with $|\Delta| = s$. Let $x \in \mathbb{C}^N$. Let Ω be a uniform + power law sampling scheme of cardinality*

$$m \gtrsim s \cdot \log(N)(1 + \log(\epsilon^{-1})).$$

Suppose that $y \in \mathbb{C}^N$ is such that $\|y - P_\Omega A x\|_2 \leq \sqrt{m} \cdot \delta$ for some $\delta \geq 0$. Let $\hat{x}$ be a minimizer of (3). Then with probability exceeding $1 - \epsilon$,

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

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

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

Remark II.3. In the presence of noise and inexact sparsity, recall from [6, Theorem 9.14] that if $f \in BV[0, 1)$ (the space of bounded variation functions), then the optimal error-decay rate for all bounded variation functions by any type of nonlinear approximation $\tilde{f}$ from s samples is $\|\tilde{f} - f\|_{L^2[0,1)} = \mathcal{O}(\|f\|_V \cdot s^{-1})$. By comparison to the optimal error bounds achievable for bounded variation function, the term of $\|x\|_{TV} \cdot s^{-1}$ is inevitable. Thus, one can improve upon this result only by removing the log factors in the error bound.

In the 2D case, we have the following analogous result.

Theorem II.4. *[5] Let $N = 2^J$ for some $J \in \mathbb{N}$. Let $x \in \mathbb{C}^{N \times N}$. Let $\epsilon \in (0, 1)$ and let $\Delta \subset \{1, \ldots, N\}^2$ with $|\Delta| = s$. Let Ω be a uniform + power law sampling scheme of cardinality m with*

$$m \gtrsim s \cdot \log(N)(1 + \log(\epsilon^{-1})).$$

and suppose that $y \in \mathbb{C}^N$ is such that $\|y - P_\Omega A x\|_2 \leq \sqrt{m} \cdot \delta$ for some $\delta \geq 0$. Then, with probability exceeding $1 - \epsilon$, any minimizer $\hat{x}$ of (6) satisfies

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

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

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

Remark II.5. Up to the log factors, the error bound is typical of compressed sensing results and as explained in [7], [8], the term of $\|P_\Delta^\perp Dx\|_1 \cdot s^{-1/2}$ in the error bound cannot be avoided and the error bounds of Theorem II.4 can only be improved by the removal of log factors.

A. Relation to previous works

Prior work relating to the use of TV in compressed sensing for the stable and robust recovery of signals in two or higher dimensions include [8], which considered the use of a linear sampling operator constructed from random Gaussians. Their techniques were later used in [3] to derive recovery results for the case of weighted Fourier samples. More recently, recovery guarantees for TV minimization from random Gaussian samples in the one dimensional case have also been derived [9].

However, to date, there has been few works directly analysing the use of TV when sampling the Fourier transform and the purpose of this paper is to extend the result of [1] to include the case of inexact gradient sparsity and noisy Fourier measurements.

In contrast to the results of [3], our results are not concerned with universal recovery where we guarantee the recovery of *all* gradient s-sparse signals from one random sampling set Ω. Instead, our results concern the recovery of one specific signal from a random choice of Ω. For this reason, we require only $\mathcal{O}(s \log N)$ samples for recovery up to sparsity level s, as opposed to $\mathcal{O}(s \log^5 N \log^3 s)$ samples as derived in [3]. Furthermore, our results assume the standard uniform noise model instead of the weighted noise model considered in [3].

III. UNIFORM RANDOM SAMPLING

The following two results show that uniform random sampling on its own can still achieve stable and robust recovery, albeit with non-optimal error estimates. For the recovery of 1D signals, we have the following result.

Theorem III.1. *[5] For $N \in \mathbb{N}$, let $x \in \mathbb{C}^N$. Let $\epsilon \in (0, 1)$ and let $\Delta \subset \{1, \ldots, N\}$ with $|\Delta| = s$. Let $x \in \mathbb{C}^N$ and let $\Omega = \Omega' \cup \{0\}$ where $\Omega' \subset \{-\lfloor N/2 \rfloor + 1, \ldots, \lceil N/2 \rceil\}$ consists of m indices chosen uniformly at random with*

$$m \gtrsim s \cdot \left(1 + \log(\epsilon^{-1})\right) \cdot \log(N) \quad (7)$$

for some numerical constant C. Suppose that $y = P_\Omega A x + \eta$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$. Then with probability exceeding $1 - \epsilon$, any minimizer

$\hat{x}$ of (3) satisfies

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

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

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

Again, an analogous result can be obtained in the 2D case:

Theorem III.2. *[5] Let $N \in \mathbb{N}$. Let $x \in \mathbb{C}^{N \times N}$. Let $\epsilon \in (0,1)$ and let $\Delta \subset \{1, \ldots, N\}^2$ with $|\Delta| = s$. Let $\Omega = \Omega' \cup \{0\}$ where $\Omega' \subset \{-\lfloor N/2 \rfloor + 1, \ldots, \lceil N/2 \rceil\}^2$ consists of m indices chosen uniformly at random with*

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

Suppose that $y = P_\Omega Ax + \eta$ where $\|\eta\|_2 \leq \sqrt{m} \cdot \delta$. Then, with probability exceeding $1 - \epsilon$, any minimizer $\hat{x}$ of (6) satisfies

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

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

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

Although it is open as to whether better error bounds are possible if one restricts to uniform random sampling, the numerical examples presented in Section V-A suggest that any improvement over the results of this section will be limited.

IV. LOW FREQUENCY SAMPLING

The final result of this paper considers the reconstruction of one dimensional vectors when we sample only the low Fourier frequencies. In particular, the result shows that if the discontinuities of the underlying signal to be recovered are sufficiently far apart, then we only need to sample from low Fourier frequencies. We first required a definition.

Definition IV.1. Let $N \in \mathbb{N}$, $\Delta = \{t_1, \ldots, t_s\} \subset \{1, \ldots, N\}$ with $t_1 < t_2 < \cdots < t_s$ and $t_0 = -N + t_s$. The minimum separation distance is defined to be

$$\nu_{\min}(\Delta, N) = \min_{j=1}^{s} \frac{|t_j - t_{j-1}|}{N}.$$

The following result provides some initial insight into how the minimum separation distance of a 1D signal should impact the sampling strategy.

Theorem IV.2. *[5] For $N \in \mathbb{N}$, let $x \in \mathbb{C}^N$. Let $\epsilon \in [0,1]$ and let $M \in \mathbb{N}$ be such that $N/4 \geq M \geq 10$. Suppose that $\nu_{\min}(\Delta, N) = \frac{1}{M}$. Let $\Omega = \Omega' \cup \{0\}$ where $\Omega' \subset \{-2M, \ldots, 2M\}$ consists of m indices chosen uniformly at random with*

$$m \gtrsim \max \left\{ \log^2 \left(\frac{M}{\epsilon} \right), s \cdot \log \left(\frac{s}{\epsilon} \right) \cdot \log \left(\frac{M}{\epsilon} \right) \right\}. \quad (8)$$

Then with probability exceeding $1 - \epsilon$, any minimizer $\hat{x}$ of (3) with $y = P_\Omega Ax + \eta$ and $\|\eta\|_2 \leq \delta \cdot \sqrt{m}$ satisfies

$$\frac{\|x - \hat{x}\|_2}{\sqrt{N}} \lesssim \frac{N^2}{M^2} \cdot \left(\delta \cdot s + \sqrt{s} \cdot \left\| P_\Delta^\perp Dx \right\|_1 \right). \quad (9)$$

Furthermore, if $m = 4M + 1$, then the error bound (9) holds with probability 1.

Fig. 1. (a) Original test image (512×512); (b), (c) & (d) Reconstruction of the test image from 35% of its noise corrupted Fourier coefficients chosen uniformly at random. The SNR's are ∞, 10 and 5 respectively and the relative errors are 0.20, 0.22 and 0.46 respectively.

Due to a commutative relationship between the discrete Fourier transform and the discrete gradient operator, this result is closely related to the idea of super-resolution, which considers the recovery of a sum of diracs from its low frequency Fourier samples [10], [11]. Even though super-resolution is studied in an infinite dimensional setting, the proof of Theorem IV.2 exploits finite dimensional versions of the results in [10], [11].

V. NUMERICAL INSIGHTS

A. Variable density sampling and stability

The theoretical results demonstrated that uniform random sampling strategies are stable to inexact sparsity and robust to noise, and this can be observed in Figure 1. However, the provable error bounds obtained for the uniform random sampling strategy are sub-optimal, whereas, by adding the samples which are chosen in accordance to the nonuniform distribution, one can guarantee near-optimal error bounds. So, this begs the question: *Does sampling in accordance to a nonuniform distribution actually improve stability, or is the improved stability between the theorems simply an artefact of the proofs?* To empirically address this question, consider the following experiment.

Given $N \in \mathbb{N}$, a gradient sparse vector $x \in \mathbb{R}^N$ and an inexact sparsity level S, perturb x by a randomly generated vector $h \in \mathbb{R}^N$, where h is such that $S = 10 \log_{10} \left(\|x\|_2 / \|h\|_2 \right)$. Note that this is the SNR of h relative to x, and smaller values of S represent larger magnitudes of perturbations. We now consider the reconstruction of the approximately sparse signal $x+h$ from $P_\Omega A(x+h)$. The sampling set Ω is such that its cardinality is $\lceil 0.15N \rceil$, and it is either chosen uniformly at random which we denote by Ω_U, or in accordance to the uniform + power law sampling scheme which we denote by Ω_P.

This experiment is performed for perturbations of two sparse signals, shown in Figure 2, and the relative errors of reconstructing the approximately sparse versions of these signals via solving (3) with $\delta = 0$ (since we are investigating stability rather than robustness) are shown in Figure 3. Observe that both sampling with Ω_P and Ω_U exhibit stability with respect to inexact sparsity, since the relative errors all decay as the SNR values increase. However, the relative errors obtained when sampling with Ω_P are much lower, suggesting that one of the benefits offered by dense sampling around the zero frequency is increased stability. Finally, it is perhaps interesting to note that the theoretical results guarantee optimal error bounds (up to log factors) on the *recovered gradient* for both uniform and uniform + power law sampling schemes, and Figure 4 confirms this result by showing that there is no substantial difference between the error on the recovered gradient between Ω_U and Ω_P. So, experimentally, it appears as though dense sampling at low frequencies will significantly improve the stability of the recovered signal, although not the stability of the recovered gradient.

This improvement in stability is particularly visible in two dimensions. Consider the recovery of the 256×256 SIPI Image Database 'Peppers' test image. Figure 5 shows the reconstruction
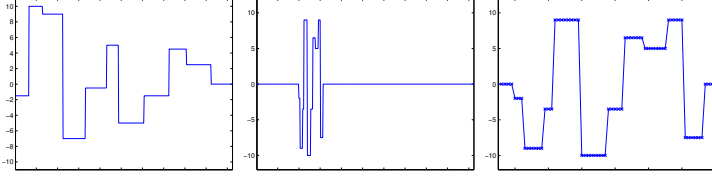
Fig. 2. The coarse signal, x_1, to be perturbed (left). The fine signal x_2 to be perturbed (centre), and a zoom in of x_2 (right) on indices between 90 and 160, for clarity, the values of x_2 on each index is marked by a cross.
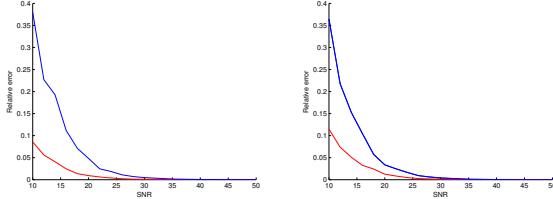

Fig. 3. Left: plot of relative error $\|\hat{x} - (x+h)\|_2 / \|x+h\|_2$ against $\text{SNR} = 10\log_{10}\left(\|x\|_2 / \|h\|_2\right)$, where $x := x_1$ is shown in Figure 2 and $\hat{x}$ is the reconstruction. The blue line refers to choosing $\Omega := \Omega_U$. The red line refers to choosing $\Omega := \Omega_P$. Right: the equivalent plot for $x := x_2$.

from sampling uniformly at random or in accordance to the uniform + power law scheme. Again, the reconstructions are obtained by solving (6) with $\delta = 0$, so we consider only the sparsity stability rather than noise robustness. The improvement in reconstruction quality is substantial and as suggested by this section and our theoretical result, one possible reason for this is that additional samples at low frequencies are required for optimal stability.

B. The price of randomness

This section elaborates on the final result of this paper; although $\mathcal{O}(s \log N)$ is the optimal sampling cardinality for s-sparse signals [1], and this can be attained through drawing samples uniformly at random, when one is interested in a subset of the possible s-sparse signals (e.g. signals whose discontinuities are sufficiently far apart), it may be unnecessary to pay the price of this log factor.

1) Sampling high frequencies is not necessary for coarse signals: Under an additional assumption that the separation of the nonzero gradient entries is at least $1/M$, Theorem IV.2 stipulates that we can sample uniformly at random from the first $4M$ samples at a slightly
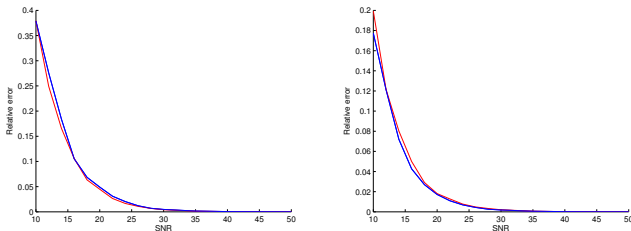

Fig. 4. Left: y-axis shows the relative error of the recovered gradient: $\|D(x+h-\hat{x})\|_2 / \|D(x+h)\|_2$, where $\hat{x}$ is the recovered signal, $x := x_1$ is the sparse signal shown in Figure 2, and h is the perturbation. The x-axis shows the SNR of h relative to x. The blue line corresponds to sampling uniformly at random, the red line corresponds to uniform plus variable sampling of Theorem II.2. Right: same as the left graph, except that $x := x_2$, where x_2 is as shown in Figure 2.

Fig. 5. Left: reconstructions from sampling uniformly at random (plus the zero frequency sample) at 10% with a relative error of 0.33. Right: reconstruction from sampling in the semi-uniform random manner described in Theorem II.4 at 10% with a relative error of 0.07.

smaller sampling order of $\mathcal{O}(s \log(M) \log(s))$ (although the provable stability and robustness bounds are worse by a factor of $\sqrt{s}$, where s is the approximate sparsity).

This firstly suggests that an understanding of the gradient structure of the underlying signal can lead to sampling patterns which will outperform uniform random sampling. Secondly, in order to recover an s sparse signal of length N, one requires $\mathcal{O}(s \log(N))$ random samples and this sampling cardinality is sharp for sparse signals [1]. Thus, although such a statement guarantees the recovery of any s-sparse signal, there is a price of $\log(N)$ associated with the randomness introduced. However, suppose that our signal of interest (denote by x) is of length N, is M-sparse in its gradient and these nonzero gradient entries have minimum separation of $1/M$. Then, Theorem IV.2 says that x can be perfectly recovered from its first $4M + 1$ Fourier samples of lowest frequencies. Note that there is no randomness in the choice of sampling set Ω and the cardinality of Ω is linear with respect to sparsity. Observe also that a uniform random choice of Ω is guaranteed to result in accurate reconstructions and allow for significant subsampling only if $M \log(N) << N$. So in the case that $M \geq N / \log(N)$, it will be better to choose Ω to index the first M samples, rather than draw the samples uniformly at random.

2) A numerical example: To illustrate the remarks above, consider the recovery of x_1 of length $N = 512$ shown on the left of Figure 2. It can be perfectly recovered by solving the following minimization problem with Ω indexing the first 20 Fourier frequencies. This accounts for 3.9% of the available Fourier coefficients. For simplicity, we present this experiment without adding noise to the samples, although similar results can be observed if noise is added.

$$\min_{z \in \mathbb{C}^{N \times N}} \|z\|_{TV} \text{ subject to } P_\Omega A z = P_\Omega A x_1. \tag{10}$$

The result of repeating this experiment over 5 trials with Ω taken to be 3.9%, 7%, 10% of the available indices, drawn uniformly at random, is shown in Table I. By sampling uniformly at random, we cannot achieve exact recovery from drawing only 3.9% and it is only when we sample at 10% that we obtain exact recovery across all 5 trials.

3) The need for further investigation: Structured sampling: To conclude, we present a numerical example to show that despite the advances in the theoretical understanding of TV for compressed sensing, there is still room for substantial improvements. Consider the reconstruction of the resolution chart of size 528×500 in Figure 6 from 6.5% of its available Fourier coefficients using different sampling maps

(i) Ω_U indexes samples drawn uniformly at random,
(ii) Ω_L which indexes the samples of lowest Fourier frequencies,

Trial	Sampling 3.9%	Sampling 7%	Sampling 9%	Sampling 10%
1	0.9096	0.2150	0	0
2	0.7739	0.1915	0	0
3	0.4388	0	0.1132	0
4	0.7287	0.4396	0.1603	0
5	0.7534	0.3044	0	0

TABLE I
RELATIVE ERROR OF RECONSTRUCTIONS OBTAINED BY SAMPLING THE
FOURIER TRANSFORM OF SIGNAL x_1 UNIFORMLY AT RANDOM.

Fig. 6. The 1951 USAF resolution test chart of size 528×500.

Fig. 7. Top row: the Fourier sampling maps, each indexing 6.5% of the available Fourier samples. Bottom row: zooms in on the reconstructions for pixels in $[180, 300] \times [220, 340]$. The relative errors are (from left to right) 0.39, 0.07, 0.09, 0.03.

(iii) Ω_P which is chosen in accordance to the uniform + power law,

(iv) Ω_V which is constructed by first dividing up the available indices into L levels in increasing order of frequency, such that $\mathbb{P}(X_j = k) = C \cdot \exp(-(bn/L)^a)$ for some appropriate constant C such that we have a probability measure, X_j is the jth element of Ω_V and k belongs to the Lth level. In this experiment, we chose $L = 25$, $a = 2.2$, and $b = 6.5$.

Conclusion of the experiment: The following observations can be made from the sampling maps and the reconstructions shown in Figure 7.

 (i) (Ω_U) Uniform random sampling yields a high relative error.

 (ii) (Ω_L) Sampling only the low Fourier frequencies recovers only the coarse details.

(iii) (Ω_P) Concentrating on low Fourier frequencies but also sampling high Fourier frequencies allowed for the recovery of both the coarse and fine details.

(iv) (Ω_V) Allowed for the recovery of both the coarse and fine details, but is substantially better than uniform + power law.

So, uniform random sampling maps are applicable only in the case of extreme sparsity due to the price of a $\log$ factor, whilst either fully sampling or subsampling the low frequencies will be applicable when we aim to only recover low resolution components of the underlying signal. This suggests that variable density sampling patterns are successful because they accommodate for a combination of these two scenarios – when there is both high and low resolution components which we want to recover and some sparsity – sampling fully at the low frequencies will allow for recovery of coarse details without the price of a $\log$ factor, whilst increasingly subsampling at high frequencies will allow for the recovery of fine details up to a $\log$ factor. One can essentially repeat this experiment for any natural image to observe the same phenomenon: by choosing the samples uniformly at random, we will be required to sample more than is necessary.

VI. CONCLUDING REMARKS

This paper showed that an uniform + power law sampling strategy achieves recovery guarantees which are optimal up to $\log$ factors. Furthermore, in the 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. These results provide some initial justification for the preference of variable density sampling patterns over uniform random sampling patterns in terms of stability and reduction in the number of samples. However, the results of this paper provide only an initial understanding towards how the distribution of the Fourier coefficients favours the recovery of certain types of signals and can allow for sub-$\mathcal{O}(s \log N)$ samples. In particular, the final numerical example suggests that there exists a much deeper connection between the gradient sparsity structure of a signal and the dist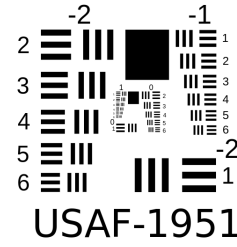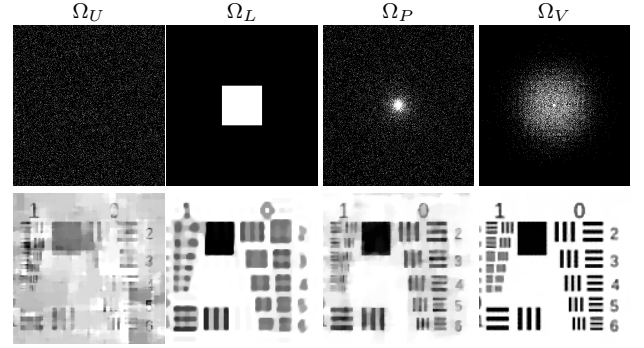ribution of the Fourier samples, and a thorough understanding of this connection could lead to the design of more efficient sampling strategies.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[2] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[3] F. Krahmer and R. Ward, "Stable and robust sampling strategies for compressive imaging," *Image Processing, IEEE Transactions on*, vol. 23, no. 2, pp. 612–622, Feb 2014.

[4] B. Adcock, A. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," *Preprint*, 2014.

[5] C. Poon, "On the role of total variation in compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 8, no. 1, pp. 682–720, 2015.

[6] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[7] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[8] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1035–1058, 2013.

[9] J.-F. Cai and W. Xu, "Guarantees of total variation minimization for signal recovery," *arXiv preprint arXiv:1301.6791*, 2013.

[10] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[11] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7465–7490, Nov 2013.