This paper considers the problem of recovering a one or two dimensional discrete signal which is approximately sparse in its discrete gradient from an incomplete set of its discrete Fourier coefficients which have been corrupted with noise. We prove that to obtain a reconstruction which is robust to noise and stable to inexact gradient sparsity with high probability, it suffices to draw $O(s \log N)$ of the available Fourier coefficients uniformly at random. We also show 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 $O(s \log M \log s)$ samples uniformly at random from the Fourier coefficients of low frequencies only.

1 Introduction

This paper revisits the theory behind one of the first instances of compressed sensing: the recovery of a gradient sparse signal from a small subset of its discrete Fourier data. This problem was first studied in [7], in which the following theorem was proved.

**Theorem 1.1** ([7]). Let $\epsilon \in (0, 1)$. Let $A \in \mathbb{C}^N$ be the discrete Fourier transform and let $x \in \mathbb{R}^N$. Suppose that $\Omega = \Omega' \cup \{0\}$ where $\Omega' \subset \{-\lfloor N/2 \rfloor + 1, \ldots, \lfloor N/2 \rfloor\}$ consists of $m$ indices chosen uniformly at random with

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

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

$$\min_{z \in \mathbb{R}^N} \|z\|_{TV} \text{ subject to } P_\Omega A z = P_\Omega A x \tag{1.1}$$

where $P_\Omega$ is a projection matrix, which restricts a vector to the index set $\Omega$ and for any $z \in \mathbb{R}^N$, $\|z\|_{TV} := \|Dz\|_1$ with $Dz = (z_j - z_{j+1})_{j=1}^N$ where $z_{N+1} := z_1$.

This result can be easily extended to two dimensions and was significant because of its close links to practical applications – many imaging devices sample the Fourier transform of the underlying object and natural images are generally thought to be ‘almost sparse’ in their gradients. The original motivation in [7] was parallel-beam tomography, but other areas in which this is of great interest include electron microscopy, magnetic resonance imaging and radio interferometry. This result suggests that a significant saving in the data acquisition process since it demonstrates that gradient sparse signals of length $N$ can be recovered from $O(s \log N)$ samples, and for $s << N$, is significantly smaller than the number specified by its Nyquist rate. However, in order to understand the role of total variation in compressed sensing for such practical applications, there are two ways in which we want to extend Theorem 1.1.

*cmhsp2@cam.ac.uk
1. For some unknown signal $x \in \mathbb{R}^N$, it is more realistic to assume that we do not observe $P_{\Omega}Ax$, but $y$ such that $\|P_{\Omega}Ax - y\|_2 \leq \delta \sqrt{m}$ with $m = |\Omega|$ and for some noise level $\delta > 0$. Furthermore, often $x$ is only approximately $s$-sparse in its gradient, i.e. there exists $\Delta \subset \{1, \ldots, N\}$ such that $|\Delta| = 1$ and $\|P_{\Delta}Dx\|_1 \ll 1$. So, in practice, the following minimization problem is solved in place of (1.1).

$$\min_{z \in \mathbb{R}^N} \|z\|_{TV} \text{ subject to } \|P_{\Omega}Az - y\|_2 \leq \sqrt{m} \cdot \delta.$$  

(1.2)

Will choosing $\Omega$ such that $0 \in \Omega$ and the remaining indices are $\mathcal{O}(s \log(N))$ indices uniformly at random guarantee robust and stable recovery of $x$? (Robustness and stability refers to control over the reconstruction error by $\delta$ and $\|P_{\Delta}Dx\|_1$ respectively).

2. Since the publication of [7], there has been much empirical work on how to choose $\Omega$ such that we minimize its cardinality, whilst still retaining stable and robust recovery. In practice, $\Omega$ is often chosen in accordance to some variable density distribution, which concentrates more a low Fourier frequencies and less on high Fourier frequencies. Furthermore, $\Omega$ is not chosen in the uniform random manner as specified by Theorem 1.1 because empirically, it has been observed to result in inferior reconstructions when compared with variable density sampling schemes. Thus, it is of interest to derive theoretical statements to understand the reconstruction qualities of solutions to (1.2) with a non-uniform choice of $\Omega$.

This paper will derive theoretical results for the first question. The second question will be considered in [16], however, we will derive a recovery result considering solutions of (1.2) when $\Omega$ indexes the first $M < N$ Fourier frequencies and discuss in Section 3 how this provides an initial understanding of the use of variable density sampling patterns.

Related results and overview

Prior work relating to the use of total variation in compressed sensing which derives results for the stable and robust recovery of signals in two or higher dimensions include [14, 13]. Their work considered the recovery of gradient sparse signals by solving (1.2) for some measurement matrix, $A$, which satisfies the restricted isometry property when composed with the discrete Haar transform. Although the discrete Fourier transform does not directly fit into this framework, [20] derived recovery results for the case of weighted Fourier samples. The purpose of this paper is to directly extend the result of [7] to include the case of inexact gradient sparsity and noisy Fourier measurements. This is presented in Section 2.1. The results are for the recovery of one or two dimensional signals, however, the techniques are applicable to higher dimensions. Note also that in contrast to the results of [14, 13, 20], the results of this paper are not concerned with universal recovery where we guarantee the recovery of all gradient $s$-sparse signals from one random sampling set $\Omega$. Instead, we derive results for the recovery of one specific signal from a random choice of $\Omega$. For this reason, the proofs in this paper do not rely on the restricted isometry property and we require only $\mathcal{O}(s \log N)$ samples for recovery as oppose to $\mathcal{O}(s \log^5 N)$ samples as derived in [20].

This paper will also consider the recovery of one dimensional signals whose sparsity pattern satisfies some minimum separation condition from low frequency Fourier samples. This is presented in Theorem 2.3. Due to the close 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 [4, 5, 18]. Even though super-resolution is studied in an infinite dimensional setting, the proof of Theorem 2.4 will make use of finite dimensional versions of the results in [4, 5, 18].

Finally, despite showing that a uniform random choice of the Fourier measurements guarantees robust and stable reconstructions, Section 3 presents some numerical examples with the aim of providing some intuition as to why uniform random sampling patterns fare poorly when compared with sampling patterns (of the same cardinality) which concentrates on samples of low frequencies.
2 Main results

2.1 Stable and robust recovery from uniform random sampling

Theorem 2.1. Let $A$ be the discrete Fourier transform on $\mathbb{R}^N$, such that given $z \in \mathbb{R}^N$,

$$Az = \left( \sum_{j=1}^{N} z_j e^{2\pi i k_j / N} \right)_{k=-[N/2]+1}^{[N/2]} \quad (2.1)$$

and let $\|z\|_{TV} := \sum_{j=1}^{N} |z_{j+1} - z_j|$ with $z_{N+1} := z_1$. Let $\varepsilon \in (0, 1)$ and let $\Delta \subset \{1, \ldots, N\}$ with $|\Delta| = s$. Let $\Omega = \Omega' \cup \{0\}$ where $\Omega' \subset \{-[N/2] + 1, \ldots, [N/2]\}$ consists of $m$ indices chosen uniformly at random with

$$m \geq C \cdot s \cdot (1 + \log(\varepsilon^{-1})) \cdot \log(N) \quad (2.2)$$

for some numerical constant $C$. Then with probability exceeding $1 - \varepsilon$, any minimizer $\hat{x}$ of

$$\min_{z \in \mathbb{R}^N} \|z\|_{TV} \text{ subject to } \|P_{\Omega}Az - y\|_2 \leq \sqrt{m} \cdot \delta$$

satisfies

$$\|x - \hat{x}\|_2 \leq C \cdot \log^{1/2}(m) \cdot \log(s) \cdot (\delta \cdot \sqrt{s} + \|P_{\Delta}Dx\|_1) \quad (2.3)$$

for some numerical constant $C$.

Remark 2.1. Let $x \in \mathbb{C}^N$ be the discretization of some function $f$ on the compact interval $[0, 1]$, with $x_j = f(j/N)$. Then it is natural to define the discrete $\ell^2$ norm (see for example, [11]) of $f$ as

$$\left( \sum_{j=1}^{N} \left| f(j/N) \right|^2 \right)^{1/2} = N^{-1/2} \cdot \|x\|_2.$$ 

Furthermore, the discrete gradient norm of $f$ is often defined to be

$$\frac{1}{N} \sum_{j=1}^{N} \left| \frac{f((j+1)/N) - f(j/N)}{1/N} \right| = \|x\|_{TV}.$$ 

So, there is in general a discrepancy of $\sqrt{N}$ between the discretized gradient and $\ell^2$ norm of a discretized signal and $N^{-1/2}$ term on the right hand side of the error estimate above is a natural occurrence. However, note that there is no such discrepancy in the two dimensional case.

Corollary 2.2. Let $A$ be the discrete Fourier transform on $\mathbb{R}^N$, such that given $z \in \mathbb{R}^{N \times N}$,

$$Az = \left( \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} z_{j_1,j_2} e^{2\pi i (j_1 k_1 + j_2 k_2) / N} \right)_{k_1,k_2=-[N/2]+1}^{[N/2]}$$

and let

$$\|z\|_{TV} := \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} \sqrt{|z_{j_1+1,j_2} - z_{j_1,j_2}|^2 + |z_{j_1,j_2+1} - z_{j_1,j_2}|^2}$$

with $z_{N+1,k} := z_{1,k}$ and $z_{k,N+1} := z_{k,1}$.

Let $\varepsilon \in (0, 1)$ and let $\Delta \subset \{1, \ldots, N\}$ with $|\Delta| = s$. Let $\Omega = \Omega' \cup \{0\}$ where $\Omega' \subset \{-[N/2] + 1, \ldots, [N/2]\}$ consists of $m$ indices chosen uniformly at random with

$$m \geq C \cdot s \cdot (1 + \log(\varepsilon^{-1})) \cdot \log(N^2)$$

for some numerical constant $C$. Then, with probability exceeding $1 - \varepsilon$, any minimizer $\hat{x}$ of

$$\min_{z \in \mathbb{R}^{N^2}} \|z\|_{TV} \text{ subject to } \|P_{\Omega}Az - y\|_2 \leq \sqrt{m} \cdot \delta$$

satisfies

$$\|x - \hat{x}\|_2 \leq C \cdot \log^{1/2}(m) \cdot \log(s) \cdot (\delta \cdot \sqrt{s} + \|P_{\Gamma}Dx\|_1)$$

where $C$ is a numerical constant.
2.2 Sampling from low Fourier frequencies

The following result considers the reconstruction obtained when we sample only the low Fourier frequencies. We first require a definition.

**Definition 2.3.** Given $N \in \mathbb{N}$ and $\Delta = \{ t_1, \ldots, t_s \} \subset \{1, \ldots, N \}$ with $t_1 < t_2 < \cdots < t_s$, let $t_0 = -N + t_s$. Then, the minimum separation distance is defined to be

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

**Theorem 2.4.** Let $A$ and $\| \cdot \|_{TV}$ be as in Theorem 2.1. 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 $\Omega = \Omega^U \cup \{ 0 \}$ where $\Omega^U \subset \{-2M, \ldots, 2M\}$ consists of $m$ indices chosen uniformly at random with

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

(2.4)

for some numerical constant $C$. Then with probability exceeding $1 - \epsilon$, any minimizer $\hat{x}$ of (2.1) satisfies

$$\frac{\| x - \hat{x} \|_2}{\sqrt{N}} \leq C \cdot \frac{N^2}{M^2} \cdot \left( \delta \cdot s + \sqrt{s} \cdot \| P_{\Delta} A x \|_1 \right)$$

(2.5)

where $C$ is a numerical constant. Furthermore, if $m = 4M + 1$, then the error bound (2.3) holds with probability 1.

3 The price of randomness

Various sampling patterns $\Omega$ have been shown to be successful for solving (1.2) [12, 10, 17, 3], but one thing which all those sampling patterns have in common is that they are not uniform random patterns. They all sample densely at low Fourier frequencies, and less densely at higher Fourier frequencies and such sampling patterns are often referred to as variable density sampling.

In fact, uniform random sampling patterns generally result in inferior reconstructions. As an example, consider the reconstruction of the test image shown on the left of Figure 1. All three of the sampling maps shown on the top row of Figure 2 index 6% of the available Fourier coefficients. The difference is that $\Omega^L$ is such that $0 \in \Omega^L$ and the rest of the samples are chosen uniformly at random, $\Omega^L$ is chosen in accordance to a variable density distribution and $\Omega^L$ indexes the samples of lowest Fourier frequencies. The reconstructions obtained by solving (1.2) for $\Omega = \Omega^U, \Omega^V, \Omega^L$ are shown in Figure 2 with the relative errors of each reconstruction denoted by $\epsilon_{rel} = \| \text{reconstruction} - \text{image} \|_2 / \| \text{image} \|_2$.

$$\min_{x \in \mathbb{C}^{N \times N}} \| z \|_{TV} \text{ subject to } \| P_{\Omega} A z - P_{\Omega} A x \|_2 \leq 10^{-6}.$$

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. Detailed analysis on the impact of non-uniform sampling patterns on the reconstruction quality in solutions to (1.2) will be covered in [16], however, Theorem 2.1 and Theorem 2.4 offers an initial understanding as to why variable density sampling can outperform uniform random sampling. Theorem 2.1 tells us that one can recover a gradient $s$-sparse signal from $O(s \log N)$ samples by choosing the samples uniformly at random regardless of where the nonzero gradient entries occur and such recovery is stable to inexact sparsity and robust to noise. However, under an additional assumption that the separation of the nonzero gradient entries is at least $1/M$, Theorem 2.4 stipulates that we can sample uniformly at random from the first $4M$ samples at a slightly smaller sampling order of $O((s \log(M)) \log(s))$ (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.

\[\text{The numerical algorithm used was the split Bregman method described in [9].}\]
Figure 1: the test image of dimension $512 \times 512$ taken from the USC-SIPI image database (left), its edge set generated using the MATLAB `edge` function with parameters ‘canny’ and 0.255 (middle), its edge set generated using the MATLAB `edge` function with parameters ‘canny’ and 0.78 (right)

| Trial | Relative Error (5.86% sampling) | Relative Error (10% sampling) | Relative Error (15% sampling) | Relative Error (20% sampling) |
|-------|---------------------------------|-------------------------------|-------------------------------|-------------------------------|
| 1     | 69.60                           | 0                             | 73.52                         | 0                             |
| 2     | 63.29                           | 0                             | 0                             | 0                             |
| 3     | 58.04                           | 68.98                         | 0                             | 0                             |
| 4     | 69.56                           | 49.83                         | 0                             | 0                             |
| 5     | 56.26                           | 0                             | 0                             | 0                             |

Table 1

The second key observation is that in most compressed sensing statements, to recover an $s$-sparse signal of length $N$, one requires $O(s \log(N))$ random samples. Thus, 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 entry have minimum separation of $1/M$. Then, Theorem 2.4 tells us 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 $\Omega$ and the cardinality of $\Omega$ is linear with respect to sparsity. Observe also that a uniform random choice of $\Omega$ 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 $\Omega$ to index the first $M$ samples, rather than draw the samples uniformly at random.

To illustrate the two observations, consider the recovery of the signal of length $N = 512$ shown on the left of Figure 3. It can be perfectly recovered by solving (1.2) with $\Omega$ indexing the first 31 Fourier frequencies and $\delta = 10^{-8}$ (the recovery will be considered exact if the relative error is no greater than $10^{-7}$). This accounts for 5.86% of the available Fourier coefficients. The result of running this experiment over 5 trial with $\Omega$ taken to be 5.86%, 10%, 15% and 20% of the available indices, drawn uniformly at random is shown in Table 1. By sampling uniformly at random, we cannot achieve exact recovery from drawing only 5.86% and it is only when we sample at 20% that we obtain exact recovery across all 5 trials.

Given the discussion on Theorem 2.1 and Theorem 2.4, consider again the example of Figure 2 – the length and width of the image is $N = 512$. The edge set of the test image shown on the middle of Figure 1 (generated using the MATLAB function `edge` with parameters ‘canny’ and 0.255) is of cardinality $15655 \approx 0.06N^2$. Suppose we let $s = 15655$, then $s \log(N^2) \approx 12.5s \approx 0.75N^2$. Thus, although Theorem 2.1 guarantees robust and stable recovery up to gradient sparsity $s$ from $O(s \log(N^2))$ measurements, for realistic images, the price of the factor of $\log(N^2)$ is significant and for reasonable subsampling ratios, $s$ is required to be very small. In order for $s \log(N^2)/N^2$ to be 0.06, we require that $s \approx 0.0048N^2$. The image shown on the right of Figure 1 is the edge set of largest support generated using the MATLAB canny edge detector for which its support is of cardinality $s \approx 0.0048N^2$. So, one cannot hope to recover sufficient sparse details using uniform random sampling to accurately represent the original image. On
Reconstruction from $\Omega_U$
$\epsilon_{rel} = 7.64\%$

Reconstruction from $\Omega_L$
$\epsilon_{rel} = 8.95\%$

Reconstruction from $\Omega_U$
$\epsilon_{rel} = 32.20\%$

Figure 2: The top row shows the Fourier sampling maps, each indexing 6% of the available Fourier samples. The bottom row shows the reconstructed images and their relative errors.

On the other hand, the result of Theorem 2.4 suggests that low resolution components of the underlying image can be recovered by fully sampling the low frequency Fourier data.

To conclude this discussion, consider the reconstruction of the resolution chart of size $256 \times 256$ on the right of Figure 3 from 8.7% of its available Fourier coefficients using different sampling maps, $\Omega_U$ indexes samples drawn uniformly at random, $\Omega_L$ which indexes the samples of lowest Fourier frequencies and $\Omega_U$ which allows for oversampling at low frequencies and significant subsampling at high Fourier frequencies. The reconstructions are shown in Figure 4: uniform random sampling yields a high relative error, sampling only the low Fourier frequencies recovers only the coarse details, whilst a sampling pattern which concentrates on low Fourier frequencies but also samples high Fourier frequencies recovers both the coarse and fine details. 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.

Relationship to the recovery of wavelet coefficients from Fourier samples

Theorem 2.4 shows that under an additional assumption on the minimum separation distance on the sparsity structure of the underlying signal, we may draw the samples from only samples of low Fourier frequencies. Furthermore, the last sentence of Theorem 2.4 implies that if the underlying signal has
gradient support \( \Delta \) with \( \nu_{\text{min}}(\Delta, N) = M^{-1} \), then the number of samples required for perfect recovery is \( \mathcal{O}(M) \), and in this case, there is no probability or log factor involved. As explained in this section, this explains why sampling only low frequency coefficients can lead to superior reconstruction quality when compared with sampling at random. This result is also reminiscent of the result from [15] which shows that the first Fourier coefficients of lowest frequencies will stably recover the first \( cM \) wavelet coefficients of lowest dilation factors for some constant \( c \leq 1 \) when solving an appropriate \( \ell^1 \) minimization problem. Note that this is a linear relationship between the number of samples and the number of recovered wavelet coefficients and there is no log factor involved. The recovery of wavelet coefficients from Fourier samples is also another application of compressed sensing in which variable density sampling patterns are preferred over uniform random sampling patterns. The work of [1] provides analysis to explain this phenomenon and shows that the strength of variable density sampling patterns is because they combine the linear relationship between the Fourier samples of lowest frequencies and the wavelets of lowest dilation factors, and exploit the benefits of randomness at the expense of log factors in the sampling cardinality.

4 Proofs

Throughout this section, given \( a, b \in \mathbb{R} \), \( a \lesssim b \) denotes \( a \leq C \cdot b \) for some numerical constant \( C \) and \( a \gtrsim b \) denotes \( a \geq C \cdot b \) for some numerical constant \( C \). Given \( x \in \mathbb{C}^N \), \( \text{sgn}(x) \in \mathbb{C}^N \) is such that \( \text{sgn}(x)_j = x_j / |x_j| \) if \( x_j \neq 0 \) and \( \text{sgn}(x)_j = 0 \) otherwise. Also, for \( j \in \mathbb{Z} \), \( e_j \) is the vector whose \( j \)th entry is 1 and is zero elsewhere.

Firstly, a minimizer to (2.1) necessarily exists for any choice of \( \Omega \) (see the appendix for a proof of this) so we will henceforth derive error bounds given a minimizer of (2.1). To prove Theorem 2.1 and Theorem 2.2, we will first derive an error bound on the gradient of the recovered signal, then use the following discrete Poincaré inequality to obtain an error bound on the recovered signal.

Lemma 4.1 (Poincaré inequality, see [13, 2]).

(i) Let \( z \in \mathbb{R}^N \) be such that \( \sum_{j=1}^{N} z_j = 0 \). Then

\[
\|z\|_2 \leq \sqrt{N} \cdot \|z\|_{TV}
\]

(ii) Let \( z \in \mathbb{R}^{N \times N} \) be such that \( \sum_{k=1}^{N} \sum_{j=1}^{N} z_{j,k} = 0 \). Then

\[
\|z\|_2 \leq \|z\|_{TV}.
\]
4.1 Proof of Theorem 2.1

We first recall some definitions from [6].

Definition 4.2. Let $U \in \mathbb{C}^{N \times N}$.

1. The coherence of $U$ is $\mu(U) := \max_{i,j} |U_{i,j}|^2 = \mathcal{O}(N^{-1})$.

2. Given $\Delta \subset \{1, \ldots, N\}$, $\delta \in (0,1)$ and $r \in \mathbb{N}$, $U$ is said to satisfy the weak restricted isometry property (RIP) if for all $v$ supported on $\Delta \cup \Gamma$ with $|\Gamma| \leq r$,
   
   $$(1 - \delta)\|v\|_2^2 \leq \|Uv\|_2^2 \leq (1 + \delta)\|v\|_2^2.$$ 

To prove Theorem 2.1, we will make use of the fact that the discrete Fourier transform satisfies $\mu(U) = 1$ and the following result, which was essentially derived in [6] (see assumptions (i)-(iii) at the start of the proof of Theorem 1.2 in [6]). The actual proof of Theorem 2.1 is similar to the proof of Theorem 1.2 in [6], however, as the minimization problem (2.1) is slightly different from the setup in [6], for completeness, we will repeat many steps of the argument from [6].

Proposition 4.3. [6] Let $\epsilon \in (0,1)$. Let $U \in \mathbb{C}^{N \times N}$ be such that $N^{-1/2}U$ is unitary. Let $m \in \mathbb{N}$ with $0 < m \leq N$. Let $\Lambda \subset \{1, \ldots, N\}$ and let $x_0 \in \mathbb{R}^N$. Let $\Gamma \subset \{1, \ldots, N\}$ be $m$ indices drawn uniformly at random. Let $U_{\Gamma,\Lambda} = m^{-1/2}P_{\Gamma}UP_{\Lambda}$. If

$$m \geq C \cdot (1 + \log(\epsilon^{-1})) \cdot \mu(U) \cdot \log(N) \cdot |\Lambda|$$

for some appropriate numerical constant $C$, then the following holds with probability exceeding $1 - \epsilon$. 

Figure 4: The top row shows the Fourier sampling maps, each indexing 8.7% of the available Fourier samples. The bottom row shows the reconstructed images and their relative errors.
(i) \( \| (U^*_{\Gamma,A} U_{\Gamma,A})^{-1} \|_{2 \to 2} \leq 2 \)

(ii) \( m^{-1/2} \max_{i \in \Lambda'} \| U^*_{\Gamma,A} P_{\Gamma} U e_i \|_2 \leq 1 \)

and there exists \( \rho = U^* P_{\Gamma} w \) such that

(iii) \( \| P_{\Lambda} \rho - \text{sgn}(P_{\Lambda} x_0) \|_2 \leq 1/4 \)

(iv) \( \| P_{\Lambda} \rho \|_\infty \leq 1/4 \)

(v) there exists some numerical constant \( C_0 \) such that \( \| w \|_2 \leq C_0 \sqrt{\| A \|} \).

(vi) \( m^{-1/2} P_{\Gamma} U \) satisfies the weak RIP with respect to \( \Lambda, \delta = 1/4 \) and

\[
 r = \min \left\{ \frac{m}{C (1 + \log(\epsilon^{-1})) \mu(U) \log(N) \log(m) \log^2(s) \cdot s}, \frac{1}{4} \right\}.
\]

**Proof of Theorem 2.7** Firstly, the discrete Fourier transform \( A \) is such that \( N^{1/2} A \) is unitary and \( \mu(A) = 1 \). So, by letting

\[
 U := A, \quad \Gamma := \Omega', \quad \Lambda := \Delta, \quad x_0 := x
\]

in Proposition 4.3 conditions (i)-(vi) of Proposition 4.3 are true with probability exceeding \( 1 - \epsilon \) provided that the number of samples \( m \) is chosen in accordance to (2.2). Let \( z = \hat{x} - x \).

**Step 1:** We will show that

\[
\| P_{\Delta}^z Dz \|_1 \leq \| P_{\Delta}^x Dz \|_1 + \delta \cdot (1 + \sqrt{s}).
\]

To do this, we first demonstrate that \( \| P_{\Delta} Dz \|_2 \) can be controlled by \( \delta \) and \( \| P_{\Delta}^z Dz \|_1 \). By properties (i) and (ii) of Proposition 4.3

\[
\| P_{\Delta} Dz \|_2 = \| (A_{\Omega', \Delta}^* A_{\Omega', \Delta})^{-1} A_{\Omega', \Delta}^* A_{\Omega', \Delta} P_{\Delta} Dz \|_2 \leq 2 \left( \frac{1}{m} \| P_{\Delta} A^* P_{\Omega'} A Dz \|_2 + \frac{1}{m} A_{\Omega', \Delta}^* A_{\Omega', \Delta} P_{\Delta} Dz \|_2 \right) \leq \sqrt{5} \cdot \frac{1}{\sqrt{m}} \| P_{\Omega'} A Dz \|_2 + 2 \max_{j \in \Delta'} \frac{1}{\sqrt{m}} \| A_{\Omega', \Delta}^* A_{\Omega', \Delta} P_{\Delta} A e_j \|_2 \| P_{\Delta} Dz \|_1 \]

\[
= \sqrt{5} \cdot \frac{1}{\sqrt{m}} \| P_{\Omega'} A Dz \|_2 + 2 \| P_{\Delta} Dz \|_1 \leq 4 \sqrt{5} \cdot \delta + 2 \| P_{\Delta} Dz \|_1
\]

In the last line of the above computation, note that for all \( k \neq 0, v_k \cdot (Az)_k = (ADz)_k \) where \( v_k = 1 - e^{2\pi i k/N} \). Also, \( (ADz)_0 = 0 \). Thus, since \( \| P_{\Omega} A z \|_2 \leq 2 \delta \sqrt{m} \) by the enforced constraint in the minimization problem, \( \| P_{\Omega} A Dz \|_2 \leq \| P_{\Omega} A Dz \|_2 = \| (v_k)_{k \in \Omega} \cdot (P_{\Omega} A z) \|_2 \leq 2 \| P_{\Omega} A z \|_2 \leq 4 \delta \sqrt{m} \) since \( |v_k| \leq 2 \) (here means pointwise multiplication).

To bound \( \| P_{\Delta}^z Dz \|_1 \), first observe that since \( \hat{x} \) is a minimizer and

\[
\| D\hat{x} \|_1 \geq \| P_{\Delta}^z Dz \|_1 - 2 \| P_{\Delta}^x Dz \|_1 + \| Dz \|_1 + \text{Re} \left( \langle P_{\Delta} Dz, \text{sgn}(P_{\Delta} Dz) \rangle \right),
\]

we have that

\[
\| P_{\Delta}^z Dz \|_1 \leq 2 \| P_{\Delta}^x Dz \|_1 + \| Dz \|_1 + \text{Re} \left( \langle P_{\Delta} Dz, \text{sgn}(P_{\Delta} Dz) \rangle \right).
\]

By properties (iii)-(v) of the dual vector from Proposition 4.3

\[
| \text{Re} \left( \langle P_{\Delta} Dz, \text{sgn}(P_{\Delta} Dz) \rangle \right) | \leq | \langle P_{\Delta} Dz, \text{sgn}(P_{\Delta} Dz) \rangle - P_{\Delta} \rho \rangle + | \langle P_{\Delta} Dz, P_{\Delta} \rho \rangle | \leq \| P_{\Delta} Dz \|_1 \cdot \frac{1}{4} + \| Dz, \rho \rangle + | \langle P_{\Delta} Dz, P_{\Delta} \rho \rangle | \leq \| P_{\Delta} Dz \|_1 \cdot \frac{1}{4} + | \langle P_{\Omega} A Dz, w \rangle | + | \langle P_{\Delta} Dz \|_1 \cdot \frac{1}{4} \leq \delta \cdot (\sqrt{5} + 4C_0 \cdot \sqrt{s}) + \frac{3}{4} \cdot \| P_{\Delta} Dz \|_1.
\]
where we use that fact that \( \| P_{\Omega'} A D z \|_2 \leq 4 \delta \sqrt{m} \). So,

\[
\| P_{\Delta} D z \|_1 \leq 8 \| P_{\Delta} D x \|_1 + \delta \cdot \left( 4 \sqrt{\delta} + 16 \cdot C_0 \cdot \sqrt{s} \right).
\]

**Step II:** Let \( h = D z \) and partition \( \Delta \) into subsets of at most cardinality \( r \) by letting \( \Delta_1 \) be the indices of the \( r \) largest entries of \( P_{\Delta} h \), \( \Delta_2 \) be the next \( r \) largest entries and so on. Let \( \Delta = \Delta_1 \cup \Delta_2 \). Since \( m^{-1/2} P_{\Omega'} A \) satisfies the weak RIP condition (vi) from Proposition 4.3,

\[
\| P_{\Delta} h \|_2^2 \leq \frac{4}{3} \| A_{\Omega'} P_{\Delta} h \|_2^2
\]

\[
= \frac{4}{3 \sqrt{m}} \left( \langle A_{\Omega'} \tilde{\Delta} P_{\Delta} h, P_{\Omega'} A h \rangle - \langle A_{\Omega'} P_{\Delta} h, P_{\Omega'} A P_{\Delta} h \rangle \right).
\]

By applying the weak RIP condition and since \( \| P_{\Omega'} A h \|_2 \leq 4 \cdot \sqrt{m} \cdot \delta \) (as shown in Step I),

\[
| m^{-1/2} \langle A_{\Omega'} \tilde{\Delta} P_{\Delta} h, P_{\Omega'} A h \rangle | \leq \sqrt{\frac{5}{4}} \cdot \| P_{\Delta} h \|_2 \cdot m^{1/2} \| P_{\Omega'} A h \|_2 \leq 2 \sqrt{\delta} \cdot \| P_{\Delta} h \|_2 \cdot \delta.
\]

Also, using the standard compressed sensing result (see [6], proof of Theorem 1.2) that

\[
\sum_{j \geq 2} \| P_{\Delta_j} h \|_2 \leq \frac{1}{\sqrt{r}} \| P_{\Delta} h \|_1,
\]

one obtains

\[
m^{1/2} \left| \langle A_{\Omega'} \tilde{\Delta} P_{\Delta} h, P_{\Omega'} A P_{\Delta} h \rangle \right| \leq \frac{1}{2 \sqrt{r}} \| P_{\Delta} h \|_2 \| P_{\Delta} h \|_1.
\]

Therefore, by combining (4.1), (4.2) and (4.3),

\[
\| P_{\Delta} h \|_2 \leq \left( \frac{8 \sqrt{\delta}}{4} \cdot \delta + \frac{2 \| P_{\Delta} h \|_1}{3 \sqrt{r}} \right)
\]

and

\[
\| h \|_2 \leq \| P_{\Delta} h \|_2 + \sum_{j \geq 2} \| P_{\Delta_j} h \|_2 \leq \left( \frac{8 \sqrt{\delta}}{4} \cdot \delta + \frac{7 \| P_{\Delta} h \|_1}{6 \sqrt{r}} \right).
\]

Combining with the result of step I and observing that \( r^{-1} \leq 1 \) yields

\[
\| h \|_2 \leq \delta (15 + 19 C_0 \sqrt{s}) + \frac{10 \| P_{\Delta} D x \|_1}{\sqrt{r}}
\]

and

\[
\| h \|_1 \leq \sqrt{\delta} \cdot \| P_{\Delta} h \|_1 + \| P_{\Delta} h \|_1 \leq \sqrt{\delta} \| P_{\Delta} h \|_2 + \| P_{\Delta} h \|_1 \lesssim \left( \delta (1 + C_0 \sqrt{s}) + \sqrt{\frac{s}{r}} \| P_{\Delta} D x \|_1 \right)
\]

**Step III:** To bound \( ||z||_2 \), note that \( \tilde{z} = (\tilde{z}_j)_{j=1}^N \), where \( \tilde{z}_j = z_j - \frac{1}{N} \sum_{j=1}^N z_j \) has mean zero. So, by the Poincaré inequality,

\[
\frac{1}{\sqrt{N}} \| \tilde{z} \|_2 \lesssim \| D z \|_1.
\]

Therefore,

\[
\frac{\| z \|_2}{\sqrt{N}} \lesssim \frac{1}{N} \sum_{j=1}^N \tilde{z}_j \| D z \|_1.
\]

Finally, since we have \( 0 \in \Omega \), we know that \( | \langle A z_0 \rangle | = \sum_{j=1}^N z_j \leq 2 \delta \sqrt{m} \). So, by the bound on \( \| D z \|_1 \) from (4.4), we have that

\[
\frac{\| z \|_2}{\sqrt{N}} \lesssim \frac{\delta \sqrt{m}}{N} + \| D z \|_1 \lesssim \left( \frac{\sqrt{m}}{N} + (1 + C_0 \sqrt{s}) \right) \cdot \delta + \sqrt{\frac{s}{r}} \| P_{\Delta} D x \|_1
\]

10
and recalling \( r \) from (vi) of Proposition 4.3, we have that
\[
\frac{s}{r} \leq \log(m) \log^2(s)
\]
which concludes this proof. \( \square \)

4.1.1 Remark on Corollary 2.2

Given \( x \in \mathbb{R}^{N \times N} \), define the vertical gradient operator as
\[
D_1 : \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N}, \quad x \mapsto (x_{j+1,k} - x_{j,k})_{j=1}^{N}
\]
with \( x_{N+1,k} = x_{1,k} \) for each \( k = 1, \ldots, N \) and the horizontal gradient operator is defined to be
\[
D_2 : \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N}, \quad x \mapsto (x_{j,k+1} - x_{j,k})_{j=1}^{N}
\]
with \( x_{j,N+1} = x_{j,1} \) for each \( j = 1, \ldots, N \). Now define the gradient operator \( D : \mathbb{R}^{N \times N} \to \mathbb{R}^{N \times N} \) as
\[
Dx = D_1 x + iD_2 x,
\]
then the isotropic total variation norm is simply
\[
\|x\|_{TV} = \|Dx\|_1.
\]

The proof of Corollary 2.2 is omitted since it is essentially the same as the proof of Theorem 2.1 - the proof of Theorem 2.1 relies on Proposition 4.3, the coherence of the discrete Fourier transform and the Poincaré inequality - in two dimensions, the proof is almost identical as the two dimensional discrete Fourier transform has coherence \( \mu(A) = 1 \) and we may apply the two dimensional discrete Poincaré inequality in Step III of the above proof.

4.2 Proof of Theorem 2.4

Throughout this section, for \( M \in \mathbb{N} \), let \([M] := \{-M, \ldots, M\}\) and let \( \Delta := \{t_1, \ldots, t_s\} \subseteq \{1, \ldots, N\} \) be such that \( t_1 < t_2 < \ldots < t_s \) and \( \nu_{\min}(\Delta, N) \geq \frac{1}{M} \), where \( \nu_{\min} \) is as defined in Definition 2.3.

The proof of Theorem 2.4 will hinge on showing that the conditions of the following Proposition are satisfied.

Proposition 4.4. Suppose that \( 0 \in \Omega \) and the following conditions hold.

(i) there exists \( L \) such that \( \hat{L} = \frac{1}{m} P_{\Delta} L P_{\Delta} \) is invertible on \( P_{\Delta}(\ell^2(\mathbb{N})) \), \( \|m^{-1/2} P_{\Delta} L P_{\Delta}\|_{2 \to 2} \leq \frac{5}{4} \)
and \( \|L^{-1}\|_{2 \to 2} \leq \frac{4}{3} \).

(ii) \( \max_{j \in \Delta} \| \frac{1}{\sqrt{m}} P_{\Delta} A P_{\Delta} e_j \|_2 \leq 6 \).

(iii) \( \|P_{\Delta}^+ \rho\|_{\infty} \leq c_0 \).

(iv) \( \|P_{\Delta} \rho - \text{sgn}(P_{\Delta} D x)\|_2 \leq c_1 \).

(v) \( \|w\|_2 \leq c_2 \frac{\sqrt{s}}{\sqrt{m}} \)

with constants \( c_0, c_1, c_2 > 0 \) such that \( C_0 = (10c_1 + c_0) < 1 \). Let \( \hat{x} \) be a minimizer of \( 2.1 \). Then
\[
\|D \hat{x} - Dx\|_2 \lesssim (1 - C_0)^{-1} \left( (c_1 + c_2 \sqrt{s}) \cdot \delta + \|P_{\Delta}^+ D x\|_1 \right).
\]
and
\[
\|\hat{x} - x\|_2 \sqrt{N} \lesssim (1 - C_0)^{-1} \left( (c_1 \sqrt{s} + c_2 s) \cdot \delta + \sqrt{s} \|P_{\Delta}^+ D x\|_1 \right).
\]
Proof. Let $z = \hat{x} - x$. We first demonstrate that $\|P_\Delta Dz\|_2$ can be controlled by $\delta$ and $\|P_\Delta^+ Dz\|_1$.

$$\|P_\Delta Dz\|_2 = \|L^{-1}P_\Delta Dz\|_2 \leq \frac{4}{3} \left( \frac{1}{m} \|P_\Delta L P_\Omega A Dz\|_2 + \frac{1}{m} \|P_\Delta L P_\Omega e_j Dz\|_2 \right)$$

$$\leq \frac{4}{3} \left( \frac{1}{m} \|P_\Omega A Dz\|_2 + \max_j \frac{1}{m} \|P_\Omega e_j Dz\|_2 \|P_\Delta Dz\|_1 \right)$$

$$= \frac{5}{3} \left( \frac{1}{m} \|P_\Omega A Dz\|_2 + 6 \|P_\Delta Dz\|_1 \right)$$

$$\leq \frac{20\delta}{3} + 10 \|P_\Delta Dz\|_1$$

In the last line of the above calculation, note that for all $k \neq 0$, $v_k \cdot (Az) = (ADz)_k$ where $v_k = 1 - e^{2\pi ik/N}$. Also, $(ADz)_0 = 0$. Thus, $\|P_\Omega ADz\|_2 = \|v_k\|_{\Omega} \cdot \|P_\Omega Az\|_2 \leq 2 \|P_\Omega Az\|_2 \leq 4\delta \sqrt{m}$ since $|v_k| \leq 2$ (here means pointwise multiplication) and $\|P_\Omega Az\|_2 \leq 2\delta \sqrt{m}$ by the enforced constraint in the minimization problem.

To bound $\|P_\Delta^+ Dz\|_1$, first observe that since $\hat{x}$ is a minimizer and

$$\|D\hat{x}\|_1 \geq \|P_\Delta^+ Dz\|_1 - 2 \|P_\Delta Dz\|_1 + \|Dz\|_1 + \text{Re}(P_\Delta Dz, \text{sgn}(P_\Delta Dz)),$$

we have that

$$\|P_\Delta^+ Dz\|_1 \leq 2 \|P_\Delta Dz\|_1 + |\text{Re}(P_\Delta Dz, \text{sgn}(P_\Delta Dz))|.$$

By properties of the dual vector,

$$|\text{Re}(P_\Delta Dz, \text{sgn}(P_\Delta Dz))| \leq |(P_\Delta Dz, \text{sgn}(P_\Delta Dz) - P_\Delta \rho)| + |(P_\Delta Dz, P_\Delta \rho)|$$

$$\leq \|P_\Delta Dz\|_1 \cdot c_1 + \|(Dz, \rho)\| + \|P_\Delta Dz\|_1 \cdot c_0$$

$$\leq \delta \left( 20c_1 + 4c_2 \cdot \sqrt{s} \right) + \left( 10c_1 + c_0 \right) \cdot \|P_\Delta Dz\|_1$$

where we use that fact that $\|P_\Omega ADz\|_2 \leq 4\delta \sqrt{m}$. So,

$$\|Dz\|_1 \leq \sqrt{s} \|P_\Delta Dz\|_2 + \|P_\Delta^+ Dz\|_1 \leq (1 - C_0)^{-1} \cdot \left( \delta \left( \frac{20c_1 \sqrt{s}}{3} + 4c_2 \cdot s \right) + 2\sqrt{s} \|P_\Delta Dz\|_1 \right)$$

Finally, the conclusion of this proposition follows from Poincaré's inequality.

By an argument in [7], the probability that one of the conditions (i)-(v) of Theorem 2.4 fails for $\Omega$ chosen uniformly at random is up to a constant bounded from above by the probability that one of these conditions fails when $\Omega$ chosen in accordance to a Bernoulli sampling model: $\Omega = \{\delta_j, j = -2M, \ldots, 2M\}$ where $\delta_j$ are independent random variables such that $\mathbb{P}(\delta_j = 1) = q$ and $\mathbb{P}(\delta_j = 0) = 1 - q$, where $q = m/M$ with $m$ as defined in (2.4). We denote such a choice of $\Omega$ by $\Omega \sim \text{Ber}(q, 2M)$. Thus, it suffices to show that the conditions hold choosing $\Omega$ in accordance to this Bernoulli sampling model. For ease of analysis, it has become customary in compressed sensing theory to prove recovery statements for Bernoulli sampling models instead of uniform sampling models. We will assume throughout this section that $\Omega \sim \text{Ber}(q, 2M)$.

Proposition 4.4 presents conditions under which the gradient of $x$ (and hence $x$ by the Poincaré inequality) can be stably recovered by solving (1.2). These conditions are essentially the conditions required for stable recovery of a discrete signal from its low frequency Fourier data, and the proof of Theorem 2.4 will heavily rely on results from the analysis of super-resolution in [5, 4, 18]. We first recall a discrete version of a result from [5] which shows that $P_{2M} A P_\Delta$ has a left inverse.

Lemma 4.5. [5] Let $K_M(t)$ be the squared Fejer kernel

$$K_M(t) = \left( \frac{\sin(\pi M t)}{M \sin(\pi t)} \right)^4 = \frac{1}{M} \sum_{j=-2M}^{2M} g_M(j) e^{-2\pi i j t}$$

12
where
\[
g_M(j) = \frac{1}{M} \min\{j+M, M\} \left( 1 - \frac{k}{M} \right) \left( 1 - \left| j - \frac{k}{M} \right| \right).
\]

Define \( L = (L_{j,k})_{j,k=1}^n \in \mathbb{C}^{n \times s} \) by \( L_{j,k} = K_M(t_j - t_k) \). Then,
1. \( L = \Delta A^*VAP_\Delta \) where \( V \) is a diagonal matrix with diagonal entries \( (g_M(l))^2_{l=-2M} \).
2. \( \|g_M\|_\infty \leq 1 \)
3. \( \|L - I\|_{2 \to 2} \leq 6.253 \times 10^{-3} \) which implies that
   (a) \( L \) is invertible and \( \|L^{-1}\|_{2 \to 2} \leq 0.993747 \)
   (b) \( \|L\|_{2 \to 2} \leq 1 + 6.253 \times 10^{-3} \).

In the following two lemmas, we will prove that conditions (i) and (ii) of Proposition \[4.4\] are satisfied with high probability under the conditions of Theorem \[2.4\] by making use of the matrix Bernstein inequality \[19\] and Lemma \[4.5\].

**Lemma 4.6.** Consider the setup of Lemma \[4.5\]. Let \( \Omega \sim \text{Ber}(q, 2M) \) with \( q = \frac{m}{M} \). For each \( \epsilon \in (0, 1] \), if
\[
m \geq \log \left( \frac{8}{\epsilon} \right) \cdot s.
\]
then with probability exceeding \( 1 - \epsilon \),
(i) \( \tilde{L} = \frac{1}{m} \Delta A^*VAP_\Omega AP_\Delta \) is invertible.
(ii) \( \|\tilde{L}\|_{2 \to 2} \leq \frac{1}{3} \) and \( \|\tilde{L}^{-1}\|_{2 \to 2} \leq \frac{1}{4} \). Note that since \( V \) is a diagonal matrix, it is self-adjoint and
\[
\tilde{L} = \frac{1}{m} \Delta A^*V_{1/2}P_\Omega V_{1/2}AP_\Delta.
\]
So, \( \|\sqrt{m} \Delta A^*V_{1/2}P_\Omega\|_{2 \to 2} = \|\tilde{L}\|_{2 \to 2} \leq \sqrt{\frac{8}{3}} \) and \( \|V_{1/2}\|_{2 \to 2} \leq 1 \)
since \( \|g_M\|_\infty \leq 1 \).

If \( q = 1 \), then (i) and (ii) hold with probability 1.

**Proof.** Firstly, if \( q = 1 \), then \( \tilde{L} = L \) and the result follows from Lemma \[4.5\]. For the case when \( q < 1 \), observe that from Lemma \[4.5\] we have that \( \|L - I\|_{2 \to 2} \leq 6.253 \times 10^{-3} \), which implies that \( \|L\|_{2 \to 2} - 6.253 \times 10^{-3} \leq \|LL^{-1}\|_{2 \to 2} \) and \( \|L^{-1}\|_{2 \to 2} \geq 1 - 6.253 \times 10^{-3} \geq 1 \). Thus, if \( \|L - I\|_{2 \to 2} \leq \frac{1}{4 \cdot 0.99} \), then \( \|\tilde{L} - L\|_{2 \to 2} \leq \frac{1}{4 \cdot 0.99} \) and
1. \( \tilde{L} \) is invertible and \( \|\tilde{L}^{-1}\|_{2 \to 2} \leq \frac{1}{4} \cdot 0.99 \).
2. \( \|\tilde{L}\|_{2 \to 2} \leq \frac{1}{4} + 0.99 \).

So, it suffices to show that with probability exceeding \( 1 - \epsilon \),
\[
\|\tilde{L} - L\|_{2 \to 2} \leq \frac{1}{4 \cdot 0.99}.
\]

We will do so using the matrix Bernstein inequality. Let \( (\delta_j)_{j=-2M}^{2M} \) be independent Bernoulli random variables such that \( \mathbb{P}(\delta_j = 1) = q \) and \( \mathbb{P}(\delta_j = 0) = 1 - q \). Let \( K_M(t) = \frac{1}{m} \sum_{j=-2M}^{2M} \delta_j g_M(j) e^{-2\pi itj} \) and observe that \( (\tilde{L})_{j,k} = K_M(t_j - t_k) \).

Let \( \eta_j = (e^{-2\pi it_1j}, e^{-2\pi it_2j}, \ldots, e^{-2\pi it_sj})^T \). Then
\[
\tilde{L} - L = \sum_{j=-2M}^{2M} \frac{g_M(j)}{m} \delta_j (\eta_j \otimes \eta_j) - \sum_{j=-2M}^{2M} \frac{g_M(j)}{M} (\eta_j \otimes \eta_j) = \sum_{j=-2M}^{2M} \frac{g_M(j)}{M} (\eta_j \otimes \eta_j) = \sum_{j=-2M}^{2M} X_j
\]
where \( X_j = \frac{g_M(j)}{M} \left( \frac{\delta_j}{q} - 1 \right) (\eta_j \otimes \eta_j) \) are independent random self-adjoint matrices of zero mean.

13
1. \[ \|X_j\|_{2 \to 2} \leq \left( \frac{|g_M(j)|}{M} \cdot \frac{1}{q} \right) \|\eta_j\|_2^2 \leq \frac{s}{m} \]

since \(\|\eta_j\|_2^2 = s\) and \(\|g_M(j)\|_\infty \leq 1\). So,

\[ R = \max_{|j| \leq 2M} \|X_j\|_{2 \to 2} \leq \frac{s}{m} \]

2. \[
\sigma^2 = \left\| \sum_{j=-2M}^{2M} E(X_j^2) \right\|_{2 \to 2} = \left\| \sum_{j=-2M}^{2M} \left( \frac{1}{q} - 1 \right) \left( \frac{g_M(j)^2\|\eta_j\|^2}{M^2} \right) (\eta_j \otimes \eta_j) \right\|_{2 \to 2} \\
\leq \left( \frac{1}{q} - 1 \right) \left( \frac{g_M(j)^2}{M} \right) \left\| \sum_{j=-2M}^{2M} \frac{g_M(j)}{M} \right\|_{2 \to 2} \leq \left( \frac{1}{q} - 1 \right) \left( \frac{s}{M} \right) \|L\|_{2 \to 2}
\]

since \(\|g_M\|_\infty \leq 1\) and \(\|\eta_j\|_2^2 = 2\). Finally, because \(\|L\|_{2 \to 2} \leq 1.1\), it follows that

\[ \sigma^2 \leq 1.1 \left( \frac{1}{q} - 1 \right) \left( \frac{s}{M} \right). \]

Let \(\gamma = \frac{1}{4 \cdot 0.99}\). Then by matrix Bernstein,

\[ \Pr \left( \|L - L\|_{2 \to 2} \geq \gamma \right) \leq s \exp \left( -\frac{\gamma^2/2}{\sigma^2 + R\gamma/3} \right) \leq \epsilon \]

provided that

\[ \log \left( \frac{s}{\epsilon} \right) \cdot (1.1 + \frac{\gamma}{2}) \cdot \frac{s}{M} \leq \frac{\gamma^2}{2} \]

\[ \square \]

**Lemma 4.7.** Let \(\epsilon \in (0, 1]\) and suppose that \(\Omega \sim \text{Ber}(q, 2M)\) with \(q = m/M\) and

\[ m \geq \log \left( \frac{N}{\epsilon} \right). \]

Then,

\[ \Pr \left( \max_{j=1}^N \| \frac{1}{m} P_{(j)}^* A^* P_{\Omega} A P_{(j)} \|_{2 \to 2} \geq 5.5 \right) \leq \epsilon. \]

If \(q = 1\), then \(\max_{j=1}^N \| \frac{1}{m} P_{(j)}^* A^* P_{\Omega} A P_{(j)} \|_{2 \to 2} \leq 5.5\) holds with probability 1.

**Proof.** Firstly, if \(q = 1\), then for each \(j = 1, \ldots, N\),

\[ \| \frac{1}{m} P_{(j)}^* A^* P_{2M} A P_{(j)} \|_{2 \to 2} = \frac{4M + 1}{M} \leq 5 \]

and so

\[ \Pr \left( \max_{j=1}^N \| \frac{1}{m} P_{(j)}^* A^* P_{\Omega} A P_{(j)} \|_{2 \to 2} \geq 5.5 \right) = 0. \]

For \(q < 1\), let \((\delta_k)_{k=-2M}^{2M}\) be independent Bernoulli random variables such that \(\Pr(\delta_j = 1) = q\) and \(\Pr(\delta_j = 0) = 1 - q\). Then

\[ \frac{1}{m} P_{(j)}^* A^* P_{\Omega} A P_{(j)} - \frac{1}{M} P_{(j)}^* A^* P_{2M} A P_{(j)} = \sum_{k=-2M}^{2M} \frac{\delta_k}{m} - \sum_{k=-2M}^{2M} \frac{1}{M} \]

\[ = \sum_{k=-2M}^{2M} \frac{1}{M} \left( \frac{\delta_k}{q} - 1 \right) = \sum_{k=-2M}^{2M} X_k \]
where \( X_k = \frac{1}{M} (\Delta k - 1) \) are independent random variables of zero mean. So, combining with \( 4.5 \) gives that for each \( j = 1, \ldots, N \),

\[
P\left(\left\| \frac{1}{qM} P(j) A^* P \Omega A P(j) \|_{2 \to 2} \geq 5.5\right\|\right) \leq P\left(\left\| \sum_{k=-M}^{M} X_k \right\| \geq 0.5\right)
\]

We will apply Bernstein’s inequality \([8]\) to bound the right hand side of the above inequality. Observe that \( |X_k| \leq \frac{5}{M} = \frac{1}{m} \) and

\[
\sum_{k=-M}^{M} \mathbb{E}(X_k^2) = \frac{2M + 1}{M^2} \cdot \left(\frac{1}{q} - 1\right) \leq \frac{3}{m}.
\]

Therefore,

\[
P\left(\left\| \sum_{k=-M}^{M} X_k \right\| \geq 0.5\right) \leq 2 \exp\left(-\frac{1/8}{3/m + 1/6m}\right).
\]

So, by applying the union bound, the conclusion follows.

\[
\square
\]

We will now show that there exists \( \rho = A^* P \Omega v \) satisfying conditions (iii) - (iv) in Proposition 4.4 with \( c_1 = 0, c_2 = 1 \) and some constant \( c_0 < 1 \). This dual certificate is actually identical to a discrete version of the dual certificate constructed in \([18]\). We simply recall a few results from \([18]\) and provide a bound for \( \|w\|_2 \).

**Lemma 4.8.** \([18, \text{Section 4.3}]\) Let \( (\delta_k)_{k=\pm 2M}^{2M} \) be independent Bernoulli random variables such that \( P(\delta_j = 1) = q \) and \( P(\delta_j = 0) = 1 - q \) with \( q = m/M \). Let \( \Delta = \{t_1, \ldots, t_s\} \) and let \( x \in \mathbb{R}^N \). If \( \nu_{\min}(\Delta, N) \geq \frac{1}{M} \) and

\[
m \geq \max\left\{ \log^2\left(\frac{M}{\epsilon}\right), s \cdot \log\left(\frac{2}{\epsilon}\right) \cdot \log\left(\frac{M}{\epsilon}\right) \right\},
\]

then with probability exceeding \( 1 - \epsilon \), there exists constants \( (\alpha_k)_{k=1}^{s}, (\beta_k)_{k=1}^{s} \in \mathbb{R}^s \) such that the trigonometric polynomial

\[
Q(t) = \sum_{k=1}^{s} \alpha_k K_M(t - t_k) + \sum_{k=1}^{s} \beta_k K_M'(t - t_k)
\]

with \( K_M(t) = \sum_{|j| \leq 2M} \delta_j g_M(j)e^{-2\pi i tj/N} \) and \( K_M' \) is the first derivative of \( K_M \), satisfies the following.

1. \( Q(t) = (\text{sgn}(P \Delta x))_t \) for each \( t \in \Delta \),
2. \( |Q(t)| < 1 \) for all \( t \notin \Delta \)
3. \( \sqrt{\sum_{k=1}^{s} |\alpha_k|^2 + \sum_{k=1}^{s} |K_M''(0)||\beta_k|^2} \leq 2\sqrt{s} \cdot q^{-1} \cdot 1.568 \)

where \( K_M'' \) is the second derivative of the squared Fejer kernel and \( K_M''(0) = -4\pi^2 (M^2 - 1)/3 \). Furthermore, for each \( k \in \{1, \ldots, N\} \setminus \Delta \),

\[
|Q(k)| \leq \max\left\{ 1 - \frac{0.92(M^2 - 1)}{N^2}, 0.99993 \right\}.
\]

**Remark 4.1** This lemma is essentially proved in \([18]\). The inequality \((4.6)\) is a result of combining the result of Proposition 4.12 and equation \((4.35)\) from \([18]\).

**Lemma 4.9.** Let \( E \) be the event that conditions (i)-(v) of Proposition 4.4 are satisfied, with

\[
c_0 = \max\left\{ 1 - \frac{0.92(M^2 - 1)}{N^2}, 0.99993 \right\}, \quad c_1 = 0, \quad c_2 = 1
\]

and \( L = A^* V \) where \( V \) is as defined Lemma 4.3. Suppose that \( M \geq 10 \) and

\[
m \geq \max\left\{ \log^2\left(\frac{M}{\epsilon}\right), s \cdot \log\left(\frac{2}{\epsilon}\right) \cdot \log\left(\frac{M}{\epsilon}\right) \right\},
\]

then \( \mathbb{P}(E) > 1 - \epsilon \).
Proof. Let $\rho = (Q(j))_{j=1}^N$ where $Q$ is as defined in Lemma 4.8. We first demonstrate that if condition (i) of Proposition 4.4 holds, then $\rho = \Lambda^* \Omega \omega$ with $\|\omega\|_2 \leq \frac{\sqrt{s}}{\sqrt{m}}$: By definition, for each $j = 1, \ldots, N$,

$$
\rho_j = \frac{1}{M} \sum_{|l| \leq 2M} \delta g_M(l) e^{2\pi i jl/N} \sum_{k=1}^s \alpha_k e^{2\pi i tkj/N} + \frac{1}{M} \sum_{|l| \leq 2M} (-2\pi il) \delta g_M(l) e^{2\pi i jl/N} \sum_{k=1}^s \beta_k e^{2\pi i tkj/N}
$$

$$
= \sum_{l \in \Omega} w_l e^{2\pi i jl/N}
$$

where

$$
w_l := \frac{g_M(l)}{M} \left( \sum_{k=1}^s \alpha_k e^{2\pi i tkj/N} - 2\pi il \sum_{k=1}^s \beta_k e^{2\pi i tkj/N} \right).
$$

$$
\frac{1}{q} \|\Omega \omega\|_2^2 = \sum_{j \in \Omega} \frac{1}{q} |w_j|^2 \leq \frac{2}{q} \sum_{j \in \Omega} \left| \frac{g_M(j)}{M} \sum_{k=1}^s \alpha_k e^{2\pi i tkj/N} \right|^2 + \frac{2}{q} \sum_{j \in \Omega} \left| \frac{g_M(j)}{M} \cdot 4\pi^2 |j|^2 \sum_{k=1}^s \beta_k e^{2\pi i tkj/N} \right|^2
$$

if we assume that $M \geq 10$. Furthermore, by definition of $\tilde{L}$ and since we are assuming condition (i) of Proposition 4.4,

$$
\frac{1}{q} \|\Omega \omega\|_2^2 \leq 2 \left\| g_M \right\|_\infty (\tilde{L}, \alpha) + 2 \left\| g_M \right\|_\infty (\tilde{L}, \alpha) \cdot 14 |K_M'(0)| \cdot \left\| \beta \right\|_2^2
$$

$$
\leq 28 \frac{M}{q} \left\| \tilde{L} \right\|_{2 \rightarrow 2} \left( \left\| \alpha \right\|_2^2 + |K_M'(0)| \left\| \beta \right\|_2^2 \right).
$$

Therefore,

$$
\|\Omega \omega\|_2^2 \lesssim \sqrt{\frac{q}{M}} \cdot \sqrt{s} = \sqrt{s}
$$

Thus, to show that $\mathbb{P}(E) > 1 - \epsilon$, it suffices to show that

$$
\mathbb{P}(E_1^c) \leq \epsilon/3, \quad \mathbb{P}(E_2^c) \leq \epsilon/3, \quad \mathbb{P}(E_3^c) \leq \epsilon/3
$$

(4.7)

where $E_1$ is the event that (i) holds, $E_2$ is the event that (ii) holds and $E_3$ is the event that $\rho$ satisfies both conditions (iii) and (iv). Observe that if (4.7) are satisfied for the index set $\Omega'$, then they are satisfied for the index set $\Omega$ since $\Omega \supseteq \Omega'$. So, since $\Omega'$ is chosen uniformly at random in accordance to (2.4), $\mathbb{P}(E_1^c) \leq \epsilon/3$ follows from Lemma 4.6, $\mathbb{P}(E_2^c) \leq \epsilon/3$ follows from Lemma 4.7 and $\mathbb{P}(E_3^c) \leq \epsilon/3$ follows from Lemma 4.8. 

5 Concluding remarks

We showed that in the reconstruction of approximately gradient sparse signals of length $N$, one is guaranteed stable and robust reconstructions from $\mathcal{O}(s \log N)$ samples drawn uniformly at random where $s$ is the approximate sparsity. In some sense, a uniform random sampling pattern is optimal, since $\mathcal{O}(s \log N)$ is the information theoretic limit for $s$-sparse signals (see [7]). However, in practice, uniform random sampling fares poorly when compared with sampling patterns which concentrate on low frequencies. Although the recovery statement for the reconstructions of signals whose gradient sparsity pattern exhibits sufficient separation provided some initial insights as to why this is the case, the link between the sparsity structure and the choice of the sampling pattern is a delicate issue which requires more detailed analysis.

6 Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/H023348/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis. The author would also like to thank Ben Adcock and Anders Hansen for useful discussions.
Appendix

A Existence of minimizers

We prove here that a minimizer to (1.2) exists. The proof that a minimizer exists in the two dimensional case is almost identical and hence omitted.

Lemma A.1. A minimizer to (1.2) exists.

Proof. For \( u \in \mathbb{C}^{\mathbb{N}} \), let \( \text{mean}(u) = \frac{1}{N} \sum_{j=1}^{N} u_j \). Let

\[
F = \{ u \in \mathbb{C}^{\mathbb{N}} : \| P_{\Omega} A u - y \|_2 \leq \delta \}.
\]

Case I:
Suppose that \( 0 \in \Omega \). Observe that given any \( u \in \mathbb{C}^{\mathbb{N}} \),

\[
\| P_{\Omega} A u - y \|_2 \geq |\langle P_{\Omega} A u - y, e_0 \rangle| = | \sqrt{N} \cdot \text{mean}(u) - \langle y, e_0 \rangle |.
\]

So, \( u \in F \) implies that \( \sqrt{N} \cdot |\text{mean}(u)| \leq | \langle y, e_0 \rangle | + \delta \). (A.1)

Also, we have as a corollary of the standard Poincaré inequality \([2, 14]\) that for all \( u \in \mathbb{C}^{\mathbb{N}} \),

\[
\| u - \text{mean}(u) \|_{[\mathbb{N}]} \|_2 \lesssim \| u \|_{TV}.
\]

(A.2)

Let \( I \) denote the minimum of (1.2) and let \( (u_n)_{n \in \mathbb{N}} \subset F \) be a minimizing sequence, so \( \lim_{n \to \infty} \| u_n \|_{TV} = I \). Then, by applying (A.1) and (A.2),

\[
\| u_n \|_2 \leq \| u_n - \text{mean}(u_n) \|_{[\mathbb{N}]} \|_2 + \left| \text{mean}(u_n) \right| \cdot \sqrt{N} \lesssim \| u_n \|_{TV} + | \langle y, e_0 \rangle | + \delta.
\]

So, \( (u_n)_{n \in \mathbb{N}} \) is a bounded sequence, and consequently, there exists a subsequence \( (u_{n_k})_{k \in \mathbb{N}} \) and \( u \in \mathbb{C}^{\mathbb{N}} \) such that

\[
\| u_{n_k} - u \|_2 \to 0, \quad k \to \infty.
\]

Since \( F \) is closed, \( u \in F \). Also, by continuity of \( \| \cdot \|_{TV} \),

\[
I = \lim_{n \to \infty} \| u_n \|_{TV} = \lim_{k \to \infty} \| u_{n_k} \|_{TV} = \| u \|_{TV} \geq I
\]

and so \( u \) is a minimizer.

Case II:
Suppose that \( 0 \notin \Omega \). Then, given any \( C \in \mathbb{C} \), \( P_{\Omega} A u = P_{\Omega} A (u + C \mathbb{1}_{[\mathbb{N}]} ) \). So, given any minimizing sequence \( (u_n)_{n \in \mathbb{N}} \subset F \),

\[
\| u_n \|_{TV} = \| u_n - \text{mean}(u_n) \mathbb{1}_{[\mathbb{N}]} \|_{TV}, \quad u_n - \text{mean}(u_n) \mathbb{1}_{[\mathbb{N}]} \in F.
\]

Thus, without loss of generality, we may assume that \( \text{mean}(u_n) = 0 \). We can now proceed as in Case I to show that any minimizing sequence \( (u_n)_{n \in \mathbb{N}} \) is bounded and hence a minimizer exists.

\( \square \)
References

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

[2] L. Ambrosio, N. Fusco, and D. Pallara. Functions of bounded variation and free discontinuity problems, volume 254. Clarendon Press Oxford, 2000.

[3] M. Benning, L. Gladden, D. Holland, C.-B. Schönlieb, and T. Valkonen. Phase reconstruction from velocity-encoded mri measurements—a survey of sparsity-promoting variational approaches. Journal of Magnetic Resonance, 238:26–43, 2014.

[4] E. J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. Journal of Fourier Analysis and Applications, 19(6):1229–1254, 2013.

[5] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. Communications on Pure and Applied Mathematics, 67(6):906–956, 2014.

[6] E. J. Candès and Y. Plan. A probabilistic and ripless theory of compressed sensing. Information Theory, IEEE Transactions on, 57(11):7235–7254, 2011.

[7] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. Information Theory, IEEE Transactions on, 52(2):489–509, 2006.

[8] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. Springer, 2013.

[9] T. Goldstein and S. Osher. The split bregman method for l1-regularized problems. SIAM Journal on Imaging Sciences, 2(2):323–343, 2009.

[10] P. E. Z. Larson, S. Hu, M. Lustig, A. B. Kerr, S. J. Nelson, J. Kurhanewicz, J. M. Pauly, and D. B. Vigneron. Fast dynamic 3D MR spectroscopic imaging with compressed sensing and multiband excitation pulses for hyperpolarized 13c studies. Magn. Reson. Med., 2010.

[11] R. J. LeVeque. Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems, volume 98. Siam, 2007.

[12] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed Sensing MRI. IEEE Signal Process. Mag., 25(2):72–82, March 2008.

[13] D. Needell and R. Ward. Near-optimal compressed sensing guarantees for total variation minimization. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society, 22(10):3941, 2013.

[14] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. SIAM Journal on Imaging Sciences, 6(2):1035–1058, 2013.

[15] C. Poon. A consistent and stable approach to generalized sampling. Journal of Fourier Analysis and Applications, pages 1–35, 2014.

[16] C. Poon. On the role of total variation in compressed sensing - structure dependence. Preprint, 2014.

[17] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. Signal Processing Letters, IEEE, 18(10):595–598, 2011.

[18] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressive sensing off the grid. In Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on, pages 778–785. IEEE, 2012.

[19] J. A. Tropp. User-friendly tail bounds for sums of random matrices. Foundations of Computational Mathematics, 12(4):389–434, 2012.

[20] R. Ward and F. Krahmer. Stable and robust sampling strategies for compressive imaging. 2013.